He brings hope to Parkinson's patients. Now he is one p. 448

Ethics of upstream choices affect downstream vaccine inequities p. 460

Guaranteed arrival over tricky terrain p. 509

Science

$15
5 MAY 2023
SPECIAL ISSUE
science.org
AAAS

AUTOIMMUNITY

The dark side of host defense p. 468

WHEN A DELICATE BALANCING ACT GOES WRONG

By **Gemma Alderton** and **Seth Thomas Scanlon**

A cardinal feature of the immune system is its ability to distinguish self from nonself. Although many early immunologists thought that its powerful defenses could rarely, if ever, be turned against the host, pioneering research on autoimmune diseases beginning in the early 1900s has documented a different reality. More than 80 different autoimmune disorders have now been described that may affect up to 5% of the population. Autoimmune diseases can be hereditary, triggered by infection and other environmental factors, or a mixture of the two. Practically any molecule, cell, tissue, or organ in the body can be a target. The immune system, it turns out, has a dark side.

This special issue surveys current progress in our understanding of autoimmunity and the regulation of immune tolerance. Recent work has elucidated how T cell tolerance is maintained in the periphery, how inborn errors of the innate immune system's nucleic acid–sensing machinery can trigger autoimmune disease, how autoimmune disease genetics can inform the development of future therapies, and how autologous hematopoietic stem cell transplantation can "reset" the immune system and treat autoimmune disease.

Beyond spurring major advances in the fields of antiviral immunity, virology, and vaccinology, the COVID-19 pandemic continues to teach scientists about how the immune system's delicate balancing act is regulated and how it sometimes goes awry. Autoantibodies against components of the antiviral immune response, for example, may underlie certain forms of severe disease. Furthermore, Long Covid is increasingly appreciated as a spectrum of virally induced autoimmune disorders. It is therefore contingent on us to better understand how the immune system is regulated and how a combination of inherited and outside forces can tip the scales.

10.1126/science.adi2318

SPECIAL SECTION

PERSPECTIVE

Resetting tolerance in autoimmune disease p. 470

REVIEWS

Spatiotemporal regulation of peripheral T cell tolerance p. 472

Innate virus-sensing pathways in B cell systemic autoimmunity p. 478

Common genetic factors among autoimmune diseases p. 485

Antibodies that target self-antigens are an important component of certain autoimmune diseases and are sometimes used as a clinical marker for these syndromes.

ILLUSTRATION: STEPHAN SCHMITZ/FOLIOART

PERSPECTIVE

Resetting tolerance in autoimmune disease

Autologous hematopoietic stem cell transplantation is effective, but mechanisms are elusive

By **Paolo A. Muraro**[1,2]

Autoimmune disorders affect the lives of millions of people worldwide. More than 80 autoimmune diseases have been described, targeting virtually any organ in the body with wide-ranging severity. The common denominator of these disorders is a deviant immune response attacking the body and causing damage (pathogenic autoimmunity). Generally, the cause of autoimmunity is unknown, and the mechanisms of the disease processes are incompletely understood. Immune-modifying treatments, which may consist of chemical or biological immunosuppressive or immune-modulatory drugs, are variably effective and are usually required long-term, with the attendant risks, burdens, and costs. An alternative strategy for the treatment of people with severe forms of autoimmune disease is autologous (the patient's own) hematopoietic stem cell transplantation (AHSCT), which has been adapted from the treatment of hematological malignancies. The clinical experience and immunological knowledge on AHSCT in autoimmune disease are expanding, yet the question remains: How does the treatment actually work?

Efforts at preventing and treating autoimmune diseases face difficult challenges: Etiologies are unknown or poorly understood, knowledge of pathophysiology is incomplete, genetic susceptibility to disease is complex and multigenic, environmental agents (infectious, chemical, physical, and behavioral) may be ubiquitous, latency between exposure to the causative agents and clinical onset may be long, and heterogeneity of disease phenotypes and variability of disease course complicate prognosis and treatment decisions. Nonetheless, treatment of autoimmune disease has made substantial progress and has evolved from immunosuppressant drugs (e.g., corticosteroid and cytotoxic agents) toward more-selective agents, such as antigen-specific tolerization or targeting a single molecule or pathway involved in the disease process. These targeted strategies have the advantage of fewer off-target effects, usually resulting in high tolerability and safety. Efficacy is predicated on the role of the target molecule in disease pathogenesis and is limited by the complexity and variability of autoimmune disease. In practice, targeted treatment strategies are only effective in a limited proportion of patients. For example, in the most common autoimmune disease of the central nervous system, multiple sclerosis (MS), the reported rates of efficacy of standard therapies at achieving complete control of disease [i.e., no evidence of disease activity (NEDA)] range between 20 and 50% (*1*).

A contrasting treatment strategy, AHSCT, is nonspecific by design: HSCT was originally developed to eradicate cancer cells of lymphoid (leukemia and lymphoma) or myeloid (myeloma) lineages regardless of their specific aberrations by infusing donor nonmalignant stem cells after ablative chemotherapy. This lack of specificity suits the treatment of autoimmune disease owing to the uncertainty on causation, individual variability, and redundancy of immune pathways. In the absence of clarity about which immune components are responsible for perpetuating the disease, unselective lymphoid and potentially myeloid ablation is an advantage. Accordingly, the broad spectrum of action of AHSCT translates to high efficacy but comes with the disadvantage of considerably higher off-target effects caused by the chemo- and immunotherapy. The risks are front-loaded, being highest in the weeks after the conditioning regimen, and include organ toxicities and infection, causing ~1% estimated treatment-related mortality in MS (*1*).

AHSCT is a multistep treatment procedure achieved in two main hospital-based stages. The first stage, usually daily hospital attendance over a few days, consists of treatments for the mobilization of hematopoietic progenitor or stem cells (HPCs) and their collection from the blood (see the figure). The hematopoietic product is frozen for subsequent use. After a few weeks of recovery, the second stage involves hospital admission for the administration of the conditioning treatment with chemotherapy and biological therapy to ablate all lymphocytes, including pathogenic cells. This step is immediately followed by the reinfusion of the hematopoietic product to initiate the regeneration of blood and immune cells. After these treatments and specialized supportive inpatient care, the patient's blood and immune cells reconstitute their numbers and reestablish immune competence. The treatment procedure is not disease specific, and the same or similar protocols have been applied across different autoimmune diseases. AHSCT protocols can be classified on the basis of their overall intensity and suppressive effects on the bone marrow. High-, intermediate-, and lower-intensity conditioning regimens are associated with a gradient of myeloablative to nonmyeloablative effects (*1*).

AHSCT is highly effective, as exemplified in MS by NEDA between 70 and 90% (*1*). Although higher efficacy and risk of adverse effects could be expected from high-intensity regimens, comparative data are lacking, and the optimal treatment intensity remains unknown. Protocols may also differ in various steps within the treatment procedure: For example, HPC mobilization to the blood can be achieved with chemotherapy or growth factors alone; the blood product can be manipulated to enrich the proportion of HPCs, or may remain unselected; the ablative conditioning can be achieved with different chemotherapeutic schemes; and antibody therapy (antithymocyte globulin or monoclonal antibodies) can be administered during or after the conditioning regimen to deplete in vivo any mature T and B cells, potentially harboring autoimmune disease–mediating cells that might be reinfused with the autologous hematopoietic product.

The autoimmune diseases that have been most often treated with AHSCT are MS, systemic sclerosis, Crohn's disease, and systemic lupus erythematosus (SLE) (*2*). For some autoimmune diseases, such as rheumatoid arthritis, biological therapies have radically improved prognosis, enabling maintenance of joint function and quality of life, and so AHSCT development was stopped. Biological therapies also benefit a large proportion of people with MS. However, for people with aggressive forms of MS, failure of standard therapy signifies high risk of poor prognosis, and AHSCT use has continued to expand (*2*).

The rationale for AHSCT is to purge with generalized immune ablation the autoimmune disease–mediating components and enable immune reconstitution with the expectation that the alterations causing autoimmune disease are not immediately, or ideally ever, reproduced. The notion of immune resetting or rebooting has been proposed in analogy with restarting a computer when it is malfunctioning, which often restores working order if the problem is not caused

[1]Department of Brain Sciences, Imperial College London, London, UK. [2]Neurology, Imperial College London NHS Trust, London, UK. Email: p.muraro@imperial.ac.uk

by a hardware failure. This analogy is useful to conceptualize a theory on how AHSCT can radically ameliorate, and in some cases cure, autoimmune disease even though the immune system is regenerated starting from autologous HPCs and not HPCs from a healthy donor. Efficacy of an autologous transplant assumes that the disease is neither strictly genetic, such as in chromosomal or monogenic diseases, nor transmissible, such as in infectious disease. These assumptions are met in autoimmune diseases, which are best explained by a multifactorial causation to which polygenic susceptibility and environmental exposures contribute. Because the cause of autoimmune diseases is unknown, studying the immune system changes that are associated with disease remission after AHSCT is an opportunity to reverse engineer knowledge of causative immune abnormalities.

AHSCT induces substantial and long-lasting changes to the immune system. Because underlying abnormalities of adaptive immune responses have been linked to autoimmune disease, this has received the most attention. A few studies have examined two broad hypotheses to explain the mechanism of action of AHSCT: Lymphoid ablation allows destruction of mature lymphocytes, regardless of their specificity, or regeneration of a new immune system overrides any memory of the previous aberrant immune activation (i.e., resetting). The experimental evidence required to support these hypotheses should include, at minimum, ablation of the old adaptive immunity, including the presumed pathogenic components; repopulation with new immune repertoire; and reinstatement or enhancement of immune tolerance. A recent study of T cell reconstitution after AHSCT for treatment of MS has shown that although some effector memory cells initially survive the ablative procedure, the naive T cell repertoire entirely renews and eventually expands in the circulation (*3*). These data confirm and extend earlier work demonstrating expansion of naive, thymus-derived T cells after AHSCT in MS (*4*), SLE (*5*), and systemic sclerosis (*6*).

Although autoimmune disease–mediating cells are not precisely defined, studies have examined the best-known candidate mediators. Proinflammatory effector molecules in MS include interferon-γ (IFN-γ) and interleukin-17 (IL-17), and both $CD4^+$ T helper 17 (T_H17) cells (*7*) and cytotoxic IFN-γ– and IL-17–producing $CD8^+$ T cells (*8*) are substantially depleted after AHSCT. To estimate the overall degree of ablation of the preexisting T cell immunity and the extent of renewal after AHSCT, T cell receptor sequencing analysis has provided evidence of extensive T cell removal and subsequent clonal renewal in MS not only in the blood (*9*) but also, and perhaps more relevantly, in the cerebrospinal fluid (*10*).

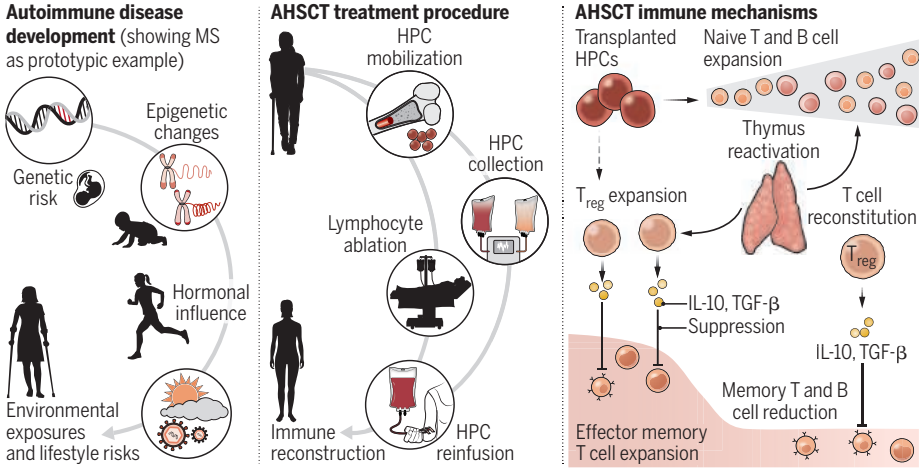
B cell reconstitution after AHSCT has been studied in systemic sclerosis (*11*), SLE (*5*), and MS (*12*). These studies converge in demonstrating regeneration and expansion of the B cell transitional and naive subsets. Disappearance of autoantibodies alongside protective antibodies suggests depletion of plasma cells (*5*), which is desirable in antibody-mediated autoimmune disease but requires revaccination against pathogens. It would be of interest to compare the B cell resetting effects and long-term clinical outcomes of AHSCT with those from autologous CD19 chimeric antigen receptor (CAR)–T cell therapy, which also achieves remission of autoimmunity in SLE and is well tolerated (*13*).

Improvement in immune regulation helps explain how immune tolerance is maintained after immune reconstitution. An increase in regulatory T cell receptor diversity, a marker of regeneration, was demonstrated after AHSCT-induced disease remission in three patients with juvenile rheumatological disease (*14*). Increased frequency of CD56-bright regulatory natural killer (NK) cells with capacity to suppress proinflammatory T_H17 and T_H1 cells has been shown in people with MS after undergoing AHSCT (*15*). Overall, the evidence indicating depletion of proinflammatory cells, T and B cell renewal, diversification of the clonal repertoire, and enhanced immune regulation after AHSCT provides important clues to the possible mechanism of action. Demonstrations of whether these immune effects are relevant to clinical outcomes, however, are sparse.

Key challenges in clinical development remain the optimal selection of patients and derisking the treatment procedure. Better understanding of which immune effects after AHSCT are essential to treatment efficacy has the potential to optimize or even revolutionize treatment protocols. Ongoing randomized controlled clinical trials comparing AHSCT with standard therapy (BEAT-MS, NCT04047628; and STAR-MS, EudraCT, 2019-001549-42) include mechanistic work packages to examine correlations of immune measurements with clinical trial end points, which are expected to provide critical answers. ■

Resetting the autoimmune cycle

Autoimmunity is initiated by a combination of genetic susceptibility, epigenetic changes, and effects of environmental exposures on the immune system. Autologous hematopoietic stem cell transplantation (AHSCT) purges autoimmune memory, inducing disease remission and potentially resetting the immune system.



HPC, hematopoietic progenitor or stem cell; IL-10, interleukin-10; MS, multiple sclerosis; TGF-β, transforming growth factor–β; T_{reg}, regulatory T cell.

GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. P. A. Muraro *et al.*, *Nat. Rev. Neurol.* **13**, 391 (2017).
2. T. Alexander, R. Greco, *Bone Marrow Transplant.* **57**, 1055 (2022).
3. J. Ruder *et al.*, *Sci. Transl. Med.* **14**, eabq1693 (2022).
4. P. A. Muraro *et al.*, *J. Exp. Med.* **201**, 805 (2005).
5. T. Alexander *et al.*, *Blood* **113**, 214 (2009).
6. L. C. M. Arruda *et al.*, *Blood Adv.* **2**, 126 (2018).
7. P. J. Darlington *et al.*, *Ann. Neurol.* **73**, 341 (2013).
8. S. V. Abrahamsson *et al.*, *Brain* **136**, 2888 (2013).
9. P. A. Muraro *et al.*, *J. Clin. Invest.* **124**, 1168 (2014).
10. K. M. Harris *et al.*, *JCI Insight* **5**, e127655 (2020).
11. J. Z. Adamska *et al.*, *Ann. Rheum. Dis.* **82**, 357 (2023).
12. V. von Niederhäusern *et al.*, *Neurol. Neuroimmunol. Neuroinflamm.* **9**, e200027 (2022).
13. A. Mackensen *et al.*, *Nat. Med.* **28**, 2124 (2022).
14. E. M. Delemarre *et al.*, *Blood* **127**, 91 (2016).
15. P. J. Darlington *et al.*, *Front. Immunol.* **9**, 834 (2018).

ACKNOWLEDGMENTS

P.A.M. is supported by the National Institute for Health and Care Research (NIHR; EME project 16/126/26), by the National Institute of Allergy and Infectious Diseases of the NIH under award no. UM1AI109565, and by the NIHR Biomedical Research Centre funding scheme to Imperial College London. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH or NIHR. P.A.M. consults for Magenta Therapeutics, Jasper Therapeutics, and Rubius Therapeutics.

10.1126/science.adg7489

REVIEW

Spatiotemporal regulation of peripheral T cell tolerance

Chrysothemis C. Brown[1,2,3]* and Alexander Y. Rudensky[2,4,5]*

The incomplete removal of T cells that are reactive against self-proteins during their differentiation in the thymus requires mechanisms of tolerance that prevent their effector function within the periphery. A further challenge is imposed by the need to establish tolerance to the holobiont self, which comprises a highly complex community of commensal microorganisms. Here, we review recent advances in the investigation of peripheral T cell tolerance, focusing on new insights into mechanisms of tolerance to the gut microbiota, including tolerogenic antigen-presenting cell types and immunomodulatory lymphocytes, and their layered ontogeny that underlies developmental windows for establishing intestinal tolerance. While emphasizing the intestine as a model tissue for studying peripheral T cell tolerance, we highlight overlapping and distinct pathways that underlie tolerance to self-antigens versus commensal antigens within a broader framework for immune tolerance.

Recognition of "altered self"—a fundamental evolutionary innovation of the vertebrate immune system—is based on the binding of T cell antigen receptors (TCRs) to foreign peptides displayed on self–major histocompatibility complex (MHC) molecules. An essential property of this sensing system is the ability to recognize and respond to a diverse array of rapidly evolving intra- and extracellular pathogens while preventing potentially fatal responses against peptides derived from self-proteins. The efficiency of this response against pathogens—afforded by an immense repertoire of clonally distributed TCRs, randomly generated through somatic recombination in developing T cells—demands an equally effective solution for preventing T cell responses directed against both self-antigens and harmless foreign antigens. Tolerance to self-antigens is first established within the thymus, the site where T cells are generated. Thymic (central) tolerance operates as a "first-pass" system, removing self-reactive T cells from the conventional T (T_{conv}) cell repertoire or directing their differentiation into functionally unresponsive cell states or specialized immunomodulatory lineages before their egress into the periphery. Here, the conundrum is one of representation—how to ensure that the entire spectrum of self-antigens, including developmentally restricted or inducible antigens, is displayed to developing thymocytes, a prerequisite for efficient establishment of tolerance. However, negative selection of self-reactive T cells is far from a complete process, with frequent self-specific T cells observed in the periphery of healthy individuals or experimental animals, which, under certain conditions, can be mobilized to mount autoimmune responses (*1*). The incomplete deletion of self-reactive T cells or their diversion to immunomodulatory regulatory T cell (T_{reg}) or intraepithelial lymphocyte (IEL) lineages necessitates peripheral mechanisms that restrain the activation of self-reactive T cells upon encounter with cognate self-antigen.

> "...the ability to discriminate 'harmless' commensal microbes from pathogens, given antigenic overlap, remains one of the most notable feats of the immune system."

Once outside of the thymus, naïve T cells encounter antigens derived from innocuous sources, including commensal microorganisms, dietary antigens, and, during pregnancy, paternal fetal antigens. Of these encounters, the ability to discriminate "harmless" commensal microbes from pathogens, given antigenic overlap, remains one of the most notable feats of the immune system. A breakdown in peripheral tolerance to commensal antigens is thought to underlie the pathogenesis of a broad range of atopic and inflammatory diseases. This notion is supported by a recent report that recurrent TCRs restricted by HLA-B27—an MHC allele with the strongest genetic association with human skeletal and oc… autoinflammatory disorders—are capabl… recognizing peptide motifs shared by commensal microbial antigens and self-antigens and that these T cell clones are expanded at sites of autoimmune inflammation (*2*).

Despite clear parallels between mechanisms of central and peripheral tolerance, each organ poses a distinctive set of challenges with different diversity and abundance of tissue-specific self-antigens and commensal and environmental antigens. Given the tools that exist for the study of peripheral tolerance to gut microbes, including genetically engineered gut commensals as well as the ability to isolate and experimentally manipulate gut microbe–induced T_{regs}, the intestine has proved a fertile area for elucidating mechanisms of peripheral T cell tolerance, as highlighted by the recent discovery of a lineage of tolerogenic RORγt$^+$ antigen-presenting cells (APCs) and their role in instructing intestinal T_{reg} differentiation (*3*–*5*). The extent to which these mechanisms are conserved across tissue sites, as well as between different classes of antigens, remains to be established, however. Within the intestine, first encounters with dietary and commensal antigens occur during defined developmental windows such as the introduction of food proteins during weaning and ordered temporal patterns of microbial colonization. Investigations of peripheral tolerance during this early-life period have yielded insights into its basic mechanisms, which may be harnessed to restore immune balance later in life in the event of a breakdown in immune tolerance.

Here, we review recent advances in investigations of mechanisms of peripheral tolerance. In our efforts to provide an integrated view of this complex network of immunoregulatory cells and molecules, we discuss shared and distinct mechanisms for establishing tolerance to self-antigens versus foreign antigens, with a particular focus on pathways of tolerance to the gut microbiota. While emphasizing the intestine as a model organ for studying peripheral T cell tolerance (Fig. 1), we highlight the importance of particular tissue environments and the developmental stage of the organism in determining tolerance.

Peripheral regulation of T cell reactivity

Mechanisms of peripheral T cell tolerance are traditionally classified as cell intrinsic and cell extrinsic. The former encompass processes of apoptosis and consequent removal of particular T cell clones from the repertoire, induction of unresponsiveness (anergy), or loss of their capacity for inflammatory effector response upon TCR ligation (*6*). Extrinsic mechanisms involve the restraint of T cell responses by specialized immunoregulatory cell types such as T_{regs} or tolerogenic APCs. These distinctions are somewhat blurred because intrinsic mechanisms

[1]Immuno-Oncology, Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Immunology and Microbial Pathogenesis Program, Weill Cornell Medicine Graduate School of Medical Sciences, New York, NY, USA. [3]Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [4]Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [5]Immunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
*Corresponding author. Email: brownc10@mskcc.org (C.C.B); rudenska@mskcc.org (A.Y.R.)

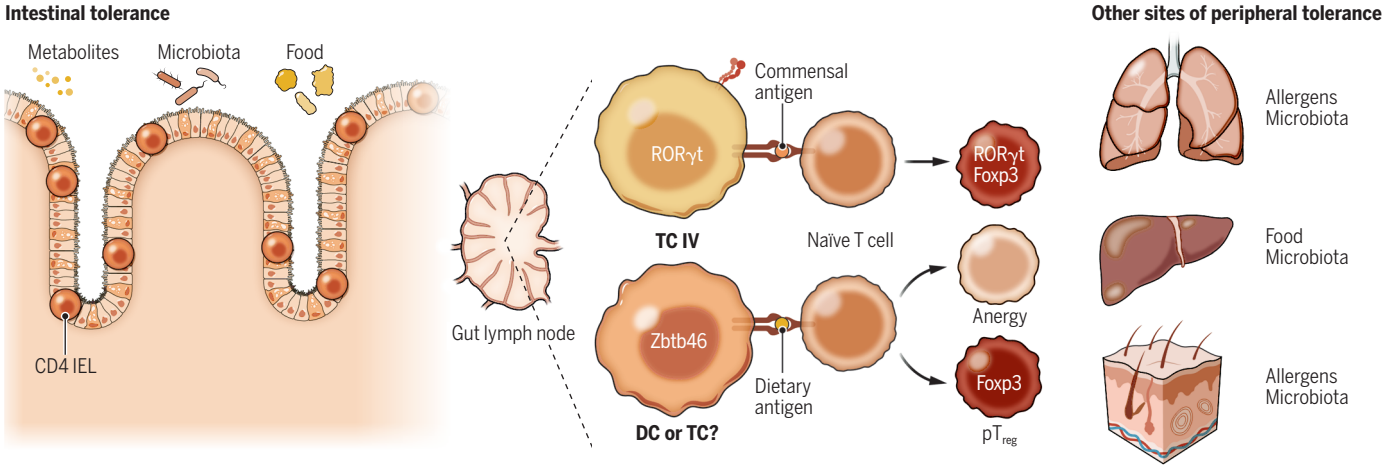


Fig. 1. The intestine as a model tissue for investigation of peripheral tolerance. Identification of the mechanisms underlying intestinal pT_{reg} generation in response to commensal and dietary antigens, including newly characterized tolerogenic APCs, provides a framework for investigation of peripheral tolerance at other tissue sites. Outside of the intestine, it is not clear whether distinct APCs and niche factors, including microbial products, regulate pT_{reg} differentiation, or whether commensal-specific T_{regs} within the skin, lung, and liver are also distinguished by expression of RORγt.

involve signaling through inhibitory receptors that raise the threshold of T cell activation downstream of TCR signaling, rendering T cells sensitive to extrinsic regulation by cell types that express corresponding ligands.

Modulation of T cell reactivity appears to be a major factor that prevents activation of peripheral self-reactive "escapees," at least at steady state. In healthy humans, circulating self-reactive CD8 T cells, although common, are hyporesponsive to ex vivo stimulation with cognate antigen (*7*). Similarly, in mice, peripheral $CD4^+$ T cells specific for a fetal or tissue-restricted self-antigen are characterized by the expression of folate receptor 4 (FR4) and CD73, two hallmarks of anergy (*8*, *9*). These cells are also prone to give rise to T_{regs} upon self-antigen challenge (*8*), suggesting a continuum of cell states that shift from intrinsic to extrinsic mechanisms of T cell tolerance. The extent to which T cell anergy is imprinted in the thymus versus periphery and whether continuous regulation is required to maintain anergy once instructed [e.g., as seen in the tumor microenvironment (*10*)] remain to be elucidated. Beyond tolerance to self-antigens, a recent study of CD4 T cell responses to oral antigens, including dietary protein gliadin (wheat), revealed a dominant population of cells that display overlapping transcriptional and phenotypic features of naïve T cells (*11*). These naïve-like T cells also exhibit hallmark features of anergy, with hyporesponsiveness to systemic immunization as well as increased propensity for peripheral T_{reg} differentiation upon continuous oral antigen exposure. In contrast to the thymus, where clonal elimination is a major mechanism of T cell tolerance, the contribution of deletion to peripheral tolerance is less clear. In this regard, one study that used tissue-restricted promoters to drive expression of Cre-recombinase peptide in the periphery reported that the numbers of Cre-specific CD4 T cells were not impaired, suggesting that deletional tolerance is minimal for antigens that are not expressed within the thymus (*12*). Similar to the thymus, where the strength of TCR signaling is a major determinant of clonal deletion versus diversion to alternative cell fates, it is likely that peripheral deletion may only remove self-reactive T cell clones with the highest-avidity TCRs that happen to escape thymic selection unscathed.

The TCR-dependent activation and differentiation potential of self-reactive peripheral T cells is modulated foremost by the prototypical costimulatory receptor CD28 and coinhibitory receptors programmed cell death 1 (PD-1) and cytotoxic T lymphocyte–associated protein 4 (CTLA-4). CTLA-4 and CD28 converge on two shared ligands, CD80 and CD86, that are expressed by APCs. In addition to competing with CD28 for binding of these ligands and potentially inducing a yet-to-be-defined inhibitory signaling cascade, CTLA-4 on T_{regs} is capable of removing CD80 and CD86 from APCs (*13*–*15*). Induction of PD-1 and CTLA-4 expression on T cells downstream of TCR signaling, coupled with dynamic changes in CD80, CD86, and PD-L1 expression after APC activation, suggests finely balanced signals between these interacting cellular and molecular partners that allow tight temporal and spatial control of T cell self-reactivity and responsiveness and complicate the assignment of an immunostimulatory or immunoinhibitory attribution to APCs based on static analyses of a few select cell-surface molecules. Antibody-mediated CTLA-4, PD-1, or PD-L1 (the ligand of PD-1) blockade, a cornerstone of cancer immunotherapy, has provided strong support for a major role of these pathways in peripheral tolerance in humans. Severe immune-related adverse events (iRAEs) that affect barrier tissues (skin, gut, and lung) and other organs are common in patients treated with PD-1– and CTLA-4–blocking antibodies (*16*). iRAEs seem to represent a breakdown in peripheral tolerance that is due to loss of PD-1–mediated restraint of self-reactive T cells. Cytotoxic CD8 T cells specific for α-myosin heavy chain, for example, are associated with myocarditis observed in $Pdcd1^{-/-}Ctla4^{+/-}$ mice and patients treated with a combination of anti–CTLA-4 and anti–PD-1 antibodies (*17*). Given the preponderance of iRAEs at microbiota-colonized barrier tissues, one would likewise expect a breakdown in tolerance to commensal and dietary antigens. Consistent with this possibility, patients with checkpoint-induced colitis show expansion of colonic lamina propria effector CD8 T cells that are clonally related to tissue-resident memory T cells of an unelucidated specificity (*18*).

Sustained stimulation of antigen-specific CD8 T cells in chronic viral infections and tumors has been linked to a terminal differentiation state termed "exhaustion," which is defined by high amounts of PD-1 expression, certain metabolic changes, and a progressive loss of inflammatory effector cytokine production and proliferative potential (*19*). The term exhaustion has negative connotations, suggesting a loss of functionality rather than a beneficial adaptation that serves to restrain inflammatory T cell responses and protect the host from excessive tissue damage. Through this lens, it is reasonable to consider that sustained antigenic stimulation of T cells—a common feature of infection, autoimmunity, and cancer—could promote their transition from inflammatory effectors into a range of noninflammatory, potentially tissue-supporting, or even immunoregulatory states. In support

CREDIT: A. FISHER/*SCIENCE*

of this notion, a recent study demonstrated that in addition to the loss of effector functions, exhausted T cells gain transcriptional and metabolic features of T_{regs}, including the ability to mediate local suppression of CD8 T cell responses by means of CD39-driven extracellular adenosine production (*20*).

A division in labor: Thymic versus peripheral T_{regs}

Foxp3$^+$ T_{regs} represent the principal immunosuppressive arm of the adaptive immune system. Although other T cell subsets—such as Tr1, CD4 IEL, Qa-1–restricted, or KIR$^+$ CD8 T cells (*21*–*24*)—have also been implicated in immune regulation, we limit this section discussion to T_{regs}, whose paramount importance in peripheral tolerance is demonstrated by the early-onset fatal autoimmunity in mice and humans that results from Foxp3 loss of function. A life-long requirement for T_{regs} in the maintenance of peripheral tolerance comes from the similarly aggressive and fatal disease observed not only in microbiota-sufficient but also adult germ-free *Foxp3^{DTR}* mice subjected to sustained ablation of T_{regs} (*25*, *26*). T_{regs} use a variety of effector mechanisms to sustain peripheral tolerance, including direct targeting of activated T cells as well as indirect effects through modulation of APCs, which have been comprehensively reviewed elsewhere (*25*).

Transcriptional and functional heterogeneity has been extensively described among T_{regs}. Arguably the greatest functional division can be attributed to the dichotomy between thymically derived T_{regs} (tT_{regs}) and those derived extrathymically in the periphery (pT_{regs}). The mechanistic studies of pT_{regs} and their origins, transcriptional regulation, and antigenic specificities have revealed distinct facets of tolerance to self-antigens and foreign antigens. tT_{regs}, by virtue of their agonist-driven selection supported by Aire-expressing medullary thymic epithelial cells (mTECs), are enriched for reactivity to self-antigens. Mice deficient in Aire or Aire$^+$ mTECs exhibit decreased T_{reg} numbers in the thymus within the first few weeks of life (*27*, *28*) and exhibit a shift in their tT_{reg} TCR repertoire owing to the retention of T_{reg}-associated tissue antigen–specific TCRs within the CD4 T_{conv} cell population (*28*, *29*). The ensuing autoimmune inflammation, despite normal numbers of T_{regs} within the periphery, suggests that the loss of specific tissue antigen–reactive tT_{reg} clones leads to insufficient control of their self-reactive T_{conv} cell counterparts. Indeed, multiple lines of evidence support the concept of "linked suppression," whereby tT_{regs} most efficiently suppress T_{conv} cells with shared antigenic or tissue specificity (*25*, *30*). Although, at present, no tool exists that allows specific ablation of tT_{regs}, the severe autoimmunity and inflammatory disease in germ-free mice that develops after ablation of Foxp3$^+$ cells, similar to that seen in microbiota-sufficient mice, provides clear evidence that tT_{regs} regulate peripheral tolerance to self (*26*).

Following the discovery that naïve T cells can up-regulate Foxp3 upon activation under particular circumstances to give rise to pT_{regs}, multiple studies have shown that pT_{regs} appear to largely develop in response to commensal microbiota and dietary antigens. Although there is a paucity of available markers to universally distinguish pT_{regs} and tT_{regs}, recent genetic lineage tracing studies of Foxp3$^+$ cells arising de novo within the periphery demonstrated that the differentiation of pT_{regs} can be triggered by microbial antigens in colon-draining lymph nodes and in the colonic lamina propria (*31*). Moreover, colonic pT_{regs} are marked by the expression of orphan nuclear receptor RORγt. Consistent with the role of microbial antigens in driving pT_{reg} generation, germ-free mice are largely devoid of intestinal RORγt$^+$ pT_{regs}, which can be readily induced upon microbial colonization (*32*). Furthermore, naïve CD4 T cells specific for defined commensal antigens give rise to RORγt$^+$ T_{regs} upon adoptive transfer into recipients colonized with the corresponding commensals (*33*–*35*). Thus, RORγt$^+$ pT_{regs} have become synonymous with gut microbe–induced pT_{regs} and represent the dominant Foxp3$^+$ population within the intestine. These microbe-specific pT_{regs} can further differentiate into immunomodulatory CD4 IELs (*24*, *36*), highlighting the importance of the tissue site in determining the mode of immunoregulation.

In keeping with their proven specificity for commensal antigens, RORγt$^+$ pT_{regs} have nonredundant roles in tolerance to commensal microbes, as evidenced by the onset of microbiota-dependent colitis in mice with genetic pT_{reg} deficiencies (*31*, *34*) or disruption of MHC class II (MHCII) antigen presentation on pT_{reg}-inducing APCs (*3*–*5*). Loss of pT_{regs} leads to a reciprocal increase in inflammatory T helper 2 (T_H2) and T_H17 cells and increased recruitment of neutrophils and monocytes, which is thought to underlie the ensuing intestinal inflammation (*3*, *31*, *32*, *37*, *38*). Beyond regulation of proinflammatory T cells, genetic or induced pT_{reg} deficiency leads to mastocytosis (*31*, *39*) as well as an increase in intestinal immunoglobulin A–positive (IgA$^+$) plasma cells (*38*). Within the small intestine and associated lymph nodes, differentiation of pT_{regs} is induced by food antigens (*11*, *40*). Mice deficient in intestinal pT_{regs} are more susceptible to allergic and inflammatory responses to oral antigens (*40*, *41*), further supporting the notion of a functional specialization of pT_{regs} in peripheral tolerance to microbe and food antigens. Although largely studied within the intestine, RORγt$^+$ T_{regs} have also been observed in lymph nodes that drain skin and other sites besides intestine, albeit at lower frequencies. At these sites, RORγt$^+$Foxp3$^+$ cells may represent pT_{regs} generated in response to the local microbial community or colonic pT_{regs}

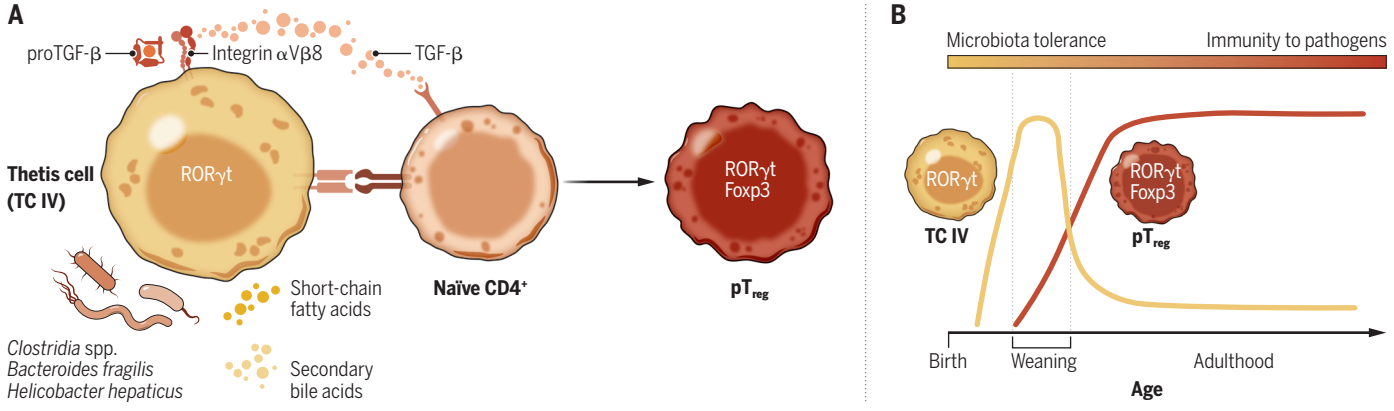


Fig. 2. Factors that promote pT_{reg} generation. (**A**) Within the intestine, commensal-specific pT_{regs} express RORγt and are instructed by a RORγt$^+$ APC (TC IV) in an ITGB8-dependent manner. A range of factors, including microbial antigens and microbial or dietary metabolites, promote pT_{reg} differentiation. (**B**) A developmental wave of TCs precedes the development of pT_{regs} upon weaning from breast milk to solid food.

that have migrated from the gut. Alternatively, RORγt may be up-regulated upon tT_{reg} activation in the presence of interleukin-6 (IL-6) and transforming growth factor–β (TGF-β) in a manner similar to that which occurs in T_H17 cells. Direct evidence supporting the migration of gut pT_{regs} comes from a recent study in which microbiota-dependent pT_{regs} generated in the gut were shown to migrate to distal sites, such as injured muscle, where they exerted immunoregulatory functions (*42*). Considering the major role that TCR specificity plays in the spatial positioning and function of T_{regs}, there are several possible scenarios for commensal-specific pT_{reg} contribution to tissue tolerance, including cross-reactivity with tissue antigens (*2*), migration of intestinal APCs bearing commensal antigens to distal sites, or low levels of colonization by translocation-competent commensals.

Although a clear role for the microbiota in pT_{reg} differentiation has been firmly established, the exact microbe-dependent signals that drive pT_{reg} differentiation upon antigen recognition are not fully known (Fig. 2). In contrast to tT_{reg} differentiation, where a direct role of TGF-β is still debatable (*43*), pT_{reg} differentiation is dependent on integrin beta 8 (ITGB8)–mediated TGF-β activation and signaling (*3*, *4*, *43*). Beyond specific examples of commensal-mediated modulation of TGF-β (*44*, *45*), a more general role for microbe-directed pT_{reg} differentiation by means of TGF-β signaling has not been established. A multitude of known and still-unidentified factors, some directly linked to microbiota, have been suggested to regulate pT_{reg} differentiation, including retinoic acid, butyrate, and secondary bile acids (*46*–*53*). How these local cues influence expression and activation of RORγt and other transcription factors implicated in microbiota-dependent differentiation of pT_{regs}, including c-Maf (*34*), remains to be determined. These transcription factors contribute to the distinctive pT_{reg} biology, which encompasses both Foxp3-dependent and -independent functions, including (for the latter) the stable expression of IL-10 (*31*, *54*). The recent finding that pT_{regs} are only partially dependent on Foxp3 for their immunosuppressive functions (*31*) sets them apart from tT_{regs}, whose function and fitness are absolutely dependent on Foxp3. These data suggest that Foxp3 acts as a late-acting factor during pT_{reg} differentiation to fine-tune and expand the immunomodulatory capacity of pT_{reg} precursors.

Tolerogenic APCs

Considering that APCs are the obligatory instigators of T cell activation and differentiation, the search for APC types or states endowed with a specific ability to induce T cell tolerance or deviation from inflammatory effector

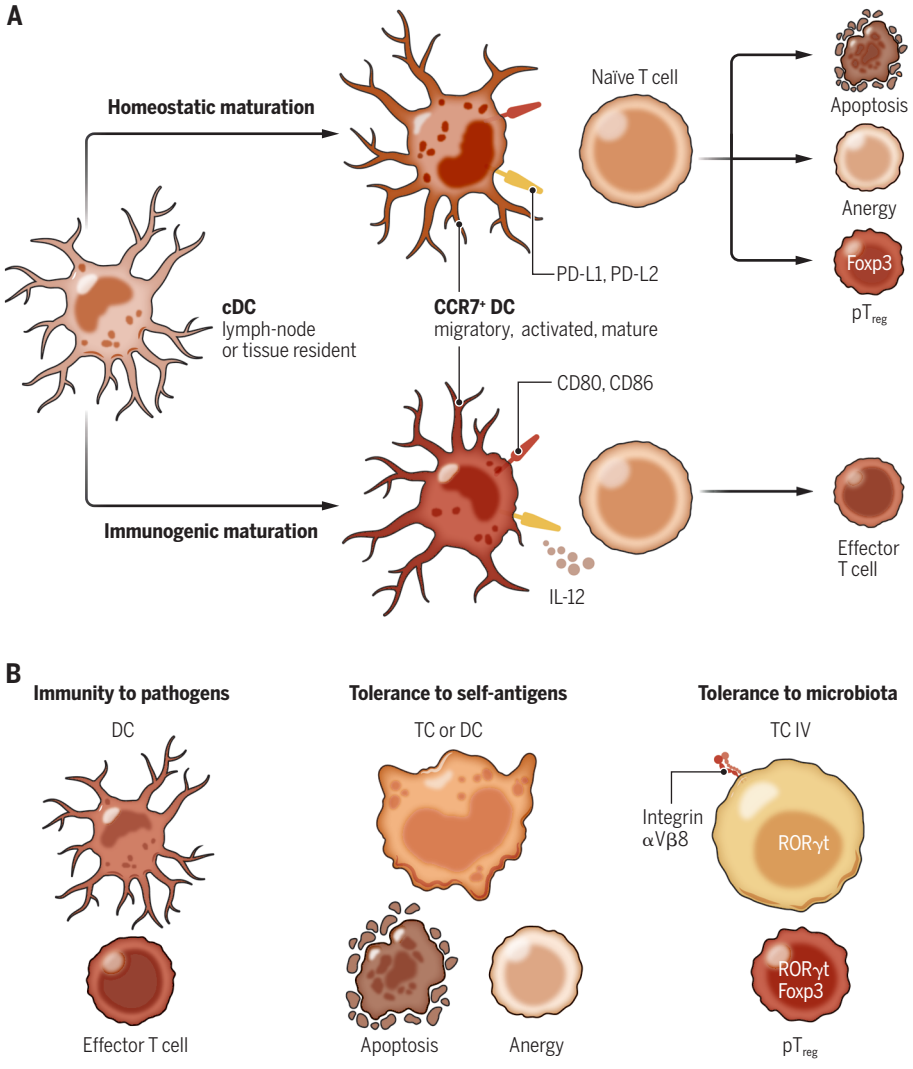


Fig. 3. Two models for tolerogenic APCs. (**A**) Homeostatic DCs are predicted to induce tolerance upon antigen presentation (homeostatic maturation), whereas DCs activated by pathogen-associated signals promote effector T cell differentiation (immunogenic maturation). The distinct "tolerogenic" versus "immunogenic" DC modules that mediate opposing T cell fates, as well as the ability of homeostatic DCs to induce all modes of T cell tolerance, have not been demonstrated. (**B**) Division of labor. Distinct APC subtypes are functionally specialized for induction of tolerance or effector T cell differentiation. Each tissue may use different APC types to direct tolerance to their microbial communities.

fate has been the holy grail of studies of peripheral tolerance. Although the existence of these cells has been widely anticipated, their identities remain poorly defined. Tolerogenic APCs, which, until recently, were assumed by default to be classical dendritic cells (cDCs), are often viewed as a monolithic entity, suggesting a single multipotent cell type with the ability to invoke both cell-intrinsic and -extrinsic mechanisms of peripheral tolerance to the entire spectrum of harmless antigens, including self-antigens and dietary, environmental, and commensal-derived antigens. However, the distinct requirements for pT_{reg} induction, such as TGF-β activation, as well as the recent finding that tolerance to the gut microbiota is instructed by a dedicated tolerogenic $ROR\gamma t^+$ APC, rather than intestinal DCs, point to a framework where distinct types of tolerogenic APCs instruct tolerance to different classes of antigens and induce particular modes of T cell tolerance.

DCs

DCs, the quintessential "professional" APCs, have evolved for highly efficient sampling, processing, and presentation of self-antigens and foreign antigens, alongside the expression of costimulatory and coinhibitory ligands. They are therefore poised to activate naïve T cells and regulate their differentiation fate. In contrast to a dedicated immunosuppressive

CREDIT: A. FISHER/*SCIENCE*

T_{reg} lineage endowed with the specific expression of a transcription factor (Foxp3) that ultimately orchestrates their differentiation and immunoregulatory function, a distinct tolerance-inducing DC type has not been identified. Instead, the view at present is that the DC activation state determines a "go" (immunity) or "no-go" (tolerance) decision (*55*) (Fig. 3). According to this model, functionally immature DCs expressing low levels of costimulatory molecules induce a hyporesponsive T cell state or T_{reg} differentiation by default. This model suggests that self-reactivity is limited in a passive manner through lack of a permissive second signal. However, the implicit assumption in this model, namely, that absent infection or sterile inflammation, self-antigens are presented by DCs lacking sufficient CD80 and CD86 levels, is not consistent with the finding that tissue-resident DCs undergo homeostatic maturation and migration to lymph nodes (Fig. 3). DC "maturation"—often used interchangeably with "activation"—refers to a switch from an antigen-acquisition and -processing mode to one of antigen presentation associated with increased expression of surface MHCII and costimulatory molecules (e.g., CD40, CD80, and CD86) (*56*, *57*). DC maturation is also associated with upregulation of chemokine receptor CCR7, which enables their subsequent migration to tissue-draining lymph nodes (*57*), where $CCR7^+$ DCs represent up to 50% of the DC population. Thus, $CCR7^+$ DCs, variably described as "activated," "mature," or "migratory," display a highly conserved transcriptional program that distinguishes them from their $CCR7^-$ counterparts. Paradoxically, among DC subsets, $CCR7^+$ DCs express the highest level of not only costimulatory molecules but also inhibitory molecules such as PD-L1, PD-L2, and CD200 (*58*), suggesting tolerogenic functions, which has led to their recent designation as "mature DCs enriched in immune regulatory molecules" (mreg DCs) (*59*). If exposure of DCs to inflammatory or pathogen-associated signals results in their shift from a tolerogenic to inflammatory response–promoting state, one would expect that $CCR7^+$ DCs would lose their immunoregulatory features in these settings. However, a comparison of unstimulated ("tolerogenic") versus Toll-like receptor–stimulated ("immunogenic") $CCR7^+$ DCs suggest convergent transcriptomes (*60*), including conserved expression of immune regulatory molecules such as PD-L1 and PD-L2. Although $CCR7^+$ DCs exhibit superior in vitro T_{reg} (iT_{reg})–inducing capacity (*59*, *61*)—likely related to their expression of the TGF-β–activating integrin αvβ8 and costimulatory molecules—an in vivo role for homeostatic $CCR7^+$ DCs in pT_{reg} differentiation (or anergy induction) has not been established. Genetic models that allow selective targeting of $CCR7^+$ DCs will therefore be required to discern the functional role of $CCR7^+$ DCs in tolerance versus immunity.

$ROR\gamma t^+$ APCs: A recently recognized lineage of tolerogenic APCs

At present, the evidence implicating DCs, which are known to express the transcription factor Zbtb46 and integrin CD11c, in peripheral tolerance stems from studies of ablation of $CD11c^+$ or $Zbtb46^+$ cells or their antigen-presenting capacity using the corresponding drivers of Cre or diphtheria-toxin-receptor expression. In these studies, CD11c or Zbtb46–expressing cells, assumed to be DCs, promoted tolerance to self-antigens (*62*, *63*) as well as oral (*64*) and microbiota-derived antigens (*35*, *65*). However, recent studies have identified a previously unappreciated lineage of APCs, distinguished by expression of the transcription factor RORγt, that share expression of these molecules (*3*–*5*). Furthermore, selective ablation of MHCII in DC precursors showed that antigen presentation by cDCs is dispensable for intestinal pT_{reg} differentiation (*3*–*5*), challenging the long-held view of DCs as the sole APC type responsible for the full spectrum of peripheral tolerance. These same studies also demonstrated that MHCII antigen presentation by $ROR\gamma t^+$ APCs is both necessary and sufficient for pT_{reg} differentiation. In two studies (*4*, *5*), the spectrum of $ROR\gamma t^+$ APCs revealed by single-cell transcriptomics included type 3 innate lymphoid cells (ILC3s) as well as previously described Aire-expressing $ROR\gamma t^+$ cells variably termed $ROR\gamma t^+$ extrathymic Aire-expressing cells (eTACs) (*5*), Janus cells (*4*, *66*), or Aire-expressing ILC3-like cells (*67*). In a third study (*3*), these $ROR\gamma t^+Aire^+$ cells were shown to belong to a broader lineage of cells, named Thetis cells (TCs), that encompasses two transcriptionally distinct $Aire^+$ subsets (TC I and TC III) as well as two $Aire^-$ subsets (TC II and TC IV) (Fig. 4). Although these cells shared markers with both DCs and ILCs, as well as mTECs, specific genetic tracing of ILC and cDC lineages demonstrated that TCs are ontogenically distinct from either one even though the specifics of TC ontogeny remain to be established (*3*, *5*). A unifying finding from the three studies, besides a requirement for MHCII presentation by $ROR\gamma t^+$ APCs, was a requirement for expression of either integrin alpha V (ITGAV) or ITGB8, the two subunits of the TGF-β–activating integrin $\alpha_v\beta_8$. Experiments using mixed chimeras demonstrated that expression of MHCII and either ITGAV or ITGB8 by the same cell was required for pT_{reg} induction (*3*, *4*), consistent with an important role for TGF-β signaling in pT_{reg} differentiation. Among $ROR\gamma t^+$ APCs, only TCs (most notably $Aire^-$ TC IVs), are distinguished by the expression of integrin $\alpha_v\beta_8$, the loss of which led to impaired pT_{reg} differentiation (*3*). The distinctive expression of ITGB8 by TCs, along with the lack of pT_{reg} deficit in mice where MHCII deficiency was restricted to ILC3, argues against a role for ILC3 in pT_{reg} generation. The finding that tolerance to the gut microbiota is instructed by a dedicated tolerogenic $ROR\gamma t^+$ APC, rather than intestinal DCs, highlights a specialized mechanism of protection. This is in contrast to the alternative error-prone scenario, whereby DCs, in an environment-dependent manner, execute both tolerogenic responses to commensals and immunogenic responses to pathogens. A key question that emerges from these findings is whether TC IVs have a role in instructing pT_{reg} differentiation and tolerance to commensal microbes at other barrier sites, such as the skin and lung, and overall pT_{reg} generation, for example, in response to dietary antigens within the intestine or to paternal fetal antigens.

"Studies of early-life immune development have revealed distinct waves of tolerogenic immune cells emerging within the thymus or periphery."

The discovery of TC-instructed pT_{reg} generation alters the paradigm for pT_{reg} generation, raising the question as to whether TC-mediated tolerance extends to other pathways of peripheral T cell tolerance, for example, clonal deletion of self-reactive T cells. Although the expression of Aire by TCs, as well as their transcriptional overlap with mTECs, is intriguing given the established role of Aire in self-tolerance, neither MHCII antigen presentation by $ROR\gamma t^+Aire^+$ cells nor the cells themselves are required for pT_{reg} differentiation (*4*, *5*). These results suggest a functional division between $Aire^+$ and $Aire^-$ TC subsets, manifesting in distinct roles in tolerance to self-antigens versus commensal antigens. A role for $Aire^+$ TCs in self-tolerance remains to be established, but previous studies examining the role of eTACs have suggested that these cells participate in extrathymic deletional tolerance and in maternal-fetal tolerance (*66*, *68*, *69*). Although low expression of Aire transcripts and an Aire reporter transgene are detectable in $CCR7^+$ DCs (*3*, *66*), only $ROR\gamma t^+Aire^+$ cells express detectable Aire protein (*3*, *70*). This, alongside the broader transcriptional overlap between TCs and mTECs, beyond expression of Aire, raises the possibility of a specialized role for TCs in peripheral

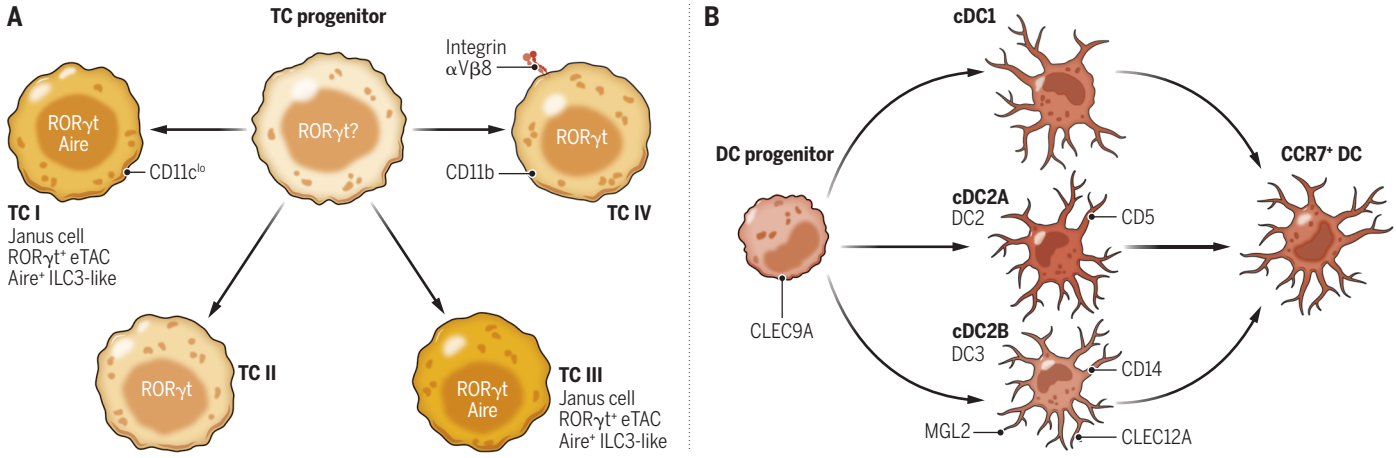


Fig. 4. The expanding spectrum of APCs. (**A**) Newly identified RORγt$^+$ TCs share cDC markers, including CD11c and Zbtb46, but are ontogenically distinct. TCs encompass four subsets that include two Aire-expressing cell types (TC I and TC III), suggesting potential roles in tolerance to self-antigens. Among Aire$^-$ cells, MHCII antigen presentation and ITGB8 expression by TC IVs is required for intestinal pT$_{reg}$ differentiation and tolerance to the gut microbiota. (**B**) DCs arise from CLEC9A$^+$ progenitors. DCs acquire expression of Aire transcript upon differentiation into CCR7-expressing migratory DCs. However, CCR7$^+$ DCs do not express detectable Aire protein.

tolerance to self, paralleling and complementing that of mTECs in central tolerance.

Developmental windows for immune tolerance

Life history of encounters with self-antigens and foreign antigens is not monotonous: Birth, weaning, sexual maturation, and pregnancy place temporal demands on the immune system as a result of the sheer number of new antigens that are encountered. Studies of early-life immune development have revealed distinct waves of tolerogenic immune cells emerging within the thymus or periphery (*3*, *27*, *71*). In addition to quantitative differences in T$_{reg}$ numbers, neonatal tT$_{regs}$ have a distinctive functionality that cannot be compensated by adult tT$_{regs}$ (*27*), suggesting not only layered ontogeny but also developmental stage–specific programs, TCR specificities, and selection in distinct environments that endow early-life tT$_{regs}$ with distinct functional properties. Genetic lineage tracing studies revealed that perinatally generated tT$_{regs}$ persist long after birth, but whether these cells provide specific functions in adulthood remain to be determined.

Within the intestine, the gut microbial community undergoes a profound diversification upon transition from breast milk to solid foods. In addition, this transition is associated with the withdrawal from a wealth of immunomodulatory factors, including growth factors, cytokines, maternal immunoglobulins, and hormones. Failure to establish intestinal tolerance during this weaning period leads to an increased risk of allergic and inflammatory diseases later in life (*71*). Although previously thought of as immature or underdeveloped, the neonatal and infant immune system is highly adapted for establishment of immune tolerance, with the aforementioned layered ontogeny that ensures enriched representation of tolerogenic APCs and T$_{regs}$ during critical windows when the host first encounters dietary and microbial antigens. Adoptive transfer studies of TCR-transgenic naïve T cells expressing a microbe-reactive TCR demonstrated increased propensity for pT$_{reg}$ generation when cells were transferred into mice before weaning (*35*, *72*). Because these adoptive transfer studies were performed with naïve T cells isolated from adult mice, this suggests that the wave of pT$_{regs}$ that emerges upon weaning is driven by extrinsic factors within the developing intestinal niche. The finding that the pT$_{reg}$-inducing APCs, TC IVs, emerge in a developmental wave within the mesenteric lymph nodes that coincides with weaning (*3*) suggests that these cells may define the temporally restricted window for pT$_{reg}$ differentiation and their homeostatic abundance, which in turn will determine the immune-regulatory tone within the intestine. In this regard, an intriguing observation is that pT$_{reg}$ abundance in adulthood is determined by maternally transmissible cues, including breast milk IgA, that are sensed during the first week of life, before the emergence of pT$_{regs}$ (*73*). Similar to the window of opportunity for intestinal tolerance to commensal microbes, neonatal encounters with skin commensal antigens lead to the development of commensal-specific pT$_{regs}$ and tolerance (*74*). Thus, dynamic changes in tolerogenic immune cell populations mirror the developmental demands of the host and perhaps serve to limit an overabundance of dominant immunosuppressive cell types that might compromise effector immunity.

Recent years have seen major advances in studies of peripheral tolerance. Yet these discoveries have also raised a number of intriguing questions: Does each barrier tissue with its particular organization, cellular composition, and microbial community have its own conventions for establishing tolerance, or do certain general rules exist with the same cell types and mechanisms that operate across distinct organs? Analogous to the functional division between tT$_{regs}$ versus pT$_{regs}$, do distinct peripheral APCs instruct tolerance to self-antigen versus foreign antigen? What determines the increased propensity for tolerogenic immune cell fates during early life, and are these same mechanisms used to restore tolerance in adulthood? An increased understanding of the cell types and molecular mediators that establish tolerance will pave the way for future studies that address these questions.

REFERENCES AND NOTES

1. D. M. Richards, B. Kyewski, M. Feuerer, *Trends Immunol.* **37**, 114–125 (2016).
2. X. Yang *et al.*, *Nature* **612**, 771–777 (2022).
3. B. Akagbosu *et al.*, *Nature* **610**, 752–760 (2022).
4. R. Kedmi *et al.*, *Nature* **610**, 737–743 (2022).
5. M. Lyu *et al.*, *Nature* **610**, 744–751 (2022).
6. M. A. ElTanbouly, R. J. Noelle, *Nat. Rev. Immunol.* **21**, 257–267 (2021).
7. W. Yu *et al.*, *Immunity* **42**, 929–941 (2015).
8. L. A. Kalekar *et al.*, *Nat. Immunol.* **17**, 304–314 (2016).
9. T. Hassler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 18537–18543 (2019).
10. R. Alonso *et al.*, *Nat. Commun.* **9**, 2113 (2018).
11. S. W. Hong *et al.*, *Nature* **607**, 762–768 (2022).
12. F. P. Legoux *et al.*, *Immunity* **43**, 896–908 (2015).
13. V. Ovcinnikovs *et al.*, *Sci. Immunol.* **4**, eaaw0902 (2019).
14. J. G. Egen, J. P. Allison, *Immunity* **16**, 23–35 (2002).
15. K. Wing *et al.*, *Science* **322**, 271–275 (2008).
16. K. P. Burke, S. Grebinoski, A. H. Sharpe, D. A. A. Vignali, *J. Exp. Med.* **218**, e20192179 (2021).
17. M. L. Axelrod *et al.*, *Nature* **611**, 818–826 (2022).
18. A. M. Luoma *et al.*, *Cell* **182**, 655–671.e22 (2020).
19. C. U. Blank *et al.*, *Nat. Rev. Immunol.* **19**, 665–674 (2019).
20. P. D. A. Vignali *et al.*, *Nat. Immunol.* **24**, 267–279 (2023).
21. M. G. Roncarolo, S. Gregori, R. Bacchetta, M. Battaglia, N. Gagliani, *Immunity* **49**, 1004–1019 (2018).
22. H. J. Kim, B. Verbinnen, X. Tang, L. Lu, H. Cantor, *Nature* **467**, 328–332 (2010).
23. J. Li *et al.*, *Science* **376**, eabi9591 (2022).
24. T. Sujino *et al.*, *Science* **352**, 1581–1586 (2016).
25. S. Dikiy, A. Y. Rudensky, *Immunity* **56**, 240–255 (2023).

CREDIT: A. FISHER/*SCIENCE*

26. T. Chinen, P. Y. Volchkov, A. V. Chervonsky, A. Y. Rudensky, *J. Exp. Med.* **207**, 2323–2330 (2010).
27. S. Yang, N. Fujikado, D. Kolodin, C. Benoist, D. Mathis, *Science* **348**, 589–594 (2015).
28. C. Haftmann *et al.*, *Sci. Immunol.* **6**, eabf3111 (2021).
29. S. Malchow *et al.*, *Immunity* **44**, 1102–1113 (2016).
30. B. Akkaya *et al.*, *Nat. Immunol.* **20**, 218–231 (2019).
31. J. van der Veeken *et al.*, *Immunity* **55**, 1173–1184.e7 (2022).
32. E. Sefik *et al.*, *Science* **349**, 993–997 (2015).
33. M. M. Wegorzewska *et al.*, *Sci. Immunol.* **4**, eaau9079 (2019).
34. M. Xu *et al.*, *Nature* **554**, 373–377 (2018).
35. K. Nutsch *et al.*, *Cell Rep.* **17**, 206–220 (2016).
36. D. Bousbaine *et al.*, *Science* **377**, 660–666 (2022).
37. C. Ohnmacht *et al.*, *Science* **349**, 989–993 (2015).
38. C. Neumann *et al.*, *Nat. Immunol.* **20**, 471–481 (2019).
39. C. Campbell *et al.*, *Immunity* **48**, 1245–1257.e9 (2018).
40. K. S. Kim *et al.*, *Science* **351**, 858–863 (2016).
41. A. Abdel-Gadir *et al.*, *Nat. Med.* **25**, 1164–1174 (2019).
42. B. S. Hanna *et al.*, *Immunity* 10.1016/j.immuni.2023.01.033 (2023).
43. J. M. Moreau, M. Velegraki, C. Bolyard, M. D. Rosenblum, Z. Li, *Sci. Immunol.* **7**, eabi4613 (2022).
44. K. Atarashi *et al.*, *Science* **331**, 337–341 (2010).
45. C. A. M. Finney, M. D. Taylor, M. S. Wilson, R. M. Maizels, *Eur. J. Immunol.* **37**, 1874–1886 (2007).
46. J. A. Hill *et al.*, *Immunity* **29**, 758–770 (2008).
47. M. J. Benson, K. Pino-Lagos, M. Rosemblatt, R. J. Noelle, *J. Exp. Med.* **204**, 1765–1774 (2007).
48. D. Mucida *et al.*, *Science* **317**, 256–260 (2007).
49. C. Campbell *et al.*, *Nature* **581**, 475–479 (2020).
50. X. Song *et al.*, *Nature* **577**, 410–415 (2020).
51. P. M. Smith *et al.*, *Science* **341**, 569–573 (2013).
52. Y. Furusawa *et al.*, *Nature* **504**, 446–450 (2013).
53. N. Arpaia *et al.*, *Nature* **504**, 451–455 (2013).
54. S. Dikiy *et al.*, bioRxiv 2022.05.16.492030 [Preprint] (2003); https://doi.org/10.1101/2022.05.16.492030.
55. R. M. Steinman, D. Hawiger, M. C. Nussenzweig, *Annu. Rev. Immunol.* **21**, 685–711 (2003).
56. R. M. Steinman, M. Pack, K. Inaba, *Adv. Exp. Med. Biol.* **417**, 1–6 (1997).
57. M. Cabeza-Cabrerizo, A. Cardoso, C. M. Minutti, M. Pereira da Costa, C. Reis e Sousa, *Annu. Rev. Immunol.* **39**, 131–166 (2021).
58. J. C. Miller *et al.*, *Nat. Immunol.* **13**, 888–899 (2012).
59. B. Maier *et al.*, *Nature* **580**, 257–262 (2020).
60. L. Ardouin *et al.*, *Immunity* **45**, 305–318 (2016).
61. A. Shiokawa, R. Kotaki, T. Takano, H. Nakajima-Adachi, S. Hachimura, *Immunology* **152**, 52–64 (2017).
62. N. Yogev *et al.*, *Immunity* **37**, 264–275 (2012).
63. C. Ohnmacht *et al.*, *J. Exp. Med.* **206**, 549–559 (2009).
64. D. Esterházy *et al.*, *Nat. Immunol.* **17**, 545–555 (2016).
65. J. Loschko *et al.*, *J. Exp. Med.* **213**, 517–534 (2016).
66. J. Wang *et al.*, *Sci. Immunol.* **6**, eabl5053 (2021).
67. T. Yamano *et al.*, *J. Exp. Med.* **216**, 1027–1037 (2019).
68. J. M. Gardner *et al.*, *Science* **321**, 843–847 (2008).
69. E. Gillis-Buck *et al.*, *Sci. Immunol.* **6**, eabf1968 (2021).
70. J. Dobeš *et al.*, *Nat. Immunol.* **23**, 1098–1108 (2022).
71. P. Brodin, *Science* **376**, 945–950 (2022).
72. K. A. Knoop *et al.*, *Sci. Immunol.* **2**, eaao1314 (2017).
73. D. Ramanan *et al.*, *Cell* **181**, 1276–1290.e13 (2020).
74. J. M. Leech *et al.*, *Cell Host Microbe* **26**, 795–809.e5 (2019).

ACKNOWLEDGMENTS

We apologize in advance to the many investigators whose work could not be cited in this manuscript because of limited space. **Funding:** C.C.B. is supported by the Parker Institute for Cancer Immunotherapy, National Institutes for Health National Institute of Allergy and Infectious Diseases (NIAID) DP2 AI171116-01, the G. Harold and Leila Y. Mathers Foundation, and a Josie Robertson Investigator award. A.Y.R. is a Howard Hughes Medical Institute investigator and is supported by the Ludwig Center at Memorial Sloan Kettering, National Cancer Institute (NCI) Cancer Center Support Grant P30 CA008748, and NIAID R01 AI034206. **Competing Interests:** A.Y.R. is a scientific advisory board member of and has equity in Sonoma Biotherapeutics, Vedanta Biosciences, Santa Ana Bio, Surface Oncology, and RAPT Therapeutics. A.Y.R. is also a scientific advisory board member of Amgen and BioInvent and holds intellectual property licensed to Takeda.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 22 February 2023; accepted 31 March 2023
10.1126/science.adg6425

REVIEW

Innate virus-sensing pathways in B cell systemic autoimmunity

Carola G. Vinuesa[1,2]*, Amalie Grenov[1], George Kassiotis[1,3]*

Although all multicellular organisms have germ line–encoded innate receptors to sense pathogen-associated molecular patterns, vertebrates also evolved adaptive immunity based on somatically generated antigen receptors on B and T cells. Because randomly generated antigen receptors may also react with self-antigens, tolerance checkpoints operate to limit but not completely prevent autoimmunity. These two systems are intricately linked, with innate immunity playing an instrumental role in the induction of adaptive antiviral immunity. In this work, we review how inborn errors of innate immunity can instigate B cell autoimmunity. Increased nucleic acid sensing, often resulting from defects in metabolizing pathways or retroelement control, can break B cell tolerance and converge into TLR7-, cGAS-STING–, or MAVS-dominant signaling pathways. The resulting syndromes span a spectrum that ranges from chilblain and systemic lupus to severe interferonopathies.

It is becoming increasingly clear that the same pathways that are required to prevent mortality because of viral infections can also cause systemic autoimmunity. This is due, in part, to the inability of the immune system to perfectly distinguish invading genomes from self–nucleic acids. Genome-wide association studies (GWASs) have identified a range of risk variants associated with severe viral infection, including COVID-19 and systemic autoimmunity. Notably, many of the genes with identified risk alleles overlap between these two conditions. An example is *TLR7*, in which loss-of-function (LOF) or gain-of-function (GOF) variants confer increased susceptibility to COVID-19 and systemic lupus erythematosus (SLE), respectively. Mechanistic studies have further described how the majority of these overlapping risk factors act in innate nucleic acid–sensing pathways in the endosome as well as in the cytoplasm. In this work, we explore how the innate sensing of viral genomes and the induction of type I interferon (IFN I) synthesis may come with the risk of autoimmunity.

Pathogen-driven selection at the risk of autoimmunity

Pathogens are among the strongest selective pressures driving the evolution of modern humans (*1*). Those that can fully integrate into the host, such as endogenous retroviruses (ERVs), may even be co-opted for further evolutionary innovation, as illustrated by regulatory rewiring of human stem cell pluripotency by the human ERV H family (*2*). Common and rare variants in genes important for sensing pathogen-derived genomes or for presenting pathogen-derived peptides to the immune system profoundly modulate susceptibility to microbial infections. Infections that affect reproductive ability will drive positive selection of genetic variants that confer resistance to infection. Such selection is now well documented for parasites and bacteria that have been infecting humans for many centuries, such as those that cause malaria, tuberculosis, and leprosy (*3*), as well as for more recent pathogens that have had devastating effects on humanity. *Yersinia pestis*, for example, produced up to 200 million deaths during the Black Death in the 14th century (*4*). In so doing, it not only created evolutionary pressure to select gene variants that enhance immunity to *Y. pestis* (e.g., *ERAP2*) but also conferred increased susceptibility to organ-specific autoimmune diseases, such as Crohn's disease (*4*).

Viruses have decimated human populations throughout history: The Greek physician Hippocrates (~460 to ~370 BCE) described a flu-like illness called the "fever of Perinthus," which is thought to include influenza and some bacterial diseases (*5*). Determining positive selection effects of viral pandemics has nevertheless been challenging, at least in part because of the enormous diversity in viral genomes even within the same order or family. Six of the seven worst pandemics in the 20th and 21st centuries were caused by six different viruses (*6*): H1N1 (1918 Spanish flu, leading to 50 million deaths), H2N2 (1957 Asian flu, leading to 1 to 2 million deaths), H3N2 (1968 Hong-Kong flu, leading to 0.5 to 2 million deaths), a novel H1N1 (2009 Swine flu, leading to 0.25 million deaths), HIV (1987 to present, with AIDS leading to more than 35 million deaths), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (2019 to present, with COVID-19 leading to 6.8 million reported deaths) [WHO Coronavirus (COVID-19) dashboard; https://covid19.who.int]. Selection of viral immunity–enhancing variants that may increase susceptibility to autoimmunity has also been postulated in HIV resistance. Strong signals under extreme positive selection have been found in major histocompatibility complex (MHC) loci, with MHC variants proposed to explain up to 20% of the phenotypic variance between HIV control and advanced disease (*7*). Among these is a variant that protects against HIV progression and correlates with increased human leukocyte antigen C (HLA-C) expression (*8*). This same variant, however, also increases the risk of Crohn's disease (*9*).

A documented signal of viral infection–driven positive selection occurs in the most ancient antiviral family member of the *IFITM* locus—the IFN-induced transmembrane protein 3 (*IFITM3*), shown to confer resistance to influenza A infection (*10*) and to inhibit HIV-1 virus fusion and cell-to-cell spread. *IFITM3* has undergone repeated duplication and N-terminal diversification in primates, allowing increased antiviral coverage of host cells. Such duplication may also come at the cost of autoimmunity: *IFITM3* expression has been reported to be higher in SLE and to negatively correlate with complement components C3 and C4 (*11*).

Innate viral immunity relies on TLRs and IFN I

All life forms use nucleic acids (DNA or RNA genomes) to store information for their replication, and all have some form of defense against parasitic genomes. The detection of different forms of DNA or RNA foreign to a host offers the means to detect an invader. Toll-like receptors (TLRs), such as TLR3, TLR7/8, and TLR9, all detect nucleic acids that are not specific to invading pathogens. Thus, discrimination between self and nonself relies on sequestration of these TLRs to endosomes. TLR7/8 and TLR9 interact with single-stranded RNA (ssRNA) and ssDNA, respectively, and induce the synthesis of IFN I and IFN-stimulated gene (ISG) products through activation of MYD88, IRAK1/4, TRAF6, and IRF7. Although IRF7-dependent induction of IFN I responses is considered the prominent mechanism for antiviral effects of TLR7 stimulation (*12*), IRF5, AP-1, and nuclear factor κB (NF-κB) can be activated through similar pathways and lead to additional proinflammatory gene transcription.

Cytoplasmic nucleic acid sensors react to double-stranded DNA (dsDNA), dsRNA, and RNA-DNA hybrid molecules that accumulate upon infection or cellular stress, including ERV dysregulation (*13*). Exogenous nucleic acids may also gain access to cytoplasmic sensors through rupture of endosomal and mitochondrial membranes or micronuclei (*13*, *14*). Cyclic guanosine 5′-monophosphate (GMP)–adenosine 5′-monophosphate (AMP) synthase (cGAS) and AIM2 are the major cytosolic DNA sensors. cGAS synthesizes 2′3′-cyclic GMP-AMP

[1]The Francis Crick Institute, London, UK. [2]China Centre for Personalised Immunology, Renji Hospital, Shanghai, China. [3]Department of Infectious Disease, Faculty of Medicine, Imperial College London, London, UK.
*Corresponding author. Email: carola.vinuesa@crick.ac.uk (C.G.V.); george.kassiotis@crick.ac.uk (G.K.)

(cGAMP) upon recognition of dsDNA molecules. cGAMP is a strong agonist for STING, located at the endoplasmic reticulum (ER) membrane, and induces oligomerization of STING dimers, leading to trafficking of STING complexes to the Golgi apparatus, where TBK1 is recruited. IRF3 phosphorylation by STING-TBK1 complexes induces IFN I and ISG transcription. Similar to TLR-induced signaling, IFN I production is considered the main antiviral effector function of the cGAS-STING pathway (*13*). AIM2 triggers a different signaling cascade from other nucleic acid sensors, inducing the formation of an inflammasome that results in caspase 1 activation (*15*). The nuclear viral genome sensor, IFI16, also forms an inflammasome that activates caspase 1 during infection with DNA viruses or damaged self-DNA (*16*). Notably, both AIM2 and IFI16 have been found bound to DNA on neutrophil extracellular traps (NETs) in the kidneys of lupus patients, which may explain why they act as autoantigens in SLE (*17*).

In addition to establishing cell-intrinsic resistance, IFN I signaling through its receptor critically contributes to antiviral protection by providing essential survival signals to CD8 T cells, influencing clonal expansion as well as memory formation (*18*). Increased MHC-I presentation by IFN I–stimulated dendritic cells (DCs) adds additional support for CD8 T cell expansion in antiviral and antitumor responses (*19*). Recent studies have also demonstrated IFN-independent functions of STING in the clearance of viral infections. These effects may include the restriction of viral propagation by induction of canonical and noncanonical pathways of autophagy (*13*).

Evidence of the importance of these nucleic acid–sensing pathways in antiviral defense comes from the identification of rare pathogenic variants in genes important for IFN I production, such as *TLR3*, *IRF7*, and *IRF3*, in familial studies of individuals with increased susceptibility to life-threatening influenza pneumonia (*20*–*23*). More recently, common and rare gene variants involved in viral RNA sensing have also been linked to susceptibility to severe COVID-19. These variants occur in *TLR7* (*12*, *24*, *25*) and in molecules involved in IFN I production and signaling, including *OAS1*, *TYK2*, and *IFNAR2* (*26*). Functional validation of private or ultrarare variants found in life-threatening COVID-19 pneumonia have demonstrated that *TLR7*, *TLR3*, *TICAM1*, *TBK1*, *IRF3*, *IRF7*, *IFNAR1*, *IFNAR2*, and *UNC93B1* variants impair IFN I responses (*27*). The demonstration that 10% of patients with severe COVID-19 pneumonia have antibodies against IFN-α and IFN-ω further reinforces the strong requirement for IFN I in antiviral immunity (*28*).

It has been estimated that rare *TLR7* LOF variants confer a fivefold increase in the odds of developing severe COVID-19 and explain >1% of cases of severe COVID-19 in males (*12*, *24*, *25*). *TLR7* is encoded in the X chromosome but partly escapes X inactivation (*29*), which means that males effectively have a lower gene dose.

Endogenous viral control by innate immunity

In addition to exogenous viruses, the host must deal with myriad retroviral species that have invaded the germ line and are now part of its genetic constitution. These include ERVs, which originate from ancestral infection of the germ line with exogenous retroviruses, and more numerous endogenous retroelements, such as long interspersed nuclear elements 1 (LINE-1s) and short interspersed nuclear elements (SINEs), which include the primate-specific *Alu* elements.

Although most human ERVs are now inactive, several still provide ssRNA, dsRNA, and cytosolic DNA ligands through transcription and reverse transcription of their genomes. ERVs can provide ligands for TLR3, TLR7/8, and TLR9. However, TLR7 has been shown to play a dominant role in controlling their expression level (*30*, *31*). ERVs that have lost envelope glycoproteins are adapted to an intracellular life cycle and may gain direct access to TLR7 by budding into endosomes (*32*). ERV control has also been shown to be dependent on antibody formation, supporting the importance of TLR7-mediated enhancement of natural antibody responses for ERV repression (*31*).

"In a species in which viruses can threaten reproductive ability, autoimmunity may be a necessary trade-off."

Beyond the provision of ligands for nucleic acid sensors, endogenous retroelements can shape innate and, by extension, adaptive immunity and susceptibility to infection through diverse mechanisms. For example, the MER41B family of ERVs provides the regulatory elements that facilitate the responsiveness to IFN of several antiviral genes, including *AIM2* and *IFI6* (*33*). IFN inducibility of several other families of ERVs is also supported by the specificity of their expression in peripheral blood mononuclear cells (PBMCs) and B cells from multiple sclerosis patients treated with IFN I or patients with SLE (*34*, *35*). By providing alternative splicing sites and exons, endogenous retroelements can create novel immune gene isoforms, such as a soluble receptor-antagonist form of PD-L1 (*36*).

The capacity of endogenous retroelements to generate replication intermediates and trigger DNA- and RNA-sensing pathways creates evolutionary trade-offs, which fine-tune antiviral responses based on nucleic acid sensing (*37*, *38*). The balance that is struck minimizes the sensing of endogenous nucleic acids that could instigate fatal autoimmunity at the expense of sensing those from exogenous infection, thus potentially compromising antiviral defenses. Conversely, heightened sensitivity to exogenous infection may break tolerance to endogenous ligands (*37*, *38*).

Nucleic acid sensing and systemic autoimmunity

Immunity operates on the edge of autoimmunity. An accumulating body of work indicates that the tipping point between antiviral immunity and autoimmunity—particularly systemic autoimmunity—hinges on the strength of nucleic acid sensing and the amount of IFN I produced. Perhaps not surprisingly, several of the genes involved in IFN I production that have been linked to viral infection susceptibility (e.g., *OAS1*, *TLR7*, *TYK2*, and *IRF7*) have also been identified as lupus-associated genes in GWASs acting in the opposite polarity (*39*, *40*). Thus, the same pathways that protect the host against viral infection can cause systemic autoimmunity when signaling is excessive or when there is an overabundance of ligands. In a species in which viruses can threaten reproductive ability, autoimmunity may be a necessary trade-off.

The ligands that trigger autoimmunity are not necessarily of foreign viral origin, however. SLE patients have been found to accumulate increased levels of ERV-derived nucleic acids, which may contribute to IFN I responses (*32*). Moreover, antibodies against ERV envelope glycoproteins have been identified in patients with SLE, juvenile SLE (JSLE), and Sjögren's syndrome and are associated with a higher IFN I gene signature (*41*). ERV-targeted antibodies can promote inflammation partly through the activation of neutrophils leading to the formation of NETs (*41*). Thus, ERV-derived nucleic acids and ERV-targeting antibodies may facilitate autoinflammation in concert.

Nonviral nucleic acid ligands can also activate nucleic endosomal or cytosolic nucleic acid sensors and constitute important triggers of systemic autoimmunity. An important source of nonviral nucleic acids in SLE are mitochondria that can release their DNA upon different forms of stress [recently reviewed in (*14*)]. Mitochondrial DNA released from neutrophils, eosinophils, and platelets can activate IFN I production by plasmacytoid DCs (pDCs) (*14*). Recently, DNA from mitochondria abnormally retained in red blood cells from SLE patients has been shown to activate type I IFN production in myeloid cells (*42*). Additional sources of interferogenic nucleic acids include nuclear antigens exposed on the surface of apoptotic cells, breakdown of micronuclei that

occur in senescent cells or because of genomic instability through missegregation of DNA during cell division, DNA and histones attached to NETs, and chromatin bridges caused by antimitotic drugs (*43*).

Innate immunity to nucleic acids shapes B cell reactivity

Although TLRs and other nucleic acid sensors can detect foreign DNA or RNA, commonalities in nucleic acids ultimately necessitated the evolution of more discerning immune sensors, such as the somatically rearranged and clonally distributed antigen receptors on B cells and T cells (BCRs and TCRs, respectively), which detect products of foreign genome translation rather than the foreign genomes directly. However, one consequence of this evolutionary development is the existence of autoreactive BCRs and TCRs, which need to be controlled or eliminated. Recent evidence suggests that the more ancient system of intrinsic immunity still plays a major role in determining the reactivity of adaptive lymphocytes, particularly of B cells.

Survival signals elicited by TLR7 signaling in B cells appear indispensable for humoral immunity to many self and certain foreign antigens (*44*), likely as a result of an evolutionary adaptation connecting BCR signaling to endosomal signals emanating from BCR-mediated antigen endocytosis. In humans, TLR7 expression increases upon activation of naive B cells and has been suggested to be a third activation signal required for their maximal expansion (*45*). Increases in nucleic acid sensing including through TLR7 have been shown to break B cell tolerance and lead to the production of autoantibodies that trigger an inflammatory cascade and end-organ damage.

Links between nucleic acid sensing by TLR7 were first described when the cause of lupus-like disease in male BXSB-*Yaa* mice was mapped to a duplication of an X chromosome segment containing *Tlr7* onto the Y chromosome (*46*). Mice made transgenic for *Tlr7* or with dysregulated endosomal trafficking of TLR7 also developed lupus-like disease (*46*). Lupus autoantibodies were also shown to be TLR7 dependent in 564lgi, MRL.*lpr*, and NZB/W mouse lupus models, whereas IFN signaling through IFNAR was not required for autoantibody formation in TLR7-driven models of SLE (*47*, *48*). Instead, TLR7 directly promotes survival of B cells activated by chromatin-containing immune complexes in a manner independent of BAFF or of production of IFN I (*49*). By contrast, cross-linking TLR9 with the BCR promotes death of self-reactive B cells (*49*), highlighting important differences between the roles of TLR7 and TLR9 in B cell tolerance.

Subsequent studies in humans have reinforced the role of TLR7 in SLE pathogenesis. In a study of East Asian men, a polymorphism affecting TLR7 expression was found to be associated with SLE and to cause more pronounced IFN I signatures (*50*). An elegant study in human PBMCs revealed that a population of B cells, expanded in SLE patients, have increased sensitivity to TLR7 signaling (*51*). Only recently, a GOF de novo mutation in *TLR7* has been shown to cause human lupus (*52*). Introduction of this variant into mice was sufficient to drive lupus-like disease.

B cell–intrinsic TLR7 stimulation is critical for the formation of spontaneous germinal centers (GCs) (*53*), seen in most lupus mouse models, and for the formation of pathogenic age-associated B cells (ABCs) in SLE (*54*, *55*). Recent evidence from humans and mice carrying the human SLE TLR7 GOF mutation has suggested that extrafollicular ABCs and their plasma cell products, rather than GCs, are the pathogenic B cell subsets in SLE (*52*). Feedback loops between TLR7 and IFN I signals may further impair B cell tolerance. For example, IFN I can increase expression of TLR7 and MHC molecules, increasing nucleic acid sensing and antigen-presenting capacity, respectively. IFN I also promotes plasma cell differentiation in an autocrine manner as well as in nearby cells, augmenting autoantibody production. Autoantibodies in the form of immune complexes can amplify IFN I responses in DCs and macrophages through FcγR engagement. IFN I–mediated cytotoxic T cell recruitment is also likely to contribute to nephritis and other end-organ damage in SLE. These studies highlight the delicate fine-tuning of TLR7 signaling, where too little nucleic acid sensing may compromise IFN transcription and antiviral immunity, whereas excessive signaling may prevent negative selection of self-reactive B cells that bind chromatin-containing immune complexes and drive systemic autoimmunity (Fig. 1).

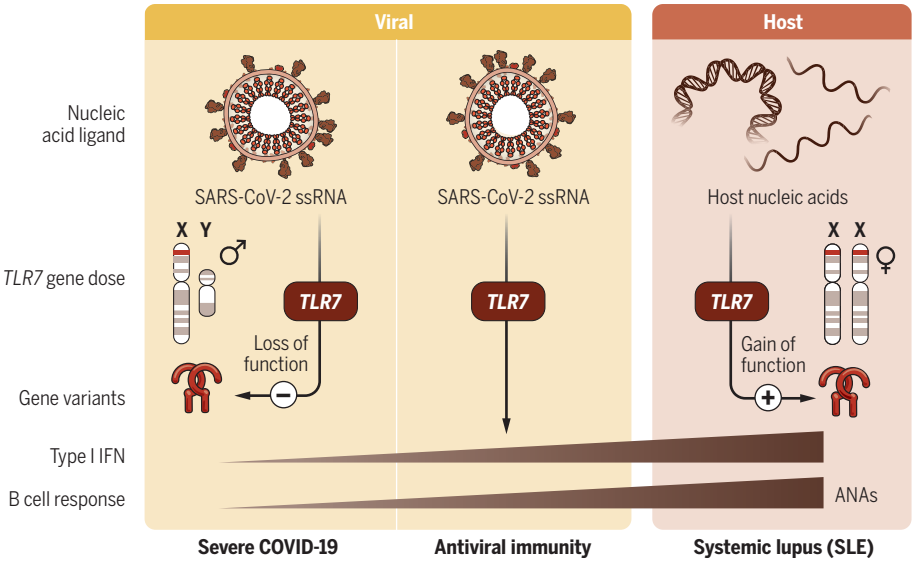


Fig. 1. *TLR7* gene dose and signaling strength affect immune function. Nucleic acid detection and signaling through TLR7 is a key player in both protection against viral infections and autoimmune responses. Loss or gain of TLR7 signaling because of genetic variants or gene copy number (chromosome XX versus XY, where the TLR7 locus escapes X inactivation) are risk factors for increased severity of SARS-CoV-2 infection and development of autoimmunity, respectively.

Endogenous nucleic acid sensing and autoimmunity

Genetic analyses have revealed that GOF mutations or copy number amplification of genes encoding nucleic acid sensors such as MDA5 and TLR7 (*56*) can trigger systemic autoimmune disease. These findings imply that there is a threshold above which nucleic acid sensing becomes pathogenic. By extension, such a threshold may also be exceeded by an overabundance of the ligands. Support for this premise stems from severe inherited autoimmune or inflammatory disorders that develop as a result of mutations in genes involved in nucleic acid metabolism, as recorded in the online catalog of human genes and genetic disorders (https://omim.org/). These include childhood-onset SLE, Aicardi–Goutières syndrome (AGS), familial chilblain lupus (FCL), and STING-associated vasculopathy with onset in infancy (SAVI) (Table 1).

AGS is an encephalopathy characterized by increased levels of IFN-α in the cerebrospinal fluid (CSF) and intracranial calcification. Despite phenotypic variability, many clinical features are similar to those found in viral infections in utero. Most patients present with

CREDIT: J. VINUESA/SCIENCE

congenital or acquired microcephaly, of which some also present with thrombocytopenia, liver complications, and/or vasculopathy as a common feature (*57*). A broad spectrum of autoantibody specificities has been identified in AGS patients with the most prevalent targets being sm/RNP, fibrinogen, and gliadin. About 43% of AGS patients experience recurrent chilblain lesions similar to those found in FCL. FCL also includes deregulated IFN I responses, and ~25% of patients have antinuclear antibodies (ANAs) (*58*).

In contrast to AGS and FCL, >95% of patients diagnosed with SLE have ANAs that include antibodies against dsDNA and DNA-associated targets. SLE is a heterogeneous disease, ranging from mild cutaneous to severe organ damage (e.g., kidney failure, pulmonary hypertension, and cardiac failure). Pathology is often driven by immune complex deposition in tissues leading to an autoimmune cascade and severe tissue damage. Anti-Ro, anti-La, and anti–Scl-70 autoantibodies, which are commonly found in autoimmune diseases such as Sjögren's syndrome and systemic sclerosis, can be detected in a subset of SLE and less frequently in AGS patients (*59*).

The source of nucleic acids appears to be a strong predictor of clinical phenotypes. For example, defects in extracellular deoxyribonucleases (DNases), such as DNASE1L*3*, cause classical SLE phenotypes in the absence of overt interferonopathy, whereas defects in cytosolic DNA and RNA digestion cause interferonopathies such as AGS and, less commonly, SLE (Fig. 2).

Extracellular and endosomal nucleic acid sensing and classical SLE

Extracellular DNases and ribonucleases (RNases) pla important roles in the removal of nucleic acids that may otherwise be immunogenic (*60*). In the absence of DNASE1, there is impaired degradation of NETs. Defects in DNASE1L*3* impair clearance of dsDNA from nucleosomes exposed on microparticles or apoptotic bodies (*14*, *43*). Uncleared nucleic acids signal through endosomal TLRs

Table 1. Diseases caused by gene variants in nucleic acid–sensing receptors and related proteins. Only autoimmune and autoinflammatory diseases recorded by the Online Mendelian Inheritance in Man (OMIM) catalog are included. The last column identifies genes for which there are alleles that either cause or have been suggested to contribute to chilblain lupus or SLE. In the sixth column, N/A indicates no disease-causing variants identified to date, and N/A (GOF) indicates no disease-causing GOF variants identified to date. Dashes indicate not applicable.

Gene	Protein	Change in function	Location	Ligand or partner	Autoimmune or autoinflammatory disease in OMIM	OMIM number	Lupus susceptibility alleles
DNASE1	DNASE1	*LOF*	Extracellular	dsDNA (NETs)	SLE	152700	Yes
DNASE1L3	DNASE1L3	*LOF*	Extracellular	Nucleosomes exposed on microparticles or apoptotic bodies	SLE 16	614420	Yes
TLR7	TLR7	*GOF*	Endosomal	ssRNA, 2′3′cGMP	SLE 17	301080	Yes
TLR9	TLR9	–	Endosomal	CpG dsDNA	N/A	–	No
MyD88	MyD88	–	Endosomal	TLRs	N/A (GOF)	–	No
ADA2	ADA2	*LOF*	Endolysosomal	Adenosine, 2′-deoxyadenosine	Sneddon syndrome	182410	No
					Vasculitis, autoinflammation, immunodeficiency, and hematologic defects syndrome (DADA2)	615688	
DNASE2	DNASE2	*LOF*	Lysosomal	Exogenous DNA	Autoinflammatory-pancytopenia syndrome	619858	No
TREX1	TREX1	*LOF*	Cytoplasmic	DNA (retroviral/ retrotransposon)	AGS 1	225750	Yes
					Chilblain lupus	610448	
					Vasculopathy, retinal, with cerebral leukoencephalopathy and systemic manifestations	192315	
					Susceptibility to SLE	152700	
TMEM173	STING	*LOF*	Cytoplasmic	cGAMP and excess DNA	STING-associated vasculopathy, infantile-onset	615934	Yes
CGAS	cGAS	–	Cytoplasmic	Cytosolic DNA	N/A (GOF)	–	No
SAMHD1	SAMHD1	*LOF*	Cytoplasmic	dNTPs	Chilblain lupus?	614415	Yes
					AGS 5	612952	
RNASEH2A	RNASEH2A	*LOF*	Cytoplasmic	RNA in RNA-DNA hybrids	AGS 4	610333	Yes
RNASEH2B	RNASEH2B	*LOF*	Cytoplasmic	RNA in RNA-DNA hybrids	AGS 2	610181	Yes
RNASEH2C	RNASEH2C	*LOF*	Cytoplasmic	RNA in RNA-DNA hybrids	AGS 3	610329	Yes
IFIH1	MDA5	*GOF*	Cytoplasmic	Long dsRNA	AGS 7	615846	Yes
					Singleton-Merten syndrome 1	182250	
RIGI	DDX58	*GOF*	Cytoplasmic	Short dsRNA	Singleton-Merten syndrome 2	616298	No
MAVS	MAVS	–	Cytoplasmic	–	N/A (GOF)	–	No
AIM2	AIM2	–	Cytoplasmic	Cytosolic dsDNA	N/A	–	No
IFI16	IFI16	–	Nuclear	Viral DNA or damaged self-DNA	N/A (GOF)	–	No

and can break B cell tolerance. In mice lacking *Dnase1l3*, anti-DNA responses and autoimmunity are driven by TLR7 and TLR9 redundantly and are only prevented by combined deletion of both endosomal receptors (*55*). A picture is emerging in which DNASE1L3 activity may be at the core of a large fraction of human SLE: More than 30% of all SLE patients and 50% of patients with lupus nephritis have neutralizing antibodies to DNASE1L3 and thus phenocopy genetic *DNASE1L3* deficiency (*61*).

Although an excess of DNA-containing immune complexes is expected to induce an IFN I response, this response is not marked in mice or humans with SLE caused by *DNASE1* or *DNASE1L3* variants and is unlikely to be the driver of B cell autoimmunity. Evidence for the latter comes from crosses of *Dnase1l3*-deficient mice to mice lacking IFN I receptors (*Ifnar1*$^{-/-}$). The emergence of anti-dsDNA is unaltered in these mice, and a significant decrease in autoantibody levels is only detected with age (*55*). This is reminiscent of the data from TLR7-driven autoimmune mice, which also develop autoantibodies independently of IFN I production (*47*, *48*). Rather than initiating the breach in B cell tolerance, IFN I more likely establishes a positive feedback loop by enhancing B cell survival and differentiation into antibody-producing cells (Fig. 3).

No *TLR9* gene variants have been associated with SLE in humans to date, and although this may be related to TLR9's ability to promote death of self-reactive B cells (*49*), the previous examples may also indicate the existence of a threshold for TLR7 activation by extracellular ligands. Such a threshold also appears to exist for intracellular ligands. For example, loss of purine nucleoside phosphorylase (PNP) activity causes imbalances in intracellular purine nucleoside levels, leading to accumulation of deoxyguanosine, a TLR7 ligand. In turn, this increases the sensitivity of TLR7 to endogenous ssRNA ligands and has been suggested as the underlying cause of systemic autoimmunity frequently reported in PNP-deficient patients (*62*). Studies of TLR trafficking to endolysosomes, a process regulated by UNC93B1, also points to a tunable threshold of TLR7 activation by endogenous ligands (*63*).

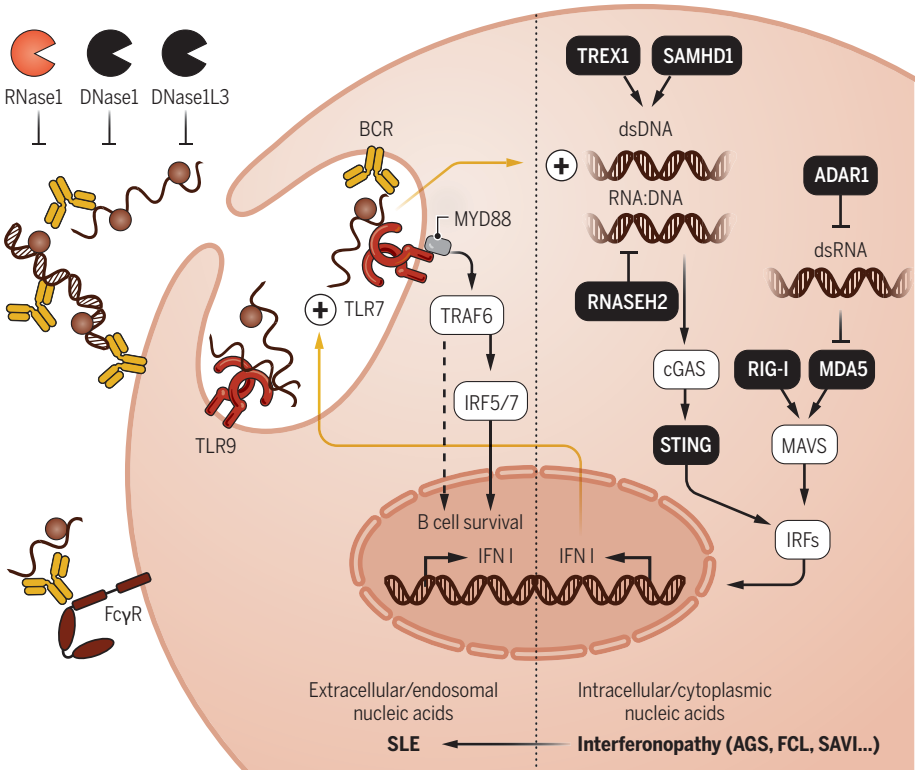


Fig. 2. Extracellular versus intracellular sources of nucleic acids primarily induce different IFN I–mediated pathologies. Exaggerated activation of TLR7/9 by endocytosed nucleic acids tends to induce autoimmune responses, such as SLE. Accumulation of cytoplasmic nucleic acids or increased sensitivity of nucleic acid sensors can trigger development of type I interferonopathies including AGS, FCL, and SAVI. IFN I induces the up-regulation of TLR7 expression and the release of nucleic acids from endolysosomes, and the rupture of endolysosomes can feed the pool of intracellular nucleic acids. Thus, these pathways bridge the intracellular and extracellular sources of nucleic acid. Enzymes in black boxes denote identified genetic risk factors for interferonopathy and/or SLE.

Intracellular nucleic acid sensing and interferonopathies

When defects occur in enzymes that digest intracellular DNA or RNA, the more prominent clinical presentations are interferonopathies, such as AGS, because of an overabundance of cytoplasmic nucleic acid intermediates that trigger the cGAS-STING pathway (DNA intermediates) or the MDA5-MAVS signaling cascade (RNA intermediates), leading to a robust IFN I response (Fig. 3). Loss of TREX1—the major cellular 3′-to-5′ DNA exonuclease that digests ssDNA or 3′-mismatched DNA—is thought to permit accumulation of cytosolic DNA from aberrant nuclear DNA transcription or replication intermediates, such as R-loops and micronuclear DNA, and of cDNA synthesized by LINE-1 reverse transcriptase (RT). Loss of the deoxynucleotide triphosphate (dNTP) triphosphohydrolase SAMHD1, which regulates cellular dNTP levels, allows reverse transcription and cDNA synthesis to occur. Loss of RNase H2 activity prevents the removal of RNA-DNA hybrids formed during DNA transcription and RNA reverse transcription. These three enzymes are additionally involved in DNA replication: TREX1 processes aberrant DNA replication intermediates, SAMHD1 promotes DNA end resection at stalled replication forks, and RNase H2 removes ribonucleosides misincorporated during DNA synthesis. The loss of such functions may also lead to an IFN I response.

Loss of dsRNA-specific adenosine deaminase activity of ADAR1 leads to accumulation of dsRNA, predominantly from inverted *Alu* repeats, whose natural complementarity would otherwise be disrupted by ADAR1-mediated adenosine-to-inosine (A-to-I) editing. Overabundance of unedited dsRNA activates the MDA5-MAVS signaling cascade, leading to an IFN I response. Inflammation is also triggered by mutations that seem to preserve the editing activity of ADAR1 but instead disrupt the function of the Zα domain, which is present in the ADAR1 p150 isoform and is responsible for binding left-handed double-helical (Z-form) dsRNA. Overabundance of Z-form dsRNA is thought to activate ZBP1, which senses Z-form DNA and RNA, ultimately inducing inflammatory cell death, as well as a MAVS-dependent IFN I response (*64*–*66*). Finally, there is increasing evidence that the cGAS-STING pathway may also respond to ligands originating from the endosomal-lysosomal pathway. Under certain conditions, membrane rupture may allow the escape of endosomal contents into the cytoplasm (*67*).

These observations paint a picture of carefully balanced sensitivity of sensors and availability of ligands, which can be tipped to excessive

CREDIT: LOLA GLORIOSO

signaling by deregulation or mutation of any one of the individual pathways (Fig. 3). Although these pathways may independently induce an IFN I response and instigate pathology, recent evidence suggests that they also intersect. For example, pathogenic 2′-deoxyguanosine 5′-triphosphate (dGTP) accumulation caused by elevated cytosolic deoxyguanosine in PNP deficiency is counterbalanced by SAMHD1, which catalyzes dGTP phosphohydrolysis, and lack of both PNP and SAMHD1 activities is lethal (*62*). Moreover, the IFN response triggered by SAMHD1 deficiency appears to require functional MDA5-mediated dsRNA sensing, linked with up-regulation of SINE transcription, that is primed by cGAS-STING sensing of cytosolic DNA (*68*). Although such intersection of metabolic pathways is cell intrinsic, the differential expression of sensors and ligands in distinct cell types allows for intersection of IFN I–inducing pathways at higher levels.

Cell type specificity of nucleic acid sensing

The host-protective function of IFN I relies on the induction of cell-intrinsic resistance to infection as well as on cellular communication that warns neighboring cells and orchestrates the immune response. Equally, the pathogenic effects of IFN I may well extend beyond the cell that produces it. Consequently, nucleic acid sensing in distinct cell types may weigh variably on the induction of autoimmunity depending on the type and magnitude of the response that it elicits. For example, TLR7 triggering induces substantial IFN I production by pDCs that is important to prevent severe COVID-19 (*12*). Although the survival signals elicited by TLR7 and/or TLR9 signaling in B cells appear indispensable for humoral immunity to self and certain foreign antigens (*49*), it is likely that pDC-, myeloid-, and B cell–intrinsic TLR7/9 signaling contribute synergistically to the development of systemic autoimmunity (*14*). Furthermore, the contribution of IFN I to autoimmunity and/or inflammation need not depend on TLR7 or on pDCs because aberrant nucleic acid sensing by a multitude of sensors may lead to overlapping phenotypes, as exemplified by AGS, FCL, and SLE. For example, studies in mice identified TREX1 deficiency in DCs as necessary for the induction of IFN I and inflammation (*69*). Thus, individual cell types may use different sensors to respond to aberrantly produced cell type–specific ligands, working in concert to instigate autoimmunity. A synergistic effect may also be facilitated by IFN I–inducible expression of certain sensors, including TLR7 in B cells, and endogenous ligands, providing a positive feedback loop (Fig. 2). IFN I produced by sensing one ligand in one cell may up-regulate the sensor for a different ligand on another cell. Negative feedback loops may also operate, as suggested

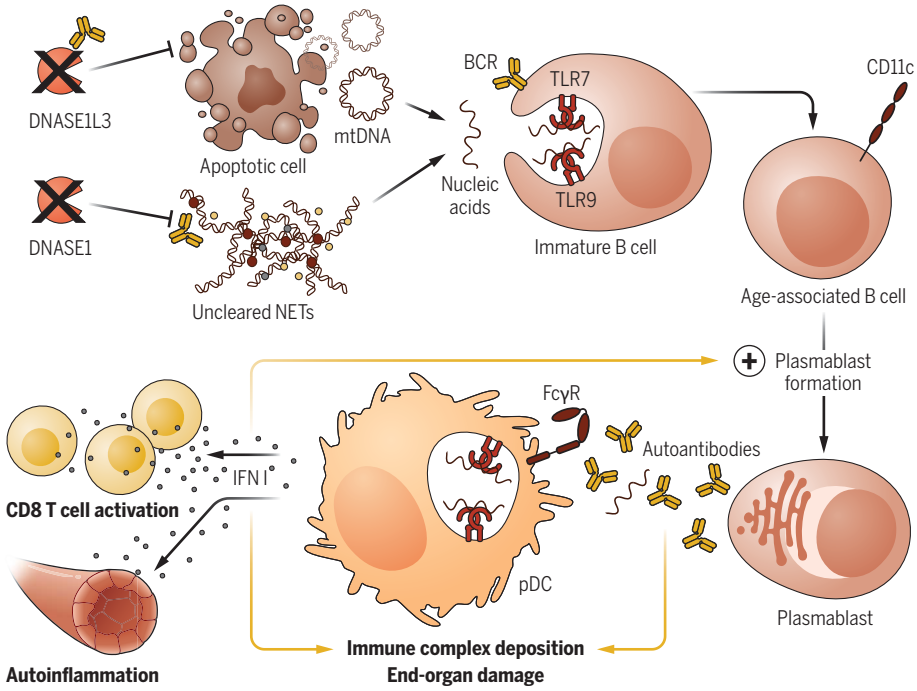


Fig. 3. DNASE1 or DNASE1L3 deficiency drives autoimmune responses and establishes a positive feedback loop. Lack of extracellular DNAse function (through genetic defects or antibodies targeting DNASE1L3) and inhibition of DNAse accessibility because of autoantibody formation cause impaired clearance of dsDNA from apoptotic cells and NETs (by DNASE1L3 and DNASE1, respectively). Along with mitochondrial DNA (mtDNA), an excess of these nucleic acid sources can promote expansion and differentiation of nuclear antigen–specific B cells. Plasmablast-derived autoantibodies drive IFN I production by pDCs through engagement with FcγR receptors. Excessive IFN I activates CD8 T cells and can cause autoinflammatory manifestations, such as vasculitis. A positive feedback loop is established through the promotion of extrafollicular B cell responses and plasmablast formation by IFN I. Autoantibodies and IFN I contribute to deposition of immune complexes and end-organ damage.

by a role for the IFN-inducible sensor AIM2, intrinsically in regulatory T cells in restraining autoimmunity (*70*).

To minimize the risk of autoimmunity, sensor sensitivity would require evolutionary tuning according to the level of endogenous or ubiquitous ligand availability. Such adaptations typically create a spectrum of sensitivities, with some leading to basal or tonic signaling in response to endogenous ligands, particularly from endogenous retroelements. Severe but rare genetic conditions, such as AGS, underscore the pathogenic potential of excessive sensor sensitivity or ligand abundance. However, recent quantitative trait locus analyses have identified variants that reduce ADAR1-mediated editing of immunogenic dsRNA, including from inverted *Alu* repeats, associated with increased risk of more common autoimmune and immune-mediated diseases (*71*), highlighting a subtler but more common contribution.

Conclusions and future directions

The underpinning of humoral immunity by innate immune signals is essential for host protection against infection but may also contribute to the development of systemic autoimmunity. Potentially important therapeutic implications can be drawn from this evolutionary trade-off. Although blockade of cGAS or IFN I signaling may be effective for the treatment of end-organ damage and general inflammation in SLE—particularly in patients with a high IFN gene signature (*72*)—this approach is unlikely to eliminate pathogenic autoantibodies that drive ongoing systemic autoimmune disease. IFN I blockade is also likely to increase susceptibility to viral infections. For example, SLE patients receiving anifrolumab (a human monoclonal antibody to IFN I receptor subunit 1) exhibited a more than sixfold and a more than threefold increase in shingles (herpes zoster) and bronchitis, respectively (*72*). In SLE, autoantibody production largely depends on TLR7 signaling in B cells; thus, effective therapies would require either elimination of pathogenic B cells or dampening of TLR7 signaling. A recent study of CAR-T cell therapy directed at CD19-expressing B cells has described impressive results (*73*) in a small number of patients with refractory SLE, suggesting that complete removal of the pathogenic B cell subset(s) is key for the success of B cell–depleting therapies.

CREDIT: K. HOLOSKI/*SCIENCE*

In nonrefractory disease, selective inhibition of TLR7 signaling in B cells, if achievable, could prevent autoantibody responses while maintaining B cells and their ability to mount protective antiviral antibody responses as well as preserve protective T cell immunity that requires TLR7 signaling in DCs.

It is now apparent that innate immune signals continue to fuel disease. Once induced, B cell autoimmunity may in turn promote further innate immune activation by enhancing neutrophil activation, DC maturation, and IFN I release by pDCs, mediated at least in part by the binding of immune complexes to Fc receptors on these cell types. This feedback, as well as additional functions of B cells, such as antigen presentation to autoreactive T cells, completes a cycle of innate-adaptive cross-talk that propagates autoimmunity. Recent genetic studies have demonstrated that this cycle can be set in motion by imbalances or immune dysregulation at any entry point. An important question for future work is whether the supporting role of innate immune stimulation continues to be important for disease perpetuation or if it becomes redundant once fulminant systemic B cell autoimmunity develops. Studies in a murine *Tlr7*-overexpression model of SLE have indicated a substantial contribution of IFN I, produced by pDCs in response to *Lactobacillus*, to disease severity, offering a therapeutic opportunity (*74*). Moreover, treatment of AGS patients with a combination of nucleoside analog reverse-transcriptase inhibitors, intended to prevent endogenous retroelement replication, significantly reduced ISG expression (*75*). Thus, different combinations of IFN I receptor, TLR7, and reverse-transcriptase inhibitors, with or without B cell depletion, may be required for optimal therapy of patients with systemic autoimmunity.

REFERENCES AND NOTES

1. M. Fumagalli *et al*., *PLOS Genet.* **7**, e1002355 (2011).
2. P. Gerdes, S. R. Richardson, D. L. Mager, G. J. Faulkner, *Genome Biol.* **17**, 100 (2016).
3. E. K. Karlsson, D. P. Kwiatkowski, P. C. Sabeti, *Nat. Rev. Genet.* **15**, 379–393 (2014).
4. J. Klunk *et al*., *Nature* **611**, 312–319 (2022).
5. P. M. V. Martin, E. Martin-Granel, *Emerg. Infect. Dis.* **12**, 976–980 (2006).
6. J. Piret, G. Boivin, *Front. Microbiol.* **11**, 631736 (2021).
7. The International HIV Controllers Study, *Science* **330**, 1551–1557 (2010).
8. R. Thomas *et al*., *Nat. Genet.* **41**, 1290–1294 (2009).
9. R. Apps *et al*., *Science* **340**, 87–91 (2013).
10. A. R. Everitt *et al*., *Nature* **484**, 519–523 (2012).
11. H. Ye *et al*., *Arthritis Res. Ther.* **21**, 70 (2019).
12. T. Asano *et al*., *Sci. Immunol.* **6**, eabl4348 (2021).
13. A. Decout, J. D. Katz, S. Venkatraman, A. Ablasser, *Nat. Rev. Immunol.* **21**, 548–569 (2021).
14. S. Caielli, Z. Wan, V. Pascual, *Annu. Rev. Immunol.* **41**, 10.1146/annurev-immunol-101921-042422 (2023).
15. V. Hornung *et al*., *Nature* **458**, 514–518 (2009).
16. N. Kerur *et al*., *Cell Host Microbe* **9**, 363–375 (2011).
17. B. Antiochos *et al*., *eLife* **11**, e72103 (2022).
18. L. J. Thompson, G. A. Kolumam, S. Thomas, K. Murali-Krishna, *J. Immunol.* **177**, 1746–1754 (2006).
19. E. Duong *et al*., *Immunity* **55**, 308–323.e9 (2022).
20. M. M. Thomsen *et al*., *Eur. J. Immunol.* **49**, 2111–2114 (2019).
21. M. M. Thomsen *et al*., *Med. Microbiol. Immunol.* **208**, 869–876 (2019).
22. S.-Y. Zhang *et al*., *Science* **317**, 1522–1527 (2007).
23. H. K. Lim *et al*., *J. Exp. Med.* **216**, 2038–2056 (2019).
24. X. Solanich *et al*., *Front. Immunol.* **12**, 719115 (2021).
25. G. Butler-Laporte *et al*., *PLOS Genet.* **18**, e1010367 (2022).
26. T. P. Velavan *et al*., *EBioMedicine* **72**, 103629 (2021).
27. Q. Zhang *et al*., *Science* **370**, eabd4570 (2020).
28. P. Bastard *et al*., *Science* **370**, eabd4585 (2020).
29. M. Souyris *et al*., *Sci. Immunol.* **3**, eaap8855 (2018).
30. P. Yu *et al*., *Immunity* **37**, 867–879 (2012).
31. G. R. Young *et al*., *Nature* **491**, 774–778 (2012).
32. A. Perl, D. Fernandez, T. Telarico, P. E. Phillips, *Curr. Opin. Rheumatol.* **22**, 483–492 (2010).
33. E. B. Chuong, N. C. Elde, C. Feschotte, *Science* **351**, 1083–1087 (2016).
34. J. Attig, G. R. Young, J. P. Stoye, G. Kassiotis, *Front. Microbiol.* **8**, 2489 (2017).
35. M. Tokuyama *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12565–12572 (2018).
36. K. W. Ng *et al*., *eLife* **8**, e50256 (2019).
37. H. E. Volkman, D. B. Stetson, *Nat. Immunol.* **15**, 415–422 (2014).
38. G. Kassiotis, *Annu. Rev. Immunol.* **41**, 10.1146/annurev-immunol-101721-033341 (2023).
39. X. Yin *et al*., *Ann. Rheum. Dis.* **80**, 632–640 (2021).
40. J. Bentham *et al*., *Nat. Genet.* **47**, 1457–1464 (2015).
41. M. Tokuyama *et al*., *J. Exp. Med.* **218**, e20191766 (2021).
42. S. Caielli *et al*., *Cell* **184**, 4464–4479.e19 (2021).
43. B. Lazzaretto, B. Fadeel, *J. Immunol.* **203**, 2276–2290 (2019).
44. A. L. DeFranco, D. C. Rookhuizen, B. Hou, *Immunol. Rev.* **247**, 64–72 (2012).
45. C. R. Ruprecht, A. Lanzavecchia, *Eur. J. Immunol.* **36**, 810–816 (2006).
46. S. Fillatreau, B. Manfroi, T. Dörner, *Nat. Rev. Rheumatol.* **17**, 98–108 (2021).
47. S. B. Chodisetti *et al*., *J. Immunol.* **204**, 796–809 (2020).
48. S. W. Jackson *et al*., *J. Exp. Med.* **213**, 733–750 (2016).
49. K. Nündel *et al*., *J. Immunol.* **194**, 2504–2512 (2015).
50. N. Shen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15838–15843 (2010).
51. S. A. Jenks *et al*., *Immunity* **49**, 725–739.e6 (2018).
52. G. J. Brown *et al*., *Nature* **605**, 349–356 (2022).
53. C. Soni *et al*., *J. Immunol.* **193**, 4400–4414 (2014).
54. A. V. Rubtsov *et al*., *Blood* **118**, 1305–1315 (2011).
55. C. Soni *et al*., *Immunity* **52**, 1022–1038.e7 (2020).
56. O. Omarjee *et al*., *Autoimmun. Rev.* **18**, 102361 (2019).
57. Y. J. Crow, N. Manel, *Nat. Rev. Immunol.* **15**, 429–440 (2015).
58. C. Fiehn, *Curr. Rheumatol. Rep.* **19**, 61 (2017).
59. M. A. Ameer *et al*., *Cureus* **14**, e30330 (2022).
60. P. Mistry, M. J. Kaplan, *Clin. Immunol.* **185**, 59–73 (2017).
61. J. Hartl *et al*., *J. Exp. Med.* **218**, e20201138 (2021).
62. E. R. Abt *et al*., *J. Clin. Invest.* **132**, e160852 (2022).
63. O. Majer, B. Liu, L. S. M. Kreuk, N. Krogan, G. M. Barton, *Nature* **575**, 366–370 (2019).
64. H. Jiao *et al*., *Nature* **607**, 776–783 (2022).
65. N. W. Hubbard *et al*., *Nature* **607**, 769–775 (2022).
66. R. de Reuver *et al*., *Nature* **607**, 784–789 (2022).
67. R. E. Gordon, J. F. Nemeth, S. Singh, R. B. Lingham, I. S. Grewal, *Trends Biotechnol.* **39**, 298–310 (2021).
68. T. Schumann *et al*., *J. Exp. Med.* **220**, e20220829 (2023).
69. K. Peschke *et al*., *J. Immunol.* **197**, 2157–2166 (2016).
70. W.-C. Chou *et al*., *Nature* **591**, 300–305 (2021).
71. Q. Li *et al*., *Nature* **608**, 569–577 (2022).
72. E. F. Morand *et al*., *N. Engl. J. Med.* **382**, 211–221 (2020).
73. A. Mackensen *et al*., *Nat. Med.* **28**, 2124–2132 (2022).
74. D. F. Zegarra-Ruiz *et al*., *Cell Host Microbe* **25**, 113–127.e6 (2019).
75. G. I. Rice *et al*., *N. Engl. J. Med.* **379**, 2275–2277 (2018).

ACKNOWLEDGMENTS

Funding: This work was supported by the Francis Crick Institute (CC2228 and CC2088), which receives its core funding from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant no. 101018670), the Lupus Research Alliance (LRA) through the Global Team Science Award, and an EMBO postdoctoral fellowship (ALTF no. 654-2022). **Author contributions:** C.G.V., A.G., and G.K. researched the literature and wrote the manuscript. **Competing interests:** G.K. is a scientific cofounder of EnaraBio and a member of its scientific advisory board. G.K. has consulted for EnaraBio and Repertoire Immune Medicines. The authors declare no other competing interests. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 31 January 2023; accepted 29 March 2023
10.1126/science.adg6427

REVIEW

Common genetic factors among autoimmune diseases

Adil Harroud[1,2,3]* and David A. Hafler[4,5,6]*

Autoimmune diseases display a high degree of comorbidity within individuals and families, suggesting shared risk factors. Over the past 15 years, genome-wide association studies have established the polygenic basis of these common conditions and revealed widespread sharing of genetic effects, indicative of a shared immunopathology. Despite ongoing challenges in determining the precise genes and molecular consequences of these risk variants, functional experiments and integration with multimodal genomic data are providing valuable insights into key immune cells and pathways driving these diseases, with potential therapeutic implications. Moreover, genetic studies of ancient populations are shedding light on the contribution of pathogen-driven selection pressures to the increased prevalence of autoimmune disease. This Review summarizes the current understanding of autoimmune disease genetics, including shared effects, mechanisms, and evolutionary origins.

Autoimmune diseases account for considerable population morbidity, particularly among young adults. Collectively, their lifetime prevalence reaches up to 9.4% in the United States (*1*) and up to 10.6% in a recent study of 78 autoimmune diseases in Catalonia (*2*). This heterogeneous group of more than 80 clinically systemic (e.g., systemic lupus erythematosus) or organ-specific (e.g., multiple sclerosis) diseases share common features, such as T cell and antibody reactivity to self-antigens and common association with certain human leukocyte antigen (*HLA*) genetic variants. It is generally thought that autoimmune diseases develop as a result of a breakdown in immune tolerance and activation of autoreactive T cells by autologous or cross-reactive microbial antigens in genetically susceptible individuals. The underlying genetic pathogenesis is evidenced by the high disease concordance among identical twins, and associations with genes in the major histocompatibility complex (MHC) locus have been recognized for over half a century. More recently, the sequencing of the human genome, followed by an understanding of its haplotype structure and the advent of technologies allowing whole-genome analysis of overrepresented haplotypes in disease populations, have led to the elucidation of a substantial degree of disease heritability. Individual haplotypes associated with disease risk have relatively low effect sizes, with MHC odds ratios in the range of 2 to 3 (with notable exceptions, such as *HLA-B*27* in ankylosing spondylitis) and other loci with lesser odds ratios. However, as will be discussed below, the molecular phenotypes associated with these genotypes are notable, and risk genes have, in a nonbiased fashion, implicated the immune system as driving these diseases. Moreover, the common pathophysiology among these diseases is indicated by their shared heritability and genetic overlap.

"The overwhelming majority (>90%) of putative causal variants associated with autoimmune diseases are in noncoding regions of the genome..."

It can be broadly stated that the underlying causes of autoimmune disease are the unfortunate outcome of gene-environment interactions. Indeed, many variants tied to autoimmune diseases evolved to protect against infectious diseases but have come to be viewed as detrimental in current low-infectious environments, as further described below. Other environmental factors that increase the risk of developing autoimmune disease include diet (*3*), smoking (*4*), Epstein-Barr virus infection (*5*), and other, as-yet-unknown environmental influences. Although the identification of a meaningful proportion of genetic variants associated with autoimmune disease risk has been accomplished, elucidating the environmental factors and their interactions with gene function is considerably more difficult. However, this is necessary to determine the reasons for the reported rise in the prevalence of autoimmunity over the past decades (*6*). This m... of autoimmunity driven by autoreactive T cells has led to highly effective therapies. Nevertheless, a paradox remains in that an immunomodulatory therapy that is highly effective in one disease may trigger another autoimmune disease. The potential underlying genetic bases of this observation will also be discussed.

Genetic insights into autoimmune diseases

Autoimmune diseases consistently demonstrate a higher concordance rate among monozygotic twins compared with dizygotic twins (35% and 6%, respectively, for multiple sclerosis) (*7*). In addition, the strongest risk factor for many autoimmune diseases is a positive family history, yet most individuals with the disease have no affected relatives. These observations provided evidence for a substantial genetic component in the etiology of these conditions while posing questions about the expected number of contributing causal variants and their effect sizes. Early studies conducted linkage analysis in families of individuals with autoimmune diseases under the assumption that variants in a small number of genes would harbor large effect sizes and drive genetic risk, similar to single-gene Mendelian disorders. These approaches allowed the discovery of some of the first and strongest susceptibility loci for autoimmune diseases, such as the MHC for type 1 diabetes (*8*) and the nucleotide-binding oligomerization domain-containing protein 2 (*NOD2*) locus in Crohn's disease (*9*). However, these experimental designs were largely unsuccessful in elucidating the genetic architecture of complex diseases.

The development of genotyping array technology and the shift from family-based studies to genome-wide association studies (GWASs) involving large collections of cases and controls has revealed the polygenic nature of autoimmune diseases (*10*). Thousands of robust and replicable genetic associations with autoimmune diseases have been identified by comparing allele frequency between affected and unaffected individuals (Fig. 1). The number of detected loci increases linearly with sample size, whereas effect sizes become smaller, with an estimation that millions of individuals (10 million for inflammatory bowel disease) would be required to fully map the associated genomic regions (*11*). The additive effects of these predominantly common variants (present in 5% or more of the population) typically account for most of the heritability, that is, the fraction of phenotypic variation caused by genetic variation (*12*). Large-scale sequencing studies are also pinpointing rare (often coding) variants associated with autoimmune diseases (*13*). These rare variants typically have larger effect sizes and often converge on the same genes as common variants (*12*, *13*).

[1]Department of Neurology and Neurosurgery, McGill University, Montréal, Quebec, Canada. [2]Department of Human Genetics, McGill University, Montréal, Quebec, Canada. [3]The Neuro (Montreal Neurological Institute and Hospital), McGill University, Montréal, Quebec, Canada. [4]Department of Neurology, Yale School of Medicine, New Haven, CT, USA. [5]Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA. [6]Broad Institute of MIT and Harvard University, Cambridge, MA, USA.
*Email: david.hafler@yale.edu (D.A.H.); adil.harroud@mcgill.ca (A.H.)

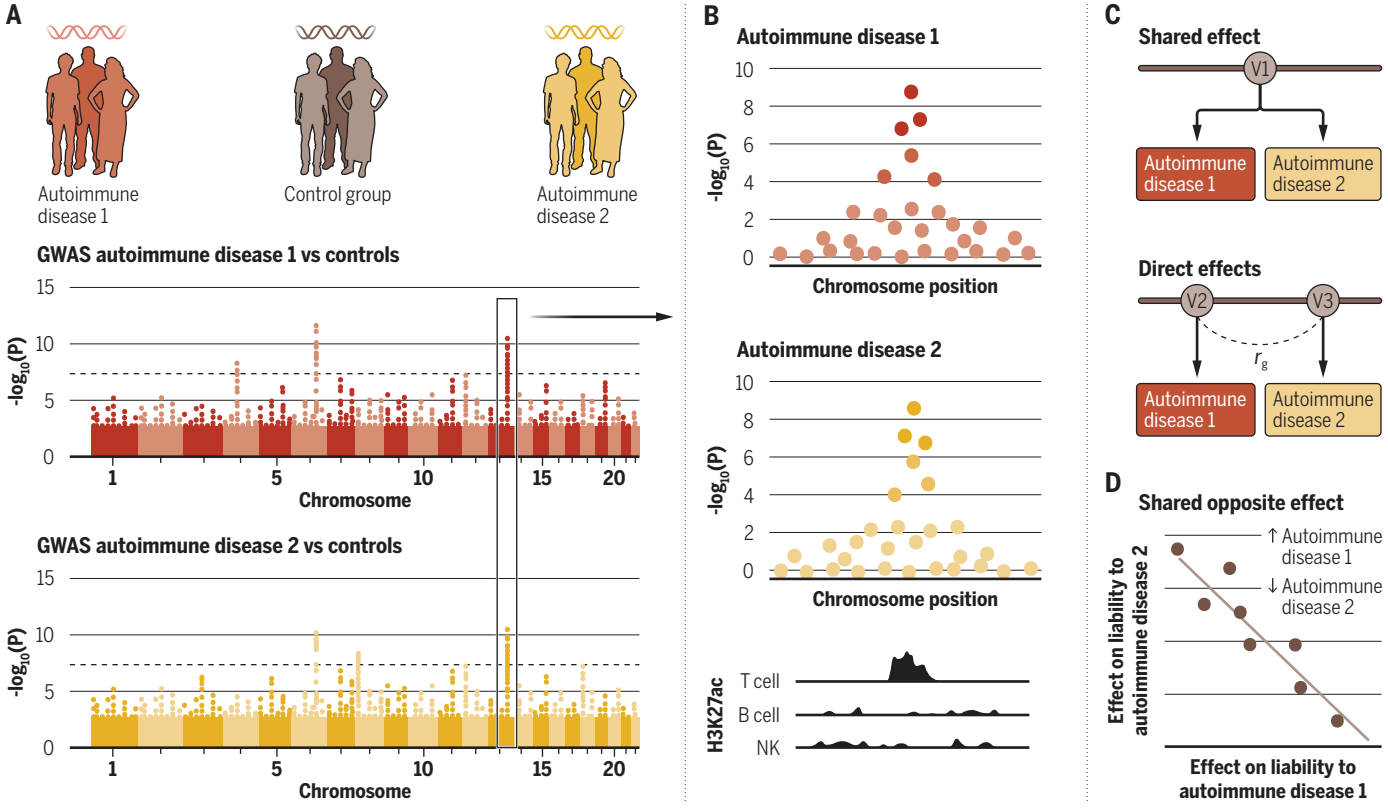


Fig. 1. Exploring shared genetic effects among autoimmune diseases. (**A**) Representation of case-control GWAS designs investigating the genetic susceptibility to two autoimmune diseases by comparing allelic frequencies between affected and unaffected individuals. Genotypes are obtained using microarrays and, increasingly, whole-exome or whole-genome sequencing. The Manhattan plots (bottom) show each variant represented by a point, with its genomic position on the *x* axis and the strength of its association on the *y* axis. The dotted line marks genome-wide significance (typically $P < 5 \times 10^{-8}$). The gray box highlights a locus with overlapping association signals between the two diseases. (**B**) Statistical methods for fine-mapping and colocalization can be applied to pinpoint likely causal variants and assess whether those are shared between the two diseases. These causal variants often localize to enhancers in noncoding regions marked by H3K27ac. These chromatin features (among other approaches) can also be used to implicate disease-causing cell types, such as the T cells shown in this illustration. (**C**) It is important to differentiate between instances in which a single variant (V1) influences multiple diseases, indicating shared molecular mechanisms, and cases in which distinct variants (V2 and V3) at the same locus affect disease risk separately. (**D**) Shared variants often increase or decrease disease risk similarly, although they may occasionally have opposing effects, as represented by the negative correlation in variant coefficients at the shared locus. r_g, genetic correlation.

Evidence from whole-genome sequencing studies in a few complex traits suggests that rare variants (allele frequency <1%) are an important source of heritability not captured by common variants (*14*). Other potential contributors to heritability include gene-environment interactions [such as with the microbiome (*15*)] and gene-gene interactions (*16*).

The overwhelming majority (>90%) of putative causal variants associated with autoimmune diseases are in noncoding regions of the genome, consistent with the observation that <2% of the human genome encodes proteins (*17*). Elucidating the biological consequences of these noncoding variants is necessary for translating genetic discoveries into the clinic, but this presents a number of challenges. The presence of many correlated variants (linkage disequilibrium) at loci implicated by GWASs complicates the identification of causal variants, which can be multiple in a single locus (*18*). Furthermore, even when causal variants are identified with reasonable confidence, such as through fine-mapping (*19*), the functional mechanisms, target genes, and cell types often remain elusive.

Over the past decade, the integration of variant-disease associations with gene regulation traits (such as gene expression, splicing, and chromatin phenotypes) has begun to shed light on the activity of these variants and their cellular contexts. For instance, a study of 21 autoimmune diseases revealed that ~60% of candidate causal variants mapped to enhancers and clusters of enhancer elements (so-called superenhancers) marked by H3 lysine 27 acetylation (H3K27ac) in $CD4^+$ T cell populations and B lymphoblastoid cells (*17*). Candidate causal variants were further enriched within stimulus-dependent enhancers, including those producing regulatory noncoding enhancer RNA (*17*). These observations suggest that most autoimmune disease–associated variants act through transcriptional regulatory mechanisms [such as through altered nuclear factor κB (NF-κB) transcription factor binding] that modulate immune cell fate and function. To illustrate this point, the Crohn's disease variant rs61839660 was mapped by CRISPR activation to a stimulation-dependent intronic interleukin-2 (IL-2) receptor alpha (*IL2RA*) enhancer, and the risk allele was found to impair the timing of *IL2RA* expression in naïve T cells upon stimulation (*20*). IL-2 signaling is essential for the maintenance of forkhead box P3–expressing ($FOXP3^+$) regulatory T cells, which suppress autoimmune responses. The *IL2RA* locus has long been known to harbor variants tied to multiple autoimmune diseases, some with opposite effects on different diseases (*21*).

Overlaying disease-associated variants onto cell type–specific regulatory elements and regions of open chromatin can also be used to

implicate relevant cell types in a disease of interest (Fig. 1). When applied to autoimmune diseases, nearly all variants preferentially mapped to $CD4^+$ T cell subpopulations (*17*, *22*). Some diseases, including multiple sclerosis and systemic lupus erythematosus (*22*), showed additional specificity for B cells, which is consistent with the use of B cell–targeted therapy for the treatment of these conditions (*7*, *23*). Additional commonly identified cell types include natural killer cells, dendritic cells, and mononuclear phagocytes. Other approaches using genome-wide summary statistics (instead of disease-associated loci) annotated on the basis of relative gene expression across different tissues or cell types have yielded similar findings (*24*). Additionally, recent genetic studies have implicated disease-specific cell types, such as mesenchymal cells in Crohn's disease (*13*) and microglia in multiple sclerosis (*25*).

To inform the biological interpretation of disease-linked variants, a common approach is to assess their effect on intermediate gene regulation traits from quantitative trait locus (QTL) studies, which examine the effect of genetic variants on gene expression, splicing, methylation, and chromatin phenotypes, among other molecular phenotypes in specific tissues and cell types. For instance, the integration of GWASs with expression QTLs (eQTLs) can prioritize causal genes in a cell type–specific manner, although only a minority (~25%) of autoimmune disease–associated loci demonstrate strong evidence of colocalization with eQTL effects (*26*). This gap is partly, but not entirely (*27*), explained by data limitations, including the lack of various immune cell types (particularly those less abundant or less well characterized), the limited level of subset resolution, and the lack of dynamic variation (*28*), which is critical given the enrichment of autoimmune disease variants in stimulus-dependent enhancers. This is consistent with a recent study that identified eQTLs active exclusively in immune cells from autoimmune patients, potentially as a result of in vivo stimulation, leading to a high degree of colocalization (63%) with systemic lupus erythematosus risk variants (*29*).

The use of multiplexed single-cell RNA sequencing for eQTL mapping holds promise for further improved cell-type resolution and identification of dynamic regulatory effects. This approach was recently used to characterize the transcriptome and genetic variation across a total of 1,267,758 peripheral blood mononuclear cells from 982 healthy individuals (*30*). The study found that most cis-eQTLs were specific to one of 14 immune cell types. Integration with GWAS risk variants for seven autoimmune diseases also identified 117 non-MHC loci where a disease-associated variant exerted a causal effect through gene expression, most of which (65%) were also cell-type specific. However, because of the differences in power across cell types driven by varying cell proportions and the number of identifiable cells per individual, it is likely that with larger sample sizes, a proportion of eQTLs currently identified as being cell-type specific will be detected in other cell types, potentially with different effect magnitudes. Within a given cell type, single-cell analysis of cell states across differentiation and activation further revealed dynamic eQTLs enriched for colocalization with autoimmune disease loci (*28*, *30*, *31*).

GWASs have identified thousands of genetic variants associated with susceptibility to various autoimmune diseases, most of which are noncoding and reside in stimulation-dependent enhancer elements. These variants may directly disrupt enhancer function, as evidenced by the *IL2RA* example, although their potential consequences are myriad. The integration of disease loci with growing maps of regulatory annotations is rapidly improving the interpretability of those noncoding variants and enabling the generation of precise hypotheses for functional follow-up (*32*). The insights that have emerged so far underscore the importance of cell-type and context specificity and the necessity to further expand the characterization of regulatory elements to encompass rare yet immunologically meaningful cell types and diverse physiological and pathological cell conditions. In light of these challenges and the large number of autoimmune disease–associated variants, it is perhaps unsurprising that downstream functional mechanisms have only been uncovered for a small fraction (Box 1). Finally, the recent development of massively parallel regulatory assays that can test tens of thousands of synthetic noncoding regulatory sequences for functional effects in relevant cell types has the potential to substantially accelerate the discovery of functional mechanisms for autoimmune disease variants (*18*, *33*).

Box 1. From genotype to molecular phenotype.

Understanding how genetic haplotypes dictate biology is the major goal of immunogenetics. Although risk haplotypes have small effects on disease risk, the phenotypes associated with these haplotypes can be substantial. An example is the transcription factor NF-κB, the central regulator of inflammation. GWASs in many autoimmune diseases have identified variants in genes encoding members of the NF-κB signaling cascade, and variants associated with increased risk for multiple sclerosis and ulcerative colitis are strongly enriched within binding sites for NF-κB (*58*, *59*). A multiple sclerosis–associated variant (rs228614) proximal to *NFKB1* was associated with increased NF-κB signaling after TNF-α stimulation and increased degradation of the NF-κB inhibitor. This variant controls signaling responses by altering the expression of NF-κB itself, with homozygous risk allele carriers expressing 20-fold more p50 NF-κB than noncarriers (*58*). NF-κB activation has also been implicated in systemic lupus erythematosus. Two associated variants (rs148314165 and rs200820567) located downstream of TNF alpha-induced protein 3 (*TNFAIP3*), which encodes a negative regulator of NF-κB, were mapped to an enhancer element and found to impair looping interaction with the *TNFAIP3* promoter, resulting in its reduced expression (*59*). The shared consequence of these genetic effects is enhanced NF-κB pathway activity and predisposition to autoimmune disease.

More recently, the immunomodulatory role of a protective variant for multiple sclerosis [rs148755202, which encodes an Arg^{166}→His missense mutation (R166H)] in histone deacetylase 7 (HDAC7) was examined (*60*). Transcriptomic analyses demonstrated that wild-type HDAC7 regulates genes essential for the function of $FOXP3^+$ regulatory T cells (T_{regs}), an immunosuppressive subset of $CD4^+$ T cells that are dysfunctional in patients with multiple sclerosis (*7*). T_{regs} transduced with the protective HDAC7 R166H variant exhibited higher suppressive capacity in in vitro functional assays, mirroring phenotypes previously observed in patients. Moreover, in vivo modeling of the human HDAC7 R166H substitution by the generation of a knock-in mouse model bearing an orthologous R150H mutant demonstrated decreased experimental autoimmune encephalitis severity linked to transcriptomic alterations of brain-infiltrating T_{regs}. Thus, genetic alterations in epigenetic modifiers, a molecular class suitable for therapeutic interventions, can mediate protection from autoimmunity.

Shared heritability and common mechanisms

Different autoimmune diseases co-occur in individuals at higher rates than would be expected given their individual prevalence, and multiple diseases potentially affecting different organ systems cluster in families (*34*). For instance, individuals having a parent with celiac disease harbored a 2.7 relative risk for systemic lupus erythematosus compared with those with unaffected parents in a nationwide Swedish registry (*34*). These observations, initially clinical and then supported by epidemiological evidence, suggested shared risk factors. As hundreds of loci associated with autoimmune diseases were discovered and their genetic architecture better understood, genetic overlaps between disorders indeed became evident.

Clustering on the basis of genetic risk loci highlights a rich network of correlations among

autoimmune diseases, more so than among diseases that affect the same organ systems or that are based on clinical similarity (*17*, *35*). A genome-wide examination of genetic effects, as opposed to statistically significant risk loci, reveals widespread evidence of shared heritability among autoimmune diseases; the pairwise genetic correlation (r_g) averaged 0.39 across a set of five chronic autoimmune disorders and was highest between Crohn's disease and ulcerative colitis (r_g = 0.78) (*36*). These estimates exclude the MHC, in which overlapping haplotypes are among the strongest risk factors for multiple autoimmune diseases. Moreover, the presence of shared genetic loci with opposite effects, increasing risk for one condition while decreasing risk for another (Box 2), can drive negative local genetic correlation and attenuate the overall genome-wide genetic correlation (*37*).

Identifying shared genetic effects and resolving their functional implications can uncover common immunopathogenic mechanisms and inform the development or repurposing of rational therapies. It is therefore noteworthy that two-thirds of the variants associated with each of 21 autoimmune diseases were in overlapping loci (*17*). However, proximity within a locus does not necessarily indicate shared underlying mechanisms (Fig. 1). The locus may instead harbor distinct causal variants associated with different diseases. This complexity is compounded by the frequent occurrence of multiple conditionally independent causal variants at a single locus for a given disease. To address these challenges, various colocalization methods have been used to identify instances of true shared molecular effects in which a single variant influences the risk of two or more diseases (*26*, *36*, *38*–*40*). These studies frequently leveraged cohorts genotyped on the Immunochip, a custom array with 196,000 variants that densely covers 186 regions associated with at least one autoimmune condition. Their main findings have been consistent: Approximately half (41 to 60%) of the genetic loci that overlap between two or more autoimmune diseases are attributable to the same underlying genetic effect (Fig. 2). This includes loci with multiple conditionally independent associations, such as signal transducer and activator of transcription 4 (*STAT4*), which harbors two independent genetic effects common to both rheumatoid arthritis and systemic lupus erythematosus (*38*). Most of these shared genetic effects are observed between pairs of diseases, with <10% occurring across four or more autoimmune conditions (*38*). A notable example is a single low-frequency missense coding variant (rs34536443) in *TYK2*, which encodes a nonreceptor tyrosine kinase that is constitutively expressed across immune cell subsets. This variant confers strong uniform protection against the 10 autoimmune conditions in which it was examined (*41*). Homozygosity for the minor allele resulted in disease odds ratios between 0.1 and 0.3 and in a near-complete loss of protein function. This in turn led to impaired signaling for type 1 interferon, IL-12, and IL-23 and consequently perturbation of $CD4^+$ T helper (T_H) type 1 and T_H17 cytokine production (*41*). This study suggested that a drug capable of similarly inhibiting TYK2 function may be beneficial across a range of autoimmune conditions. Indeed, the first selective tyrosine kinase 2 (TYK2) inhibitor, deucravacitinib, was approved by the US Food and Drug Administration (FDA) in September 2022 for the treatment of plaque psoriasis, with ongoing trials in psoriatic arthritis, systemic lupus erythematosus, Crohn's disease, and ulcerative colitis. On the basis of this genetic evidence, TYK2 inhibitors may yet be extended to additional autoimmune conditions. Homozygosity for the same missense variant (rs34536443) also confers a higher risk of mycobacterial disease, including primary tuberculosis, driven by IL-23 disruption (*42*). Therefore, evaluation for active and latent tuberculosis is recommended before initiating deucravacitinib.

These data reveal the existence of a substantial genetic overlap between autoimmune diseases, as evidenced by the presence of pleiotropic variants shared across multiple diseases, occasionally with opposing effects. By leveraging this genetic overlap, cross-disease analysis of existing autoimmune disease cohorts has

Fig. 2. Shared risk variants in autoimmune diseases. This chord diagram shows the pairwise shared genetic associations driven by the same underlying allele in an Immunochip analysis of six autoimmune diseases. Data are from (*32*). The number of pairwise shared effects for each disease is indicated in parentheses. CeD, celiac disease; IBD, inflammatory bowel disease; MS, multiple sclerosis; T1D, type 1 diabetes; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

Box 2. Genetic basis of treatment outcomes and differential effects in autoimmune diseases.

Inhibitors of TNF-α are widely used and highly effective treatments for a variety of autoimmune diseases, including rheumatoid arthritis, inflammatory bowel disease, and ankylosing spondylitis, with the notable exception of multiple sclerosis. Preclinical studies in an experimental autoimmune encephalomyelitis model suggested that TNF-α depletion would be beneficial for multiple sclerosis. However, a phase 2 randomized trial of TNF-α capture using lenercept, a recombinant TNFα receptor p55-immunoglobulin fusion protein, resulted in dose-dependent disease worsening, leading to early trial discontinuation (*61*). A subsequent GWAS identified a variant (rs1800693) intronic to TNF receptor superfamily member 1A (*TNFRSF1A*). Functional experiments revealed that the risk variant led to exon 6 skipping and premature transcription termination in ~10% of mRNAs, resulting in loss of the transmembrane and intracellular domains and increased expression of a soluble TNF receptor isoform that neutralizes circulating TNF-α, similar to lenercept (*62*). Diseases that benefit from TNF inhibitor treatment either have no association with variants in *TNFRSF1A* (such as rheumatoid arthritis and inflammatory bowel disease) or an opposite association; the risk allele in multiple sclerosis correlates with a protective effect for ankylosing spondylitis of similar but opposite magnitude (*63*). Recently, epidemiological studies have also demonstrated that treatment with TNF-α inhibitors is associated with an increased incidence of multiple sclerosis (*64*).

This example highlights the potential of genetics to inform and predict treatment response and toxicity, including differential effects across diseases. It also illustrates that even though genetic effects may appear small (rs1800693 increases the odds of multiple sclerosis by 15%), pharmacological interventions targeting the same mechanisms can have a substantial impact on outcomes (lenercept increases the relapse rate in multiple sclerosis by up to 68%) (*61*).

been successful in identifying new genetic associations and improving causal variant identification (*36*, *38*). Compelling evidence suggests that some of these shared genetic factors, and consequently shared immunopathology, have evolutionary roots shaped by adaptation to pathogens.

Evolutionary origins of autoimmune diseases

Genetic diversity and variation are shaped by evolutionary forces, including negative selection, a process by which alleles with detrimental effects on fitness are removed from the population. Therefore, the presence of common genetic variants that substantially increase the risk of autoimmune diseases implies that these variants may have provided a beneficial evolutionary trade-off. Considering the immune system's role in combating infections and the strong impact of host genetics on susceptibility to infectious diseases, it has been proposed that autoimmune risk alleles have been preserved at high frequency in the population because of their role in improving resistance to infections (*43*). The canonical example of this antagonistic pleiotropy, genotypes with opposing effects on different traits, is variation of the hemoglobin subunit-β (*HBB*) locus that protects against malaria but causes sickle cell anemia when inherited in a recessive manner. Similarly, genetic variations in the MHC region, which is responsible for a large proportion of the inherited risk for autoimmune diseases, are expectedly also linked with susceptibility to various infections because of the roles of the encoded proteins in antigen presentation and T cell receptor composition (*44*). Recently, a systematic analysis of genetic effects on infectious and autoimmune disorders confirmed that variants associated with both trait categories were significantly more prevalent than expected [by >100-fold (*43*)] (Fig. 3).

Another example is the tumor necrosis factor (TNF) ligand superfamily member 13B (*TNFSF13B*) locus, which was found to be associated with the risk of multiple sclerosis and systemic lupus erythematosus in people from Sardinia (*45*). The causal variant at the locus (rs200748895) results in higher expression of B cell activating factor (BAFF, encoded by *TNFSF13B*) and in turn higher humoral immunity. This variant was several-fold more frequent in people from Sardinia compared with those in mainland Europe (26.5% in Sardinia versus 1.8% in the United Kingdom and Sweden), and the haplotype displayed signatures consistent with positive selection, as opposed to higher frequency arising from random chance (i.e., genetic drift). Because high BAFF expression and antibody production may protect against severe malarial disease, the authors proposed that this positive selection was driven by adaptation to malaria, which was endemic in Sardinia until the 1950s and more prevalent than in mainland Europe. Even though malaria has been eliminated from Sardinia, its genetic imprint may help to explain why the island population currently has some of the highest rates of multiple sclerosis and systemic lupus erythematosus. This example also highlights the potential benefits of studying genetic variation in isolated and diverse populations with varying haplotype frequencies and structures in informing disease biology (*46*).

Although previous evidence of the selective pressures imposed on human populations by specific pathogens has been largely circumstantial, recent advances in ancient DNA preparation and sequencing have provided a direct means of assessing human adaptation (*47*). For instance, a recent study examined the genetic adaptations to the Black Death, a devastating pandemic of plague caused by *Yersinia pestis* that killed 30 to 50% of the populations of Europe, the Middle East, and North Africa in the Middle Ages (*48*). To identify genetic loci under selection from the pandemic, the authors analyzed ancient DNA samples from individuals buried in London cemeteries before (850 to 1250 AD), during (1348 and 1349 AD), and after (1350 to 1539 AD) the Black Death. A cohort of individuals before and after the plague in Denmark provided replication. Overall, 206 individuals were analyzed for 356 immune-related genes and 496 immune disorder GWAS loci. The results showed large changes in allele frequency at immune loci, but not in regions under neutral evolution. The strongest evidence was found for a variant (rs2549794) in endoplasmic reticulum aminopeptidase 2 (*ERAP2*), which conferred an estimated 40% reduction in Black Death mortality in individuals homozygous for the protective allele. The same allele has been associated with increased risk of Crohn's disease, and the corresponding haplotype also increased the risk of other autoimmune diseases, including ankylosing spondylitis, psoriasis, and juvenile idiopathic arthritis. *ERAP2* encodes an aminopeptidase that adapts antigens for binding to MHC class I molecules and presentation to $CD8^+$ T lymphocytes. Likewise, another Black Death–protective allele near cytotoxic T lymphocyte–associated protein 4 (*CTLA4*) increases the risk for celiac disease, rheumatoid arthritis, and systemic lupus erythematosus. It therefore appears that the selection of advantageous alleles in survivors of the plague, and the preservation of their frequencies in descendants, have led to an increased risk of various autoimmune diseases in modern populations.

Just as past selection favored host resistance alleles, it has also selected against variants that weakened immune responses and increased the risk of infection, even when they offered protection against autoimmune diseases. Recent studies have revealed that the *TYK2* missense variant (rs34536443), which reduces the risk of various autoimmune conditions,

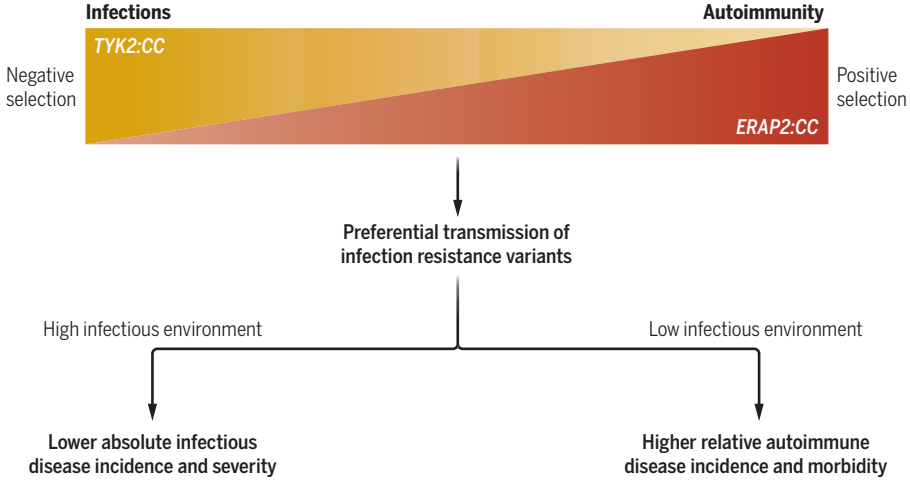


Fig. 3. Antagonistic pleiotropy between risk of autoimmunity and infection. Many genetic variants associated with autoimmune diseases also influence infectious disease risk and severity. Selective pressures from pathogens over millennia have favored variants, such as in *ERAP2*, that provide infection resistance even at the expense of higher autoimmunity (positive selection). Conversely, the same pressures have resulted in decreased frequency of variants, such as in *TYK2*, that protect against autoimmune diseases but increase infectious risk (negative selection). This suggests that the benefits of infection resistance outweigh the risk of autoimmune diseases, especially in environments with high rates of infection (most of human history). These evolutionary pressures have resulted in a higher genetic predisposition to autoimmune diseases, contributing to their high disease burden in modern societies relative to historically low infection rates.

underwent negative selection over the past 2000 years because of an increased predisposition to tuberculosis, resulting in a lower frequency in the population (*43*). This variant also increased the risk of severe COVID-19, consistent with the role of *TYK2* in the type 1 interferon pathway implicated in this form of the disease (*49*).

Throughout history, human immune systems have adapted in response to environmental pressures, perhaps none greater than those imposed by infectious agents. Their genetic imprint favoring more robust immune responses came at the cost of a higher shared genetic risk of various autoimmune diseases. The transition from high levels of infection to lower pathogen exposure in recent times has led to a mismatch between this history of selection and our current environments, contributing to the high disease burden of chronic autoimmune conditions. Furthermore, ancient DNA has also been used to trace the historical origins of autoimmune disease–associated variants (*50*) and revealed that loci from extinct hominins such as Neanderthals and Denisovans are enriched for innate and adaptive immunity genes (*51*).

Concluding remarks

Over the past 15 years, GWASs have substantially advanced our understanding of the genetic architecture of autoimmune diseases. The discovery of hundreds of disease-associated variants, including many overlapping across autoimmune conditions, confirmed their shared genetic and immunopathological etiology. Some of these shared variants have opposing effects on distinct autoimmune diseases, mirroring and offering insight into the differential outcomes observed with TNF-α inhibitors (Box 2). Ancient DNA sequencing analyses are revealing how selection pressure from pathogens over millennia has shaped the human immune genetic repertoire, contributing to higher frequencies for some common genetic factors and consequently to a higher prevalence of autoimmunity. The interpretation of these mostly noncoding risk variants has been challenging, although this is being addressed through multidimensional integration with gene regulation traits at cell-type– and context-specific resolution using single-cell sequencing and data from individuals with autoimmune diseases. Functional studies have begun to shed light on the molecular consequences of several autoimmunity variants, and the development of high-throughput approaches to functional characterization will accelerate this further. As evidenced by the observation that two-thirds of drugs approved by the FDA in 2021 were supported by genetic evidence (*52*), unraveling the biological mechanisms behind these variants has the potential to improve prevention and treatment options for these diseases. Furthermore, the use of polygenic predictors from GWAS results can help to identify healthy individuals at elevated risk of disease, potentially enhancing screening and allowing for presymptomatic interventions, such as in type 1 diabetes prevention trials (*53*, *54*).

Many autoimmune diseases affect population groups differently (*1*), and examining the role of genetic variation in these differences may yield valuable insights, as exemplified by the study of *BAFF* in Sardinia (*45*). However, most genomic studies thus far have been in participants of European ancestry. Going forward, major efforts will be required to expand genetic and epigenetic data collections to diverse ancestries around the world. Moreover, large collections of longitudinally followed and deeply phenotyped cohorts are required to identify genetic variants that contribute, not only to disease risk, but also to heterogeneity in disease progression, because these may differ (*55*, *56*). Finally, although this Review has focused on classic autoimmune diseases, genetic studies have revealed a role for autoimmunity in other conditions. For instance, GWASs in Parkinson's disease have identified HLA associations, and subsequent research reported frequent T cell responses to α-synuclein, which forms pathogenic protein aggregations in this disease (*57*). These and other findings in Alzheimer's disease, atherosclerosis, and cancer are broadening the view of inflammation and autoimmunity across human diseases.

REFERENCES AND NOTES

1. G. S. Cooper, M. L. K. Bynum, E. C. Somers, *J. Autoimmun.* **33**, 197–207 (2009).
2. A. Sisó-Almirall *et al.*, *Autoimmun. Rev.* **19**, 102448 (2020).
3. G. S. Temba *et al.*, *Nat. Immunol.* **22**, 287–300 (2021).
4. K. H. Costenbader, E. W. Karlson, *Lupus* **15**, 737–745 (2006).
5. K. Bjornevik *et al.*, *Science* **375**, 296–301 (2022).
6. G. E. Dinse *et al.*, *Arthritis Rheumatol.* **74**, 2032–2041 (2022).
7. M. Filippi *et al.*, *Nat. Rev. Dis. Primers* **4**, 43 (2018).
8. S. S. Rich, L. R. Weitkamp, J. Barbosa, *Am. J. Hum. Genet.* **36**, 1015–1023 (1984).
9. J. P. Hugot *et al.*, *Nature* **379**, 821–823 (1996).
10. E. Uffelmann *et al.*, *Nat. Rev. Methods Primers* **1**, 59 (2021).
11. L. Yengo *et al.*, *Nature* **610**, 704–712 (2022).
12. D. J. Weiner *et al.*, *Nature* **614**, 492–499 (2023).
13. A. Sazonovs *et al.*, *Nat. Genet.* **54**, 1275–1283 (2022).
14. P. Wainschtein *et al.*, *Nat. Genet.* **54**, 263–273 (2022).
15. D. R. Plichta, D. B. Graham, S. Subramanian, R. J. Xavier, *Cell* **178**, 1041–1056 (2019).
16. J. Domingo, P. Baeza-Centurion, B. Lehner, *Annu. Rev. Genomics Hum. Genet.* **20**, 433–460 (2019).
17. K. K.-H. Farh *et al.*, *Nature* **518**, 337–343 (2015).
18. N. S. Abell *et al.*, *Science* **375**, 1247–1254 (2022).
19. D. J. Schaid, W. Chen, N. B. Larson, *Nat. Rev. Genet.* **19**, 491–504 (2018).
20. D. R. Simeonov *et al.*, *Nature* **549**, 111–115 (2017).
21. L. M. Maier *et al.*, *PLOS Genet.* **5**, e1000322 (2009).
22. A. T. Satpathy *et al.*, *Nat. Biotechnol.* **37**, 925–936 (2019).
23. J. T. Merrill *et al.*, *N. Engl. J. Med.* **386**, 1034–1045 (2022).
24. H. K. Finucane *et al.*, *Nat. Genet.* **50**, 621–629 (2018).
25. N. A. Patsopoulos *et al.*, *Science* **365**, eaav7188 (2019).
26. S. Chun *et al.*, *Nat. Genet.* **49**, 600–605 (2017).
27. H. Mostafavi, J. P. Spence, S. Naqvi, J. K. Pritchard, bioRxiv 491045 [Preprint] (2022); https://doi.org/10.1101/2022.05.07.491045.
28. B. Soskic *et al.*, *Nat. Genet.* **54**, 817–826 (2022).
29. M. Ota *et al.*, *Cell* **184**, 3006–3021.e17 (2021).
30. S. Yazar *et al.*, *Science* **376**, eabf3041 (2022).
31. A. Nathan *et al.*, *Nature* **606**, 120–128 (2022).
32. K. Kundu *et al.*, *Nat. Genet.* **54**, 251–262 (2022).
33. K. Mouri *et al.*, *Nat. Genet.* **54**, 603–612 (2022).
34. K. Hemminki, X. Li, K. Sundquist, J. Sundquist, *Arthritis Rheum.* **60**, 2845–2847 (2009).
35. A. Verma *et al.*, *Am. J. Hum. Genet.* **104**, 55–64 (2019).
36. D. Ellinghaus *et al.*, *Nat. Genet.* **48**, 510–518 (2016).
37. W. van Rheenen, W. J. Peyrot, A. J. Schork, S. H. Lee, N. R. Wray, *Nat. Rev. Genet.* **20**, 567–581 (2019).
38. M. R. Lincoln *et al.*, bioRxiv 2021.05.13.21257044 [Preprint] (2021); https://doi.org/10.1101/2021.05.13.21257044.
39. M. D. Fortune *et al.*, *Nat. Genet.* **47**, 839–846 (2015).
40. H.-J. Westra *et al.*, *Nat. Genet.* **50**, 1366–1374 (2018).
41. C. A. Dendrou *et al.*, *Sci. Transl. Med.* **8**, 363ra149 (2016).
42. S. Boisson-Dupuis *et al.*, *Sci. Immunol.* **3**, eaau8714 (2018).
43. G. Kerner *et al.*, *Cell Genomics* **3**, 100248 (2023).
44. K. Ishigaki *et al.*, *Nat. Genet.* **54**, 393–402 (2022).
45. M. Steri *et al.*, *N. Engl. J. Med.* **376**, 1615–1626 (2017).
46. G. Sirugo, S. M. Williams, S. A. Tishkoff, *Cell* **177**, 1080 (2019).
47. Y. Liu, X. Mao, J. Krause, Q. Fu, *Science* **373**, 1479–1484 (2021).
48. J. Klunk *et al.*, *Nature* **611**, 312–319 (2022).
49. M. E. K. Niemi, M. J. Daly, A. Ganna, *Nat. Rev. Genet.* **23**, 533–546 (2022).
50. W. Barrie *et al.*, bioRxiv 2022.09.23.509097 [Preprint] (2022); https://doi.org/10.1101/2022.09.23.509097.
51. A. Liston, S. Humblet-Baron, D. Duffy, A. Goris, *Nat. Immunol.* **22**, 1479–1489 (2021).
52. D. Ochoa *et al.*, *Nat. Rev. Drug Discov.* **21**, 551 (2022).
53. L. A. Ferrat *et al.*, *Nat. Med.* **26**, 1247–1255 (2020).
54. K. C. Herold *et al.*, *N. Engl. J. Med.* **381**, 603–613 (2019).
55. J. C. Lee *et al.*, *Nat. Genet.* **49**, 262–268 (2017).
56. S. Baranzini, S. Sawcer, International Multiple Sclerosis Genetics Consortium, MultipleMS Consortium, Research Square rs.3.rs-1723574/v1 [Preprint] (2023).
57. D. Sulzer *et al.*, *Nature* **546**, 656–661 (2017).
58. W. J. Housley *et al.*, *Sci. Transl. Med.* **7**, 291ra93 (2015).
59. S. Wang, F. Wen, G. B. Wiley, M. T. Kinter, P. M. Gaffney, *PLOS Genet.* **9**, e1003750 (2013).
60. P.-P. Axisa *et al.*, *Sci. Transl. Med.* **14**, eabl3651 (2022).
61. The Lenercept Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group, *Neurology* **53**, 457–465 (1999).
62. A. P. Gregory *et al.*, *Nature* **488**, 508–511 (2012).
63. A. Cortes *et al.*, *Nat. Genet.* **45**, 730–738 (2013).
64. L. Li *et al.*, *Neurology* **100**, e558–e567 (2023).

ACKNOWLEDGMENTS

Funding: This work was supported by the National Institutes of Health (grants P01 AI073748, U24 AI11867, R01 AI22220, UM 1HG009390, P01 AI039671, P50 CA121974, and R01 CA227473 to D.A.H.). **Competing interests:** D.A.H. has received research funding from Bristol-Myers Squibb, Novartis, Sanofi, and Genentech and has been a consultant for Bayer Pharmaceuticals, Bristol Myers Squibb, Compass Therapeutics, EMD Serono, Genentech, Juno Therapeutics, Novartis Pharmaceuticals, Proclara Biosciences, Sage Therapeutics, and Sanofi Genzyme. A.H. has been a consultant for Biogen.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 13 February 2023; accepted 29 March 2023
10.1126/science.adg2992

EDITORIAL

The College Board can't be trusted

H. Holden Thorp
Editor-in-Chief,
Science journals.
hthorp@aaas.org;
@hholdenthorp

An institution that has long influenced academia in the United States has lost credibility. The College Board—a nonprofit organization that oversees the development of Advanced Placement (AP) precollege courses and the administration of the SAT exam that is used in college admissions—has been caught in an outright lie about its practices, raising questions about its vulnerability to political pressures. With the integrity of the College Board now in question, academia must decide whether it can be trusted.

In February, a major storm erupted in academia around the launch of an AP course in African American studies. Prior to its release, Florida governor Ron DeSantis and other conservatives complained that the curriculum promulgated liberal orthodoxy. To the dismay of scholars, the Board then released a revised version, stating that the changes were made through normal processes. Now, emails have surfaced belying that claim. The implications of this deception are tremendous not only for the humanities but also for science. An incursion into academic freedom for one is an incursion for all.

> "The College Board... has been caught in an outright lie about its practices, raising questions about its vulnerability to political pressures."

The College Board had it right in the fall of 2022 when it released a framework for this AP course. As historian Gerald Early told me, "Black studies was always meant to be a challenge to the white assumptions about the construction of knowledge and about the theoretical claims and political uses of the white deployment of knowledge." The initial framework was true to this vision, and included material on critical race theory and intersectionality, concepts crucial to understanding the activist nature of the field. But just as DeSantis criticized the course as "woke indoctrination," the curriculum was refined, making much of this material optional. Critics claimed that the College Board appeased a future presidential candidate, but the organization flooded the media with interviews and statements insisting that this was not the case.

Now the *Wall Street Journal* has published private emails showing that statements by the College Board were false. The courageous actor in the saga is Professor Nishani Frazier of the University of Kansas, who was involved in developing the portion of the course in which revisions were made. In response to the spin that the College Board's chief executive officer put on the controversy, Frazier wrote in one of the emails, "We all know this is a blatant lie. In fact, the major changes which occurred came from my unit—and not once did AP speak with me about these changes. Instead, it rammed through revisions, pretended course transformation was business as usual, and then further added insult to injury by attempting to gaslight the public with faux innocence." The course, she said, was "edited behind our backs." She put the blame for this on "the failure of AP to recognize both its own institutional racism and how its own lies and capitulation precipitated the creation of a monster of its own making." When I asked the College Board for a comment about Frazier's email, it declined to provide one.

In a column I wrote in February, I suggested that the reason the College Board could deny that it bowed to political pressure was that it had already known what changes the Florida governor (and perhaps other conservative states) would want for Black studies. Whatever the case, we now know from Frazier's email that the Board did not follow its normal procedures in editing the course, as they had repeatedly insisted.

The College Board now says that it will "rectify" the matter by again revising the course in a manner that will involve experts and a development committee. With eyes fixed upon it, the organization cannot afford to diminish any topics in the curriculum. But what about the other AP courses? There are nearly 40 of them, for which there is a large high school and college appetite. Are they in jeopardy? If the College Board knew what DeSantis desired for Black studies, might they also anticipate that he and other conservatives would favor adding intelligent design to AP Biology or removing climate change from AP Environmental Science? The College Board can no longer be trusted to defend any discipline.

The corruption of the College Board is appalling. It simply cannot be trusted, and academia must stop relying on it to make important decisions about education.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

Published online 27 April 2023; 10.1126/science.adi4588

NEWS

IN BRIEF

Edited by **Michael Price**

Pinyon-juniper woodl
are the most widespread
type of old-growth forest
on U.S. federal land.

FOREST ECOLOGY

U.S. boosts tally of old forests

Last year, President Joe Biden surprised forest scientists when he ordered an inventory of the government's holdings of mature and old-growth forests by this Earth Day. It triggered a scramble by the U.S. Forest Service and Bureau of Land Management to create a formal definition of what constitutes "mature" and "old-growth" forests and to apply those definitions across millions of hectares. Meeting the 22 April deadline last month, the agencies released their findings in a report, noting that of the nearly 72 million hectares of forest they manage, 45% are mature and 18% are old growth. The figures, which exceed estimates published by nonfederal researchers, include 9 million hectares of pinyon-juniper woodlands (pictured here in Utah)—a forest type that was previously rarely categorized as old growth. The report's findings are likely to fuel a raging debate about how to manage older forests and make them resilient to climate change.

Chemist gets home confinement

RESEARCH SECURITY | A U.S. district judge last week sentenced former Harvard University chemist Charles Lieber to 6 months of home confinement and fined him $50,000 for lying to federal agencies about his interactions with a Chinese university and failing to report payments from it. The ruling ended the most publicized case of some two dozen recent criminal prosecutions of U.S. academic scientists with research ties to China. In December 2021, Lieber's ties to the Wuhan University of Technology led to a guilty verdict in his trial. Prosecutors had asked for a 90-day prison sentence and a $150,000 fine for the 64-year-old Lieber, who has an incurable blood cancer and retired from Harvard earlier this year. His lawyers had requested no jail time because of his poor health. The case was brought under the government's China Initiative, which aimed to stop economic espionage by the U.S. rival. The campaign was renamed last year to clarify that it applies to malign actors from anywhere in the world. The government has had a mixed record in prosecuting academic scientists; several were acquitted or had their cases dropped, whereas a handful were found guilty of offenses similar to Lieber's and received prison sentences.

The unknown under the sea

MARINE BIOLOGY | Scientists suspect they have described less than 10% of the marine species on Earth. To learn more about the remaining ocean dwellers, researchers, businesses, and philanthropists have teamed up to identify within the next decade some 100,000 new sea creatures out of the estimated 2 million still unidentified species. The Ocean Census, launched last week, will combine DNA sequencing with machine learning to build a sort of cyber taxonomy, classifying organisms collected on expeditions across the world's seas. The results could aid conservation and give scientists a better understanding of the role of marine life in oxygen and food production, as well as carbon cycling. With funding from the Nippon Foundation, Japan's largest philanthropic organization, a U.K. marine science and conservation institute called Nekton will coordinate collections by ships, divers, submarines, and deep-sea robots. Ocean Census will make its data, along with 3D digital images of all new species, available to both researchers and the public. Given

the loss of coral, sharks, and other marine species in recent decades, "We are in a race against time," says the project's scientific director, Alex Rogers, a marine biologist at the University of Oxford.

Indian classrooms shed Darwin

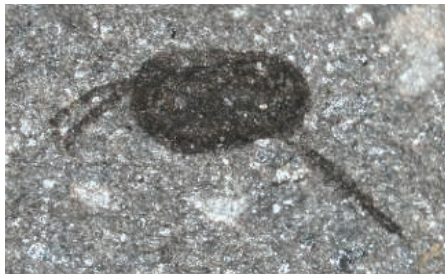
EDUCATION | Scientists in India are protesting a decision to cut discussion of Charles Darwin's theory of evolution from textbooks used by millions of ninth and 10th graders. More than 4000 people have signed a plea from the Breakthrough Science Society to restore the material. The nonprofit science advocacy group reports that the National Council of Educational Research and Training, an autonomous government group that sets curricula for India's 256 million primary and secondary students, dropped the topic as part of a "content rationalization" process. The removal makes "a travesty of the notion of a well-rounded secondary education," says evolutionary biologist Amitabh Joshi of the Jawaharlal Nehru Centre for Advanced Scientific Research. Others fear it signals a growing embrace of pseudoscience by Indian officials and think it's unlikely NCERT will relent.

Trial EU defense fund blasted

SCIENCE POLICY | The European Union was ill-prepared to ramp up defense research funding, according to a report released last week by its own financial watchdog. In 2017–19, the EU spent about €90 million on 18 projects under the Preparatory Action on Defence Research, a fund designed to "pave the way" for a much larger €8 billion European Defence Fund, which kicked off in 2021 and will run until 2027. But the report from the European Court of Auditors says the earlier trial fund didn't deliver fully as a "testbed" for the bigger program, as projects were delayed and made "limited progress." Auditors also warned that the European Commission is too understaffed to manage the surge in defense-research spending.

Welsh fossils highlight early life

PALEONTOLOGY | In Wales, paleontologists have uncovered a rich source of 462-million-year-old fossils that reveal more overlap than expected between animals that evolved during the Cambrian Explosion 40 million years earlier and the ancestors of modern species. Researchers had thought those ancestors had replaced the Cambrian creatures, but the new site—a small quarry in a sheep field—shows a much more



This 462-million-year-old fossil represents a new species, a clamlike creature with long appendages.

gradual transition, say Joe Botting and Lucy Muir of Amgueddfa Cymru-National Museum Wales. Among a bonanza of fossils, the married duo has cataloged 170 marine species, including glass sponges, crustaceans called horseshoe crab shrimp, and a six-legged arthropod that may have given rise to insects. Almost all the animals are tiny, with many ranging in size from a sesame seed to a pencil eraser, and their soft bodies are exquisitely preserved, providing insights into what they ate and how they lived, the research team reports this week in *Nature Ecology & Evolution*. This quarry, says Julien Kimmig, a paleontologist at the State Museum of Natural History Karlsruhe who was not involved with the work, "could definitely be as famous" as the renowned Burgess Shale in Canada, a plentiful source of Cambrian fossils from 500 million years ago.

Canadian Ph.D.s demand raises

SCIENTIFIC COMMUNITY | Thousands of scientists across Canada walked off the job on 1 May to protest low wages for graduate students and postdocs. At one event on Canada's Parliament Hill, Sarah Laframboise, a University of Ottawa biochemistry Ph.D. student and executive director of the grassroots organization Support Our Science, cited research suggesting that 86% of graduate students are stressed and anxious about their finances. The organization, which staged the 1-day protest, is asking the federal government to increase pay for graduate students and postdocs who are funded through federal scholarships and fellowships. In August 2022, it delivered an open letter to the government requesting increased investment in the next generation of scientists. But this year's federal budget, released in March, didn't include any such changes.

Andean condors' nests record a messy archive of their diets stretching back thousands of years.

ECOLOGY

Condors' poo reveals their history

To find out how the Andean condor's diet has shifted over millennia of environmental change, researchers scaled a cliff in Argentina's Patagonia region to collect samples from a doughnut-shaped mound of the birds' droppings. Based on radiocarbon dating and other clues, the scientists discovered that condors have nested on this cliffside for some 2200 years. However, the guano revealed that, between about 300 C.E. and 1300 C.E., Andean condors became scarce, as ash from nearby volcanic eruptions blanketed the landscape and killed off animals whose carcasses they preyed upon. Scientists also learned the condors' carrion of choice has changed over the years. Traces of llama DNA dominated the guano deposit's older layers, whereas introduced sheep and cattle are more prominent in fresher layers. Researchers say the findings illustrate the value of studying long-term nesting sites for reconstructing a species' ecological history.

PHOTOS: (TOP TO BOOTOM) JOE BOTTING; JACK DYKINGA/NPL/MINDEN PICTURES

IN DEPTH

Samoylov station in Siberia is no longer a hub of Russian-German collaboration on permafrost research.

SCIENCE DIPLOMACY

Russia tensions keep Arctic research on ice

Collaboration stifled as many scientists look to establish fieldwork elsewhere

By **Warren Cornwall**

In 2010, Russian President Vladimir Putin—then the prime minister—visited a remote research station on Samoylov Island, in Siberia's far north. Beginning in the 1990s, the facility on the Lena River had become a hub for German and Russian collaboration on change in the region's permafrost. "I see a good example of international cooperation here," Putin told a group of scientists.

Today, as Arctic researchers prepare for their summer fieldwork, Samoylov station instead highlights the collapse of international cooperation. German scientists have not been to Samoylov since 2021, and a regular stream of permafrost data has dried up. After Russia invaded Ukraine in 2022, Germany and other European countries joined the United States and Canada in barring their scientists from collaborating on most projects with Russia, which controls half of the Arctic's coastline. "We are cut off from this part of the Arctic," says Anne Morgenstern, a permafrost scientist at Germany's Alfred Wegener Institute who coordinated the German presence at Samoylov. "It's just a tragedy, the whole situation, and everybody hopes that this horrible war ends soon."

The abrupt rupture is jeopardizing data on climate change, oceanography, and ecology that stretch back decades. On Samoylov Island, sensors that measure carbon dioxide and methane emissions from the warming permafrost are likely still operating, Morgenstern says. But none of the data is flowing to German scientists, and she questions how long the observations can be sustained without spare parts or specialized knowledge from Germany. "We suspect it's going to be very hard for them to maintain the measurements in the long term," she says.

Rebuilding severed ties could be slow. Mike Sfraga, chair of the U.S. Arctic Research Commission, described it as a "tectonic shift" at a February conference in Norway. "If the war stopped tomorrow miraculously, things don't just turn back on."

Russia continues to pursue Arctic research. It is building a permafrost monitoring network and recently launched the *Severny Polyus* (*North Pole*), a research vessel designed for yearslong Arctic deployments, the press office of the St. Petersburg–based Arctic and Antarctic Research Institute said in a statement. Most collaborations with the West "have now been suspended, not on our initiative. Of course, we regret this decision of foreign colleagues, but we continue to work on our own."

Some Western scientists have found workarounds. Jan van Gils, an ecologist at the Royal Netherlands Institute for Sea Research, studies red knots, shorebirds that winter in Africa, then fly 9000 kilometers to Siberia to breed in the summer. He was planning an intensive 4-year field campaign in Russia. Instead, during fieldwork last month in Mauritania, he attached GPS trackers to 80 of the birds in hopes of following them remotely. "Of course the science would even be better if we could study it onsite," he says.

Agencies are also improvising. For years, Russian vessels had serviced a set of Arctic Ocean sensors, funded by the U.S. National Science Foundation, that are moored just

outside Russian waters. Instead, the U.S. Coast Guard icebreaker *Healy* will travel this summer to work on them.

Many scientists are also reorienting their research to more accessible territory in Scandinavia and North America. Morgenstern is working to establish a program on the Mackenzie River in the Canadian Arctic to monitor its water chemistry, including the effects of thawing permafrost. That would take the place of her past work on the Lena River. Bruce Forbes, a geographer at Finland's University of Lapland, is moving a planned study of the cues that guide reindeer migration across the tundra from Siberia to northern Finland. "I hope I can go back eventually," Forbes says.

As individual scientists and institutions try to move forward, the Arctic Council, an intergovernmental policymaking body, remains in limbo. Composed of officials from the eight Arctic nations, the council doesn't directly fund research. But it sets research priorities and coordinates scientific activities. After the invasion, most work was frozen when officials from the seven Western countries refused to meet with their Russian counterpart, even though Russia held the council's chair.

Norway takes over as chair on 11 May. It may try to get all the nations to agree to ground rules that would allow lower level council committees to resume activities that require Russian involvement, says Jennifer Spence, a Harvard University Arctic policy expert who recently served on the council's Sustainable Development Working Group. That could include publishing scientific reports and supporting researchers monitoring environmental conditions. "Norway is looking for a meaningful way to keep the work of the Arctic Council alive," she says.

In the Barents Sea, Norway and Russia continue to exchange data about fisheries they managed together. But it's just a sliver of the past cooperation on this shared water body, says Ole Arve Misund, executive director of the Norwegian Polar Institute. "This is the only official contact," he says.

Individual researchers are also struggling to keep lines of communication open. Last year, Norwegian organizers of the largest Arctic science conference, the Arctic Science Summit Week, barred Russians from participating. This year, Austrian organizers of the event allowed Russians to attend as long as they displayed no official affiliation with a Russian institution, such as listing it on a name tag. But only six Russian scientists ended up attending a gathering that drew more than 800 people, and five of those were online.

Vladimir Romanovsky, a Russian-born permafrost expert at the University of Alaska Fairbanks who carries both U.S. and Russian passports, traveled via Turkey to a conference in Siberia's Yakutsk region this winter, where a good friend leads a permafrost research program. He says he's glad he went, in part to encourage young researchers he met. "I will continue to support these relations just for those young people to have some hope."

Yet he worries that a program he coordinates tracking permafrost temperatures across the Arctic is imperiled. He has had to abandon his practice of sending money to Russian colleagues to help support their work. He says they are reluctant to accept research money from abroad, fearing they will be labeled a "foreign agent"—which could put them at risk of government scrutiny or arrest. "The coming field season is in big question," he says. ■

PHOTO: THOMAS OPEL/ALFRED WEGENER INSTITUTE

In 2010, Vladimir Putin, then Russia's prime minister, visited the Samoylov research station.

SCIENTIFIC COMMUNITY

NIH stumbles on 'pass the harasser'

Agency allowed transfer of grants and handed new money to David Gilbert, despite "severe" behavior

By **Meredith Wadman**

When genome researcher David Gilbert left Florida State University (FSU) in 2021 for the San Diego Biomedical Research Institute (SDBRI), he took two large National Institutes of Health (NIH) grants with him. The biomedical agency approved the transfer and went on to award Gilbert a new, $2.5 million grant last year.

None of this would be out of the ordinary—except that, in 2020, prior to any of these moves, FSU had completed a far-reaching investigation prompted when Gilbert emailed a description of his erotic dream to a graduate student. The probe revealed a yearslong history and concluded Gilbert's "gendered, sexualized and invasive behaviors were severe and pervasive." NIH learned the full extent of his misconduct at FSU before making the new award but after his move to San Diego—where his behavior elicited a new probe, *Science* has learned, and drove at least one woman scientist from the institute.

The sequence of events that allowed Gilbert to continue a well-funded career unfolded at the same time as NIH very publicly said it was cracking down on sexual harassers who tried to move their bad behavior from one institution to another. In 2020, NIH instituted a requirement that then-Director Francis Collins said in a *Science* editorial was aimed in part at "preventing 'passing the harasser,' in which a scientist who changed institutions could evade the consequences of findings of sexual harassment."

The policy required grantee institutions to notify the agency about sexual harassment findings if they drove requests to make certain changes to a grant, including transferring it to a new institution, so NIH could weigh this information in its decision about whether to approve the change or take other action. In March 2022, Congress wrote a tightened policy into law, making it mandatory for

outside Russian waters. Instead, the U.S. Coast Guard icebreaker *Healy* will travel this summer to work on them.

Many scientists are also reorienting their research to more accessible territory in Scandinavia and North America. Morgenstern is working to establish a program on the Mackenzie River in the Canadian Arctic to monitor its water chemistry, including the effects of thawing permafrost. That would take the place of her past work on the Lena River. Bruce Forbes, a geographer at Finland's University of Lapland, is moving a planned study of the cues that guide reindeer migration across the tundra from Siberia to northern Finland. "I hope I can go back eventually," Forbes says.

As individual scientists and institutions try to move forward, the Arctic Council, an intergovernmental policymaking body, remains in limbo. Composed of officials from the eight Arctic nations, the council doesn't directly fund research. But it sets research priorities and coordinates scientific activities. After the invasion, most work was frozen when officials from the seven Western countries refused to meet with their Russian counterpart, even though Russia held the council's chair.

Norway takes over as chair on 11 May. It may try to get all the nations to agree to ground rules that would allow lower level council committees to resume activities that require Russian involvement, says Jennifer Spence, a Harvard University Arctic policy expert who recently served on the council's Sustainable Development Working Group. That could include publishing scientific reports and supporting researchers monitoring environmental conditions. "Norway is looking for a meaningful way to keep the work of the Arctic Council alive," she says.

In the Barents Sea, Norway and Russia continue to exchange data about fisheries they managed together. But it's just a sliver of the past cooperation on this shared water body, says Ole Arve Misund, executive director of the Norwegian Polar Institute. "This is the only official contact," he says.

Individual researchers are also struggling to keep lines of communication open. Last year, Norwegian organizers of the largest Arctic science conference, the Arctic Science Summit Week, barred Russians from participating. This year, Austrian organizers of the event allowed Russians to attend as long as they displayed no official affiliation with a Russian institution, such as listing it on a name tag. But only six Russian scientists ended up attending a gathering that drew more than 800 people, and five of those were online.

Vladimir Romanovsky, a Russian-born permafrost expert at the University of Alaska Fairbanks who carries both U.S. and Russian passports, traveled via Turkey to a conference in Siberia's Yakutsk region this winter, where a good friend leads a permafrost research program. He says he's glad he went, in part to encourage young researchers he met. "I will continue to support these relations just for those young people to have some hope."

Yet he worries that a program he coordinates tracking permafrost temperatures across the Arctic is imperiled. He has had to abandon his practice of sending money to Russian colleagues to help support their work. He says they are reluctant to accept research money from abroad, fearing they will be labeled a "foreign agent"—which could put them at risk of government scrutiny or arrest. "The coming field season is in big question," he says. ■

PHOTO: THOMAS OPEL/ALFRED WEGENER INSTITUTE

In 2010, Vladimir Putin, then Russia's prime minister, visited the Samoylov research station.

SCIENTIFIC COMMUNITY

NIH stumbles on 'pass the harasser'

Agency allowed transfer of grants and handed new money to David Gilbert, despite "severe" behavior

By **Meredith Wadman**

When genome researcher David Gilbert left Florida State University (FSU) in 2021 for the San Diego Biomedical Research Institute (SDBRI), he took two large National Institutes of Health (NIH) grants with him. The biomedical agency approved the transfer and went on to award Gilbert a new, $2.5 million grant last year.

None of this would be out of the ordinary—except that, in 2020, prior to any of these moves, FSU had completed a far-reaching investigation prompted when Gilbert emailed a description of his erotic dream to a graduate student. The probe revealed a yearslong history and concluded Gilbert's "gendered, sexualized and invasive behaviors were severe and pervasive." NIH learned the full extent of his misconduct at FSU before making the new award but after his move to San Diego—where his behavior elicited a new probe, *Science* has learned, and drove at least one woman scientist from the institute.

The sequence of events that allowed Gilbert to continue a well-funded career unfolded at the same time as NIH very publicly said it was cracking down on sexual harassers who tried to move their bad behavior from one institution to another. In 2020, NIH instituted a requirement that then-Director Francis Collins said in a *Science* editorial was aimed in part at "preventing 'passing the harasser,' in which a scientist who changed institutions could evade the consequences of findings of sexual harassment."

The policy required grantee institutions to notify the agency about sexual harassment findings if they drove requests to make certain changes to a grant, including transferring it to a new institution, so NIH could weigh this information in its decision about whether to approve the change or take other action. In March 2022, Congress wrote a tightened policy into law, making it mandatory for

institutions to inform NIH of disciplinary actions taken against harassers, regardless of whether they prompted changes to grants. By then, NIH in concert with institutions had already removed dozens of principal investigators (PIs) from grants.

Gilbert was not one of them. In his case, say harassment experts and activists who reviewed related documents for *Science*, the agency was either remiss or ineffectual, compounding a failure of Gilbert's two institutions to adequately protect victims. "It's outrageous" that Gilbert still has a job and three big NIH grants, says Angie Rasmussen, a virologist at the University of Saskatchewan who in 2019 served on a working group advising the NIH director on combating sexual harassment. "I would just like to know what NIH plans to do to resolve this. Because when I was on that committee, we were assured over and over again, first by Francis Collins, then by [acting NIH Director] Larry Tabak, that this was a priority."

FSU launched its investigation of Gilbert early in 2020 after a biology graduate student filed a complaint triggered by an email from Gilbert, a professor with an endowed chair and a high profile. "You were in my dream last night," he wrote before describing the erotic dream in the email, which concluded: "I hope … you appreciate that i [sic] am not some 20-something who is done in 30 minutes … Us 'mature' guys like to work slowly - take our time and savor every minute!!!!"

The investigation resulted in a 131-page report detailing behavior that stretched back to at least 2017. Incidents documented in emails and witness testimonies, and admitted to by Gilbert, include kissing a student on the neck and emailing his lab members to ask which one of them had spelled "cock" with refrigerator magnets during a party at Gilbert's home. He added in a follow-up email: "I guess it is something that is available at my house, if the right person were to ask nicely." In an exchange with a lab technician who asked for help writing a website summary of her work, he joked, unsolicited, about her sticking things up her nose, then asked whether "you stick things up the other end of your body," adding, "The things I thought of would get me fired for sexual harassment."

One undergraduate who worked in Gilbert's lab reported her unease when he pulled her in to dance at a party at his home, where he grew angry at her for refusing to drink tequila shots; she was 19. "Severely uncomfortable" whenever she had to meet with him, she told investigators, "I would go to people and ask for help and their response was, 'Dave makes so much money for FSU so they're not going to talk to him.'" After her experience, she added, "I'm definitely not doing research again."

Gilbert did not respond to repeated interview requests from *Science*. But when asked by FSU investigators about the email regarding the erotic dream, he said he felt "immediate remorse" and acknowledged it "wasn't appropriate or welcome." Of some other witness testimony, he said: "These inferences are distorted, out of context and in some cases completely fabricated accusations."

After its probe concluded, FSU told *Science*, it removed Gilbert's endowed chair and his access to the funds the endowment generated. He was suspended for 10 days without pay. He also received "follow-up training" and a warning that any repeat of the behavior could cost him his job, FSU later wrote to NIH. These sanctions were "commensurate with the level of offense," FSU told *Science*.

When David Gilbert was at Florida State University, an investigation found his sexual harassment was "severe and pervasive."

Others disagree. The university's response "just seems so insufficient" and sends "a very distinct message that there is very little consequence for his violation," says Jennifer Freyd, a psychologist who studies institutional betrayal as president of the Center for Institutional Courage.

NIH didn't learn of the FSU probe until April 2021, 1 year after it was completed. As Gilbert was planning to move to San Diego at the end of June 2021, NIH policy required the university to alert the agency of the sexual harassment investigation and its conclusion in the process of asking NIH to approve the transfer of Gilbert's two grants, worth $867,000 annually at the time. FSU did not share its lengthy written report with NIH.

NIH Deputy Director for Extramural Research Michael Lauer soon wrote to FSU to find out whether the probe and any resulting disciplinary actions had affected Gilbert's ability to perform the funded research in the preceding year. Laurel Fulkerson, FSU's interim vice president for research, wrote to Lauer that Gilbert had not lost access to NIH funding, to campus, or to his lab. She added that Gilbert was "contrite" and that he meant his behavior as "funny." In a June 2021 teleconference, FSU administrators and NIH agreed that his grants would be reassigned to SDBRI.

And with that, Lauer later wrote, "we were under the impression that this issue was resolved"—until November 2021 when Tristan Wood, a reporter at FloridaPolitics.com, published an article linking to the full 131-page Gilbert report.

Lauer soon wrote to FSU the bureaucratic-language equivalent of an outraged letter. He cited vivid details of Gilbert's behavior from the report and said the article "raises concerns about the severity of Dr. Gilbert's violations, compared to what the University previously reported to NIH." The correspondence went on to discuss what actions the university was taking to, as Lauer put it, "assure that NIH-funded research and training is conducted in a civil, safe, and respectful environment, free from discrimination and unlawful harassment." Gilbert, however, was already long gone.

Two months after Lauer sent that irate missive to FSU, on 10 March 2022, NIH's National Cancer Institute sent Gilbert a new, $2.5 million award, to support his work at SDBRI. Whether NIH imposed any conditions on the new grant award is not clear. In a statement responding to detailed questions for Lauer about the Gilbert case, NIH's Office of Extramural Research (OER) said it "does not discuss grant deliberations." Asked who permitted Gilbert's new award, OER said, commenting generally on its procedures, the institute or center funding the grant "conducts a pre-award risk assessment" and sends the award out the door. It added that the agency requires institutions to have in place "effective internal controls" for assuring safe work spaces. OER can prevent grant transfers or hold a pending award to a "PI of concern" until it is satisfied that "compliance issues are resolved." If they can't be, Tabak said last year, it is within NIH's power to terminate harassers' grants.

FSU did not tell the San Diego institute about Gilbert's behavior, and by one account, it continued there. In early 2021, a trainee landed in Gilbert's new SDBRI lab a few

months ahead of him. She had met Gilbert at a conference, and he was "very nice" and "seemed supportive" as she began corresponding with him about a possible job, she says. He was a big name and she was excited to work for him.

But when Gilbert arrived, in July 2021, the abuse began, she says. He shouted at her, called her stupid, and complained about her to a colleague—then told her he had done so. He was by turns dismissive and patronizing and she left every conversation feeling humiliated, the trainee says. Sometimes he would allude to the "great night" he had just had with his girlfriend. One time, she says, he came into the small, dark microscopy room where she was working and stood extremely close to her. Intimidated, she asked him forcefully to leave. The last straw came in July 2022, when Gilbert texted her an image of a young woman's breasts in a strapless bra; he soon explained he had meant it for his girlfriend. Within days, the trainee complained to SDBRI President Joanna Davies, who launched an investigation.

In late August, Davies and the external investigator she had hired to conduct the probe called in the trainee to tell her the results: Gilbert had not violated any legal standard of sexual harassment. The trainee insisted on receiving that finding in writing. Davies and an SDBRI human resources professional wrote a terse memo stating that, despite the lack of a legal finding, "the Institute recognizes that steps need to be taken to ensure that individuals who work with Dave are working in what they believe is a respectful environment." (Davies did not respond to repeated interview requests.) The trainee was devastated, she says, and soon moved to a new institution.

Gilbert continues to work at the San Diego institute. His three NIH grants provide him and the institute $1.5 million in annual funding—$527,000 of which goes to SDBRI in the form of indirect costs.

"For people who are high-flying scientists—particularly well-funded ones—institutions seem to be remarkably willing to look the other way to obtain their employment," says Jeremy Berg, a computational and systems biologist at the University of Pittsburgh and former head of NIH's National Institute of General Medical Sciences, which funds one of Gilbert's grants. (Berg is also former editor-in-chief of the *Science* family of journals.)

Given this kind of payout, Rasmussen asks, "If NIH isn't going to put a stop to this money getting passed around, what incentive is there for the institutions to do anything about it themselves?" ■

This story was supported by the *Science* Fund for Investigative Reporting.

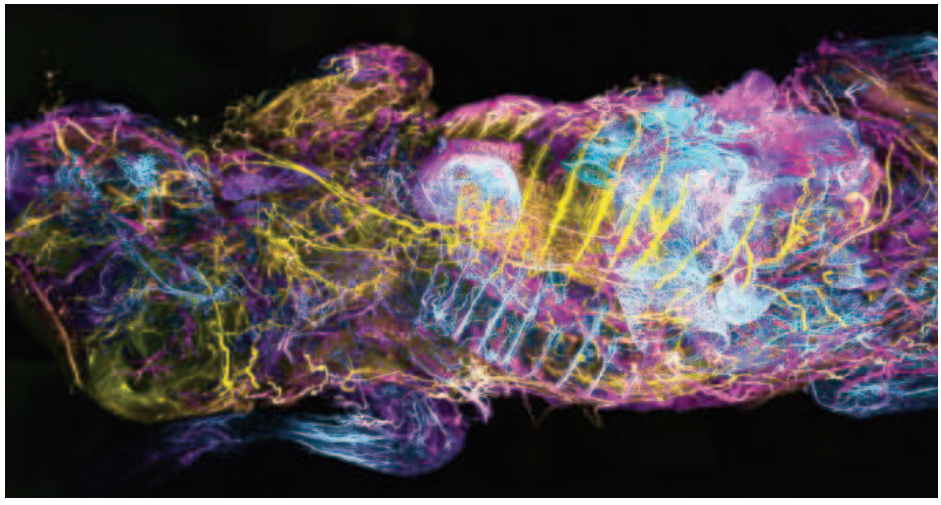


IMAGE: ERTÜRK LAB/HELMHOLTZ MUNICH

Neuron-specific antibodies, colored based on depth within a mouse, help visualize a peripheral nervous system.

BIOLOGY

'Game changer' method lets scientists peer into mice

Imaging technique uses standard cell-specific antibodies

By **Esra Öz**

A research team has turned the bodies of dead mice into vivid 3D maps of anatomy, with tissues, nerves, and vessels highlighted in color. They and other researchers say the technique, which renders the corpses transparent and then exposes them to fluorescent antibodies that label distinct cell types, could help everything from drug development to understanding the spread of cancer.

The developers, at the Helmholtz Munich research institute, call their technique wildDISCO—wild because it can work on any "wild type," or normal, mice, and DISCO for 3D Imaging of Solvent-Cleared Organs. Building on previous success at making mouse bodies transparent and labeling cell types, their newest technique removes cholesterol from the bodies so that a vast array of existing antibodies can penetrate deep into the animals. "wildDISCO is a game changer—it allows us to see the hidden highways and byways in the body," says Muzlifah Haniffa, a dermatologist and immunologist at the Wellcome Sanger Institute and Newcastle University's Biosciences Institute.

The method should let scientists map a mouse at the cellular level and explore previously hidden links between tissues, like neural connections between organs, says neuroscientist Ali Ertürk, director of Helmholtz Munich, who led the work, posted recently as a preprint. His group in Germany has already posted eye-catching videos of "flying" through the 3D anatomy of a mouse with different tissues labeled.

The method is a variation of work his team started more than a decade ago, when various scientists were developing novel ways, such as one called CLARITY, to chemically treat tissue samples and isolated organs so they became transparent and the cells inside could be seen. Ertürk and colleagues developed a similar method that made entire mouse bodies transparent. Seeking a way to pick out different cell types in the see-through bodies, they turned to nanobodies, synthetic antibodies that are much smaller than standard ones and can slip more easily into the tissue. But only a handful of nanobodies have been designed for specific cell types.

Ertürk's team has now found that treating mouse bodies for 2 weeks with a chemical called beta-cyclodextrin dissolves cholesterol in the cell membrane. This creates spongelike holes in the whole organism through which standard immunoglobulin G antibodies can slip, including markers of the nervous system, vasculature, immune system, and proliferating cells.

"This is exciting, and the immunolabeling data are quite impressive," says Alain Chédotal, a neuroscientist at the Vision Institute in Paris. ■

Esra Öz is a journalist based in Turkey.

months ahead of him. She had met Gilbert at a conference, and he was "very nice" and "seemed supportive" as she began corresponding with him about a possible job, she says. He was a big name and she was excited to work for him.

But when Gilbert arrived, in July 2021, the abuse began, she says. He shouted at her, called her stupid, and complained about her to a colleague—then told her he had done so. He was by turns dismissive and patronizing and she left every conversation feeling humiliated, the trainee says. Sometimes he would allude to the "great night" he had just had with his girlfriend. One time, she says, he came into the small, dark microscopy room where she was working and stood extremely close to her. Intimidated, she asked him forcefully to leave. The last straw came in July 2022, when Gilbert texted her an image of a young woman's breasts in a strapless bra; he soon explained he had meant it for his girlfriend. Within days, the trainee complained to SDBRI President Joanna Davies, who launched an investigation.

In late August, Davies and the external investigator she had hired to conduct the probe called in the trainee to tell her the results: Gilbert had not violated any legal standard of sexual harassment. The trainee insisted on receiving that finding in writing. Davies and an SDBRI human resources professional wrote a terse memo stating that, despite the lack of a legal finding, "the Institute recognizes that steps need to be taken to ensure that individuals who work with Dave are working in what they believe is a respectful environment." (Davies did not respond to repeated interview requests.) The trainee was devastated, she says, and soon moved to a new institution.

Gilbert continues to work at the San Diego institute. His three NIH grants provide him and the institute $1.5 million in annual funding—$527,000 of which goes to SDBRI in the form of indirect costs.

"For people who are high-flying scientists—particularly well-funded ones—institutions seem to be remarkably willing to look the other way to obtain their employment," says Jeremy Berg, a computational and systems biologist at the University of Pittsburgh and former head of NIH's National Institute of General Medical Sciences, which funds one of Gilbert's grants. (Berg is also former editor-in-chief of the *Science* family of journals.)

Given this kind of payout, Rasmussen asks, "If NIH isn't going to put a stop to this money getting passed around, what incentive is there for the institutions to do anything about it themselves?" ■

This story was supported by the *Science* Fund for Investigative Reporting.

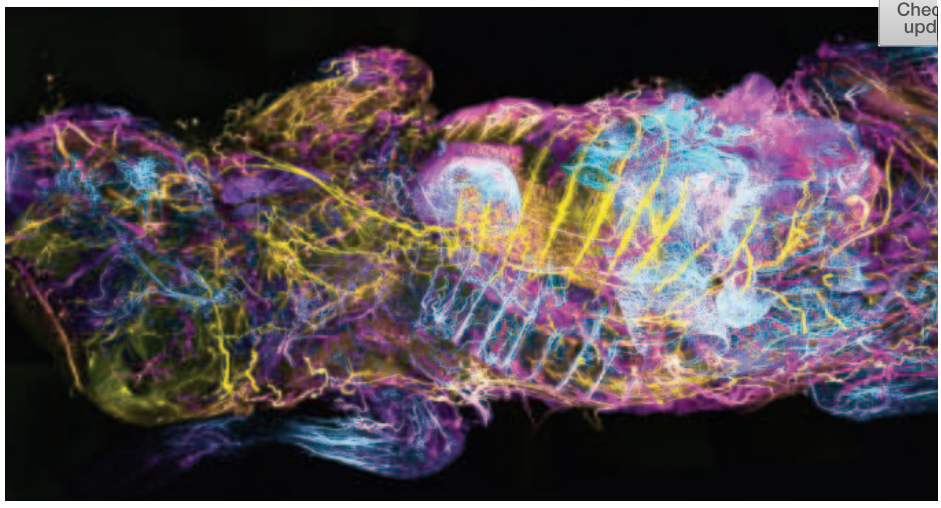


IMAGE: ERTÜRK LAB/HELMHOLTZ MUNICH

Neuron-specific antibodies, colored based on depth within a mouse, help visualize a peripheral nervous system.

BIOLOGY

'Game changer' method lets scientists peer into mice

Imaging technique uses standard cell-specific antibodies

By **Esra Öz**

A research team has turned the bodies of dead mice into vivid 3D maps of anatomy, with tissues, nerves, and vessels highlighted in color. They and other researchers say the technique, which renders the corpses transparent and then exposes them to fluorescent antibodies that label distinct cell types, could help everything from drug development to understanding the spread of cancer.

The developers, at the Helmholtz Munich research institute, call their technique wildDISCO—wild because it can work on any "wild type," or normal, mice, and DISCO for 3D Imaging of Solvent-Cleared Organs. Building on previous success at making mouse bodies transparent and labeling cell types, their newest technique removes cholesterol from the bodies so that a vast array of existing antibodies can penetrate deep into the animals. "wildDISCO is a game changer—it allows us to see the hidden highways and byways in the body," says Muzlifah Haniffa, a dermatologist and immunologist at the Wellcome Sanger Institute and Newcastle University's Biosciences Institute.

The method should let scientists map a mouse at the cellular level and explore previously hidden links between tissues, like neural connections between organs, says neuroscientist Ali Ertürk, director of Helmholtz Munich, who led the work, posted recently as a preprint. His group in Germany has already posted eye-catching videos of "flying" through the 3D anatomy of a mouse with different tissues labeled.

The method is a variation of work his team started more than a decade ago, when various scientists were developing novel ways, such as one called CLARITY, to chemically treat tissue samples and isolated organs so they became transparent and the cells inside could be seen. Ertürk and colleagues developed a similar method that made entire mouse bodies transparent. Seeking a way to pick out different cell types in the see-through bodies, they turned to nanobodies, synthetic antibodies that are much smaller than standard ones and can slip more easily into the tissue. But only a handful of nanobodies have been designed for specific cell types.

Ertürk's team has now found that treating mouse bodies for 2 weeks with a chemical called beta-cyclodextrin dissolves cholesterol in the cell membrane. This creates spongelike holes in the whole organism through which standard immunoglobulin G antibodies can slip, including markers of the nervous system, vasculature, immune system, and proliferating cells.

"This is exciting, and the immunolabeling data are quite impressive," says Alain Chédotal, a neuroscientist at the Vision Institute in Paris. ■

Esra Öz is a journalist based in Turkey.

Opposition lawmakers seized the dais of Mexico's Senate last week to protest actions by the majority.

SCIENCE POLICY

Amid controversy, Mexico passes major science reform bill

Many researchers oppose law, which could face court test

By **Alejandra Manjarrez**

A controversial effort by Mexico's president to reshape the nation's research funding and governance system has reached a chaotic conclusion that clears the way for the bill to become law. Many scientists strongly oppose the measure, saying it will tighten government control over science. But it was approved in a rapid-fire series of votes last week, including one on 29 April when senators from the ruling party abandoned their chamber because of a protest and passed the measure after just 14 minutes of debate, with no opposition lawmakers present.

The frenzied maneuvering could expose the bill to legal challenges, observers say. The process was "full of legislative procedural flaws" that will likely "be fought in the courts," says Alma Cristal Hernández Mondragón, president of the Mexican Association for the Advancement of Science and a policy specialist at the National Polytechnic Institute's Center for Research and Advanced Studies.

The Senate move capped several years of often fierce debate over the legislation, which President Andrés Manuel López Obrador's administration has said is needed to streamline policymaking and bolster government support for basic and applied research. After many scientists objected to a first draft released in late 2020—in part because they said it gave too much power to Mexico's central funding agency, the National Council of Science and Technology (Conacyt)—the administration said it would revise the plan. But that revision, presented to congress in December 2022, did little to quell many scientists' concerns. And critics say the now-approved version retains a number of problematic provisions.

One that has drawn notice eliminates a semi-independent panel that helps the government set research priorities and includes a relatively wide range of representatives from the public and private sectors. It would be replaced by a new group with a narrower membership that sits beneath a new agency, the National Council for the Humanities, Sciences, Technologies and Innovation (Conahcti), which will replace Conacyt. Hernández Mondragón fears the new panel will exclude "voices that can contribute" to policy debates. And the Mexican Academies of Sciences, Engineering, and Medicine has expressed concern that the law could restrict "freedom of research" by giving the new agency greater control.

Other researchers worry about provisions that make it harder for researchers at Mexico's private universities to compete for government grants. The law favors public universities and will result in "discriminatory treatment" of researchers and students at private institutions, says David René Romero Camarena, a genome scientist at the National Autonomous University of Mexico, which is a public institution.

The law also abandons a goal, backed in previous legislation, of spending 1% of Mexico's gross domestic product on research. Although the nation has never reached that goal, researchers worry the move signifies a rollback of Mexico's scientific ambitions.

In a statement after the vote, Conacyt defended the measure, calling it "an historic milestone" that will protect "the human right to science and knowledge" and help address "urgent national problems." It rejected charges that the law would centralize power and said "it does not disdain research done at private universities."

Some scientists are encouraged by provisions that call for work on behalf of the environment and disadvantaged communities. "It will help many who were not privileged before," says Martha Ileana Espejel Carbajal, a social ecologist who recently retired from the Autonomous University of Baja California. "It is the first time in 40 years of being a research professor that the law asks to do what my team has always struggled to defend: the work with the most marginalized communities, with women, with youth, with social and ecological systems." The measure may have flaws, she adds, but "every law is perfectible."

Science groups, however, fault lawmakers for breaking promises to consult with the public and scientists. They note that Mexico's lower house, the Chamber of Deputies, held just two of seven promised public hearings before moving to rapidly approve the bill on 26 April by a vote of 257 to 209.

Researchers were still surprised—and angry—about that vote when the Senate took up the bill 2 days later. Initially, it appeared the Senate would not vote on the measure before it recessed on 30 April because opposition lawmakers seized the dais to protest actions by the ruling bloc. But senators from López Obrador's Morena party and their allies moved to a nearby venue, where they passed the science bill and 18 other measures in a few frenzied hours.

The new law is likely to face court challenges. But whatever the outcome, Romero Camarena says he's optimistic. "Mexican scientists have had to face a great number of obstacles," he says. "This is one more obstacle that we will be able to overcome." ■

Alejandra Manjarrez is a journalist in Mexico City.

OCEANOGRAPHY

Critical El Niño monitor gets an upgrade

Revamped tropical Pacific buoys could aid in forecasts of atmospheric rivers

By **Paul Voosen**

For 3 years in a row, cool La Niña conditions have reigned in the tropical Pacific Ocean, suppressing the steady march of global warming. But warm waters are now rolling east and gathering off the west coast of South America, signaling the likely arrival of El Niño later this year and, next year, a surge in heat that could push the planet past 1.5°C of warming. These fluctuations in the Pacific—the greatest short-term control on global climate—once caught the world off guard. But they are now predictable months in advance, largely because of the Tropical Atmosphere Ocean (TAO) array, a series of 55 U.S. buoys, moored to the sea floor, that stretch some 13,000 kilometers along the equator.

Now, the TAO array is getting a $23 million overhaul, the first since it was set up in the mid-1990s, the National Oceanic and Atmospheric Administration (NOAA) says. The revamped buoys, the first of which was deployed on 13 April, will be more robust and able to monitor the ocean below in more detail, potentially allowing earlier and more accurate El Niño forecasts. Some will be moved to locations north of the equator, to enable better forecasts of cyclones and atmospheric rivers, the parades of storms that can inundate coastal regions such as California. "This is the most significant change the array has undergone," says Karen Grissom, an oceanographer at NOAA's National Data Buoy Center, which runs TAO. "It's going to be completely different."

During El Niño, the Pacific trade winds that blow along the equator from east to west weaken, allowing warm water from the western Pacific to flow east; during La Niña, the trade winds strengthen, trapping the warm water in the west and dragging cool water up from the depths in the east. The shifts not only change global ocean and land temperatures, but also reshape weather patterns, fueling flooding and drought as far away as India and Australia. "It's like an anchor of the global climate system," says Susan Wijffels, a physical oceanographer at the Woods Hole Oceanographic Institution.

Predicting these conditions is more than an academic exercise: The unprecedented "triple-dip" La Niña between 2020 and 2022 led to severe drought and food insecurity in eastern Africa. But by feeding TAO data into long-range weather models, researchers were able to forecast the event, allowing aid agencies to prepare while the Kenya Meteorological Department advised farmers to plant drought-tolerant crops, says Chris Funk, a climate scientist at the University of California, Santa Barbara. "That all leads back to the TAO array."

A prototype of an upgraded Tropical Atmosphere Ocean array buoy was deployed last year on a cruise.

PHOTO: NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

The TAO overhaul stems from a crisis a decade ago, when NOAA decommissioned the ship that serviced the 5-meter-tall buoys and had trouble financing the replacement of broken ones. Although funding was restored, it prompted a reassessment of what the tropical Pacific observing system, or TPOS, should look like with the advent of satellite observations and the Argo array, a fleet of nearly 4000 drifting, autonomous floats that profile ocean temperatures down to 2000 meters.

Each part of the observing system should focus on what it does best, says William Kessler, an oceanographer at NOAA's Pacific Marine Environmental Laboratory who co-chaired the international TPOS 2020 committee. For satellites, that's a broad view of the ocean surface currents and temperatures. For Argo, it's the deep vertical profiles. And for the TAO moorings, which measure temperatures to a depth of 500 meters, it's providing a high-resolution view of fast-changing tropical ocean flows.

NOAA expects to finish deploying the new moorings by 2027, Grissom says. "As far as government time, that's just right around the corner." Each will beam back data every 10 minutes rather than once an hour. They'll be less power hungry and more resilient, allowing engineers to remotely access and reset instrumentation.

The moorings will also collect much more data. Each will have up to seven new sensors for measuring the shallow waters just below the surface, helping efforts to understand how the deep heat of El Niño rises through this "mixed layer" to cause weather at the surface. Meanwhile, other instruments suspended from the buoys will measure currents to depths of 315 meters to better understand the propagation of El Niño's heat. And several moorings will be relocated about 1000 kilometers north of the equator, to a cloud-covered region where cyclones and atmospheric rivers form.

Overall, NOAA's plans are "very exciting," and follow the TPOS 2020 recommendations, says Sophie Cravatte, a physical oceanographer at France's National Research Institute for Sustained Development. But she has a few concerns. Because the array is being reduced by seven moorings, the committee recommended compensating by doubling the number of Argo floats released in the tropical Pacific. But so far, NOAA and its partners have only upped those releases by 20%, Kessler says.

Perhaps the biggest hole is in the western Pacific, where Japan used to operate the Triangle Trans-Ocean Buoy Network (Triton), which completed TAO's equatorial line of sensors. Only one Triton buoy remains operational, Kessler says. China has promised to fill that gap, but the first buoys in its Ding array failed soon after deployment. "This western part is key for weather and [El Niño] forecasting," Cravatte says, "and it is missing." ■

ARCHAEOLOGY

The woman with the deer pendant

Pioneering technique gleans DNA from a Stone Age ornament, revealing its last wearer

By **Ann Gibbons**

Twenty thousand years ago, someone dropped a deer-tooth pendant in a cave in southwestern Siberia, where it lay until archaeologists excavated it in 2019. Now, researchers have caught a glimpse of its last wearer. After years of effort, Elena Essel, a graduate student at the Max Planck Institute for Evolutionary Anthropology (EVA), developed a way to extract DNA embedded in an artifact's porous surface by sweat and skin cells. Her team's analysis of the ornament, reported this week in *Nature*, shows it once adorned a woman whose ancestry lay far east of the cave.

"It's the first time to my knowledge that we have a nondestructive way to extract DNA from Paleolithic artifacts," says co-author Marie Soressi, an archaeologist at Leiden University. The technique promises a window into how, and by whom, ancient ornaments and tools were used. Human DNA gleaned from their surfaces could offer "new insight into cultural practices and social structure in ancient populations," says evolutionary biologist Beth Shapiro of the University of California, Santa Cruz.

Geneticists have tried for 20 years to extract human DNA from Stone Age tools, ornaments, and other artifacts. But although they can get ancient human DNA from bones, teeth, and hair, and even from sediments, they have failed with artifacts more than a few hundred years old.

Working in the lab of EVA biochemist Matthias Meyer, Essel took up the challenge in 2017 as part of her thesis. She and Meyer reasoned it should be possible to coax ancient human DNA from tools made of porous bone or teeth. (Stone tool surfaces are too smooth to bind with DNA.)

After 5 years of trial and error with different chemicals and methods, the EVA team found that submerging an entire artifact in a mild sodium phosphate buffer bath while slowly heating it from room temperature to more than 90°C worked best. This released the ancient DNA trapped in the bone gradually, starting with DNA on the surface and ending with DNA from deeper within the bone matrix, while leaving the artifact's surface texture unharmed. Like a "washing machine, you use higher temperatures for big stains," Essel says.

The method worked on test materials Soressi provided from sites in France. But it failed when Essel applied it to tools from sites in Europe where ancient humans and Neanderthals both lived. She hoped to learn which kind of human made the sophisticated artifacts, but the tools were either contaminated with modern human DNA or came from soils containing substances that suppressed the enzymes used to extract DNA.

A deer-tooth pendant from a Siberian cave yielded DNA left by a woman who lived some 20,000 years ago.

In 2019, however, archaeologist Maxim Kozlikin at the Siberian Branch of the Russian Academy of Sciences visited the lab. He pulled a plastic bag full of dirt out of his pocket, said it contained an artifact, and asked whether Meyer's team would like to try its luck. The sample came from a famed site called Denisova Cave, where ancient DNA from modern humans, Neanderthals, and Denisovans has been found in fossils and sediments, preserved by the cave's cool, consistent conditions.

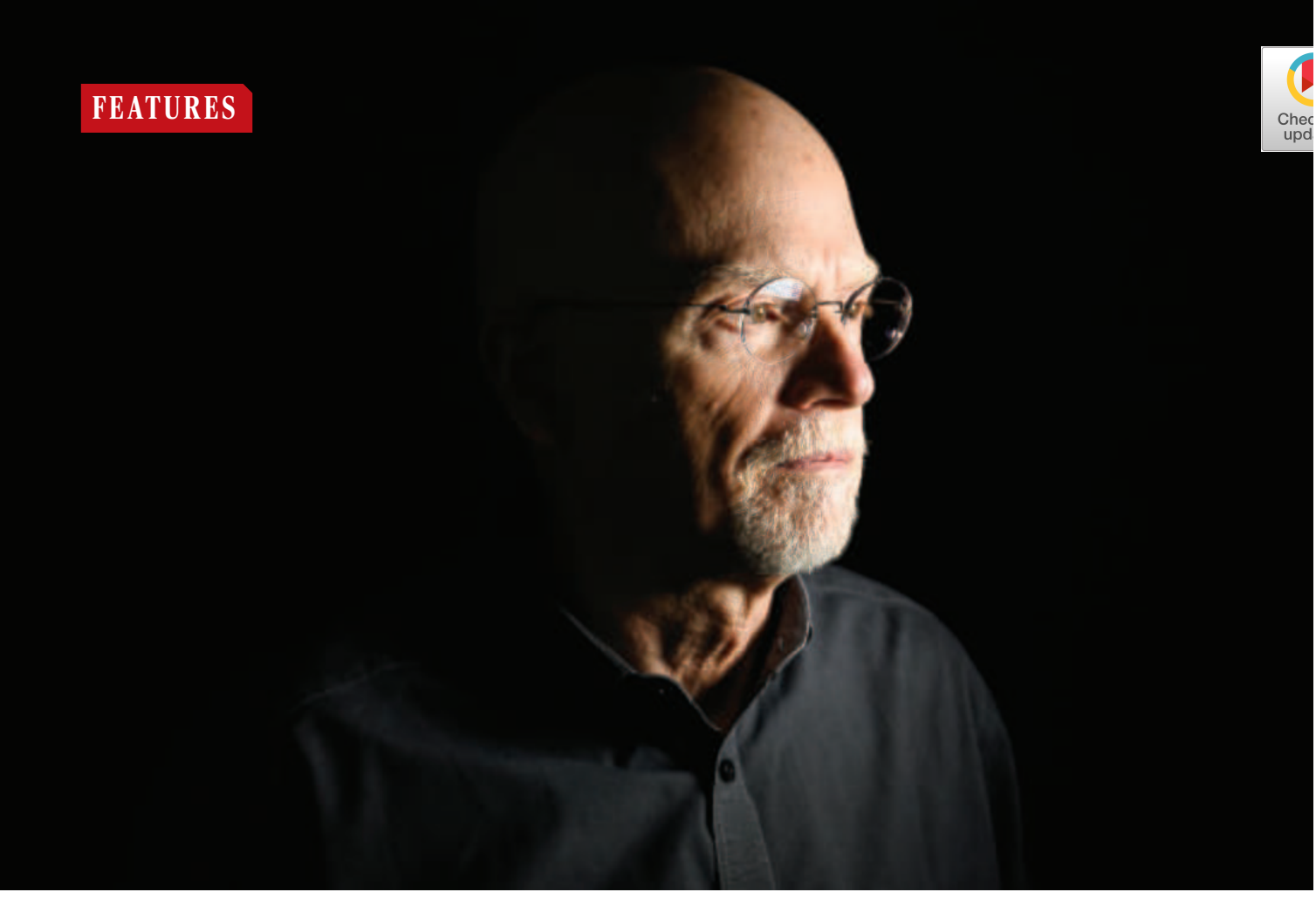
When Essel opened the bag in a clean room, she could see the tip of a deer tooth pendant encased in sediment. It turned out to be her lucky piece, yielding a "surprisingly high" amount of mitochondrial and nuclear DNA, she says. Two factors may have been key: The archaeologists wore gloves, masks, and coveralls when they excavated the pendant, and they put it in a plastic bag without removing the sediment clinging to it, which reduced contamination.

The EVA team found some of the ancient DNA came from wapiti, the species of elk whose tooth was used to make it. But some DNA was from a female modern human; its sequence revealed she was most closely related to people of the Maltinsko-buretskaya culture, known to have lived 2000 kilometers farther east near Lake Baikal—and who are among the ancestors of Siberians, Native Americans, and Bronze Age steppe herders. By comparing the DNA from both the woman and the elk with other ancient samples, the researchers dated the pendant to between 19,000 and 25,000 years ago.

The method is "a really important contribution to the field," says archaeologist Ron Pinhasi of the University of Vienna, who recently used a minimally invasive DNA-extraction method to determine which mammals' antlers were used to make bone tools. Archaeologist Solange Rigaud of CNRS and the University of Bordeaux says human DNA from artifacts could reveal whether men or women used tools or ornaments, or trace trade between populations.

Rigaud cautions, however, that the DNA on an artifact's surface reveals only the last person to wear or handle it, not necessarily its maker, because ornaments are traded and passed down generations. "You would need DNA from many tools in a tool workshop," she says. Ultimately, though, such studies could help answer questions researchers have debated for decades—such as whether Neanderthals or modern humans made the famous Châtelperronian ornaments in France.

The paper has come out just in time for a new field season, Essel says. She has a request for colleagues as they unearth new artifacts: "Please, please, please wear gloves and face masks if you want to look at the DNA of the people who were making these things and using them." ■



Several years ago, Tim Greenamyre, a neuroscientist and physician who directs the Pittsburgh Institute for Neurodegenerative Diseases at the University of Pittsburgh (Pitt), began to notice unsettling symptoms in his own body. He couldn't smell things. He was constipated. He was shouting and kicking in his sleep. His left arm didn't swing when he walked.

In July 2021, Greenamyre turned to a neurologist colleague to confirm the diagnosis he already suspected. He had Parkinson's disease, an illness he has devoted himself to treating and trying to cure. Over the course of his long and productive career, the 67-year-old has not only won the admiration of his patients and clinical colleagues, but also developed a widely used animal model of Parkinson's and contributed key insights into environmental triggers. That work exposed him to chemicals that induce the disease in rodents, a possible factor in his own illness.

"The irony is obvious," says Greenamyre, a shy man with a dry sense of humor and a penchant for practical jokes who, to the unpracticed eye, shows few if any signs of the disease. (For now, he says, medication is helping.)

For colleagues, the news has been shocking and heartbreaking. "I was so deeply affected I could not respond right away," says Laurie Sanders, a neuroscientist who studies Parkinson's at Duke University School of Medicine and is a former Greenamyre postdoc. "All I really honestly wanted to do was just drive to Pittsburgh and give him a hug."

Nearly all of Greenamyre's 200 or so Parkinson's patients, some of whom he has looked after for more than a decade, will learn of his diagnosis from this article, in which he is reluctantly making his condition public because it's becoming the subject of rumors. He worries it will distract them. "I want their visits to be focused on them, not on me," he says.

Compounding the irony, Greenamyre's diagnosis comes at a time of fresh optimism among Parkinson's researchers, who believe they might at last be closing in on treatments that could slow or stop the progression of the second most common neurodegenerative disease after Alzheimer's. Parkinson's disease afflicts roughly 1 million people in the United States, with nearly 90,000 new cases diagnosed each year (see graph, p. 452). Globally, more than 8.5 million people have the disease, which is the fastest growing neurological disease in the world by one estimate.

Tim Greenamyre sees new hope on the horizon for treating Parkinson's disease.

The classic motor symptoms of Parkinson's disease, described in 1817 by English surgeon James Parkinson, result from the degeneration of cells that produce the neurotransmitter dopamine in the substantia nigra, a midbrain area involved in movement control. This causes a variety of symptoms including, most commonly, tremors, muscle stiffness, and trouble with balance and coordination. As the disease advances, people may have difficulty speaking and initiating movement. Later, many develop dementia. Parkinson's itself doesn't kill, but its complications, particularly aspiration pneumonia due to difficulty swallowing, often do.

Dopamine, delivered in an oral drug combination of levodopa (L-Dopa) and carbidopa, has been the front-line therapy since it was approved in the U.S. in the 1970s, although it remains unavailable or unaffordable for many globally. The drug improves motor symptoms, but with time it begins to wear off more quickly, and sometimes-intolerable

TWIST OF FATE

A physician-scientist has probed Parkinson's disease for more than 30 years. Now, he has it

By **Meredith Wadman**

side effects—including involuntary, jerking movements called dyskinesia—frequently develop. Fifty years on, the failure of science to produce a therapy that stops the disease instead of targeting its symptoms has been painful and frustrating for patients and their families.

But now, "We are at an inflection point," says Todd Sherer, a neuroscientist and former Greenamyre postdoc who ran the Michael J. Fox Foundation for Parkinson's Research until 2021 and is now its chief mission officer. Such optimism is the outcome of decades of research, hastened by the revelations, beginning in 1997, of genetic mutations linked to the disease. Those discoveries opened the doors for scientists including Greenamyre to probe the disease's molecular mechanisms, and a gusher of publications resulted (see graph, p. 452). The new knowledge has in turn allowed companies to develop drugs and other experimental therapies that aim to slow, stop, or even prevent the disease (see graphic, p. 450). Today those therapies are entering clinical trials at a striking rate.

"There has been a shift because we are better understanding the disease and the targets," Sanders says, noting that more than 50 clinical trials attacking the roots of the disease are in progress. It's an extraordinary increase, she says, from the handful that were underway when she was starting out in the field 15 years ago.

Despite such hopeful signs, the anticipated breakthroughs may come too late for Greenamyre and his patients with Parkinson's. By the time the disease is diagnosed via a shaking hand or a dragging foot, it is thought to have been active under the radar for decades, causing the quieter symptoms such as constipation and loss of smell that Greenamyre experienced—and destroying roughly half of the dopamine-producing neurons in the substantia nigra. But Greenamyre remains determined to break the back of the disease.

"There's no good time to be diagnosed with Parkinson's disease," he said last fall when he won a $100,000 prize for research leadership from the Fox foundation. "But this is the best time in history to be diagnosed with Parkinson's disease."

GREENAMYRE GREW UP in New York's Westchester County, the son of a chemical engineer and a homemaker. In medical school at the University of Michigan, one of his first lecturers was physician Anne Buckingham Young, an enthusiastic new assistant professor whom Greenamyre recalls as a "force of nature." As Young described her Ph.D. work illuminating how the poison strychnine exerts its effects at receptors for the neurotransmitter glycine in the spinal cord, Greenamyre became entranced. "I thought it was so cool, the space she was in, where she could talk about basic pharmacological, physiological mechanisms that have clinical implications," he says. He joined her lab and developed a technique for visualizing neurotransmitter receptors in the brain, which he and Young used to investigate the mechanisms of Alzheimer's and Huntington diseases. He earned both an M.D. and a Ph.D., and by 1990, as he began running his own lab and seeing patients at a top-notch movement disorders clinic at the University of Rochester, he had co-authored 20 papers, two of them in *Science*.

He also began to study Parkinson's disease. Like others in the field, he was inspired by a 1983 *Science* paper that described a startling cluster of young people with sudden-onset Parkinson's who had turned up in Northern California hospitals. All had taken a street drug contaminated with a chemical called MPTP. Its toxic metabo-

lite, MPP+, had destroyed dopaminergic neurons in the substantia nigra, and scientists soon showed that MPP+ homes in on and inhibits complex I, the first enzyme in a crucial biochemical chain that converts food to energy in mitochondria, the tiny powerhouses inside cells. It was a tantalizing clue that mitochondrial damage might play a role in the disease.

Greenamyre and others knew MPP+ wasn't the only inhibitor of complex I. There were many, including some chemicals that studies were beginning to implicate in Parkinson's disease. But the classical inhibitor of this vital mitochondrial enzyme was a pesticide used by home vegetable gardeners called rotenone.

Rotenone was considered an "organic" insecticide because it was derived from the roots of certain plants; people also used it to kill fleas and ticks on their pets, and wildlife agencies used it to control invasive fish populations. But for Greenamyre, it was a way to plumb the mysteries of Parkinson's. By 1990, he had begun using radiolabeled rotenone, a fat-loving chemical that readily crosses biological membranes, including the blood-brain barrier, to map locations in the brain of complex I.

Around the same time, several labs found evidence in postmortem brain samples as well as platelets from patients that people with Parkinson's disease had defective complex I activity in their mitochondria. Greenamyre realized he might be able to mimic the disease in rats by exposing them to rotenone. He expected the chemical would addle mitochondria in every organ. And it did. But it had outsize toxic effects on just one cell type.

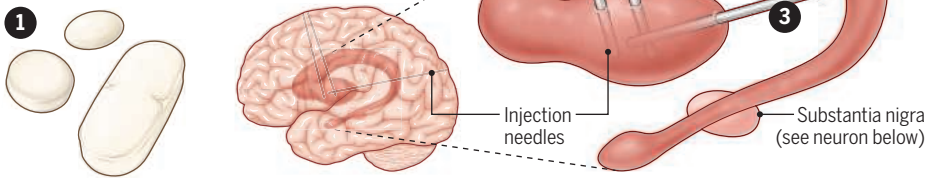
In 2000 Greenamyre, who by then was at Emory University, and his team published a seminal paper in *Nature Neuroscience*. They reported that giving rats a chronic intravenous infusion of rotenone selectively destroyed the same dopamine-producing neurons in the substantia nigra that degenerate in Parkinson's patients. Moreover, the surviving neurons contained fibrils, or tiny threads, of the protein alpha-synuclein, aggregated in clumps. Those aggregations closely resembled structures called Lewy bodies that are a signature of Parkinson's in human brains. What's more, the rats developed parkinsonian symptoms: unsteady movements and hunched postures, shaking paws and severe rigidity.

The work gave researchers the first animal model that captured both the classic motor symptoms and the hallmark pathology of the disease. It also increased suspicions that rotenone and other pesticides could trigger Parkinson's.

Taking aim at Parkinson's

Decades of research identifying genes and probing the biological mechanisms of Parkinson's disease have led to dozens of ongoing clinical trials that aim to slow or stop the disease. Some act on the dopamine-producing neurons in the substantia nigra that degenerate in Parkinson's. These cells project to nearby structures called the caudate nucleus and putamen.

Therapeutic strategies

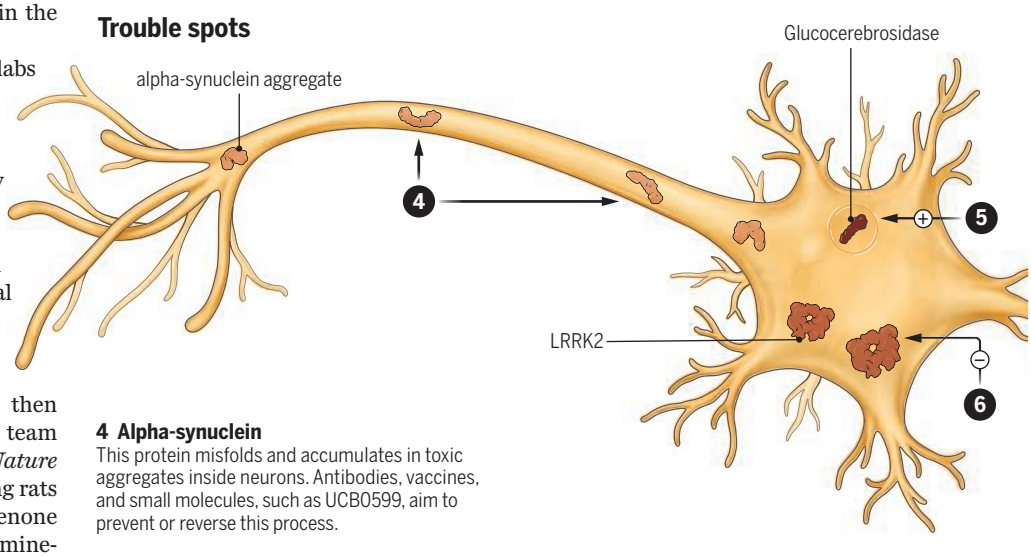


1 Oral and injectable drugs
Some drugs now in trials have specific targets within neurons (see below). Another approach repurposes self-injected diabetes drugs called GLP-1 receptor agonists. These may improve neuronal health by stimulating insulin secretion in the brain, which is often impaired in Parkinson's.

2 Gene therapies
This approach aims to boost survival or regeneration of dopaminergic neurons. One trial injects a nonreplicating virus carrying the gene for glial cell line–derived neurotrophic factor into the putamen.

3 Cell replacement therapies
An early trial is injecting healthy dopamine neurons derived from human embryonic stem cells into the putamen. Another approach uses mesenchymal stem cells to promote neuronal survival and regeneration.

Trouble spots



4 Alpha-synuclein
This protein misfolds and accumulates in toxic aggregates inside neurons. Antibodies, vaccines, and small molecules, such as UCB0599, aim to prevent or reverse this process.

5 Glucocerebrosidase
Suppressed activity of this enzyme in Parkinson's disease has been linked to accumulation of toxic alpha-synuclein. Researchers are trying gene therapy and drugs such as ambroxol, a repurposed cough medicine, to restore the activity.

6 LRRK2 (leucine-rich repeat kinase 2)
This enzyme is overactive in some patients, likely driving lysosomal damage, the buildup of toxic byproducts inside cells, and neurodegeneration. Drugs and synthetic genetic snippets called antisense oligonucleotides aim to tamp down LRRK2 activity.

TODAY, GREENAMYRE wonders whether his decades of research with rotenone and similar compounds may have caused his disease. "Because we didn't know as much, we weren't as careful," he says. "And I got exposed to things, and particularly rotenone, quite a bit."

It's February, and Greenamyre is musing aloud over a lunch bowl of rice, spinach, and goat cheese in his sunlit office, which adjoins his lab in the heart of Pitt's medical campus downtown. He has just come from a lab meeting where neuroscientist Emily Rocha, a former postdoc, presented a gnarly challenge she had encountered in her research, and a stop to look over some newly collected images of dopamine neurons with grad student Matthew Keeney. Former grad students and postdocs say Greenamyre has an encouraging but hands-off style of mentoring. "He had

a great ability to let you explore but then he knew when to pull you back when you were going down a rabbit hole," Sherer recalls.

Over lunch, Greenamyre explains that rotenone has to be dissolved in a solvent like dimethyl sulfoxide (DMSO) to make an infusible solution. That solution would spill on his gloves from time to time, and "the DMSO takes it right through the gloves and into your skin and right through your skin." He didn't make much of it at the time, he says, because he and his colleagues didn't really think there was much danger. "Till we did a little research on it," he adds. (A high-quality epidemiological study published in 2011 associated rotenone use with a 2.5-fold increase in risk of developing Parkinson's disease in farmers and their spouses.)

If rotenone did play a role in causing Greenamyre's disease, it likely added to underlying genetic vulnerabilities. He is one of the 90% of people with so-called idiopathic disease: Their Parkinson's has no clear genetic cause but almost certainly results from some combination of ill-defined genetic susceptibilities and environmental triggers. Greenamyre, for example, once had red hair, which epidemiological studies have associated with an elevated risk of Parkinson's. (Why this might be isn't understood.)

About 10% of Parkinson's cases, however, are clearly due to mutations in specific genes. After a 1997 *Science* publication identified the first of them, in the gene for alpha-synuclein, many scientists focused on genes and only genes. Soon they unearthed more inherited mutations—in genes encoding a protein called Parkin, an enzyme called LRRK2, and another enzyme, glucocerebrosidase. Researchers began to drill down on how the mutated genes were doing their damage.

These discoveries coincided with a tricky time in Greenamyre's personal life. A difficult divorce led to his move from Emory to Pitt in 2004. He also faced funding woes. Greenamyre's primary funder at the time, the Picower Foundation, had invested tens of millions with Bernie Madoff, the notorious Ponzi scheme operator; it was all lost. "Anything from $20 on up is welcome," Greenamyre told *Neurology Today* in 2009 when asked how he would keep his 13-member lab afloat.

Despite such obstacles, Greenamyre has spent the past 20 years trying to define the critical mechanisms that destroy neurons in Parkinson's, rather than chasing new causative genes as many others have done. He has stayed fixed on the complicated dance between genes and the environment, with an emphasis on pesticides and solvents that interact with key Parkinson's-associated genes.

In recent years, Greenamyre has become laser-focused on *LRRK2* (pronounced "lark 2"), which stands for leucine-rich repeat kinase 2. The enzyme it encodes is a master traffic controller, regulating the movement of proteins and compartments called vesicles inside cells. Several mutations in the gene send the enzyme's activity into overdrive. That ultimately damages the function of lysosomes, the cellular trash collectors that degrade unwanted proteins, and the damage is thought to contribute to the disease.

Mutations in *LRRK2* account for about 3% to 4% of all cases of Parkinson's. But in 2021, Greenamyre's group, led by postdoc Briana de Miranda, now an assistant professor at the University of Alabama at Birmingham, found that toxins can mimic their effects. They reported that the solvent trichloroethylene, a mitochondrial toxin used in dry-cleaning and degreasing metal, ramped up LRRK2 enzyme activity and induced Parkinson-like pathology in the brains of older rats. The work has provided grist for an incipient class action lawsuit from people who decades ago drank contaminated water at the Marine Corps base at Camp Lejeune in North Carolina. They allege that trichloroethylene in that water caused their Parkinson's disease.

Greenamyre's decades of work on environmental toxins also helped form the basis for a lawsuit filed by the Fox foundation and others against the U.S. Environmental Protection Agency (EPA) after the agency in 2021 reapproved the widely used herbicide paraquat. Like rotenone, paraquat causes rodents to develop parkinsonian symptoms and pathology, and other researchers have used it to make animal models of the disease. EPA argued that with safety measures in place, paraquat's benefits in agriculture outweighed its risks to human health, saying it had found "insufficient epidemiological evidence" to suggest a causal link to Parkinson's. Greenamyre filed an amicus brief in the case, chiding EPA for discounting numerous epidemiological and animal studies linking paraquat exposure to Parkinson's. (At least 50 other countries have banned its use.)

Since Greenamyre's groundbreaking 2000 paper using rotenone to create Parkinson-like disease in rats, EPA has twice examined the safety of that pesticide as well. In 2007, it restricted rotenone's use to control of invasive fish species—and manufacturers voluntarily withdrew it from the market for residential home and garden use. Last year, EPA imposed further safety restrictions on its use in fish kills and reiterated that the occupational risks of rotenone exposure, driven by skin exposure, "are of concern." But the agency in 2022 also reviewed the literature and concluded there is "insufficient evidence" to suggest a causal link between rotenone and Parkinson's.

Other scientists say epidemiology studies and lab work by Greenamyre and others are persuasive. "He has made a really compelling case for rotenone and other pesticides as very significant and relevant triggers for idiopathic Parkinson's disease," says Malú Gámez Tansey, a neuroscientist at the University of Florida College of Medicine. "It was a very important contribution."

Tim Greenamyre as a teenager. Red hair may be a risk factor for Parkinson's.

PHOTO: J. TIMOTHY GREENAMYRE

IN A CLINIC in the next building over from his lab, Greenamyre has donned a white coat and mentally prepared his list of questions for Barbara Frieze, 71, a Parkinson's patient he has been looking after for nearly 10 years. "How's your puppy?" he asks, an easy opener. She confides that the puppy has kept up her spirits as her husband, who has dementia, declines in a nursing home. Frieze visits him every day, which makes exercising tough, she explains when he gently prods her about her activity. (Aerobic exercise probably slows the progression of motor symptoms, and Greenamyre himself works out 7 days a week.)

He goes on to ask Frieze about her sleep, her balance, her mood, and her tremors. The latter are painfully visible when she extends her arms straight in front of her; her left hand continues to shake as she walks down a nearby hall at Greenamyre's request.

Back in the exam room, he says: "For 11 years in, you're doing great. Your walking is really good. The only way I would know you had Parkinson's was the tremor."

"When I leave here, I feel hopeful," Frieze says. "If I'm feeling dubious about the Parkinson's and I come here, I always feel good leaving."

The next patient, Carl Maskiewicz, is a mostly retired accountant in his late 60s who attends boxing and yoga classes six times a week. He praises Greenamyre's patience and ability to explain things. "I think sometimes he has a better idea of what I'm going through than people that have Parkinson's," Maskiewicz says.

A third patient, Robert Hannan, calls in by Zoom from Florida. The 84-year-old former CEO of a drugstore chain has had the disease for 25 years. He takes dopamine every 2 hours now, but as the drug wears off, his words become slurred and the tremor in his left hand reappears.

On this February day, Greenamyre did not tell the patients he saw about his own diagnosis. He emailed them soon after, though, not wanting them to be blindsided when the news came out. "I told him he was in my thoughts and prayers," says Maskiewicz, who has been Greenamyre's patient for 9 years. "And I offered to chat, 'If you ever just want somebody to talk to'—because sometimes that's the difficult part of this disease."

Like the million other Americans with Parkinson's, Greenamyre's patients are all waiting for something beyond dopamine. "It's still a disease without a cure," Hannan says. "I keep asking every time I come: 'Is there something close?'"

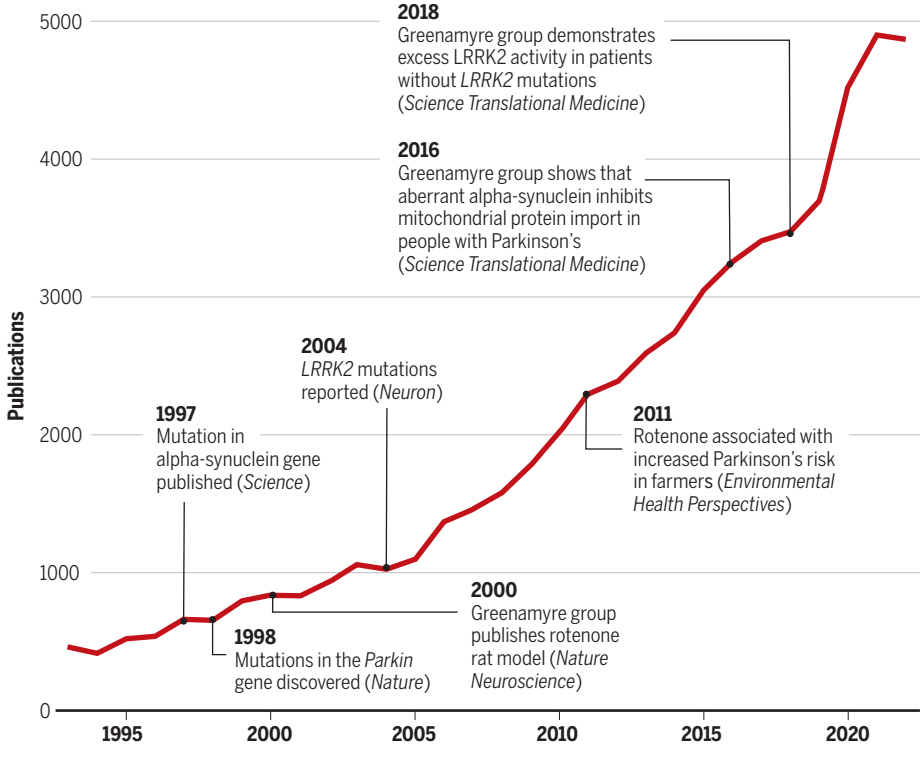
IN 2018, A NEW PAPER from Greenamyre's group, led by research scientist Roberto di Maio, pointed to one way forward. They had developed an assay to gauge LRRK2 enzyme activity in different kinds of brain cells and used it to show that the enzyme is overactive in dopamine-producing neurons in the substantia nigra—even in patients with no mutations in the gene.

Additional experiments revealed the underlying cascade of molecular events, culminating with the crippling of the cell's garbage disposal system. This, the team posited, resulted in accumulation of aberrant alpha-synuclein.

"This was the first data in actual patient tissues demonstrating … in patients with

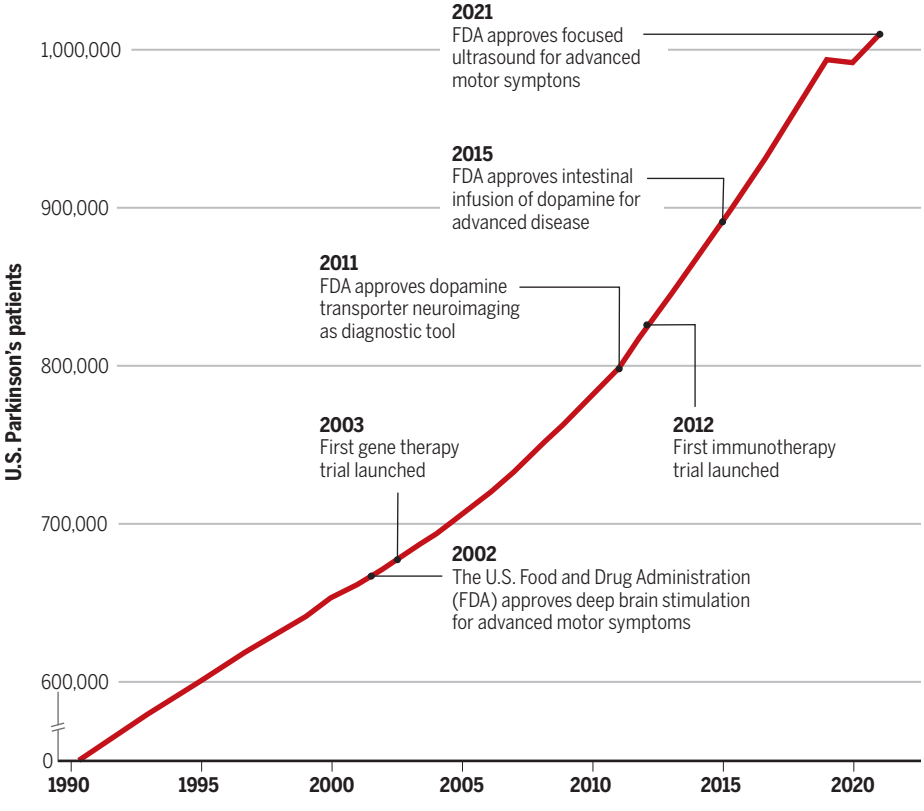
An explosion of research

After genetic discoveries opened new windows on Parkinson's, the number of papers probing its biology soared, as gauged by the number of PubMed publications with "Parkinson's" in their titles.



A growing problem

The aging of the United States has helped drive a big increase in Parkinson's prevalence, but the only approved therapies target symptoms rather than disease progression.



idiopathic Parkinson's disease that they do have increased LRRK2 kinase activity," says Carole Ho, chief medical officer and head of development at Denali Therapeutics.

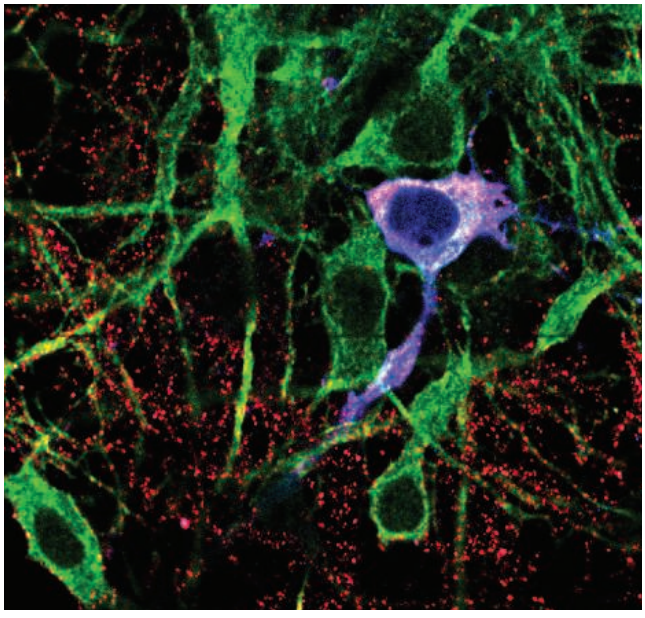
The paper strongly suggested inhibiting that activity might help many patients—and not just the 3% to 4% who harbor inherited *LRRK2* mutations. In rats given rotenone, the researchers found, an LRRK2 inhibitor blocked all the abnormal events typically caused by the pesticide. The finding gave new impetus to efforts by companies to test LRRK2 inhibitors in people; Denali's candidate is the furthest along. It has partnered with Biogen, which is now enrolling more than 1000 Parkinson's patients with and without *LRRK2* mutations to see whether an inhibitor of the enzyme slows progression of the disease.

Labs and clinical trials around the world are pursuing other drugs, gene therapy, and stem cells to replace lost dopamine neurons. Some target demonstrated biological culprits, for instance boosting activity of the enzyme glucocerebrosidase, which can be suppressed in Parkinson's, or blocking aggregation of toxic alpha-synuclein. Others aim to give general ammunition to the brain, such as pharmacological tools for better repair and maintenance of dopaminergic neurons.

There have been early disappointments. In 2021 and 2022, respectively, Biogen and AbbVie pulled the plug on clinical trials of antibodies that attack alpha-synuclein. Roche and Prothena Biosciences are pressing ahead despite similarly disappointing trial results last summer showing that their monoclonal antibody, prasinezumab, had no meaningful impact on disease progression.

Many researchers hope such conspicuous failures will become rarer. The underlying biology of Parkinson's varies widely between subsets of patients, and efforts are afoot to match the right people to the right trials so that each therapy is tested in those most likely to respond to it.

Identifying at-risk people long before motor symptoms appear could also help. Antibodies to alpha-synuclein might still work, for example, if given early enough, before extensive neuron loss. The Fox foundation is running a huge, long-term study looking for imaging, biologic, and genetic markers that could catch people in the earliest stages of the disease. That study produced landmark results last month: A report in *The Lancet Neurology* found that a spinal tap test for misfolded alpha-synuclein in the fluid bathing the spinal cord accurately diagnosed Parkinson's 88% of the time, including identifying it before people started to have motor symptoms. Although the test's invasiveness makes it unlikely to be used routinely, it will be a valuable tool for sorting patients for clinical trials and probing the disease's biology, scientists say. And a similar blood test, if promising early results bear out, may be available by the time therapies are ready to treat early-stage disease.



Oxidative damage (magenta) in a rat dopaminergic neuron (blue) after treatment with the pesticide rotenone.

IMAGE: MATTHEW T. KEENEY/UNIVERSITY OF PITTSBURGH

AFTER MONTHS of worrying about his own lack of smell and other symptoms, Greenamyre finally turned to a trusted colleague, Edward Burton, a Pitt neurologist and neuroscientist who studies Parkinson's. Burton says he is sometimes burdened by his training; he analyzes people without meaning to, with an eye expertly attuned to the most subtle abnormalities of movement.

In the summer of 2019, Burton was returning from a Gordon Research Conference on Parkinson's Disease that Greenamyre had chaired, and the pair was sitting in an airport lounge. Someone else asked Greenamyre a question, and Burton thought his friend's head turned toward the questioner just a shade too slowly.

Soon after, the COVID-19 pandemic descended and Burton did not see his colleague for the better part of 2 years. Then in July 2021, when both men were back in the lab and the clinic, Greenamyre asked Burton to meet confidentially with him.

"He was clearly quite worried," Burton recalls. (Greenamyre gave Burton permission to discuss his case for this article.) The sleep disturbances Greenamyre had been having for several years, rapid eye movement sleep behavior disorder (RBD), did not bode well. RBD is an early symptom of Parkinson's, but it can also be a harbinger of rarer, more rapidly progressive neurodegenerative diseases.

Burton listened carefully to his friend's account of his symptoms, then gave him the same motor tests that Greenamyre had given his own patients countless times. When Greenamyre tried to quickly open and close his index fingers and thumbs, his left hand lagged subtly. Burton had already noted in casual passings in the hall that his colleague's gait was asymmetrical; he was not swinging his left arm as he did his right.

"With it being somebody that I like very much and respect a great deal, examining him, every physical sign was like having my stomach ripped out," Burton recalls.

Greenamyre's symptoms and the slow progression of his illness led Burton to conclude this was likely Parkinson's and not something still worse. He started Greenamyre on dopamine therapy, and his symptoms rapidly improved, clinching the diagnosis.

"There was some slowness. And that totally changed," recalls Greenamyre's partner, Pitt neuroscientist Teresa Hastings. "His whole posture seemed to change once he started on the L-Dopa. It's like his muscles were tuned up."

As their interview ended that summer day in 2021, Greenamyre reached out to shake Burton's hand. "He thanked me for evaluating him," Burton says. "Tim has really handled this with dignity and aplomb."

IN A BREAK between patients in February, the typically reserved Greenamyre admitted to some worries about what comes next. "When I have some weird sensation or something—you know, everybody has these things. But now you wonder whether that's part of the disease … a tingling here or a tingling there."

Still, he remains optimistic, as he was when his patient Hannan asked whether a better therapy was finally coming into sight. "I think we're getting to the heart of it. But it's not ready for prime time," Greenamyre said. "Things are just beyond the horizon." ■

ECOLOGY

The chemical ecology of locust cannibalism

An anticannibalistic signaling pathway offers a new understanding of locust swarm formation

By **Iain D. Couzin**[1,2,3] and **Einat Couzin-Fuchs**[2,3]

Locust plagues threaten food security across the globe, with infestations estimated to affect the livelihood of 1 in 10 people (*1*). Cannibalistic interactions have been implicated in the formation and maintenance of these swarms (*2*–*4*). The threat of consumption by others is thought to have resulted in selection pressure for flightless juvenile locusts to use both visual and tactile information to move away from those approaching and to move toward those moving away (*4*). The outcome is large-scale, coordinated motion out of nutrient-poor areas, with the benefit of conspecifics being a source of essential, yet often scarce, nutrients along the way (*2*, *3*). On page 537 of this issue, Chang *et al*. (*5*) report that chemical communication plays a central role in regulating cannibalistic interactions in the migratory locust (*Locusta migratoria*). This discovery contributes new understanding about the role of cannibalism in locust ecology and suggests possible targets for their control.

Cannibalistic behavior is ubiquitous in the natural world (*6*) and may once have been common among human ancestors (*7*). Cannibalism has the potential to act as a key modulator in population growth and dispersal (*8*), an ecological role that may be especially profound for certain swarm-forming insects. In Mormon crickets (*Anabrus simplex*), for example, cannibalism is rife, with insects marching in unison not through some cooperative tendency, but rather with individuals being on what is effectively a "forced march" to avoid the threat of attack (*9*).

Even when cannibalism is less prevalent, such as among the two principal locust pest species—the migratory locust and desert locust (*Schistocerca gregaria*)—it has the potential to exert strong selection pressure on behavior. Thus, the proximate mechanisms that regulate social interactions among locusts, such as visually mediated attraction and avoidance, likely reflect this cannibalistic aspect of their evolutionary ecology (*4*). If food supplies are abundant, individuals will tend to persist as shy and cryptic "solitarious" grasshoppers. Population upsurges start with atypical rainfall that provides temporarily good breeding conditions, followed by food limitation. The increase in local density experienced by individuals, which ever closely aggregate on limited food resources, causes them to become actively "gregarious" (*10*). Moving in the same direction as others minimizes contact, and thus cannibalism (*4*), while facilitating collective motion out of areas of low food availability.

It is this extreme phenotypic plasticity, which involves rapid and substantial changes in physiology and behavior, that allows locusts to be so well suited to highly variable and unpredictable environments. Yet despite behavioral similarities, the differences in the molecular mechanisms by which the phylogenetically distant *S. gregaria* and *L. migratoria* regulate density-dependent plasticity could provide valuable insights into the evolutionary ecology of swarm formation (*10*, *11*). Both species exhibit a striking density-dependent color change when gregarious. This acts as a warning signal to predators (*12*). However, it is only the gregarious migratory locust that is known to synthesize a potent toxin—hydrogen cyanide. Doing so is likely to be costly, owing to energetic costs and the risk of self-poisoning, and only occurs under conditions of stress (*13*). However, by emitting a precursor in the hydrogen cyanide biosynthesis pathway, phenylacetoni-

[1]Department of Collective Behavior, Max Planck Institute of Animal Behavior, Konstanz, Germany. [2]Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany. [3]Department of Biology, University of Konstanz, Konstanz, Germany.
Email: icouzin@ab.mpg.de

A migratory locust (*Locusta migratoria*) consumes a conspecific. Juveniles deter cannibalism by emitting the chemical phenylacetonitrile (PAN).

trile (PAN), they provide an "honest" (i.e., truthful) olfactory signal of their toxicity to potential predators (*13*). Gregarious desert locusts, by contrast, bias their diet toward toxic plants, and so their gut contents confer protection. This difference may imply that desert locusts are incapable of producing PAN, but this is not the case. Gregarious sexually mature adult male *S. gregaria* emit PAN as an aversive volatile to rival males during courtship (*14*, *15*). Thus, the emission of PAN has evolved, likely independently, as an aversive signal in both species. Neither species produces PAN when solitarious, using instead both cryptic coloration and behavior to avoid predation.

Chang *et al*. identified a PAN-sensitive olfactory receptor (LmOR70a) and edited mutations in the encoding gene that make this receptor nonfunctional, thereby creating a means to test how the reception of PAN affects behavior. They also engineered insects that are unable to produce PAN by introducing a loss-of-function mutation in *CYP305M2*, a crucial gene for PAN biosynthesis (*13*). Behavioral experiments with these gene-edited insects demonstrated not only that PAN is a generally aversive compound associated with crowding but also that it plays a specific role in suppressing cannibalism among juvenile migratory locusts. (It is notable that cannibalism is expected to play a greater role in the swarms of flightless juveniles that inevitably precede the adult flying swarms.) Juveniles that were unable to produce PAN were both more frequently attacked and preferentially consumed compared with wild-type nymphs. Conversely, *LmOR70a*-edited locusts, which are unable to detect PAN, readily consumed PAN-releasing individuals. Together, this suggests that PAN may be an honest signal to deter cannibalism (warning of the locusts' capability to produce hydrogen cyanide if attacked), but whether this results from direct toxicity or an association with another physiological and/or behavioral mechanism remains to be determined.

The finding that PAN mediates within-species communication in *L. migratoria* narrows the disparities between its role in the two locust species, which is thus-far associated with male-male antiharassment in *S. gregaria* and antipredation in *L. migratoria*. It could be that migratory locusts, which are adapted to grassland biomes in which plant toxins are less prevalent, cannot use gut contents as a chemical deterrent. Why juvenile *S. gregaria* do not use the PAN pathway to suppress cannibalism is less clear.

The work of Chang *et al*. serves as an important step in understanding the mechanisms that mediate the intricate balance between aggregation and competition in locusts. Because the PAN pathway regulates cannibalistic interactions, which in turn can drive mass migration, it may also prove to be a promising target for the future development of locust control agents. ■

REFERENCES AND NOTES

1. A. Steedman, *Locust Handbook* (Natural Resources Institute, ed. 3, 1990).
2. S. Bazazi *et al*., *Curr. Biol.* **18**, 735 (2008).
3. S. Bazazi *et al*., *Proc. R. Soc. London Ser. B* **278**, 356 (2011).
4. V. Guttal, P. Romanczuk, S. J. Simpson, G. A. Sword, I. D. Couzin, *Ecol. Lett.* **15**, 1158 (2012).
5. H. Chang *et al*., *Science* **380**, 537 (2023).
6. M. A. Elgar, B. J. Crespi, *Cannibalism: Ecology and Evolution Among Diverse Taxa* (Oxford Univ. Press, 1992).
7. T. D. White, *Sci. Am.* **285**, 58 (2001).
8. L. R. Fox, *Annu. Rev. Ecol. Syst.* **6**, 87 (1975).
9. S. J. Simpson, G. A. Sword, P. D. Lorch, I. D. Couzin, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4152 (2006).
10. M. P. Pener, S. J. Simpson, *Adv. Insect Physiol.* **36**, 1 (2009).
11. H. Song, *J. Orthoptera Res.* **14**, 235 (2005).
12. G. A. Sword, S. J. Simpson, O. T. M. El Hadi, H. Wilps, *Proc. Biol. Sci.* **267**, 63 (2000).
13. J. Wei *et al*., *Sci. Adv.* **5**, eaav5495 (2019).
14. K. Seidelmann, H. J. Ferenz, *J. Insect Physiol.* **48**, 991 (2002).
15. D. A. Cullen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2200759119 (2022).

ACKNOWLEDGMENTS

The authors thank G. Sword for discussions and the Deutsche Forschungsgemeinschaft (DFG) (grant 2117-422037984) for support.

10.1126/science.adh5264

BIOCHEMISTRY

Some animals make plant sterols

Most animals abandoned plant sterols early in evolution, but some still depend on them

By **Jochen Brocks** *and* **Ilya Bobrovskiy**

Cholesterol is a flat, rigid molecule that slots into animal cell membranes, controlling physical properties such as membrane rigidity, fluidity, permeability, and curvature. Other eukaryotes have taken sterol biosynthesis further by adding alkyl groups to the side chain, such as phytosterols found in plants. However, it remains an enigma why the membranes of nearly all animals use cholesterol alone, whereas most other eukaryotes, including plants and fungi, contain complex mixtures of various sterols in their membranes. On page 520 of this issue, Michellod *et al*. (*1*) report the discovery of gutless worms (*Olavius algarvensis*) that endogenously produce cholesterol as well as phytosterols that are usually found in plants. By searching for the remains of such sterols in fossils, it may be possible to map the evolution of sterol biosynthesis in the animal kingdom and eventually understand why humans exclusively produce cholesterol.

Cholesterol is a lipid composed of 27 carbon atoms forming four rings and an eight-carbon side chain. The side chain interacts with other lipids in the interior of the membrane. Sitosterol, the major sterol of land plants, has two additional carbon units in the side chain at position C-24. The plant enzyme responsible for the production of this distinct feature is the C-24 sterol methyltransferase (SMT). SMT can add a methyl ($-CH_3$) group to sterol side chains but can also act twice to yield sterols with ethyl functionality ($-CH_2CH_3$). Such alkylated sterols are widely distributed in all major branches of Eukarya, including algae, plants, fungi, and most single-celled

Research School of Earth Sciences, The Australian National University, Canberra ACT 2601, Australia.
Email: jochen.brocks@anu.edu.au

A migratory locust (*Locusta migratoria*) consumes a conspecific. Juveniles deter cannibalism by emitting the chemical phenylacetonitrile (PAN).

trile (PAN), they provide an "honest" (i.e., truthful) olfactory signal of their toxicity to potential predators (*13*). Gregarious desert locusts, by contrast, bias their diet toward toxic plants, and so their gut contents confer protection. This difference may imply that desert locusts are incapable of producing PAN, but this is not the case. Gregarious sexually mature adult male *S. gregaria* emit PAN as an aversive volatile to rival males during courtship (*14*, *15*). Thus, the emission of PAN has evolved, likely independently, as an aversive signal in both species. Neither species produces PAN when solitarious, using instead both cryptic coloration and behavior to avoid predation.

Chang *et al*. identified a PAN-sensitive olfactory receptor (LmOR70a) and edited mutations in the encoding gene that make this receptor nonfunctional, thereby creating a means to test how the reception of PAN affects behavior. They also engineered insects that are unable to produce PAN by introducing a loss-of-function mutation in *CYP305M2*, a crucial gene for PAN biosynthesis (*13*). Behavioral experiments with these gene-edited insects demonstrated not only that PAN is a generally aversive compound associated with crowding but also that it plays a specific role in suppressing cannibalism among juvenile migratory locusts. (It is notable that cannibalism is expected to play a greater role in the swarms of flightless juveniles that inevitably precede the adult flying swarms.) Juveniles that were unable to produce PAN were both more frequently attacked and preferentially consumed compared with wild-type nymphs. Conversely, *LmOR70a*-edited locusts, which are unable to detect PAN, readily consumed PAN-releasing individuals. Together, this suggests that PAN may be an honest signal to deter cannibalism (warning of the locusts' capability to produce hydrogen cyanide if attacked), but whether this results from direct toxicity or an association with another physiological and/or behavioral mechanism remains to be determined.

The finding that PAN mediates within-species communication in *L. migratoria* narrows the disparities between its role in the two locust species, which is thus-far associated with male-male antiharassment in *S. gregaria* and antipredation in *L. migratoria*. It could be that migratory locusts, which are adapted to grassland biomes in which plant toxins are less prevalent, cannot use gut contents as a chemical deterrent. Why juvenile *S. gregaria* do not use the PAN pathway to suppress cannibalism is less clear.

The work of Chang *et al*. serves as an important step in understanding the mechanisms that mediate the intricate balance between aggregation and competition in locusts. Because the PAN pathway regulates cannibalistic interactions, which in turn can drive mass migration, it may also prove to be a promising target for the future development of locust control agents. ■

REFERENCES AND NOTES

1. A. Steedman, *Locust Handbook* (Natural Resources Institute, ed. 3, 1990).
2. S. Bazazi *et al*., *Curr. Biol.* **18**, 735 (2008).
3. S. Bazazi *et al*., *Proc. R. Soc. London Ser. B* **278**, 356 (2011).
4. V. Guttal, P. Romanczuk, S. J. Simpson, G. A. Sword, I. D. Couzin, *Ecol. Lett.* **15**, 1158 (2012).
5. H. Chang *et al*., *Science* **380**, 537 (2023).
6. M. A. Elgar, B. J. Crespi, *Cannibalism: Ecology and Evolution Among Diverse Taxa* (Oxford Univ. Press, 1992).
7. T. D. White, *Sci. Am.* **285**, 58 (2001).
8. L. R. Fox, *Annu. Rev. Ecol. Syst.* **6**, 87 (1975).
9. S. J. Simpson, G. A. Sword, P. D. Lorch, I. D. Couzin, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4152 (2006).
10. M. P. Pener, S. J. Simpson, *Adv. Insect Physiol.* **36**, 1 (2009).
11. H. Song, *J. Orthoptera Res.* **14**, 235 (2005).
12. G. A. Sword, S. J. Simpson, O. T. M. El Hadi, H. Wilps, *Proc. Biol. Sci.* **267**, 63 (2000).
13. J. Wei *et al*., *Sci. Adv.* **5**, eaav5495 (2019).
14. K. Seidelmann, H. J. Ferenz, *J. Insect Physiol.* **48**, 991 (2002).
15. D. A. Cullen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2200759119 (2022).

ACKNOWLEDGMENTS

The authors thank G. Sword for discussions and the Deutsche Forschungsgemeinschaft (DFG) (grant 2117-422037984) for support.

10.1126/science.adh5264

BIOCHEMISTRY

Some animals make plant sterols

Most animals abandoned plant sterols early in evolution, but some still depend on them

By **Jochen Brocks** *and* **Ilya Bobrovskiy**

Cholesterol is a flat, rigid molecule that slots into animal cell membranes, controlling physical properties such as membrane rigidity, fluidity, permeability, and curvature. Other eukaryotes have taken sterol biosynthesis further by adding alkyl groups to the side chain, such as phytosterols found in plants. However, it remains an enigma why the membranes of nearly all animals use cholesterol alone, whereas most other eukaryotes, including plants and fungi, contain complex mixtures of various sterols in their membranes. On page 520 of this issue, Michellod *et al*. (*1*) report the discovery of gutless worms (*Olavius algarvensis*) that endogenously produce cholesterol as well as phytosterols that are usually found in plants. By searching for the remains of such sterols in fossils, it may be possible to map the evolution of sterol biosynthesis in the animal kingdom and eventually understand why humans exclusively produce cholesterol.

Cholesterol is a lipid composed of 27 carbon atoms forming four rings and an eight-carbon side chain. The side chain interacts with other lipids in the interior of the membrane. Sitosterol, the major sterol of land plants, has two additional carbon units in the side chain at position C-24. The plant enzyme responsible for the production of this distinct feature is the C-24 sterol methyltransferase (SMT). SMT can add a methyl ($-CH_3$) group to sterol side chains but can also act twice to yield sterols with ethyl functionality ($-CH_2CH_3$). Such alkylated sterols are widely distributed in all major branches of Eukarya, including algae, plants, fungi, and most single-celled

Research School of Earth Sciences, The Australian National University, Canberra ACT 2601, Australia.
Email: jochen.brocks@anu.edu.au

protists. It is thus likely that the Last Eukaryotic Common Ancestor (LECA) possessed SMT (*2*). The most ancient animals, the sponges, likewise possess SMTs and make a large variety of alkylated sterols. However, for all other animals (eumetazoans), the activity of SMTs has been controversial. In the 1970s and 1980s, these animals were extensively screened for their sterol composition, revealing that although most only contain cholesterol, some eumetazoan species among echinoderms, cnidarians, and possibly bivalves are capable of synthesizing C-24-alkylated sterols as major membrane components (*3*, *4*). Yet, such experiments may be prone to contamination through sterol-producing microorganisms and gut contents (*4*).

Michellod *et al*. chose an ideal species to investigate sterol biosynthesis: *O. algarvensis* are gutless worms that do not need external food sources because they are sustained by bacterial symbionts. The authors found that sitosterol is a major lipid in the worms, comprising up to 60% of the total sterol content, alongside cholesterol. Intriguingly, they also found that *O. algarvensis* possesses not only a full biosynthetic pathway for cholesterol but also a functional SMT that can perform two consecutive methylations on sterol intermediates, yielding sitosterol. Furthermore, this *SMT* gene is widespread among animals: It occurs in other annelids; in a subgroup of rotifers (microscopic aquatic invertebrates); and in a variety of cnidarians, mollusks, and sponges. These genetic results, together with an unpublished preprint (*5*), have finally confirmed that many eumetazoans can produce more than just cholesterol.

To investigate whether phytosterol biosynthesis is performed by the gutless annelids, Michellod *et al*. measured the stable carbon isotopic composition of sterols and detected intriguingly large offsets between cholesterol and sitosterol and between two *Olavius* species. This may indicate that the two sterols are synthesized from distinct pools of substrate or that gutless annelids supplement endogenously synthesized sterols with sterols from the environment. Measuring the carbon isotopic composition of sterol intermediates and substrates provided by the symbionts may help trace the origins of the annelid sterols in the future.

Sterols modify the physical properties of membranes and are involved in the formation of lipid rafts, which are distinct regions in membranes that function as sorting platforms for proteins and may facilitate signal transduction (*6*). However, little is known about how particular sterol structures affect these properties. There are hints that methylated fungal sterols play a role in the protection of yeast during desiccation and rehydration (*7*). The ethyl group of phytosterols may enhance cohesion with other lipids, providing improved membrane ordering over large temperature ranges, and specific phytosterols in plants are involved in the response to gravity, helping shoots to find their way up, and pathogen attack (*8*). The incorporation of two or more different types of sterols in the same membrane may further extend the temperature range of membrane-associated processes and may also facilitate cell asymmetry (*6*, *9*). Yet, much of this is uncertain because it is difficult to disentangle the numerous impacts caused by the manipulation of sterol compositions in living organisms. *O. algarvensis* may become the ideal model with which to study the functions of particular sterols in animals as well as possible roles of sterols in host-symbiont interactions.

With the findings of Michellod *et al*., a major question is whether phytosterol production in worms was directly inherited from the last common ancestor of living animals, followed by multiple *SMT* losses, as suggested in an unpublished preprint (*5*), or conversely, whether *SMT* was lost early in animal evolution and later regained by several instances of horizontal gene transfer—for example, from algae and protists to animals, as suggested by Michellod *et al*. A more comprehensive phylogenetic tree of *SMT* is needed to resolve exactly where phytosterol production in various animals comes from and to determine when other eumetazoans started to rely on cholesterol alone.

The fossil record provides some clues. It may be possible to estimate the timing of *SMT* gene losses and gains by analyzing sterol remains extracted from individual fossils (*10*, *11*). Steroid biomarkers from ~550-million-year-old fossils of the Ediacara biota, which comprise some of the oldest known animals, indicate that the possible stem-group eumetazoan *Dickinsonia*, the stem-bilaterian *Kimberella*, and the annelid *Calyptrina* already predominantly or solely contained cholesterol, with no evidence for C-24–alkylated sterols (*12*, *13*). *Dickinsonia* contained >99% cholesteroids (*13*), and *Kimberella* and *Calyptrina* already had digestive systems that preferentially absorbed cholesterol from food (*12*). The disuse of side chain alkylated sterols may thus have emerged very early in the history of animals, and the analysis of biomarkers from younger fossils may help to map the use of *SMT* genes through the evolution of animals, including worms that had the guts to defy sterol conventions. ■

Molecular analysis of *Dickinsonia*, one of the earliest animals on Earth, has revealed the presence of cholesteroids, indicating an early abandonment of phytosterols in animal evolution.

REFERENCES AND NOTES

1. D. Michellod *et al*., *Science* **380**, 520 (2023).
2. E. Desmond, S. Gribaldo, *Genome Biol. Evol.* **1**, 364 (2009).
3. A. Kanazawa, *Fish. Sci.* **67**, 997 (2001).
4. L. J. Goad, *Pure Appl. Chem.* **53**, 837 (1981).
5. D. Gold *et al*., *Res. Square* 10.21203/rs.3.rs-1686449/v1 (2022).
6. E. J. Dufourc, *J. Chem. Biol.* **1**, 63 (2008).
7. S. Dupont, L. Beney, T. Ferreira, P. Gervais, *Biochim. Biophys. Acta Biomembr.* **1808**, 1520 (2011).
8. S. I. Aboobucker, W. P. Suza, *Front. Plant Sci.* **10**, 354 (2019).
9. A. Rogowska, A. Szakiel, *Phytochem. Rev.* **19**, 1525 (2020).
10. I. Bobrovskiy, J. M. Hope, A. Krasnova, A. Ivantsov, J. J. Brocks, *Nat. Ecol. Evol.* **2**, 437 (2018).
11. K. Grice, A. I. Holman, C. Plet, M. Tripp, *Minerals* (Basel) 9, 158 (2019).
12. I. Bobrovskiy, A. Nagovitsyn, J. M. Hope, E. Luzhnaya, J. J. Brocks, *Curr. Biol.* **32**, 5382 (2022).
13. I. Bobrovskiy *et al*., *Science* **361**, 1246 (2018).

10.1126/science.adh8097

HIGHER EDUCATION

A brief intervention to improve college success and equity

A large-scale experiment focused on social belonging shows impacts on first-year full-time completion

By **Nicholas A. Bowman**

Around the world, substantial proportions of students who begin postsecondary education do not complete their degree (*1*, *2*). Overall graduation rates often mask considerable differences by race, gender, age, and institutional type (*1*, *2*). In an effort to promote student success and equity, social psychologists have created interventions that seek to reframe challenges that students may face in their college environments to help them overcome obstacles and persist in school. On page 499 of this issue, Walton *et al.* (*3*) describe an ambitious randomized controlled experiment that tested a social-belonging intervention among >26,000 incoming students at 22 US colleges and universities. This brief intervention increased the proportion of students who completed their first year as full-time students, which is a key leading indicator of future college graduation (*4*).

This social-belonging intervention sought to reduce students' concerns about whether they will belong in college by conveying two messages: First, many students are concerned about belonging and may experience substantial doubts and challenges early on, such as difficulty making friends or receiving poor grades. Second, students generally overcome these challenges and ultimately feel a sense of belonging. The intervention delivered these messages by sharing with participants a combination of survey results and personal stories from older students. Study participants were then asked to reflect on these themes and write about why belonging concerns are likely to diminish over time during college, perhaps drawing upon their own personal experiences with transitions. A focus on belonging constitutes one potential mechanism for improving success and equity, and this type of belonging intervention has been used in prior research. Other psychological interventions have focused on students' perceptions that they can improve their achievement and intelligence through hard work and appropriate strategies or that their coursework will help them achieve their own valued goals (*5*, *6*).

Walton *et al.* expand on prior research in three important ways. First, whereas previous social-psychological interventions have often been implemented in classrooms or one-on-one settings, this intervention was conducted online and sent to entire cohorts of incoming college students (usually as one of several administrative tasks that students completed before their initial college enrollment). Despite potential limitations (e.g., difficulty in fully maintaining students' attention to the intervention), this online administration provides a straightforward way to reach thousands of students before they start college.

"A focus on belonging constitutes one potential mechanism for improving success and equity..."

Next, whereas previous research has often occurred with students at selective universities, this study examined a rather broad range of institutions, which is notable when a meta-analysis of various psychological interventions and academic outcomes found considerably larger effect sizes at selective institutions (*6*). Walton *et al.* conducted a formal analysis of the generalizability of their findings, which suggested that these can be extrapolated to a population of 749 colleges and universities that enroll more than one million incoming full-time, degree-seeking students each year.

And lastly, from a theoretical perspective, this study provides direct insights into when and for whom these social-belonging interventions are most likely to work. Researchers have frequently demonstrated that students who face negative stereotypes based on their identities (e.g., Black and Latinx students, students from lower socioeconomic backgrounds, women in math and science coursework) exhibit notably larger benefits from psychological interventions than do other students (*6*–*8*). This is consistent with prevailing theory that highlights how these groups of students receive implicit and explicit messages casting doubt on their potential to belong in college and succeed academically (*9*).

However, a broad focus on students' identities may only tell part of the story. Walton *et al.* argue for the importance of "local-identity groups" for determining who may benefit most from the intervention; these groups consist of a particular combination of race, first-generation status, college, and cohort. Walton *et al.* suggest that two criteria must be met for the intervention to be successful with each of these local-identity groups: (i) students face vulnerability to worries about belonging in college, and (ii) students have meaningful opportunities to belong in college. This second criterion has often been overlooked in previous research; a psychological intervention may try to persuade students that they will eventually belong, but students may not believe or internalize that message if they consistently encounter a hostile campus environment. Walton *et al.* demonstrated that the intervention was most successful when both of these criteria were met, which therefore led to greater equity in full-time completion.

Among local-identity groups who fit the two criteria, the intervention yielded roughly a 2 percentage-point increase in the likelihood of completing the first year as a full-time student. This effect could be perceived as trivial, but it may be practically important. Based on the generalizability analyses, this intervention would lead to approximately 12,000 additional students completing full-time enrollment every year if it were administered at those 749 colleges and universities. Social-psychological interventions frequently have substantial effect sizes if the outcomes are closely related to the theoretical mechanism of interest (*5*, *7*), but effect sizes are much smaller when the outcomes measure academic achievement and are even smaller for college retention or persistence (*6*–*8*). Given the myriad factors that may influence whether a student leaves or stays in college, 2 percentage points for a short precollege intervention seems notable.

Future research is necessary to further explore the conditions in which psychological interventions may be effective. In terms of institutional type, a national study showed that a sense of belonging in college was only related to mental health and college persistence at 4-year institutions (*10*), which suggests that a social-belonging intervention may not work at 2-year institutions. Moreover, although substantial

Department of Educational Policy and Leadership Studies, University of Iowa, Iowa City, IA, USA.
Email: nick-bowman@uiowa.edu

racial disparities in college graduation occur at predominantly white institutions, Black students graduate at the same rate as white students from 4-year institutions where at least half of undergraduates identify as Black, and there is also no graduation gap between Latinx and white students at majority-Latinx 4-year institutions (*11*). As a result, the social-belonging intervention may not be effective for these well-represented local-identity groups by race and college.

Additional research may also help explore features of interventions that are most effective. Walton *et al.* used a well-established and theoretically grounded approach in which the stories of previous students highlighted the presence of belonging challenges among those holding different racial and gender identities. Some previous evidence suggests that stories linking specific challenges and adjustment strategies to students' identities or backgrounds (e.g., first-generation status) can provide an especially effective approach for promoting college success among students who hold minoritized identities (*12*). However, this strategy also has the potential to backfire if it gives the impression that students from a particular identity group are unlikely to ultimately belong or succeed in college. Fortunately, colleges and universities can implement this established intervention as part of their efforts to bolster student retention and graduation, because the online materials in the Walton *et al.* study are freely available (*13*). ■

REFERENCES AND NOTES

1. J. Causey *et al.*, "Completing college: National and state report with longitudinal data dashboard on six- and eight-year completion rates" (Signature Report 21, National Student Clearinghouse Research Center, 2022).
2. Organisation for Economic Co-operation and Development (OECD), "Education at a glance 2022: OECD indicators" (OECD Publishing, 2022); https://doi.org/10.1787/3197152b-en.
3. G. M. Walton *et al.*, *Science* **380**, 499 (2023).
4. C. Adelman, "The toolbox revisited: Paths to degree completion from high school through college" (US Department of Education, 2006); https://www2.ed.gov/rschstat/research/pubs/toolboxrevisit/index.html.
5. R. A. Lazowski, C. S. Hulleman, *Rev. Educ. Res.* **86**, 602 (2016).
6. S. Solanki, D. Fitzpatrick, M. R. Jones, H. Lee, *Educ. Res. Rev.* **31**, 100359 (2020).
7. J. L. Burnette *et al.*, *Psychol. Bull.* 10.1037/bul0000368 (2022).
8. V. F. Sisk *et al.*, *Psychol. Sci.* **29**, 549 (2018).
9. D. S. Yeager, G. M. Walton, *Rev. Educ. Res.* **81**, 267 (2011).
10. M. Gopalan, S. T. Brady, *Educ. Res.* **49**, 134 (2020).
11. N. A. Bowman, N. Denson, *J. Higher Educ.* **93**, 399 (2022).
12. N. M. Stephens, M. G. Hamedani, M. Destin, *Psychol. Sci.* **25**, 943 (2014).
13. PERTS (Project for Education Research That Scales), Social-belonging for college students; https://www.perts.net/orientation/cb.

10.1126/science.adh7681

COOLING TECHOLOGY

Sustainable cooling with water generation

Dual-use devices offer a different path for more-sustainable living

By **Primož Poredoš** *and* **Ruzhu Wang**

A fivefold increase in cooling-related greenhouse gas emissions is expected by 2050 (*1*) from a combination of improved standards of living, especially in countries with humid climates, and the ongoing overall warming of the planet. Alternative approaches to efficient cooling are required to meet these needs. Although a large effort is underway to improve the energy efficiency of cooling more generally, a potentially important strategy is found by looking at the small water droplets that emerge on cold surfaces. Most people have experience with this process, from the condensation formed from air conditioners to cold drinks taken from a refrigerator. Although conventional chillers are becoming more efficient, integrating a passive radiative cooling strategy in which the sky is used as a cold sink could not only improve performance but also produce fresh water. This may help with cooling and potentially also with water scarcity issues in many regions of the world.

The water vapor dewing process has a large barrier owing to its inefficiency. An established figure of merit known as a coefficient of performance (COP) represents the ratio of useful cooling energy generated to the work provided [$COP = Q_{cool}(T_E)/W(T_E,T_C)$]. In general, the COP of chillers depends on the temperature difference between the hot (T_C) and cold (T_E) sides. Air conditioning requires air dehumidification, but water vapor dewing introduces a larger temperature difference and thus decreases the COP. This likely requires a reconceptualization of humidity management.

Synergistic and multifunctional effects can be realized by focusing on the air-water-energy nexus, primarily driven by water separation from the air. Pulling the water vapor from the air achieves air dehumidification while also creating a source of fresh water. Several studies already have demonstrated more than a twofold energy efficiency increase by decoupling latent from sensible cooling, predominantly in humid climate conditions (*2*). Splitting the water vapor–capturing process from sensible cooling sparked a pertinent question concerning the viability of dual-use devices that promote efficient cooling and freshwater generation (*3*). Although the dual nature of devices is relatively straightforward in terms of economic and performance efficiency for hot and humid regions, severe challenges exist for deploying these systems in hot and dry places.

Arid regions typically have abundant natural sunlight under clear sky conditions, offering a sustainable way to either increase or decrease the internal energy of a material relative to the ambient. The elevated temperature of a wet material accelerates a drying process, with water vapor released to the ambient. By contrast, the reversibility of that process is facilitated by hygroscopic porous materials that capture water from the atmosphere owing to the water concentration difference, according to Fick's law. In 2017, a proof of concept showed water harvesting from the air in outdoor arid conditions by using a porous metal-organic framework (*4*). In just a few years, atmospheric water harvesting has come a long way, from diurnal devices to ones using multiple cycles per day (*5*), even in the desert environment (*6*), swiftly converging to continuously operated devices not only by sorbent–desiccant but also using daytime radiative sky cooling materials (*7*). Attaining the dual-use feature of existing water harvesters with cooling capabilities could potentially be accomplished by reengineering existing devices rather than reinventing materials.

Sustainable cooling in arid regions requires a two-step approach (see the figure) owing to intrinsic material and heat-transfer characteristics. Sorbent at subambient temperature lowers outdoor air temperature while simultaneously promoting water capture, culminating in cooled and dehumidified air. As the adsorption process generates heat, both sensible and adsorption heat (Q_{ADS}) can be taken away by using radiative cooling ($Q_{C,RC}$) (*8*) or a desiccant-based heat pump ($Q_{C,HP}$) (*2*). The addition of the latter device with its capability of adsorbing moisture from the handled

Institute of Refrigeration and Cryogenics, MOE Engineering Research Center of Solar Power and Refrigeration, Shanghai Jiao Tong University, Shanghai 200240, China. Email: rzwang@sjtu.edu.cn

racial disparities in college graduation occur at predominantly white institutions, Black students graduate at the same rate as white students from 4-year institutions where at least half of undergraduates identify as Black, and there is also no graduation gap between Latinx and white students at majority-Latinx 4-year institutions (*11*). As a result, the social-belonging intervention may not be effective for these well-represented local-identity groups by race and college.

Additional research may also help explore features of interventions that are most effective. Walton *et al.* used a well-established and theoretically grounded approach in which the stories of previous students highlighted the presence of belonging challenges among those holding different racial and gender identities. Some previous evidence suggests that stories linking specific challenges and adjustment strategies to students' identities or backgrounds (e.g., first-generation status) can provide an especially effective approach for promoting college success among students who hold minoritized identities (*12*). However, this strategy also has the potential to backfire if it gives the impression that students from a particular identity group are unlikely to ultimately belong or succeed in college. Fortunately, colleges and universities can implement this established intervention as part of their efforts to bolster student retention and graduation, because the online materials in the Walton *et al.* study are freely available (*13*). ■

REFERENCES AND NOTES

1. J. Causey *et al.*, "Completing college: National and state report with longitudinal data dashboard on six- and eight-year completion rates" (Signature Report 21, National Student Clearinghouse Research Center, 2022).
2. Organisation for Economic Co-operation and Development (OECD), "Education at a glance 2022: OECD indicators" (OECD Publishing, 2022); https://doi.org/10.1787/3197152b-en.
3. G. M. Walton *et al.*, *Science* **380**, 499 (2023).
4. C. Adelman, "The toolbox revisited: Paths to degree completion from high school through college" (US Department of Education, 2006); https://www2.ed.gov/rschstat/research/pubs/toolboxrevisit/index.html.
5. R. A. Lazowski, C. S. Hulleman, *Rev. Educ. Res.* **86**, 602 (2016).
6. S. Solanki, D. Fitzpatrick, M. R. Jones, H. Lee, *Educ. Res. Rev.* **31**, 100359 (2020).
7. J. L. Burnette *et al.*, *Psychol. Bull.* 10.1037/bul0000368 (2022).
8. V. F. Sisk *et al.*, *Psychol. Sci.* **29**, 549 (2018).
9. D. S. Yeager, G. M. Walton, *Rev. Educ. Res.* **81**, 267 (2011).
10. M. Gopalan, S. T. Brady, *Educ. Res.* **49**, 134 (2020).
11. N. A. Bowman, N. Denson, *J. Higher Educ.* **93**, 399 (2022).
12. N. M. Stephens, M. G. Hamedani, M. Destin, *Psychol. Sci.* **25**, 943 (2014).
13. PERTS (Project for Education Research That Scales), Social-belonging for college students; https://www.perts.net/orientation/cb.

10.1126/science.adh7681

COOLING TECHOLOGY

Sustainable cooling with water generation

Dual-use devices offer a different path for more-sustainable living

By **Primož Poredoš** *and* **Ruzhu Wang**

A fivefold increase in cooling-related greenhouse gas emissions is expected by 2050 (*1*) from a combination of improved standards of living, especially in countries with humid climates, and the ongoing overall warming of the planet. Alternative approaches to efficient cooling are required to meet these needs. Although a large effort is underway to improve the energy efficiency of cooling more generally, a potentially important strategy is found by looking at the small water droplets that emerge on cold surfaces. Most people have experience with this process, from the condensation formed from air conditioners to cold drinks taken from a refrigerator. Although conventional chillers are becoming more efficient, integrating a passive radiative cooling strategy in which the sky is used as a cold sink could not only improve performance but also produce fresh water. This may help with cooling and potentially also with water scarcity issues in many regions of the world.

The water vapor dewing process has a large barrier owing to its inefficiency. An established figure of merit known as a coefficient of performance (COP) represents the ratio of useful cooling energy generated to the work provided [$COP = Q_{cool}(T_E)/W(T_E,T_C)$]. In general, the COP of chillers depends on the temperature difference between the hot (T_C) and cold (T_E) sides. Air conditioning requires air dehumidification, but water vapor dewing introduces a larger temperature difference and thus decreases the COP. This likely requires a reconceptualization of humidity management.

Synergistic and multifunctional effects can be realized by focusing on the air-water-energy nexus, primarily driven by water separation from the air. Pulling the water vapor from the air achieves air dehumidification while also creating a source of fresh water. Several studies already have demonstrated more than a twofold energy efficiency increase by decoupling latent from sensible cooling, predominantly in humid climate conditions (*2*). Splitting the water vapor–capturing process from sensible cooling sparked a pertinent question concerning the viability of dual-use devices that promote efficient cooling and freshwater generation (*3*). Although the dual nature of devices is relatively straightforward in terms of economic and performance efficiency for hot and humid regions, severe challenges exist for deploying these systems in hot and dry places.

Arid regions typically have abundant natural sunlight under clear sky conditions, offering a sustainable way to either increase or decrease the internal energy of a material relative to the ambient. The elevated temperature of a wet material accelerates a drying process, with water vapor released to the ambient. By contrast, the reversibility of that process is facilitated by hygroscopic porous materials that capture water from the atmosphere owing to the water concentration difference, according to Fick's law. In 2017, a proof of concept showed water harvesting from the air in outdoor arid conditions by using a porous metal-organic framework (*4*). In just a few years, atmospheric water harvesting has come a long way, from diurnal devices to ones using multiple cycles per day (*5*), even in the desert environment (*6*), swiftly converging to continuously operated devices not only by sorbent–desiccant but also using daytime radiative sky cooling materials (*7*). Attaining the dual-use feature of existing water harvesters with cooling capabilities could potentially be accomplished by reengineering existing devices rather than reinventing materials.

Sustainable cooling in arid regions requires a two-step approach (see the figure) owing to intrinsic material and heat-transfer characteristics. Sorbent at subambient temperature lowers outdoor air temperature while simultaneously promoting water capture, culminating in cooled and dehumidified air. As the adsorption process generates heat, both sensible and adsorption heat (Q_{ADS}) can be taken away by using radiative cooling ($Q_{C,RC}$) (*8*) or a desiccant-based heat pump ($Q_{C,HP}$) (*2*). The addition of the latter device with its capability of adsorbing moisture from the handled

Institute of Refrigeration and Cryogenics, MOE Engineering Research Center of Solar Power and Refrigeration, Shanghai Jiao Tong University, Shanghai 200240, China. Email: rzwang@sjtu.edu.cn

air in the evaporator introduces cooling-adsorption dehumidification that results in comfortable air conditioning. By leveraging the intrinsic heat pump capability with simultaneous cooling and heating energy production, the condensing heat ($Q_{H,HP}$) as a by-product is not wasted, but rather is used for the heating-desorption process (Q_{DES}) at a reduced temperature. Hence, elevated evaporation ($T_E \rightarrow T_{E,HP}$), as well as reduced condensation ($T_C \rightarrow T_{C,HP}$) temperatures, creates a positive impact on the COP. Thus, the latent heat load handled by low-temperature evaporation cooling (condensing dehumidification) in conventional air-conditioning devices can be managed by a system based on sorption technology in which sorbent materials could be regenerated by solar heating or condensing heat from a heat pump, having a positive impact on the device performance.

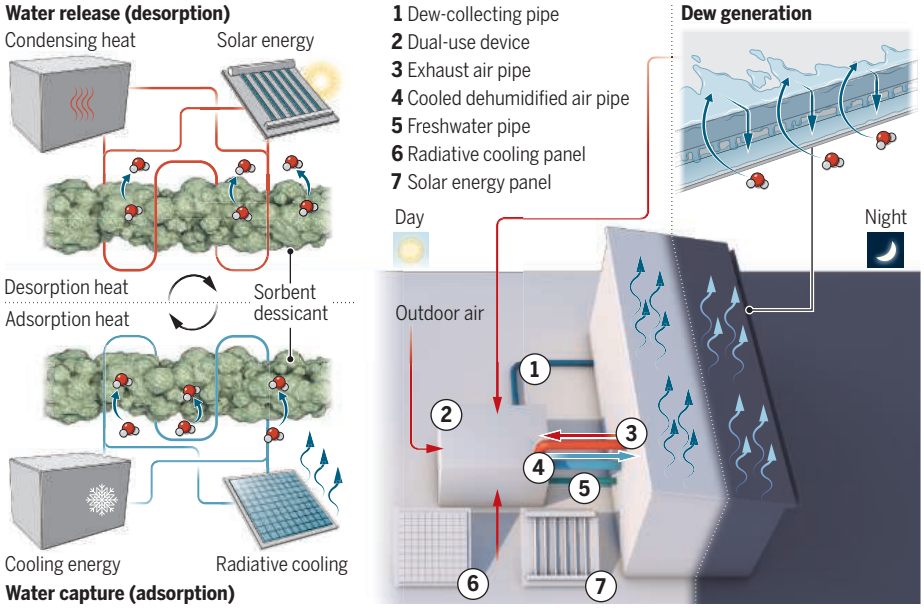
Alternatively, solar energy ($Q_{H,SOL}$), coupled with technological advances in thermal energy storage, emerges as another potential source, capable of ejecting the water stored in sorbent material in a continuous manner. The final step in collecting a liquid form of water is established through a thermal pathway ($Q_{V,C}$) between the high-temperature water vapor and a heat sink, served by the ambient, residual cooling capacity of a heat pump ($Q_{C,HP}$), or radiative cooler ($Q_{C,RC}$). Solar-powered harvesters could play a large role in water generation (*9*). Radiative coatings deployed on rooftops in general may be helpful in dry regions and would take some of the burden off either dual-use devices or conventional air conditioning for buildings (*10*). The reduced cooling requirements during the night offer an effective solution for providing liquid water by using photon flow to outer space (*11*). In addition, mass recovery of exhaust indoor air with latent heat load from humans or other wet sources emerges as a different source of water in extremely dry ambient conditions. Such air-humidity recovery has already been used in spacecraft or space stations for months-long or year-round water supply to astronauts.

Variable ambient and indoor conditions throughout the year require a flexible strategy concerning device adaptability. For instance, temperature-enabled precise regulation of the metal-organic framework's stepwise position showed promise with the prospect of operation under harsh ambient conditions with relative humidity down to 10% (*12*). This approach seems suitable for desiccant-based heat pumps owing to the straightforward regulation of evaporation temperature. At micro- and nanometer scales, real-time manipulation of the sorbent pore volume is poised to bring tunable water-uptake behavior as well as desorption temperature (*13*). Radiative coatings, by contrast, can potentially utilize switchable temperature-modulated thermal emittance (*14*), extending the usage scenario of the dual-use devices in providing all-year thermal comfort. One issue that makes it more difficult to chart a path forward is the need for a consensus in evaluating devices with relevant metrics to associate useful products such as generated cooling energy and water with the provided work. Other factors, such as total and sorbent weight, besides radiative cooling, solar, and total areal footprint, might add another layer in ensuring a fair basis for comparison of competing devices. In addition, dual-use devices ought to be globally evaluated for arid and semi-arid regions to clearly show potential for widespread deployment. Such an engineering opportunity requires further quantitative analysis and deeper physical insights, with clear descriptions about the theoretical and practical improvements over conventional single-use devices.

The widespread implementation of the dual-use approach may encounter various challenges. One of these is a decreased potential of sorbents to capture the water caused by lower relative humidity. An effective solution to enhance the adsorption rates and kinetics of sorbents is to use composite materials, such as MOF+LiCl (*15*), and regulate the evaporation temperature (*12*). By combining these material- and system-level strategies, the potential of the sorbents to adsorb water can be substantially improved. By contrast, lower moisture content in the atmosphere can improve the performance of radiative sky cooling materials (*10*). Because sustainable cooling with atmospheric water harvesting in a single dual-use device is interfaced with diverse technologies, a special effort needs to be addressed to finding the best solution for commercial attractiveness, along with a low-carbon path for sustainable living. ■

Dual-use device for cooling and water generation in arid places

Hygroscopic sorbent material captures water vapor from the outdoor air, while also lowering air temperature with a heat pump or radiative cooler. Upon saturation, sorbent materials are regenerated by using solar thermal energy or a heat pump. Released high-temperature water vapor is dewed. Humid indoor exhaust air and vapor in ambient air are also water sources, where the latter (vapor in ambient air) is dewed during the nighttime by using radiative roof coating. The continuous operation of the dual-use device relies on solar thermal storage. Radiative sky cooling materials provide a cooling effect continuously owing to the atmospheric window (8 to 13 μm) during the day.



GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. J. Woods *et al.*, *Joule* **6**, 726 (2022).
2. Y. D. Tu, R. Z. Wang, T. S. Ge, X. Zheng, *Sci. Rep.* **7**, 40437 (2017).
3. M. Ejeian, R. Z. Wang, *Joule* **5**, 1678 (2021).
4. H. Kim *et al.*, *Science* **356**, 430 (2017).
5. J. Xu *et al.*, *Energy Environ. Sci.* **14**, 5979 (2021).
6. N. Hanikel *et al.*, *ACS Cent. Sci.* **5**, 1699 (2019).
7. I. Haechler *et al.*, *Sci. Adv.* **7**, eabf3978 (2021).
8. D. Zhao *et al.*, *Joule* **3**, 111 (2019).
9. J. Lord *et al.*, *Nature* **598**, 611 (2021).
10. X. Yin, R. Yang, G. Tan, S. Fan, *Science* **370**, 786 (2020).
11. Y. Zhai *et al.*, *Science* **355**, 1062 (2017).
12. Y. Feng, T. Ge, B. Chen, G. Zhan, R. Wang, *Cell Rep. Phys. Sci.* **2**, 100561 (2021).
13. N. Hanikel *et al.*, *Science* **374**, 454 (2021).
14. K. Tang *et al.*, *Science* **374**, 1504 (2021).
15. J. Xu *et al.*, *Angew. Chem. Int. Ed.* **59**, 5202 (2020).

10.1126/science.add1795

POLICY FORUM

GLOBAL HEALTH

Cascading failures in COVID-19 vaccine equity

Ethical analysis should encompass upstream decisions and their downstream consequences

By **James V. Lavery[1,2], Rachael M. Porter[1], David G. Addiss[3]**

Equity is a central tenet of global health, a recognition of the profound health disparities between the world's rich and poor. But the meaning assigned to equity is often figurative or metaphorical: a call to action, a framing for advocacy. The proliferation of equity rhetoric does not appear to be matched by corresponding rates of progress in reducing global disparities. If unaddressed, inequities can undermine public trust in and support for science. Failure to achieve anything close to equitable global distribution of COVID-19 vaccines provides a case in point. Simply measuring the ultimate outcomes of vaccine distribution is not sufficient. Upstream decisions that create downstream distributive outcomes, and the complex interests and policies that lie behind them, deserve greater ethical scrutiny. Empirically driven analysis of the entire "equity-deficit cascade" is essential for equity to gain the attention and currency it deserves—beyond its metaphorical value.

Collectively, we have yet to engage meaningfully with the full complexity and implications of the concept of equity (*1*). In essence, a claim of equity is an assessment that two or more individuals or groups have been treated fairly in some transaction or intervention and that their treatment should be considered fair, even if it is not equal, because morally relevant differences have been considered in the justification of any treatment differences. Conversely, claims of inequity require the observation of unequal treatment—for example, unequal access to COVID-19 vaccines between two different populations—and inadequate moral justification for the unequal treatment. Moral justification for different treatment is the cornerstone of equity. Without it, appeals to equity are uninterpretable.

COVID-19 vaccine discovery, development, and delivery involved many steps and decisions (*1*) within a complex policy environment. Many of these decisions have differentially affected access to vaccines for different populations, but without adequate moral justification. For example, the decision to pursue an mRNA vaccine platform, despite the enormous success of the strategy and the extraordinary benefits it has delivered for many, complicated the planning and distribution of the vaccines to a substantial proportion of the global population because of limited capacity for manufacturing, high costs, and demanding cold-chain requirements, among other factors. The tendency to invoke equity casually obscures the central importance of assessing the moral relevance, implications, and justification of each of these decisions. And despite their potentially profound implications, these strategic decisions are not adequately accounted for in prominent ethical frameworks, which tend to focus on the articulation of guiding values, principles, and models for how fairness should be achieved in the allocation of COVID-19 vaccines rather than on factors that limit their acquisition or justifications for different treatment (*2*).

UPSTREAM ANALYSIS

More than a decade ago, De Vries likened the field of bioethics to an "emergency department" for ethical crises: "Bioethics as currently constituted waits until a problem arises and then brings its powers to bear on trying to solve that problem…puzzling out the problem, defining it clearly, and deciding which aspects should be addressed. What is needed is a re-focusing on the upstream causes of the ethical problems" [(*3*), p. 425]. This view has been reflected more recently with respect to the equitable translation of climate science: "environmental decision-making…demands that the views of affected ('vulnerable') parties regarding the principles and criteria for fair processes and outcomes be explored upstream in the innovation process, and not to be a matter for a narrow set of privileged experts to delineate" [(*4*), p. 56].

The limitations of a "downstream" focus are well illustrated in a recent analysis of the impact of the COVID-19 Vaccines Global Access (COVAX) Facility (*5*). COVAX was created to facilitate COVID-19 vaccine distribution globally, with distribution to follow the allocation framework of the World Health Organization (WHO) for fair and equitable access to COVID-19 health products (*5*). COVAX largely fulfilled its commitment to equitable distribution by income status of participating countries and by the proportion of older people in their populations. However, by January 2022—the end date for the analysis—only 9.7% (about 63 million) of people in low-income countries had received at least one dose of COVID-19 vaccine (*5*). Many of the mechanisms driving the inequitable distribution of COVID-19 vaccines operated outside and upstream from the distributive logic of the COVAX system.

Where they have been applied, upstream analyses have focused on a narrow set of issues. Critics have highlighted the role of the COVID-19 vaccine manufacturers' reluctance to license their intellectual property (IP) to manufacturers in low- and middle-income countries and to share their technologies, which are already well understood to limit the equitable distribution of vaccines (*6*). In addition, changes in global trade agreements, IP practice conventions, and technology transfer have been slow to materialize. However, these mechanisms alone do not provide a full account of the causes of inequitable distribution.

A broader analysis of the entire equity-deficit cascade would move beyond the identification of relevant ethical principles. It would function as a pragmatic ethical due-diligence test. This analysis would begin with four main steps: (i) describe the major steps and decisions associated with the vaccines' discovery, development, and delivery and the policies that shape these decisions; (ii) describe the relative contributions of key decisions at each of these points with respect to their potential to undermine equitable vaccine distribution, through reviews of existing literature and in-depth interviews with key actors in the vaccine pipelines; (iii) identify what morally relevant differences among potential vaccine recipient populations might, or might not, justify different levels of vaccine access at each of these decision points; and (iv) explore how adjustments in each or

[1]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [2]Center for Ethics, Emory University, Atlanta, GA, USA. [3]Focus Area for Compassion and Ethics, Task Force for Global Health, Atlanta, GA, USA. Email: jlavery@emory.edu

any of these decisions might alter the final outcomes in ways that would make the ultimate distribution, including timeliness of access, more equitable.

Applying this type of systematic diagnostic analysis to COVID-19 vaccines illustrates how traditional ethical analyses must be expanded upstream and focused more explicitly on decisions and mechanisms that function as critical determinants of inequity (*7*). It also elucidates the connections between upstream decisions and the constraints and scarcities they create for decision-makers downstream. Making the genesis of these downstream circumstances more apparent could also help to frame appropriate scrutiny on the decisions made by governments and institutions at these downstream points along the cascade. What equity deficits are inherited at each step or decision point in the vaccine production pathway? What is the nature of these deficits, that is, how do specific product profiles or pricing decisions constrain downstream opportunities for vaccine acquisition, use, or distribution and for whom? And what are the ultimate implications of these deficits for health systems and for individuals and their communities around the world?

EQUITY-DEFICIT CASCADE

Determining who should answer these questions, and how, will require careful attention to minimize the potential of distorting influences of the companies and organizations that control the very processes in question. The processes of investigation and analysis must be designed and conducted to ensure that they offer actionable insights for actors across the vaccine discovery, development, and delivery pathway to improve practices and governing policies in ways that will close the gap between professed commitments to equity and the disappointing reality. Possible models include various types of innovation systems evaluation, value-chain analyses, and various approaches to complex systems analysis to inform policy innovation. Independence and legitimacy for the necessary analyses might require that the work be undertaken by bodies with sufficient global status and transdisciplinary influence so that the findings will have some sway at the level of global policy and practice, such as National Academies' studies or a Lancet Commission.

We refer to the key decisions involved in the discovery, development, and delivery of COVID-19 vaccines, and key features of the policy environment that shapes these decisions, as the "equity-deficit cascade," because each upstream decision has implications for equitable distribution downstream (see the table). To avoid such an equity failure in fu-

The "equity-deficit cascade"

Key decision points and equity challenges for the distribution of COVID-19 vaccines

DECISION POINT	EQUITY CHALLENGES
PUBLIC POLICY	
Intellectual property (IP) laws and regulations	• IP laws and regulations for licensing and transfer of technologies may not adequately reflect the equity interests of public investment in vaccine science and manufacturing • Business interests of manufacturers may be in conflict with equitable vaccine distribution
Trade agreements and barriers	• Import tariffs may create barriers to equitable vaccine distribution • Established trade agreements and conventions tend to consolidate vaccine distribution among high-income countries (HICs) • Incentive structures discourage companies from using the World Trade Organization's Trade-Related Aspects of Intellectual Property Rights (TRIPS) flexibilities to increase equitable vaccine distribution
National policies for vaccine procurement	• Advance purchase agreements between companies and HICs limit supply for global distribution
VACCINE DISCOVERY	
Investment in vaccine science and manufacturing	• Investments in vaccine research occur disproportionately in the Global North • Vaccine business models may not adequately reflect equity interests of public and nonprofit investors in vaccine science
Vaccine performance, e.g., efficacy, effectiveness, and adverse reactions	• Methods for determining safety and efficacy of vaccines underrepresent important and potentially high-risk groups, such as people with complex comorbidities, pregnant women, and children • Clinical trial results drive implementation decisions but do not adequately predict public health value
VACCINE DEVELOPMENT	
Vaccine pipeline management	• Vaccine product profiles—e.g., cold-chain requirements, mode of administration, dose regimens, cost, and pricing estimates—may not account for equity implications • Vaccine business models may not adequately reflect equity interests of public and nonprofit investors in vaccine science
Production and technical requirements to manufacture at scale	• Vaccine manufacturing capacity is inadequate to meet global demand • Existing manufacturing capacity is disproportionately located in the Global North • Vaccine business models may not adequately reflect equity interests of public and nonprofit investors in vaccine manufacturing
VACCINE DELIVERY	
Procurement and accessibility	• Priority vaccine procurement for HICs limits available supply for low- and middle-income countries (LMICs) • Priority by "preferred customer" and purchasing power inhibits distribution according to public health goals
Global vaccine distribution governance mechanisms	• Global distribution mechanisms must compete with HIC priority for vaccine purchasing, which creates a "two-tiered" system of vaccine distribution
Supply chain	• Capacity limits constrain the kinds and volumes of vaccines that can be distributed to LMICs
Global COVID-19 epidemiology and disparities	• Persons most at risk of infection, disease, and death may not be prioritized for vaccination
Global public health and public health goals	• Different public health goals may require different vaccine-distribution strategies • Limited vaccine access among some groups may result in loss of vaccine efficacy more generally because of the proliferation of COVID-19 variants globally
Allocation considerations	• Vaccine distribution may not be informed by appropriate ethical principles or values • Procedures used in vaccine allocation may not be fair • There is limited transparency about the rationales and justifications for distribution decisions
Last mile	• Within-country distribution to vaccine delivery sites may favor more-densely-populated areas, which could lead to systematically missing more-vulnerable or marginalized groups • Decisions about where and how much vaccine to distribute may be driven by political considerations rather than public health priority and epidemiological data
Last inch	• Lack of trust in health systems and interventions may be greatest among marginalized communities, increasing vaccine hesitancy and reducing vaccine uptake and coverage • Expectations in LMICs of late and limited distribution of vaccines may create attitudes of resignation and suppress demand, especially among more-marginalized communities

ture pandemics, we need an accounting of these deficits, in particular, where ethical justifications do not support claims of equity, and an explanation of the conditions that produced them. This cascade draws on previous frameworks that have highlighted key policies and practices in the vaccine discovery, development, and delivery pathway (*8*, *9*). We expect the framework to be expanded and refined as it is used to guide the collection of relevant insights. And although the equity-deficit cascade focuses here on vaccine production, we expect its logic to be highly transferrable to other complex contexts in translational science.

PUBLIC INTERESTS AND ACCOUNTABILITY

One of the unresolved questions that might be informed by empirical findings from an equity-deficit cascade analysis is how obligations to ensure the equitable distribution of technologies, such as the COVID-19 vaccines, should be distributed across the vaccine enterprise, from publicly-funded basic science to private-equity financing. In the United States, science policy is strongly oriented toward the commercialization of scientific discoveries into products and services. But it does not specify how publicly funded, university-based discoveries should be used or applied. Many such discoveries are acquired by private, for-profit corporations. And whether, or how, those companies realize value with the acquisitions is determined either at the sole discretion of the company or by the terms of a transfer agreement with the university, with little or no input from the public and little evidence of a sense of obligation to share "the bounties of federally funded biomedical research" [(*10*), p. 795].

The value of the private sector has been abundantly clear in the rapid production of COVID-19 vaccines. But the inequitable distribution of COVID-19 vaccines globally illuminates a noteworthy shortcoming of present science policy and market models. A vaccine for a pandemic disease has a different value proposition than, for example, a new biologic for psoriasis. Inequitable distribution of vaccines does not simply create disparities in protection among individuals. It also facilitates the perpetuation of disease transmission that creates profound collective effects on health systems, economies, politics, and social relationships. Aside from the incalculable value of the decades of public investment in science and technology (*11*), the considerable public investments in government subsidies and tax credits have helped manufacturers ramp up COVID-19 vaccine production; secure purchasing agreements and regulatory approvals; gain extraordinary revenue and profit and market shares; and mitigate risk. But despite the prominence of equity rhetoric, these public investments have not come with the requirement of equitable distribution of the resulting vaccines globally, which could be viewed as a requirement not simply of fairness, but for realizing the full potential value of the vaccines (*9*).

A more constructive dialogue is needed about the role and obligations of these private corporations as central players in global public health and as partners and beneficiaries of public investments in the scientific enterprise. Companies have intentionally and strategically embraced equity in their marketing campaigns for its public relations cachet, most recently at the World Economic Forum in Davos, Switzerland, where 39 organizations signed the Zero Health Gaps Pledge (*12*). Drug and vaccine manufacturers deserve to be celebrated and rewarded for generating safe and effective vaccines against COVID-19 in record time. But their proclamations about equity, which are widely echoed across the spectrum of global-health actors, simply do not match the reality of the inequitable global distribution of their COVID-19 vaccines.

At a minimum, it seems disingenuous and unfair for companies to celebrate their commitments to equity in their public communications when these professions do not appear to be backed up by the kinds of investments in new policies and practices that would be more likely to improve the equitable distribution of their products. There are simply too many business practices that appear to be firmly aligned against equitable distribution, and these need to be more thoroughly understood and challenged (*6*). Our equity-deficit–cascade analysis might actually offer value for private companies by improving their awareness and understanding of where and how their development and manufacturing processes are creating irretrievable equity deficits—insights that conventional ethical analyses are not delivering. And an equity-deficit–cascade analysis could actually help companies demonstrate their commitments to equity in their public communications, if they are able to provide the necessary ethical justifications.

Shareholders and boards of directors of drug and vaccine manufacturers could play an important role in encouraging and shaping their companies' actions with respect to vaccine equity (*13*). The same logic applies to the collective force that citizens can exert on their governments, which negotiate the terms of corporate subsidies on behalf of constituents and also hold considerable sway over many of the levels of the vaccine equity-deficit cascade. Many scientists and engineers are already extending these concerns, arguing that science policy has failed to keep pace with social changes and new attitudes toward social justice (*14*).

EQUITY MUST HAVE MEANING

The stark gap between the pervasive rhetoric about equity and the dismal reality of the global vaccine distribution for severe acute respiratory coronavirus 2 (SARS-CoV-2) demands a collective reckoning. A robust explanatory analysis of the causes of vaccine inequity in the COVID-19 vaccine distribution system, and shortcomings in the ethical justifications necessary to support the pervasive claims of equity, represents a constructive first step forward for all stakeholders. It might help private vaccine manufacturers to identify opportunities for process, practice, and strategy improvements and innovations. It might also serve as a useful guide for advocates for global COVID-19 vaccine equity and help inform present deliberations for a more genuinely equitable and effective response to future pandemics.

Despite earnest efforts by scientists and others to advance distributive equity, invoking the concept of equity has little real meaning in the absence of a comprehensive understanding of the forces, structures, and decisions that give rise to inequity and the ethical reasoning that gives the term its normative weight. Expansive rhetoric and empty promises have surprising staying power. If we wish equity to have anything more than allegorical value, we must take the concept more seriously, beginning with a disciplined and deliberate examination of the equity-deficit cascade. ■

REFERENCES AND NOTES

1. M. J. Smith, *Public Health Ethics* **8**, 173 (2015).
2. E. J. Emanuel *et al.*, *Science* **369**, 1309 (2020).
3. R. De Vries, *Ethical Theory Moral Pract.* **14**, 419 (2011).
4. J. A. Flegal, A. Gupta, *Int. Environ. Agreement Polit. Law Econ.* **18**, 45 (2018).
5. K. J. Yoo *et al.*, *Bull. World Health Organ.* **100**, 315 (2022).
6. A. Maxmen, *Nature* **607**, 226 (2022).
7. S. R. Benatar, *S. Afr. J. Sci.* **118**, 14995 (2022).
8. P. L. Stern, *Ann. Allergy Asthma Immunol.* **125**, 17 (2020).
9. O. J. Wouters *et al.*, *Lancet* **397**, 1023 (2021).
10. H. Markel, *N. Engl. J. Med.* **369**, 794 (2013).
11. J. M Spector *et al.*, *Sci. Transl. Med.* **438**, eaaq1787 (2018).
12. World Economic Forum, The Zero Health Gaps Pledge (2023); https://initiatives.weforum.org/global-health-equity-network/pledge.
13. P. Singer, "Boards and investors can help solve global vaccine inequity," LinkedIn (2021); https://www.linkedin.com/pulse/boards-investors-can-help-solve-global-vaccine-inequity-peter-singer/.
14. K. Chintam *et al.*, "Science policy can't be simply about science," *Scientific American,* 22 April 2021; https://www.scientificamerican.com/article/science-policy-cant-be-simply-about-science.

ACKNOWLEDGMENTS

J.V.L. is a member of the bioethics advisory council for Pfizer, Inc.

10.1126/science.add5912

BOOKS *et al.*

Lithium, an essential but limited natural resource, lies below salt piles in Bolivia's Uyuni Salt Flat.

PHOTO: BLOOMBERG/GETTY IMAGES

ECONOMICS

Making the most of scarcity

A pair of historians dig into the deep ideological roots of our planetary predicament

By **Erle C. Ellis**

Written by historians Fredrik Albritton Jonsson and Carl Wennerlind, *Scarcity* connects, dissects, and narrates the history of Western economic ideas about the natural limits to human societies, from the deep roots beneath Thomas Malthus's 1798 book *An Essay on the Principle of Population* to Johan Rockström and Will Steffen's contemporary Planetary Boundaries framework. The book's goal is clear from the beginning—to use a deep historical understanding of these ideas to challenge the central thinking behind neoclassical economics, depicted as being "at the heart of the planetary crisis we now face."

The book's eight core chapters explore five centuries of Western intellectual movements and their conceptions of human–nature relations, including the Enlightenment, Romanticism, socialism, and neoclassical economics. The authors introduce a framework of "Finitarianism" versus "Cornucopianism" as a means to clarify ideological debates about natural limitations and unlimited possibilities, with "each side defin[ing] itself by rejecting the other."

Anyone interested in the history of ideas will enjoy this book, with its detailed and fascinating narratives connecting important thinkers on human–nature relations from Aristotle to Rousseau, Adam Smith, Malthus, Ricardo, Mill, Marx, Keynes, Marshall Sahlins, Rachel Carson, and many more, together with a host of influencers and influences not usually connected with environmental thinking, from Émile Zola to Hannah Arendt.

Scientists will be drawn in by the authors' efforts to connect historical narratives with environmental and social conditions, from the "Little Ice Age" to the space age. But in the end, this is a book with a mission: to convince us that "business as usual will bring disaster" and that the answer is an ideology of "planetary scarcity" focused on "the limits of human ingenuity, the power of unintended consequences, and the fragility of all earthly things."

Environmentalists will find this territory familiar, and there is little but critique for other perspectives—and realities. The book completely ignores two of the most important thinkers on human–nature relations since Malthus—the agricultural economist Ester Boserup and Nobel Prize–winning economist of the commons Elinor Ostrom. How are we to understand the failed predictions of both Malthus and Ehrlich without Boserup's model of increasing food production driven by increasing demand? Can we understand sustainable governance of common-pool resources—like Earth's atmosphere—without Ostrom's polycentric governance? The book is nearly silent on the subjects of demographic transitions and rapid declines in population growth, ongoing intensification of agriculture and increased productivity, and the forest transitions and striking forest recoveries of developed nations.

Although the book spares no effort attacking neoclassical economics—the subject of much debate within economics itself—one feels distant from the action, with very little direct examination of these critical debates. Even the recent "degrowth" movement, which advocates shrinking economies to address environmental limits, is only mentioned in passing—a surprising omission for a book focused on planetary scarcity.

While the authors recognize their limited focus on "a narrow range of elite, white, male thinkers," they call for others to write a "more expansive history." It is difficult to understand the absence of at least some effort to go beyond the Western canon in order to present the full conceptual range of thinking on human–nature relations. The reader will find no inspiration herein from the many diverse and broader views of human–nature relations across cultures, whether Eastern, African, or Indigenous. Nor does the book stray far from Western history, conventional environmentalism, or capitalist critique. Within such limits, it is no surprise that the book's narrative begins with the Christian dialectic of Eden's bliss versus toil in the garden and ends with binary thinking of planetary boundaries—safety on one side and disaster on the other.

Scarcity
Fredrik Albritton Jonsson and Carl Wennerlind
Harvard University Press, 2023. 304 pp.

Planetary scarcity may indeed represent aspects of planetary reality, but as a new ideology it has little to offer those seeking to achieve the main goal of this book—"to reconsider fundamentally how we organize our economy." Beyond the doom and sacrifice of conventional environmental Finitarianism, there is a growing world of ideas and action for transformative change. These include the developmental and conservation possibilities of clean energy, declining populations, new models of social cooperation and exchange, and biodiversity conservation in working landscapes.

While this book is perhaps a new classic for historians of ideas—particularly those focused on the Western world—those seeking a deeper and broader understanding of how people and planet might interact more beneficially in the future must look elsewhere. ■

10.1126/science.adg0493

The reviewer is at the Department of Geography and Environmental Systems, University of Maryland, Baltimore County, Baltimore, MD 21250, USA. Email: ece@umbc.edu

SCIENCE WRITING

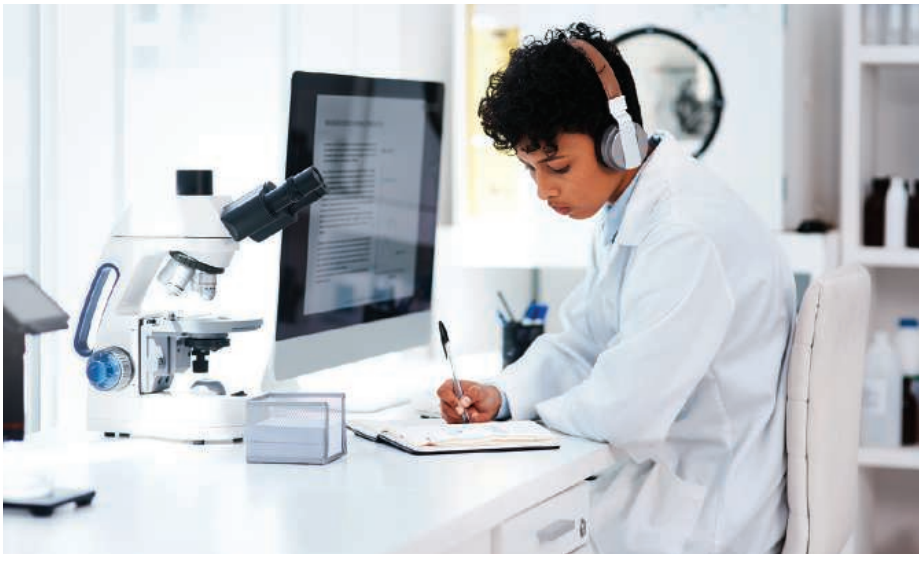
Writing for the reader

Clarity and brevity are key to readable research papers

By **Thom Scott-Phillips**

Scientific Papers Made Easy
Stuart West and Lindsay Turnbull
Oxford University Press, 2023. 208 pp.

Undergraduate science students must complete many types of writing assignments: methods for experiments, reports on field trips, essays surveying a field. The person who reads this work does so because—let's be blunt—they are given money to do so. It is a requirement of the job. This means that students become skilled at writing for people who are paid to assess what they have written. But this creates a quirky inflection point in scientific careers, where something that was once rewarded suddenly becomes bad practice.



Young scientists must appreciate that their readers are not instructors evaluating an assignment.

For those who go on to complete a PhD and embark on a scientific career, a new audience emerges—other scientists. These new readers are busy and distracted, and if they cannot quickly see why they should care about the work an author is attempting to describe, they will not do so. Many scientists never fully adjust to this change. They maintain the habit of showcasing their knowledge and talents to a reader who, they imagine, will proceed diligently through their every word. Scientific papers produced this way are less read and have less impact (*1*).

The reviewer is at the Institute for Logic, Cognition, Language, and Information, 20018 Donostia, Gipuzkoa, Spain. Email: thomascharles.scottphillips@ehu.eus

In *Scientific Papers Made Easy*, Oxford biologists Stuart West and Lindsay Turnbull seek to impart the importance of simplicity and clarity in scientific writing on early-career life scientists. Past books have sought to help readers develop other writing skills, such as creativity and the development of an individual style, but I know of no other book dealing with the basics of scientific writing as cleanly and as directly as this one does.

West and Turnbull identify many common habits that hinder readability and add unnecessary length to scientific papers, and they provide specific guidance for improving readability while decreasing length. These include sticking to one major point per paragraph and avoiding excessive use of the passive voice. Experienced writers might disagree with some of their advice—for instance, that figure legends should be short—but very few writers should dispute the book's most basic principle: "The reader must come first."

The book's central chapters each address one part of a standard research project: methods, results, figures, introductions, discussions, abstracts, titles, and cover letters. With this ordering, the book begins with material tasks and moves toward the theoretical. It also provides new scientists writing their first papers with a clear path. It may be best to begin not with the beginning of the paper itself, the book suggests, but rather with the specifics of what one did and found out.

The book does not offer much guidance about other types of scientific writing. There is some advice about how to present mathematical models but almost nothing on review articles or big-picture perspectives, nor about the specific challenges of interdisciplinary audiences, for example. Some readers may consider this a limitation, but it is consistent with the book's theme of maintaining a clear, sharp focus on doing the basics very well. Indeed, the book sticks admirably to its own principles on two fronts: It is short and to the point.

An especially useful feature of this book is the presence of a number of detailed, annotated passages demonstrating both good and bad writing practice. These comparisons make the authors' generic advice concrete and offer readers specific examples on which to reflect. QR codes throughout the book link to additional material, such as quizzes and lesson plans, should the book be used in scientific writing courses. As the title suggests, most example passages come from the biological sciences, but the lessons generalize, and no knowledge of this subject matter is necessary for one to make the most of the book.

How we write scientific papers may be about to change. In the near future, large language models (LLMs) such as ChatGPT may be able to summarize research fields or describe standard methods from simple prompts. Yet solid mastery of writing basics will not become a redundant skill for the successful scientist, not least because some scientific journals do not currently permit the use of these technologies in article creation (*2*). But even where the technology is put to use, it will always be necessary to write about the specific features of one's own studies or, at the very least, to edit the output of a LLM to ensure its relevance to one's target audience. Great scientists still need to know how to write cleanly and plainly, and *Scientific Papers Made Easy* is an excellent guide for acquiring that skill. ■

REFERENCES AND NOTES

1. B. Freeling, Z. A. Doubleday, S. D. Connell, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 341 (2019).
2. H. H. Thorp, *Science* **379**, 313 (2023).

10.1126/science.adh1867

Europe's old-growth forests, such this one in Romania's Fagaras Mountains, lack sufficient protection.

Edited by **Jennifer Sills**

Protect old-growth forests in Europe now

Old-growth forests harbor high and unique biodiversity, store vast amounts of carbon, are important for water and nutrient cycling, and constitute a unique element of natural heritage (*1*). In the European Union, old-growth forest protection is now mandated by the EU Biodiversity Strategy for 2030. However, almost 3 years after the strategy's adoption, stakeholders and policymakers are still discussing definitions and legislative mechanisms, while old-growth forests continue to decline at alarming rates (*2*–*4*).

Many old-growth forests are logged before their identification and protection. In Sweden, for example, unprotected boreal old-growth forests have been cut at a rate that could lead to their disappearance within the next 50 years (*2*). Similarly, Romania harbors up to 738,000 ha of potential old-growth forest, but more than 90% of this area lacks strict protection (*5*). In Romania and elsewhere in Eastern Europe, logging continues across some of the continent's few remaining large landscapes dominated by temperate old-growth forests (*4*). Even protected old-growth forests are often salvage logged after natural disturbances (*6*).

In March, the European Commission suggested guidelines to map and protect old-growth forests by the end of 2029 (*7*). However, these guidelines are neither binding nor prescriptive. Given current widespread logging of old-growth stands, the EU is on track to fail its 2030 goals.

Pressure on Europe's biomass-rich old-growth forests is high and rising. Timber prices have increased (*8*). Compensation would encourage forest owners to adopt strict protection, but there are insufficient resources and tools to provide financial incentives (*9*). Because landowners anticipate that forest protection will increase in the future, and forest monitoring is sparse, they are motivated to log as much as possible before regulation tightens.

To improve protection, the EU should immediately implement a logging moratorium on areas potentially harboring old-growth forests, make resources available to detect old-growth forests, require member states to include old-growth protection in their national strategies, and provide equitable financial tools to ensure their effective protection (*10*). Exemptions from strict conservation could be considered only for stands managed by well-documented practices that support biodiversity. Without bold and swift action, Europe risks irreparable loss to its natural heritage.

Martin Mikoláš[1]*, Gianluca Piovesan[2], Anders Ahlström[3], Daniel C. Donato[4], Rhiannon Gloor[1], Jeňýk Hofmeister[1], William S. Keeton[5], Bart Muys[6], Francesco M. Sabatini[7,1], Miroslav Svoboda[1], Tobias Kuemmerle[8]

[1]Czech University of Life Sciences Prague, Faculty of Forestry and Wood Sciences, Prague, Czech Republic. [2]Department in Ecological and Biological Sciences, University of Tuscia, Viterbo, Italy. [3]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden. [4]University of Washington, School of Environmental and Forest Sciences, Seattle, WA, USA. [5]Rubenstein School of Environment and Natural Resources and Gund Institute of Environment, University of Vermont, Burlington, VT, USA. [6]Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium. [7]Alma Mater Studiorum–University of Bologna, Bologna, Italy. [8]Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany. *Corresponding author. Email: mikolasm@fld.czu.cz

REFERENCES AND NOTES

1. F. Hua *et al*., *Science* **376**, 839 (2022).
2. A. Ahlström, J. G. Canadell, D. B. Metcalfe, *Earth. Fut.* **10**, e2022EF003221 (2022).
3. F. M. Sabatini *et al*., *Sci. Data* **8**, 220 (2021).
4. "Illegal logging of Romania's natural forests increases despite court threat—new report," *Save Paradise Forests* (2022); https://www.saveparadiseforests.eu/en/illegal-logging-of-romanias-natural-forests-increases-despite-court-threat-new-report/.
5. C. Munteanu *et al*., *Conserv. Biol*. **36**, e13820 (2022).
6. M. Żmihorski, P. Chylarecki, A. Orczewska, T. Wesołowski, *Science* **361**, 238 (2018).
7. "Commission staff working document: Commission guidelines for defining, mapping, monitoring, and strictly protecting EU primary and old-growth forests" (2023); https://ec.europa.eu/transparency/documents-register/detail?ref=SWD(2023)62&lang=en.
8. A. Dikins, "Wooden market forecast Q1/2023" (Kronus, 2023); https://www.kronus.eu/wooden-market-forecast-q1-2023/.
9. M. Sotirov, T. Schulz, G. Winkel, in *How to Balance Forestry and Biodiversity Conservation—A View across Europe*, F. Krumm, A. Schuck, A. Rigling, Eds. (2020), pp. 62–75.
10. N. Selva, P. Chylarecki, B. G. Jonsson, P. L. Ibisch, *Science* **368**, 1438 (2020).

10.1126/science.adh2303

Credit credibility threatens forests

Addressing global warming requires increased investment in conserving and restoring carbon-dense natural habitats. Some companies that emit carbon have turned to certified carbon credits to offset their environmental impact. However, the effectiveness of carbon credits depends on the methods used to quantify them. If carbon credits do not accurately represent their environmental benefits, relying on them could exacerbate climate change (*1*). To ensure that carbon credits are robust, the methods used to calculate them must be improved.

So far, credits from tropical forest conservation have been generated by estimating project effects through comparisons with historical trends in reference areas identified by project proponents (*2*). However, there is considerable evidence that these methods substantially overestimate the degree to which projects are changing outcomes (i.e., their "additionality") (*3*–*5*).

Carbon crediting procedures also need to get better at accounting for the risk of increased emissions elsewhere. If food or timber production is prevented by a carbon offset project, that production and its carbon impacts may simply move to a different location (*6*, *7*). Most assessments of this

PHOTO: KAROL KALISKÝ; AROLLA FILM

0505_Letters_17071208.indd 466 5/1/23 4:12 PM

carbon "leakage" focus on rough estimates of small-scale leakage and ignore or greatly underestimate longer-range displacement of production (*8*).

Finally, the carbon credit system needs more robust ways of accounting for the impermanence of carbon held in vegetation and soils. Current certification methods try to underwrite credit permanence claims by maintaining a shared pool of nontradable credits which can be drawn from in the event of reversals (*9*). However, this system provides no incentive for future stakeholders to prevent carbon releases after credits have been issued (*10*).

Better methods are available. Additionality can now be estimated more reliably using statistical techniques developed over many decades in economics and public health to eradicate bias in estimating the counterfactual—i.e., what would have happened without the intervention (*11*, *12*). Likewise, there are new methods to better adjust for leakage effects (*8*) and properly value impermanent storage (*10*). All of these improvements should be routinely deployed in credit quantification. Carbon credits can be a valuable tool for climate change mitigation and forest conservation, but their success depends on improving their credibility.

Andrew Balmford[1]*, Pedro H. S. Brancalion[2], David Coomes[3], Ben Filewod[4], Ben Groom[5], Alejandro Guizar-Coutiño[3], Julia P. G. Jones[6], Srinivasan Keshav[7], Andreas Kontoleon[8], Anil Madhavapeddy[7], Yadvinder Malhi[9], Erin O. Sills[10], Bernardo B. N. Strassburg[11], Frank Venmans[4], Thales A. P. West[12], Charlotte Wheeler[3], Tom Swinfield[1]

[1]Department of Zoology and Conservation Research Institute, University of Cambridge, Cambridge CB2 3EJ, UK. [2]Department of Forest Sciences, University of São Paulo, 13.418-900 Piracicaba-SP, Brazil. [3]Department of Plant Sciences and Conservation Research Institute, University of Cambridge, Cambridge CB2 3EA, UK. [4]Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, London WC2A 2AE, UK. [5]Land, Environment, Economics, and Policy Institute, Department of Economics, University of Exeter Business School, Exeter EX4 4PU, UK. [6]School of Natural Sciences, Bangor University, Bangor, UK. [7]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, UK. [8]Department of Land Economy, University of Cambridge, Cambridge CB3 9EP, UK. [9]Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, UK. [10]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695–8008, USA. [11]Rio Conservation and Sustainability Science Centre, Department of Geography and the Environment, Pontifical Catholic University, Rio de Janeiro, Brazil. [12]Institute for Environmental Studies (IVM), VU University Amsterdam, Amsterdam, Netherlands.
*Corresponding author.
Email: a.balmford@zoo.cam.ac.uk

REFERENCES AND NOTES

1. P. Greenfield, "Revealed: More than 90% of rainforest carbon offsets by biggest provider are worthless, analysis shows," *The Guardian* (2023).
2. Verra, "VCS Standard 4.4" (2022).
3. T. A. P. West *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24188 (2020).
4. A. Guizar-Coutiño *et al*., *Conserv. Biol.* **36**, e13970 (2022).
5. P. Delacotte, G. le Velly, G. Simonet, *Res. Energ. Econ.* **67**, 101277 (2022).
6. C. Streck, *Clim. Pol.* **21**, 843 (2021).
7. N. Villoria *et al*., *Nat. Comm.* **13**, 5476 (2022).
8. B. Filewod, G. McCarney, "Avoiding leakage from nature-based offsets by design" (London School of Economics and Political Science, 2023); http://eprints.lse.ac.uk/117927/.
9. Verra, "Proposal to create a long-term reversal monitoring system" (2021).
10. A. Balmford *et al*., "Realising the social value of impermanent carbon credits," Cambridge Open Engage 10.33774/coe-2023-5v93l-v3 (2023).
11. G. W. Imbens, J. D. Angrist, *Econometrica* **62**, 467 (1994).
12. P. J. Ferraro, M. M. Hanauer, *Ann. Rev. Env. Res.* **39**, 495 (2014).

COMPETING INTERESTS

A.B., D.C., S.K., A.M., and T.S. are directors of the Cambridge Centre for Carbon Credits (4C), a research partnership aimed at improving the robustness of estimates of the performance of carbon credit–generating projects. 4C directors also serve as unpaid members of the Cambridge Offsetting Working Group advising the University of Cambridge on their carbon offsetting decisions. P.H.S.B. and B.B.N.S. are respectively partner at and founder of Re.green, a company dedicated to restoring native ecosystems, and have an equity stock. C.W. and T.S. are funded by the Tezos Foundation gift to 4C (grant code NRAG/719).

10.1126/science.adh3426

List child dependents on death certificates

Models suggest that at least 10.5 million children experienced COVID-19–associated orphanhood and caregiver loss in the first 2 years of the pandemic (*1*), and the numbers continue to grow (*2*). However, public health data cannot identify children experiencing orphanhood. More information about such children could facilitate the delivery of support, services, and loving care, minimizing the negative effects of orphanhood (*3*–*5*). To obtain this data, death certificates should include the number of children under the age of 18 living in the home of the deceased parent or caregiver.

Recording dependent children on death certificates has led to support for children in Brazil, where collecting names and ages of dependent children is standard practice for all adult deaths (*6*). In response to data on the numbers of children who lost caregivers during the pandemic, one municipality in São Paulo, Brazil passed legislation to identify children who lost one or both parents to COVID-19, assess their needs, and connect them with services such as grief counseling and psychosocial, educational, and economic support (*7*). Furthermore, 11 Brazilian states have ratified or are considering bills providing economic support for such children (*8*).

Other countries should adopt this model, which could be expanded to allow identification of children orphaned as a result of any cause of death. The World Health Organization (WHO) publishes an "International form of medical certificate of cause of death" with two parts (*9*). The first section, cause of death, must be completed by a doctor. The second section can be filled out by others and includes questions about where the death occurred and whether the deceased was pregnant. The WHO should add a question to the second section asking for the number of children under the age of 18 residing in the home.

To maximize the effectiveness of reporting dependents, countries must work to collect comprehensive death records. Only about two-thirds of the 55 million annual global deaths are registered in civil registration and vital statistics (CVRS) systems, but the WHO CVRS Strategic Implementation Plan aims to increase their use (*10*). Ensuring complete records that include dependents, regardless of the caregiver's cause of death, could standardize aid for orphaned children across the world.

Seth Flaxman[1]*, Lackson Kasonka[2], Lucie Cluver[1], Andrea Santos Souza[3], Charles A. Nelson III[4], Alexandra Blenkinsop[5], H. Juliette T. Unwin[5], Susan Hillis[1]

[1]University of Oxford, Oxford, UK. [2]Zambia Ministry of Health, Lusaka, Zambia. [3]Ministério Público do Estado de São Paulo, São Paulo, Brazil. [4]Harvard University, Cambridge, MA, USA. [5]Imperial College London, London, UK.
*Corresponding author.
Email: seth.flaxman@cs.ox.ac.uk

REFERENCES AND NOTES

1. S. Hillis *et al*., *JAMA Pediatr.* **176**, 1145 (2022).
2. Imperial College London, "Imperial College London–COVID-19 orphanhood calculator" (2021); https://imperialcollegelondon.github.io/orphanhood_calculator/#/country/Global.
3. M. Dhaliwal *et al*., *BMJ* **376**, e068123 (2022).
4. S. Handa, C. T. Halpern, A. Pettifor, H. Thirumurthy, *PLOS One* **9**, e85473 (2014).
5. T. Thomas, M. Tan, Y. Ahmed, E. L. Grigorenko, *Ann. Behav. Med.* **54**, 853 (2020).
6. "Artigo 80 da Lei n° 6.015 de Dezembro de 31 de 1973," *Jusbrasil* (1973); https://www.jusbrasil.com.br/topicos/11326617/artigo-80-da-lei-n-6015-de-31-de-dezembro-de-1973 [in Portuguese].
7. Leis Municipais, São Paulo, Campinas 16.200 2022 de 16 de Março de 2022 (2022); https://leismunicipais.com.br/a/sp/c/campinas/lei-ordinaria/2022/1620/16200/lei-ordinaria-n-16200-2022-institui-plano-de-acao-destinado-as-criancas-e-aos-adolescentes-em-situacao-de-orfandade-causada-pela-covid-19-no-municipio-de-campinas [in Portuguese].
8. Centers for Disease Control and Prevention, "Global orphanhood associated with COVID-19" (2022); https://www.cdc.gov/globalhealth/covid-19/orphanhood/index.html.
9. World Health Organization, "Cause of death" (2023); https://www.who.int/standards/classifications/classification-of-diseases/cause-of-death.
10. "WHO civil registration and vital statistics strategic implementation plan 2021–2025" (World Health Organization, 2021); https://apps.who.int/iris/handle/10665/342847.

10.1126/science.adh8784

0505_Letters_17071208.indd 467 5/1/23 4:12 PM

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PHOTO: NOAA

MARINE PRODUCTIVITY

A dusting of nutrients

Most of the nutrients that fuel primary production in the oceans come from water upwelled to the surface from deeper depths, but atmospheric aerosols have long been recognized as another potential source. Westberry *et al*. used measurements of ocean color made by satellites to show that the global distribution of phytoplankton is affected by dust deposition, with impacts that vary from region to region. Climate change is expected to alter the relative importance of this mechanism. —HJS *Science*, abq5252, this issue p. 515

Satellite imagery shows winds carrying dust from the Sahara desert to the northern Atlantic Ocean, providing nutrients that support phytoplankton.

CELL DEATH

Cytotoxic pores require the right splice

Granzyme A (GzmA) is one of the cytotoxic granule proteins that killer lymphocytes deploy to kill targeted tumor cells. Whereas GzmA-dependent cytotoxicity is reported to depend on pyroptosis activated by cleavage of gasdermin B (GSDMB), controversy has arisen over whether all splice variants of *GSDMB* can support the formation of cytotoxic pores. Kong *et al*. assessed the ability of *GSDMB* isoforms to induce tumor cell death and found that only the *GSDMB3* and *GSDMB4* splice variants retaining exon 6 supported pyroptosis; the other isoforms interfered with pyroptosis through a dominant-negative mechanism. These findings better define the structural requirements for GzmA-dependent pyroptosis involved in antitumor immunity and reveal a mechanism by which some tumor cells can evade this killing pathway through modified splicing of *GSDMB* mRNA. —IRW

Sci. Immunol. (2023) 10.1126/sciimmunol.adg3196

ROBOTICS

Moving objects like information

Locomotion over rugged terrain, whether human or robotic, generally requires extensive feedback to allow for adjustments in stride to compensate for cracks, inclines, or changes in surface composition. This ability typically requires a network of sensors to detect changes in terrain. Chong *et al*. show that an alternative approach requiring minimal environmental awareness can guarantee a successful arrival using information theory. The authors draw a parallel between having multiple, connected legs on the robot and having signal transmission protocols that minimize error in transmission—in this case, the "signal" being transmitted is the body of the robot. —MSL

Science, ade4985, this issue p. 509

MITORIBOSOME

Terminating mitochondrial translation

Termination of translation in human mitochondria involves recognition of noncanonical stop codons that in other genomes code for amino acids. To reveal which factor is responsible for this process and to understand the molecular mechanism of stop codon recognition, Saurer *et al.* used a combination of gene editing, ribosomal profiling and cryo–electron microscopy to investigate the termination process in vivo and in vitro. The specific recognition of the noncanonical stop codon by the identified release factor results

0505_ISIO_17060199.indd 491 5/1/23 4:14 PM

in a distortion of the messenger RNA and requires participation of the ribosomal RNA, a recognition mechanism that was not observed for canonical termination complexes. —DJ

Science, adf9890, this issue p. 531

NEUROSCIENCE

The maturation of proper memory formation

The hippocampal episodic memory system is not present at birth but develops during childhood. Ramsaran *et al*. examined hippocampal engrams in juvenile and adult mice and identified a cascade of events that underlie the emergence of episodic-like memory precision. The immature hippocampus forms dense engrams and imprecise memories. The ability to form sparse engrams does not emerge until the fourth postnatal week with the maturation of inhibitory circuits in the hippocampus. The maturation of perineuronal nets, extracellular matrix structures primarily ensheathing the soma and proximal dendrites of parvalbumin-containing interneurons, helps to drive inhibitory interneuron maturation in the cortex and hippocampus. —PRS

Science, ade6530, this issue p. 543

GALAXIES

Recycled gas feeding a massive galaxy

Galaxies accrete gas from the surrounding intergalactic medium and then turn this gas into stars. Feedback processes such as supernova explosions enrich the gas with elements heavier than helium and can impart enough momentum to eject some gas out of the galaxy. Zhang *et al*. observed the intergalactic medium around a massive galaxy at redshift 2.3. In addition to emission lines caused by hydrogen and helium, they observed lines for carbon, indicating that the gas has been enriched with heavier elements. The kinematics are consistent with streams of gas spiraling toward the massive galaxy. The authors propose that the enriched gas has been recycled from an earlier period of star formation. —KTS

Science, abj9192, this issue p. 494

CANCER

Exploiting a stress response

Pancreatic ductal adenocarcinoma (PDAC) is associated with extremely poor survival, and better treatments are urgently required. Kartha *et al.* found that high sirtuin 6 (SIRT6) abundance defines classical PDAC and controls activating transcription factor 4 (ATF4) by regulating its stability. ATF4 is known to control the integrated stress response, which is constitutively active in classical PDAC. The more aggressive basal subtype was characterized by low SIRT6 expression, leading to low ATF4 abundance and therefore poor activation of the integrated stress response. Basal PDAC xenografts were therefore sensitive to the inhibition of two cyclin-dependent kinases, suggesting a potential treatment avenue for this extremely aggressive cancer subtype. —MN

Sci. Transl. Med. (2023) 10.1126/scitranslmed.abn9674

PETROLOGY

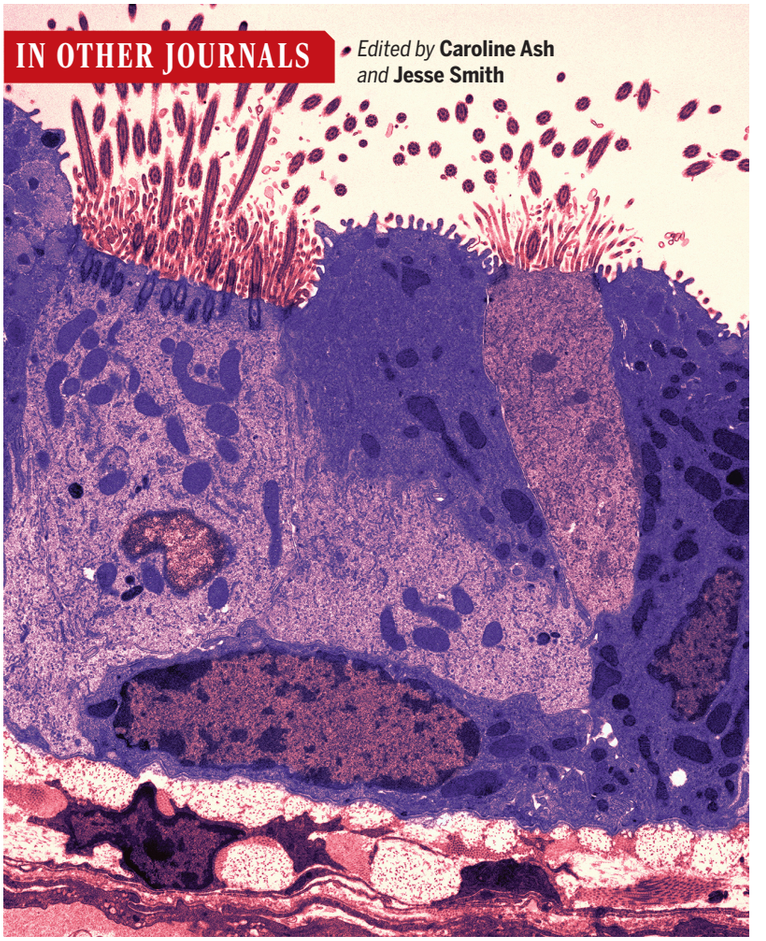
Getting rid of garnet

The formation of garnet-bearing cumulates is one leading candidate to explain the chemistry of oceanic arc lavas, which are ultimately responsible for building continents. Holycross and Cottrell tested this hypothesis by running a series of experiments to determine how garnet formation changes the amount and speciation of iron. They found that garnet does not have a large impact and is unlikely to be the sole cause of the oxidation of arc lavas. The observed chemistry requires a different explanation. —BG

Science, ade3418, this issue p. 506

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



METABOLISM

All protein is equal

High protein consumption has been shown to be beneficial for endurance athletes by promoting myofibrillar protein synthesis. However, whether the source of protein (plants or animals) affects the rate of myofibrillar protein synthesis remains strongly debated. Monteyne *et al.* evaluated myofibrillar protein synthesis before and after leg resistance exercise in 16 young adults testing high-protein omnivorous or plant-based diets for 3 days. The two diets contained similar amounts of macronutrients, including proteins. Resting and exercised daily myofibrillar protein synthesis was comparable between the two dietary regimens. The results support the idea that the source of protein is not a major factor in modulating skeletal muscle remodeling induced by exercise. —MMa

J. Nutr. (2023) 10.1016/j.tjnut.2023.02.023

EVOLUTION

Bug-eyed vertebrates

The camera-like vertebrate eye was considered by Darwin to be one of the greatest challenges to the theory of evolution through gradual modifications. Kalluraya *et al.* show that an essential protein in the vertebrate eye acquired a domain from a bacterial gene more than 500 million years ago through horizontal gene transfer. Interestingly, the protein, called interphotoreceptor retinoid-binding protein, facilitates the physical separation between

PHOTO: AMI IMAGES/SCIENCE SOURCE

0505_ISIO_17060199.indd 492 5/1/23 4:14 PM

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

SOCIAL BELONGING

Improving social belonging among college students

The rate of earning university degrees in the United States is very unequal across social class, race, and ethnicity. Interventions that promote academic persistence also work differently for students from different backgrounds and operate differently across contexts. Walton *et al.* conducted a randomized controlled trial to systematically explain and understand these heterogenous effects in a brief online intervention across 22 universities and colleges (see the Perspective by Bowman). The intervention was designed to remedy students' concerns about belonging through a reading-and-writing activity that emphasized how worries about fitting in, struggling in class, and feeling homesick during the college transition are common and improve over time. They found that the intervention improved retention and persistence in school, particularly among historically underrepresented students, when the school context offered students opportunities to belong. The findings have policy implications for academic institutions that strive to better support and retain diverse students. —EEU

Science, ade4420, this issue p. 499; see also adh7681, p. 457

LOCUST SWARMING

Don't eat me

Locusts spend much of their lives living as peaceful herbivores. However, in response to various cues, they change morphologies, aggregate, and create huge swarms. Swarming locusts look different and behave differently, including engaging in cannibalism. Chang *et al*. identified a mechanism though which young locusts can limit their potential to be preyed upon by conspecifics (see the Perspective by Couzin and Couzin-Fuchs). A specific compound, phenylacetonitrile, is produced by locust nymphs raised in crowded conditions and protects the animals from their cannibalistic relatives. Characterization of this mechanism could contribute to an improved understanding of locust swarming behavior. —SNV

Science, ade6155, this issue p. 537; see also adh5264, p. 454

MAGNETISM

Imaging emergent strings

Frustrated magnetic systems, such as spin ices, do not have a ground state in which each spin is at an energy minimum. In an artificial spin ice geometry called Santa Fe ice, this frustration can be viewed through the magnetic moment configurations at the vertices of the magnetic array. Zhang *et al*. monitored the kinetics of the strings formed by "unhappy" vertices—those not in their lowest energy state. Using x-ray magnetic circular dichroism photoemission electron microscopy, the researchers took images of the system once a second and analyzed the observed changes in string configurations. The system underwent a crossover between a low-temperature regime with limited string motion to a high-temperature one characterized by changes in string topology. —JS

Science, add6575, this issue p. 526

STEROL BIOSYNTHESIS

Sitosterol is not just for plants

For most animals, cholesterol is an essential membrane lipid and signaling molecule. Plants and fungi use chemically similar molecules with one or two additional carbon atoms. Michellod *et al*. studied two species of gutless marine annelids and found that they produced sitosterol, a common plant lipid, as the main sterol in their lipid membranes (see the Perspective by Brocks and Bobrovskiy). A noncanonical methyltransferase that is known in plants but not previously known in bilaterian animals appears to be key in generating this sterol. Additional analyses of genomic and transcriptomic data from annelids and several other animal phyla demonstrated that this biosynthesis gene is widespread and has a complex evolutionary history. —MAF

Science, add7830, this issue p. 520; see also adh8097, p. 455

CORONAVIRUS

Taking the STING out of host defenses

The papain-like proteases (PLPs) from coronaviruses remove the proteins ubiquitin and ISG15 from host proteins to suppress antiviral signaling. Xiong *et al*. found that in contrast to those from the highly pathogenic coronaviruses, PLPs from mildly pathogenic coronaviruses preferred ubiquitin over ISG15. Coronavirus pathogenicity also correlated with the PLP's ability to inhibit antiviral signaling. In a related paper, Cao *et al*. showed that the SARS-CoV-2 PLP deubiquitinated the host defense protein STING. A STING agonist and a PLP inhibitor synergistically reduced SARS-CoV-2 replication in human lung cells. —AMV

Sci. Signal. (2023) 10.1126/scisignal.ade1985, 10.1126/scisignal.add0082

in a distortion of the messenger RNA and requires participation of the ribosomal RNA, a recognition mechanism that was not observed for canonical termination complexes. —DJ

Science, adf9890, this issue p. 531

NEUROSCIENCE

The maturation of proper memory formation

The hippocampal episodic memory system is not present at birth but develops during childhood. Ramsaran *et al*. examined hippocampal engrams in juvenile and adult mice and identified a cascade of events that underlie the emergence of episodic-like memory precision. The immature hippocampus forms dense engrams and imprecise memories. The ability to form sparse engrams does not emerge until the fourth postnatal week with the maturation of inhibitory circuits in the hippocampus. The maturation of perineuronal nets, extracellular matrix structures primarily ensheathing the soma and proximal dendrites of parvalbumin-containing interneurons, helps to drive inhibitory interneuron maturation in the cortex and hippocampus. —PRS

Science, ade6530, this issue p. 543

GALAXIES

Recycled gas feeding a massive galaxy

Galaxies accrete gas from the surrounding intergalactic medium and then turn this gas into stars. Feedback processes such as supernova explosions enrich the gas with elements heavier than helium and can impart enough momentum to eject some gas out of the galaxy. Zhang *et al*. observed the intergalactic medium around a massive galaxy at redshift 2.3. In addition to emission lines caused by hydrogen and helium, they observed lines for carbon, indicating that the gas has been enriched with heavier elements. The kinematics are consistent with streams of gas spiraling toward the massive galaxy. The authors propose that the enriched gas has been recycled from an earlier period of star formation. —KTS

Science, abj9192, this issue p. 494

CANCER

Exploiting a stress response

Pancreatic ductal adenocarcinoma (PDAC) is associated with extremely poor survival, and better treatments are urgently required. Kartha *et al.* found that high sirtuin 6 (SIRT6) abundance defines classical PDAC and controls activating transcription factor 4 (ATF4) by regulating its stability. ATF4 is known to control the integrated stress response, which is constitutively active in classical PDAC. The more aggressive basal subtype was characterized by low SIRT6 expression, leading to low ATF4 abundance and therefore poor activation of the integrated stress response. Basal PDAC xenografts were therefore sensitive to the inhibition of two cyclin-dependent kinases, suggesting a potential treatment avenue for this extremely aggressive cancer subtype. —MN

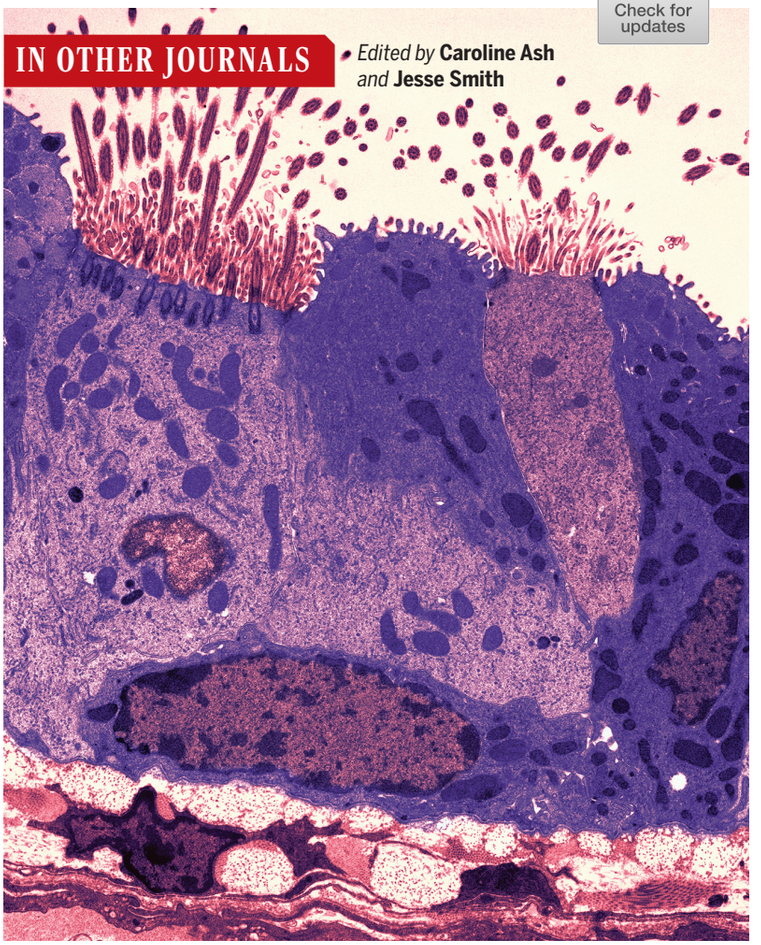
Sci. Transl. Med. (2023) 10.1126/scitranslmed.abn9674

PETROLOGY

Getting rid of garnet

The formation of garnet-bearing cumulates is one leading candidate to explain the chemistry of oceanic arc lavas, which are ultimately responsible for building continents. Holycross and Cottrell tested this hypothesis by running a series of experiments to determine how garnet formation changes the amount and speciation of iron. They found that garnet does not have a large impact and is unlikely to be the sole cause of the oxidation of arc lavas. The observed chemistry requires a different explanation. —BG

Science, ade3418, this issue p. 506



METABOLISM

All protein is equal

High protein consumption has been shown to be beneficial for endurance athletes by promoting myofibrillar protein synthesis. However, whether the source of protein (plants or animals) affects the rate of myofibrillar protein synthesis remains strongly debated. Monteyne *et al.* evaluated myofibrillar protein synthesis before and after leg resistance exercise in 16 young adults testing high-protein omnivorous or plant-based diets for 3 days. The two diets contained similar amounts of macronutrients, including proteins. Resting and exercised daily myofibrillar protein synthesis was comparable between the two dietary regimens. The results support the idea that the source of protein is not a major factor in modulating skeletal muscle remodeling induced by exercise. —MMa

J. Nutr. (2023) 10.1016/j.tjnut.2023.02.023

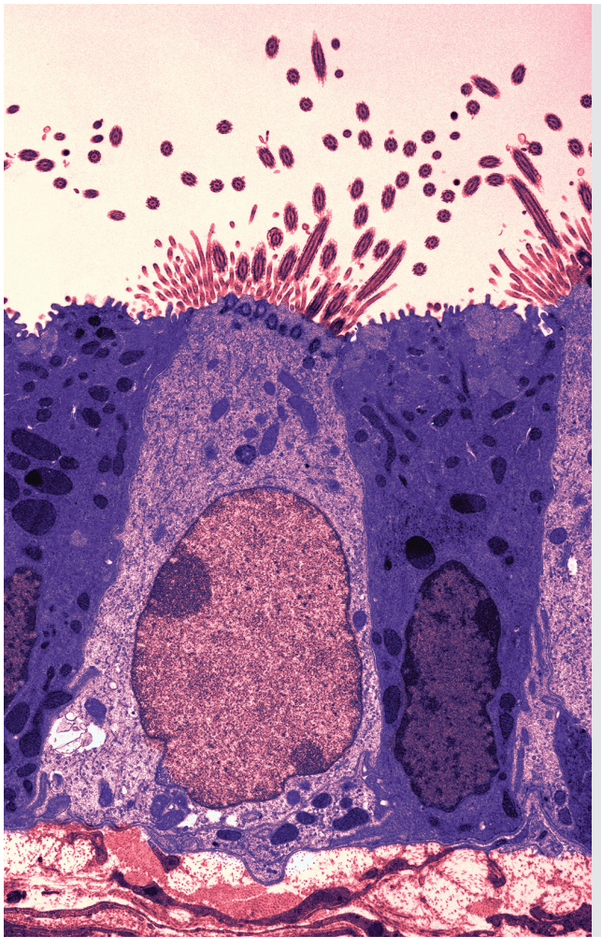
EVOLUTION

Bug-eyed vertebrates

The camera-like vertebrate eye was considered by Darwin to be one of the greatest challenges to the theory of evolution through gradual modifications. Kalluraya *et al.* show that an essential protein in the vertebrate eye acquired a domain from a bacterial gene more than 500 million years ago through horizontal gene transfer. Interestingly, the protein, called interphotoreceptor retinoid-binding protein, facilitates the physical separation between

PHOTO: AMI IMAGES/SCIENCE SOURCE

0505_ISIO_17060199.indd 492 5/1/23 4:14 PM



DIFFERENTIATION

Maintaining lung epithelia

The lung contains a variety of alveolar epithelial cell types, including alveolar type 1 (AT1) cells, which are required for efficient gas exchange, and AT2 cells, which recycle pulmonary surfactant and can differentiate into AT1 cells. Shiraishi *et al.* identified AT1 cells as the important mechanosensor in the alveolar gas-exchange niche of the lung. The authors combined cell and molecular biology with cell type–specific lineage tracing to examine AT1 and AT2 cells from mouse and human lung tissues. Cyclical tissue strain from normal breathing movements was required to maintain cell fate in the lung alveolus, and loss of these movements led to AT1-AT2 cell reprogramming. The underlying mechanism involved actin-generated constraint of the nucleus, which affected nuclear lamina-chromatin interactions. Therefore, these results confirm that biophysical forces are important for maintaining lung cell fate and identity. —SMH

Cell (2023) 10.1016/j.cell.2023.02.010

Cell fate in the lung epithelium, shown here in a transmission electron micrograph, is sustained by the mechanical strain of breathing.

light sensing and retinoid regeneration. This is a unique feature of the vertebrate eye enabling sensing in low light conditions. These authors have uncovered a key transition in building the complex vertebrate eye and highlight the ways genes from the bacterial world have been coopted in the evolution of higher organisms. —DJ

Proc. Natl. Acad. Sci. U. S. A. (2023) 10.1073/pnas.2214815120

HEART DISEASE

Taking a toll on the aorta

Calcification of the aortic valve is a major cause of aging-associated cardiovascular disease. In this condition, fibroblasts residing within the aortic valve acquire a bone-forming phenotype and deposit calcium, which damages the valve. By studying samples from human aortic valves, as well as those from mice and zebrafish, Gollmann-Tepeköylü *et al*. identified a pathway involving toll-like receptor 3 and interferon signaling that contributes to aortic calcification. In young animals, this pathway is necessary for normal bone formation, but it acquires a pathogenic activity later in life and may be amenable to pharmacological intervention. —YN

Circulation (2023) 10.1161/CIRCULATIONAHA.122.063481

OPTICAL TWEEZERS

An atomic game of throw-and-catch

Optical tweezing is a powerful technique for trapping and manipulating single particles, molecules, and atoms. Arrays of atoms trapped in a periodic landscape of optical potential where the interactions between atoms can be controlled could be useful for the development of quantum simulators. Hwang *et al*. demonstrate the ability to throw and catch single rubidium atoms with optical tweezers. The thrower is an accelerating optical tweezer, and the catcher is a decelerating tweezer. With the quantum state persevered as the atom makes its flight between thrower and catcher, these literal "flying qubits" could be used to transport atoms across atom chips being developed for dynamic quantum computing and information processing architectures. —ISO

Optica (2023) 10.1364/OPTICA.480535

ARCTIC CLIMATE

Switching drivers

Anthropogenic activity has warmed the Arctic precipitously, resulting in less sea-ice coverage during more of the year. Before warming began to cause sea ice coverage to decrease, summer clouds had little direct effect on maximum annual sea surface temperatures because the sea surface was largely insulated by its ice cover. Sledd *et al.* now report that this situation has changed. As the ocean loses its sea ice cover earlier in summer, clouds have begun to exert a stronger influence on how much sunlight reaches the sea surface. Clouds will continue to have a greater role in moderating summer high sea surface temperature extremes in the Arctic in a warmer future. —HJS

Geophys. Res. Lett. (2023) 10.1029/2023GL102850

BLACK HOLES

AI sharpens an image of a black hole

Radio interferometry has previously been used to image the supermassive black hole in the nearby galaxy Messier 87. It appeared as a ring at submillimeter wavelengths, consistent with the predictions of general relativistic magnetohydrodynamic (GRMHD) simulations. Medeiros *et al.* have harnessed an artificial intelligence (AI) approach to reanalyze the interferometric data using a machine learning algorithm trained on GRMHD simulations. They found a narrower ring of emission than the previous analysis, with a more clearly defined gap at the location of the black hole. The method provides a substantially sharper image and could be applied to other nearby supermassive black holes, but it implicitly assumes that general relativity is correct. —KTS

Astrophys. J. Lett. (2023) 10.3847/2041-8213/acc32d

0505_ISIO_17060199.indd 493 5/1/23 4:14 PM

RESEARCH ARTICLE

GALAXIES

Inspiraling streams of enriched gas observed around a massive galaxy 11 billion years ago

Shiwu Zhang[1]†, Zheng Cai[1,2]*†, Dandan Xu[1], Rhythm Shimakawa[3,4], Fabrizio Arrigoni Battaia[5], Jason Xavier Prochaska[6,7], Renyue Cen[8,9], Zheng Zheng[10], Yunjing Wu[1], Qiong Li[11,12], Liming Dou[13], Jianfeng Wu[14], Ann Zabludoff[15], Xiaohui Fan[15], Yanli Ai[16], Emmet Gabriel Golden-Marx[1], Miao Li[17], Youjun Lu[18,19], Xiangcheng Ma[20], Sen Wang[1], Ran Wang[11], Feng Yuan[21]

Stars form in galaxies, from gas that has been accreted from the intergalactic medium. Simulations have shown that recycling of gas—the reaccretion of gas that was previously ejected from a galaxy—could sustain star formation in the early Universe. We observe the gas surrounding a massive galaxy at redshift 2.3 and detect emission lines from neutral hydrogen, helium, and ionized carbon that extend 100 kiloparsecs from the galaxy. The kinematics of this circumgalactic gas is consistent with an inspiraling stream. The carbon abundance indicates that the gas had already been enriched with elements heavier than helium, previously ejected from a galaxy. We interpret the results as evidence of gas recycling during high-redshift galaxy assembly.

Simulations of galaxy formation in the early Universe indicate that low-mass galaxies grow by the direct accretion of gas from the circumgalactic medium (CGM) and intergalactic medium (IGM) (*1*). Both simulations and observations have shown that galaxies in low-mass dark matter halos—those with a halo mass $M_h < 10^{12}$ solar masses ($M_\odot$) (*2*, *3*)—can accrete streams of gas at temperatures of ~10^4 K. These streams link the galaxy to the surrounding CGM and IGM with a filamentary web of pristine gas (gas containing almost no elements heavier than helium) (*1*, *2*). Transport of gas along these streams prevents it from being shock-heated while falling into the potential well of the dark matter halo, so this process is referred to as "cold-mode" accretion (*2*). Cold-mode accretion could explain the high star formation rate (SFR) of high-redshift galaxies and the increase in angular momentum of galaxy halos over time (*4*).

Although cold-mode accretion is expected for pristine gas, predictions differ for metal-enriched gas (gas with higher abundances of elements heavier than helium, referred to as its metallicity). Cosmological simulations predict potentially observable quantities of metal-enriched CGM around galaxies with $M_h > 10^{12}\ M_\odot$ (*5*, *6*). Because metal-enriched CGM gas can cool more efficiently than pristine gas, metal-enriched accretion (recycled inflow) could provide additional gas and boost the SFR of galaxies in massive halos ($M_h > 10^{12}\ M_\odot$) at redshift $z > 2$ (*7*–*10*). Observations of absorption lines toward background sources have implied the presence of metal-enriched CGM around galaxies (*11*), but these only provide information at a single point. To determine the spatial distribution of CGM gas requires studying its emission lines.

Observations of MAMMOTH-1

Lyman alpha (Lyα) is an emission line of neutral hydrogen that has a rest-frame wavelength of 1216 Å. Observations of enormous Lyα nebulae with wide-field integral field spectrographs could potentially determine the physical properties and kinematics of CGM at $z > 2$. One such Lyα nebula is the MAMMOTH-1 nebula at $z \approx 2.31$ (coordinates: right ascension $14^h41^m24^s.42$, declination +40°03′09″.7, J2000 equinox), which has Lyα emission with a projected spatial extent of 442 kpc and Lyα luminosity $L_{Ly\alpha} = 5.1 \pm 0.1 \times 10^{44}$ erg s^{-1} (*12*). It resides in an overdense galaxy environment (*12*–*14*).

Our team observed MAMMOTH-1 with the Keck Cosmic Web Imager (KCWI) on the 10-m Keck II telescope in imaging spectroscopy mode centered on the Lyα, C IV 1548/1550 Å, and He II 1640 Å emission lines. We also performed narrowband imaging of redshifted Hα (another line of neutral hydrogen with a rest-frame wavelength of 6563 Å) emission using the Multi-Object InfraRed Camera and Spectrograph (MOIRCS) on the 8-m Subaru telescope. We supplement these data with archival observations at near-infrared, x-ray, and radio wavelengths (*15*).

We optimally extracted images from the KCWI data (Fig. 1, A to C) (*13*, *15*). These show that the flux peaks of the emission lines Lyα, C IV, and He II coincide with a quasar (at $14^h41^m24^s.42$, +40°03′09″.7, J2000) detected in the x-ray data (fig. S1B), which we designate as MAMMOTH-1:G-2 (hereafter G-2) (*15*). G-2 is located inside the nebula and provides the ionizing photons that excite the gas emission lines. The Lyα, C IV, and He II emission regions are asymmetrically distributed around G-2. Both He II and C IV are spatially extended, with projected scales of 88 and 108 kpc (2σ emission), respectively, and have luminosities of $L_{He\ II} = 5.48 \pm 0.13 \times 10^{42}$ erg s^{-1} and $L_{C\ IV} = 10.50 \pm 0.16 \times 10^{42}$ erg s^{-1}, respectively. The MOIRCS narrowband imaging (Fig. 1J) shows that the Hα emission has a projected scale of 97 kpc, with a luminosity of $L_{H\alpha} = 2.33 \pm 0.15 \times 10^{43}$ erg s^{-1}. From the archival radio observations of CO ($J = 1 \rightarrow 0$, where J is the rotational quantum number) (*14*), we measure the redshift of G-2 to be $z = 2.3116 \pm 0.0004$ (*15*). The CO ($J = 1 \rightarrow 0$) and CO ($J = 3 \rightarrow 2$) observations (fig. S1A) (*14*, *16*) show that the sources around G-2 we designate as G-1, G-3, G-4, G-5, and G-6 (Fig. 1) also have redshifts of $z \approx 2.31$ and thus are located within the nebula.

We used the spectral information to construct a flux-weighted Lyα velocity map (Fig. 1D). This shows two regions of gas, each with gas at a similar velocity extending for ≥100 kpc, which we refer to as regions A and B. The ionizing source, G-2, is located within region A. One-dimensional (1D) spectra (Fig. 2) extracted from selected apertures (labeled in Fig. 1E) have double- or triple-peaked structures, which we fitted with models consisting of multiple Gaussians. The Lyα and He II emission spectra have similar double- or triple-peaked profiles at each location, indicating that the line profiles are attributable to the motion of the ionized gas rather than Lyα radiative transfer effects (*17*). Using smaller apertures (fig. S2) confirms this result (*15*). In the large-scale CGM where no He II is detected,

[1]Department of Astronomy, Tsinghua University, Beijing 100084, China. [2]School of Mathematics and Physics, Qinghai University, Xining 810016, China. [3]National Astronomical Observatory of Japan, National Institutes of Natural Sciences, Tokyo 181-8588, Japan. [4]Waseda Institute for Advanced Study, Waseda University, Tokyo 169-0051, Japan. [5]Max-Planck-Institut für Astrophysik, D-85748 Garching bei München, Germany. [6]Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA. [7]Kavli Institute for the Physics and Mathematics of the Universe, The University of Tokyo, Kashiwa 277-8583, Japan. [8]Department of Physics, Zhejiang University, Hangzhou 310027, China. [9]Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA. [10]Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA. [11]Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China. [12]Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester M13 9PL, UK. [13]Department of Astronomy, Guangzhou University, Guangzhou 510006, China. [14]Department of Astronomy, Xiamen University, Xiamen 361005, China. [15]Steward Observatory, University of Arizona, Tucson, AZ 85721, USA. [16]College of Engineering Physics, Shenzhen Technology University, Shenzhen 518118, China. [17]Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA. [18]National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China. [19]School of Astronomy and Space Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [20]Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA, 94720, USA. [21]Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China.
*Corresponding author. Email: zcai@mail.tsinghua.edu.cn
†These authors contributed equally to this work.

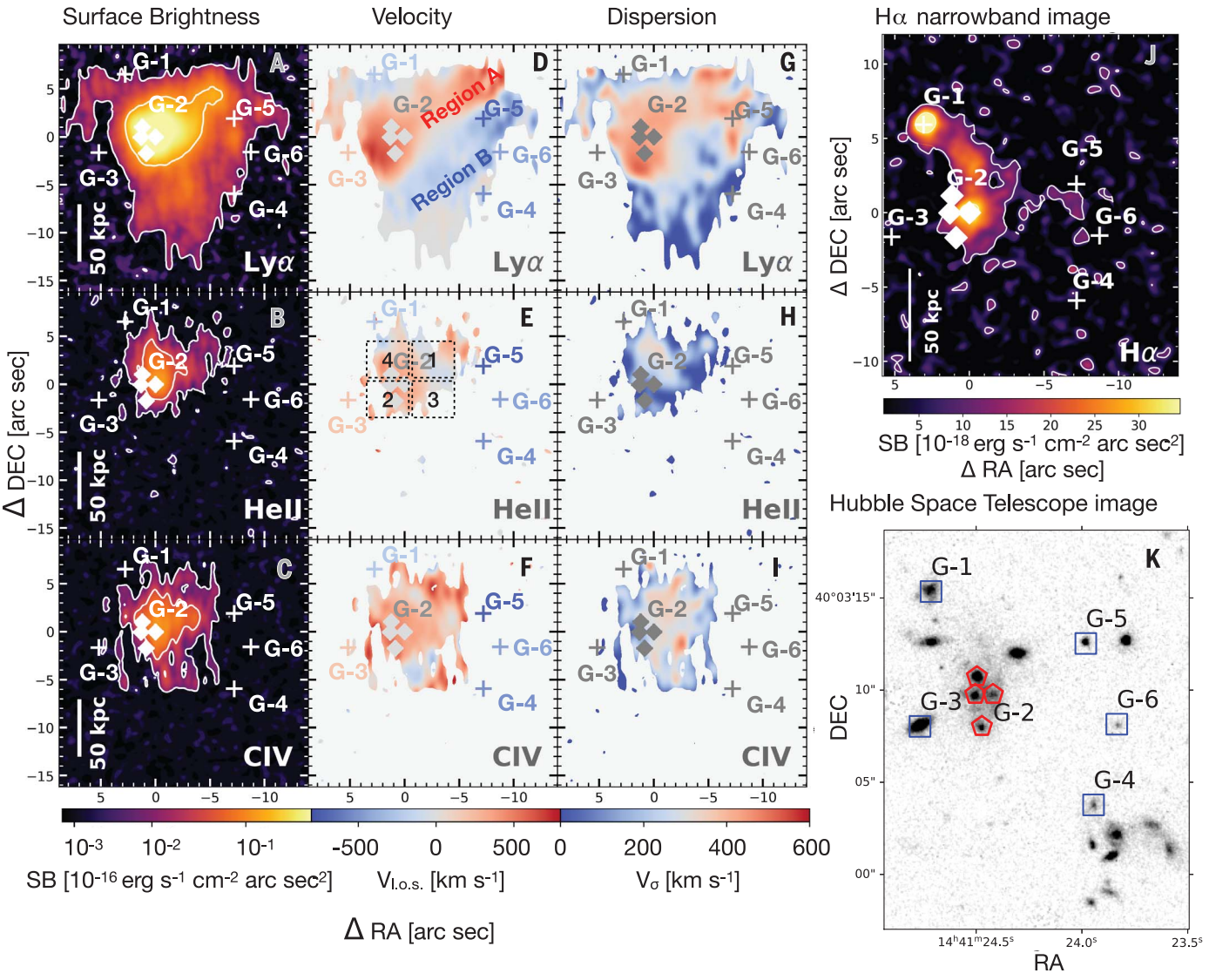


Fig. 1. Images of the extended line emission. (**A** to **C**) Optimally extracted surface brightness (SB) images of Lyα, He II, and C IV emission from the KCWI datacube. ΔRA and ΔDEC are the coordinate offsets in right ascension and declination, respectively, relative to the position of G-2 ($14^h41^m24^s.42$, +40°03′09″.7, J2000). Cross and diamond symbols labeled G-1 to G-6 are galaxies at redshift z = 2.31, measured in both the CO (J = 1→0) (*14*) and CO (J = 3→2) observations (*16*). White contours outline regions of 2σ and 10σ detection of Lyα emission and regions of 2σ and 8σ detection for He II and C IV. In each case, the emission line extends over CGM scales. (**D** to **F**) Flux-weighted line-of-sight velocity ($v_{l.o.s.}$) maps. The Lyα velocity map is labeled to indicate the large-scale red strip (region A) and blue strip (region B) (see also Fig. 4A). The four dashed squares in the He II velocity map are each 4″ by 4″ wide, and the extracted 1D spectra from each of the four apertures are shown in Fig. 2. (**G** to **I**) Flux-weighted velocity dispersion (v_σ) maps. (**J**) Hα narrowband image (seeing 0.4″). Symbols are the same as in (A) to (C). The flux peak of Hα coincides with the Lyα, He II, and C IV emission. (**K**) Hubble Space Telescope image of the galaxy group, using the filter F160W, which is rest-frame optical. Four clumps within 2″ of G-2 are circled in red, and galaxies at $z \approx 2.3$ are marked with blue boxes. Coordinates are J2000 equinox.

we expect radiative transfer effects to be weaker, so the Lyα also traces the cool gas kinematics in those regions (*15*).

Analysis of line ratios

Using the diffuse C IV, He II, and Hα emission lines, we studied the properties of the CGM with spatially resolved line ratio diagnostics (*15*). Hα emission is better suited for line ratio diagnostics than Lyα because it is less affected by resonant scattering (*18*). Hα is also less affected by dust attenuation; we estimate that the potential effect of dust on the line ratios is <15% (*15*). We consider two possible emission mechanisms: In the pure photoionization scenario, the gas is highly ionized by strong ultraviolet radiation from the quasar, causing the line emission to be dominated by recombination. In the alternative shock-with-precursor scenario, the gas moves at a velocity higher than the local sound speed, so a shock front forms at the leading edges of the clouds. Such shocks could heat the gas to sufficiently high temperatures that it cools by emitting soft x-rays (*19*). Model line ratios are shown in fig. S5 for the pure photoionization scenario and in Fig. 3 for the shock-with-precursor scenario. The observed C IV/Hα ratio is consistent with both scenarios, whereas the He II/Hα ratio can only be produced by the shock-with-precursor scenario—it is an order of magnitude smaller than predicted for the pure photoionization model. We examined alternative assumptions

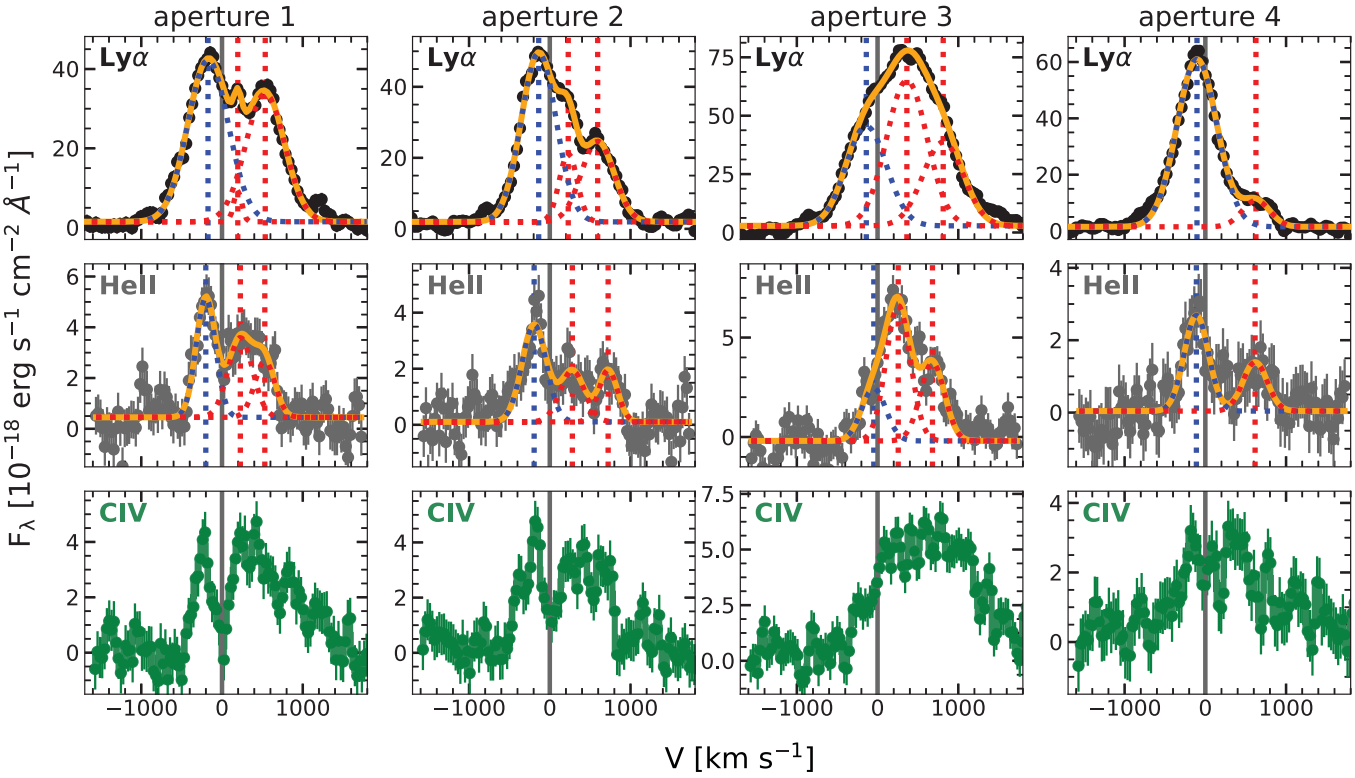


Fig. 2. One-dimensional spectra of the Lyα, He II, and C IV emission. These spectra were extracted from the apertures shown in Fig. 1E. F_λ is the flux density at wavelength λ. Dots (with 1σ error bars) show the observed spectra, while the orange lines are a model consisting of multiple Gaussians fitted to the data. The individual Gaussian components are shown with dashed lines. Numerical values are listed in table S1. The vertical dashed lines mark the peaks of these Gaussian components. Each model requires two or three components to fit the data. Because He II is a nonresonant line, the multiple components indicate gas motion of the CGM rather than resonant scattering effects (*17*). C IV is a doublet, so it was not modeled with Gaussians.

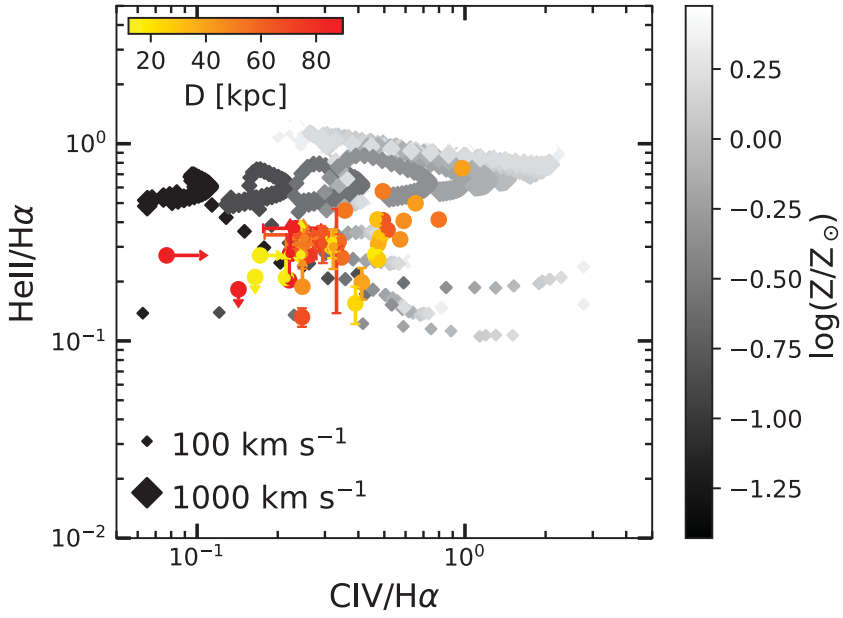


Fig. 3. Line ratios He II/Hα and C IV/Hα for the observations and the shock-with-precursor scenario. Colored dots are the observations, and gray diamonds are predictions from the model (*19*). Error bars are 2σ, and arrows denote lower or upper limits. Colors indicate the projected distance (*D*) between the southeast aperture and each of the other apertures used to extract line emission (fig. S4). For the model, the gray scale represents metallicity (*Z*), and symbol sizes scale with the model shock velocity (100 to 1000 km s^{-1} in steps of 25 km s^{-1}), as indicated in the legend. The observed line ratios of He II/Hα and C IV/Hα are consistent with the shock-with-precursor scenario at CGM metallicity of 0.1 to 1.0 $Z_\odot$, across a distance of ≈100 kpc.

about the quasar emission and dust attenuation, finding that a photoionization scenario is unlikely to produce the observed line ratios in all cases (*15*). We therefore conclude that the shock-with-precursor scenario is the mechanism responsible for the observed emission line.

From the line ratio diagnostics, we find that the CGM metallicities are within 2σ of the solar metallicity ($Z_\odot$) (Fig. 3). This is an order of magnitude higher than previous measurements of the metallicity of the interstellar medium in other galaxies with a stellar mass, $M_\star$, of ≤10^9 $M_\odot$ at $z \approx 2$ to 3 (*20*, *21*) but consistent within 2σ with the CGM metallicity

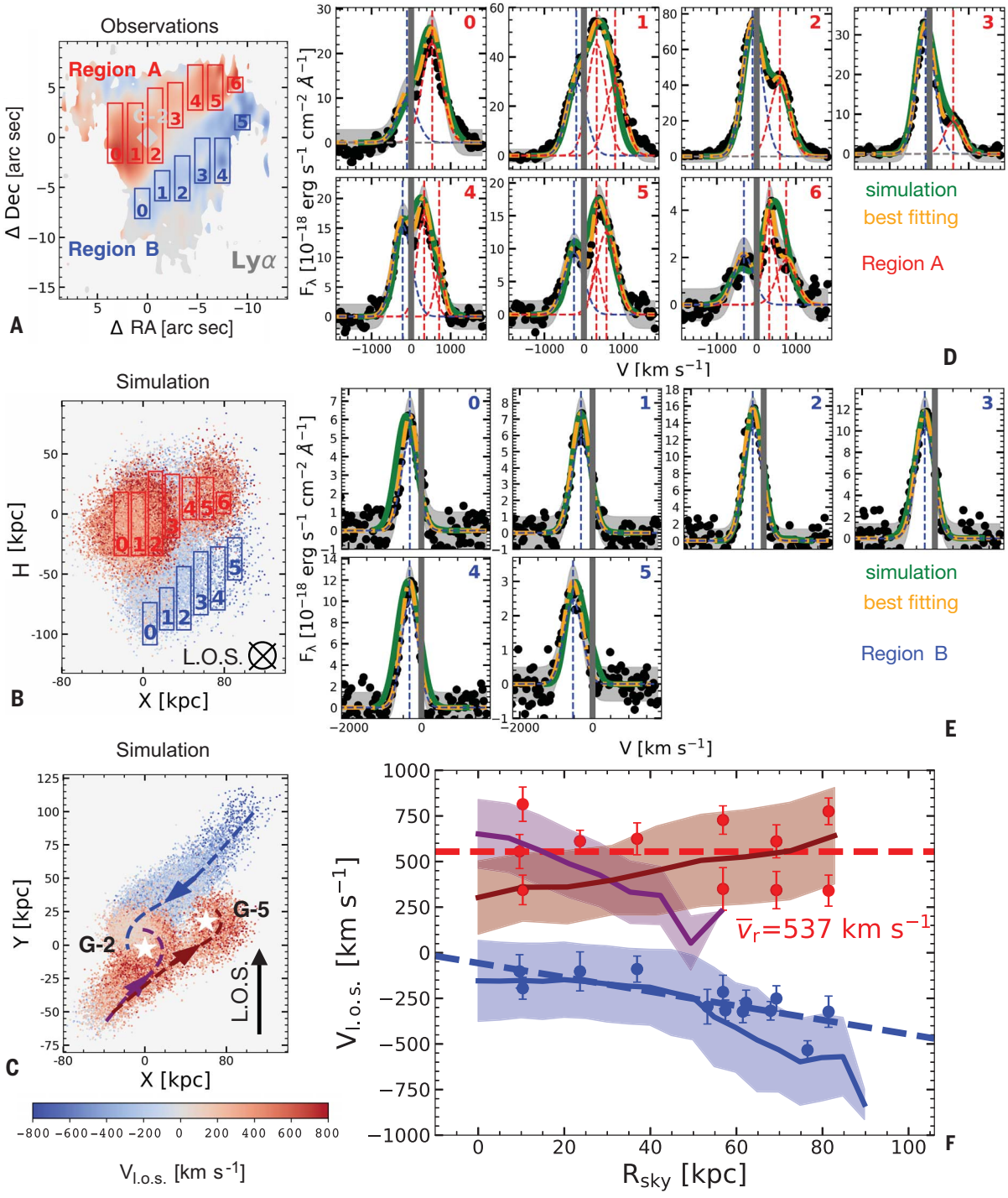


Fig. 4. Interpretation as an inspiraling stream. (**A**) Same as Fig. 1D, except regions A and B are divided into subregions indicated by the boxes, within which the spectra in (D) and (E) were extracted. (**B**) Same as (A), but for our simulation of the inspiraling streams model (*15*). The line of sight (L.O.S.) is along the *y* axis. (**C**) Projection of our model on the *x*–*y* plane. The two white stars mark the positions of G-2 and G-5. Three main inspiraling streams are shown with dashed lines, with arrows indicating flow directions. Two streams are around G-2, and the third stream is around G-5. Other projections of this model are shown in fig. S9. (**D** and **E**) Observed 1D spectra extracted from the regions shown in (A). The orange lines include multiple Gaussians, with each individual component shown as dashed lines, as in Fig. 2. Black dots represent the data, and the gray shaded region indicates the 1σ scatter. The green lines are the simulated spectra extracted from the regions shown in (B). The data and the simulation are consistent. (**F**) Projected line-of-sight velocity profile of the observations (dots) obtained from the red and blue components of the spectra in (D) and (E). R_{sky} is the projected distance from each region to G-2 on the plane of the sky. The 11 red points correspond to the 11 redshifted Gaussians shown in (D), and the 13 blue points correspond to the 13 blueshifted Gaussians shown in (D) and (E). Error bars are 1σ. The dashed lines are linear models fitted to the data points. The solid lines with shaded areas are our simulated velocity profiles, with colors corresponding to the streams in (C). The shaded areas denote the 1σ scatter. The red components have a roughly constant velocity profile, whereas the absolute values of the blue components have an increasing profile. $\overline{v}_r$ is the mean of the observed redshifted velocities.

of some galaxies at $z \approx 0$ (*22*). Previous studies have found that metal-enriched gas, traced by C IV absorption toward background sources, is distributed within ≈125 kpc around galaxies at $z = 2$ to 3 (*23*). Other studies have shown that CGM metallicity could be higher than $Z_\odot$ at $z = 1$ (*24*). Our measurements show that the CGM in MAMMOTH-1 has been enriched to between 0.1 and 1.0 $Z_\odot$ on ≈100-kpc scales.

Interpretation of the kinematics

We investigated whether the observed kinematics could be produced by an active galactic nucleus (AGN) outflow (see supplementary text in the supplementary materials). We find that they cannot, for two reasons: First, we expect any outflow to decelerate as it propagates into the CGM, as a result of energy loss (*25*–*27*). Instead, the observed line-of-sight velocity profile (Fig. 4F) shows that the redshifted velocity is roughly constant and that the absolute value of the blueshifted velocity increases with increasing distance from G-2. Secondly, if the kinematics was due to an AGN outflow, the observations would require an implausibly high coupling efficiency between the outflow power and the AGN luminosity (supplementary text) that is almost one order of magnitude greater than that found by previous observations or simulations (fig. S13).

Alternatively, the gas metallicity, spatial distribution, and kinematics could be due to metal-enriched inspiraling streams. Gas in the CGM could have been enriched and expelled by previous outflows, from the AGN or starbursts in other galaxies. Cosmological simulations predict that inspiraling cool streams allow accreting galaxies to gain high angular momentum (*4*, *28*). In cosmological simulations (*29*), ~46% of halos with masses $M_h \geq 10^{13}\,M_\odot$ at $z = 2$ have inspiraling streams that are metal-enriched on CGM scales (*15*). Those simulations predicted that, in addition to pristine gas accretion, accretion of recycled, metal-enriched gas could be a common process.

Motivated by those simulations, we constructed a simple kinematic model of metal-enriched inspiraling streams to interpret our observations (*15*). The model consists of three inspiraling streams: two that surround G-2 and a third that surrounds G-5. The geometry of the model is shown in Fig. 4C. Comparing this model to the data, we find that both the simulated velocity map (Fig. 4B) and line profiles (Fig. 4, D and E) are consistent with the observations. The reduced χ^2 between the simulated and observed spectra is ~0.9. The kinematics is also reproduced by the model (Fig. 4F), within the 1σ scatter.

Implications for gas accretion

Cosmological simulations have also shown that recycled inflows can provide gas accretion at $z \sim 0$ (*8*, *9*). Other simulations (*7*) have shown that, at $z = 2$, the fraction of stellar mass contributed by recycled gas is 40% in a $10^{12.8}\,M_\odot$ halo. Our interpretation of the MAMMOTH-1 observations is consistent with the latter scenario (*7*).

We calculate that the streams provide a mass inflow rate $\dot{M}_{in} = 703^{+101}_{-78}\,M_\odot\ \mathrm{year}^{-1}$ (*15*), which is higher than the SFR of G-2 ($81 \pm 18\,M_\odot\ \mathrm{year}^{-1}$) derived from its far-infrared emission (*15*). We suggest that this indicates a link between the accretion rate of the recycled gas inflow and the SFR of G-2 (*15*). The observed line-of-sight velocity profile (Fig. 4F) is consistent with cool gas undergoing deceleration as it falls into the dark matter halo, as predicted by semianalytic models of cool CGM gas (*30*, *31*). In those semianalytic models, the deceleration is interpreted as a consequence of drag forces exerted by the hot coronal gas on the cool gas stream. Because MAMMOTH-1 resides in the peak of overdensity that is also a galaxy group, the diffuse metal-enriched gas could also arise from galaxy interactions, such as tidal stripping. We estimate that this scenario is negligible on the CGM scale (supplementary text). A Hubble Space Telescope image of MAMMOTH-1 (Fig. 1K) shows no evidence of tidal stripping on large scales.

Our spectroscopic observations show that the group of galaxies within MAMMOTH-1 has a redshift gradient qualitatively consistent with that of the CGM (Fig. 1D and table S3). This indicates that the large-scale orbital angular momentum of the galaxy group aligns with the CGM angular momentum. Any satellite galaxies moving around G-2 could impart angular momentum to the enriched cool CGM gas, which would then flow back to the galaxy in an inspiraling stream (fig. S8A). The CGM gas flow would then induce a shock, facilitating gas cooling through line emission. From the estimated inflow rate of the recycled gas, the recycled inflow could sustain star formation in G-2 at this redshift.

REFERENCES AND NOTES

1. J. Tumlinson, M. S. Peeples, J. K. Werk, *Annu. Rev. Astron. Astrophys.* **55**, 389–432 (2017).
2. D. Kereš, N. Katz, D. H. Weinberg, R. Davé, *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).
3. J. Stern, D. Fielding, C.-A. Faucher-Giguère, E. Quataert, *Mon. Not. R. Astron. Soc.* **492**, 6042–6058 (2020).
4. K. R. Stewart *et al*., *Astrophys. J.* **843**, 47 (2017).
5. J. Suresh, D. Nelson, S. Genel, K. H. R. Rubin, L. Hernquist, *Mon. Not. R. Astron. Soc.* **483**, 4040–4059 (2019).
6. N. Lehner *et al*., *Astrophys. J.* **887**, 5 (2019).
7. D. Anglés-Alcázar *et al*., *Mon. Not. R. Astron. Soc.* **470**, 4698–4719 (2017).
8. R. J. J. Grand *et al*., *Mon. Not. R. Astron. Soc.* **490**, 4786–4803 (2019).
9. B. D. Oppenheimer *et al*., *Mon. Not. R. Astron. Soc.* **406**, 2325–2338 (2010).
10. R. Brennan *et al*., *Astrophys. J.* **860**, 14 (2018).
11. J. X. Prochaska, M. W. Lau, J. F. Hennawi, *Astrophys. J.* **796**, 140 (2014).
12. Z. Cai *et al*., *Astrophys. J.* **837**, 71 (2017).
13. F. Arrigoni Battaia *et al*., *Astron. Astrophys.* **620**, A202 (2018).
14. B. H. C. Emonts, Z. Cai, J. X. Prochaska, Q. Li, M. D. Lehnert, *Astrophys. J.* **887**, 86 (2019).
15. Materials and methods are available as supplementary materials.
16. Q. Li *et al*., *Astrophys. J.* **922**, 236 (2021).
17. Y. Yang, A. Zabludoff, K. Jahnke, R. Davé, *Astrophys. J.* **793**, 114 (2014).
18. F. Arrigoni Battaia *et al*., *Astrophys. J.* **804**, 26 (2015).
19. M. G. Allen, B. A. Groves, M. A. Dopita, R. S. Sutherland, L. J. Kewley, *Astrophys. J. Suppl. Ser.* **178**, 20–55 (2008).
20. C. C. Steidel *et al*., *Astrophys. J.* **795**, 165 (2014).
21. J. Tumlinson *et al*., *Science* **334**, 948–952 (2011).
22. C. B. Wotta *et al*., *Astrophys. J.* **872**, 81 (2019).
23. C. C. Steidel *et al*., *Astrophys. J.* **717**, 289–322 (2010).
24. N. Lehner, J. M. O'Meara, J. C. Howk, J. X. Prochaska, M. Fumagalli, *Astrophys. J.* **833**, 283 (2016).
25. A. C. Fabian, *Annu. Rev. Astron. Astrophys.* **50**, 455–489 (2012).
26. A. J. Richings, C.-A. Faucher-Giguère, *Mon. Not. R. Astron. Soc.* **478**, 3100–3119 (2018).
27. C. Circosta *et al*., *Astron. Astrophys.* **620**, A82 (2018).
28. K. R. Stewart *et al*., *Astrophys. J.* **738**, 39 (2011).
29. D. Nelson *et al*., *Comput. Astrophys. Cosmol.* **6**, 2 (2019).
30. T.-W. Lan, H. Mo, *Mon. Not. R. Astron. Soc.* **486**, 608–622 (2019).
31. A. Afruni, F. Fraternali, G. Pezzulli, *Astron. Astrophys.* **625**, A11 (2019).

ACKNOWLEDGMENTS

We thank the anonymous referees for insightful comments that substantially improved the manuscript. Z.C. and S.Z. thank S. Cantalupo, H.-W. Chen, M. Gronke, K. Kakiichi, C. Martin, M. Matuszewski, V. Springel, Y. Su, and H. Zhou for useful discussions. Z.C. and S.Z. also thank J. Zou and B. Wang for proofreading the manuscript. **Funding:** Z.C., S.Z., and Y.W. are supported by the National Key R&D Program of China (grant 2018YFA0404503), the National Science Foundation of China (grant 12073014), and Tsinghua University Initiative Scientific Research Program (grants 20223080023 and 20221080006). D.X. and S.W. are supported by the Tsinghua University Initiative Scientific Research Program (grant 2019Z07L02017). R.S. acknowledges support from Grants-in-Aid for Scientific Research (KAKENHI; 19K14766) through the Japan Society for the Promotion of Science (JSPS). Z.Z. is supported by NSF grant AST-2007499. A.Z. acknowledges funding from NSF grant AAG-1715609. **Author contributions:** Z.C. conceived of the project. S.Z. and Z.C. led the data reduction, analysis, and manuscript writing. Z.C., R.S., J.X.P., X.F., and Q.L. led the telescope proposals, with Z.C. as the principal investigator of most of the proposals. R.S. contributed to the observations and data reduction of the MORICS data. D.X., F.A.B., and R.C. contributed to the theoretical interpretation. L.D., J.W., and Y.A. contributed the analysis of the x-ray observations. R.C., Z.Z., A.Z., E.G.G.-M., M.L., Y.L., X.M., S.W., R.W., Y.W., and F.Y. contributed to the scientific interpretation and writing of the manuscript. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The observational data are available in the Keck/KCWI archive (https://koa.ipac.caltech.edu/cgi-bin/KOA/nph-KOAlogin) under program ID N052; the ALMA and VLA archive (https://data.nrao.edu/portal/#/) under project codes 17A-174 and 2018.1.00859.S; the HST archive (https://mast.stsci.edu/search/ui/#/hst) under proposal ID 14760; the Chandra data archive (https://cda.harvard.edu/chaser/) under observation ID 20357; and the Subaru archive (https://smoka.nao.ac.jp/fssearch.jsp) by selecting MOIRCS under "Instruments" and then entering "MAMMOTH-1" as the object name. The results of our spectral fitting are provided in tables S1 and S2, our measured redshifts in table S3, and the parameters of our streams model in table S4.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj9192
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 to S4
References (*32*–*78*)

Submitted 18 June 2021; accepted 31 March 2023
10.1126/science.abj9192

SOCIAL BELONGING

Where and with whom does a brief social-belonging intervention promote progress in college?

Gregory M. Walton[1]*, Mary C. Murphy[2], Christine Logel[3], David S. Yeager[4], J. Parker Goyer[1], Shannon T. Brady[5], Katherine T. U. Emerson[2], David Paunesku[1,6], Omid Fotuhi[7], Alison Blodorn[1]†, Kathryn L. Boucher[8]†, Evelyn R. Carter[9]†, Maithreyi Gopalan[10]†, Amy Henderson[1]†, Kathryn M. Kroeper[11]†, Lisel Alice Murdock-Perriera[12]†, Stephanie L. Reeves[13]†, Tsotso T. Ablorh[14]‡, Shahana Ansari[15]‡, Susie Chen[13]‡, Peter Fisher[16]‡, Manuel Galvan[17]‡, Madison Kawakami Gilbertson[18]‡, Chris S. Hulleman[19]‡, Joel M. Le Forestier[20]‡, Christopher Lok[21]‡, Katie Mathias[2]‡, Gregg A. Muragishi[22]‡, Melanie Netter[4]‡, Elise Ozier[2]‡, Eric N. Smith[23]‡, Dustin B. Thoman[24]‡, Heidi E. Williams[2]‡, Matthew O. Wilmot[25]‡, Cassie Hartzog[1], X. Alice Li[1], Natasha Krol[1]

A promising way to mitigate inequality is by addressing students' worries about belonging. But where and with whom is this social-belonging intervention effective? Here we report a team-science randomized controlled experiment with 26,911 students at 22 diverse institutions. Results showed that the social-belonging intervention, administered online before college (in under 30 minutes), increased the rate at which students completed the first year as full-time students, especially among students in groups that had historically progressed at lower rates. The college context also mattered: The intervention was effective only when students' groups were afforded opportunities to belong. This study develops methods for understanding how student identities and contexts interact with interventions. It also shows that a low-cost, scalable intervention generalizes its effects to 749 4-year institutions in the United States.

Upward mobility and economic growth depend substantially on postsecondary attainment (*1*, *2*). Yet the likelihood of earning a college degree is highly unequal across racial-ethnic and socioeconomic groups, even among students with similar preparation (*3*, *4*). A challenge to stemming intergroup inequality is that programs to promote college persistence have heterogeneous effects—they work differently for different people in different contexts. A critical goal is to systematically explain and anticipate this heterogeneity (*5*, *6*) so that programs can be designed to respond to the broad diversity of higher-education students and institutions (*7*).

Here we report the findings of a large team-science partnership involving a double-blind, randomized controlled trial conducted with two cohorts of students at 22 institutions (N = 26,911 students). This partnership between experimental psychologists and university administrators, called the College Transition Collaborative (CTC), was designed specifically to understand variability across contexts in the effects of a promising and scalable approach to reducing postsecondary inequality. We focus on a core outcome on the path to graduation: the likelihood that students complete their first year of college enrolled full-time. Participating institutions were highly diverse and selected for their potential to advance an understanding of heterogeneity. We applied advanced analytic methods to this dataset, such as preregistration of complex analyses, and a conservative, Bayesian multilevel machine-learning analysis method. Ultimately, the CTC trial makes two primary contributions. First, it provides a systematic examination of how different groups of students in different postsecondary contexts do or do not benefit from a social-psychological intervention. In doing so it reveals insights into how psychological and contextual factors come together to affect intergroup inequality and how this inequality can be remedied. Second, it presents a framework for understanding how student identities interact with contexts and, thus, the principled evaluation of heterogeneous intervention effects.

The intervention examined by the CTC was the social-belonging intervention (*8*). This is a brief (10 to 30 min) reading-and-writing act that can be self-administered by college students over the internet. It features three elements: (i) results of a survey of older students, showing that everyday worries about belonging—such as feeling homesick, struggling academically, or having difficulty interacting with professors—are normal in the transition to college and can lessen with time (*9*); (ii) carefully curated stories from older students describing these worries and how they improved for them; and (iii) an opportunity to reflect on these stories in writing to help future students as they come to college, including how concerns about belonging are normal and typically improve with time ("saying-is-believing" exercises).

The social-belonging intervention has been tested in more than a dozen rigorous randomized controlled trials. Benefits are typically observed with regard to core academic outcomes for racial-ethnic minority and first-generation college students (*8*). In one study published in *Science*, the intervention delivered in the first year at a selective, predominantly white university raised the grade point average (GPA) of Black students relative to multiple randomized control groups, reducing by half the racialized achievement gap over the next 3 years (*10*). Further, a long-term follow-up revealed substantial gains in participants' career success and life satisfaction 7 to 11 years later. These gains were mediated by greater development of mentor relationships in college, consistent with the theory that addressing belonging concerns can unlock relationships of enduring value (*11*). Other trials have found increased progress (e.g., first-year full-time completion rates) (*12*), including among students attending broad-access institutions (*13*). In male-dominated STEM (science, technology, engineering, and mathematics) fields, the intervention has improved experiences and raised grades among women (*14*, *15*). Some rigorous trials have not shown significant effects, however, demonstrating heterogeneity (*16*).

One reason the social-belonging intervention can have beneficial effects is because higher-education outcomes are subject to psychological threats rooted in group identities. These threats evoke concerns about belonging (*17*, *18*). These concerns can become acute when students are underrepresented on campus; when their group is struggling or faces stereotypes

[1]Department of Psychology, Stanford University, Stanford, CA, USA. [2]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA. [3]Department of Social Development Studies, Renison University College, University of Waterloo, Waterloo, ON, Canada. [4]Department of Psychology, University of Texas at Austin, Austin, TX, USA. [5]Department of Psychology, Wake Forest University, Winston-Salem, NC, USA. [6]The Project for Education Research that Scales (PERTS), San Francisco, CA, USA. [7]Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA. [8]Department of Psychological Sciences, University of Indianapolis, Indianapolis, IN, USA. [9]Paradigm, Palo Alto, CA, USA. [10]Department of Education Policy Studies, The Pennsylvania State University, State College, PA, USA. [11]Department of Psychology, Sacred Heart University, Fairfield, CT, USA. [12]School of Education, Sonoma State University, Rohnert Park, CA, USA. [13]WGU Labs, Salt Lake City, UT, USA. [14]Department of Clinical Psychology, University of Massachusetts Boston, Boston, MA, USA. [15]Department of Psychology, University of Hawai'i at Mānoa, Honolulu, HI, USA. [16]Department of Psychology, University of California, Los Angeles, CA, USA. [17]Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill, NC, USA. [18]School of Psychology, Fuller Theological Seminary, Pasadena, CA, USA. [19]School of Education and Human Development, University of Virginia, Charlottesville, VA, USA. [20]Department of Psychology, University of Toronto, Toronto, ON, Canada. [21]Department of Psychology, University of Waterloo, Waterloo, ON, Canada. [22]Department of Psychology, University of Washington, Seattle, WA, USA. [23]Population Research Center, University of Texas at Austin, Austin, TX, USA. [24]Department of Psychology, San Diego State University, San Diego, CA, USA. [25]Department of Psychology, The Ohio State University, Columbus, OH, USA.
***Corresponding author. Email: gwalton@stanford.edu** †These authors contributed equally to this work. ‡These authors contributed equally to this work.

that allege that it is less able or less deserving of educational opportunities than others; when they experience a campus culture that does not match their cultural values or curricula that do not value their experiences; or when they experience racism, among other factors (*19*–*23*). In turn, belonging concerns predict less persistence in college above and beyond other factors (*12*, *24*). In part, this is because these concerns function as a lens through which students make sense of adversities in college. When students are uncertain of their belonging, even common challenges experienced by students in all social groups (e.g., feeling homesick, academic struggles) can appear as evidence that people like them do not belong in college in general (*19*). However, when an intervention represents everyday challenges as normal and as improving over time, students are better able to sustain a sense of belonging on campus in the face of everyday adversities (*10*, *13*, *15*, *19*). The social-belonging intervention uses diverse stories from diverse students to convey this message as a common truth, with variation [see supplementary materials (SM)]. In representing worries about belonging as initially normal in the transition to college and as improving with time, the intervention does not deny that students can also have different experiences along group lines. In at least some circumstances, this can free students to build critical relationships, improving academic outcomes and life success long into the future (*11*, *12*, *15*).

Sociological and psychological theories do not lead to an expectation that members of a given identity group—for example, Latina students—should experience the same concerns about belonging in all contexts, nor that they will have the same opportunities to belong. A core theoretical tenet is that group identities do not have inherent meaning but rather meaning that arises in context (*25*). Experimental studies have underappreciated this fact, however, treating racial, ethnic, and socioeconomic groups as invariant in statistical analyses (*26*). For instance, even as Steele introduced stereotype threat as a "situational predicament" (*18*), both laboratory and field-experimental research on identity threat (e.g., social-belonging, values-affirmation, and difference-education interventions) have treated specific racial and gender groups as canonical and then used them as a static grouping variable in analyses, with rare exceptions (*14*, *15*, *27*).

This practice obscures variation within groups across contexts. One source of this variation involves sorting mechanisms. For instance, scholars emphasize the differences in racial socialization and, therefore, identity threat that can emerge between Black Americans who are descended from people who were enslaved in the US, relative to Black Americans who are the children of more recent immigrants (*28*). Postsecondary admissions and recruitment practices, in turn, may be influenced by this variation (*4*), altering the meaning of racial-ethnic identities in college contexts.

Institutions also serve different students differently, including in ways that shape the opportunities students have to belong (*23*). Such belonging affordances can vary as a function of classroom practices (*22*, *29*), institutional messaging (*30*), and campus cultures (*20*). One belonging affordance is the opportunity to participate in cultural and affinity groups and to develop "pride" for one's racial or ethnic group (*21*). Yet opportunities to do so, in courses, student groups, and residential communities, vary. Another belonging affordance is the opportunity for high-quality intergroup interactions, such as to develop friendships with majority-group members on campus in residences, courses, and elsewhere, which can support belonging and reduce identity threat (*22*, *31*–*33*). These, too, vary across contexts. Among other qualities, schools vary in the degree of social bias present. If a student of color seeks out mentorship, for instance, will a professor respond enthusiastically or not (*34*)? In general, the complexity of identities-in-context raises the need for both analytic methods and theoretical frameworks that can detect and understand differences in both identity threats and belonging affordances across contexts.

One contribution of the CTC trial is to overcome the limitations of the dominant static-identity approach by introducing the concept of "local-identity groups." Given our focus on social-identity threat, here we define local-identity groups as students of a given race-ethnicity, with a given first-generation status, in a given college, in a given cohort. Because some students within each of the 374 local-identity groups in our sample were randomly assigned to the treatment and some to the control, our models estimate 374 different treatment effects; we can then examine how these effects vary by qualities of these groups (table S6). This approach offers far greater statistical power and far more theoretical nuance than comparing a few canonical groups (e.g., first-generation versus continuing-generation students). It is also preferable to the common practice of clustering groups broadly defined as "disadvantaged" (e.g., underrepresented minority and first-generation students) (*12*, *13*, *22*, *29*), which can imply that disadvantage is inherent to these groups while not specifying the factors that cause a group to experience disadvantage within a specific context. Our institutional sampling plan was also designed to support well-powered, replicable, and generalizable inferences about how different local-identity groups respond to the intervention, including by partnering with highly diverse institutions (e.g., admissions rates ranged from 6 to 90%; table S4).

We used two factors to characterize variation in local-identity groups and examine contextual heterogeneity in treatment effects. At a conceptual level, these are (i) the degree to which students experience vulnerability to worries about belonging that could undermine persistence and (ii) the opportunities students have to come to belong. We expected the largest effects in contexts that prompt identity-threatening questions ("Do 'people like me' belong here?") but that make belonging possible, or afford belonging, for students in a given local-identity group. Only in these more supportive contexts can the positive answer the treatment offers to questions about belonging ("People like me can come to belong here") be "locally true" and, thus, legitimate, sustained, and useful in guiding students' interpretation of everyday experiences and continued engagement on campus (Fig. 1).

Our theorizing grew out of the emerging science on treatment heterogeneity. Past research found that growth-mindset interventions produce greater academic gains among academically struggling students (those vulnerable to asking "Am I dumb?") in school environments that afford a growth mindset (opportunity), such as schools with positive peer norms around challenge-seeking (*35*) and classrooms where teachers endorse a growth mindset (*36*).

In this study, we examined vulnerability to worries about belonging that could undermine academic progress at the group level. To do so, we assessed the local-identity group's historic level of attainment along the primary outcome: first-year full-time completion rates, in preexperimental cohorts (*27*) (table S7). This approach reflects several streams of reasoning: seeing fewer members of one's group persist could provide a basis for belonging concerns; this circumstance could reflect the presence of other factors in the college environment that undermine belonging and persistence; and, finally, students in more poorly performing groups have greater room for improvement (see SM).

To operationalize the opportunity to belong ("belonging affordances"), we assessed survey responses to measures of belonging in the spring term in the control condition only. Absent treatment, to what extent do students in each local-identity group experience a feeling of belonging toward the end of their first year? In prioritizing students' own experiences, we obtain a direct measure of the extent to which each local-identity group had an opportunity to come to belong at their college.

Our theory brings these factors together. It predicts greater treatment effects for local-identity groups that have historically performed

A

Term	Conceptual Definition	Operationalization
Local-Identity Group	Students with a shared social identity along dimensions of relevance.	Students of a given race-ethnicity, with a given first-generation status, in a given college, in a given cohort.
Vulnerability to Worries About Belonging That Could Undermine Persistence	Does the post-secondary context evoke the question, "Do people 'like me' belong here?" in ways that can undermine persistence?	Students in local-identity groups with lower rates of first-year full-time completion in prior cohorts.
Belonging Affordances	Is the hopeful message about belonging provided by the intervention legitimate in the post-secondary context and one a student can maintain and act on to succeed?	Students in local-identity groups report, absent treatment, at least a moderate level of belonging toward the end of the first year of college.

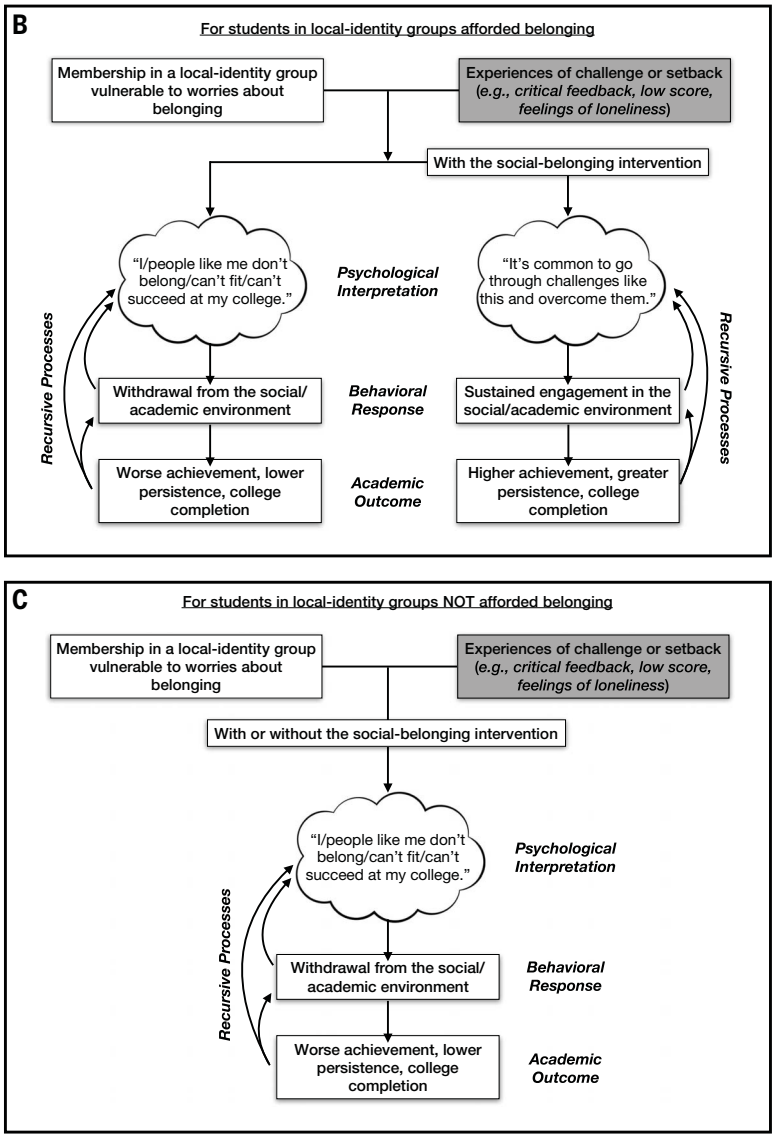


Fig. 1. Key concepts and theory of change. (**A**) Key concepts. (**B**) Theory of change for students in local-identity groups afforded belonging [adapted from figure 1 in (*12*)]. (**C**) Theory of change for students in local-identity groups not afforded belonging. Past research provides primary evidence for the theory of change depicted in (B) (*8*, *10*–*15*, *19*). The present research introduces the concept of local-identity groups and the moderating role of belonging affordances, comparing (B) and (C).

poorly yet realize sufficient levels of belonging in the spring term. If so, the intervention would help groups at risk of poor performance take advantage of supportive environments. Competing hypotheses are also possible. For instance, could the belonging intervention compensate for a hostile climate? This is a legitimate hypothesis. Yet the intervention does not create institutional structures that support belonging; all it does is offer students a positive way to make sense of common challenges to belonging that arise in college. It helps students navigate college to develop their belonging, but college still has to be navigable. From this theoretical perspective, only in contexts that make it possible for a group to overcome belonging worries will students be able to use this idea effectively (see SM).

In a study such as this, what kind of effects would be meaningful? Typical benchmarks for understanding effect sizes in laboratory research are inappropriate in field settings focused on changing educational outcomes such as college achievement (*37*). Instead, it is important to calibrate against other potential reforms. Past research finds percentage-point gains of 0.3 to 3.4 points in the probability of college attendance and college persistence from such investments as improving teacher quality in secondary school and increased financial aid (see SM). Relative to the cost of changing instruction or financing scholarships, the investments needed to implement the social-belonging intervention are minimal, particularly for a brief online form completed before matriculation (it is freely available online: https://www.perts.net/orientation/cb). Indeed, the present implementation is especially brief: In the key writing task, participating students wrote for just 7 min 36 s (median), which is considerably shorter than prior smaller trials (tables S1 to S3). If an experience this brief can produce even a 1 to 2 percentage-point gain in the rate at which students complete their first year of college enrolled full-time, it would be remarkable from both a theoretical and a practical perspective.

Results

A total of 102,792 students were invited to participate. Primary analyses were conducted with the intent-to-treat sample, 26,911 students across 22 postsecondary institutions. A 23rd institution was excluded, as it was not in the United States and used a different design (see Table 1 and tables S4 and S5 for student and institutional sample characteristics; see table S9 for baseline equivalence). The trial also included a third condition, which adapted belonging materials for each campus. That condition is not reported here, given our focus on contextual rather than material heterogeneity. For the CONSORT diagram, see Fig. 2.

Preliminary analyses

Manipulation check

A manipulation check was included to ensure that the intervention achieved its intended initial impact, which it did. The belonging intervention led students to anticipate greater growth in belonging over time compared with the control group $b = 0.289$ [95% confidence interval (CI): 0.261, 0.318], SE = 0.014, $t = 20.13$,

Table 1. Intent-to-treat sample by first-generation status, race-ethnicity, advantage status, and gender. "Advantaged" students are white and Asian continuing-generation students. "Disadvantaged" students are all other race-ethnicity × first-generation status combinations.

Classification	First-generation status	Race-ethnicity		*N*	Percent of total sample
By first-generation status and race-ethnicity	First generation	Of Hispanic/Latinx origin (of any race)		3,934	15%
		Not of Hispanic/Latinx origin	Black/African/African American	1,121	4%
			White/European American	3,001	11%
			Asian/Asian American	640	2%
			Native American/Native Hawaiian/Other Pacific Islander	54	0.2%
			Other	737	3%
			Total first generation	**9,487**	**35%**
	Continuing generation	Of Hispanic/Latinx origin (of any race)		1,487	6%
		Not of Hispanic/Latinx origin	Black/African/African American	1,166	4%
			White/European American	10,832	40%
			Asian/Asian American	1,925	7%
			Native American/Native Hawaiian/Other Pacific Islander	73	0.27%
			Other	1,941	7%
			Total continuing generation	**17,424**	**65%**
By canonical "advantage" status		Advantaged		12,757	47%
		Disadvantaged		14,154	53%
By gender		Male		10,754	40%
		Female		15,743	58.5%
		I prefer another term/transgender		410	1.5%
		Not provided		4	0.01%
		Grand total		**26,911**	**100%**

$P < 0.001$, as expected. Although there was some heterogeneity in this effect across local-identity groups (see SM), it was not explained by the moderators of interest: the treatment effect did not interact with local-identity groups' historic achievement level or belonging affordance, Ps > 0.187. Thus, the primary results here are not attributable to differences in how persuasive the intervention was for a given local-identity group.

Heterogeneity of the treatment effect on first-year full-time completion rates

There was meaningful heterogeneity in treatment effects on the primary outcome, first-year full-time completion rates. The standard deviation of this effect across local-identity groups was 2.4 percentage points, $b = 0.024$ [0.014, 0.042], SE = 0.007, $t = 3.60$, $P < 0.001$, $Q = 215.87$, df = 49, $P < 0.001$. Thus, we proceeded to test our primary multilevel interaction-effect hypothesis.

Independence of moderators

The local-identity group belonging affordance and historic achievement level reflect separate factors [i.e., they were not significantly correlated at the local-identity group level (see SM)]. Thus, the historic circumstance that, we theorize, can lead to vulnerability to worries about belonging was largely independent of the opportunity students had to belong in their college in their cohort.

Primary analyses

Three-way interaction

The preregistered linear mixed-effects model showed that the effect of the belonging intervention on first-year full-time enrollment rates depended on the local-identity group's historic achievement and belonging affordances, three-way interaction, $b = 0.013$ [0.004, 0.023], SE = 0.005, $t = 2.81$, $P = 0.005$ (see model details in the SM).

Bayesian multilevel analyses

To interpret and visualize the three-way interaction and to guide follow-up hypothesis tests, we estimated a flexible Bayesian multilevel model using machine learning tools called Bayesian causal forest (BCF). An advantage of BCF is that we can make decisions about subgroups and functional forms that are not possible in a conventional regression by using an algorithmic decision rule to examine the posterior distribution of treatment effects across nonparametric effects of moderators (*38*). This helps avoid the undue influence of researcher degrees of freedom (see Fig. 2 caption and SM).

Using this hands-off, data-driven approach, we found that, among local-identity groups low in afforded belonging [<−0.5 SD, number of groups (k) = 135, number of students (n) = 4078], there was no discernible treatment effect overall and no moderation by group historic achievement level. By contrast, among local-identity groups medium to high in afforded belonging (k = 239, n = 22,833), signifying sufficient belonging affordances, we found a linear moderation effect by group historic achievement, with larger effects for lower-achieving local-identity groups. The BCF model's results are plotted in Fig. 3.

Two-way interactions and simple effects

Guided by the BCF model's decision rules for subgroups, we conducted follow-up hypothesis tests of our preregistered linear mixed-effects model among groups that were medium to high in afforded belonging versus low in afforded belonging. Within the medium-to-high belonging category (85% of students and 64% of local-identity groups), we found an overall conditional average treatment effect (CATE) of 1.1 percentage points, $b = 0.011$ [0.0005, 0.002], SE = 0.005, $t = 2.06$, $P = 0.040$, that was qualified by a condition × historic achievement interaction, $b = 0.013$ [0.002, 0.024], SE = 0.006, $t = 2.34$, $P = 0.020$.

We decomposed this two-way interaction into subgroup effects (tables S10 and S11). Among students whose local-identity groups had lower historic achievement (k = 61, n = 5212), in which just 49% (SD = 8%) of students had historically maintained full-time enrollment through the first year, we observed the largest CATE, 2.0 percentage points, b = 0.020 [0.003, 0.038], SE = 0.009, t = 2.27, P = 0.023. Among students whose local-identity groups had medium historic achievement (k = 120, n = 12,460; 77% historic full-time rates, SD = 10%), the CATE was 1.3 percentage points, b = 0.013 [0.002, 0.024], SE = 0.006, t = 2.27, P = 0.023. Among students from groups with high historic achievement (k = 58, n = 5161; 96% full-time rates, SD = 2%), the CATE was essentially zero, b = −0.003, P = 0.681. Among students whose local-identity groups were not afforded opportunities for belonging, the simple effects of treatment at low, medium, and high levels of historic achievement were all nonsignificant, Ps > 0.112.

Robustness

Multiple tests confirmed the robustness of the primary results. First, similar results obtained using a static societal disadvantage classifier in lieu of historic achievement, despite some differences in two-way interactions (see SM). Notably, these analyses had a higher standard error for the three-way interaction term (disadvantaged status: SE = 0.010; historic group-level achievement SE = 0.005), consistent with our theory that treating social disadvantage as a static grouping variable masks contextual variability. Second, results were similar when including poststratification weights and when using treatment-on-treated analyses (table S10).

Generalizability

What do these effect sizes imply for the broader population of colleges? While the sample of schools is a convenience one, it was recruited to be diverse along multiple dimensions (e.g., geography, selectivity, size). Indeed, it reasonably generalizes to 749 4-year nonprofit degree-granting colleges and universities in the United States (Tipton b index = 0.887) (see SM for full specification of the generalizability sample). In 2015 and 2016, these 749 institutions welcomed an average of 1,019,790 first-time full-time–degree seeking undergraduates annually. When applying poststratification weights to force the composition of our sample to resemble the population of inference, the positive treatment effect among students in local-identity groups medium to high in belonging affordances (85% of our sample) was similar, 1.4 percentage points (as compared with 1.1 percentage points, reported above, for the unweighted sample; table S10). Thus, the results imply that if the full set of institutions

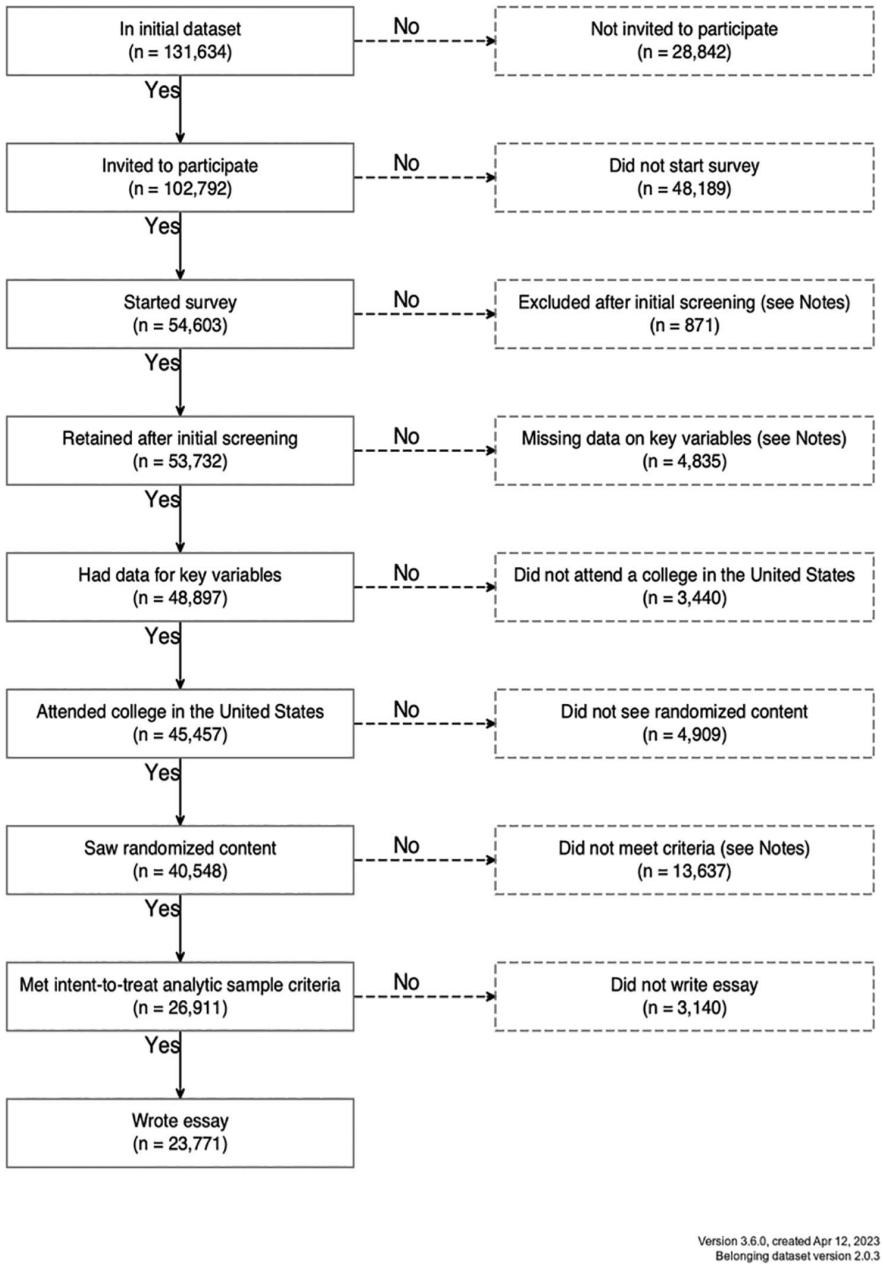


Fig. 2. CONSORT diagram. "In initial dataset" is the number of students who were either in the intervention roster, started the intervention (t1) survey, or had academic data provided by their college. "Invited to participate" is the number of students who were either in the intervention roster or started the intervention (t1) survey. "Started survey" is the number of students who started the intervention (t1) survey. "Retained after initial screening" is the number of first-time students who met all prescreening criteria. (Exclusion details: missing condition: 612; multiple conditions: 185; invitation issue: 74. Exclusions were processed sequentially as listed.) "Had data for key variables" is the number of students who had data for variables used in analytic models. (Exclusion details: missing cohort/college name: 1836; missing all academic outcome data: 167; missing race-ethnicity, first-generation status, or local-identity group variable: 2832. Exclusions were processed sequentially as listed.) "Attended college in the United States" is the number of students who attended a US college (one Canadian university participated in the study but used a different design and so will be reported separately). "Saw randomized content" is the number of students who proceeded through the survey far enough to see randomized content, meeting the intent-to-treat sample criterion. "Met intent-to-treat analytic sample criteria" is the number of students who met the intent-to-treat analytic sample criteria. (Exclusion details: in customized treatment condition, with school-specific content: 13,522; local-identity group did not have at least one participant in each condition: 88; missing year 1 full-time completion variable: 25; missing study moderators: 2. Exclusions were processed sequentially as listed.) "Wrote essay" is the number of students who wrote an essay in either condition of any length (i.e., met the treatment-on-treated sample criterion). Version 3.6.0, created 12 April 2023. Belonging dataset version 2.0.3.

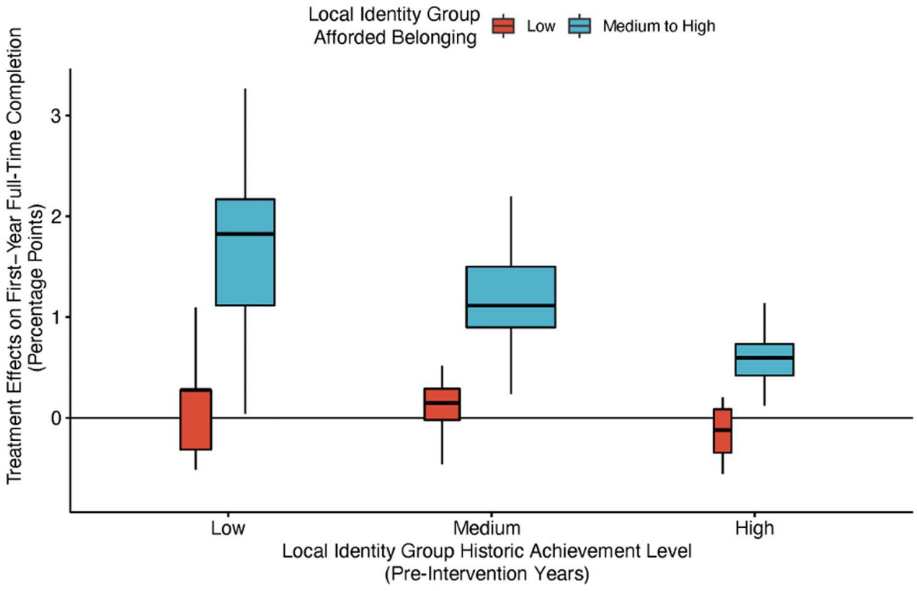


Fig. 3. Conditional average treatment effects (CATEs) on the probability of completing the first year enrolled full-time in the intent-to-treat sample (*N* = 26,911, *k* = 374 at 22 colleges and universities), from the Bayesian causal forest analysis. Each of the 374 local-identity groups had an estimated treatment effect (i.e., random slope, in the multilevel model, with two student-level covariates: gender and standardized test scores). This plot depicts the interquartile ranges (IQR) of the distributions of local-identity groups' average treatment effects, at different levels of the moderators. The dark line is the median (which will be similar, but not identical, to the mean treatment effects listed for the CATEs in the text), and the whiskers are the 95% percentile ranges. The width of each box is proportional to the number of local-identity groups in that category. The figure shows positive treatment effects for students in groups medium-to-high in afforded belonging (*N* = 22,833, *k* = 239), especially groups with lower historic achievement along the outcome measure, first-year full-time completion rate. The thinner, red boxes show null effects: Groups not afforded belonging did not tend to benefit from the treatment. To select the belonging cut-point used in this figure, we fit a random forest model to the posterior distribution of treatment effects and allowed the algorithm to choose the belonging cut-point that best minimized mean-squared error. That value was −0.5 SD for afforded belonging (36th percentile among local-identity groups). The cut-points for the three achievement groups were set at the 25th and 75th percentiles (Low: 0th to 24th historic percentile; Medium: 25th to 74th percentile; High: 75th to 100th percentile). Generalizability sample: 1.02 million new students annually entering 749 US 4-year colleges and universities. For mean completion rates by condition, see table S11.

in the generalizability sample implemented the 10- to 30-min online belonging module with their incoming students they would see an additional 12,136 students (1,019,790 × 0.85 × 0.014) complete the first year enrolled full-time each year. Moreover, these gains would be concentrated in groups with lower historic achievement, reducing inequality. Insofar as institutions take steps to convey the belonging message in more effective ways or support greater belonging affordances, they may see larger gains.

Discussion

The present results confirm the potential of even a brief social-belonging intervention to improve students' progress in college, especially students from groups that have historically achieved at lower rates. Yet this potential is not invariant. It depends on the opportunity students have to belong in their college contexts.

These results are consistent with the emerging theory that positive change requires planting "high-quality seeds" (hopeful answers to threatening questions) in "fertile soil" (contexts in which these answers are true) (*39*). Above a certain threshold, college contexts make belonging possible for diverse student groups—they afford belonging. The belonging intervention helped students realize the benefit of this affordance in terms of improved persistence. Conversely, when affordances for belonging were inadequate, which was the case for 15% of our student sample, the hopeful way of making sense of belonging offered by the intervention did not increase persistence.

For colleges and universities striving to better support students, the findings point to the importance of twin goals: (i) simultaneously conveying that belonging concerns are common in the transition to college for students from all backgrounds and improve with time and (ii) acting to ensure that this message is legitimate and reflected in opportunities to belong in the lived experience of all groups (*8*, *23*, *40*). To convey this message, colleges may complement brief online modules with broader efforts to create a culture on campus that normalizes challenges and worries about belonging and emphasizes opportunities for growth, whether through welcome addresses, residential programming, pedagogy, or other mutually reinforcing institutional communications and conversations. To ensure that this message is legitimate for all students, colleges may increase efforts to ensure an adequate representation of both students and faculty from diverse groups on campus (*41*); support ethnic-themed clubs, events, activities, and coursework that cultivate an understanding of and pride in group identities (*21*, *42*); prioritize pedagogy that emphasizes growth rather than identifying the ostensibly smart people (*29*, *43*); and create residential (*32*) and classroom communities that offer opportunities for positive intra- and intergroup interactions, including by establishing and communicating norms that encourage supportive relationships and discourage bias (*22*). Psychological and structural reforms are not substitutes for one another but can work together to promote positive change.

Belonging concerns are primary for many students as they enter college. Moreover, the history and reality of racism and social-class exclusion in higher education means that everyday challenges such as feeling excluded or having a hard time finding a lab partner can take on a racialized or social class–laden meaning for specific identity groups: "People like me don't belong here." Because such fixed, global attributions can become self-confirming, it is important to forestall them. Yet even as the social-belonging intervention focuses on everyday adversities experienced by students from all backgrounds at one time or another, it does not deny that students can also experience racial bias, stereotyping, and discrimination; nor does it deny the pride that students may feel in their racial-ethnic group. It is both true that students have similar challenges and experiences (e.g., worries about belonging) in the transition to college and true that students in different identity groups experience distinct challenges. Indeed, some adaptations of the belonging intervention for specific populations specifically raise and normalize group differences in experience (*13*, *15*). More broadly, it is important that messaging does not suppress the racialized and social class–informed experience of higher education. Surfacing these experiences can also be powerful (*20*, *21*, *23*, *44*).

Correspondingly, for researchers, these findings point to a new generation of work to

understand belonging affordances: What is needed to make belonging possible, for whom, and where? This may include the structured opportunities students have to interact with and build relationships with each other and with faculty (e.g., in residential life, first-year classes, student groups) as well as the social and academic culture on campus, including the degree to which instructors express their beliefs and engage in practices that affirm students' distinctive identities and strengths (*42*) and promote a growth mindset (*12*, *16*, *29*, *43*), among other factors (*17*, *45*).

As intervention science matures and we move to increasingly large-scale studies, identifying contextual boundary conditions around promising practices will increasingly come to the fore (*7*). For research on intergroup inequality, it is important to study variation in identity-group experiences, for we cannot assume that a given identity group has the same meaning or the same opportunities in different contexts. The local-identity group methodology developed here allows quantitative social scientists to relax the assumption that identity-group experiences are static and, correspondingly, to develop theory about how and why these experiences vary in ways that, for instance, create vulnerabilities to belonging concerns and belonging affordances. In doing so, we can learn what aspects of school contexts we should change to better realize their promise for all students.

REFERENCES AND NOTES

1. R. Chetty, J. N. Friedman, E. Saez, N. Turner, D. Yagan, "Mobility report cards: The role of colleges in intergenerational mobility" (NBER Working Paper 23618, National Bureau of Economic Research, 2017); http://www.nber.org/papers/w23618.
2. J. Rothwell, "What colleges do for local economies: A direct measure based on consumption," Brookings Institution, 17 November 2015; https://www.brookings.edu/research/what-colleges-do-for-local-economies-a-direct-measure-based-on-consumption/.
3. E. Aucejo, Z. Tobin, "Assessing racial disparities in postsecondary education," Federal Reserve Bank of Boston, 5 October 2021; https://www.bostonfed.org/-/media/Documents/events/2021/racial-disparities-in-todays-economy/Assessing-Racial-Disparities-in-Postsecondary-Education.pdf?la=en.
4. P. Tough, *The Inequality Machine: How College Divides Us* (Mariner Books, 2021).
5. J. A. Berlin, *Am. J. Epidemiol.* **142**, 383–387 (1995).
6. B. B. McShane, J. L. Tackett, U. Böckenholt, A. Gelman, *Am. Stat.* **73** (suppl. 1), 99–105 (2019).
7. C. J. Bryan, E. Tipton, D. S. Yeager, *Nat. Hum. Behav.* **5**, 980–989 (2021).
8. G. M. Walton, S. T. Brady, in *Handbook of Wise Interventions: How Social Psychology Can Help People Change*, G. M. Walton, A. J. Crum, Eds. (The Guilford Press, 2021), pp. 36–62.
9. D. A. Prentice, D. T. Miller, *Adv. Exp. Soc. Psychol.* **28**, 161–209 (1996).
10. G. M. Walton, G. L. Cohen, *Science* **331**, 1447–1451 (2011).
11. S. T. Brady, G. L. Cohen, S. N. Jarvis, G. M. Walton, *Sci. Adv.* **6**, eaay3689 (2020).
12. D. S. Yeager *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3341–E3348 (2016).
13. M. C. Murphy *et al*., *Sci. Adv.* **6**, eaba4677 (2020).
14. K. R. Binning *et al*., *Psychol. Sci.* **31**, 1059–1070 (2020).
15. G. M. Walton, C. Logel, J. M. Peach, S. J. Spencer, M. P. Zanna, *J. Educ. Psychol.* **107**, 468–485 (2015).
16. M. Broda *et al*., *J. Res. Educ. Eff.* **11**, 317–338 (2018).
17. J. M. Harackiewicz, S. J. Priniski, *Annu. Rev. Psychol.* **69**, 409–435 (2018).
18. C. M. Steele, *Am. Psychol.* **52**, 613–629 (1997).
19. G. M. Walton, G. L. Cohen, *J. Pers. Soc. Psychol.* **92**, 82–96 (2007).
20. N. M. Stephens, S. A. Fryberg, H. R. Markus, C. S. Johnson, R. Covarrubias, *J. Pers. Soc. Psychol.* **102**, 1178–1197 (2012).
21. T. N. Brannon, A. Lin, *Am. Psychol.* **76**, 488–501 (2021).
22. S. Murrar, M. R. Campbell, M. Brauer, *Nat. Hum. Behav.* **4**, 889–897 (2020).
23. D. L. Gray, E. C. Hope, J. S. Matthews, *Educ. Psychol.* **53**, 97–113 (2018).
24. M. Gopalan, S. T. Brady, *Educ. Res.* **49**, 134–137 (2020).
25. E. Goffman, *Stigma: Notes on the Management of Spoiled Identity* (Prentice-Hall, 1963).
26. M. Cikara, J. E. Martinez, N. A. Lewis Jr., *Nat. Rev. Psychol.* **1**, 537–549 (2022).
27. G. D. Borman, J. Grigg, C. S. Rozek, P. Hanselman, N. A. Dewey, *Psychol. Sci.* **29**, 1773–1784 (2018).
28. K. Deaux *et al*., *Soc. Psychol. Q.* **70**, 384–404 (2007).
29. E. A. Canning, K. Muenks, D. J. Green, M. C. Murphy, *Sci. Adv.* **5**, eaau4734 (2019).
30. J. G. Starck, S. Sinclair, J. N. Shelton, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2013833118 (2021).
31. A. M. Locks, S. Hurtado, N. A. Bowman, L. Oseguera, *Rev. Higher Educ.* **31**, 257–285 (2008).
32. N. J. Shook, R. Clay, *J. Exp. Soc. Psychol.* **48**, 1168–1172 (2012).
33. R. M. Carey, N. M. Stephens, S. S. M. Townsend, M. G. Hamedani, *J. Pers. Soc. Psychol.* **123**, 889–908 (2022).
34. K. L. Milkman, M. Akinola, D. Chugh, *J. Appl. Psychol.* **100**, 1678–1712 (2015).
35. D. S. Yeager *et al*., *Nature* **573**, 364–369 (2019).
36. D. S. Yeager *et al*., *Psychol. Sci.* **33**, 18–32 (2022).
37. M. A. Kraft, *Educ. Res.* **49**, 241–253 (2020).
38. S. Woody, C. M. Carvalho, J. S. Murray, *J. Comput. Graph. Stat.* **30**, 144–161 (2021).
39. G. M. Walton, D. S. Yeager, *Curr. Dir. Psychol. Sci.* **29**, 219–226 (2020).
40. N. M. Stephens, H. R. Markus, S. A. Fryberg, *Psychol. Rev.* **119**, 723–744 (2012).
41. N. A. Bowman, N. Denson, *J. Higher Educ.* **93**, 399–423 (2022).
42. D. M. Silverman, R. J. Rosario, I. A. Hernandez, M. Destin, *Pers. Soc. Psychol. Rev.* 10.1177/10888683221145243 (2023).
43. K. M. Kroeper, A. C. Fried, M. C. Murphy, *Soc. Psychol. Educ.* **25**, 371–398 (2022).
44. N. M. Stephens, M. G. Hamedani, M. Destin, *Psychol. Sci.* **25**, 943–953 (2014).
45. E. N. Smith, D. S. Yeager, C. S. Dweck, G. M. Walton, *Educ. Psychol. Rev.* **34**, 2197–2219 (2022).
46. J. P. Goyer *et al*., Does a Social-belonging Intervention Reduce the Effects of Social Identity Threat on Group-based Inequalities in Academic Outcomes? Results from a Large, Multi-site, Randomized Trial, Open Science Framework (2020); https://doi.org/10.17605/OSF.IO/ZT653.

ACKNOWLEDGMENTS

This manuscript uses data from the Social-Belonging Trial conducted by the College Transition Collaborative [https://www.collegetransitioncollaborative.org/; principal investigators (PIs): C. Logel, M. Murphy, G. Walton, and D. Yeager]. Methods and data systems were created by the Project for Education Research That Scales (PERTS, https://www.perts.net/; PI: D. Paunesku). We are grateful for the collaboration and institutional data provided by partner colleges and universities and individual staff collaborators: Albion College (D. Dunham, B. Van Eck); Allegheny College (R. Holmgren, R. Pickering); Bowling Green State University (A. Alt, K. Brooks, J. Matuga, D.-L. Stewart); California State University, Dominguez Hills (B. Driscoll); California State University, Northridge (J. Oh, R. Feyk-Miney, B. Huber); The College of Wooster (E. Falduto, G. Holmes); Cornell University (E. Stephenson, M. Clarkberg); Dartmouth College (J. T. Davis, B. Tillotson); DePauw University (J. Miranda, K. Hall, C. Setchell); the Great Lakes College Association (R. Detweiler); Hope College (M. Hudgins, M. Inman, L. Smith); Indiana State University (J. Powers, L. Ferguson); Indiana University (M. Payne, J. Ouimet, J. Teague, L. Shepard, D. Anderson); Kalamazoo College (A. Dueweke, T. Webb); Lewis & Clark College (J. Detweiler-Bedell, B. Detweiler-Bedell, R. Orlick, K. McFaddin, C. Harcleroad, M. Figueroa); Ohio Wesleyan University (L. Hall, B. Andereck, D. Swartzentruber, C. Stinemetz); Southern Oregon University (D. DeNeui, C. Stanek); University of California, Santa Cruz (J. Padgett); University of Central Arkansas (K. Boniecki, A. Hall); University of Oregon (G. Schoonover, K. Frazee); University of Pittsburgh (K. Binning, P. Ruz); University of Waterloo (H. Westmorland, V. Young, L. Brackenridge); Wabash College (R. Horton); and Yale University (H. Peck, M. Poe, T. Pavlis). We thank T. Dee, S. Gaither, C. Muller, S. Porter, L. Quay, S. Roberts, C. Smith, B. Tipton, R. Urstein, and S. Woodruff for feedback and support; the Center for Advanced Study in the Behavioral Sciences (CASBS) for meeting and planning space; and the Mindset Scholars Network for community, feedback, and support. **Funding:** Funding was provided by the participating partner colleges and universities; National Institute of Health: grants R01HD08477 and P2CHD042849; National Science Foundation: grants DRL-1450755, HRD-1661004, and 1761179; the Higher Education Quality Council of Ontario (HEQCO); the Raikes Foundation; the William T. Grant Foundation: grants 189706, 184761, and 182921; an Advanced Research Fellowship from the Jacobs Foundation to D.S.Y.; and fellowships from the Center for Advanced Study in the Behavioral Sciences (CASBS) to G.M.W., M.C.M., and D.S.Y. **Author contributions:** Conceptualization: G.M.W., M.C.M., C.Log., D.S.Y., S.T.B., and O.F. Data curation: J.P.G., T.T.A., P.F., M.Ga., C.H., A.H., and X.A.L. Formal analysis: D.S.Y., J.P.G., and K.T.U.E. Investigation (primary): S.T.B., O.F., A.B., K.L.B., E.R.C., M.Go., A.H., K.M.K., L.A.M.-P., and S.L.R. Investigation (secondary): T.T.A., S.A., S.C., P.F., M.Ga., M.K.G., C.S.H., J.M.L.F., C.Lok, K.M., G.A.M., M.N., E.O., E.N.S., D.B.T., H.E.W., and M.O.W. Methodology: G.M.W., M.C.M., C.Log., D.S.Y., and S.T.B. Funding acquisition: G.M.W., M.C.M., C.Log., O.F., and N.K. Project administration: A.B., X.A.L., and N.K. Software: J.P.G., D.P., and X.A.L. Supervision: G.M.W., M.C.M., C.Log., and D.S.Y. Validation: J.P.G. and K.T.U.E. Visualization: D.S.Y. Writing – original draft: G.M.W. Writing – review & editing: G.M.W., M.C.M., C.Log., D.S.Y., J.P.G., S.T.B., K.L.B., K.M.K., T.T.A., C.S.H., and D.B.T. **Competing interests:** G.M.W., M.C.M., and D.S.Y. have disseminated the findings from past intervention experiments, including social-belonging, through paid speaking appearances or consulting for educational institutions or private companies. All consulting and speaking appearances have been disclosed to their respective university offices of research support and compliance, and no financial conflicts of interest have been identified under university policies. D.P. is the cofounder and executive director of PERTS, which offers free social-belonging interventions. None of the other authors has an ongoing financial relationship with any private entity that provides social-belonging products or services. **Data and materials availability:** Code and data will be available upon publication to researchers who agree to terms of data use, including institutional review board approvals and analysis in a secure computing environment and prohibitions against any analysis that risks exposing the identity of participating students (i.e., deductive disclosure) or of participating institutions in reports of school-specific results without permission from the partner institution. Please contact belonging@accelerateequity.org. The preregistered analysis plan can be found in the Open Science Framework (OSF) (*46*). The intervention module will not be commercialized and is available at no cost to all postsecondary schools in the US or Canada that wish to use it: https://www.perts.net/orientation/cb. Researchers wanting to access the intervention materials can see sample content in the supplementary materials and may contact G.M.W. or C.Log. and must agree to terms of use, including noncommercialization of the intervention and prohibitions on independent use. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade4420
Materials and Methods
Supplementary Text
Tables S1 to S22
References (*47*–*88*)

Submitted 26 August 2022; accepted 16 March 2023
10.1126/science.ade4420

STEROL BIOSYNTHESIS

De novo phytosterol synthesis in animals

Dolma Michellod[1], Tanja Bien[2], Daniel Birgel[3], Marlene Violette[4], Manuel Kleiner[4], Sarah Fearn[5], Caroline Zeidler[1], Harald R. Gruber-Vodicka[1]†, Nicole Dubilier[1,6]*, Manuel Liebeke[1]*

Sterols are vital for nearly all eukaryotes. Their distribution differs in plants and animals, with phytosterols commonly found in plants whereas most animals are dominated by cholesterol. We show that sitosterol, a common sterol of plants, is the most abundant sterol in gutless marine annelids. Using multiomics, metabolite imaging, heterologous gene expression, and enzyme assays, we show that these animals synthesize sitosterol de novo using a noncanonical C-24 sterol methyltransferase (C_{24}-SMT). This enzyme is essential for sitosterol synthesis in plants, but not known from most bilaterian animals. Our phylogenetic analyses revealed that C_{24}-SMTs are present in representatives of at least five animal phyla, indicating that the synthesis of sterols common to plants is more widespread in animals than currently known.

Sterols are lipids that play essential roles in all multicellular eukaryotes. Their distribution and synthesis differs across eukaryotic kingdoms. Fungi and plants mainly synthesize sterols with 28 to 29 carbon atoms (C_{28} and C_{29}) called ergosterols and phytosterols (*1*, *2*), whereas animals predominantly produce the C-27 sterol cholesterol. These interkingdom differences reflect the complex evolutionary history of sterol synthesis. Phylogenetic analyses suggest that most enzymes for the biosynthesis of plant, fungal, and animal sterols were present in the last eukaryotic common ancestor (LECA) (*3*, *4*), with the distribution observed in most extant eukaryotes evolving through multiple events of enzyme losses and different pathways for sterol synthesis.

Cholesterol differs from C_{28} and C_{29} sterols by only one methyl or ethyl group at position C_{24}. This alkylation is catalyzed by C-24 sterol methyltransferase (C_{24}-SMT), an enzyme widely distributed in plants, microbial eukaryotes, and fungi but not known from nearly all animals (*5*, *6*). The exceptions are some marine sponges and the marine annelid *Capitella teleta* (*7*, *8*). Marine sponges have an unusual sterol composition enriched in highly branched alkylated sterols (*9*–*12*), of which some are synthesized by C_{24}-SMT homologs (*13*). As with most other enzymes required for sterol synthesis, an ancestral C_{24}-SMT was likely present in the LECA (*3*) and is assumed to have been lost early in animal evolution following the divergence of sponges, explaining why eumetazoans are not able to produce C_{28} and C_{29} sterols (*6*, *8*).

Sitosterol is the main sterol in the marine gutless annelid *Olavius algarvensis*

O. algarvensis belongs to a group of gutless marine annelids found worldwide, mainly in coral reef and seagrass sediments. These annelid worms lack a digestive system and are obligately associated with bacterial endosymbionts that provide them with nutrition (*14*–*17*). As part of our ongoing research on the *O. algarvensis* symbiosis, we analyzed the metabolome of single worm individuals using both gas chromatography mass spectrometry and high-performance liquid chromatography mass spectrometry. These analyses revealed an unusual sterol composition, with sitosterol accounting for most of the sterols detected (60%), and the remainder consisting of cholesterol (Fig. 1A and fig. S1). This was unexpected, as cholesterol generally dominates the sterol pool in bilaterians, often making up more than 90% of the total sterol content (*18*, *19*). Sitosterol is abundant in most plants but among bilaterians has only been reported as the most abundant sterol in a few phytoparasitic nematodes that are not capable of de novo sterol synthesis (*20*–*22*). In these plant parasites, it is unclear whether their sitosterol is only present in the nematode gut content or incorporated into their cells and tissues. The absence of a gut in *O. algarvensis* excludes sterol contamination from plant matter in the digestive tract.

O. algarvensis sterols have an isotopic composition that is distinct from their environment

We next asked whether *O. algarvensis* could acquire its sterols from the environment through passive diffusion. Chemical analyses of porewater collected in the vicinity of seagrass meadows—the habitat of many gutless annelids (including *O. algarvensis*)—revealed that sterols were present in the environment in concentrations sufficient to sustain the growth of small sterol-auxotrophic invertebrates (supplementary text and fig. S2). Therefore, we further investigated the origin of sterols in *O. algarvensis* by analyzing the carbon isotopic signature ($\delta^{13}C$) of sterols in the worms, their environment (which includes the seagrass *Posidonia oceanica*) and the porewater of the sediments these worms live in. Carbon isotopic signatures are used to reveal carbon sources and their paths through the food web. As a rule, the bulk $\delta^{13}C$ values of animals reflect their dietary sources [0.5 per mil (‰) to 2 ‰ difference] (*23*, *24*), but sterols are typically depleted in ^{13}C relative to bulk biomass by as much as 5 ‰ to 8 ‰ (*25*, *26*). Results from gas chromatography isotope ratio mass spectrometry (GC-IRMS) with single metabolite resolution showed that sitosterol in the seagrass and porewater had $\delta^{13}C$ values ranging from −30 ‰ to −15 ‰ (Fig. 1E and supplementary text). The sterols in *O. algarvensis*, as well as those in another co-occurring gutless annelid species, *Olavius ilvae*, had much lower $\delta^{13}C$ values: −38 ‰ to −36 ‰ for sitosterol and −40 ‰ to −31 ‰ for cholesterol (Fig. 1E). The difference in the isotopic signature of sterols in both *Olavius* species and their environment excludes that these worms acquired sterols from their environment, and instead indicates an endogenous origin. *O. algarvensis*, as all other *Olavius* and *Inanidrilus* species, derives all its nutrition from its chemosynthetic bacterial symbionts, and this is reflected in its bulk isotopic composition with $\delta^{13}C$ values of −30.6 ‰ (*27*). The ^{13}C-depleted signatures of both cholesterol and sitosterol by 1 ‰ to 10 ‰ compared with bulk biomass in *O. algarvensis* and *O. ilvae* led us to hypothesize that these animals synthesize both sterols de novo, using carbon derived from their chemosynthetic symbionts.

Bacterial symbionts are not the source of sitosterol in *O. algarvensis*

Having shown that sitosterol in *O. algarvensis* and *O. ilvae* did not originate from the uptake of plant sterols from their environment, we next investigated whether their bacterial symbionts were the source of this C_{29} sterol. The bacterial symbionts of these worms form a thick layer between the cuticle and the epidermis of the animal (Fig. 1B). To localize the distribution of sterols in *O. algarvensis*, we used two high spatial resolution metabolite imaging techniques. Time-of-flight secondary ion mass spectrometry (TOF-SIMS) data revealed that, at a spatial resolution of 0.4 μm, both sitosterol and cholesterol were uniformly distributed throughout the animals' tissues (Fig. 1, C and D). We found no evidence for a tissue-specific distribution of these two sterols—that is, there was no correlation between symbiont location and sitosterol distribution. These findings are supported by a second mass spectrometry imaging method, matrix-assisted laser

[1]Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany. [2]Institute of Hygiene, University of Münster, Robert-Koch-Str. 41, 48149 Münster, Germany. [3]Institute for Geology, Center for Earth System Research and Sustainability, University of Hamburg, Bundesstraße 55, 20146 Hamburg, Germany. [4]Department of Plant and Microbial Biology NC State University, Raleigh, NC 27695, USA. [5]Department of Materials, Imperial College London, London SW7 2AZ, UK. [6]MARUM, Center for Marine Environmental Sciences, University of Bremen, 28359 Bremen, Germany.
***Corresponding author. Email: mliebeke@mpi-bremen.de (M.L.); ndubilier@mpi-bremen.de (N.D.)**
†Present address: Zoologisches Institut, University of Kiel, 24118 Kiel, Germany.

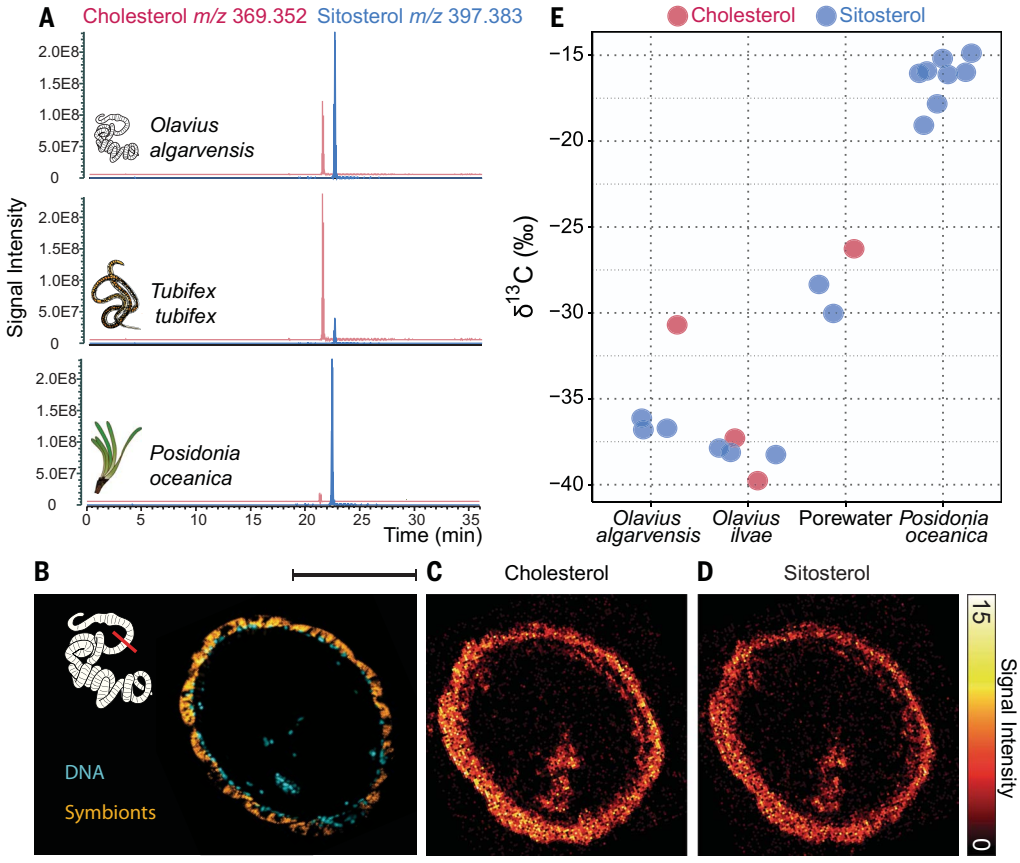


Fig. 1. *Olavius algarvensis* has an unusual sterol profile dominated by sitosterol, a common plant sterol. (**A**) Extracted-ion chromatograms (XIC) of cholesterol [$[M\text{-}H_2O+H]^+$ $C_{27}H_{45}$ at *m/z* 369.352 (red)] and sitosterol [$[M\text{-}H_2O+H]^+$ $C_{29}H_{49}$ at *m/z* 397.383 (blue)]. The XICs were generated from lipid extracts of (from top to bottom): the gutless marine annelid *O. algarvensis*, the freshwater annelid *Tubifex tubifex*, and the seagrass *Posidonia oceanica*. (**B**) Chemosynthetic symbiotic bacteria are located just below the cuticle of *O. algarvensis*. 16S rRNA fluorescence in situ hybridization (FISH) image of a cross section through a worm showing the symbionts in yellow (general eubacterial probe) and host nuclei in blue (DAPI). (**C** and **D**) Distribution of sitosterol and cholesterol in *O. algarvensis* measured by TOF-SIMS (C) Summed intensity of cholesterol ions (*m/z* 369.38, 385.34, 401.35) measured with TOF-SIMS. (D) Summed intensity of sitosterol ions (*m/z* 397.47, 383.37, 413.45) as measured by TOF-SIMS. (**E**) The ^{13}C isotopic composition of sterols in gutless annelids differed from that of the neighboring seagrass (*P. oceanica*) and sediment porewater. Scale bar in (B) to (D) 100 μm.

desorption/ionization mass spectrometry imaging (MALDI-2-MSI), of cross and longitudinal sections at a spatial resolution of 5 μm. The MALDI imaging data of longitudinal worm sections confirmed a uniform distribution of sitosterol and cholesterol throughout the animal (figs. S3 and S4) and the identity of these sterols (table S1). This homogeneous sterol distribution suggests that the bacterial symbionts are not the source for sitosterol in *O. algarvensis*.

As a second approach to investigate whether the bacterial symbionts are the source of sitosterol, we sequenced and assembled the genomes of the *O. algarvensis* symbionts and screened them for enzymes involved in de novo sterol synthesis. These analyses revealed that the symbionts, as in most bacteria, do not encode enzymes involved in de novo sterol synthesis (for more details see materials and methods).

O. algarvensis encodes and expresses enzymes involved in sitosterol synthesis that overlap with those of cholesterol synthesis

Having ruled out a symbiotic origin and an environmental source of sitosterol in *O. algarvensis*, we next investigated whether the animals themselves can synthesize this typical plant sterol. To identify and characterize the biosynthetic pathways involved in sterol production, we sequenced and assembled the genome of *O. algarvensis* and analyzed metatranscriptomic and metaproteomic data to search for enzymes involved in de novo sterol synthesis. The host possessed the full enzymatic toolbox required for cholesterol and sitosterol synthesis. For cholesterol synthesis, the genome of *O. algarvensis* encodes homologs of 11 enzymes known to be involved in the synthesis of this sterol (fig. S5). The intron-exon structure of these genes confirms their eukaryotic origin and excludes bacterial contamination (Fig. 2A and table S2). The cholesterol biosynthesis pathway, starting with squalene, is a series of 10 connected enzymatic reactions encoded by 11 genes (fig. S5 and table S3). Homologs of all enzymes were transcribed (11 out of 11 enzymes) and 5 out of 11 proteins were detected in the proteome of *O. algarvensis* (fig. S5 and tables S4 and S5), indicating active expression of the genes involved in cholesterol synthesis. Phylogenetic analysis allowed us to assign each homolog to an ortholog group and thus to a potential function (figs. S6 to S15). Collectively, these data show that *O. algarvensis* has all the enzymes required for de novo cholesterol synthesis, which in combination with the isotopic signature of their cholesterol suggests that these annelids are able to synthesize cholesterol.

Notably, our analyses also identified a homolog of C_{24}-SMT in the genome of *O. algarvensis*, an enzyme essential to sitosterol synthesis in plants (Fig. 2). As described above, most bilaterians are assumed to lack C_{24}-SMT. C_{24}-SMT catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to the sterol side chain and is essential for the biosynthesis of sitosterol and other C_{28}-C_{29} sterols commonly found in plants and fungi. As with the cholesterol synthesis genes, the intron-exon structure of the putative C_{24}-SMT gene confirmed its eukaryotic origin and excluded bacterial contamination (Fig. 2 and table S2). The putative C_{24}-SMT gene is a 1071-base pair (bp) open reading frame (ORF) encoding a 356-amino acid polypeptide and contains all the conserved residues characteristic of C_{24}-SMT as well as the 4 conserved signature motifs responsible for substrate binding (fig. S16) (*28*–*31*). We identified the C_{24}-SMT gene in *O. algarvensis* transcriptomes and proteomes, confirming that these animals express this enzyme (Fig. 2 and tables S4 and S5). Our findings suggest that the *O. algarvensis* C_{24}-SMT gene encodes a functional enzyme involved in sitosterol metabolism.

The *O. algarvensis* C_{24}-SMT homolog is bifunctional and consecutively transfers methyl groups to sterol intermediates

Two methylation reactions are required for the final steps of sitosterol synthesis, one that adds a methyl group at C-24 and one at C-28.

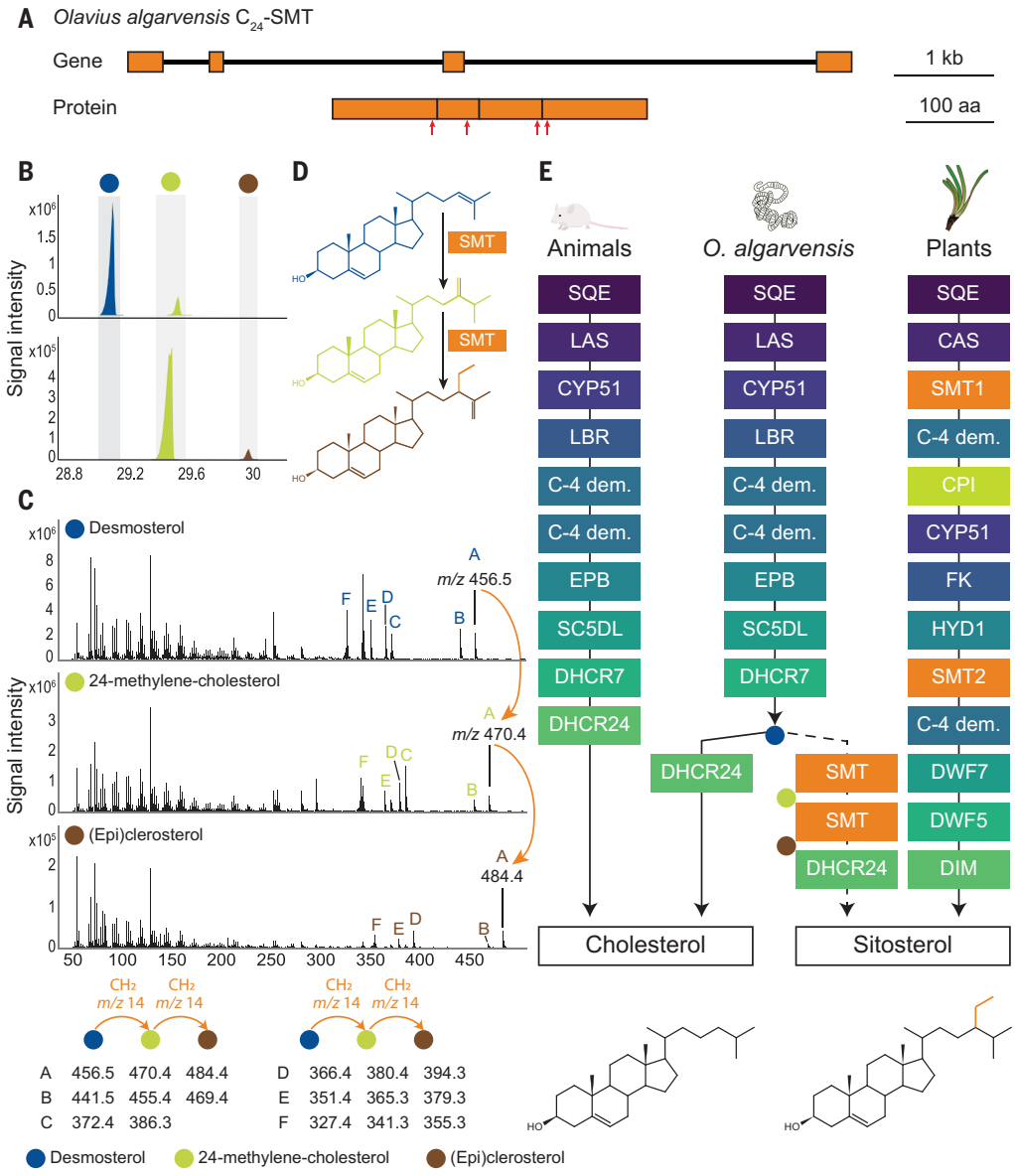


Fig. 2. *Olavius algarvensis* encodes and expresses a C_{24}-SMT that catalyzes two consecutive methylations, using desmosterol as the first and 24-methylene-cholesterol as the second substrate. (**A**) The *O. algarvensis* C_{24}-SMT gene consists of four exons, separated by three introns. The exons form a 1071-bp open reading frame encoding a 356 amino-acid polypeptide. The four conserved regions of the enzyme are highlighted by red arrows. (**B**) Chromatograms of enzymatic assays with desmosterol (top) and 24-methylene-cholesterol (bottom) as substrates. *O. algarvensis* C_{24}-SMT, after overexpression in *E. coli*, added a methyl group to the side chains of desmosterol and 24-methylene-cholesterol. In the first methylation step, desmosterol, an intermediate of cholesterol synthesis, was methylated to produce 24-methylene-cholesterol (C_{28} sterol). In the second methylation, 24-methylene-cholesterol was methylated to produce a C-29 sterol, most likely (epi) clerosterol (C_{29} sterol). (**C**) Mass spectra of the different substrates and methylated products from the enzymatic assays. Sterol intermediates differ by the number of methyl groups (CH_2 at *m/z* 14) attached to their side chain. The side chain of desmosterol is not methylated, 24-methylene-cholesterol has a methyl group at C-24, and (epi)clerosterol has two methyl groups at C-24 and C-28. The substrates and methylated products were identified by MS, retention time, and comparison with standards. The fragmentation pattern suggests that the methyl groups were added to the side chain of the sterols. The stereochemistry of the methyl groups was not determined. (**D**) Structural representation of the two methylation steps in *O. algarvensis*. (**E**) Comparison of the enzyme used in the proposed sterol synthesis pathways in *Olavius* to the canonical cholesterol and sitosterol synthesis pathways (similar enzymatic reactions are colored similarly). The first six steps are common to both cholesterol and sitosterol synthesis pathways. This trunk pathway branches off after the synthesis of desmosterol. For sitosterol synthesis, desmosterol is first methylated by C_{24}-SMT to 24-methylene-cholesterol, which is then methylated in a second, consecutive step by C_{24}-SMT to (epi)clerosterol. (Epi)clerosterol is reduced to sitosterol by a sterol C24-reductase (DHCR24, DIM). Squalene monooxygenase (SQE), oxydosqualene cyclase (LAS, CAS), sterol 14 demethylase (CYP51), sterol 14-reductase (LBR, FK), C-4 demethylation (C-4 dem.), Sterol Δ7-Δ8 isomerase (EBP, HYD1), sterol 5-desaturase (SC5DL, DWF7), sterol Δ7 reductase (DHCR7, DWF5), and C-24 sterol methyltransferase (C_{24}-SMT, SMT1, SMT2).

These reactions can be catalyzed by the same or different enzymes (*6*, *13*, *32*–*34*) (supplementary text). Because *O. algarvensis* encodes a single C_{24}-SMT homolog, we hypothesized that this enzyme mediates both C-24 and C-28 methylation.

To test this hypothesis, we overexpressed *O. algarvensis* C_{24}-SMT in *Escherichia coli* and examined its enzymatic activity, substrate preferences, and products by assaying crude protein extracts with the methyl-donor SAM and different sterol substrates. The heterologously expressed C_{24}-SMT from *O. algarvensis* was not able to methylate classical plant sterol substrates (table S6). However, the enzyme was able to methylate zymosterol and desmosterol, two intermediates of the cholesterol biosynthetic pathway. When incubated with either of these sterol substrates and SAM, the *O. algarvensis* C_{24}-SMT produced a methylated sterol product (C_{28}) (Fig. 2 and figs. S17 to S20). Zymosterol was methylated to fecosterol and desmosterol to 24-methylene-cholesterol. The shift in retention times and changes in mass spectra of the products indicated that a methyl group was added to their sterol side chain, likely at the C_{24}-position (Fig. 2, and figs. S19 and S20). These results suggest that the cholesterol and sitosterol synthesis pathways overlap in *O. algarvensis*, as the two C_{24}-SMT substrates, zymosterol and desmosterol, are intermediates produced in the second half of the animal cholesterol synthesis pathway (Fig. 2).

After confirming the first methylation step at C-24, we next searched for potential substrates for the second methylation step at C-28. This second methylation is essential as sitosterol is a C_{29} compound, characterized by the

presence of an ethyl group on its C-24 position. To test our hypothesis that both of these methylations are catalyzed by the *O. algarvensis* C_{24}-SMT, we selected the product of the first methylation, 24-methylene-cholestrol, as well as campesterol, as potential substrates for the second methylation (fig. S21). Only 24-methylene-cholesterol, but not campesterol, was methylated by the *O. algarvensis* C_{24}-SMT, producing a C_{29} sterol compound, most likely (epi)clerosterol (Fig. 2 and fig. S22). 24-methylene-cholesterol is the product of the methylation of desmosterol, providing evidence to support our hypothesis that in *O. algarvensis*, the C-24 and C-28 methylations are catalyzed by the same enzyme and occur consecutively. That is, the *O. algarvensis* C_{24}-SMT first methylates desmosterol at C-24 to produce the C_{28} sterol 24-methylene-cholesterol, and then adds a second methyl group to 24-methylene-cholesterol at C-28, to produce the C_{29} sterol (epi)clerosterol. (Epi)clerosterol differs from sitosterol by the presence of a double bond at position C-25(26 or 27). This double bond is most likely removed by delta(24)-sterol reductase (DHCR24), which was expressed based on its presence in *O. algarvensis* transcriptomes and proteomes (figs. S5 and S15). Our results provide evidence for an animal C_{24}-SMT that catalyzes the two methylation steps needed to synthesize sitosterol from a cholesterol intermediate, revealing a previously unknown pathway for C_{28}-C_{29} sterol synthesis in animals (Fig. 2).

C_{24}-SMT homologs are widespread in annelids

Having demonstrated the activity of an animal C_{24}-SMT homolog that enables *O. algarvensis* to synthesize sitosterol de novo, we asked whether other gutless annelids also encode functional C_{24}-SMTs. To answer this question, we analyzed the sterol contents of six additional gutless annelid species. All six species had lipid profiles similar to that of *O. algarvensis*, with sitosterol as their major sterol (table S7 and supplementary text).

In addition to lipid profiling, we screened the transcriptomes of nine *Olavius* and *Inanidrilus* species and found that all expressed a C_{24}-SMT homolog, including *O. ilvae* (table S8). Moreover, heterologous gene expression analyses confirmed the C-24 methylation ability of a C_{24}-SMT homolog from another gutless annelid, *O. clavatus*. The C_{24}-SMT from *O. clavatus* is also a bifunctional sterol methyltransferase, capable of methylating zymosterol, desmosterol, and 24-methylene-cholesterol (figs. S17, S18, and S22).

We next asked whether C_{24}-SMT homologs are present in other annelids. We screened published transcriptomes and genomes and identified C_{24}-SMT homologs in 3 deep-sea gutless tubeworm species and in 17 gut-bearing annelid species from marine, limnic, and terrestrial environments (table S9 and fig. S23). Despite the presence of C_{24}-SMT homologs in these annelids, the published sterol profiles of annelids—including four gut-bearing species analyzed in this study—were dominated by cholesterol (table S9 and Fig. 1A). However, sitosterol as well as other C_{28} and C_{29} sterols accounted for a considerable proportion of total sterols (15 to 30%) in some of these species, including the vent and seep tubeworms *Riftia pachyptila* and *Paraescarpia echinospica* (table S9). These deep-sea siboglinid annelids are only distantly related to *Olavius* and *Inanidrilus*, but also lack a gut and gain all of their nutrition from their chemosynthetic symbionts (*35*). The sterol contents of these tubeworms are dominated by cholesterol and desmosterol, but the C_{28} sterol campesterol as well as other C_{28} and C_{29} sterols make up as much as nearly one-third of their sterol contents (*36*–*38*). Our discovery of C_{24}-SMT homologs in these deep-sea annelids suggest that these tubeworms are able to synthesize C_{28} and C_{29} sterols as well. Further support that annelid C_{24}-SMT homologs methylate sterols was recently shown for the enzyme encoded by the nonsymbiotic annelid *C. teleta* (*39*). Taken together, these results indicate that functional C_{24}-SMT homologs are widespread in the annelid phylum and that the presence of this gene is not restricted to hosts that harbor symbiotic bacteria.

C_{24}-SMT homologs are present in at least five animal phyla: Cnidaria, Porifera, Rotifera, Annelida, and Mollusca.

To assess the broader distribution of C_{24}-SMT homologs in animals, we performed protein searches against public databases (see materials and methods for details). Hits were found in six additional animal phyla: sponges, cnidarians, rotifers, mollusks, nematodes, and chordata (supplementary text and table S10). For the latter two, we concluded they are not animal C_{24}-SMTs for the following reasons. Among the chordata, we found a single homolog in a metatranscriptome from a fruit bat. It had 99.1% identity to sequences from a plant and fell in a clade with these, suggesting that this C_{24}-SMT is a contamination (supplementary text). The nematode sequences belonged to a group of SMTs called C-4 sterol methyltransferases (C_4-SMTs) that are specific to nematodes (*40*, *41*). Our phylogenetic analyses showed that these nematode C_4-SMTs are not closely related to the monophyletic clade of C_{24}-SMTs from plants, fungi, microbial eukaryotes, sponges, cnidarians, rotifers, annelids, and mollusks (Fig. 3B).

C_{24}-SMTs are widespread across the tree of life

We reconstructed the evolutionary relationships of animal C_{24}-SMTs from sponges, cnidarians, rotifers, mollusks, and annelids and compared these with previously described C_{24}-SMTs from bacteria, microbial eukaryotes, fungi, and plants (Fig. 3A). Our phylogenetic analyses revealed that C_{24}-SMTs are widespread across the tree of life and fall into nine well-supported clades (Fig. 3B, A to I). These clades were not congruent with the phylogeny of bacteria and eukaryotes, indicating a complex evolutionary history for C_{24}-SMTs, which we discuss below. Many eukaryotes had more than one C_{24}-SMT homolog, with members of some groups such as the choanoflagellates and cnidarians encoding as many as four homologs that belonged to different, phylogenetically distant clades (Fig. 3C). Most animal C_{24}-SMTs clustered in seven clades (Fig. 3A and figs. S24 to S27), with each clade consisting of animal sequences only. The exceptions were animal C_{24}-SMTs that fell on single, long branches that formed sister lineages to other eukaryotic clades (Fig. 3A, supplementary text, and figs. S25 and S26).

For the animal C_{24}-SMTs, two of the seven clades (2.F and 3.F) contain homologs from sponges and annelids whose C-24 methylation function has been experimentally verified [this study and (*13*, *39*)]. The five remaining clades consist of predicted C_{24}-SMTs, but their high sequence homology to verified C_{24}-SMTs together with conserved protein domains suggest that these homologs also methylate sterol intermediates at C-24.

The complex evolutionary history of C_{24}-SMTs

In two animal phyla, Annelida and Rotifera, we observed strong congruence between their C_{24}-SMT homologs and their evolutionary history (Fig. 3D and figs. S27 and S28). For annelids, in addition to the gutless *Olavius* and *Inanidrilus* sequences from our study, we found C_{24}-SMT homologs in nine annelid orders from limnic, terrestrial, and marine environments. Annelid C_{24}-SMTs fell within two sister clades, one from Errantia and the other from Sedentaria, corresponding to two major subgroups of Annelida (Fig. 3D). The only exception was *Megasyllis neponica*, which belongs to the Errantia but has a C_{24}-SMT that falls in the Sedentaria (Fig. 3D). Congruence was also visible within the two annelid clades (Fig. 3D and fig. S28). These results suggest that C_{24}-SMTs from extant annelids evolved through direct inheritance from a common ancestor. Similarly, the rotifer C_{24}-SMT clade consisted largely of sequences from the Bdelloidea, a phylogenetically well-defined class within the phylum Rotifera (supplementary text). Within this clade, the C_{24}-SMT topology corresponded well to the phylogeny of bdelloid rotifers, consistent with direct inheritance driving the evolution of this gene within this animal phylum as well (fig. S27).

At longer evolutionary time scales, there was no congruence between C_{24}-SMT homologs and the evolutionary history of the lineages they came from, not even between closely related

phyla. For example, Annelida and Mollusca are more closely related to each other than to Porifera, yet their C_{24}-SMTs often clustered with those from Porifera. The lack of congruence at deeper nodes of the C_{24}-SMT tree with the evolutionary history of eukaryotes is also visible in the interspersed phylogeny of C_{24}-SMT homologs from fungi, plants, and animals throughout the tree (Fig. 3A). The consistent incongruencies throughout most of the C_{24}-SMT tree are hard to reconcile with direct inheritance from a common eukaryotic ancestor. Repeated, independent events of lateral gene transfer (LGT) may provide a more likely explanation for the C_{24}-SMT tree topology.

LGT is widespread and well-studied in bacteria, but less is known about LGT within extant eukaryotic lineages. Recent evidence from high quality genome sequencing suggests that LGT may play a more important role in

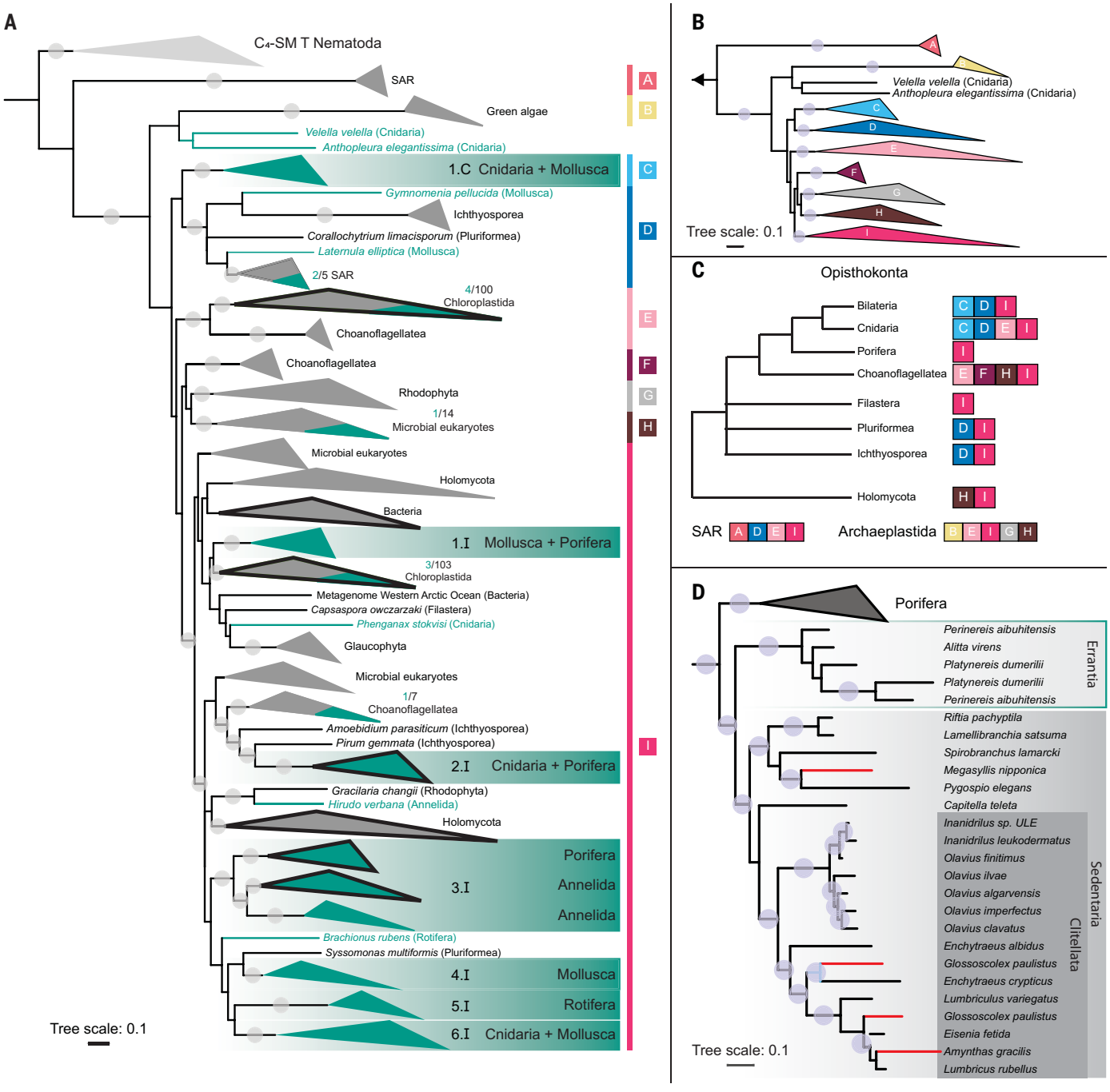


Fig. 3. C_{24}-SMT homologs are widely distributed across the tree of life including at least five animal phyla. (**A**) Maximum likelihood amino acid tree of eukaryotic and bacterial C_{24}-SMTs. Sequences were clustered at 90% identity. Nodes with ultrafast bootstrap (UFB) values >95% are shown with gray circles. Nematode C_4-SMTs were used as an outgroup. The homologs found in animal datasets are highlighted in green. Clades with experimentally verified C_{24}-SMTs are outlined in black. (**B**) Collapsed version of the tree shown in (A). C_{24}-SMTs fall into nine well-supported clades (UFB >95%). (**C**) Most eukaryotic clades had several phylogenetically distinct homologs. (**D**) Detailed view of the Annelida [clade 3.I from (A)]. The phylogeny of annelid C_{24}-SMTs largely corresponds to their evolutionary history, exceptions are shown in red.

eukaryotes than previously assumed (*42*–*45*). LGT within eukaryotes is assumed to be most commonly mediated by viruses and bacteria, particularly in species that are intimately associated with beneficial or pathogenic microorganisms (*46*). However, we ruled out LGT mediated by associated bacteria in the gutless *Olavius* and *Inanidrilus*, as we exhaustively searched their symbiont metagenomes for C_{24}-SMTs. LGT from bacteria to eukaryotes of C_{24}-SMTs is also unlikely to have played a role in the more recent evolution of other animals, as bacterial C_{24}-SMTs were not closely related to those from animals, with the exception of sponges (Fig. 3A). Alternatively, LGT mediated by viruses could explain how genes are transferred between eukaryotes (*47*). Independent of the precise route for LGT across eukaryotes, our study provides evidence for rampant LGT of gene homologs within the animal kingdom and suggests that the acquisition of C_{24}-SMTs provides animals with a strong selective advantage.

Conclusions

Although cholesterol is commonly the most abundant sterol in animals, through studies of non-model organisms and technological advances, evidence is growing that C_{28} and C_{29} sterols are also often present, sometimes in considerable amounts (*48*–*51*). Our findings highlight the value of reconsidering the source of C_{28} and C_{29} sterols in both extant animals and the fossil record and determining whether these originated from the animal's diet as is often assumed, or if they were synthesized by the animals themselves. It is also timely to reconsider the widespread use of the term phytosterols for C_{28} and C_{29} sterols, as we now know that microbial eukaryotes, plants, and animals can synthesize these lipids.

How can we explain the unusually high abundance of sitosterol in gutless *Olavius* and *Inanidrilus*? Studies have shown that C_{28} and C_{29} sterols can be incorporated into animal membranes (*52*, *53*) and provide beneficial effects when added to animal diets or cell lines. For example, they act as cholesterol-lowering agents, have anti-tumor, anti-inflammatory, antibacterial, and antifungal properties (*54*, *55*), and modulate interactions between bacterial pathogens and eukaryotic hosts (*56*). Therefore, the anti-inflammatory and antibacterial properties of sitosterol, as well as its ability to protect animal cells against toxins that target cholesterol (*57*), might play a role in the symbiosis between *Olavius* and *Inanidrilus* and their chemoautotrophic symbionts, by preventing the symbionts from entering the host cytoplasm. Changes in sterol composition also affect the fluidity and permeability of membranes, and these physical changes in turn affect many cellular processes. For example, sitosterol has been shown to enhance mitochondrial energy metabolism in a mouse cell line (*58*) and might enable *Olavius* and *Inanidrilus* to gain more energy under low oxygen concentrations in their environment. Our findings highlight the need for studies that elucidate the physiological and ecological roles of sitosterol in animals. *Olavius* and *Inanidrilus* are valuable model systems for studying the impact of C_{28} and C_{29} sterols on animal membrane properties in vivo and furthering our understanding of the roles sterols play in eukaryotic cells.

REFERENCES AND NOTES

1. M. J. Lagarda, G. García-Llatas, R. Farré, *J. Pharm. Biomed. Anal.* **41**, 1486–1496 (2006).
2. J. D. Weete, M. Abril, M. Blackwell, *PLOS ONE* **5**, e10899 (2010).
3. E. Desmond, S. Gribaldo, *Genome Biol. Evol.* **1**, 364–381 (2009).
4. R. E. Summons, A. S. Bradley, L. L. Jahnke, J. R. Waldbauer, *Philos. Trans. R. Soc. B* **361**, 951–968 (2006).
5. J. K. Volkman, *Org. Geochem.* **36**, 139–159 (2005).
6. B. A. Haubrich *et al*., *Phytochemistry* **113**, 64–72 (2015).
7. J. Germer, N. Cerveau, D. J. Jackson, *Front. Mar. Sci.* **4**, (2017).
8. D. A. Gold *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2684–2689 (2016).
9. A. Aiello, E. Fattorusso, M. Menna, *Steroids* **64**, 687–714 (1999).
10. P. R. Bergquist, P. Karuso, R. C. Cambie, D. J. Smith, *Biochem. Syst. Ecol.* **19**, 17–24 (1991).
11. R. G. Kerr, B. J. Baker, *Nat. Prod. Rep.* **8**, 465–497 (1991).
12. J. A. Zumberge, thesis, UC Riverside (2019).
13. M. O. Brown, B. O. Olagunju, J. L. Giner *et al*., *Nat Commun* **14**, 1859 (2023).
14. O. Giere, *Mar. Ecol. Prog. Ser.* **5**, 353–357 (1981).
15. O. Giere, *Zool. Scr.* **14**, 279–286 (1985).
16. M. Kleiner *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, E1173–E1182 (2012).
17. T. Woyke *et al*., *Nature* **443**, 950–955 (2006).
18. L. J. Goad, *Pure Appl. Chem.* **53**, 837–852 (1981).
19. N. H. Sissener, G. Rosenlund, I. Stubhaug, N. S. Liland, *Br. J. Nutr.* **119**, 599–609 (2018).
20. D. J. Chitwood, P. A. Hutzell, W. R. Lusby, *J. Nematol.* **17**, 64–68 (1985).
21. D. J. Chitwood, M. A. McClure, M. F. Feldlaufer, W. R. Lusby, T. E. Oliver, *J. Nematol.* **19**, 352–360 (1987).
22. R. J. Cole, L. R. Krusberg, *Exp. Parasitol.* **21**, 232–239 (1967).
23. J. H. McCutchan Jr., W. M. Lewis Jr., C. Kendall, C. C. McGrath, *Oikos* **102**, 378–390 (2003).
24. A. V. Tiunov, *Izv. Akad. Nauk Ser. Biol.* **34**, 475–489 (2007).
25. E. A. Canuel, K. H. Freeman, S. G. Wakeham, *Limnol. Oceanogr.* **42**, 1570–1583 (1997).
26. J. M. Hayes, *Reviews in Mineralogy and Geochemistry* **43**, 225–277 (2001).
27. M. Kleiner *et al*., *Environ. Microbiol.* **17**, 5023–5035 (2015).
28. K. Ganapathy *et al*., *Biochim. Biophys. Acta* **1781**, 344–351 (2008).
29. P. Jayasimha, W. D. Nes, *Lipids* **43**, 681–693 (2008).
30. R. M. Kagan, S. Clarke, *Arch. Biochem. Biophys.* **310**, 417–427 (1994).
31. W. D. Nes *et al*., *Biochemistry* **43**, 569–576 (2004).
32. P. Bouvier-Navé, T. Husselstein, P. Benveniste, *Eur. J. Biochem.* **256**, 88–96 (1998).
33. M.-A. Hartmann, in *Lipid Metabolism and Membrane Biogenesis*, G. Daum, Ed. (Springer, 2004); pp. 183–211.
34. A. K. Neelakandan *et al*., *Phytochemistry* **70**, 1982–1998 (2009).
35. M. Bright, O. Giere, *Symbiosis* **38**, 1–45 (2005).
36. H. Guan *et al*., *Deep Sea Res. Part I Oceanogr. Res. Pap.* **174**, 103556 (2021).
37. C. F. Phleger *et al*., *Comp. Biochem. Physiol. B* **141**, 196–210 (2005).
38. G. Rieley, C. L. van Dover, D. B. Hedrick, D. C. White, G. Eglinton, *Spec. Publ. Geol. Soc.* **87**, 329–342 (1995).
39. D. Gold *et al*., Sterol methyltransferases in annelid worms rewrite the molecular fossil record. Research Square [Preprint] (2022); doi:10.21203/rs.3.rs-1686449/v1.
40. D. J. Chitwood, in *Physiology and Biochemistry of Sterols*, G. W. Patterson, W. D. Nes, Eds. (American Oil Chemists' Society, 1991), pp. 257–293.
41. S. Darnet, S. J. Fliesler, H. Schaller, *J. Lipid Res.* **61**, 129–132 (2020).
42. L. Boto, *Proc. R. Soc. B* **281**, 20132450 (2014).
43. C. Conaco *et al*., *PLOS ONE* **11**, e0151092 (2016).
44. F. Husnik, J. P. McCutcheon, *Nat. Rev. Microbiol.* **16**, 67–79 (2018).
45. S. M. Soucy, J. Huang, J. P. Gogarten, *Nat. Rev. Genet.* **16**, 472–482 (2015).
46. Y. Li *et al*., *Cell* **185**, 2975–2987.e10 (2022).
47. S. A. Widen, I. C. Bes, A. Koreshova, D. Krogull, A. Burga, bioRxiv 2022.07.12.499685 [Preprint] (2022); doi:10.1101/2022.07.12.499685.
48. L. Carreón-Palau, N. Ş. Özdemir, C. C. Parrish, C. Parzanini, *Mar. Drugs* **18**, 598 (2020).
49. E. A. Hambleton *et al*., *eLife* **8**, e43923 (2019).
50. T. Li *et al*., *Nat. Commun.* **12**, 4343 (2021).
51. M. Takeshima, M. H. Ogihara, H. Kataoka, *Int. J. Mol. Sci.* **20**, 4840 (2019).
52. O. G. Mouritsen, M. J. Zuckermann, *Lipids* **39**, 1101–1113 (2004).
53. F. Xu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14551–14556 (2005).
54. M. S. Bin Sayeed, S. S. Ameen, *Nutr. Cancer* **67**, 1214–1220 (2015).
55. S. Saeidnia, A. Manayi, A. R. Gohari, M. Abdollahi, *European J. Med. Plants* **4**, 590–609 (2014).
56. Y. P. M. van der Meer-Janssen, J. van Galen, J. J. Batenburg, J. B. Helms, *Prog. Lipid Res.* **49**, 1–26 (2010).
57. H. Li *et al*., *Sci. Rep.* **5**, 17668 (2015).
58. C. Shi, F. Wu, J. Xu, *J. Bioenerg. Biomembr.* **45**, 301–305 (2013).
59. J. A. Vizcaíno *et al*., *Nucleic Acids Res.* **44**, D447–D456 (2016).
60. D. Michellod *et al*., Supplementary materials from "De novo phytosterol synthesis in animals" - C24-SMTs alignment file and tree raw file, figshare (2023); https://doi.org/10.6084/m9.figshare.c.6443729.v1.
61. D. Michellod *et al*., Supplementary materials from "De novo phytosterol synthesis in animals" - Sterol profiles, figshare (2023); https://doi.org/10.6084/m9.figshare.c.6448546.v1.
62. D. Birgel *et al*., Compound specific isotopes of sitosterol/cholesterol in gutless annelids, seagrass, porewater from Elba, Italy, figshare (2023); doi:10.6084/m9.figshare.22189252.v1.
63. D. Michellod *et al*., Supplementary materials from "De novo phytosterol synthesis in animals" - Sterol imaging - MALDI2-MSI and TOF-SIMS, figshare (2023); doi:10.6084/m9.figshare.c.6443741.v1.
64. D. Michellod *et al*., Supplementary materials from "De novo phytosterol synthesis in animals" - C24-SMT assay results, figshare (2023); doi:10.6084/m9.figshare.c.6450151.v1.

ACKNOWLEDGMENTS

We thank A. Garsdal, J. Beckmann, M. Weinhold, M. Meyer, S. Wetzel, T. Markowski, T. Gulstad, K. Caspersen, and F. Fojt from the Max Planck Institute for Marine Microbiology (MPI-MM) for support with data acquisition and sample preparation. We thank M. Sadowski and G. D'Angelo for proofreading the manuscript. We are grateful to A. Gruhl, Y. Sato, N. Leisch, A. Mankowski, S. Bulgheresi, C. Bergin, E. M. Sogin, and K. Schmidt for sample collections and help in the field. We thank the Meyer Lab for providing us with *C. teleta* worms. We also thank B. Huettel (Max Planck Genome Center, Cologne) for his support with sequencing. LC-MS/MS–based proteomics measurements were made in the Molecular Education, Technology, and Research Innovation Center (METRIC) at North Carolina State University. We thank Bruker Daltonics for providing access to the prototype timsTOF flex MALDI-2 instrument and J. Soltwisch and K. Dreisewerd for their advice and support during the sterol imaging. We are grateful for fruitful discussions with L. Hambleton, C. Laumer, and colleagues in the Department of Symbiosis (MPI-MM). We thank M. Weber, C. Lott, and HYDRA staff for sample collections. This work is contribution 1071 from the Carrie Bow Cay Laboratory, Caribbean Coral Reef Ecosystem Program, National Museum of Natural History, Washington DC. **Author contributions:** D.M., M.L., and N.D. conceived the study. D.M. collected, processed, and analyzed the metabolomic, metatranscriptomic, and metagenomic data. H.G.V. advised and helped with the metatranscriptomic and metagenomic analyses. T.B. and S.F. collected mass spectrometry imaging data. D.B. performed the GC-IRMS measurements. M.V. and M.K. collected and analyzed the proteomics data. C.Z. and D.M. performed the heterologous gene expression. D.M. performed the enzyme assay and analyzed the data. D.M. and M.L. wrote the manuscript together with N.D. and contributions from H.G.V., T.B., D.B., M.K., and C.Z. **Funding:** We are grateful to the Max-Planck Society for their generous financial support over the many years it took to bring this work to fruition. Additional financial

support to N.D. from the Gottfried Wilhelm Leibniz Prize from the German Research Foundation (DFG) and a Gordon and Betty Moore Foundation Marine Microbiology Initiative Investigator Award (grant GBMF3811), and to M.L. from sea4society, a CDRmare campaign in the German Marine Research Alliance under Award number 03F0896D. M.K. received financial support from the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM138362 and the US National Science Foundation (grant IOS 2003107). **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The metaproteomic mass spectrometry data have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository (*59*) with the following dataset identifier: PXD014881. The sequencing data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers PRJEB52460 (PolyA libraries) and PRJEB52678 (PacBio reads). The C24-SMT protein alignment and the original tree file can be found on Figshare (*60*). The metabolomics and other MS-based data are on Figshare. The sterol profiles (*61*), the sterol GC-IRMS data (*62*), the metabolite imaging data (*63*), and the enzyme assay results (*64*) are available on Figshare. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add7830
Materials and Methods
Supplementary Text
Figs. S1 to S29
Tables S1 to S16
References (*65*–*90*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 13 July 2022; accepted 3 March 2023
10.1126/science.add7830

ROBOTICS

Multilegged matter transport: A framework for locomotion on noisy landscapes

Baxi Chong[1,2], Juntao He[3], Daniel Soto[3], Tianyu Wang[3], Daniel Irvine[4], Grigoriy Blekherman[4], Daniel I. Goldman[1,2,3]*

Whereas the transport of matter by wheeled vehicles or legged robots can be guaranteed in engineered landscapes such as roads or rails, locomotion prediction in complex environments such as collapsed buildings or crop fields remains challenging. Inspired by the principles of information transmission, which allow signals to be reliably transmitted over "noisy" channels, we developed a "matter-transport" framework that demonstrates that noninertial locomotion can be provably generated over noisy rugose landscapes (heterogeneities on the scale of locomotor dimensions). Experiments confirm that sufficient spatial redundancy in the form of serially connected legged robots leads to reliable transport on such terrain without requiring sensing and control. Further analogies from communication theory coupled with advances in gaits (coding) and sensor-based feedback control (error detection and correction) can lead to agile locomotion in complex terradynamic regimes.

The transport of matter across land is crucial to societies and groups (*1*, *2*, *3*), as well as to individuals during locomotion (*4*). Engineered self-propulsion as a means of terrestrial matter transport has been studied across scales from enormous multiwheeled trains (Fig. 1A) to small few-wheeled and legged robots (*5*–*7*). For large devices, low dissipation, inertia-dominated locomotion is a commonly used matter-transport scheme. Specifically, locomotion in wheeled systems on smooth surfaces such as tracks and roads will persist over long distance unless acted on by dissipative internal or external forces (Fig. 1A).

On natural terrain, dissipation attributable to external forces can occur through interactions with terrain heterogeneities such as obstacles, gaps, or inclines (*8*), as well as through interactions with flowable materials (*9*). In dissipation-dominated applications, such as those encountered in robot movement in certain agricultural (e.g., crop fields) or confined and crowded search-and-rescue (e.g., collapsed buildings) scenarios, a system must continuously and actively generate forces and/or reduce dissipation. In part, because of our lack of understanding of the terradynamic (*9*, *10*) interactions with the environments listed above, principles by which locomotors can be designed and controlled to guarantee reliable and predictable matter transport are lacking.

One engineering solution (*11*) to facilitate matter transport in complex terradynamic regimes is to use structures such as limbs to periodically make and break contact with the environment (*12*). Such dynamics can potentially simplify the thrusting interactions into a collection of discrete units, which minimize unexpected interference (*4*), thus providing an alternative to wheeled carriers on "noisy" landscapes. There have been two basic approaches of limb use in dissipation-dominated environments. The first relies heavily on sensors (*13*) to detect and respond to terrain heterogeneity in real time (*5*, *14*). This approach is used for the increasingly agile locomotion in state-of-the-art legged robots (mostly bipedal or quadrupedal) (*5, 12, 14, 15*). However, the use of sensors and high bandwidth control can be expensive and restricted to specific applications.

The second approach is to instill legged robots with "mechanical intelligence" such that locomotion can be performed with minimal environmental awareness. This has been most effective with devices with more than four legs, such as hexapods (*6*) and myriapods (*16*, *17*). Whereas more limbs help avoid catastrophic failures (e.g., loss of stability), terrain heterogeneity can still cause deficiencies in thrusting interactions, which substantially reduce locomotor performance (movie S1) (*18*–*20*). This raises the question of how variable numbers of limbs and sensors should be arranged such that one can guarantee that a locomotor can go from point A to point B in a specified time and across an arbitrarily complex landscape, and furthermore, how much sensing, feedback, bandwidth, and/or control are needed.

This question is analogous to that of information and signal transmission over noisy channels as first analyzed by Shannon nearly a century ago (*21*). Over a noiseless channel, a continuous analog signal is, in principle, able to convey an infinite amount of information (*22*). Despite its efficiency, an analog signal can be distorted by channel noise inherent in communication modalities, a property similar to heterogeneities introduced to smooth surfaces in inertia-driven matter transport (Fig. 1A). To counter channel noise in communication, Shannon (*21*) constructed a scheme in which the central idea was to digitize (encode) information into binary bit sequences and "buffer" (correct) the transmission error through redundancy (Fig. 1C).

With the analogy to information theory, it is reasonable—at least, conceptually—to anticipate reliable matter transport over noisy landscapes given sufficient redundancy of terrestrial interaction. In this work, we show that this anticipation is correct and leads to an open-loop framework for matter transport by which, for a complex terradynamic task, we can guarantee that sufficiently redundant multilegged robots can reliably and predictably self-transport over a given distance through "buffer and tolerate" dynamics without the need for sensing and feedback control or environmental awareness (Fig. 1B and movie S1) (*16*).

Development of the matter-transport framework

Our framework proceeds as follows (Fig. 2): We define a transport task as a physical entity moving to a specific destination D at a fixed time T. As shown in Fig. 2B (i), this is analogous to the intended message being transmitted across a noisy channel at a given rate. Similar to the bit-based digital signal transmission, we focus on a dissipation-dominated system in which locomotion is driven by thrusting interactions from basic active contacts (bacs, our analogy to bits), which are discrete units of active terradynamic interaction. Examples of bacs include limbs making contact with the environment (*23*) or vertical waves of contact in limbless robots (*24*). We quantify the temporal and spatial distribution of bacs as a binary sequence X^m, in which 1 denotes contact and 0 denotes noncontact [Fig. 2B (ii)]. As the locomotor implements the desired bac sequence over a noisy landscape (the analog of a noisy channel) [Fig. 2B (iii)], the terrain uncertainty can introduce contact noise to the actual bac sequence, Y^m [Fig. 2B (iv)]. This leads to a discrepancy between the actual destination $\hat{D}$ (evaluated at the scheduled time T) and the desired destination D [Fig. 2B (v)].

We next discuss our characterization and quantification of noisy landscapes. A dissipation-dominated terrain can have different types of heterogeneities, each with different complex terradynamic effects on bacs (*9*). Consider a terrain characterized by a height map, $h(x,y)$. Depending on the scale of the gradient, $[\partial h/\partial x, \partial h/\partial y]$, the terrain heterogeneity can affect the locomotion in the form of slopes, walls, or obstacles (Fig. 1B), which directly impact the thrust-generation process in the plane parallel to the terrestrial

[1]Interdisciplinary Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, North Avenue, Atlanta, GA 30332, USA. [2]School of Physics, Georgia Institute of Technology, 837 State St NW, Atlanta, GA 30332, USA. [3]Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, 801 Atlantic Dr NW, Atlanta, GA 30332, USA. [4]School of Mathematics, Georgia Institute of Technology, 686 Cherry St NW, Atlanta, GA 30332, USA. *Corresponding author. Email: daniel.goldman@physics.gatech.edu

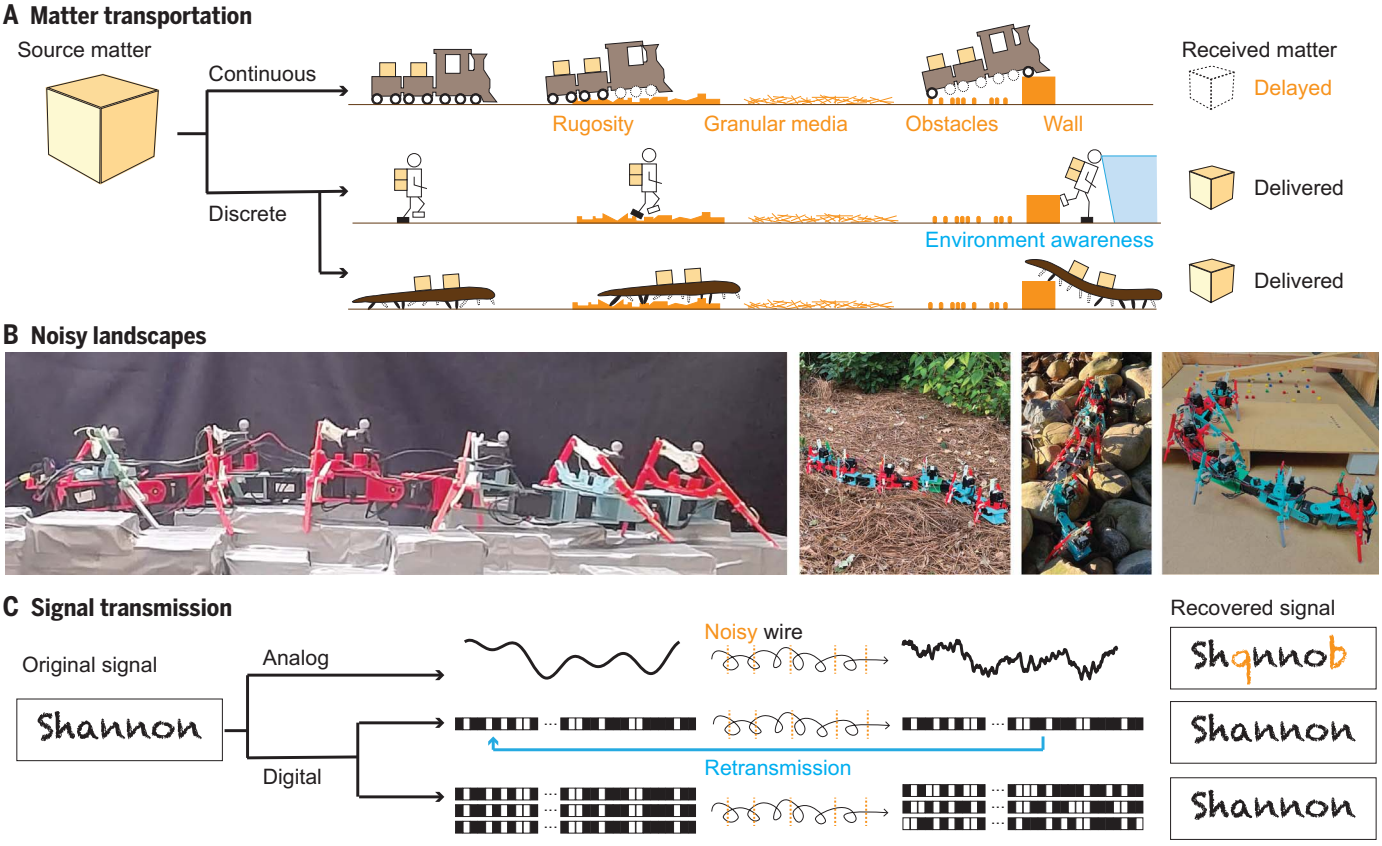


Fig. 1. Signal transmission and matter transport. (**A**) Matter transport with either continuous or discrete active contacts can be effective on "noise-free" tracks. Discrete redundant contacts enable effective matter transport over rugose tracks through redundancy or environmental awareness. (**B**) Multi-segmented robophysical locomotors with directionally compliant legs traverse noisy landscapes: (from left to right) a laboratory model of rugose terrain, entangled granular media, boulders, and steps. (**C**) The transmission of analog and digital signals through noisy wires. A digital signal allows reliable transmission through a noisy wire through either redundancy or a retransmission channel.

surface (e.g., a stumble) (*25*). Parallel thrust disturbances can be minimized by proper design of mechanical structures or passively compliant mechanisms [supplementary materials (SM), section 1.2] (*16*, *26*). In this study, we focus on a class of noisy landscapes (rugose terrains) in which the height distribution, $h(x,y)$, can affect the supporting force distribution (e.g., missing steps) in the direction perpendicular to the terrestrial plane and therefore contaminate the intended bac sequence ($X^m \rightarrow Y^m$).

With the notion of bacs and contact noise established, we can now model matter transport as a stochastic process. We first consider an abstract characterization of thrust generation if given one pair of legs. We quantify the instantaneous thrust over a bac, $f(t)$, as the instantaneous external force required to keep the locomotor in place at time $t \in [0, \tau)$, where τ is the duration of the bac. An example of thrust function $f(t)$ is illustrated in Fig. 3A. The nominal (undisturbed on flat terrain) average thrust is $f_n = \frac{1}{\tau}\int_0^\tau f(t)dt$.

Next, we introduce a coefficient function which encapsulates the uncertainty in the bac, $c(t)$. The terrain-disturbed thrust can be formulated by $\hat{f} = \frac{1}{\tau}\int_0^\tau c(t)f(t)dt$. We assume $c(t)$ has the property $\frac{1}{\tau}\int_0^\tau c(t) = 1$ so that the supporting force balances gravity. Further, we assume that the initiation of a bac is delayed by some time, c_1, and the duration of a bac is shortened to $\tau_u : \{c(t) = 0, t \notin [c_1, c_1 + \tau_u]\}$. Specifically, we assume c_1 to be a random variable having a uniform distribution of $c_1 \sim U(0, \tau)$, and the duration of the bac, τ_u, is assumed to be a random variable determined by the terrain rugosity. We sample τ_u from the cumulative distribution function given by $G(\tau_u) = (1-b)\tau_u/\tau + b, \tau_u \in [0, \tau]$ so that there is a finite probability of complete bac loss, $p(\tau_u = 0) = b$, and $b < 1$ characterizes the contact noise level and offers an approximation to the rugosity of the terrain (Fig. 3C). Whenever $c_1 + \tau_u > \tau$, we extend the excessive contact duration $(c_1 + \tau_u - \tau)$ into the next bac [Fig. 2B (iii)] (SM, section 1.4). For simplicity, we assume that $c(t)$ is otherwise uniformly distributed during the bac: $\{c(t) = \tau_u^{-1}\tau, t \in [c_1, c_1 + \tau_u]\}$. In this way, the terrain-disturbed average thrust reduces to $\hat{f} = sign(\tau_u)\tau_u^{-1}f_u$, where $f_u = \int_{c_1}^{c_1+\tau_u} f(t)dt$ is the thrust disturbance. The sign function $sign(\tau_u)$ implies that no thrust will be generated $(\hat{f} = 0)$ with complete bac loss (τ_u = 0).

As demonstrated in previous studies of dissipation-dominated multilegged locomotion on flat ground (*27*, *28*), because of the periodic limb lifting and landing, an effective viscous (rate-dependent) cycle-averaged thrust-velocity (the average thrust and velocity over a period, respectively) relationship emerges in frictional environments, despite such thrusts being instantaneously independent of velocity. Specifically, the relationship of cycle-averaged robot locomotion velocity is derived to be linearly correlated with the cycle-averaged thrust: $\hat{v} = \gamma^{-1}\hat{f}$, where γ is the effective viscous drag coefficient (Fig. 3D). In this way, the terrain-disturbed velocity can be approximated by $\hat{v} = \gamma^{-1}sign(\tau_u)\tau_u^{-1}f_u$.

Taking the analogy from information theory in which redundant bits can bound the uncertainty from channel noise, we hypothesize that locomotors with redundant bacs can offer robustness over terrain uncertainty. A straightforward scheme to include redundancy is to decrease the transport rate by allowing more

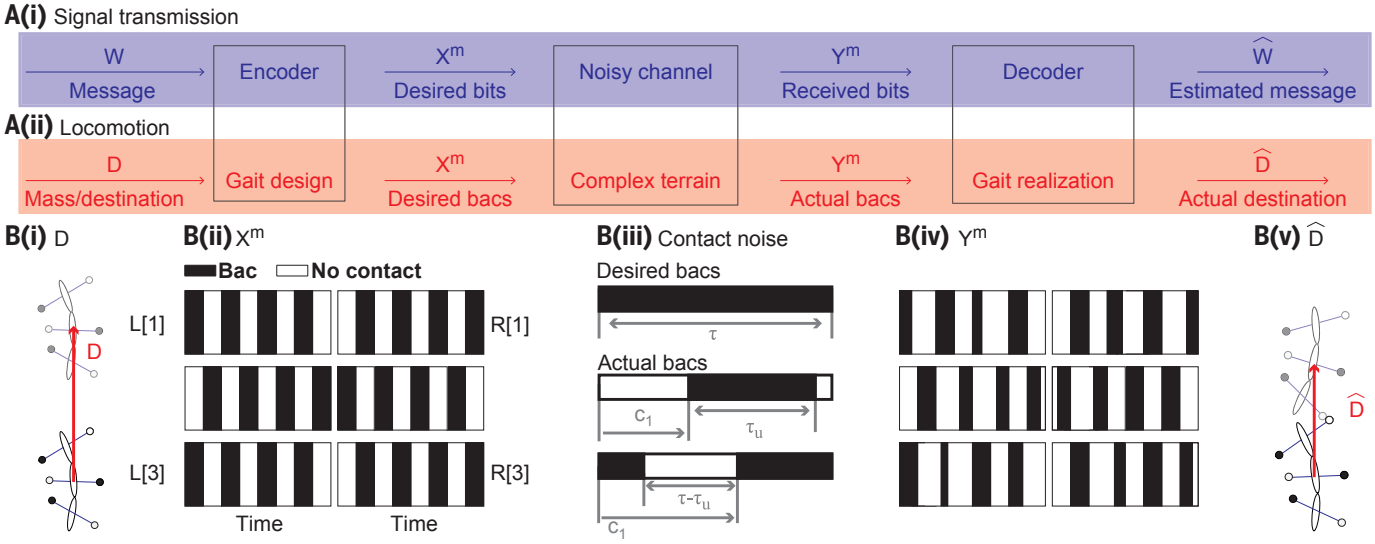


Fig. 2. Framing the matter-transport problem as a sequence of basic active contacts (bacs). (**A**) The correspondence of processes in (**i**) signal transmission [adapted from (*21*)] and (**ii**) locomotion (matter transport). (**B**) (**i**) A multilegged robot, the matter to be transported to a destination *D*. (**ii**) The desired bac sequence to reach the locomotion destination. (**iii**) Noisy landscapes can introduce contact errors such as delaying bacs and shortening the duration of bacs. We compare the desired bac (which spans a duration τ) and two terrain-contaminated bacs (each begins at c_1) with shorter duration (τ_u). (**iv**) A bac sequence contaminated by contact errors leads to a (**v**) locomotion destination $\hat{D}$ smaller than the expected *D*.

transport time (temporal redundancy). Thus, we have:

$$\hat{v}_T^{[1]} = \frac{1}{\gamma T}\sum_{i=1}^{T} sign(\tau_u^i)\frac{f_u^i}{\tau_u^i} \quad (1)$$

where $\hat{v}_T^{[1]}$ is the average terrain-disturbed velocity over T periods, τ_u^i and f_u^i are the contact and thrust disturbance, respectively, over the ith period. Here, T represents the order of temporal redundancy. We expect the variance of the average terrain-disturbed velocity, $\sigma^2\left(\hat{v}_T^{[1]}\right)$, to decrease as T increases. Further, $\hat{v}_T^{[1]}$ converges to a Dirac delta function as T approaches infinity (proof given in the SM, proposition 3). Moreover, the expected average terrain-disturbed velocity, $\left\langle \hat{v}_T^{[1]} \right\rangle$, remains constant (by the law of large numbers).

We next evaluate the effectiveness of this temporal redundancy scheme. For simplicity, we assume D is one-dimensional, $\hat{D}_T^{[1]} = T\hat{v}_T^{[1]}$, and we note that there is no reason to expect that $\hat{D}_T^{[1]}$ should converge (to D) as T increases. Therefore, temporal redundancy can only guarantee the completion of a matter-transport task, but the exact transport duration can be indefinite.

Given the inefficiency of temporal redundancy, we develop a framework, analogous to Shannon's encoding scheme, to remove inefficient redundancy and compensate it with "redundancy of the right sort" [(*29*) p. 164] for more effective locomotion. In particular, the appropriate redundancy facilitates the simultaneous "communication" (e.g., redistribution) of bacs in response to contact noise. To develop a specific scenario for legged systems, we consider redundancy in the form of repeating serially-connected modules, in which a module is defined as a pair of legs. With proper coordination, the effect of contact noise will be shared among all bacs instead of acting on an individual bac. Because such redundancy is distributed in the spatial domain, we refer to it as spatial redundancy. Effectively, this spatial redundancy serves as a moving average filter over the contact-noise profile. For simplicity, we consider a simple module coordination that the instantaneous thrust $f(t)$ on each module is identical and invariant to the number of modules. The average terrain-disturbed velocity for N serially connected modules over T periods is:

$$\hat{v}_T^{[N]} = \frac{1}{\gamma T}\sum_{i=1}^{T}\left(sign\left(\sum_{j=1}^{N}\tau_u^{ij}\right)\frac{\sum_{j=1}^{N} f_u^{ij}}{\sum_{j=1}^{N}\tau_u^{ij}}\right) \quad (2)$$

where τ_u^{ij} and f_u^{ij} are disturbances on the jth module over the ith temporal repetition. Intuitively, in the case where there are M complete bac losses in the ith temporal repetition, $M = \left|\left\{j, \tau_u^{ij} = 0\right\}\right|$, the locomotor with N modules will essentially reduce to the configuration with $N - M$ modules. In other words, locomotors with spatial redundancy N can afford up to $N - 1$ complete bac losses without substantial thrust deficiency, which indicates that spatial redundancy can also serve to bound the uncertainty in thrust generation. We show that for fixed $T \geq 1$, $\hat{v}_T^{[N]}$ will also converge to a Dirac delta function as spatial redundancy N approaches infinity (SM, proposition 4). The expected average terrain-disturbed velocity, $\left\langle \hat{v}_T^{[N]} \right\rangle$, can be approximated by $(1 - b^N)C_s$, where C_s is a constant determined by $f(t)$, γ, and b (SM, proposition 4). Therefore, greater spatial redundancy not only reduces variance, but also improves the expected average terrain-disturbed velocity, a feature otherwise not possible with only temporal redundancy.

For any fixed T, the distribution of actual destinations $\hat{D}_T^{[N]} = T\hat{v}_T^{[N]}$ will also converge to a Dirac delta function as N approaches infinity. Given a matter-transport task over desired distance D at scheduled time T, we consider it successful if

$$\left\|\left(\hat{D}, \hat{T}\right) - (D, T)\right\| = \left|\hat{D}_T^{[N]} - D\right| < \varepsilon \quad (3)$$

where ε is the tolerance. In this way, for arbitrary $\varepsilon > 0$ and $p_0 < 1$, there exists a finite N such that the probability of successful matter transport (subject to ε) is greater than p_0 (proof given in the SM, proposition 5). The "minimal spatial redundancy" required to achieve the probability of p_0 for tolerance $\varepsilon > 0$ is bounded by $N_{\varepsilon,p_0} \leq \frac{\log(1-\sqrt[kD]{p_0})}{\log(b)}$, where k is a constant that describes the locomotor speed on flat terrain, and D is the desired destination specified by distance, for one-dimensional locomotion (proof given in the SM, propositions 5 and 6).

Numerical tests of the framework

We first tested our theoretical predictions by numerical simulations. We chose a saptiotemporal

bac distribution pattern that was based on limb-stepping patterns of biological centipedes [and whose efficacy in generating locomotion in multilegged robots was previously studied in (*27, 30*)]. Each bac generates an instantaneous thrust given by $f(t)$, which is independent of our choice of spatial redundancy (proof given in the SM, proposition 1). We illustrate $f(t)$ in Fig. 3B. Assuming $b = 0.5$, we compared the distribution of normalized (by the nominal speed, v_{open}) terrain-disturbed velocity, v/v_{open}, for different combinations of temporal and spatial redundancy in Fig. 3E. In this example, $C_s = v_{open}$ indicates that the terrain-disturbed velocity converges to the nominal velocity (dashed black curve) given sufficient spatial redundancy.

To illustrate that spatial redundancy allows the matter-transport process (defined by the actual distance achieved, $\hat{D}$, and duration to achieve this displacement, $\hat{T}$) to converge to the desired transport-distance and duration pair (D, T), we calculate the 90% confidence interval (CI) of $\hat{D}$ as a function of T for $N = 8$ (red curves) and $N = 1$ (blue curves) (Fig. 3F). With greater spatial redundancy, the variance of $\hat{D}$ is substantially reduced, approaching the limit of nominal distance-duration relationship (dashed black curve) (Fig. 3F). Further, we numerically calculate the probability of successful matter transport (evaluated at $T = 1$) as a function of spatial redundancy N subject to different choices of tolerance ε (Fig. 3G). For all choices of ε, the success probability converges to one as N increases. Figure 3G also shows that the marginal benefit of having more legs decreases as N increases; the predicted limit is given by the dashed black curve. Finally, the locomotion performance can be affected by the terrain rugosity. We calculated N_{ε,p_0} as a function of b, the contact noise level, and plot this in Fig. 3H. We compare three combinations of ε and p_0 and notice that more rugose terrains require greater (yet finite) N_{ε,p_0}; the theoretical bound is given by the dashed black curve when $p_0 = 0.9$.

Experimental tests of the framework

Because the model assumes simplified environmental interactions, we next tested if our framework could predict locomotor performance in a physical system. We chose to work with a well-controlled laboratory multilegged robophysical model (movie S1) (design and control details are provided in the SM, sections 1.2 and 1.3), similar in design to those in (*27, 30, 31*). We measured the average speed and variance in such robots with different leg numbers and terrain complexity. Given that our framework indicates that matter transport can be achieved without the need for environmental awareness, we controlled all robots such that they executed their pre-programmed stepping patterns (open loop) and did not sense or respond to features of the environment. To facilitate comparison across different spatial redundancies N, our chosen bac sequence (the same as that in the numerical tests) has the property that, in theory, all robots share the same thrust function $f(t)$, the same performance on flat terrain, and the same thrust-velocity relationship (proof given in the SM, proposition 2).

We constructed laboratory models of rugose terrains composed of (10×10 cm^2) blocks with variation in heights (Fig. 4A). The block heights, $h(x,y)$, are randomly distributed (SM, section 1.1). Such rugose terrains ensure that limbs will experience thrust deficiency from

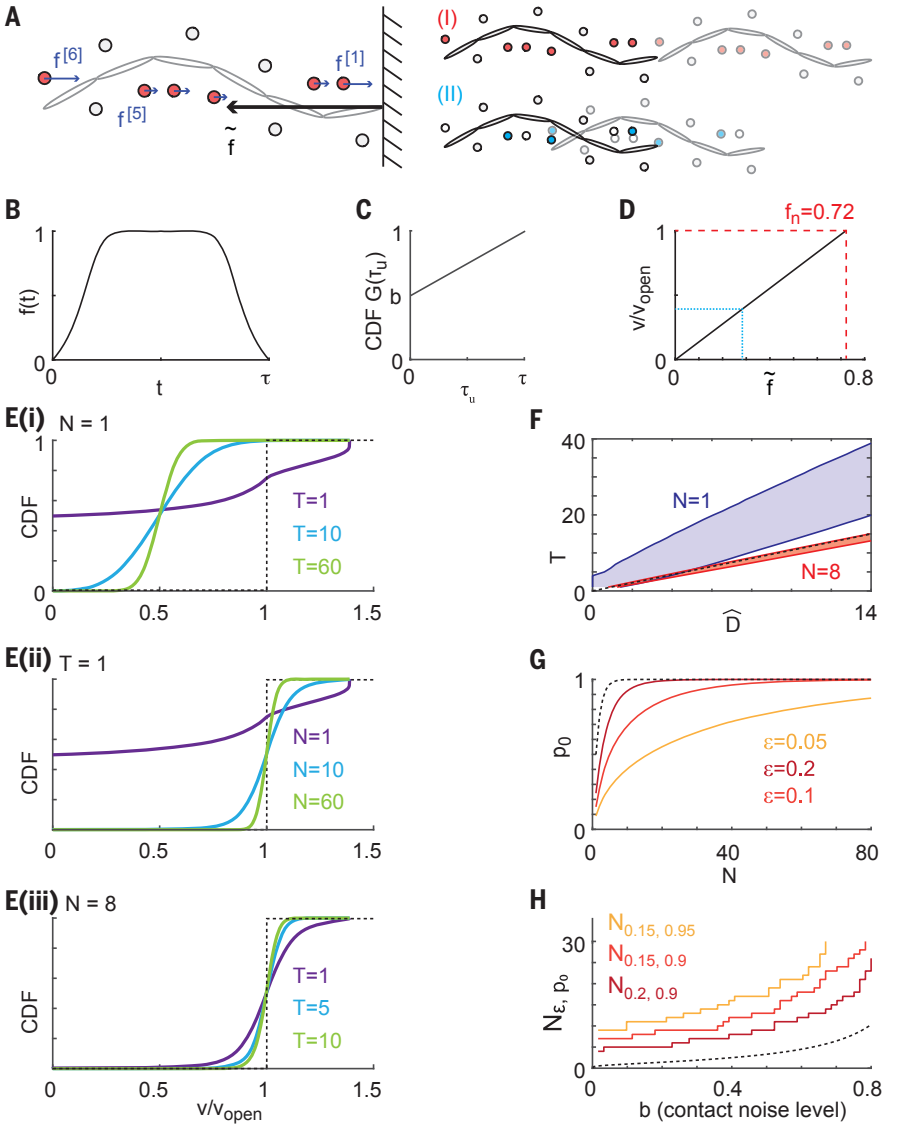


Fig. 3. Numerical simulation to test the matter-transport framework. (**A**) Illustrations of (left) thrust generation from bacs and (right) the thrust-velocity relationship. Self-propulsion with (I) nominal contact and (II) contact errors are compared. (**B**) The instantaneous thrust $f(t)$ as a function of time, derived from (*28*). (**C**) The cumulative distribution function (CDF) of τ_u. (**D**) The thrust-velocity (normalized by the nominal velocity) relationship. (**E**) The numerical CDF of terrain-disturbed velocity for robots with different combinations of temporal (T) and spatial (N) redundancy. (i to iii) Black dashed curves indicate the nominal CDF. (**F**) Numerically calculated 90% CI of actual destination $\hat{D}$ as a function of T. We compare two spatial redundancies: $N = 1$ (blue) and $N = 8$ (red). The nominal $\hat{D}$-T relationship is shown with the black dashed curve. (**G**) The probability of successful scheduled matter transport (p_0) as a function of spatial redundancy evaluated at different tolerance ε. We plot the theoretical predicted limit as a black dashed curve. (**H**) The relationship between N_{ε,p_0}, the minimal spatial redundancy required to achieve the desired success probability (p_0) subject to tolerance ε, and b, contact noise level. The black dashed line indicates the theoretical predicted limit when $p_0 = 0.9$.

stochastic contact (*32*). The contact error can also arise from robot motor noise because of actuation delay, insufficient torque, or body compliance. We define the terrain rugosity, R_g, as the standard deviation of heights normalized by block side length. We tested the performance of 3-8 segmented [each segment has two directionally compliant limbs (*31*)] robophysical models on rugose terrains and recorded the bac duration (τ_u) on each leg. The distributions of τ_u measured from 225 and 309 bacs for terrain with rugosity 0.17 and 0.32, respectively, are shown in fig. S5. For $0 < \tau_u < \tau$, the measured cumulative distribution function of τ_u can be approximated by a linear function, a feature in accord with (and therefore, justifying) our assumed τ_u bac duration distribution in Fig. 3B.

We also recorded the normalized interstep average velocity (the average velocity over a step that spans the duration of a bac) of the 12-legged robot on rugose terrains during multiple steps (v/v_{open}) and compared this with the corresponding normalized bac duration (τ_u/τ) (Fig. 4B). Specifically, in each trial (five trials on each terrain), we programmed the robot to run for two periods such that there was at least one bac generated by each leg (6 legs visualizable from side-view camera). The correlation in these quantities (Fig. 4B) indicates that bac contamination is an important source of locomotor speed variability, which is in accord with our model assumption. Further, we measured the distributions of average v/v_{open} with different combinations of temporal and spatial redundancy (Fig. 4D). Specifically, we measured the average displacement (over T cycles; T is specified in Fig. 4D) of a 6-legged and a 12-legged robot on each terrain. The measured cumulative distribution functions of average v/v_{open} were in qualitative agreement with predictions from numerical tests (Fig. 3E). The sources for discrepancies between numerical and experimental tests include over-simplification of terrain characterization and contact error from robot motor noise.

As discussed earlier, for a given transport distance D, the average and the variance of transport duration are both important metrics for evaluating matter-transport performance. We measured the transport duration (in units of numbers of periods) as a function of D for a 6-legged robot and a 14-legged robot over a rugose terrain (R_g = 0.32). Similar to our numerical prediction in Fig. 3F, experimental results also show that spatial redundancy reduces the variance (illustrated by error bars) in transport duration (Fig. 4C). Thus, we can guarantee predictable and reliable speed on a noisy terrain during open-loop noninertial matter transport, analogous to that of the predictability of inertia-driven dynamics on noiseless (e.g., rails and roads) terrain.

To determine how spatial redundancy affects transport duration, we recorded $T_{[D=60]}$, the time required for a robot to locomote 60 cm over rugose terrains, for robots with varying number of legs on R_g = {0, 0.17, 0.32} (10 trials in each condition) (Fig. 4E). A hexapod can eventually self-transport 60 cm, but there is a large variance in $T_{[D=60]}$. By contrast, systems with 16 legs can finish the self-transport task with short average time and small variance. Moreover, for systems with sufficiently high spatial redundancy (e.g., $N \geq 5$), further increases in N do not result in substantial changes in $T_{[D=60]}$, including both the average and the

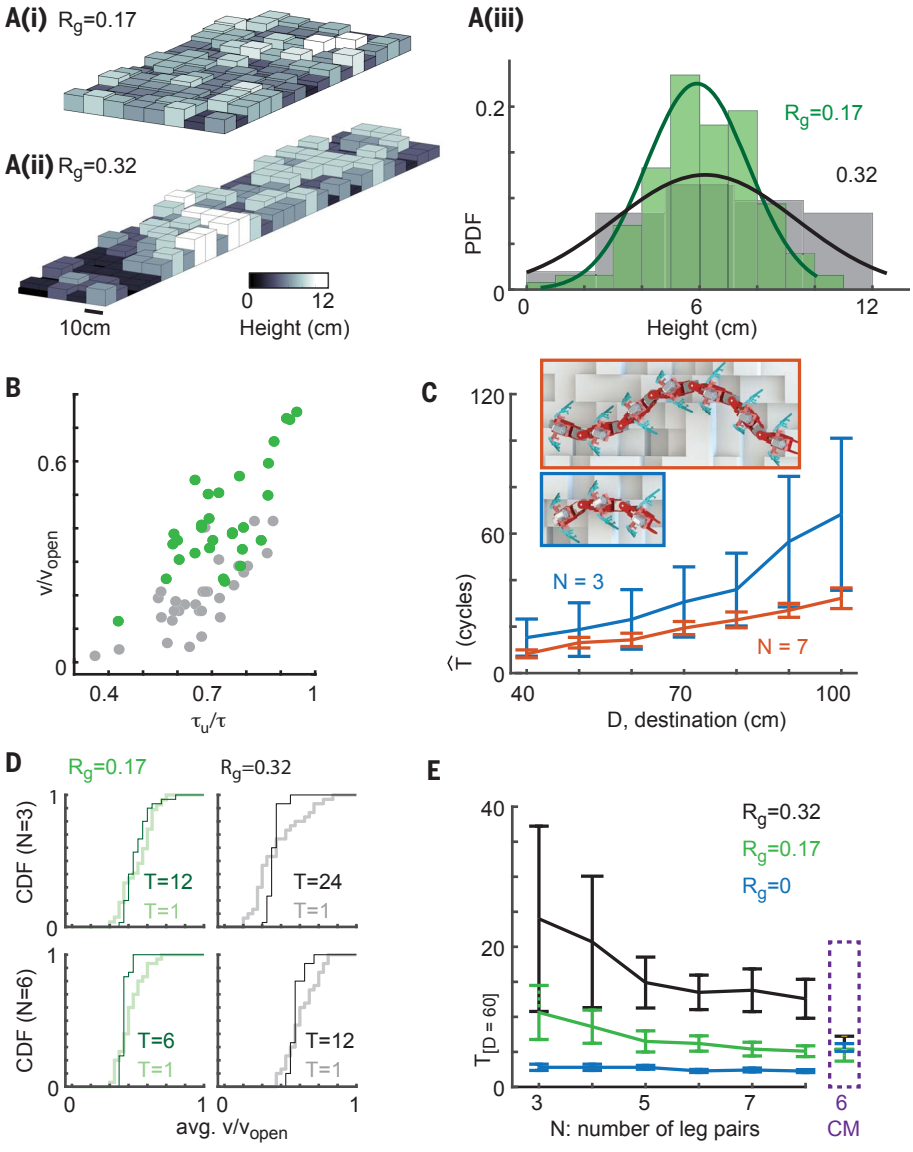


Fig. 4. Experimental test of the matter transport framework on multilegged robots. (**A**) Renderings of rugose terrains with rugosity (**i**) R_g = 0.17 and (**ii**) R_g = 0.32. The block-height distributions are shown in (**iii**). (**B**) Bac contamination leads to speed degradation in a 12-legged robot on rugose terrains. Each point denotes the robot's normalized interstep averaged velocity (v/v_{open}) and a corresponding contact duration (τ_u/τ) in one bac. Color schemes are identical to those in (A). (**C**) The empirical transport duration ($\hat{T}$, in units of gait periods) as a function of destination distance D. We compare two robots with 14 (red) and 6 (blue) legs (renderings shown as insets). There is a large variance in transport duration for the six-legged robot, and the variance grows as destination distance increases. The 14-legged robot has a comparably tightly bounded transport duration. (**D**) The empirical distribution of average velocity on terrains with (left) R_g = 0.17 and (right) R_g = 0.32 on (Top) the six-legged robot and (Bottom) the 12-legged robot. The empirical distributions were obtained from 30 trials. (**E**) For robots with different numbers of leg pairs N, we recorded $T_{[D=60]}$, the number of periods required to transport D = 60 cm on terrains with (blue) R_g = 0, (green) R_g = 0.17, and (black) R_g = 0.32. The error bar was calculated from at least 10 trials. $T_{[D=60]}$ for contact-modulated gaits are illustrated in the dashed purple rectangle.

variance. This finding is consistent with our numerical prediction of the marginal benefit from increases in spatial redundancy after a large enough N (Fig. 3G).

From the Shannon scheme for signal transmission, it is reasonable to anticipate improved performance with more elaborate coding schemes, which we define in matter transport as designing the terradynamic interaction profile of bacs (e.g., the instantaneous thrust function $f(t)$ and thrust-velocity relationship). One straightforward method to modulate the bac-interaction profile is to change the body wave amplitude (*27*) or the number of waves on the body (*33*), which alters the temporal and spatial distribution of bacs and the associated body postures. To illustrate the potential of appropriate coding, we show one example of gait design modulation. Specifically, we imposed a head-to-tail vertical travelling wave along the body with twice the spatial frequency as the horizontal wave. This had the effect that the duration of bacs (τ) was actively and systematically shortened (contact modulation; SM, section 1.5). We tested the performance of contact-modulated (CM) multilegged robots over our rugose terrains and observed improved locomotion robustness over terrain rugosity (indicated by smaller error bars), although with some reduction of the nominal velocity v_{open} for locomotion on flat ground (Fig. 4E, dashed purple rectangle). Furthermore, with sufficient spatial redundancy ($N = 6$) as well as contact modulation, our multilegged robot was capable of traversing diverse laboratory (obstacles, inclines, and walls) environments (Fig. 1B and movie S1) and field-like environments (granular media, pebbles, and rock piles) with completely open-loop operations.

Discussion

One value of our framework lies in its codification of the benefits of redundancy, which lead to locomotor robustness over environmental contact errors without requiring sensing. This contrasts with the prevailing paradigm of contact-error prevention in the conventional sensor-based closed-loop controls that take advantage of visual, tactile, or joint-torque information from the environment to change the robot dynamics (*5*, *14*). In this way, the complexity of matter transport can be transferred from the real-time feedback-based control (e.g., dealing with the flow of sensor information) to preprogrammed gait design. Thus, our framework could simplify matter transport tasks such as search-and-rescue (*34*), extraterrestrial exploration (*35*), or even microrobotics (*36*), in which robot deployments are often preferred but are challenging because of unpredictable terradynamic interactions and unreliable sensors.

However, sensory feedback can clearly be of value when a robot becomes "stuck," when terradynamic interactions vary substantially (e.g., moving from low to high rugosity terrain) or when large spatial redundancy is undesirable. In such cases, the robot must understand its state (through proprioception or exteroception) (*13*) and change dynamics accordingly. An analogy from information theory, error-detection coding, facilitates the augmentation of our framework with sensory-based control: in so-called error-detection coding, the presence of a "reverse channel" can facilitate the retransmission of signals, thereby improving signal transmission accuracy (*21*, *29*). The introduction of feedback can increase the capacity of a noisy channel and can therefore decrease the coding complexity (*37*). We posit that a sensor-based framework in locomotion for a fixed number of bacs shares a mechanism similar to error-detection coding. If performance improvement with increased redundant bacs becomes marginal, then sensors (e.g., monitoring foot contact) that detect bac contamination could facilitate rapid environmental adaptation (e.g., whole-body gait adaptation or local leg placement adaptation) and further improve locomotion performance. Thus, a combination of redundancy-based and sensor-based mechanisms can offer unique advantages over challenging terrains, a feature similar to hybrid automatic repeat-request method (hybrid ARQ) in signal transmission (*38*).

In addition to the importance of locomotion in artificial locomotors, we posit that our matter-transport framework can give insights into aspects of neuromechanical and morphological evolution (*39*) from a physics of living systems perspective. That is, animals ranging from those which generate propulsion through a single bac-pair (i.e., bipeds) (*40*, *41*), to those which use many bacs (i.e., myriapods) (*42*), are capable of traversing complex natural terrains. The importance of environmental awareness and whole-body coordination is hypothesized to diminish as the number of bacs (redundancy) increases (*42*–*44*). Thus, in biological terrestrial locomotors, there appears to be a shift toward either advanced neuromechanical control with reduced body appendages, or redundant body appendages with simplified neuromechanical control. Integration of our framework with advances in biological experimentation could yield insights into the benefits and trade-offs of diverse control architectures.

REFERENCES AND NOTES

1. J. Aguilar *et al*., *Science* **361**, 672–677 (2018).
2. J. Mindell, H. Rutter, S. Watkins, *Encyclopaedia of Environmental Health* (Elsevier Masson, 2011), pp. 578–589.
3. H. Kreutzmann, *Mod. Asian Stud.* **25**, 711–736 (1991).
4. R. M. Alexander, Principles of animal locomotion (Princeton Univ. Press, 2003).
5. P. Holmes, R. J. Full, D. Koditschek, J. Guckenheimer, *SIAM Rev.* **48**, 207–304 (2006).
6. M. Raibert, K. Blankespoor, G. Nelson, R. Playter, *IFAC Proceedings Volumes* **41**, 10822 (2008).
7. U. Saranli, M. Buehler, D. E. Koditschek, *Int. J. Robot. Res.* **20**, 616–631 (2001).
8. M. G. Bekker, *Off-the-road locomotion: Research and development in terramechanics* (Univ. of Michigan Press, 1960).
9. C. Li, T. Zhang, D. I. Goldman, *Science* **339**, 1408–1412 (2013).
10. C. Li *et al*., *Bioinspir. Biomim.* **10**, 046003 (2015).
11. W. Weber, E. Weber, *Mechanik der menschlichen Gehwerkzeuge* (Verlag der Dietrichschen Buchhandlung, Göttingen, Germany, 1836).
12. J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, M. Hutter, *Sci. Robot.* **5**, eabc5986 (2020).
13. J. Borenstein, H. R. Everett, L. Feng, D. Wehe, *J. Robot. Syst.* **14**, 231–249 (1997).
14. M. Hutter *et al*., *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (IEEE, 2016), pp. 38–44.
15. E. R. Westervelt, J. W. Grizzle, C. Chevallereau, J. H. Choi, B. Morris, *Feedback control of dynamic bipedal robot locomotion* (CRC press, 2018).
16. Y. Ozkan-Aydin, D. I. Goldman, *Science Robotics* **6**, eabf1628 (2021).
17. S. Aoi *et al*., *Sci. Rep.* **6**, 30199 (2016).
18. E. Moore, D. Campbell, F. Grimminger, M. Buehler, *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, Washington, DC, 11 to 15 May, 2002 (IEEE, 2002), vol. 3, pp. 2222–2227.
19. C. Ordonez, J. Shill, A. Johnson, J. Clark, E. Collins, *Proc SPIE Int Soc Opt Eng* **8741**, 240–251 (2013).
20. X. A. Wu, T. M. Huh, R. Mukherjee, M. Cutkosky, *IEEE Robot. Autom. Lett.* **1**, 1125–1132 (2016).
21. C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379–423 (1948).
22. A. Feinstein, *IEEE Trans. Inf. Theory* **4**, 2– 22 (1954).
23. R. J. Full, D. E. Koditschek, *J. Exp. Biol.* **202**, 3325–3332 (1999).
24. H. C. Astley *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6200–6205 (2015).
25. H. Forssberg, *J. Neurophysiol.* **42**, 936–953 (1979).
26. P. E. Schiebel *et al*., Robophysical modeling of bilaterally activated and soft limbless locomotors, *Biomimetic and Biohybrid Systems: 9th International Conference, Living Machines*, Freiburg, Germany, 28 to 30 July, 2020 (Springer International Publishing, 2020), pp. 300–311.
27. B. Chong *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2213698120 (2023).
28. D. Zhao, B. Bittner, G. Clifton, N. Gravish, S. Revzen, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113222119 (2022).
29. J. R. Pierce, *An Introduction to Information Theory: Symbols, Signals and Noise* (Courier Corporation, 2012).
30. B. Chong *et al*., *Bioinspir. Biomim.* **17**, 046015 (2022).
31. Y. Ozkan-Aydin, B. Chong, E. Aydin, D. I. Goldman, *2020 3rd IEEE International Conference on Soft Robotics (RoboSoft)*, New Haven, CT, USA, 15 May to 15 July, 2020 (IEEE, 2020), pp. 156–163.
32. A. Jacoff, A. Downs, A. Virts, E. Messina, *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, Gaithersburg, MD, USA, 19 to 21 August, 2008 (NIST, 2008), pp. 29–34.
33. B. Chong *et al*., *Int. J. Robot. Res.* **40**, 1547–1562 (2021).
34. J. Whitman, N. Zevallos, M. Travers, H. Choset, *2018 IEEE International Symposium on Safety, Security, and Rescue rRobotics (SSRR)*, Philadelphia, PA, USA, 6 to 8 August, 2018 (IEEE, 2018), pp. 1–6.
35. N. D. Naclerio *et al*., *Sci. Robot.* **6**, eabe2922 (2021).
36. J. J. Abbott, Z. Nagy, F. Beyeler, B. J. Nelson, *IEEE Robot. Autom. Mag.* **14**, 92–103 (2007).
37. S. Tatikonda, S. Mitter, *IEEE Trans. Inf. Theory* **55**, 323–349 (2008).
38. C. Lott, O. Milenkovic, E. Soljanin, *2007 IEEE Information Theory Workshop on Information Theory for Wireless Networks*, Bergen, Norway, 1 to 7 July, 2007 (IEEE, 2007), pp. 1–5.
39. B. Chong, T. Wang, E. Erickson, P. J. Bergmann, D. I. Goldman, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2118456119 (2022).
40. D. S. Marigold, *Exerc. Sport Sci. Rev.* **36**, 145–151 (2008).
41. T. Krasovsky, A. Lamontagne, A. G. Feldman, M. F. Levin, *Gait Posture* **39**, 378–385 (2014).
42. K. Diaz, E. Erickson, B. Chong, D. Soto, D. I. Goldman, *J. Exp. Biol.* **226**, jeb244688 (2023).
43. S. Sponberg, R. J. Full, *J. Exp. Biol.* **211**, 433–446 (2008).
44. B. Bläsing, H. Cruse, *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **190**, 173–183 (2004).

45. D. Soto, B. Chong, Code for rugose terrain generation, Zenodo (2021); https://doi.org/10.5281/zenodo.7121219

ACKNOWLEDGMENTS

We thank T. Murphey, P. Umbanhowar, G. A. Sartoretti, D. Luo, T. Berrueta, E. Flores, and X. Sun for helpful discussion. We thank K. Diaz for proofreading. **Funding:** The authors received funding from NSF-Simons Southeast Center for Mathematics and Biology (Simons Foundation SFARI 594594), Georgia Research Alliance (GRA.VL22.B12), Army Research Office (ARO) MURI program, Army Research Office grant W911NF-11-1-0514, and a Dunn Family Professorship. **Author contributions:** B.C. designed the study and model and performed numerical simulations. J.H. collected the raw data for the robophysical experiments. J.H., T.W., D.S., and B.C. built the experimental platform. D.I., G.B., and B.C. performed the mathematical proof. D.I.G. oversaw the study. All authors contributed to the preparation of the manuscript and were involved in the interpretation of results. **Competing interests:** Some of the subject matter herein may be implicated in one or more pending patent applications such as PCT Patent Application No. PCT/US2022/043362, entitled "DEVICES AND SYSTEMS FOR LOCOMOTING DIVERSE TERRAIN AND METHODS OF USE", which claims the benefit of priority to US Provisional App. No. 63/243,435 filed 09/13/2021, and US Provisional App. No. 63/318,868 filed 03/11/2022. The authors declare that they have no other competing interests. **Data and materials availability:** All data that support the claims in this manuscript are available on the Zenodo repository (*45*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade4985
Materials and Methods
Supplementary Text
Figs. S1 to S6
Movie S1
References (*46*, *47*)

Submitted 29 September 2022; resubmitted 7 February 2023
Accepted 3 April 2023
10.1126/science.ade4985

MARINE PRODUCTIVITY

Atmospheric nourishment of global ocean ecosystems

T. K. Westberry[1]*, M. J. Behrenfeld[1], Y. R. Shi[2,3], H. Yu[3], L. A. Remer[2,4], H. Bian[2,3]

Over the vast open ocean, vital nutrients for phytoplankton growth in the sunlit surface layer are largely provided through physical transport from deep waters, but some nutrients are also provided through atmospheric deposition of desert dust. The extent and magnitude of dust-mediated effects on surface ocean ecosystems have been difficult to estimate globally. In this work, we use global satellite ocean color products to demonstrate widespread responses to atmospheric dust deposition across a diverse continuum of phytoplankton nutritional conditions. The observed responses vary regionally, with some areas exhibiting substantial changes in phytoplankton biomass, whereas in other areas, the response reflects a change in physiological status or health. Climate-driven changes in atmospheric aerosols will alter the relative importance of this nutrient source.

Earth's atmosphere and ocean ecosystems are intimately linked. A complex milieu of dissolved and particulate organic compounds is mediated by surface ocean plankton and, through wave breaking and bubble bursting, can be injected into the atmosphere, where these compounds influence aerosol and cloud properties and, thus, the radiative budget of the planet (*1*, *2*). In turn, terrigenous matter lofted into the atmosphere over land may be carried out to sea for thousands of kilometers before ultimately being deposited on the surface, where it can stimulate photosynthesis and plankton growth (*3*). To date, documented responses of marine ecosystems to atmospheric nourishment have largely been restricted to anomalously large local events associated with volcanic eruptions (*4*, *5*), wildfires (*6*, *7*), and extreme dust storms (*8*). Even mesoscale iron fertilization experiments, where dissolved iron is deposited locally into the ocean, use high concentrations of nutrient unlikely to be matched during typical atmospheric deposition. Nevertheless, the far more common lower-concentration atmospheric depositions that occur across the global ocean have long been anticipated to have widespread effects on ocean productivity because, in most open regions, phytoplankton division rates are limited by macro- or micronutrient availability (*9*, *10*). Although this global response has previously eluded detection, regional in situ studies have demonstrated long-term variations in ecosystem functioning in response to climate oscillations affecting low deposition rates (*11*). We use a 14-year global time series of modeled dust deposition coupled with satellite ocean color data to reveal consistent, broad-scale responses of marine plankton communities to atmospheric nourishment. In stable, low-latitude ocean regions, the signature of these responses is predominantly expressed as an improvement in the physiological status, or health, of phytoplankton communities, whereas in the more seasonally varying, high-latitude seas, an enhancement in phytoplankton biomass is also often observed. When integrated globally, we estimate that deposition of dust onto the ocean supports 255 Tg of C per year of primary production, which represents 4.5% of the global annual export production. Regional variation in this contribution can be much higher, approaching 20 to 40% of the annually exported particulate carbon flux from the surface ocean. Global warming is anticipated to alter atmospheric dust burdens and deposition patterns (*12*, *13*), making their contributions to the future nourishment of ocean ecosystems uncertain.

Patterns in dust deposition over the ocean correlate strongly with atmospheric transport paths downwind from dust sources (Fig. 1A and fig. S1) (*14*, *15*). The nutrient complement carried in this dust is highly variable and dependent on source region (*16*). Furthermore, chemical transformation during atmospheric transport modifies the bioavailability of the different nutrients in the dust (iron, phosphorus, and nitrogen) (*17*–*19*). Thus, at the time of wet or dry deposition (*20*) on the ocean, the stoichiometry of the different nutrients can vary spatially and temporally (*11*, *18*). Global distributions of surface ocean phytoplankton biomass and chlorophyll concentration (Fig. 1B and fig. S1) exhibit little semblance to these dust deposition patterns (Fig. 1A) because the primary source of nutrients supporting marine ecosystems is through vertical transport from depth (*21*–*23*). Nevertheless, the atmosphere represents a source of additional nutrients to the ocean, and the stoichiometric composition of dust nutrients is sufficiently distinct from vertically transported nutrients that it can affect the dominant nutrient that limits plankton productivity (*11*, *24*). For example, in low-dust regions with substantial upwelling (e.g., the eastern Equatorial Pacific) or deep mixing (e.g., the Southern Ocean), phytoplankton communities tend to be iron limited, at least seasonally (fig. S2). By contrast, tropical and subtropical regions outside upwelling zones—where the surface mixed layer is permanently stratified above the sunlit photic depth [hereafter referred to as the permanently stratified ocean (PSO)]—are generally viewed as being primarily limited by nitrogen or phosphorus (fig. S2). Yet even in these regions, dust deposition can represent an important source of additional nitrogen and phosphorus and, perhaps equally important, can stimulate nitrogen fixation through its supplementation of iron. PSO waters with high dust inputs (e.g., the North Atlantic subtropical gyre) tend toward phosphorus-limited communities rather than nitrogen-limited phytoplankton (fig. S2).

Detecting a widespread influence of dust deposition on plankton communities is challenging because, across much of the open ocean, mean deposition rates are small (e.g., red line in Fig. 1C). Thus, the responses that they elicit are difficult to decipher from other sources of ecological variability (e.g., changes in mixed-layer growth conditions and stochastic processes in marine food webs). However, a distinct feature of long-term atmospheric deposition records is their frequent punctuation by short-term dust events (black spikes in Fig. 1C). Similarly, temporal records of surface chlorophyll concentration typically exhibit repeated annual cycles (red line in Fig. 1D) interrupted by brief departures from these mean patterns (black spikes in Fig. 1D). These transient dust and chlorophyll features may provide an opportunity to distinguish atmospheric deposition effects from other sources of ecosystem variability. To illustrate the impact of ocean nourishment from the atmosphere, we therefore combined 5°-by-10° gridded 4-day-resolution dust deposition data and satellite-observed ocean color products and evaluated ecosystem responses to the largest 10% of deposition events (N = 127) at each grid point between 2003 and 2016 (see materials and methods). These deposition events constitute ~40% of the annual dust mass deposited into the global ocean.

The average mass of dust deposited on the ocean during the top 10% of events varies among our 5°-by-10° bins by nearly five orders of magnitude globally (~10^{-2} to 10^{3} mg m^{-2} day^{-1}) (Fig. 2A) and exhibits a spatial pattern closely corresponding with that for total annual dust deposition (Fig. 1A). In other words, the largest events in low-dust ocean areas are still far smaller than those in high-dust regions.

[1]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA. [2]Joint Center for Earth Systems Technology, University of Maryland, Baltimore County, Baltimore, MD, USA. [3]Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA. [4]Airphoton Inc., Baltimore, MD, USA.
*Corresponding author. Email: toby.westberry@oregonstate.edu

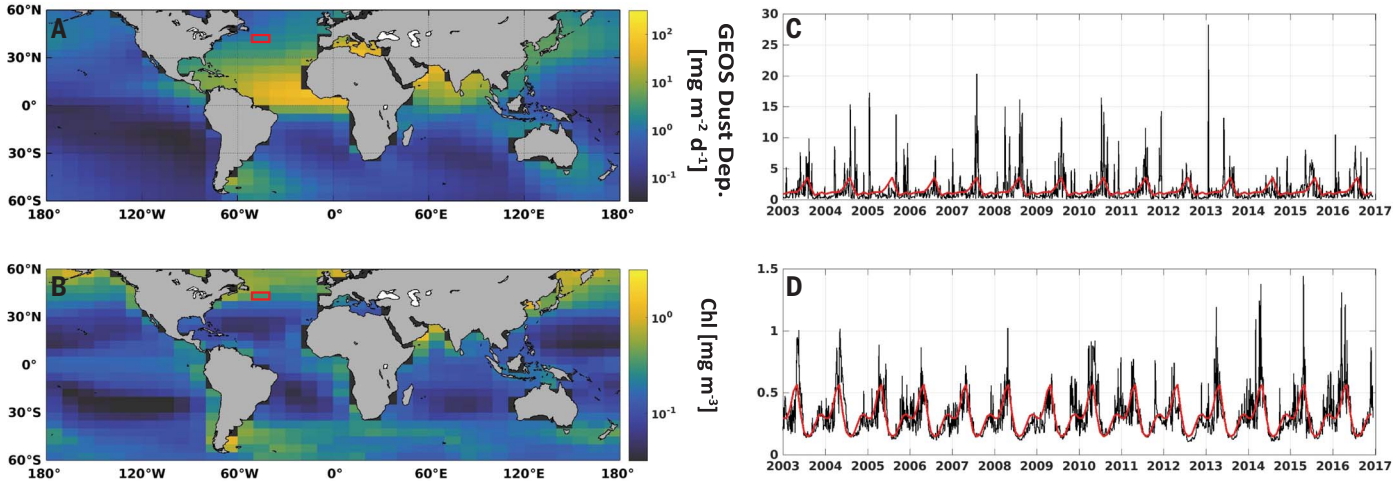
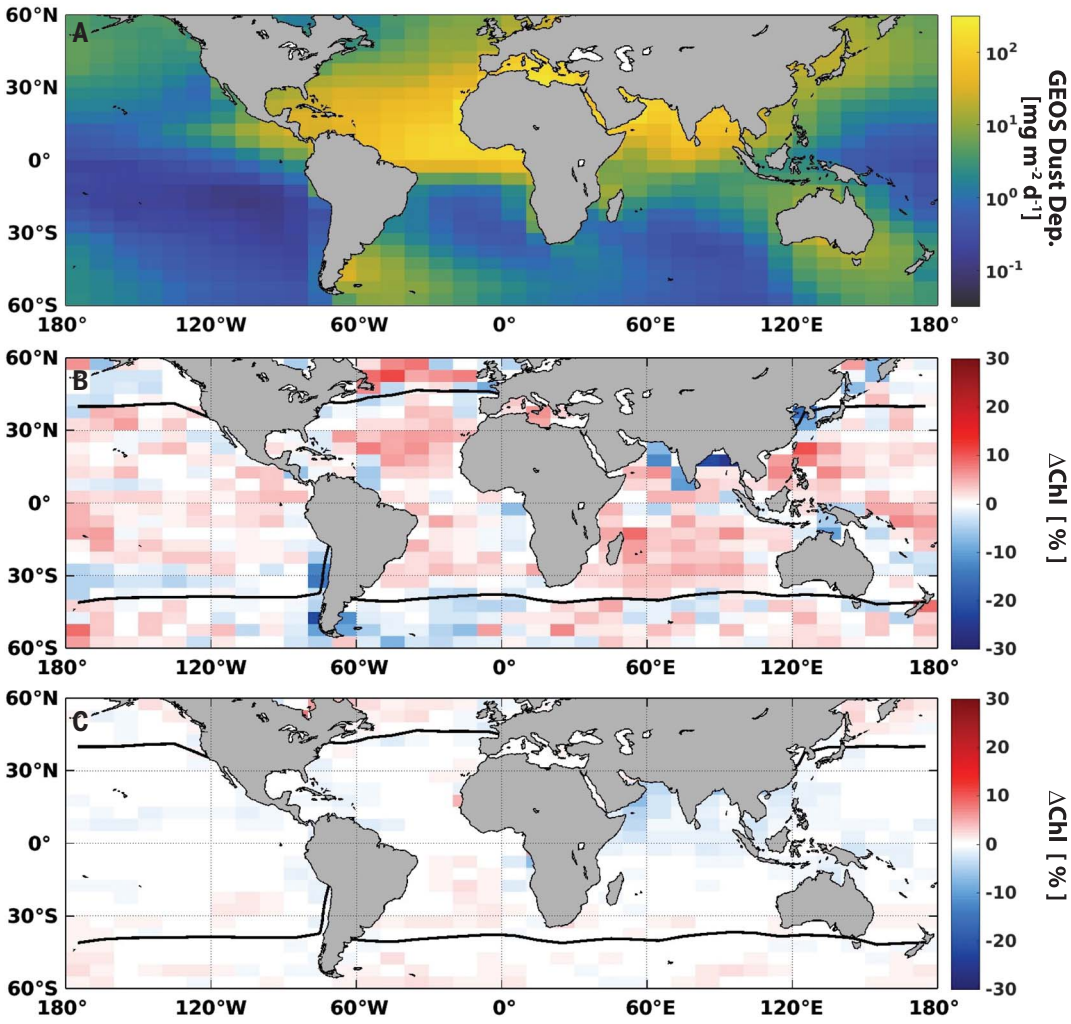


Fig. 1. Model-based dust deposition and satellite-observed chlorophyll concentration over the global ocean. (**A**) Average GEOS dust deposition rates (milligrams per square meter per day) over the period 2003 to 2016. (**B**) Average MODIS-Aqua chlorophyll concentration (milligrams per cubic meter) over the period 2003 to 2016. Red boxes outline a single 5°-by-10° grid cell centered on 42.5°N, 45°W, whose time series are shown in subsequent panels. (**C**) Fourteen-year time series of 4-day average dust deposition (thin black line) over red-outlined region in (A). Also shown is a repeating average annual cycle of dust deposition (solid red line) for the same location. (**D**) Similar to (C), but for MODIS-Aqua chlorophyll concentration (milligrams per cubic meter).

Fig. 2. High dust deposition and subsequent ocean color response. (**A**) Average GEOS dust deposition rate to the surface ocean (milligrams per square meter per day) during top 10% of record (2003 to 2016). (**B**) Median relative change in chlorophyll concentration evaluated between the 4-day period after top 10% of dust events (N = 127) and the 4-day period before the events (materials and methods). (**C**) Median relative change in chlorophyll concentration evaluated similarly to (B), but for the 4-day periods before and after a randomly (bootstrap with replacement, N = 1000) sampled 10% of the dust deposition time series at each location. Solid black contour in (B) and (C) shows the 15°C isotherm calculated from the average MODIS Sea Surface Temperature during 2003 to 2016. This boundary is used to demarcate the PSO with annual average sea surface temperature (SST) > 15°C.



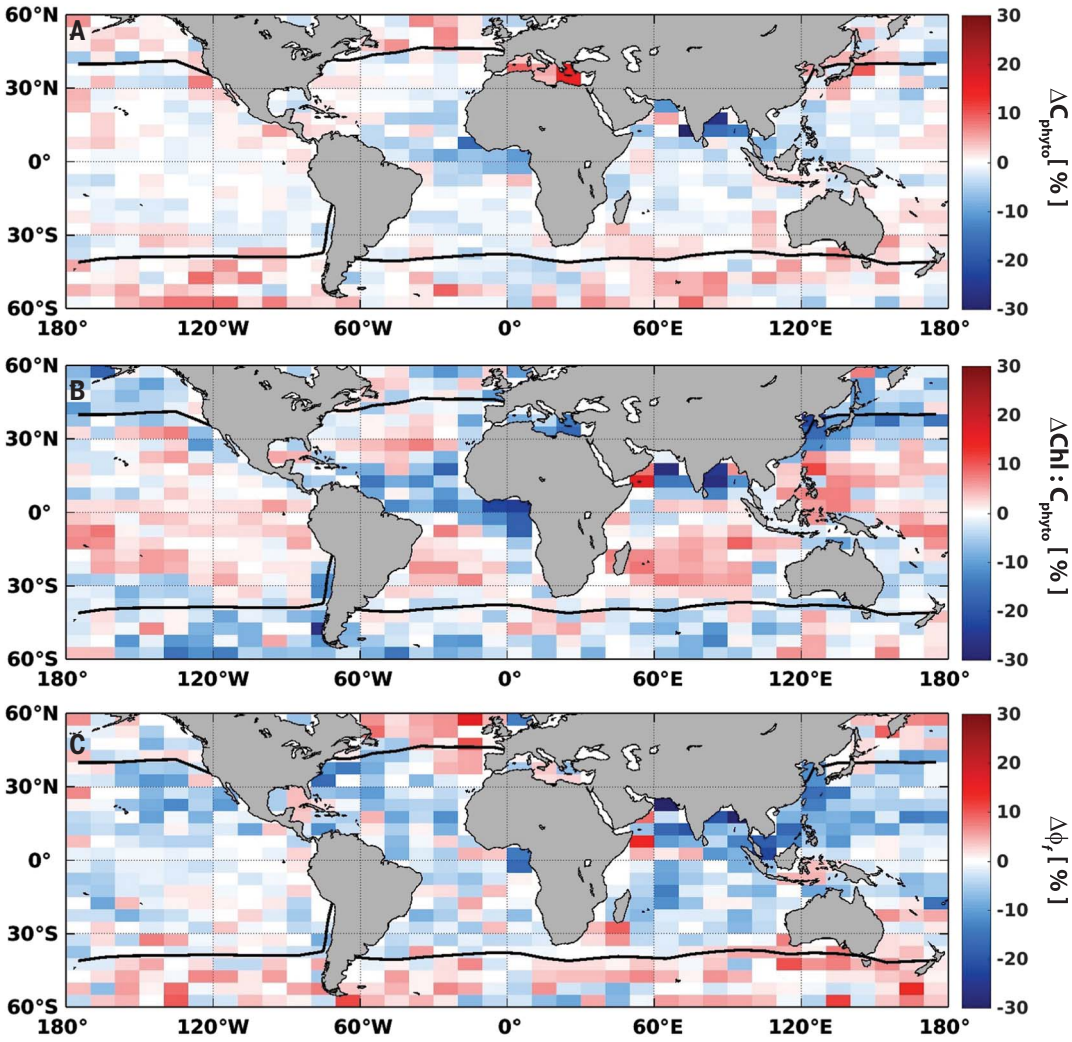


Fig. 3. Biological response to dust deposition in various satellite ocean color–based properties. (**A**) Median relative change in phytoplankton carbon biomass (C_{phyto}) evaluated between the 4-day period after top 10% of dust events (N = 127) and the 4-day period before the events. (**B**) Median relative change in phytoplankton chlorophyll to carbon ratio (Chl:C_{phyto}) evaluated between the 4-day period after top 10% of dust events and the 4-day period before the events. (**C**) Median relative change in chlorophyll fluorescence quantum yield (ϕ_f) evaluated between the 4-day period after top 10% of dust events and the 4-day period before the events.

For each of these events, we first quantified an ocean ecosystem response as the relative change (in percent) in surface chlorophyll (ΔChl) 4 days after the event relative to 4 days before the event (materials and methods). This analysis revealed widespread mean enhancements in ΔChl (65% of total bins) across the global ocean resulting from the dust events (red pixels in Fig. 2B). These responses are of modest magnitude (generally ≤20%) and are comparable (as a percentage) between the PSO and higher-latitude seasonal seas [delineated by annual mean sea surface temperature ≥15°C (*25*); black lines in Fig. 2B]. Ocean regions deviating from the predominant positive ΔChl values are generally restricted to continental shelf waters and a limited region of the Southern Ocean east of Argentina (Fig. 2B, blue pixels). Chlorophyll anomalies in our 14-year record calculated for randomly sampled time points not associated with high dust deposition events consistently exhibit mean ΔChl values close to zero (Fig. 2C). This result thus supports the direct connection between the top 10% of dust events and surface chlorophyll changes (Fig. 2B).

Traditionally, dust deposition has been envisioned as primarily relevant to iron-limited, high-nutrient low-chlorophyll (HNLC) ocean regions (*26*). This is because dust is generally high in iron content, phytoplankton have low cellular iron requirements (i.e., it is a micronutrient), and HNLC waters have ample macronutrients to support significant biomass increases when amended with iron (*27*, *28*). Our results (Fig. 2B) instead suggest a far broader stimulatory effect of atmospheric depositions. This finding could reflect a variety of factors, including (as noted above) stimulation of diazotroph nitrogen fixation by dust iron and input of macronutrients during deposition events. In addition, iron-stressed phytoplankton assemblages are now recognized as commonplace outside of HNLC regions, including in central ocean gyres (*11*, *29*, *30*), coastal upwelling areas (*31*, *32*), and classical bloom-forming areas (*33*–*35*). Finally, variations in limiting nutrient for different phytoplankton groups can occur within a given community (*9*, *36*) such that stimulation of specific groups by dust deposition contributes to broadly observed bulk ΔChl enhancements (Fig. 2B).

Notably, chlorophyll concentrations change in the surface ocean as a result of both variations in phytoplankton biomass and physiological status. With respect to the latter, phytoplankton increase cellular chlorophyll concentrations as nutrient stress (or light availability) decreases. Accordingly, a broadly consistent increase in ΔChl after dust deposition events (Fig. 2B) does not necessarily imply an equally consistent underlying basis. Fortunately, additional satellite ocean color products are available to disentangle observed changes in chlorophyll and have now been used to illustrate the pervasive effect of physiology on the satellite-observed chlorophyll record (*37*, *38*). One of these properties is the particulate backscattering coefficient, b_{bp}, which has been quantitatively related to phytoplankton carbon biomass, C_{phyto} (*39*, *40*). Unlike chlorophyll concentration, C_{phyto} is insensitive to light- and nutrient-driven physiological adjustments

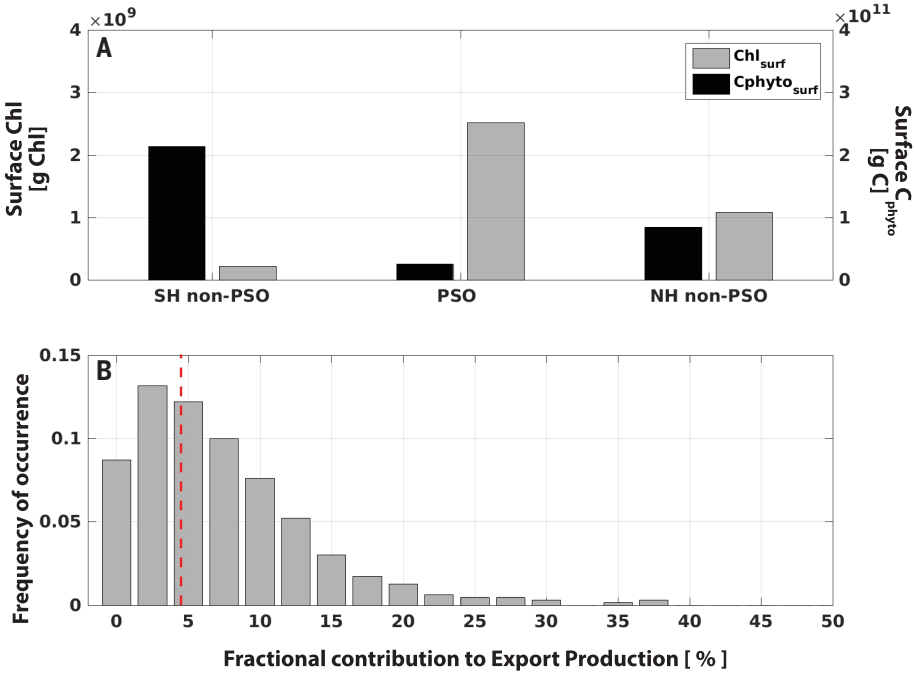


Fig. 4. Globally integrated effects of dust deposition on bulk ocean color properties and net primary production. (**A**) Median net effect of top 10% of dust events in three broad regions of the ocean on surface ocean chlorophyll (in units of grams of chlorophyll) and phytoplankton biomass (in units of grams of carbon). The three regions correspond to the PSO (annual SST ≥ 15°C), and the northern (NH non-PSO) and southern (SH non-PSO) areas of ocean outside the PSO. Values have been integrated over the top 10 m as a conservative estimate of the surface layer directly detected by satellite ocean color measurements. (**B**) Frequency distribution of the contribution of annual dust-mediated enhancement in NPP relative to corresponding export production (expressed as percentage) over all 5°-by-10° bins (N = 631). Annual, integrated global contribution is indicated by the dashed red line and is equal to 4.5% (255 Tg of C per year). Corresponding global map is shown in fig. S5.

(*41*, *42*). Reevaluating our top 10% of deposition events in terms of C_{phyto} anomalies (ΔC_{phyto}) (materials and methods), we find that observed changes in chlorophyll (Fig. 2B) are largely not indicative of increased biomass in the PSO (Fig. 3A). The two most notable apparent ΔC_{phyto} features in the PSO are the modest decreases in the tropical Atlantic west of Africa and in the northernmost Indian Ocean and Bay of Bengal (Fig. 3A). However, these two regions are notoriously dusty and are prone to errors in satellite b_{bp} retrievals (*43*, *44*). In contrast to the PSO, higher-latitude regions exhibit notable increases in ΔC_{phyto} after dust events (poleward of black lines in Fig. 3A) in conjunction with the increases in ΔChl (Fig. 2B).

Results shown in Fig. 2B and Fig. 3A imply that dust depositions to the PSO primarily elicit improvements in phytoplankton physiological status without affecting standing stocks. More specifically, phytoplankton increase cellular pigment levels when dust events improve surface layer nutrient conditions—a response revealed in satellite Chl:C_{phyto} anomalies (ΔChl:C_{phyto}) (Fig. 3B). We see, aside from the above-noted problematic regions west of Africa and south of India, broad regions of elevated ΔChl:C_{phyto} after depositions (red pixels in Fig. 3B), indicative of increased phytoplankton photosynthesis and division rates (*45*, *46*). A physiological interpretation of the ΔChl response in the PSO is further supported by concurrent and widespread decreases in observed chlorophyll fluorescence quantum yields ($\Delta\varphi_f$) (Fig. 3C) (materials and methods). As nutrient stress commonly enhances fluorescence yields in phytoplankton (*47*), these observed decreases in $\Delta\varphi_f$ are consistent with improved surface nutrient conditions. At latitudes poleward of the PSO (black lines in Fig. 3), ΔChl:C_{phyto} was modestly depressed after dust events (Fig. 3B) and $\Delta\varphi_f$ was slightly elevated (Fig. 3C). Although further investigation is warranted, these unexpected findings could reflect a preferential increase in accessory photosynthetic pigments in the diverse phytoplankton communities found at these higher latitudes (*48*). Similar to results for ΔChl (Fig. 2C), random sampling of low-deposition periods during our 14-year record yielded ΔC_{phyto}, ΔChl:C_{phyto}, and $\Delta\varphi_f$ values close to zero (fig. S3).

Results presented here provide observational evidence for global ocean ecosystem responses to prevalent but modest-scale dust depositions. By investigating a variety of phytoplankton properties now retrieved from ocean color remote sensing, we find a distinctly different signature of deposition responses between the PSO and higher-latitude seasonal seas. When integrated across the PSO, the deposition events were associated with an average elevation in surface chlorophyll levels of 2.5×10^9 g of Chl while having a negligible effect on phytoplankton biomass (Fig. 4A). This finding is consistent with current understanding of lower-latitude plankton ecosystems. Across the open ocean PSO, seasonal changes in growth conditions are muted, and variations in phytoplankton division rates are tightly coupled to loss rates. Accordingly, modest dust-stimulated increases in phytoplankton division rates are paralleled by increased loss rates, thereby rapidly conveying enhanced primary production to higher trophic levels. A similar finding was reported for phytoplankton responses to climate variability in the PSO, wherein observed variations in chlorophyll were associated predominantly with physiological adjustments in Chl:C_{phyto} rather than C_{phyto} (*38*). By contrast, growth conditions at higher latitudes are constantly changing over the seasons, and this variability sustains a greater degree of decoupling between phytoplankton division and loss rates (*49*). Stimulation of phytoplankton division by atmospheric nutrient inputs has a greater chance of affecting phytoplankton stocks, which, for the dust events evaluated here, amounted to average increases in C_{phyto} and Chl of 8.5×10^{10} g of C and 1.1×10^9 g of Chl for the north polar and subpolar region and 2.1×10^{11} g of C and 2.3×10^8 g of Chl for the south polar and subpolar region, respectively (Fig. 4A).

The satellite ocean color measurements used in our analysis only detect properties of plankton a few meters below the surface (e.g., Fig. 2B and Fig. 4A). If the Chl and C_{phyto} responses to atmospheric deposition reported here are applied to a contemporary ocean net primary productivity (NPP) model and scaled to the total annual dust deposition, we find that dust-mediated NPP amounts to 255 Tg of C per year. This enhancement results from externally provided ("'new"') nutrients and can be compared with the total annual ocean exportable production of 5.7 Pg of C per year (*50*) (see supplementary materials for discussion). Thus, 4.5% (red dashed line in Fig. 4B) of globally exportable carbon is attributable to dust deposition, but this dust contribution varies widely by location from just a few percent to ~40% of the annual total within any 5°-by-10° bin (Fig. 4B and fig. S2). Although considerable uncertainty remains in this assessment, it does provide an observation-based, order-of-magnitude quantification for the contribution

of atmospheric nourishment to global ocean ecosystems.

The productivity and health of Earth's biosphere are intimately linked to the functioning of the coupled land-ocean-atmosphere systems as well as to human activities. For global plankton, climate warming poses direct threats through its effects on upper ocean thermal gradients, currents, and seasonal cycles. However, human emissions and warming may also influence ocean ecosystems through their effect on atmospheric aerosols. Changes to atmospheric circulation, precipitation patterns, and soil moisture and the advance or retreat of freshwater bodies associated with climate warming or other human activity may be associated with substantial future change in aerosol deposition to the ocean. Concurrent alterations in the atmospheric load of pollutants and smoke emissions may also influence chemical transformations of dust during transport from source regions to deposition sites, further complicating predictions of impacts on ocean biology. The current analysis demonstrates measurable ocean biological responses to a wide dynamic range in atmospheric inputs (Fig. 2A). Although these responses represent only a modest contribution to total plankton stocks or annual net primary production, they nevertheless are an important component of ocean production that is likely to change in the face of a warming planet.

REFERENCES AND NOTES

1. N. Meskhidze, A. Nenes, *Science* **314**, 1419–1423 (2006).
2. A. Lana, R. Simo, S. M. Vallina, J. Dachs, *Atmos. Chem. Phys.* **12**, 7977–7993 (2012).
3. T. D. Jickells *et al*., *Science* **308**, 67–71 (2005).
4. R. C. Hamme *et al*., *Geophys. Res. Lett.* **37**, L19604 (2010).
5. T. K. Westberry, Y. R. Shi, H. Yu, M. J. Behrenfeld, L. A. Remer, *Geophys. Res. Lett.* **46**, 11270–11280 (2019).
6. W. Tang *et al*., *Nature* **597**, 370–375 (2021).
7. Y. Wang *et al*., *Sci. Total Environ.* **807**, 150775 (2022).
8. J. K. B. Bishop, R. E. Davis, J. T. Sherman, *Science* **298**, 817–821 (2002).
9. C. M. Moore *et al*., *Nat. Geosci.* **6**, 701–710 (2013).
10. L. J. Ustick *et al*., *Science* **372**, 287–291 (2021).
11. R. M. Letelier *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12720–12728 (2019).
12. N. M. Mahowald, C. Luo, *Geophys. Res. Lett.* **30**, 1903 (2003).
13. T. Yuan *et al*., *Geophys. Res. Lett.* **47**, GL089711 (2020).
14. H. Yu *et al*., *Science* **337**, 566–569 (2012).
15. H. Yu *et al*., *J. Geophys. Res. Atmos.* **124**, 7975–7996 (2019).
16. B. J. Krueger, V. H. Grassian, J. P. Cowin, A. Laskin, *Atmos. Environ.* **38**, 6253–6261 (2004).
17. N. Meskhidze, W. L. Chameides, A. Nenes, G. Chen, *Geophys. Res. Lett.* **30**, 2085 (2003).
18. A. R. Baker, M. French, K. L. Linge, *Geophys. Res. Lett.* **33**, L07805 (2006).
19. R. Paris, K. V. Desboeufs, P. Formenti, S. Nava, C. Chou, *Atmos. Chem. Phys.* **10**, 4273–4282 (2010).
20. M. van der Does *et al*., *Geophys. Res. Lett.* **47**, e2019GL086867 (2020).
21. D. J. McGillicuddy Jr. *et al*., *Nature* **394**, 263–266 (1998).
22. A. F. Michaels *et al*., *Deep Sea Res. Part I Oceanogr. Res. Pap.* **41**, 1013–1038 (1994).
23. A. F. Michaels *et al*., *Biogeochemistry* **35**, 181–226 (1996).
24. B. A. Ward, S. Dutkiewicz, C. M. Moore, M. J. Follows, *Limnol. Oceanogr.* **58**, 2059–2075 (2013).
25. M. J. Behrenfeld *et al*., *Nature* **444**, 752–755 (2006).
26. J. H. Martin, *Paleoceanography* **5**, 1–13 (1990).
27. S. T. Lindley, R. T. Barber, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **45**, 1135–1150 (1998).
28. P. W. Boyd *et al*., *Science* **315**, 612–617 (2007).
29. T. J. Browning *et al*., *Nature* **551**, 242–246 (2017).
30. S. Bonnet *et al*., *Biogeosciences* **5**, 215–225 (2008).
31. K. W. Bruland, E. L. Rue, G. J. Smith, *Limnol. Oceanogr.* **46**, 1661–1674 (2001).
32. D. V. Biller, T. H. Coale, R. C. Till, G. J. Smith, K. W. Bruland, *Cont. Shelf Res.* **66**, 58–72 (2013).
33. A. J. Birchill *et al*., *Sci. Rep.* **9**, 1435 (2019).
34. T. J. Ryan-Keogh *et al*., *Limnol. Oceanogr.* **58**, 533–545 (2013).
35. M. C. Nielsdóttir, C. M. Moore, R. Sanders, D. J. Hinz, E. P. Achterberg, *Global Biogeochem. Cycles* **23**, GB3001 (2009).
36. J. K. Moore, S. C. Doney, K. Lindsay, *Global Biogeochem. Cycles* **18**, GB4028 (2004).
37. D. A. Siegel *et al*., *Remote Sens. Environ.* **135**, 77–91 (2013).
38. M. J. Behrenfeld *et al*., *Nat. Clim. Change* **6**, 323–330 (2016).
39. J. R. Graff *et al*., *Deep Sea Res. Part I Oceanogr. Res. Pap.* **102**, 16–25 (2015).
40. V. Martinez-Vicente, G. Dall'Olmo, G. Tarran, E. Boss, S. Sathyendranath, *Geophys. Res. Lett.* **40**, 1154–1158 (2013).
41. R. J. Geider, H. L. MacIntyre, T. M. Kana, *Limnol. Oceanogr.* **43**, 679–694 (1998).
42. R. J. Geider, H. L. MacIntyre, T. M. Kana, *Mar. Ecol. Prog. Ser.* **148**, 187–200 (1997).
43. C. Moulin, H. R. Gordon, V. F. Banzon, R. H. Evans, *J. Geophys. Res.* **106**, 18239–18249 (2001).
44. C. Moulin, H. R. Gordon, R. M. Chomko, V. F. Banzon, R. H. Evans, *Geophys. Res. Lett.* **28**, 5–8 (2001).
45. E. A. Laws, T. T. Bannister, *Limnol. Oceanogr.* **25**, 457–473 (1980).
46. K. H. Halsey, B. M. Jones, *Annu. Rev. Mar. Sci.* **7**, 265–297 (2015).
47. M. J. Behrenfeld *et al*., *Biogeosciences* **6**, 779–794 (2009).
48. L. M. Bolaños *et al*., *ISME J.* **14**, 1663–1674 (2020).
49. M. J. Behrenfeld *et al*., *Nat. Geosci.* **10**, 118–122 (2017).
50. D. A. Siegel *et al*., *Global Biogeochem. Cycles* **28**, 181–196 (2014).

ACKNOWLEDGMENTS

Funding: This study was supported by NASA grants 80NSSC18K0953 (L.A.R., Y.R.S., and H.Y.) and 80NSSC18K0616 (T.K.W. and M.J.B.). **Author contributions:** Conceptualization: T.K.W., M.J.B., and L.A.R. Methodology: T.K.W., M.J.B., L.A.R., Y.R.S., H.Y., and H.B. Investigation: T.K.W., M.J.B., L.A.R., Y.R.S., and H.Y. Visualization: T.K.W. and Y.R.S. Funding acquisition: T.K.W., M.J.B., L.A.R., and H.Y. Project administration: L.A.R. Supervision: M.J.B., L.A.R., and H.Y. Writing – original draft: M.J.B. Writing – review & editing: T.K.W., M.J.B., L.A.R., Y.R.S., H.Y., and H.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All global satellite ocean color data used in this work are freely available through the NASA Ocean Biology Distributed Active Archive Center (OB.DAAC). Daily mixed-layer depths were calculated from freely available hindcast simulations of the HYCOM model and are accessible on the Oregon State University Ocean Productivity Website (http://sites.science.oregonstate.edu/ocean.productivity). The working 4-day, 5°-by-10° dataset including all ocean color products and dust deposition rates used for this analysis can be found at the same website. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq5252
Materials and Methods
Supplementary Text
Figs. S1 to S5
References (*51*–*74*)

Submitted 15 April 2022; accepted 31 March 2023
10.1126/science.abq5252

PETROLOGY

Garnet crystallization does not drive oxidation at arcs

Megan Holycross[1,2]* and Elizabeth Cottrell[2]

Arc magmas, the building blocks of continental crust, are depleted in total iron (Fe), have higher ratios of oxidized Fe to total Fe ($Fe^{3+}/\sum Fe$), and record higher oxygen fugacities (fO_2's) compared with magmas erupted at mid-ocean ridges. Garnet crystallization could explain these observations if garnet removes substantial amounts of Fe^{2+}, but not Fe^{3+}, from magma, yet this model for continental crust generation has never been tested experimentally. Analysis of garnets and melts in laboratory experiments show that the compatibilities of Fe^{2+} and Fe^{3+} in garnet are of similar magnitudes. Our results indicate that fractional crystallization of garnet-bearing cumulates will remove 20% of total Fe from primary arc basalts but negligibly alter the $Fe^{3+}/\sum Fe$ ratio and fO_2 of the melt. Garnet crystallization is unlikely to be responsible for the relatively oxidized nature of basaltic arc magmas or the Fe-depletion trend observed in continental crust.

The trace-element signatures of arc lavas mirror those of the continental crust, suggesting that subduction zone magmatism is linked to the growth of continents (*1*). Arc lavas are also oxidized relative to lavas erupted in other tectonic settings (*2*). Resolving the source of oxidation in the subduction system is then critical for understanding the evolution of continental crust and formation of ore deposits. The oxidized nature of arc magmas may arise from melting of mantle oxidized by material liberated from subducted lithosphere (*3*–*10*). Conversely (or in addition), arc magmas may oxidize in the crust as a result of magmatic differentiation (*11*, *12*). In the latter category, high-pressure fractional crystallization of garnet-bearing cumulates may act to remove more ferrous than ferric iron from melts, such that a "crustal redox filter" oxidizes arc magmas after they have separated from their mantle sources (*13*–*15*). In the "crustal redox filter" scenario, the crystallization of Fe-enriched garnet would also result in the generation of melts depleted in their total Fe contents—a common feature of calc-alkaline magmas found at continental margins. We provide direct measurements of the oxidation state of Fe in experimental garnets and melts equilibrated under controlled conditions to determine if the crystallization of garnet is compatible with the production of oxidized, Fe-depleted arc magmas that are characteristic of the continental crust.

Cumulates bearing garnet+pyroxene±rutile assemblages (i.e., garnet pyroxenites) may crystallize in magmas formed at continental arc volcanoes where the overriding crust is ≥35 km thick [e.g., (*16*–*18*)]. Previous work has suggested that low $Fe^{3+}/\sum Fe$ [$Fe^{3+}/\sum Fe$ = 0.04 to 0.08; (*15*)], Fe-enriched garnet pyroxenites are the complements to high $Fe^{3+}/\sum Fe$ [$Fe^{3+}/\sum Fe$ = 0.15 to 0.35 (*2*, *4*, *6*, *9*)], Fe-depleted arc magmas (*15*). Garnet fractionation has been proposed to be a key aspect for forming large Cu ore bodies (*19*) because porphyry copper deposits are typically associated with thick-crusted, mature arc volcanoes with oxidized magmas (*20*, *21*).

These hypotheses require that the compatibility of Fe^{2+} in garnet be much greater than the compatibility of Fe^{3+} in garnet as it crystallizes from primitive arc magmas in the crust, resulting in an increase in the $Fe^{3+}/\sum Fe$ ratio and fO_2 of the magma. However, if the compatibilities of Fe^{3+} and Fe^{2+} in garnet are similar, little fractionation of Fe species between garnet and liquid will occur during cumulate crystallization (or later remelting of cumulate phases).

Partitioning of Fe species between garnet and melt

With new laboratory measurements of Fe^{3+} and Fe^{2+} partitioning between garnet and melt, we directly evaluate the role that garnet fractionation could play in generating Fe-depleted melts with $Fe^{3+}/\sum Fe$ ratios elevated above those derived from low-pressure fractionation in the absence of garnet (as found at ridges and thin-crusted arcs). We equilibrated garnet and silicate melts from hydrous basaltic bulk compositions in 13 piston-cylinder experiments that simulated conditions in deep crustal magma chambers (T = 950° to 1230°C, P = 1.5 to 3.0 GPa). Experiment fO_2's, ranging from 0.8 logarithmic units below to 6.4 logarithmic units above the quartz-fayalite-magnetite (QFM) buffer, were set using multiple O_2-buffering approaches (*22*). Experiments (Fig. 1) cosaturated in pyrope-rich garnets with subequal almandine and grossular components, clinopyroxene (cpx) ± rutile, and quartz [see (*22*, *23*) for full details].

We measured the $Fe^{3+}/\sum Fe$ ratios of experimental garnets and silicate melts with Fe K-edge micro x-ray absorption near-edge structure (μ-XANES) spectroscopy. We calculated the concentrations of ferric and ferrous iron in both phases from measurements of glass and garnet $Fe^{3+}/\sum Fe$ ratios and FeO_T contents (i.e., all Fe present expressed as Fe^{2+}). We calculated coefficients for the partitioning of Fe^{2+} and Fe^{3+} between garnet (grt) and melt as

$$D_{Fe^{3+}}^{grt/melt} = \frac{Fe^{3+}grt}{Fe^{3+}melt} \quad (1)$$

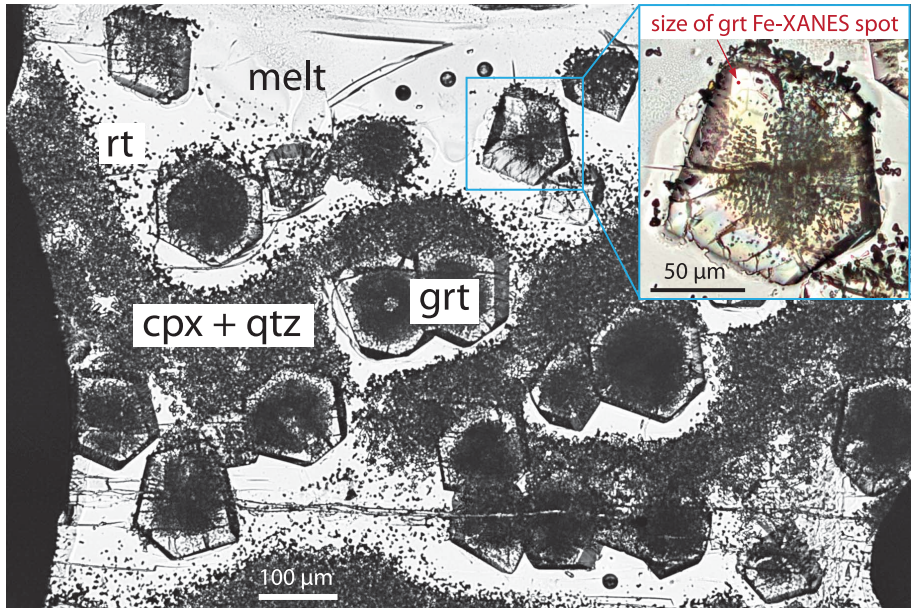


Fig. 1. Transmitted light micrograph of garnet-melt partitioning experiment. This experiment (experiment 22) contains a large fraction of silicate melt, garnet (grt), clinopyroxene (cpx), rutile (rt), and quartz (qtz). Garnet cores contain abundant mineral inclusions. Enlarged inset image shows the size of a garnet XANES spot (2 μm by 2 μm) for comparison.

[1]Cornell University, Ithaca, NY 14853, USA. [2]National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA.
*Corresponding author. Email: holycross@cornell.edu

$$D^{grt/melt}_{Fe^{2+}} = \frac{Fe^{2+}grt}{Fe^{2+}melt} \quad (2)$$

Garnet $Fe^{3+}/\sum Fe$ ratios increase as fO_2 and temperature increase at constant pressure (fig. S8). Three experiments equilibrated at 1100°C and ΔQFM+6 show that an increase in pressure from 2.0 to 3.0 GPa decreases garnet $Fe^{3+}/\sum Fe$ ratios by 36% (relative). Garnet $Fe^{3+}/\sum Fe$ ratios are lower than melt $Fe^{3+}/\sum Fe$ ratios in all experiments; however, garnets contain 2 to 10 times more total Fe compared with melts. Consequently, the calculated partition coefficients show that Fe^{3+} is slightly incompatible to compatible in garnet (D = 0.7 to 2.4) with the greatest values of $D^{grt/melt}_{Fe^{3+}}$ measured in our coldest, most reducing experiments (Fig. 2). The decrease in $D^{grt/melt}_{Fe^{3+}}$ as fO_2 increases at constant pressure and temperature may result from a steric limit on the substitution of Fe^{3+} in pyroxenitic garnets, not unlike that observed for clinopyroxenes (*24*). The partitioning of Fe^{2+} between garnet and melt is constant across changing fO_2. Limited data from the pressure-series experiments at 1100°C and ΔQFM+6 suggest that increasing pressure decreases $D^{grt/melt}_{Fe^{3+}}$ but increases $D^{grt/melt}_{Fe^{2+}}$.

We performed multiple linear regressions to parameterize $D^{grt/melt}_{Fe^{3+}}$ and $D^{grt/melt}_{Fe^{2+}}$ in our experiments. At constant fO_2 and pressure, increasing temperature causes $D^{grt/melt}_{Fe^{2+}}$ to decrease more rapidly than $D^{grt/melt}_{Fe^{3+}}$, such that $D^{grt/melt}_{Fe^{3+}}$ and $D^{grt/melt}_{Fe^{2+}}$ converge to similar values when T > 1100°C (fig. S9). Consequently, Fe^{3+} will be minimally fractionated from Fe^{2+} in garnet-bearing melts at high temperatures and fO_2's near the QFM buffer, where the compatibilities of both iron species in garnet are similar. More substantial fractionation could occur during processing under colder ($T \leq$ 950°C) and/or very oxidizing ($fO_2 \geq$ ΔQFM+4) conditions in which $D^{grt/melt}_{Fe^{2+}} \approx 10 D^{grt/melt}_{Fe^{3+}}$.

Deep arc magma chambers

Our data allow us to test the hypothesis that the fractionation of garnet at high pressures in the lower crust can generate Fe-depleted magmas with elevated $Fe^{3+}/\sum Fe$ ratios and high fO_2's (*14*, *15*). Previous work (*14*) suggested that the fractionation of a pyroxenite assemblage with 30 modal % garnet from a hydrous primitive magma can remove up to 60% of the Fe from the melt while increasing its fO_2 by as much as six logarithmic units. In the absence of measured ferrous-ferric garnet–melt partition coefficients, the model of Tang *et al.* (*14*) assumed perfect incompatibility of Fe^{3+} in garnet (i.e., $D^{grt/melt}_{Fe^{3+}} = 0$), and $D^{grt/melt}_{Fe^{2+}}$ was described as "significantly >1." We adopt a similar modeling approach to evaluate how our new calculated values for $D^{grt/melt}_{Fe^{3+}}$ and $D^{grt/melt}_{Fe^{2+}}$ affect this finding.

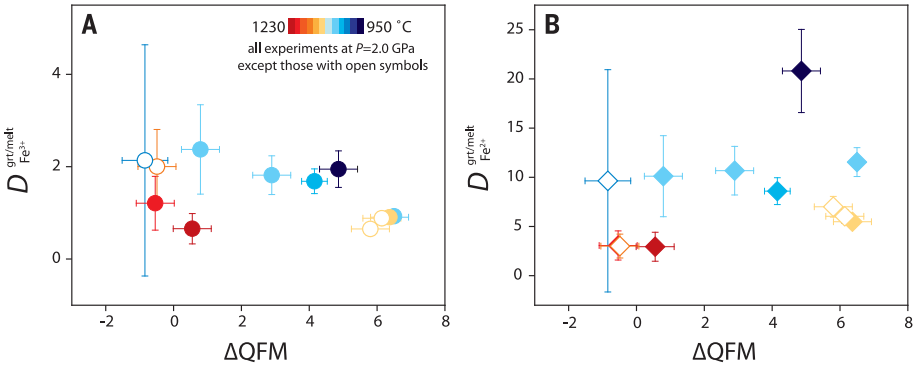


Fig. 2. Calculated partition coefficients for Fe^{2+} and Fe^{3+} in garnet. (**A**) $D^{grt/melt}_{Fe^{3+}}$ decreases as fO_2 and temperature increase. (**B**) $D^{grt/melt}_{Fe^{2+}}$ is dependent on temperature only. The total Fe content of garnet increases as temperature decreases (data S1). Data error bars are ± calculated standard errors of partition coefficients (*23*). Calculated standard errors are greatest for the most reduced experiments in which the uncertainty of the XANES calibration is large relative to the measured $Fe^{3+}/\sum Fe$ ratios of garnet.

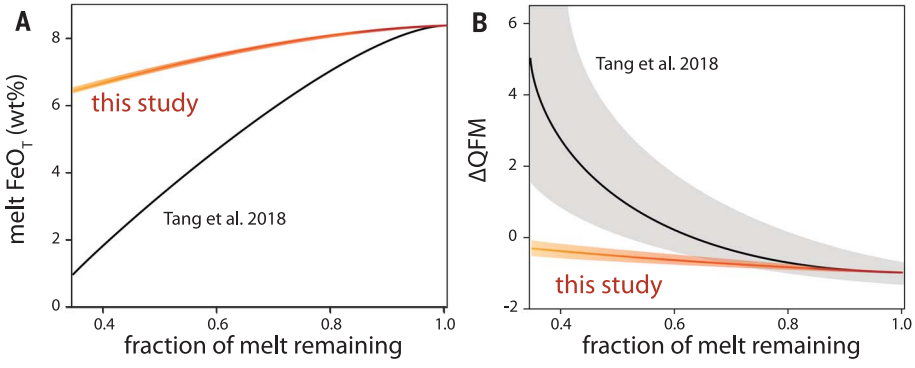


Fig. 3. Impact of garnet pyroxenite crystallization on the total Fe and fO_2 of arc magmas. We compare our model results to those of Tang *et al.* (*14*) (black curve with gray error envelope), modeled under equivalent conditions with assumed partition coefficients. The red-orange gradient of our model represents cooling from 1200°C at the onset of crystallization to 1035°C at 35% melt remaining. (**A**) The magma becomes depleted in Fe as crystallization progresses, but the trend modeled using our measured partition coefficients is moderate compared to previous work (*14*). (**B**) Our new data suggest that garnet pyroxenite fractionation will increase the fO_2 of primitive arc magmas by only 0.7 logarithmic units after >60% crystallization. Our result contrasts with the model of (*14*) that assumed $D^{grt/melt}_{Fe^{2+}} \gg D^{grt/melt}_{Fe^{3+}}$, leading to a significant increase in the $Fe^{3+}/\sum Fe$ ratio and fO_2 of residual melts during crystallization.

Our modeled fractionation of a cumulate with 70 modal % cpx and 30 modal % garnet begins at fO_2 = ΔQFM − 1 in a hydrous basaltic magma with $Fe^{3+}/\sum Fe$ = 0.09 (Fig. 3). We chose model conditions to facilitate direct comparison with the results of Tang *et al.* (*14*). Both ferrous and ferric iron are compatible in garnet at these conditions, but the difference between their partition coefficients is minimal ($D^{grt/melt}_{Fe^{2+}} = 2.1$; $D^{grt/melt}_{Fe^{3+}} = 1.4$; this study). We applied $D^{cpx/melt}_{Fe^{3+}}$ between 0.3 and 0.8 (*25*, *26*), represented by the width of the red-orange envelope around the modeled value (solid line) of $D^{cpx/melt}_{Fe^{3+}} = 0.6$ in Fig. 3. Ferrous iron and Mg occupy the same sites in cpx and garnet, so we tie the partitioning of Fe^{2+} to Mg in both phases using an Fe-Mg exchange coefficient, K_D Fe/Mg. We modeled the composition of the melt until ~65% fractional crystallization of garnet pyroxenite [i.e., previously published estimates of the proportion of mafic cumulates generated during arc magma differentiation in the lower crust (*27*)]. Because the physical mechanism for crystallization is likely to be cooling, we varied temperature and pressure over the fractional crystallization interval in our model beginning at 1200°C, 2.0 GPa to 1035°C, 1.3 GPa at a residual melt fraction of 0.35. Coefficients for the partitioning of ferric iron between garnet and melt and our K_D^{grt} Fe/Mg values were recalculated at each crystallization step. We assumed clinopyroxene-melt partition coefficients to be insensitive to changing pressure and temperature, and K_D^{cpx}

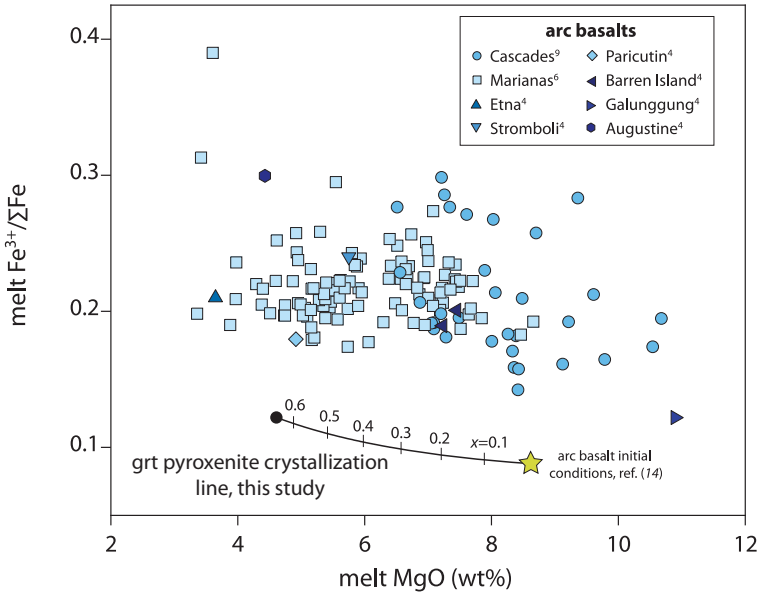


Fig. 4. Comparison between $Fe^{3+}/\sum Fe$ ratios of modeled arc magmas and measured $Fe^{3+}/\sum Fe$ ratios of arc basalts. The modeled $Fe^{3+}/\sum Fe$ ratios of primitive arc basalts that fractionate garnet pyroxenites (black line) do not approach the measured $Fe^{3+}/\sum Fe$ ratios of arc basalts (blue data points). The black line shows the $Fe^{3+}/\sum Fe$ ratio and MgO content of the residual melt modeled in Fig. 3 using garnet-melt partition coefficients for Fe species measured in this study. Crystal fractionation intervals (x) are indicated for 0.1 steps on the black line. Model initial conditions are the same as in Tang *et al*. (*14*).

Fe/Mg was set at 0.25 (*25*, *26*). Additional assumptions, parameters, and other details can be found in (*23*).

The compatibility of ferrous iron in both cpx and garnet, and ferric iron in garnet, leads to a decrease in melt FeO_T as crystallization progresses in our model (Fig. 3A). However, the degree of Fe depletion in the melt is moderate (~20%). An extreme Fe-depletion trend [≥60%, e.g., (*14*)] can only be produced through the input of unrealistically high Fe-Mg exchange coefficients (>7× measured values at relevant conditions), which results in bulk cumulate compositions with extremely low Mg numbers (Mg number of the first cumulate = 0.07; fig. S12). Our model reproduces the Mg numbers [0.43 to 0.89 (*27*)] and garnet compositions of natural pyroxenite xenoliths (fig. S10) (*23*). Pyroxenitic garnets contain subequal Fe and Mg components and only become almandine-rich during crystallization at low temperatures. Our measurements and the literature data show that the compatibility of Fe^{2+} is greater than that of Fe^{3+} in garnet [and is likely greater in cpx (*25*, *26*)], but the partition coefficients for both species in each phase are of similar magnitudes at the conditions relevant for cumulate fractionation in primitive arc magmas. Consequently, the crystallization of garnet pyroxenites, modeled using our measured partition coefficients, will only increase the $Fe^{3+}/\sum Fe$ ratio of the magma by 0.04, corresponding to a 0.8 logarithmic unit increase in the fO_2 of the residual melt after 65% fractional crystallization (Fig. 3B). The $Fe^{3+}/\sum Fe$ ratio of the residual melt does not approach the measured $Fe^{3+}/\sum Fe$ ratios of olivine-hosted arc melt inclusions and submarine glasses (*2*, *4*, *6*, *9*) over this crystallization interval (Fig. 4). The arc basalt oxidation trends of (*14*) can only be achieved if Fe^{2+} is assumed to be outstandingly compatible in garnet and cpx. Even if Fe^{3+} is modeled as perfectly incompatible in garnet, the $Fe^{3+}/\sum Fe$ ratio of the residual melt does not increase by more than 0.06, mirroring the olivine fractionation trend observed in mid-ocean ridge basalt (*23*, *28*).

Garnet crystallization does not drive oxidation at arcs

Fractionation of garnet + cpx + rutile residues has been invoked as an essential process to "auto-oxidize" the magmas associated with large copper porphyries (*19*) and to lower the Nb/Ta ratios of arc magmas generally (*29*). Our finding does not question that formation of garnet pyroxenites can be an important part of arc petrogenesis (*16*–*18*, *30*); however, it cannot "auto-oxidize" nor lower the Nb/Ta ratios of magmas. Measured rutile-melt partition coefficients demonstrate that the fractionation of rutile-bearing residues cannot explain the Nb/Ta ratios of continental crust (*31*), and it is clear from the results of this study that the crystallization of garnet pyroxenites from basaltic magmas cannot be responsible for the elevated $Fe^{3+}/\sum Fe$ ratios and high fO_2's of arc magmas and Cu porphyries, which must be produced by a different process. Our results also indicate that the Fe-depleted calc-alkaline differentiation trend cannot be attributed solely to garnet crystallization in the base of thick-crusted arcs.

REFERENCES AND NOTES

1. J. B. Gill, *Orogenic Andesites and Plate Tectonics* (Springer, 1981).
2. E. Cottrell *et al*., in *Magma Redox Geochemistry*, R. Morretti, D. R. Neuville, Eds. (Wiley, 2021), pp. 33–61.
3. B. J. Wood, L. T. Bryndzia, K. E. Johnson, *Science* **248**, 337–345 (1990).
4. K. A. Kelley, E. Cottrell, *Science* **325**, 605–607 (2009).
5. K. A. Evans, A. G. Tomkins, *Earth Planet. Sci. Lett.* **308**, 401–409 (2011).
6. M. N. Brounce, K. A. Kelley, E. Cottrell, *J. Petrol.* **55**, 2513–2536 (2014).
7. F. Gaillard, B. Scaillet, M. Pichavant, G. Iacono-Marziano, *Chem. Geol.* **418**, 217–233 (2015).
8. B. Debret, D. A. Sverjensky, *Sci. Rep.* **7**, 10351 (2017).
9. M. J. Muth, P. J. Wallace, *Geology* **49**, 1177–1181 (2021).
10. J. J. Ague *et al*., *Nat. Geosci.* **15**, 320–326 (2022).
11. C. T. Aeolus Lee, W. P. Leeman, D. Canil, Z. X. A. Li, *J. Petrol.* **46**, 2313–2336 (2005).
12. C. T. A. Lee *et al*., *Science* **336**, 64–68 (2012).
13. E. J. Chin, K. Shimizu, G. M. Bybee, M. E. Erdman, *Earth Planet. Sci. Lett.* **482**, 277–287 (2018).
14. M. Tang, M. Erdman, G. Eldridge, C. A. Lee, *Sci. Adv.* **4**, eaar4444 (2018).
15. M. Tang, C. T. A. Lee, G. Costin, H. E. Höfer, *Earth Planet. Sci. Lett.* **528**, 115827 (2019).
16. O. Müntener, P. B. Kelemen, T. L. Grove, *Contrib. Mineral. Petrol.* **141**, 643–658 (2001).
17. O. Jagoutz, O. Müntener, M. W. Schmidt, J. P. Burg, *Earth Planet. Sci. Lett.* **303**, 25–36 (2011).
18. O. Jagoutz, P. B. Kelemen, *Annu. Rev. Earth Planet. Sci.* **43**, 363–404 (2015).
19. C. T. A. Lee, M. Tang, *Earth Planet. Sci. Lett.* **529**, 115868 (2020).
20. R. H. Sillitoe, *Econ. Geol.* **105**, 3–41 (2010).
21. J. P. Richards, *Ore Geol. Rev.* **40**, 1–26 (2011).
22. M. Holycross, E. Cottrell, *Contrib. Mineral. Petrol.* **177**, 21 (2022).
23. Materials and methods are available as supplementary materials.
24. M. C. McCanta, M. D. Dyar, M. J. Rutherford, J. S. Delaney, *Am. Mineral.* **89**, 1685–1693 (2004).
25. F. A. Davis, E. Cottrell, *Contrib. Mineral. Petrol.* **176**, 67 (2021).
26. A. Rudra, M. M. Hirschmann, *Geochim. Cosmochim. Acta* **328**, 258–279 (2022).
27. C. T. A. Lee, in *Treatise on Geochemistry*, H. D. Holland, K. K. Turekian, Eds. (Elsevier, ed. 2, 2014), pp. 423-456.
28. E. Cottrell, K. A. Kelley, *Earth Planet. Sci. Lett.* **305**, 270–282 (2011).
29. M. Tang *et al*., *Nat. Commun.* **10**, 235 (2019).
30. P. Ulmer, R. Kaegi, O. Müntener, *J. Petrol.* **59**, 11–58 (2018).
31. M. Holycross, E. Cottrell, *Am. Mineral.* **105**, 244–254 (2020).

ACKNOWLEDGMENTS

Thanks to A. Woodland for garnet reference materials; F. Davis, M. Muth, M. Brounce, and J. Ague for helpful discussion; T. Gooding, T. Rose, E. Gazel, J. B. Balta, J. Thomas, and G. Macpherson for laboratory help; and A. Lanzirotti, M. Newville, K. Kelley, J. Andrys, and R. Tappero for beamline assistance. We are grateful to L. Hale and C. Brown for assistance with reference materials borrowed from the National Rock Collection at the Smithsonian. We thank two

anonymous reviewers for detailed and constructive comments that improved the manuscript. **Funding:** National Science Foundation grant EAR 1855208 (M.H.), Smithsonian Scholarly Studies and the Lyda Hill Foundation (E.C.), and NMNH Peter Buck Fellowship (M.H.). This research used resources of the National Synchrotron Light Source II, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under contract no. DE-SC0012704, and resources of the Advanced Photon Source, a US DOE Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. **Author contributions:** Conceptualization, Methodology, Investigation, Visualization, and Funding Acquisition: M.H. and E.C.. First Draft: M.H.. Review and Editing: M.H. and E.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.
License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade3418
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (*32*–*69*)
Data S1 and S2

Submitted 10 August 2022; accepted 30 March 2023
10.1126/science.ade3418

MAGNETISM

Topological kinetic crossover in a nanomagnet array

Xiaoyu Zhang[1], Grant Fitez[1], Shayaan Subzwari[1], Nicholas S. Bingham[1]†, Ioan-Augustin Chioar[1], Hilal Saglam[1], Justin Ramberger[3], Chris Leighton[3], Cristiano Nisoli[2]*, Peter Schiffer[1,4]*

Ergodic kinetics, which are critical to equilibrium thermodynamics, can be constrained by a system's topology. We studied a model nanomagnetic array in which such constraints visibly affect the behavior of the magnetic moments. In this system, magnetic excitations connect into thermally active one-dimensional strings whose motion can be imaged in real time. At high temperatures, our data showed the merging, breaking, and reconnecting of strings, resulting in the system transitioning between topologically distinct configurations. Below a crossover temperature, the string motion is dominated by simple changes in length and shape. In this low-temperature regime, the system is energetically stable because of its inability to explore all possible topological configurations. This kinetic crossover suggests a generalizable conception of topologically broken ergodicity and limited equilibration.

Our understanding of equilibrium thermodynamics relies on simplifying assumptions about kinetics—for example, the ergodic hypothesis, which postulates that a system can explore all energetically equivalent configurations (*1*). When kinetics are constrained (*2*–*7*), however, ergodicity can break down on relevant timescales, causing memory effects, glassiness, nonequilibration, or slow equilibration (*8*).

Topological constraints on kinetics are of particular interest in the context of the critical role of topology in much of modern physics and have been examined within the context of various mathematical models (*2*, *4*). The constraints can arise from a partition of the accessible states (or phase space) into topologically distinct subsets, often called topological sectors (*9*, *10*). Kinetics within a sector are topologically trivial; they do not alter the system's topology, contrasting with nontrivial kinetics that allow the system to cross through sectors and change the system's topology. If the latter are constrained, then the system cannot explore all of its configurations and should show deviations from reversible equilibrium thermodynamics. In experimental systems, topologically constrained kinetics were invoked early in the context of soft-matter systems—for example, in macromolecule elasticity, foams, cellular patterns, and the kinetics of continuous networks (*2*)—and this broad topic is closely connected to fundamental questions around equilibration, ergodicity, and memory.

[1]Department of Applied Physics, Yale University, New Haven, CT 06511, USA. [2]Theoretical Division and Center for Nonlinear Studies, MS B258, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [3]Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA. [4]Department of Physics, Yale University, New Haven, CT 06511, USA.
***Corresponding author. Email: peter.schiffer@yale.edu (P.S.); cristiano@lanl.gov (C.N.)**
†Present address: Department of Physics, University of Maine, Orono, ME 04469, USA.

Strings in Santa Fe ice

We report the direct visualization of topologically constrained kinetics in a designed nanomagnet array. Such arrays, known broadly as artificial spin ice systems, have displayed a range of exotic collective phenomena and technological potential (*11*, *12*). The particular array geometry that we studied is Santa Fe ice (SFI) (*13*, *14*). The structure of SFI is shown in Fig. 1, A and B, where each element is a single-domain ferromagnetic island with the moment oriented along the long axis as a result of shape anisotropy. The islands are arranged in rectangular plaquettes, with pairs of interior plaquettes (Fig. 1, A, C, and D, light blue shaded areas) surrounded by six peripheral plaquettes. Much of the physics of this system can be understood through the moment configurations at the vertices where islands converge in the lattice (*13*, *14*). The SFI lattice geometry requires that some fraction of the vertices are not in the lowest local energy state, even in the collective ground state. These local excitations have previously been dubbed "unhappy" vertices (*13*, *15*) and are indicated with red dots in Fig. 1, C and D. SFI lends itself to the study of topological kinetics because, as we previously showed (*14*), the lattice geometry forces unhappy vertices to form one-dimensional (1D) strings that can either end within interior plaquettes or form closed loops [(*16*), section S3]. In Fig. 1, C and D, we show these strings as olive lines connecting the unhappy vertices in two realizations of the highly degenerate, disordered ground state of the system (*13*, *14*). Such strings have naturally complex topological properties (*17*), as is readily apparent in a bowl of spaghetti or udon (*18*, *19*).

To probe the physics of SFI, we studied samples composed of stadium-shaped permalloy islands with designed lateral dimensions of 470 by 170 nm, thickness of ~2.5 nm, and lattice spacing of a = 600, 700, and 800 nm (Fig. 1A). The specifics of sample fabrication and characterization are discussed in (*16*), section S1, and have been described previously in (*14*), in which we also showed that the string population and length distributions are thermally activated at the highest experimentally accessible temperatures. We measured the moment configuration in our samples using x-ray magnetic circular dichroism photoemission electron microscopy (XMCD-PEEM) (*16*), which yielded real-space images of the island magnetic moments (Fig. 1B). When we increased temperature, thermal excitations caused a fraction of the moments to visibly flip in orientation between images, with the rate of flipping increasing with increasing temperature, and we could therefore observe the thermal kinetics of this system. We show results for the a = 600 nm sample, in which the interisland interactions are strongest, in the section on quantifying string motion; the data are consistent for the other lattice spacings (*16*). All data are derived from the same XMCD-PEEM images taken on the same samples that were analyzed in (*14*).

Because the energy cost of a string increases with its length, the disordered ground state corresponds to strings of three unhappy vertices, each on three-island vertices. Therefore, in the ground state, strings are constrained to connecting near-neighbor interior plaquettes (Fig. 1, C and D). In excited states of the system, strings can be longer and can connect interior plaquettes that are not neighboring, or they can form closed loops. In our experimental data, the system approached the disordered ground state at the lowest temperatures, with a residual population of excitations that increased in number for the larger lattice spacings (*14*).

Types of string motion

In considering string kinetics, one must consider the possible motions of the strings within the geometrical constraints imposed by the SFI structure. For any string configuration, the strings can be continuously deformed by bending, elongating, or shrinking them while keeping the endpoints the same. Such deformations do not change the topology of the string configuration and thus keep the system within a single topological sector of available configurations, also known as a homotopy class. In the case of SFI, a homotopy class is defined as the set of all string configurations connecting the same interior plaquettes, in which any member of the set can be transformed to any other member of the set through continuous deformations only. Because the SFI ground state can be realized with many different sets of string connections among the plaquettes, the low-energy configurations of the system can be partitioned into different homotopy classes (Fig. 1, C and D).

By contrast, string configurations that cannot be obtained by continuous bending or deformation of the strings are in different

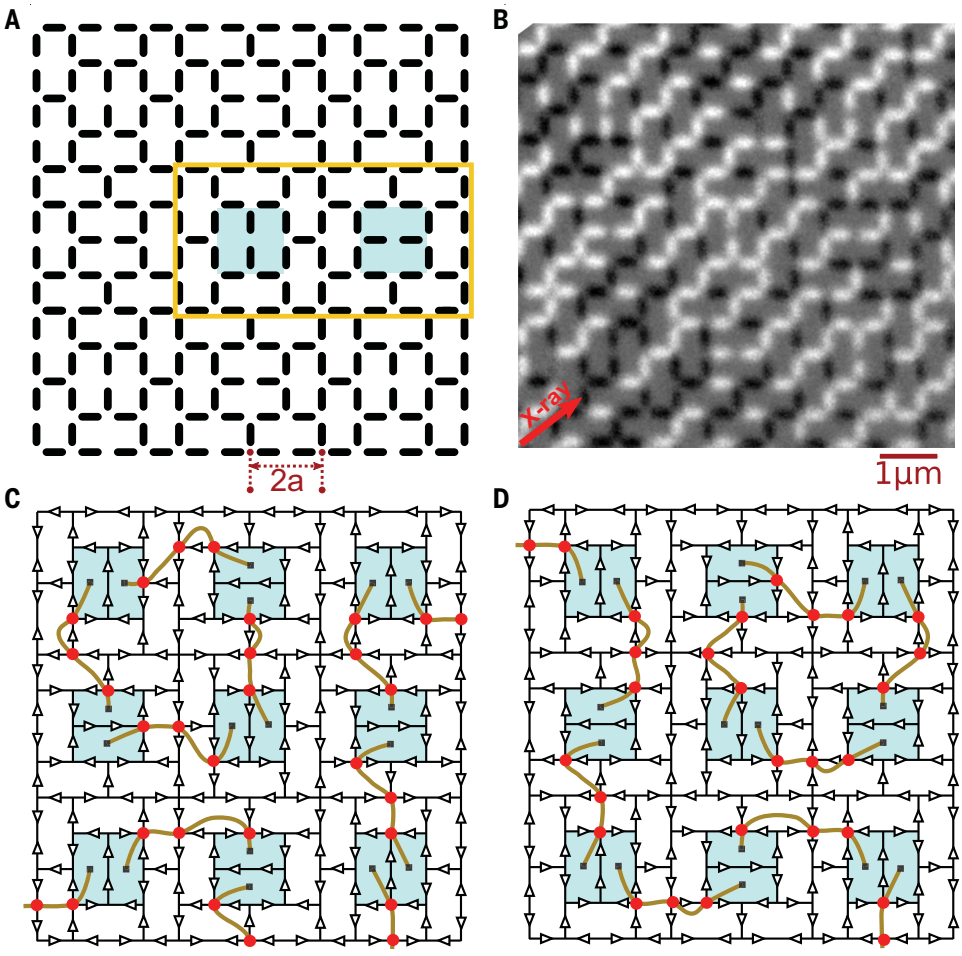


Fig. 1. The Santa Fe ice (SFI) geometry. (**A**) Schematic of the lattice, where each element represents a single-domain nanomagnet and the lattice spacing is *a*. The unit cell (indicated in yellow) has four interior plaquettes (indicated by light blue shading) that are separated by two-island vertices and surrounded by 12 peripheral plaquettes. (**B**) XMCD-PEEM image of SFI, in which all of the islands are either black or white, indicating the magnetic moment direction through its component projected onto the incident x-ray beam direction (red arrow). The lattice is slightly offset from the structure in (A). (**C** and **D**) Illustrations of two SFI disordered ground states, where the excited vertices are indicated by circular red dots. They are in different homotopy classes because the interior plaquettes that are connected by strings are distinct.

homotopy classes. For example, if two string configurations have different interior plaquettes attached to each other, there is no continuous way to transform the strings from one configuration to another. Strings must be created, eliminated, combined, or cut to transform between such configurations, thus breaking continuity and changing the topology.

In terms of energy, each homotopy class has an energy minimum obtained by continuously deforming and shrinking all the strings in any configuration belonging to the class, so that strings have the minimum number of lowest-energy vertices connecting with the string endpoints. If strings in a homotopy class extend to connect interior plaquettes that are not in close proximity, the energy minimum of that homotopy class is necessarily higher than the global ground state energy. To relax to the ground state, the system requires a topological change in the strings to change the homotopy class. Such a process, in which a string changes in such a way that the system transits from one homotopy class to another, is called "topological surgery" (*20*, *21*). The specific process of topological surgery in SFI occurs when two or more strings merge and then separate into a new configuration. This concept of topological surgery—the connection and separation of proximate 1D objects—can also be used to describe diverse phenomena such as chromosome meiosis, DNA recombination, magnetic flux reconnection in plasmas, vortex line fusion and reconnection in classical and quantum liquids, and dislocation lines in metallurgy (*22*–*26*). SFI provides an unusual opportunity for experimentally tracking this process in real time in a thermal system.

On the basis of the topological character of the strings in SFI, we considered the kinetics of our string configurations. We analyzed these kinetics by comparing sequential XMCD-PEEM images (*27*, *28*) and recording the motions (we use the term "motion" to describe any change to a string, including the creation and annihilation of loops). We classified string motions into two categories: trivial and nontrivial. These two categories correspond to motions that continuously deform a string or to motions that include a change in string topology, respectively. In other words, trivial motions do not change the homotopy class of the string configuration in the system, whereas nontrivial changes do. The distinction is demonstrated in Fig. 2, where we show simple examples of each, and in Fig. 3, where we give a taxonomy of trivial and nontrivial motions.

The trivial motions of strings correspond to continuous deformations of strings that retain the same interior plaquettes as endpoints. The different variations of trivial motions are shown in the top schematic of Fig. 3 and are labeled "wiggle" [red (T1), dark blue (T2), and purple (T3)]; "grow" and "shrink" [orange (T4)]; "loop": "wiggle" [green (T5)]; "loop": "grow" and "shrink" [light blue (T6)]; and "loop": "creation" and "annihilation" [pink (T7)]. The loop motions are included as trivial because loops are contractible to zero, and thus loop creation and annihilation are topologically trivial.

The nontrivial motions of strings are those changes that represent a change in the homotopy class of the system—that is, a change in how the interior plaquette endpoints are connected by strings. The different variations of nontrivial motions are shown in the bottom schematic of Fig. 3 and are labeled "merge" and "split" [magenta/red (N1/N2) and light

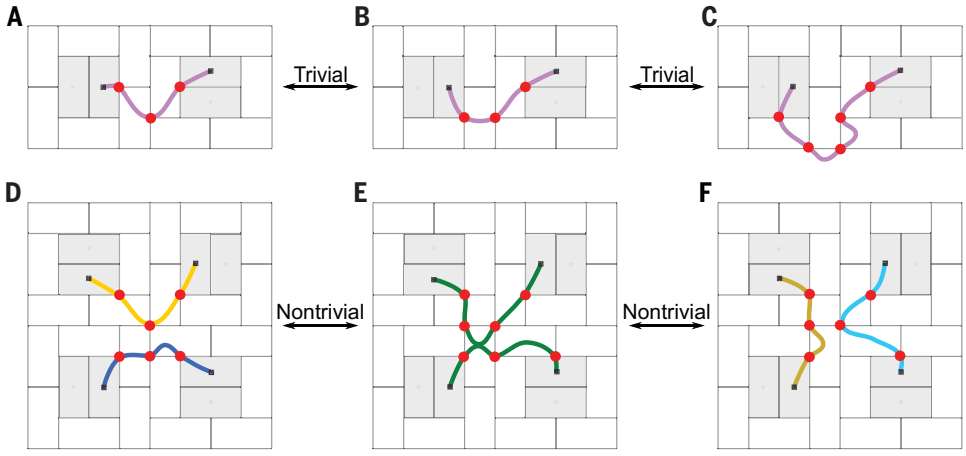
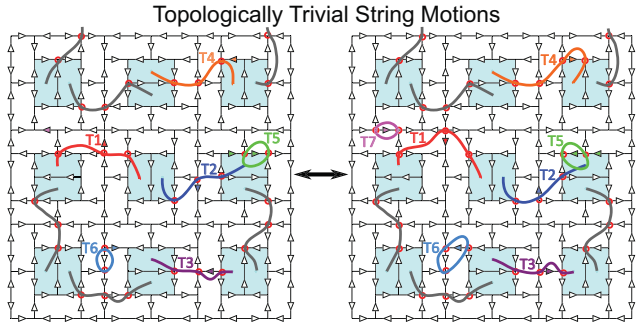
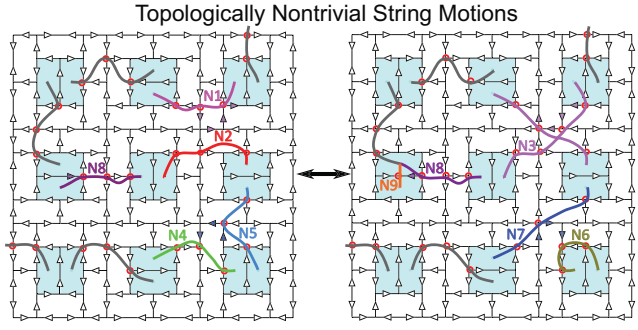


Fig. 2. Trivial and nontrivial string motions. Red dots indicate unhappy vertices. (**A** to **C**) Examples of trivial string motions. Between configurations (A) and (B), there is a "wiggle" motion with no energy change. Between configurations (B) and (C), there is a "grow" motion with energy change because (C) is longer. (**D** to **F**) Examples of nontrivial string motions. Configurations (D) and (F) have pairs of strings that have different pairs of endpoints, and configuration (E) has a merged crossed string with four endpoints. Each of the three configurations represents topologically distinct states, that is, different homotopy classes.



Topologically trivial motions	
Wiggle	Energy increase (T1)
	Energy decrease (T1)
	No energy change (T2,T3)
Grow (T4)	
Shrink (T4)	
Loop	Wiggle (T5)
	Grow (T6)
	Shrink (T6)
	Creation (T7)
	Annihilation (T7)



Topologically nontrivial motions	
Merge (N1,N2→N3)	
Split (N3→N1,N2)	
Reconnection (N4,N5↔N6,N7)	
Adjacent reconnection (N8)	
Strings on adjacent interior plaquette	Creation (N9)
	Annihilation (N9)

Fig. 3. Illustration and taxonomy of various string motions. Island moments are shown as open arrows, and the flipped island moments (solid arrows) correspond to the strings. (Top) Trivial string motions. (Bottom) Nontrivial string motions. The tables list the motions illustrated in the schematics.

purple (N3)]; "reconnection" [green/light blue (N4/N5) and olive/dark blue (N6/N7)]; "adjacent reconnection" [purple (N8)]; and "strings on adjacent interior plaquettes": "creation" and "annihilation" [orange (N9)]. Merge and split are necessarily intermediate steps of a reconnection, but because of the limited time resolution, we labeled a full reconnection distinctly, when we could witness a full reconnection between two consecutive frames. Adjacent reconnection and interior creation and annihilation are in a separate category in that they involve only adjacent interior plaquettes with a shared side and the flip of a single moment along that side. We therefore focus on the topologically nontrivial kinetics associated with strings that traverse the peripheral plaquettes.

In addition to the trivial and nontrivial string motions that we classified, there are strings that naturally do not change between sequential images (that have no motion). There are also complex motions that involve multiple nontrivial changes in moment orientations between two sequential XMCD-PEEM images, as a consequence of the finite time resolution of image acquisition. Furthermore, there are other trivial and nontrivial motions associated with a small number of nanoislands whose magnetizations could not be determined from our XMCD-PEEM images (<0.1%) or that lie on the edges of those images; these motions resulted in incomplete strings in our image recognition program. These other motions and complex motions are not listed in the Fig. 3 tables.

Quantifying string motion

To quantitatively and unambiguously characterize the string motions from our sequential experimental maps of the moments, we used an analysis program that labels each string with two sets: One set comprises the interior plaquettes at which the string ends, and the other set comprises the unhappy vertices it connects. By comparing the elements in these two sets between sequential images, the algorithm can therefore collect the temperature-

dependent prevalence of each type of string motion (*16*). We quantified the string-motion prevalence as N, the average number of string motions between subsequent XMCD-PEEM images per unit cell. N effectively gives the rate of string motions because the images are taken at 1-s intervals, and it can be separated by the type of motion involved.

We then turned to the temperature dependence of our data. We first examined the total vertex energy of the system as a function of temperature, for example, the sum of the magnetostatic energy at each vertex, averaged over all images at a given temperature, as determined by micromagnetic calculations (*16*, *29*). The energy, quantified as the excess above the ground state, has a clear crossover temperature of around T_X = 330 K (Fig. 4A), which is reflected in the kinetics of the strings and is also visible in our previously published data on average string length in the system (*14*).

In Fig. 4, B to D, we show the statistics of motions as a function of temperature, with the temperature regime below T_X indicated by shading. Below T_X, trivial string motions dominate, whereas nontrivial string motions dominate above T_X (the number of trivial motions at higher temperatures is suppressed by a fraction of them being subsumed within the nontrivial motions). The system energy shown in Fig. 4A is nearly flat below T_X; the system does not substantially change its energy as a function of temperature (even though string motions retain considerable temperature dependence). This temperature independence is consistent with the observed crossover in the predominant types of string motions from nontrivial to trivial when crossing through T_X. Substantial reductions in system energy require topological changes in the strings—that is, nontrivial motions—and therefore the system is unable to relax to a lower-energy overall configuration.

Among the nontrivial motions above T_X, the complex motions are predominant (Fig. 4C). Because these represent multiple updates of the moments among XMCD-PEEM images at higher temperature, they also include most other topologically nontrivial motions. By contrast, local processes among adjacent interior plaquettes dominate below T_X (Fig. 4D). This is consistent with the system's topological constraints because such motions require only a single moment flip and make a minimal change to the system's topology.

We next considered the trivial kinetics. As shown in Fig. 5A, the loop processes (light green curves) and wiggles (red curves) are the highest contributors. The grow (blue) and shrink (violet) curves always follow each other, as expected from detailed balance. The same can be said in general for all reversible inelastic processes, as seen also in wiggles corresponding to energy increase and decrease (Fig. 5B) (the increase in energy being caused by the strings occupying a vertex of coordination 4, whose excitations have slightly higher energy than vertices of coordination 3) and loop creation and annihilation (Fig. 5C). Most of the wiggles, however, do not change the total system energy (Fig. 5B), and thus they dominate (Fig. 5A, red curve).

Last, in Fig. 5D we show a log-linear plot of the number of trivial processes, $N_{trivial}$, versus the reciprocal temperature. Within our fairly narrow accessible temperature range below T_X, the data display apparent thermally activated behavior: $N_{trivial}(T) = \frac{1}{\tau_o} e^{-\frac{E_S}{T}}$, where τ_o is the inverse attempt frequency for the thermal processes and E_S is an energetic barrier for changing states. This apparent activated behavior is also observed for the other lattice spacings (*16*). Fits of that form to the data, however, depend very much on our choice of fitting range, with $\tau_o \sim 10^{-5}$ seconds and $E_S \sim 3000$ K for the a = 600 nm sample. We can qualitatively understand this behavior in that the motion of strings, even those with no net energy change, requires moments to flip in direction and overcome the energy barrier associated with the island's shape anisotropy. The energy barrier to flipping a moment, as calculated with micromagnetics at zero temperature, is considerably higher than E_S (*16*), but we expect that the energy barrier for flipping the moment in a real system will be affected by the magnetic texture of the island at elevated temperature (*30*). Similarly, the effective attempt frequency, $1/\tau_o$, is small compared with the typical spin wave frequency, which is in the gigahertz regime. Again, this is perhaps not surprising in that the string motions, even almost all of the trivial ones, are associated with changes in multiple islands.

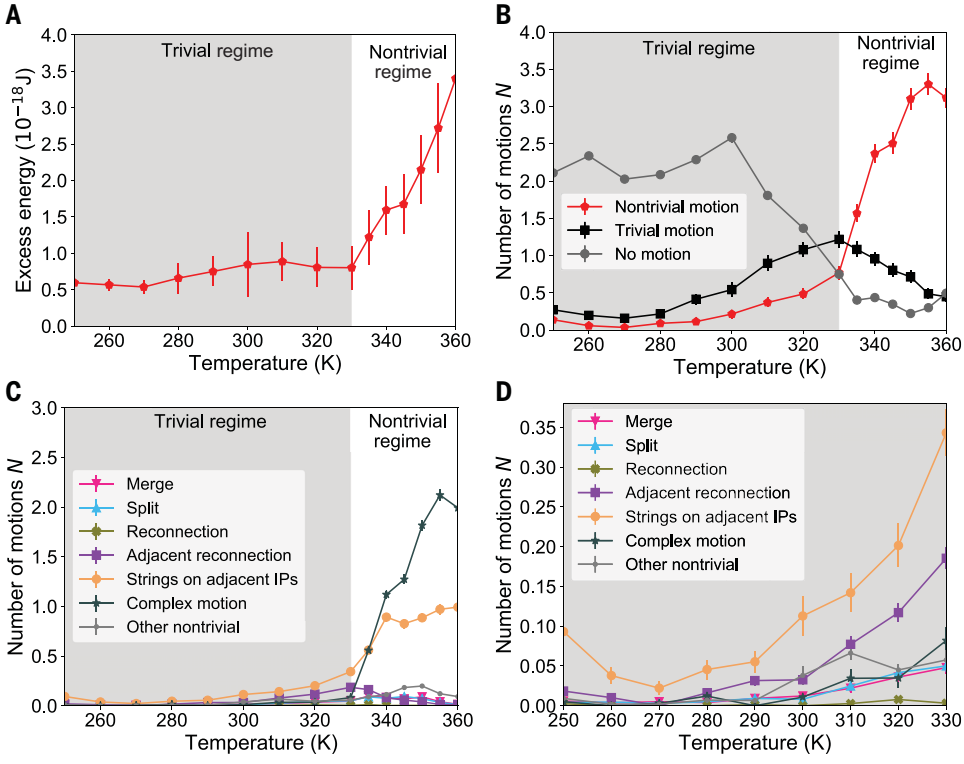


Fig. 4. Temperature-dependent string properties. The shaded regions indicate the low-temperature regime below T_X. (**A**) The excess energy per unit cell versus temperature. (**B**) The temperature dependence of the average number of string motions per XMCD-PEEM image. We also tracked the number of strings that did not change between subsequent images; we plot the total number of these strings as "no motion." (**C**) The temperature dependence of the average number of string motions per XMCD-PEEM image for all types of nontrivial motions. (**D**) An expanded plot of the low-temperature regime from (C). All data are shown per unit cell and are for a = 600 nm SFI; error bars are the SEM from all XMCD-PEEM images taken at the given temperature.

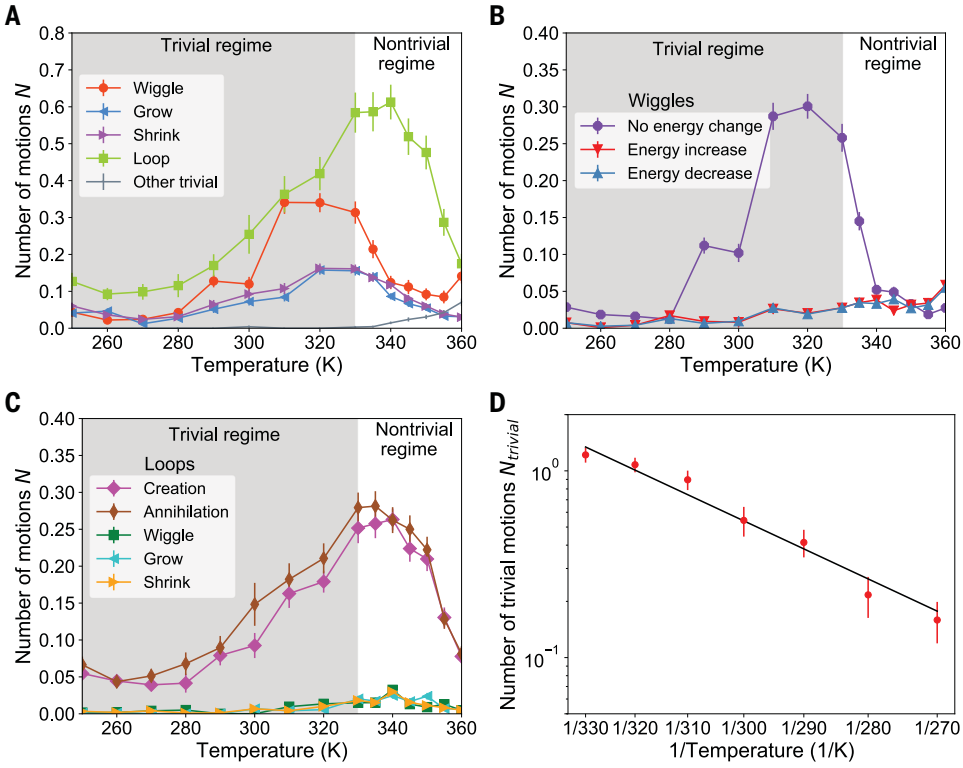


Fig. 5. Temperature dependence of trivial string motions. Shown are the average number of string motions per XMCD-PEEM image for trivial motions. Shading indicates the low-temperature regime in which trivial string motions dominate. (**A**) All types of trivial motions. (**B**) String-wiggle subtypes of trivial motions. (**C**) Loop subtypes of trivial motions. (**D**) The number of total trivial motions as a function of inverse temperature with fit. All data are shown per unit cell and for the *a* = 600 nm sample. The error bars are the SEM from all XMCD-PEEM images taken at the given temperature.

Discussion and outlook

Our observed crossover leads to a consistent picture of how string kinetics evolve. Below T_X, the system does not change its energy, but it is kinetically active through topologically trivial motions (bending and stretching modes). In this regime, the system has a low likelihood of undergoing topological surgery and changing between homotopy classes. In other words, the system's limited ability to explore homotopy classes likely prevents full ergodicity on experimental timescales at low temperatures, and thus the system is impeded from further reducing its global energy to fully realize the ground state. The trivial kinetic processes are thermally equilibrated below T_X, but within a limited phase space of possible string configurations. This state is evidenced by the close tracking of energy-increasing and -decreasing motions (Fig. 5, A to C), which indicates that detailed balance holds, and by the activated behavior shown in Fig. 5D.

By contrast, at high temperature, above T_X, the topologically nontrivial motions become predominant, and the system appears to be fully ergodic. This allows the total energy of the system to change substantially because the different homotopy classes can have radically different energies; this is the regime in which the activated string length was demonstrated previously (*14*).

The observation of a clear crossover between the two regimes correlates somewhat to glass transitions and other dynamic slowdowns, but here it is driven by topology rather than by a disordered potential landscape. We can broadly understand the crossover in that the homotopy class–changing nontrivial motions require substantially longer strings and thus more moments to reverse directions than do the trivial motions, and therefore are kinetically suppressed with decreasing temperature. Although our present data does not allow us to identify the crossover with a dynamic phase transition and associated nonanalytic behavior, the relative sharpness of the crossover is notable relative to most glass transitions, suggesting future studies with frequency-sensitive techniques (*31*–*33*).

Our findings should be generalizable in defining and characterizing kinetic crossovers to nonergodicity beyond systems with topological constraints (*34*). Comparing experimental results with existing theoretical methods to ascertain ergodicity could elucidate slow relaxation after quenches and memory effects and aid in exploring the relation with other kinetic crossovers (*8*, *34*–*36*). Such studies could also have relevance for new forms of computing (*37*) and possibly for quantum tunneling through homotopy classes within qubit realizations (*38*) of SFI or similar structures. Our work might also prove pertinent to systems such as structural glasses, for which dislocation lines ending in vertices proved to be a useful description (*39*–*42*). The advantage of our physical realization of SFI is that it allows direct experimental measurements of these phenomena, suggesting future examinations of other bespoke topologically interesting structures.

REFERENCES AND NOTES

1. K. Huang, *Statistical Mechanics* (John Wiley and Sons, 2000).
2. D. Weaire, N. Rivier, *Contemp. Phys.* **25**, 59–99 (1984).
3. P. Goldbart, N. Goldenfeld, *Phys. Rev. Lett.* **58**, 2676–2679 (1987).
4. T. Aste, D. Sherrington, *J. Phys. Math. Gen.* **32**, 7049–7056 (1999).
5. F. Ritort, P. Sollich, *Adv. Phys.* **52**, 219–342 (2003).
6. J.-Y. Xiong *et al.*, *J. Phys. Chem. B* **109**, 5638–5643 (2005).
7. P. Oikonomou, P. Cluzel, *Nat. Phys.* **2**, 532–536 (2006).
8. R. G. Palmer, *Adv. Phys.* **31**, 669–735 (1982).
9. C. L. Henley, *J. Phys. Condens. Matter* **23**, 164212 (2011).
10. C. Nisoli, *Europhys. Lett.* **132**, 47005 (2020).
11. S. H. Skjærvø, C. H. Marrows, R. L. Stamps, L. J. Heyderman, *Nat. Rev. Phys.* **2**, 13–28 (2020).
12. P. Schiffer, C. Nisoli, *Appl. Phys. Lett.* **118**, 110501 (2021).
13. M. J. Morrison, T. R. Nelson, C. Nisoli, *New J. Phys.* **15**, 045009 (2013).
14. X. Zhang *et al.*, *Nat. Commun.* **12**, 6514 (2021).
15. R. L. Stamps, *Nat. Phys.* **10**, 623–624 (2014).
16. Materials and methods are available as supplementary materials.

17. M. D. Frank-Kamenetskii, A. V. Lukashin, A. V. Vologodskii, *Nature* **258**, 398–402 (1975).
18. D. Michieletto, M. S. Turner, *Phys. World* **27**, 28–31 (2014).
19. D. Michieletto, M. S. Turner, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5195–5200 (2016).
20. J. W. Milnor, in *Proceedings of Symposia in Pure Mathematics*, vol. 3, *Differential Geometry*, C. B. Allendorfer, Ed. (American Mathematical Society, 1961), pp. 39–55.
21. S. Antoniou, S. Lambropoulou, *PLOS ONE* **12**, e0183993 (2017).
22. M. A. Berger, G. B. Field, *J. Fluid Mech.* **147**, 133 (1984).
23. J. Koplik, H. Levine, *Phys. Rev. Lett.* **71**, 1375–1378 (1993).
24. S. Kida, M. Takaoka, *Annu. Rev. Fluid Mech.* **26**, 169–177 (1994).
25. U. F. Kocks, H. Mecking, *Prog. Mater. Sci.* **48**, 171–273 (2003).
26. M. Yamada, R. Kulsrud, H. Ji, *Rev. Mod. Phys.* **82**, 603–664 (2010).
27. A. Farhan *et al*., *Phys. Rev. Lett.* **111**, 057204 (2013).
28. I. Gilbert *et al*., *Nat. Phys.* **12**, 162–165 (2016).
29. A. Vansteenkiste *et al*., *AIP Adv.* **4**, 107133 (2014).
30. S. D. Slöetjes, B. Hjörvarsson, V. Kapaklis, *Appl. Phys. Lett.* **118**, 142407 (2021).
31. M. Pohlit *et al*., *Phys. Rev. B* **101**, 134404 (2020).
32. M. Goryca *et al*., *Phys. Rev. X* **11**, 011042 (2021).
33. M. Goryca *et al*., *Phys. Rev. B* **105**, 094406 (2022).
34. G. Miloshevich, T. Dauxois, R. Khomeriki, S. Ruffo, *Europhys. Lett.* **104**, 17011 (2013).
35. F. Borgonovi, G. L. Celardo, M. Maianti, E. Pedersoli, *J. Stat. Phys.* **116**, 1435–1447 (2004).
36. D. Mukamel, S. Ruffo, N. Schreiber, *Phys. Rev. Lett.* **95**, 240604 (2005).
37. A. Lucas, *Front. Phys.* (Lausanne) doi:10.3389/fphy.2014.00005 (2014).
38. A. D. King, C. Nisoli, E. D. Dahl, G. Poulin-Lamarre, A. Lopez-Bezanilla, *Science* **373**, 576–580 (2021).
39. D. R. Nelson, M. Rubinstein, F. Spaepen, *Philos. Mag. A Phys. Condens. Matter Defects Mech. Prop.* **46**, 105–126 (1982).
40. D. R. Nelson, M. Widom, *Nucl. Phys. B* **240**, 113–139 (1984).
41. S. Sachdev, D. R. Nelson, *Phys. Rev. Lett.* **53**, 1947–1950 (1984).
42. D. R. Nelson, in *Applications of Field Theory to Statistical Mechanics*, L. Garrido, Ed. (Springer, 1985), pp. 13–30.
43. X. Zhang *et al*., Topological kinetic crossover in a nanomagnet array, Dryad (2023); https://doi.org/10.5061/dryad.w0vt4b8w9.
44. X. Zhang *et al*., Topological kinetic crossover in a nanomagnet array, version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.7641526.

ACKNOWLEDGMENTS

We thank C. Castelnovo, N. Goldenfeld, and Y. Shokef for helpful feedback on the manuscript. We also thank R. Chopdekar and R. Koch for support at the PEEM-3 endstation at beamline 11.0.1.1 of the Advanced Light Source during our data acquisition. **Funding:** Work at Yale University was funded by the US Department of Energy (DOE) Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under grant DE-SC0020162. This research used resources of the Advanced Light Source, a DOE Office of Science user facility, under contract DE-AC02-05CH11231. Work at the University of Minnesota was supported by the NSF through grant DMR-2103711. Work at Los Alamos National Laboratory was carried out under the auspices of the DOE through Los Alamos National Laboratory, operated by Triad National Security, (contract 892333218NCA000001), and financed by DOE LDRD. **Author contributions:** J.R. performed film depositions under the guidance of C.L. X.Z. and N.S.B. oversaw the lithography. X.Z., N.S.B., H.S., and I.-A.C. performed the XMCD-PEEM characterization of the thermally active samples. X.Z., G.F., and S.S. analyzed the string structures. C.N. developed the theory and wrote the first draft. P.S. supervised the entire project. All authors contributed to the discussion of results and to the finalization of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Additional experimental data and underlying data from the plots generated in this study are available at Dryad (*43*) and Zenodo (*44*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add6575
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 and S2

Submitted 25 June 2022; accepted 31 March 2023
10.1126/science.add6575

MITORIBOSOME

Molecular basis of translation termination at noncanonical stop codons in human mitochondria

Martin Saurer[1], Marc Leibundgut[1], Hima Priyanka Nadimpalli[2], Alain Scaiola[1], Tanja Schönhut[1], Richard G. Lee[3,4,5], Stefan J. Siira[3,4], Oliver Rackham[3,4,5,6], René Dreos[2], Tea Lenarčič[1], Eva Kummer[7], David Gatfield[2], Aleksandra Filipovska[3,4,5]*, Nenad Ban[1]*

The genetic code that specifies the identity of amino acids incorporated into proteins during protein synthesis is almost universally conserved. Mitochondrial genomes feature deviations from the standard genetic code, including the reassignment of two arginine codons to stop codons. The protein required for translation termination at these noncanonical stop codons to release the newly synthesized polypeptides is not currently known. In this study, we used gene editing and ribosomal profiling in combination with cryo–electron microscopy to establish that mitochondrial release factor 1 (mtRF1) detects noncanonical stop codons in human mitochondria by a previously unknown mechanism of codon recognition. We discovered that binding of mtRF1 to the decoding center of the ribosome stabilizes a highly unusual conformation in the messenger RNA in which the ribosomal RNA participates in specific recognition of the noncanonical stop codons.

Mitochondria maintain their own protein-synthesis machinery required to translate mRNAs transcribed from their highly reduced genome that encodes membrane proteins required for oxidative phosphorylation (OXPHOS) (*1*, *2*). Components of the mammalian mitochondrial translational machinery, such as mitochondrial ribosomes (mitoribosomes) (*3*, *4*), tRNAs, and protein factors, have diverged considerably from their bacterial predecessors (*5*, *6*). Furthermore, several changes to the standard genetic code have occurred in vertebrate mitochondria, including the emergence of unconventional stop codons (*1*, *7*).

Most prokaryotic and eukaryotic translation systems use the stop codons UGA, UAG, and UAA, which are recognized by release factors to terminate protein synthesis. Release factors recognize stop codons in the A-site of the small ribosomal subunit (SSU) to hydrolyze the ester bond between the P-site tRNA and the nascent polypeptide chain in the active site of the large ribosomal subunit (LSU) (*8*–*11*). In human mitochondria, UAG and UAA have retained their function as stop codons, whereas UGA encodes tryptophan (*1*, *12*). Furthermore, two additional stop codons, AGA and AGG, have emerged from reassigned arginine codons and are found at the end of the cytochrome C oxidase subunit 1 (COX1) and the nicotinamide adenine dinucleotide (reduced) (NADH)–ubiquinone oxidoreductase chain 6 protein (ND6) open reading frames (ORFs), respectively (*7*).

In bacteria, the stop codons UAG and UAA are recognized by release factor 1 (RF1), whereas UGA and UAA are read by release factor 2 (RF2) (*9*–*11*). In cytoplasmic translation, eukaryotes use a single factor, eRF1, that binds to all three stop codons (*8*). In human mitochondria, mitochondrial release factor 1a (mtRF1a or mtRF1L) releases ribosomes at UAG or UAA stop codons (*13*–*15*), yet it does not recognize the two noncanonical stop codons AGA and AGG (*13*, *14*) (Fig. 1A). Therefore, which factor releases nascent chains at noncanonical stop codons has remained elusive (*16*). Among previously considered candidates were the mitochondrial release factors ICT1 (mL62, MRPL58) and mtRF-R (C12ORF65) (*13*, *17*–*19*); however, it was subsequently established that they are responsible for the rescue of stalled mitoribosomes on truncated mRNAs or after aberrant translation termination (*8*, *13*, *18*, *20*). It was also proposed that termination at noncanonical stop codons in human mitochondria occurs by mtRF1a, following a −1 ribosomal frameshifting event (*21*). Lastly, mtRF1 was identified as a release factor on the basis of homology, but it was predicted that mtRF1 would not be able to bind to the decoding center because it contains sequence insertions in the stop codon–recognition domain that would prevent interactions with the mRNA (*22*, *23*). Several in vitro studies did not detect binding of mtRF1 to ribosomes programmed with AGA or AGG codons (*13*, *15*, *17*, *19*, *24*). Recently, mtRF1 has been shown to participate in the translation termination of COX1 but no[t] ND6; however, the mechanistic role of [the] factor in translation termination remained elusive (*25*).

In this study, we used biochemical and structural approaches to establish that mtRF1 is responsible for termination at noncanonical stop codons in human mitochondria. We also show that the specific recognition of the stop codon occurs through a network of interactions between the codon, mtRF1, and the ribosomal RNA (rRNA).

Results and discussions

mtRF1 is required for translation termination of COX1

It is currently unclear which factor is responsible for the release at noncanonical stop codons. We used genome editing to delete each of the four mitochondrial release factors mtRF1, mtRF1a, mtRF-R, and ICT1, as well as for the simultaneous deletion of both mtRF1 and mtRF1a in human CAL51 cells. Loss of mtRF1 led to a reduction of the steady state (Fig. 1B) and de novo synthesis (Fig. 1C) of the mitochondrially encoded polypeptide COX1, whereas loss of mtRF1a and the double deletion ("knockout," KO) of mtRF1 and mtRF1a caused global reduction in mitochondrial protein synthesis (Fig. 1C). This indicates that mtRF1 is specifically required for efficient translation termination of COX1. The observed residual production of COX1 in the absence of mtRF1 (Fig. 1C) is likely because of the activity of rescue factor mtRF-R (*20*, *25*). Deletion of mtRF-R or ICT1 resulted in marked and general reduction of mitochondrial protein synthesis (fig. S1A), which was consistent with their role as rescue factors in translation (*15*, *20*). COX1 is needed for the initial assembly of the catalytic core of the OXPHOS complex IV, and diminished levels of COX1 are likely to disrupt complex IV assembly and reduce the stability of members of complex IV (*26*). Accordingly, we observed reduced steady-state abundance of COX2 in mtRF1-KO cells (Fig. 1B). The decreased levels of complex IV proteins in the absence of mtRF1 resulted in reduced mitochondrial respiration (fig. S1B). Moreover, we conducted ribosome footprinting experiments (*27*, *28*), which identified that loss of mtRF1 leads to the stalling of mitoribosomes on the noncanonical stop codons at the end of the COX1 and ND6 mRNAs (Fig. 1, D and E), but generally not on the canonical stop codons UAA and UAG (fig. S2). Codon occupancy analyses further corroborated the increased stalling of mitoribosomes at noncanonical codons AGA and AGG in the mtRF1-KO cells (Fig. 1F). These results reveal that noncanonical AGA or AGG termination at the end of COX1 and ND6 genes is specifically compromised in the absence of mtRF1. Next, we addressed whether the catalytic activity of

[1]Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zürich, 8093 Zürich, Switzerland. [2]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. [3]Harry Perkins Institute of Medical Research, The University of Western Australia, Nedlands, Western Australia 6009, Australia. [4]ARC Centre of Excellence in Synthetic Biology, QEII Medical Centre, The University of Western Australia, Nedlands, Western Australia 6009, Australia. [5]Telethon Kids Institute, Northern Entrance, Perth Children's Hospital, 15 Hospital Avenue, Nedlands, Western Australia, Australia. [6]Curtin Medical School and Curtin Health Innovation Research Institute, Curtin University, Bentley, Western Australia 6102, Australia. [7]Novo Nordisk Foundation Center for Protein Research, Protein Structure and Function Program, Blegdamsvej 3B, 2200 København N, Denmark.
*Corresponding author. Email: aleksandra.filipovska@uwa.edu.au (A.F.); ban@mol.biol.ethz.ch (N.B.)

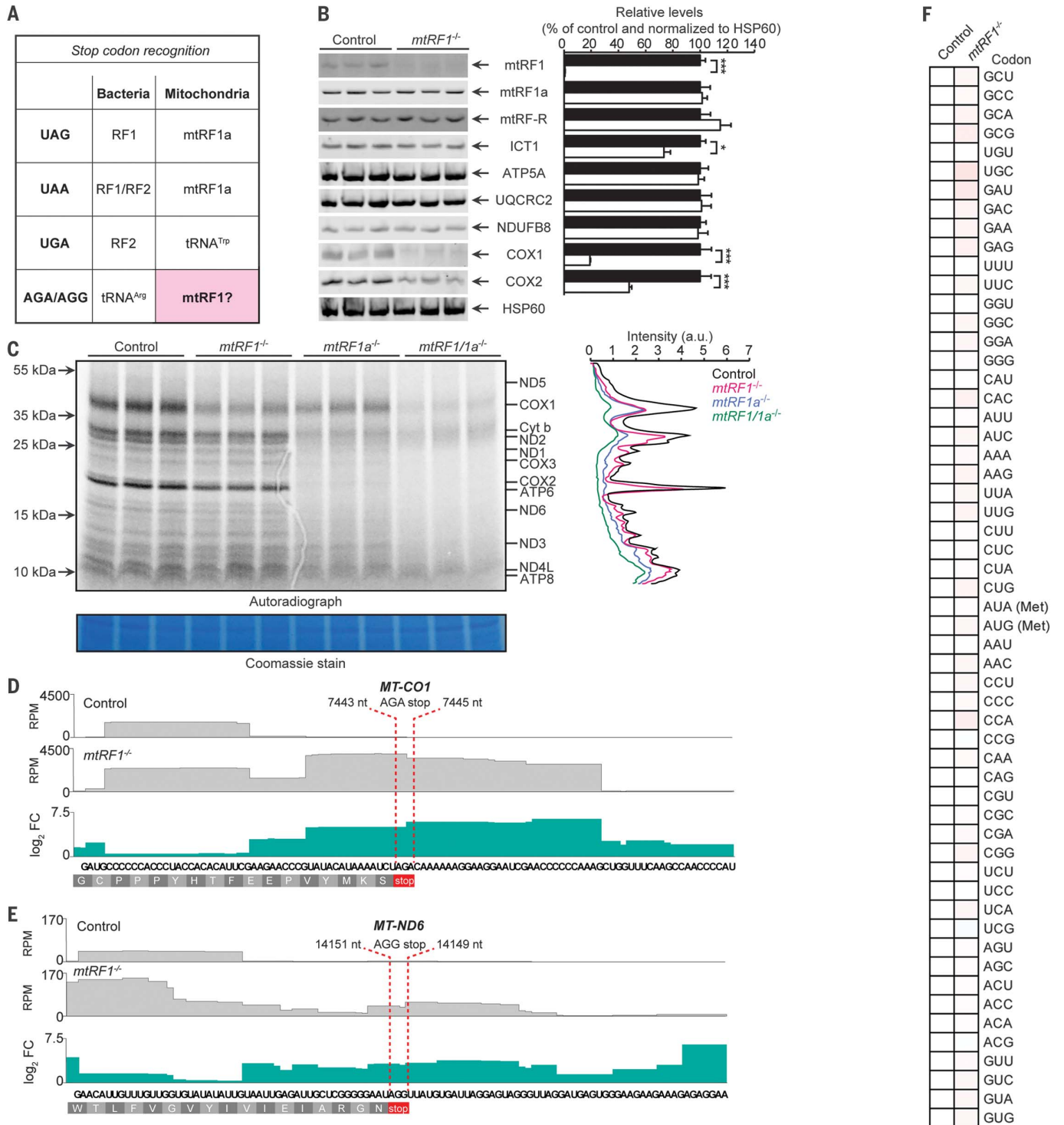


Fig. 1. Loss of mtRF1 reduces the translation of specific mitochondrially encoded mRNAs. (**A**) Comparison of stop-codon occurrence and recognition in bacteria and human mitochondria. (**B**) Immunoblotting against mitochondrial proteins in control and *mtRF1*$^{-/-}$ cells; HSP60 was used as a loading control. Relative abundance of proteins was measured with the Image Studio Lite 5.2 software and normalized to the loading control. Data are mean ± SEM. *$P < 0.05$; ***$P < 0.001$, determined using Student's *t* test. (**C**) De novo mitochondrial protein synthesis was measured by ^{35}S[methionine] and ^{35}S[cysteine] incorporation in control, *mtRF1*$^{-/-}$ and *mtRF1a*$^{-/-}$ cells ($n = 3$). 20 μg of cell lysates were separated by SDS-PAGE and visualized by autoradiography. Coomassie-stained gels are shown as loading controls. (**D** to **F**) Ribosomal profiling was carried out on translating mitoribosome fractions isolated from *mtRF1*$^{-/-}$ and control cells. Normalized read counts are shown in gray at the termination regions of *MT-CO1* (D) and *MT-ND6* (E) in *mtRF1*$^{-/-}$ and control cells, and the $\log_2$ fold-changes between them show increased (turquoise) and decreased (orange) coverage in specific regions of the mRNA ends. (F) Heatmap showing the $\log_2$ fold change difference in codon occupancy in the mtRF1 KO compared to control mitoribosome profiles. The $\log_2$ fold-change shown ranges from −10 to 10.

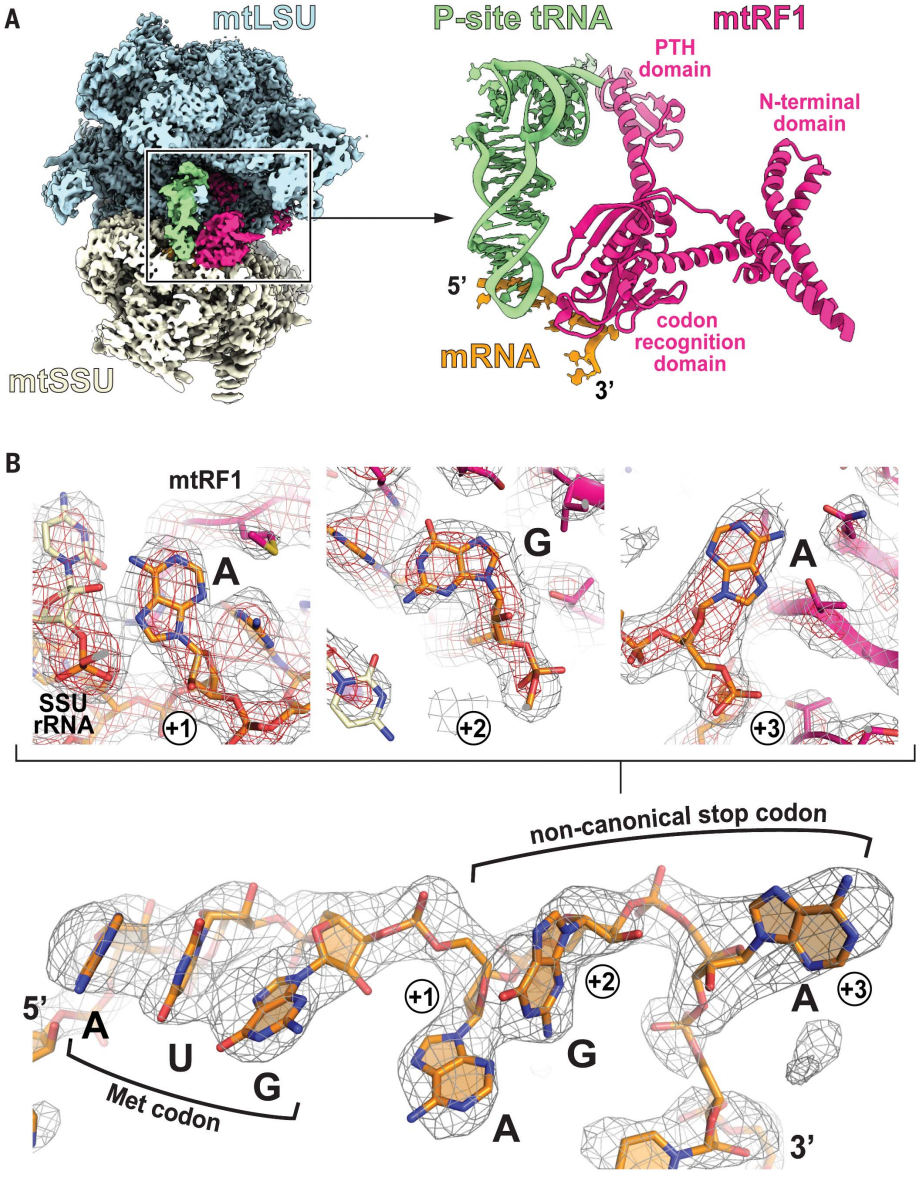


Fig. 2. Cryo-EM densities of the reconstituted mammalian mitoribosome–mtRF1^{wt} complex. (**A**) Overview of mtRF1 (magenta) bound to the A-site of the 55*S* mitoribosome with mRNA (orange) and tRNA^{Met} (green) in the P-site shown in the filtered cryo-EM reconstruction. A model of the visualized release factor, the mRNA, and the tRNA is shown in the same orientation. (**B**) Detailed views of cryo-EM densities that represent the noncanonical AGA stop codon in the mRNA shown at two different thresholds (gray and red). Whereas tRNA^{Met} base-pairs with the AUG codon in the P-site, the three purines of the noncanonical stop codon AGA interact with the codon recognition domain of mtRF1 in the A-site.

mtRF1 is required for translation termination. The activity of the conserved peptidyl-tRNA hydrolysis (PTH) domain of release factors can be suppressed by mutation of the catalytic GGQ motif to AAQ, which renders them unable to release the nascent chain (*8*). When expressed in *mtRF1*$^{-/-}$ cells, mtRF1^{AAQ} was not able to rescue translation of COX1 (fig. S1C), which indicates that the catalytic activity of mtRF1 is required for the synthesis of COX1. Concomitant with the strong decrease in COX1 levels in *mtRF1*$^{-/-}$ cells, we observed that mitochondrial protein biosynthesis was globally diminished, although to a lesser degree than COX1 itself. Conceivably, this global reduction was caused by a reduced pool of translation-competent mitoribosomes, because COX1 polypeptides cannot be efficiently released in the absence of mtRF1. These results show that mtRF1 is the release factor responsible for the recognition of noncanonical stop codons. This conclusion is supported by the work of two independent research groups who published their findings during the revision of this manuscript (*25*, *29*). However, these results also raise the question of how the unusual codon-recognition domain of mtRF1 accomplishes this task.

Structure determination of the mtRF1 termination complex

To gain insights into the mechanism of codon recognition by mtRF1, we solved the structure of an in vitro–reconstituted complex of the mitoribosome and mtRF1 by single-particle cryo-EM. Human embryonic kidney 293 (HEK293) cells were transfected with a plasmid that encoded the human wild-type (WT) mtRF1 protein with a C-terminal FLAG tag. Following the isolation and lysis of mitochondria, mtRF1^{wt}-FLAG was affinity-purified with anti-FLAG resin. Mitoribosomal subunits were purified from pig liver tissue and combined with a short mRNA (5′-A AUG AGA CAA-3′) and the human mitochondrial formylmethionine (fMet)–tRNA^{Met} to obtain translation termination complexes with the noncanonical stop codon AGA in the A-site of the mitoribosome. Purified mtRF1^{wt}-FLAG was added and after a short incubation, the mixture was prepared for analysis by single-particle cryo-EM. We found 55*S* mitoribosomes with bound mRNA, the tRNA in the P-site, and mtRF1 bound to the A-site. We were able to calculate a map for this complex that reached an overall resolution of 3.6 Å, and additional refinement focused on either the SSU- or the LSU-yielded maps at overall resolutions of 3.6 Å and 3.5 Å, respectively (Fig. 2 and figs. S3 and S4). To build an atomic model for mtRF1, we used an AlphaFold2 model (*30*) and a recently published structure of mtRF1a (*13*) as a guide.

mtRF1 binds to a highly distorted mRNA conformation

Although mtRF1 and mtRF1a share a similar fold, regions with considerable differences can be found in the N-terminal domain, where an extension containing positively charged amino acids is found in mtRF1 (Fig. 3, A and C, and fig. S5A). In our reconstruction, we see this extension protruding into a pocket below the L7/L12 stalk of the LSU that is formed by the tips of helices of the LSU rRNA, which is an interaction that might help to stabilize mtRF1 binding (Fig. 2A).

Whereas the catalytic PTH domain of mtRF1 that releases the peptide in the active site of the ribosome is highly conserved, the codon-recognition domain features insertions at the tip of helix α5 and in the codon-recognition loop that are incompatible with canonical codon recognition by mtRF1 (Fig. 3). The two additional amino acids (threonines 194 and 195) extend the tip of helix α5 in the A-site of the SSU, thereby distorting the mRNA backbone and reorienting the first two bases of the noncanonical stop codon (Fig. 3B). The structure of the codon-recognition loop also markedly differs from canonical mtRF1a because it

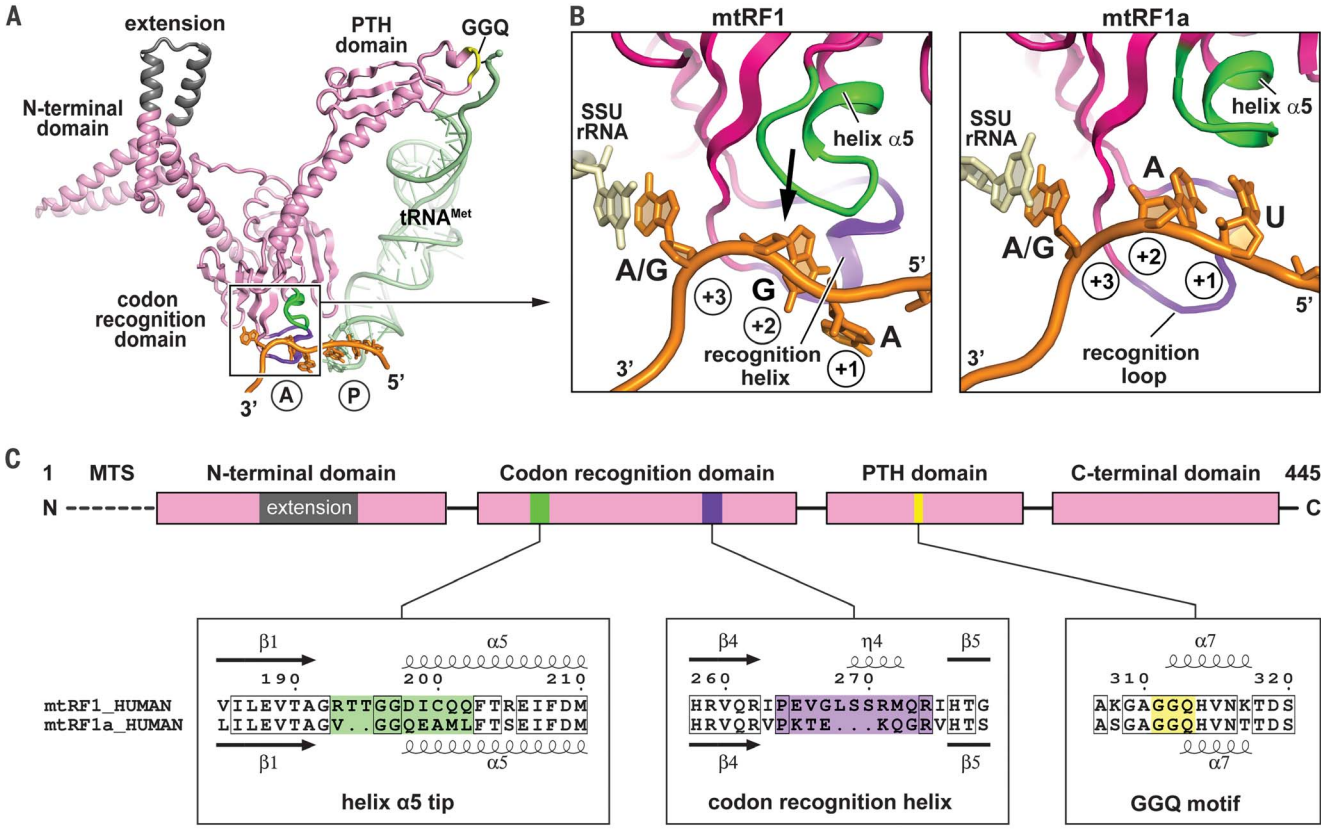


Fig. 3. Comparison between the mRNA recognition in the A-site of the SSU by the noncanonical mtRF1 and the canonical mitochondrial release factor mtRF1a. (**A**) Overview of release factor mtRF1 (pink) bound to mRNA (orange) in the A-site of the decoding center in the SSU, with $tRNA^{Met}$ in the P-site. The extension of the N-terminal domain is highlighted in gray. The location of the catalytic GGQ motif is shown in yellow. (**B**) Detailed view of the A-site of mtRF1 (this study) and comparison with mtRF1a (*13*). The loop-preceding helix α5 (green) is extended in mtRF1 and pushes the purine bases at the first two positions away from the recognition helix (purple), whereas in mtRF1a, the corresponding bases are embraced by the recognition loop. The purine in the third position is held in a similar position by amino acids that are conserved in both factors, as well as by a stacking base of the rRNA. (**C**) Schematic domain overview of mtRF1. Alignments between the functionally important regions of mtRF1 and mtRF1a are shown below. MTS, mitochondrial targeting sequence. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

forms a short α-helix with several key residues oriented toward the first and second base of the noncanonical stop codon (Fig. 3B and fig. S5). Consequently, we refer to it as the "recognition helix." Instead of embracing the codon in the same way that the recognition loop of mtRF1a does, the recognition helix of mtRF1 pushes the first two bases of the codon away such that they point toward the head of the SSU (Fig. 3B). Binding of mtRF1 further results in a distortion in helix 44 of the SSU rRNA, with the decoding bases alternatively flipped (fig. S6).

As an independent validation, we solved the structure of the human translation termination complex purified from mitochondria. We expressed FLAG-tagged, catalytically defective mutant $mtRF1^{AAQ}$ in HEK293 cells to affinity-purify stalled noncanonical termination complexes for structure determination by cryo-EM. The obtained reconstruction of the human termination complex at 3.0 Å overall resolution (figs. S7 to S9) revealed the same conformation of the mRNA and of the codon-recognition domain of $mtRF1^{AAQ}$ as in the in vitro–reconstituted complex. On the basis of the visual inspection of the densities of the mRNA and the P-site tRNA, we could establish that the ribosomes contained the mRNA sequence AU(G/A)AG(G/A) with $tRNA^{Met}$ in the P-site interacting with AU(G/A) and the release factor interacting with AG(G/A) (figs. S9 and S10). Because these sequences neither correspond to the end of the *MT-COX1* nor to the *MT-ND6* ORFs, we performed ribosome profiling on pulled-down $mtRF1^{AAQ}$ complexes to identify which sequences the ribosomes may be interacting with. In addition, we also profiled whole mitochondria from $mtRF1^{AAQ}$-expressing cells to evaluate the specificity of the obtained footprint signals. This analysis revealed ribosome footprints at several AU(G/A)AG(G/A) sites located on the (+) strand transcript, two of which were located in the −1 frame within *MT-ND4* (fig. S11, C and D) and one in the −1 frame within *MT-ND5* (fig. S11, C and E). In comparison with whole-mitochondria footprints, a local peak in coverage on these sites was apparent in pulled-down $mtRF1^{AAQ}$-complex data. The preferential identification of complexes with $tRNA^{Met}$ in the P-site of the mitoribosome may be explained by the conserved, stabilizing features of its anticodon stem-loop (*31*) and the availability of AU(G/A)AG(G/A) sites located outside of known reading frames (fig. S11).

mtRF1 recognizes noncanonical stop codons with help from rRNA

The structures of translation termination complexes allowed us to establish the chemical basis of stop-codon recognition. The distortion of the mRNA that occurs owing to the unusual structure of the mtRF1 codon-recognition domain positions the noncanonical stop codons such that they are specifically recognized by a combination of chemical features in mtRF1, the surrounding bases of the SSU rRNA, and neighboring bases of the mRNA. We describe

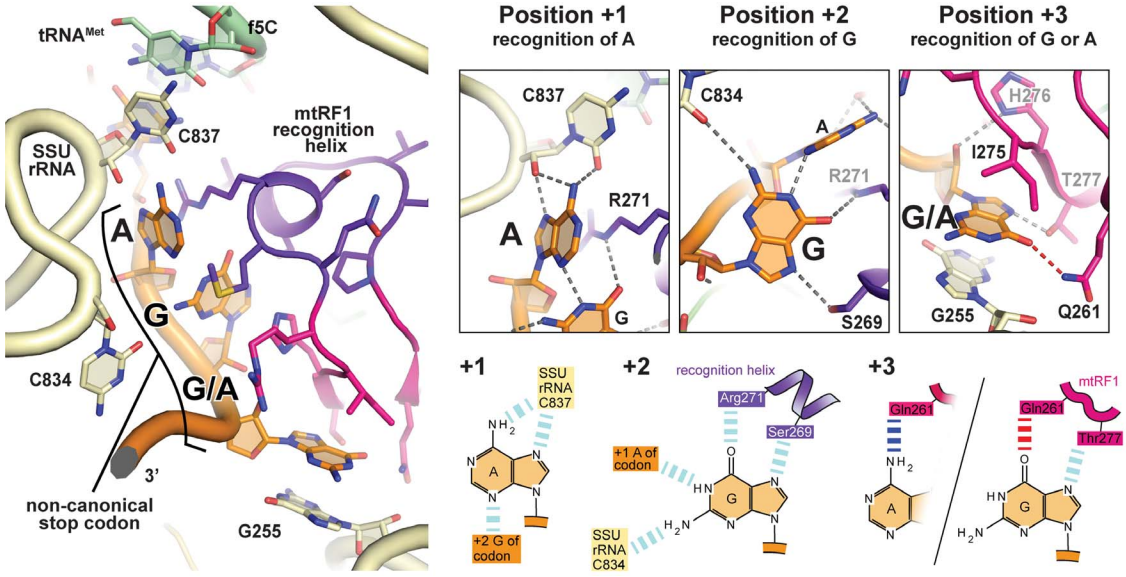


Fig. 4. The structural basis of noncanonical stop codon recognition. The A-site of the human mitoribosome in complex with mtRF1 is shown. A network of interactions that includes residues of mtRF1, the SSU rRNA, and the mRNA itself define the identity of the first two purine bases as A and G, respectively, whereas at the third position, both purine bases can be specifically recognized. Important interactions at all three stop codon positions are shown in detailed views, with corresponding schematics below.

these interactions as visualized for the human translation-termination complex.

The adenosine at the first position is stabilized by stacking onto arginine 271 of the codon-recognition helix (Fig. 4 and fig. S5F). In this orientation, the Hoogsteen edge of the adenosine forms hydrogen bonds with the sugar edge of cytosine C837 (human nomenclature) of the SSU rRNA. Owing to the number of possible hydrogen bonds and the complementary partial charges, adenosine is favored in this position over all other bases. In particular, the carbonyl group of guanosine, the other purine, would not be able to form favorable interactions with C837. In this way, mtRF1 does not directly detect the identity of the first base through amino acid side chains; rather, the helix α5 loop and the codon-recognition loop orient the first base such that it forms specific hydrogen bonds with the rRNA only in case of an adenosine as the first base.

The side chain of arginine 271 is also involved in the recognition of the base in the +2 position. Its positively charged guanidinium group interacts with the carbonyl group of the +2 guanosine, thereby arranging the guanosine almost perpendicularly relative to the +1 adenosine (Fig. 4). This conformation allows a hydrogen bond to form between the N1 atom of guanosine and the N3 atom of the +1 adenosine. As is the case for the first adenosine, a base of the SSU rRNA is also involved in the recognition of the second guanosine: The amino group of the guanosine faces the 2-hydroxyl group of C834 (human nomenclature) and form a hydrogen bond. Additionally, the +2 guanosine is sterically hindered from rotating away by the extended loop at the tip of helix α5 and an additional hydrogen bond with serine 269 that resides in the codon-recognition helix. Altogether, these extensive and specific interactions strongly favor a guanosine over all other bases in the second position.

In the third position of the A-site codon, both guanosine and adenosine can be recognized without any conformational changes of the factor or the mRNA. This is consistent with our observation that the structure of the in vivo–purified sample with a mixture of A and G in the third position was identical to the structure of the reconstituted complex that contained the AGA codon. The purine base at the last position of the stop codon is sandwiched between G255 (human nomenclature) of the SSU rRNA and isoleucine 275 of mtRF1 and is held in place by hydrogen bonds with threonine 277 and with the backbone carbonyl atom of isoleucine 275. Finally, both the carbonyl group of a guanosine or the amino group of an adenosine would be able to form hydrogen bonds with glutamine 261 depending on the orientation of the side chain. This mode of third-base recognition is highly similar to what was previously observed for mtRF1a and bacterial RF1, release factors that can recognize both G and A in the third position of stop codons (*10*, *13*, *32*). Therefore, mtRF1 recognizes both purines in the third position of the stop codon, which is in agreement with our observation that the stalling of mitoribosomes occurs at both noncanonical stop codons in the absence of mtRF1 (Fig. 1, D to F, and fig. S2).

In conclusion, our cryo-EM reconstructions reveal how mtRF1 interacts with noncanonical stop codons in a highly specific manner. In contrast with all other described release factors, mtRF1 has to recognize two stop codons with adenosine at the first position instead of uracil. To meet this requirement, mtRF1 has evolved a codon-recognition mechanism that is distinct from all other canonical release factors. Rather than identifying the codon in the A-site by direct interactions of amino acid side chains with the bases, as observed in bacterial RF1 and mtRF1a (*10*, *13*), mtRF1 repositions the first two bases of the stop codon to use an intricate network of interactions that includes residues of the release factor, the rRNA of the SSU, as well as neighboring bases of the mRNA. These interactions are only established if the identity of the first two bases is "AG," whereas both G and A are recognized in the third position. It follows that mtRF1 specifically recognizes the codons AGA and AGG and selects against all other sense or stop codons that are unable to form the observed network of interactions. The role of mtRF1 in translation termination on AGA and AGG codons is likely broadly conserved in vertebrates owing to a high level of sequence similarity in the key regions of the release factor involved in codon recognition (*32*). The results presented here reveal the molecular machinery that resulted from an extremely rare evolutionary recoding event, which affected codon usage across the entire genome and required replacement of a particular tRNA by a remodeled release factor.

REFERENCES AND NOTES

1. S. Anderson *et al.*, *Nature* **290**, 457–465 (1981).
2. O. Rackham, A. Filipovska, *Nat. Rev. Genet.* **23**, 606–623 (2022).
3. A. Amunts, A. Brown, J. Toots, S. H. W. Scheres, V. Ramakrishnan, *Science* **348**, 95–98 (2015).
4. B. J. Greber *et al.*, *Science* **348**, 303–308 (2015).
5. E. Kummer, N. Ban, *Nat. Rev. Mol. Cell Biol.* **22**, 307–325 (2021).
6. P. Bieri, B. J. Greber, N. Ban, *Curr. Opin. Struct. Biol.* **49**, 44–53 (2018).
7. S. Osawa, T. Ohama, T. H. Jukes, K. Watanabe, *J. Mol. Evol.* **29**, 202–207 (1989).
8. A. Brown, S. Shao, J. Murray, R. S. Hegde, V. Ramakrishnan, *Nature* **524**, 493–496 (2015).
9. A. Korostelev *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19684–19689 (2008).
10. M. Laurberg *et al.*, *Nature* **454**, 852–857 (2008).
11. A. Weixlbaumer *et al.*, *Science* **322**, 953–956 (2008).

12. B. G. Barrell, A. T. Bankier, J. Drouin, *Nature* **282**, 189–194 (1979).
13. E. Kummer, K. N. Schubert, T. Schoenhut, A. Scaiola, N. Ban, *Mol. Cell* **81**, 2566–2582.e6 (2021).
14. Y. Nozaki, N. Matsunaga, T. Ishizawa, T. Ueda, N. Takeuchi, *Genes Cells* **13**, 429–438 (2008).
15. H. R. Soleimanpour-Lichaei *et al*., *Mol. Cell* **27**, 745–757 (2007).
16. F. Nadler, E. Lavdovskaia, R. Richter-Dennerlein, *RNA Biol.* **19**, 117–131 (2021).
17. S. Akabane, T. Ueda, K. H. Nierhaus, N. Takeuchi, *PLOS Genet.* **10**, e1004616 (2014).
18. H. A. Feaga, M. D. Quickel, P. A. Hankey-Giblin, K. C. Keiler, *PLOS Genet.* **12**, e1005964 (2016).
19. C. Lind, J. Sund, J. Aqvist, *Nat. Commun.* **4**, 2940 (2013).
20. N. Desai *et al*., *Science* **370**, 1105–1110 (2020).
21. R. Temperley, R. Richter, S. Dennerlein, R. N. Lightowlers, Z. M. Chrzanowska-Lightowlers, *Science* **327**, 301 (2010).
22. M. A. Huynen, I. Duarte, Z. M. Chrzanowska-Lightowlers, S. B. Nabuurs, *Biol. Direct* **7**, 14 (2012).
23. Y. Zhang, L. L. Spremulli, *Biochim. Biophys. Acta* **1443**, 245–250 (1998).
24. I. Duarte, S. B. Nabuurs, R. Magno, M. Huynen, *Mol. Biol. Evol.* **29**, 3497–3512 (2012).
25. F. Nadler *et al*., *Nat. Commun.* **13**, 6406 (2022).
26. S. Dennerlein, P. Rehling, *J. Cell Sci.* **128**, 833–837 (2015).
27. R. J. Morscher *et al*., *Nature* **554**, 128–132 (2018).
28. E. Schöller *et al*., *Mol. Cell* **81**, 4810–4825.e12 (2021).
29. A. Krüger *et al*., *Nat. Commun.* **14**, 30 (2023).
30. J. Jumper *et al*., *Nature* **596**, 583–589 (2021).
31. E. Kummer *et al*., *Nature* **560**, 263–267 (2018).
32. D. J. Young *et al*., *RNA* **16**, 1146–1155 (2010).

ACKNOWLEDGMENTS

We are grateful to the members of the Ban lab, in particular K. Schubert, A. Picenoni, R. Jost, and D. Yudin, for help with sample purification and helpful discussions. We thank M. Stentenbach from the Filipovska group for help with the respiration experiments. We are indebted to M. Peterek, B. Qureshi, D. Boehringer, and the ETH Scientific Center for Optical and Electron Microscopy (ScopeM) for support during cryo-EM data collection. **Funding:** Swiss National Science Foundation (SNSF) individual grant 179190 (D.G.); SNSF individual grants 31003A_182341 and 310030_212308 (N.B.); Swiss National Center of Competence in Research (NCCR), RNA & Disease program, project funding 141735 (D.G.), 51NF40-182880, and 51NF40-205601 (N.B.); Roessler Prize, Ernst Jung Prize, and Otto Naegeli Prize for Medical Research (N.B.); European Molecular Biology Organization (EMBO) long-term fellowship grant 1074-2019 (T.L.); National Health and Medical Research Council (NHMRC), Australian Research Council (ARC), Western Australian Child Research Fund (WACRF), and Cancer Council of Western Australia (CCWA) fellowships and project grants (O.R. and A.F.).
Author contributions: E.K. and T.L. carried out initial experiments. R.G.L., S.J.S., O.R., and A.F. carried out cell biology and *mtRF1*$^{-/-}$ ribo-profile experiments. M.S., T.S., and E.K. purified and prepared samples for cryo-EM studies. A.S. and M.S. collected and processed cryo-EM data. M.S. and M.L. performed molecular model building and refinement. M.S., M.L., and N.B. interpreted the density. H.P.N., R.D., and D.G. designed, executed, and interpreted the ribo-profile experiments with mitochondria and purified mitoribosomes from mtRF1^{AAQ}-expressing cells. M.S., M.L., and N.B. drafted the manuscript. H.P.N., A.S., R.G.L., S.J.S., O.R., R.D., E.K., T.L., D.G., and A.F. reviewed and edited the manuscript.
Competing interests: Authors declare that they have no competing interests. **Data and materials availability:** The coordinates and corresponding cryo-EM map of the in vitro reconstituted mammalian 55*S*-mtRF1^{wt}-mRNA-tRNA^{Met} complex are available in the Protein Data Bank (PDB) as PDB ID 8OIN and in the Electron Microscopy Data Bank (EMDB) as EMD-16894, respectively. The structures and maps focused on either the SSU (28S-mtRF1^{wt}-mRNA-tRNA^{Met}) or the LSU (39S-mtRF1^{wt}-tRNA^{Met}) are deposited under the accession codes PDB 8OIP and EMD-16895 and PDB 8OIQ and EMD-16896, respectively. The coordinates and corresponding cryo-EM map of the human 55*S*-mtRF1^{AAQ}-mRNA-tRNA^{Met} complex are available as PDB ID 8OIR and as EMD-16897, respectively. The structures and maps focused on either the SSU (28S-mtRF1^{AAQ}-mRNA-tRNA^{Met}) or the LSU (39S-mtRF1^{AAQ}-tRNA^{Met}) are deposited under the accession codes PDB 8OIS and EMD-16898, and PDB 8OIT and EMD-16899, respectively. Sequencing data for ribo-profiling from isolated mitochondria of control and mtRF1^{AAQ}-expressing cells have been deposited at Gene Expression Omnibus (GEO) GSE216779. Sequencing data for ribo-profiling from *mtRF1*$^{-/-}$ and control mitochondria have been deposited at GEO GSE216979. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf9890
Materials and Methods
Figs. S1 to S11
Table S1 and S2
References (*33*–*61*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 25 November 2022; accepted 1 April 2023
10.1126/science.adf9890

LOCUST SWARMING

A chemical defense deters cannibalism in migratory locusts

Hetan Chang[1], Sina Cassau[2], Jürgen Krieger[2], Xiaojiao Guo[3], Markus Knaden[1], Le Kang[3]†, Bill S. Hansson[1]*†

Many animals engage in cannibalism to supplement their diets. Among dense populations of migratory locusts, cannibalism is prevalent. We show that under crowded conditions, locusts produce an anticannibalistic pheromone called phenylacetonitrile. Both the degree of cannibalism and the production of phenylacetonitrile are density dependent and covary. We identified the olfactory receptor that detects phenylacetonitrile and used genome editing to make this receptor nonfunctional, thereby abolishing the negative behavioral response. We also inactivated the gene underlying phenylacetonitrile production and show that locusts that lack this compound lose its protection and are more frequently exposed to intraspecific predation. Thus, we reveal an anticannibalistic feature built on a specifically produced odor. The system is very likely to be of major importance in locust population ecology, and our results might therefore provide opportunities in locust management.

Cannibalism is defined as the killing and consumption of all or parts of a conspecific. Although it can sometimes function to eliminate competitors (*1*, *2*) or be the result of interactions connected to mating (*3*), in most cases it is a means of supplementing nutrition (*4*, *5*). When practiced frequently, cannibalism has important ecological consequences for population dynamics and stability, interspecific trophic interactions, and pathogen transmission and epidemiology (*6*–*8*). For predatory animals, cannibalism is a straightforward extension of diet, and many predators include conspecifics among their prey (*9*). A broader range of taxa, including insects, engage in egg cannibalism, which is facilitated by the eggs' small relative size, immobility, and vulnerability (*10*), or in predation on much smaller juvenile conspecifics (*6*). For prey of cannibalism, strong selection pressures favor the development of anticannibalistic strategies to deter potential predators of their own kind (*11*). Among insects, anticannibalistic strategies, although having evolved independently in species across taxa, are mostly reported for egg protection (*12*–*14*). Beyond egg protection, animals can deploy direct defense against conspecifics. For example, in a particular predatory nematode, cannibalism seems to be avoided by a chemically based self-recognition process (*15*). However, anticannibalistic strategies, in general, remain largely unknown.

[1]Department of Evolutionary Ecology, Max Planck Institute for Chemical Ecology, Jena, Germany. [2]Department of Animal Physiology, Martin-Luther-Universitat Halle-Wittenberg, Halle (Saale), Germany. [3]State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, P. R. China.
*Corresponding author. Email: hansson@ice.mpg.de
†These authors contributed equally to this work.

Among species of locusts, cannibalism is a common characteristic, especially when they are deficient in protein and carbohydrates (*16*, *17*). Like several other locust behavioral traits, cannibalism shows phase polyphenism in response to changes in local population density. At low densities, locusts live as solitary individuals that avoid physical contact. As the local density increases beyond a critical value, behavioral repulsion declines and the locusts begin to be mutually attracted, thereby increasing the likelihood of encountering conspecifics and engaging in cannibalistic interactions (*16*, *18*). To individuals in a group, cannibalizing vulnerable conspecifics offers the dual benefits of surviving longer and traveling farther than a solitary individual without the opportunity to cannibalize (*18*, *19*). Furthermore, cannibalistic interactions have been suggested as one of the driving forces behind collective mass movement. The threat of cannibalistic attacks, especially from behind, was shown to be one of the factors that potentially underlies swarm behavior and movement in migratory bands of nymphs (*16*). Locusts, in general, are extremely important agricultural pests, with swarms devastating crops valued at billions of dollars. At present, the most serious problems are caused by the desert locust, *Schistocerca gregaria*, but the migratory locust, *Locusta migratoria*, also causes substantial damage in Africa and Asia (*20*–*22*). Our recent study on the migratory locust identified hundreds of volatile compounds emanating from all life stages (*23*). Several of these compounds were shown to be behaviorally repulsive, but their biological functions remain largely unknown. Here, we used *L. migratoria* as a model system, revealing the importance of one specific olfactory cue that is involved in protecting individuals against cannibalism at biological levels, from detection to behavior.

Cannibalism is frequent among crowded gregarious nymphs

Before setting out to study the olfactory background of anticannibalism, we established specific traits of cannibalism in *L. migratoria*. We observed that cannibalism is a characteristic of gregarious locusts at all life stages, even if the insects are supplied with sufficient plant-based food sources. Locusts are hemimetabolous insects, with five juvenile instars before reaching adulthood. In our experiments, the highest rates of cannibalism occurred among nymphs at the fourth-to-fifth instar (fig. S1), when the locusts, in general, are more aggressive toward conspecifics (*24*). We therefore chose this stage to investigate density-dependent variations in the rate of cannibalism. We found that cannibalism among fourth-to-fifth instar nymphs displayed density-dependent changes (Fig. 1). At densities of 5 and 25 individuals per cage (cage volume of 9.4 liters), the behavior was reminiscent of solitary locusts, with no individuals attacking or biting each other. The interactions at these two lower densities were observed visually over the course of 1 hour, and no biting attacks occurred (n = 5 cages). When the density was increased beyond 50 individuals per cage, increasing rates of cannibalism, which leveled out at a density of 250 nymphs per cage, were noted. Nymphs of the migratory locust thus display a clear, density-dependent rate of cannibalism. Our results are consistent with earlier observations in the desert locust (*18*). On the basis of these initial results, a clear selection pressure for anticannibalistic agents and behavior could be expected, and we set out in search of potential olfactory cues.

Crowded locusts emit phenylacetonitrile

To establish candidate odors that possibly protect against cannibalism, we conducted a chemical analysis of all volatile compounds emitted from the body of solitary and gregarious migratory locusts (Fig. 2A). Among these, we identified 17 compounds that are only emitted by the gregarious phase, during which cannibalism occurs. Four of the compounds identified—4-vinylanisole, 2-methyl-5-isopropylpyrazine, 2,4,6-trimethylpyridine, and phenylacetonitrile (PAN)—had been reported earlier (*23*, *25*, *26*), and 4-vinylanisole, 2-methyl-5-isopropylpyrazine, and 2,4,6-trimethylpyridine were identified as intraspecifically attractive compounds (*23*, *26*); the remaining compounds were shown to be behaviorally neutral in our tests (fig. S2). However, PAN, a compound also identified in a number of other organisms (*27*–*29*), has been shown to be a repellent against bird predators in *L. migratoria* and also a male antiaphrodisiac in the desert locust (*30*, *31*). Furthermore, it has been shown to be an honest signal for toxicity because its breakdown can form the toxic product hydrogen cyanide (*30*). Thus, the distinctive function of PAN as a repellent and

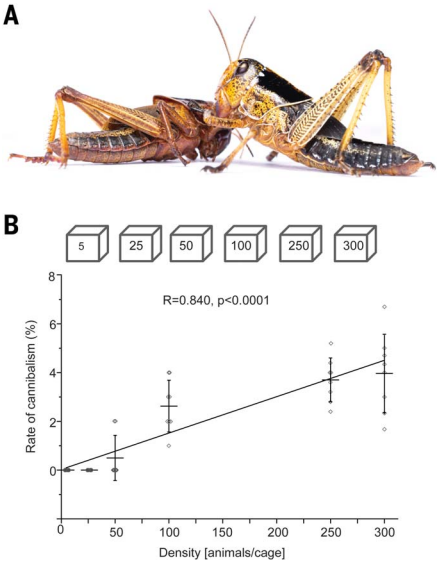


Fig. 1. Cannibalism in *L. migratoria*. (**A**) A fifth-instar migratory locust (right) consuming the body of a conspecific. (**B**) The rate of cannibalism displayed by gregarious fifth-instar nymphs at different population densities. The rate of cannibalism was calculated as the percentage of animals that were completely eaten at each density. Minor attacks also occurred but were impossible to distinguish from other injuries, for example, legs lost from getting stuck in cage ventilation holes. We therefore chose the clearest and most conservative measure, that is, complete consumption. Data are plotted as the mean and standard deviation, with raw data points in the background, and n = 8 replicates per density [displayed Spearman's rank correlation coefficient (R) and p values are based on a Spearman's rank correlation analysis]. The boxes with numbers represent cages and the number of individuals per cage. [Photo credit: B. Fabian]

an honest signal of toxicity in the migratory locust raised our interest in this compound, and we asked if it might indeed be an intraspecific behavioral antagonist, inhibiting cannibalism.

Consistent with earlier studies (*25*), we found that PAN is the dominant volatile component emitted from the body of gregarious nymphs. Moreover, we found that PAN emission by gregarious nymphs covaries with density and thereby parallels the earlier observations of cannibalism rates (Fig. 1B). Nymphs thus begin to release PAN when the population density exceeds 25 individuals per cage (Fig. 2B), whereafter release rates increase gradually with density. Because both PAN release and cannibalism displayed a density-dependent pattern, we hypothesized that PAN could play a role in nymph interactions at high densities.

Locusts exhibit aversion to PAN

Next, we subjected gregarious fourth-to-fifth instar nymphs to dual-choice olfactometer tests to determine the behavioral valence of PAN (Fig. 2C). First, we applied paraffin oil to both sides of the olfactometer as a control and found that nymphs did not show any behavioral preference for a specific side when an olfactory stimulus was absent (Fig. 2D). Applying PAN, we found that the gregarious nymphs displayed significant repulsion at concentrations ranging from 10^{-4} to 10^{-2} dilution (1:10,000 to 1:100 volume:volume dilution). The nymphs significantly preferred to move and remain in the zone suffused with vapor of paraffin oil than the zone smelling of PAN. At a lower concentration of 10^{-5}, the negative effect of PAN was absent (Fig. 2D). We proceeded to evaluate the behavioral performance across developmental stages, phases, and sexes and found that PAN also evoked aversion among adults of both sexes and both phases (Fig. 2E). We thus conclude that PAN is an aversive compound to migratory locusts regardless of age, sex, or phase.

PAN is mainly detected by one olfactory receptor that is necessary for aversive behavior

We now turned to the olfactory system of *L. migratoria* in search of the olfactory receptors (ORs) involved in the detection of PAN. Initially, we cloned 49 ORs and functionally expressed them in the antennal trichoid 1 empty neuron system of *Drosophila melanogaster*, in which the endogenous receptor OR67d is lacking (*32*). We chose this system because it involves the SNMP1 coreceptor that was earlier shown to be expressed in locust olfactory sensory neurons (*33*). The alternative Δ-halo system lacks SNMP1 (*34*). To establish the molecular specificity of the locust ORs, we performed single-sensillum recording (SSR) measurements in the transgene fly lines while stimulating with PAN. These recordings revealed that the fly line bearing the *L. migratoriaOR70a* (*LmOR70a*) gene displayed the highest response to PAN among all the fly lines tested (Fig. 3A). The PAN response of LmOR70a was almost fourfold and fivefold higher than that of the second- and third-best ORs LmOR20 and LmOR75, respectively (Fig. 3B). We then asked whether LmOR70a was exclusively tuned to PAN and screened 204 locust volatiles reported previously from our laboratory (*23*) (tested at 10^{-1} dilution) in continued recording experiments (table S1). We identified two additional compounds, which are structurally similar to PAN (benzaldehyde and cinnamaldehyde), that activated neurons expressing LmOR70a, though at higher concentrations (Fig. 3, C and D). However, the vapor pressure of the second-best ligand, benzaldehyde, is about 14 times higher than that of PAN, meaning that the dose-response curve should be corrected accordingly (Fig. 3D). The fact that two additional compounds activated LmOR70a, albeit to a much lower degree, prompted us to return to behavioral experiments. We found that none of these odors elicited repulsion or preference (Fig. 3E), suggesting that these two compounds might also be detected by other ORs that modulate the input of LmOR70a-expressing neurons to the locust brain and thereby change the behavioral response. Specific receptors that detect both benzaldehyde and cinnamaldehyde were indeed identified in a parallel study (*23*). Another possible explanation for the specific response to PAN could be the presence of other receptors that are specific to this compound among those receptors that were not tested. However, the fact that all electroantennogram (EAG) responses were abolished in locusts when the LmOR70a receptor had been made nonfunctional (see next paragraph and Fig. 3H) is contrary to this explanation. After testing 49 ORs and more than 200 relevant odors, we thus established LmOR70a as a highly sensitive and specific detector of PAN.

Olfactory sensory neurons that bear different types of ORs are located in olfactory hairs, or sensilla, on the locust antenna. To determine in which type of sensilla the neurons detecting PAN were located, we carried out SSR experiments with two types of locust sensilla (basiconic and trichoid) while stimulating with PAN. These are the sensillum types that house olfactory sensory neurons that express the olfactory coreceptor Orco, which is known to be present in all OR-expressing neurons (*35*, *36*). Basiconic sensilla house 20 to 50 sensory neurons, whereas trichoid sensilla house two or three (*35*). The high number of neurons present in the basiconic sensilla prevented us from discriminating responses from individual neurons on the basis of action potential (spike) amplitudes, and the response was calculated as the total number of spikes elicited from the complete population of neurons housed in the sensillum. We found a strong response from neurons present in basiconic sensilla, whereas no response was observed from neurons in trichoid sensilla (Fig. 3F). We then asked whether it was LmOR70a-expressing neurons that were housed in basiconic sensilla and contributed to the response to PAN. First, we performed RNA in situ hybridization to determine the distribution of neurons expressing LmOR70a in the locust nymph antenna and found that LmOR70a-expressing neurons were indeed housed exclusively in basiconic sensilla (Fig. 3G). Next, we generated a loss-of-function allele of the *LmOR70a* gene using CRISPR-Cas9 genome editing. A single guide RNA (sgRNA), targeting the first exon of the *LmOR70a* gene, introduced an 8–base pair (bp) deletion, resulting in a truncated LmOR70a protein (fig. S3A). Through EAG recordings, we could show that the response to PAN was abolished in the *LmOR70a*$^{-/-}$ line, whereas the response to another odor remained unaffected. (Fig. 3H and fig. S3, B and C). When we then recorded the

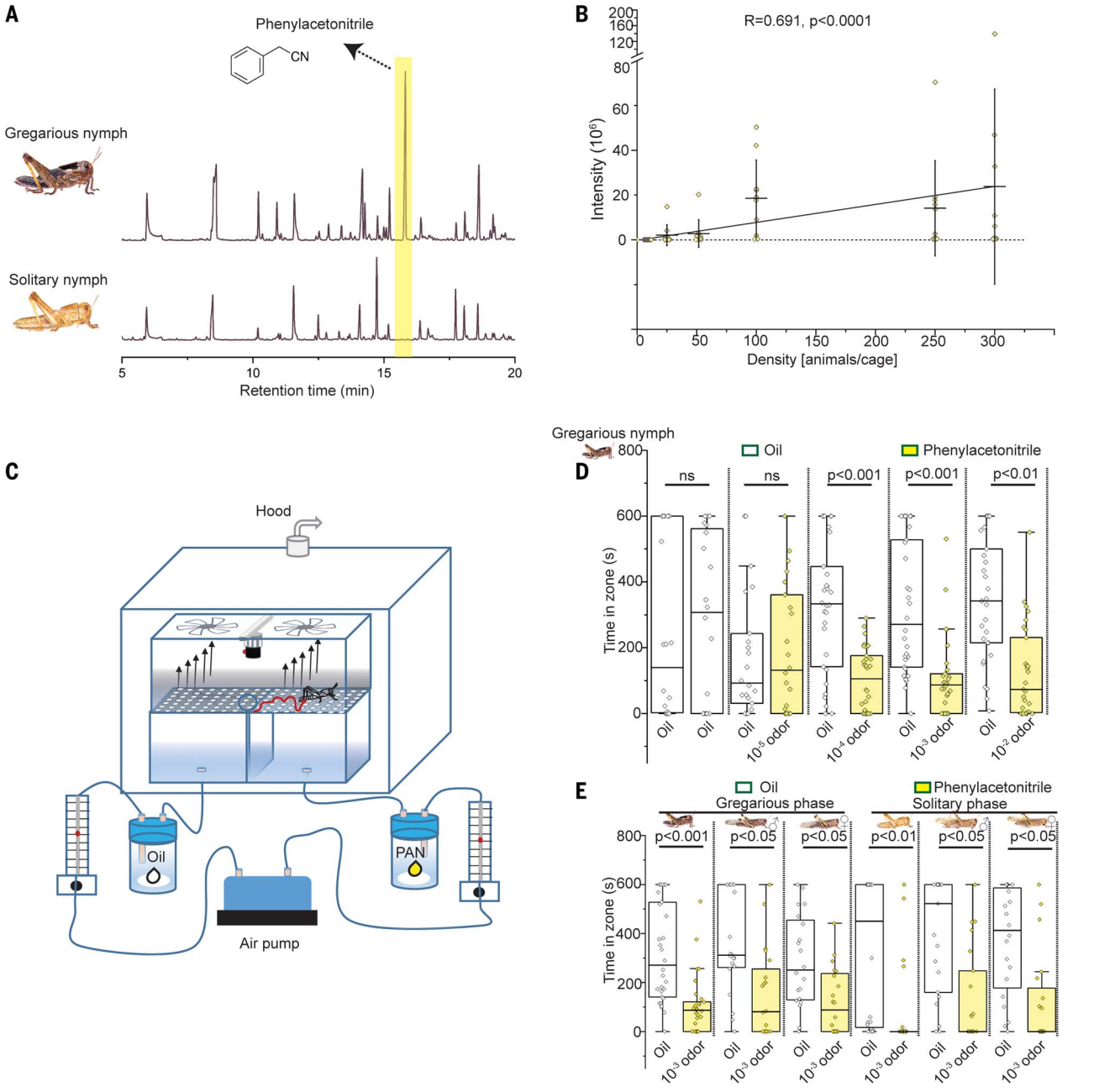


Fig. 2. Gregarious nymphs release PAN to repel others. (**A**) Gas chromatograms of locust body volatiles. PAN is indicated by the yellow bar. (**B**) PAN emitted by gregarious nymphs at different population densities. Data are plotted as the mean and standard deviation, with raw data points in the background, and n = 10 replicates for each population density (displayed R and p values are based on a Spearman's rank correlation analysis). (**C**) Schematic drawing of the dual-choice olfactometer experiment. Details of the experiments are described in the materials and methods. (**D**) Behavioral responses of gregarious nymphs to different dosages of synthetic PAN, with n = 20 to 30 replicates for each concentration. (**E**) Behavioral responses of locusts across phase, sex, and developmental stage to 10^{-3} dilution PAN, with n = 18 to 30 replicates for each type. Data in (D) and (E) are plotted as box and whisker plots (boxes depict median and upper and lower quartiles; whiskers depict quartiles ± 1.5 times the interquartile range). In (D) and (E), p values were determined by a paired-sample Wilcoxon signed-rank test; ns is not significant. [Photo credits: B. Fabian]

SSR response to PAN of neurons housed in basiconic sensilla, we found that in the *LmOR70a*$^{-/-}$ line, the response was close to abolished compared with that of wild-type (WT) nymphs (Fig. 3I). Moreover, we tested gregarious nymphs of the *LmOR70a*$^{-/-}$ line in the dual-choice olfactometer for their behavioral response to PAN and found that the homozygous mutant nymphs had completely lost their aversion to PAN as compared with WT insects (Fig. 3J). We conclude that olfactory sensory neurons that express LmOR70a are present in basiconic sensilla and are responsible for the negative behavioral response to PAN in migratory locusts. This pathway is thus necessary for the aversion to occur.

PAN suppresses cannibalism

When observing locust behavior, we found that cannibalism could be categorized into two forms: one that entails the consumption

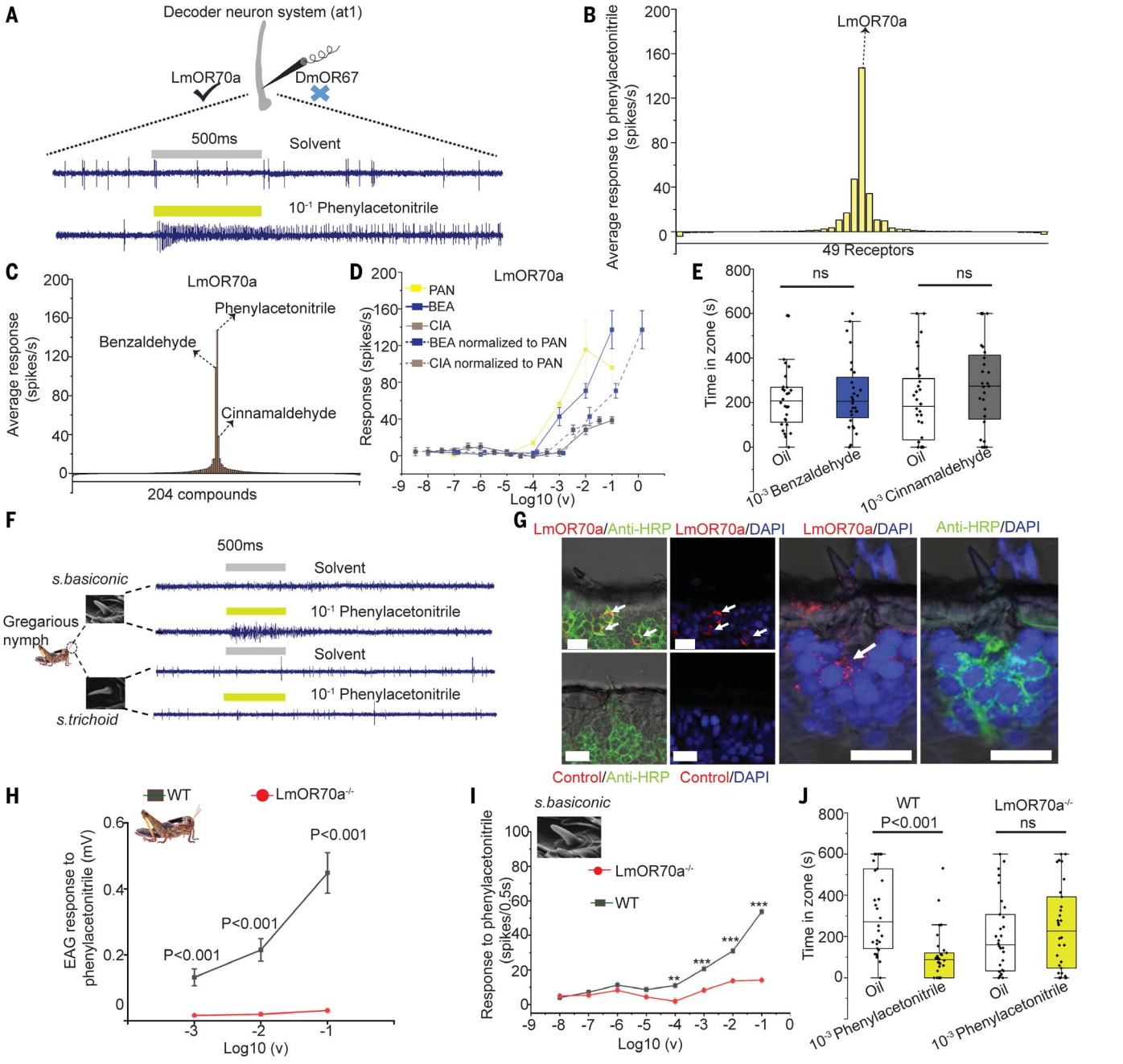


Fig. 3. LmOR70a detects PAN and mediates aversiveness in locusts. (**A**) Representative SSR traces of LmOR70a-expressing *Drosophila* neurons to 10^{-1} dilution PAN. DmOR67 is absent in the at1 neurons of these flies. (**B**) Response to PAN from *Drosophila* decoder neurons expressing 49 different LmORs, with $n = 5$ or 6 recordings for each receptor (*41*, *42*). (**C**) Response of *Drosophila* decoder neurons expressing LmOR70a to 204 odorants (each odor was applied at 10^{-1} dilution), with $n = 6$ recordings for each odorant. (**D**) Dose-response curves of *Drosophila* decoder neurons expressing LmOR70a to the three most active odor molecules, with $n = 6$ recordings for each compound. Data points are mean ± SEM. Dashed curves indicate responses to the two less active compounds after correction according to vapor pressure relative to PAN [the vapor pressure at 25°C of PAN is 0.089 mmHg, of benzaldehyde (BEA) is 1.27 mmHg, and of cinnamaldehyde (CIA) is 0.0289 mmHg]. (**E**) Behavioral responses of gregarious nymphs to benzaldehyde and cinnamaldehyde, with $n = 27$ or 28 biological replicates for each odor. (**F**) Representative spike traces from SSR recordings from locust basiconic and trichoid sensilla in response to 10^{-1} dilution PAN. Also shown are scanning electron micrographs of a basiconic and a trichoid sensillum from the locust antenna. (**G**) In situ fluorescence hybridization revealing that LmOR70a-expressing olfactory sensory neurons are housed in basiconic sensilla. White arrows in all images indicate LmOR70a-positive cells, and the white scale bars represent 20 μm. DAPI, 4′,6-diamidino-2-phenylindole; HRP, horseradish peroxidase. (**H**) EAG dose-response curves to PAN at different concentrations recorded in WT and *LmOr70a*$^{-/-}$ locusts, with $n = 10$ to 12 recordings for each locust type. Data points are mean ± SEM. v, dilution. (**I**) SSR dose-response curves in WT and *LmOr70a*$^{-/-}$ locusts, with $n = 35$ (WT) and 27 (*LmOr70a*$^{-/-}$) sensilla. Data points are mean ± SEM. The inset shows a scanning electron micrograph of a basiconic sensillum. (**J**) Behavioral responses to PAN in WT and *LmOr70a*$^{-/-}$ locusts, with $n = 29$ or 30 replicates for each assay. Data in (E) and (J) are plotted as box and whisker plots (boxes depict median and upper and lower quartiles; whiskers depict quartiles ± 1.5 times the interquartile range). In (E) and (J), p values were determined by a paired-sample Wilcoxon signed-rank test, whereas p values in (H) and (I) were determined by a two-tailed unpaired Student's t test. ***$p < 0.001$; **$p < 0.01$; ns is not significant. [Photo credits: B. Fabian]

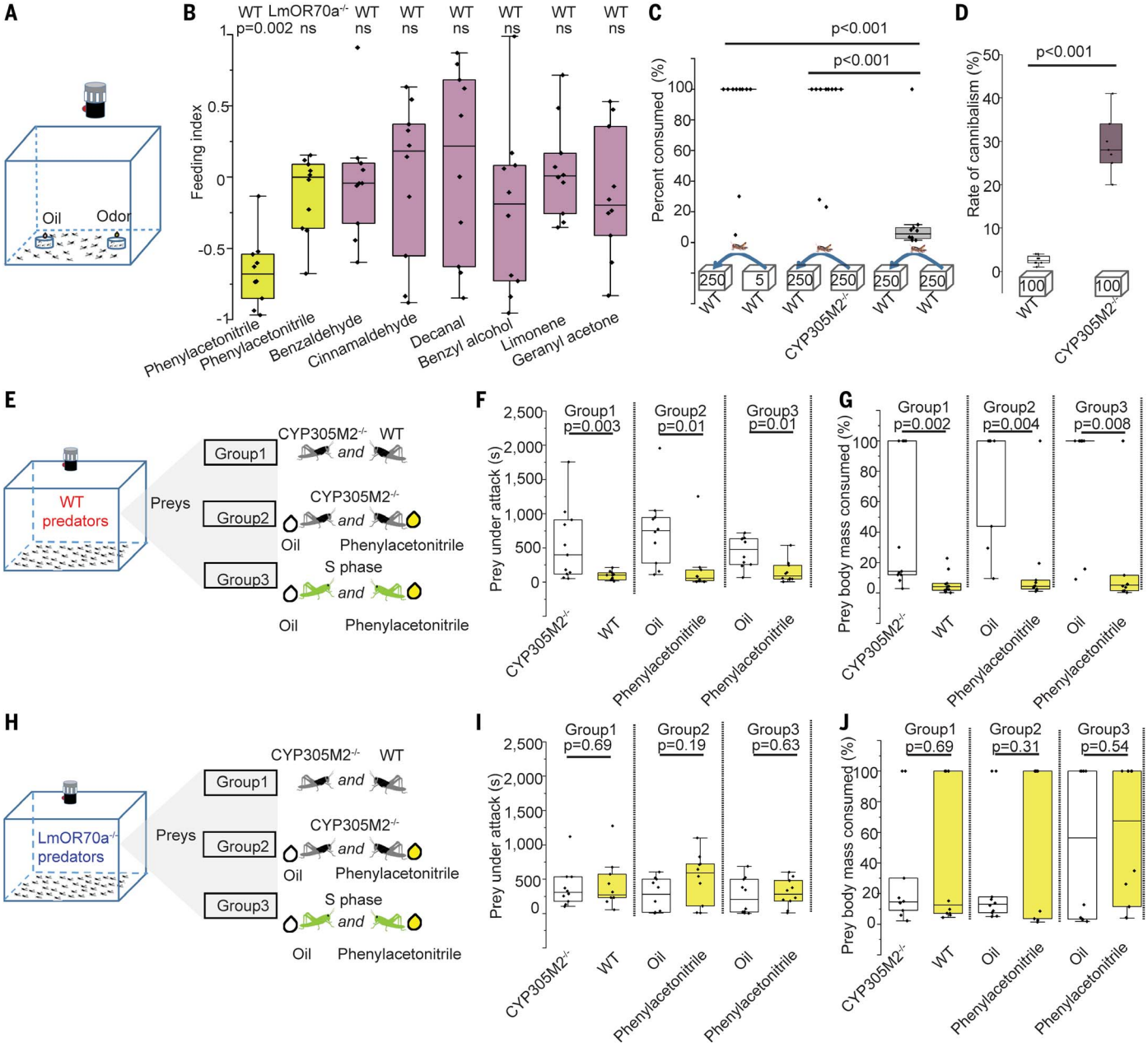


Fig. 4. PAN deters locust cannibalism. (**A**) Schematic of the dual-choice predation experiments using nymph corpses. (**B**) Results of the dual-choice experiments in which starved nymphs chose between corpses with odor and corpses without odor, with n = 10 biological replicates for each assay. The feeding index was calculated as (O − C)/(O + C), where O is the consumed mass of corpses perfumed with odor and C is the consumed mass of corpses perfumed with paraffin oil as a control. (**C**) The rate of cannibalism on living individual nymphs transferred from low-density to high-density cages and from high-density to high-density cages and on *CYP305M2*$^{-/-}$ mutants transferred from high-density to high-density cages. The rate of cannibalism as a percentage was calculated as (B − A)/B, where B is the mass of the prey before the assay and A is the mass of the prey after the assay, with n = 10 biological replicates for each assay. The boxes with numbers represent cages and the number of individuals per cage. (**D**) The rate of cannibalism displayed by WT and *CYP305M2*$^{-/-}$ mutants at a density of 100 individuals per cage. The rate of cannibalism was calculated as the percentage of animals that were completely eaten, with n = 8 (WT) and 7 (*CYP305M2*$^{-/-}$). (**E**) Schematic of dual-choice cannibalism experiments on three pairs of nymphs that were introduced to 50 WT starved nymphs: group 1, a *CYP305M2*$^{-/-}$ individual (no PAN) and a WT individual (producing PAN); group 2, a *CYP305M2*$^{-/-}$ individual (no PAN) and a *CYP305M2*$^{-/-}$ individual perfumed with PAN; and group 3, a solitary-phase individual (no PAN) and a solitary-phase individual perfumed with PAN. (**F** and **G**) Quantification of the time that prey nymphs were under attack (F) and percentage of the prey body that was consumed (G) when prey nymphs were introduced to WT nymphs. (**H**) Schematic of dual-choice cannibalism experiments on three pairs of living gregarious *CYP305M2*$^{-/-}$ or solitary nymphs treated with paraffin oil (control) or with PAN and then introduced to 50 *Lm0r70a*$^{-/-}$ starved nymphs. (**I** and **J**) Same as for (F) and (G) but for the experiment shown in (H) with *Lm0r70a*$^{-/-}$ nymphs as potential predators. In (F), (G), (I), and (J), n = 10 or 11 replicates for each assay. In (B), (F), (G), (I), and (J), p values were determined by a paired-sample Wilcoxon signed-rank test; in (C) and (D), p values were determined by a two-tailed unpaired Student's t test. All data were plotted as box and whisker plots (boxes depict median and upper and lower quartiles; whiskers depict quartiles ± 1.5 times the interquartile range). ns, not significant.

of an already dead body and one that involves more active predation on juvenile conspecifics while still alive. Given this, we first wanted to know whether the presence of PAN influences cannibalistic feeding on a dead nymph body. We chose the corpses of fifth-instar gregarious nymphs as potential food sources and noted that these corpses completely stopped releasing PAN 24 hours after death (fig. S4A). We then compared the consumption of natural corpses, only scented with paraffin oil, with those perfumed with PAN in a group of 20 starved, healthy, fifth-instar gregarious nymphs (Fig. 4A). The results showed that the tested insects refused to approach and feed on PAN-scented corpses and instead feasted on the oil-treated ones (Fig. 4B). Weight-wise, the amount of PAN-perfumed corpses that were consumed by feeding nymphs was five times lower than the amount of oil-treated corpses that were consumed. Next, we repeated these experiments but replaced the WT feeding nymphs with *LmOR70a*$^{-/-}$ insects. The *LmOR70a*$^{-/-}$ nymphs did not show any biting or feeding preference between the corpses loaded with oil or PAN (Fig. 4B). We thus conclude that PAN deters locust nymphs from feeding on dead bodies of conspecifics and that LmOR70a-expressing sensory neurons are responsible for this distinction.

To test the distinctive anticannibalistic role of PAN, we tested another four locust volatiles that were previously reported as aversive to gregarious nymphs (*23*) for their effects on preventing feeding on conspecific corpses. None of these compounds reduced feeding on dead nymph bodies, suggesting that PAN might be the only compound that specifically suppresses locust cannibalism (Fig. 4B). The two compounds that activated LmOR70a-expressing neurons to a lower extent than PAN did not reduce feeding on dead nymph bodies either, pointing again at the possibility that the specific ORs involved in the detection of these odors (*23*) modulate their importance.

Next, we wanted to know whether the emission of PAN by living gregarious locust nymphs protects them from conspecific predation. The ideal test would be to have a living gregarious locust devoid of PAN but otherwise intact in every way. Taking advantage of our earlier results, which showed that nymphs from low-density cages produced no PAN, whereas those bred in high-density cages produced high levels, we performed a first test. Nymphs transferred from low-density to high-density cages were directly and completely cannibalized, whereas those transferred from high-density to high-density cages were left untouched (Fig. 4C).

To further investigate the effect of PAN production, we took advantage of the findings from earlier investigations that showed that one member of the cytochrome P450 gene family called CYP305M2 is crucial for PAN biosynthesis in *L. migratoria* (*30*). Building on these insights, we generated a loss-of-function allele of *CYP305M2* using CRISPR-Cas9 genome editing. A sgRNA targeting the second exon of the *CYP305M2* gene introduced a 26-bp deletion, resulting in a truncated CYP305M2 protein (fig. S4B). As shown earlier, PAN biosynthesis covaries with population density. Therefore, to eliminate density effects, we raised *CYP305M2*$^{-/-}$ locusts and WT locusts under identical conditions. Chemical analysis of the emissions of *CYP305M2*$^{-/-}$ and WT locusts showed that the mutant locusts were unable to produce PAN (fig. S4A), whereas the release levels of other major compounds, body color, and mobility characteristics remained unchanged as compared with WT insects. An interesting characteristic of these mutant locusts was that they exhibited such a high degree of cannibalism that production in high numbers became extremely hard. However, because we had access to nymphs that did not produce PAN, we repeated the experiment shown in Fig. 1B but with *CYP305M2*$^{-/-}$ locusts at the lowest density at which significant degrees of cannibalism were shown among the WT locusts (100 individuals per cage). The rate of cannibalism among the *CYP305M2*$^{-/-}$ locusts was found to be almost 10 times higher than that among WT animals (Fig. 4D). We then performed a series of dual-choice predation experiments by introducing a pair of nymphs, one either emitting or perfumed with PAN and the other devoid of the smell of PAN, into an observation chamber containing 50 starved gregarious nymphs (Fig. 4E). First, we examined the predation choice between *CYP305M2*$^{-/-}$ and WT nymphs and found that *CYP305M2*$^{-/-}$ nymphs were frequently attacked and preferentially consumed over WT nymphs. Both the time being under attack and the amount of the body that was consumed were five- and sevenfold higher in *CYP305M2*$^{-/-}$ nymphs than in WT insects, respectively (Fig. 4, F and G). Second, we used *CYP305M2*$^{-/-}$ locusts scented with PAN or paraffin oil to test predation choice. *CYP305M2*$^{-/-}$ locusts were each perfumed with 300 ng PAN, an amount similar to the biosynthesis level in WT nymphs (fig. S4C) (*30*). In predation experiments as described above, oil-treated *CYP305M2*$^{-/-}$ locusts were, in most cases, completely consumed by starved conspecifics, whereas PAN-treated *CYP305M2*$^{-/-}$ nymphs were seldom injured or consumed (Fig. 4, F and G). In a final experiment, we replaced the *CYP305M2*$^{-/-}$ nymphs with WT solitary ones because solitary nymphs are also unable to produce PAN. We found that starved nymphs also refused to attack and feed on PAN-treated individuals in this setup, whereas they very frequently fed on the nonscented, solitary locusts (Fig. 4, F and G). To determine whether LmOR70a was also required for these PAN-dependent interactions, we used *LmOR70a*$^{-/-}$ locusts as starved predators and repeated the choice experiments, including both *CYP305M2*$^{-/-}$ and solitary locusts (Fig. 4H). We found that *LmOR70a*$^{-/-}$ locusts displayed no predation preference for prey individuals with or without PAN (Fig. 4, I and J). These results together confirm that PAN produced by gregarious nymphs deters intraspecific predation and that the detection of PAN by LmOR70a-expressing neurons governs the suppression of cannibalism among gregarious locust nymphs.

Discussion

Locusts are infamous for their habit of forming giant swarms, causing severe damage to many crops (*17*, *37*). Such enormous densities of individuals raise questions regarding the interactions that occur within the swarm. In our study, we reveal an interesting impact of the compound PAN. The production of PAN in different locust species was reported decades ago, and several hypotheses have been postulated regarding its function (*30*, *31*, *38*, *39*). In the desert locust, it has been proposed to be both an aggregation pheromone and a male courtship inhibitor (*31*, *38*, *39*). In another exhaustive study, Wei *et al*. described that in *L. migratoria*, PAN acts as an olfactory aposematic signal in defense against general predators, for example, birds (*30*), and that it is an honest signal of toxicity. However, insects living in dense groups have been shown to be less vulnerable to predation than solitary individuals (*37*, *40*) but still face a substantial risk of being eaten by conspecifics. Cannibalism has been hypothesized as being one of the factors that underlies swarm dynamics (*16*). Irrespective of the specific swarm situation, cannibalism occurs frequently among crowded locust nymphs, and avoiding it should be beneficial. Here, we report that PAN has a clear anticannibalistic function among locust nymphs. Taking all results under consideration, we suggest that this might be the original function of the compound, because a more widespread and less density-dependent occurrence might be expected in a purely antipredatory compound. Indeed, only crowded gregarious nymphs of the migratory locust release high amounts of PAN, whereas solitary individuals do not. This dichotomy further emphasizes that the production of PAN comes with specific costs and has a specific function that seems to be restricted to crowded conditions. Under these conditions, the level of PAN production varies between individual nymphs, which might explain why some degree of cannibalism remains. However, when PAN production was abolished in mutant locusts, cannibalism increased manifold, revealing that cannibalism is strongly suppressed in WT animals.

REFERENCES AND NOTES

1. D. H. Wise, *Annu. Rev. Entomol.* **51**, 441–465 (2006).
2. M. R. Crossland, M. N. Hearnden, L. Pizzatto, R. A. Alford, R. Shine, *Anim. Behav.* **82**, 775–782 (2011).
3. K. W. Welke, J. M. Schneider, *Anim. Behav.* **83**, 201–207 (2012).
4. D. Mayntz, S. Toft, *J. Anim. Ecol.* **75**, 288–297 (2006).
5. G. A. Polis, *Annu. Rev. Ecol. Syst.* **12**, 225–251 (1981).
6. M. L. Richardson, R. F. Mitchell, P. F. Reagel, L. M. Hanks, *Annu. Rev. Entomol.* **55**, 39–53 (2010).
7. A. Lomnicki, *Population Ecology of Individuals* (Princeton Univ. Press, 1988), chap. 1.
8. V. H. W. Rudolf, *Ecology* **89**, 3116–3127 (2008).
9. O. Kishida, G. C. Trussell, K. Nishimura, T. Ohgushi, *Ecology* **90**, 3150–3158 (2009).
10. F. K. Ho, P. S. Dawson, *Ecology* **47**, 318–322 (1966).
11. M. A. Elgar, B. J. Crespi, Eds., *Cannibalism. Ecology and Evolution Among Diverse Taxa* (Oxford Univ. Press, 1992).
12. S. Narasimha *et al*., *PLOS Biol.* **17**, e2006012 (2019).
13. L. K. Thomas, A. Manica, *Anim. Behav.* **66**, 205–210 (2003).
14. Z. Růžička, *Eur. J. Entomol.* **93**, 161–166 (1997).
15. J. W. Lightfoot *et al*., *Science* **364**, 86–89 (2019).
16. S. Bazazi *et al*., *Curr. Biol.* **18**, 735–739 (2008).
17. S. Bazazi *et al*., *Proc. Biol. Sci.* **278**, 356–363 (2011).
18. V. Guttal, P. Romanczuk, S. J. Simpson, G. A. Sword, I. D. Couzin, *Ecol. Lett.* **15**, 1158–1166 (2012).
19. S. Bazazi *et al*., *PLOS ONE* **5**, e15118 (2010).
20. G. A. Matthews, *Agronomy* **11**, 1052 (2021).
21. A. T. Showler, S. Shah, S. Khan, S. Ullah, F. Degola, *J. Integr. Pest Manag.* **13**, 1 (2022).
22. L. Zhang, M. Lecoq, A. Latchininsky, D. Hunter, *Annu. Rev. Entomol.* **64**, 15–34 (2019).
23. H. Chang *et al*., bioRxiv 2022.2006.2021.496967 [Preprint] (2022); https://doi.org/10.1101/2022.06.21.496967.
24. M. Miladi, K. Abdellaoui, A. B. Hamouda, I. Boughattas, *Tunisian J. Plant Protect.* **13**, 201–216 (2018).
25. J. Wei *et al*., *Insect Sci.* **24**, 60–72 (2017).
26. X. Guo *et al*., *Nature* **584**, 584–588 (2020).
27. R. R. Junker *et al*., *Plant Biol.* **13**, 918–924 (2011).
28. S. Irmisch *et al*., *Plant J.* **80**, 1095–1107 (2014).
29. T. Bellas, B. Hölldobler, *J. Chem. Ecol.* **11**, 525–538 (1985).
30. J. Wei *et al*., *Sci. Adv.* **5**, eaav5495 (2019).
31. K. Seidelmann, H.-J. Ferenz, *J. Insect Physiol.* **48**, 991–996 (2002).
32. A. Kurtovic, A. Widmer, B. J. Dickson, *Nature* **446**, 542–546 (2007).
33. R.-S. Lemke *et al*., *Cell Tissue Res.* **379**, 275–289 (2020).
34. A. A. Dobritsa, W. van der Goes van Naters, C. G. Warr, R. A. Steinbrecht, J. R. Carlson, *Neuron* **37**, 827–841 (2003).
35. Y. Yang, J. Krieger, L. Zhang, H. Breer, *Int. J. Biol. Sci.* **8**, 159–170 (2012).
36. Y. Li *et al*., *Insect Biochem. Mol. Biol.* **79**, 27–35 (2016).
37. S. J. Simpson, G. A. Sword, P. D. Lorch, I. D. Couzin, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4152–4156 (2006).
38. M. P. Pener, S. J. Simpson, *Adv. Insect Physiol.* **36**, 1–272 (2009).
39. A. L. Deng, B. Torto, A. Hassanali, E. Ali, *J. Insect Physiol.* **42**, 771–776 (1996).
40. G. A. Sword, P. D. Lorch, D. T. Gwynne, *Nature* **433**, 703–703 (2005).
41. A. F. Carey, G. Wang, C. Y. Su, L. J. Zwiebel, J. R. Carlson, *Nature* **464**, 66–71 (2010).
42. E. A. Hallem, J. R. Carlson, *Cell* **125**, 143–160 (2006).

ACKNOWLEDGMENTS

We thank S. Brase for locust rearing, K. Weniger for assistance with chemical analysis, R. Stieber-Rödiger for assistance with cloning, and S. Trautheim for assistance with fly crossing. **Funding:** This research was supported through funding by the Max Planck Society, the National Natural Science Foundation of China (NSFC) (no. 31920103004), and the Basic Science Center Project of the NSFC (no. 32088102). **Author contributions:** Conceptualization: H.C., B.S.H., M.K., X.G., L.K.; Methodology: All authors; Investigation: H.C., S.C.; Visualization: H.C.; Funding acquisition: B.S.H., L.K., M.K.; Project administration: B.S.H., J.K., M.K., L.K., X.G.; Supervision: B.S.H., J.K., M.K.; Writing – original draft: H.C., B.S.H.; Writing – review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Correspondence and material requests should be directed to B.S.H. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade6155
Materials and Methods
Figs. S1 to S4
Tables S1 and S2
References (*43*–*45*)
MDAR Reproducibility Checklist
Data S1 to S7

Submitted 27 August 2022; accepted 6 March 2023
10.1126/science.ade6155

NEUROSCIENCE

A shift in the mechanisms controlling hippocampal engram formation during brain maturation

Adam I. Ramsaran[1,2], Ying Wang[1,3], Ali Golbabaei[1,4], Stepan Aleshin[5], Mitchell L. de Snoo[1,4], Bi-ru Amy Yeung[1,3], Asim J. Rashid[1], Ankit Awasthi[1], Jocelyn Lau[1,3], Lina M. Tran[1,3,6], Sangyoon Y. Ko[1,3], Andrin Abegg[1,7]†, Lana Chunan Duan[1,4], Cory McKenzie[1,2], Julia Gallucci[1]‡, Moriam Ahmed[1], Rahul Kaushik[5,8], Alexander Dityatev[5,8,9], Sheena A. Josselyn[1,2,3,4,10], Paul W. Frankland[1,2,3,4,11]*

The ability to form precise, episodic memories develops with age, with young children only able to form gist-like memories that lack precision. The cellular and molecular events in the developing hippocampus that underlie the emergence of precise, episodic-like memory are unclear. In mice, the absence of a competitive neuronal engram allocation process in the immature hippocampus precluded the formation of sparse engrams and precise memories until the fourth postnatal week, when inhibitory circuits in the hippocampus mature. This age-dependent shift in precision of episodic-like memories involved the functional maturation of parvalbumin-expressing interneurons in subfield CA1 through assembly of extracellular perineuronal nets, which is necessary and sufficient for the onset of competitive neuronal allocation, sparse engram formation, and memory precision.

The episodic memory system is absent or immature at birth and develops during childhood. Accordingly, early event memories are imprecise or gist-like until ~5 to 8 years of age, when mnemonic precision increases (*1*–*5*). Hippocampal maturation is thought to underlie the emergence of precise episodic memories in humans and episodic-like memories in nonhuman animals (*5*–*8*), but the specific processes regulating memory precision during hippocampal development are unknown.

Age-dependent increases in engram sparsity track changes in the precision of hippocampus-dependent memories

The binding of events to their surrounding spatial context is a core feature of episodic and episodic-like memory that may be studied in animals using spatial or contextual learning tasks (*9*). To assess when this ability emerges during mouse development, we trained mice of different ages in contextual fear conditioning and tested their memory 24 hours later in either the same (context A) or a distinct (context B) testing apparatus (Fig. 1, A and B). Younger mice [postnatal day 16 (P16) to P20] expressed imprecise contextual fear memories, freezing at equivalent levels in the training context A and the novel context B. By contrast, older mice (≥P24) expressed context-specific memories, freezing more in context A than in context B (Fig. 1C). This shift in memory precision parallels similar shifts in rats (*10*), was independent of the animals' sex or weaning status, and did not depend on prior experience with contexts, potential age-dependent differences in learning rate, or ability to perceptually discriminate the contexts (fig. S1, A to S). Memory imprecision in juvenile mice scaled with the similarity between the training and testing contexts (fig. S1, T and U). Shifts in memory precision also occurred between P20 and P24 in a related aversive contextual learning task (inhibitory avoidance; fig. S1, V to X) and an appetitively motivated spatial foraging task (Fig. 1, D to F, and fig. S2, A to H).

Although contextual fear memories depend on the hippocampus in adult rodents (*11*), it is possible that this type of learning is supported by extrahippocampal structures in juvenile mice (*12*). This is consistent with proposals that the hippocampus does not support early event memories in children but instead "comes online" during childhood to allow the emergence of episodic memory (*12*, *13*). We tested the hippocampal dependency of contextual fear memories in juvenile and adult mice by microinjecting adeno-associated viruses (AAVs) encoding inhibitory opsins into the dorsal CA1 of the hippocampus (fig. S3, A and B). CA1 may support both precise and imprecise memories across development (*14*, *15*). Optogenetic silencing of CA1 pyramidal neurons impaired precise memory recall in P60 mice (*11*) (fig. C and D). In P20 mice, silencing CA1 neurons reduced freezing in both the A and B contexts, indicating that CA1 supports imprecise contextual memories at this developmental stage.

Although these results indicate that the immature hippocampus supports early memories, age-dependent differences in memory precision suggest that these memories may be encoded differently in the hippocampi of juvenile versus adult mice. In adults, context memories are encoded by sparse ensembles of neurons, known as engrams, in the hippocampus (*16*, *17*). Given the imprecision of juvenile memories, we wondered whether CA1 engrams supporting contextual memories in juvenile mice lack sparsity. To identify putative engram neurons, we examined the expression of the activity-regulated immediate-early gene (IEG) c-Fos in the dorsal CA1 of P20, P24, and P60 mice after contextual fear conditioning (Fig. 1G). Training induced Fos expression in ~20% of CA1 neurons in P24 and P60 mice (*18*–*20*). By contrast, training induced c-Fos expression in ~40% of CA1 neurons in P20 mice, suggesting that engrams are more densely encoded in juvenile mice (Fig. 1H). Moreover, a similarly high proportion of c-Fos$^+$ CA1 neurons was observed in juvenile mice after testing (fig. S4, A to C) and in P20 "home cage" mice (fig. S4, D to H), consistent with observations that IEG expression is transiently elevated in the hippocampus of experimentally naïve rodents during the third postnatal week as the activity-dependent assembly of hippocampal neural circuitry nears completion (*21*–*23*).

CA1 engram sparsity controls memory precision

To test whether there is a causal relationship between engram size and memory precision, we investigated whether artificially shrinking engrams in juvenile mice would promote adult-like memory precision. We expressed the inhibitory designer receptor exclusively activated by designer drugs (DREADD) hM4Di in a subset of CA1 neurons and injected the DREADD ligand C21 before training to inhibit infected neurons and prevent their inclusion in the engram (Fig. 2A and fig. S5A). C21 treatment reduced engram size (training-induced c-Fos expression; Fig. 2, B and C), and these juvenile mice precociously exhibited adult-like memory precision (freezing more in context A than in context B) (Fig. 2, D and E, and fig. S5, B and C).

Conversely, we investigated whether artificially expanding the engram in adult mice would induce juvenile-like memory imprecision. We expressed the excitatory DREADD construct hM3Dq in pyramidal layer neurons (Fig. 2F and fig. S5D) and injected C21 before training to increase the activity of hM3D-infected neurons. C21 treatment increased engram size (training-induced c-Fos expression;

[1]Program in Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, Ontario, Canada. [2]Department of Psychology, University of Toronto, Toronto, Ontario, Canada. [3]Department of Physiology, University of Toronto, Toronto, Ontario, Canada. [4]Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada. [5]Molecular Neuroplasticity, German Center for Neurodegenerative Diseases, Magdeburg, Germany. [6]Vector Institute, Toronto, Ontario, Canada. [7]Department of Biology, ETH Zürich, Zürich, Switzerland. [8]Center for Behavioral Brain Sciences, Magdeburg, Germany. [9]Medical Faculty, Otto von Guericke University, Magdeburg, Germany. [10]Brain, Mind, & Consciousness Program, Canadian Institute for Advanced Research, Toronto, Ontario, Canada. [11]Child & Brain Development Program, Canadian Institute for Advanced Research, Toronto, Ontario, Canada.
*Corresponding author. Email: paul.frankland@sickkids.ca
†Present addresses: Institute of Pharmacology and Toxicology, University of Zürich, Zürich, Switzerland, and Switzerland Neuroscience Center, University of Zürich, Zürich, Switzerland.
‡Present addresses: Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada, and Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada.

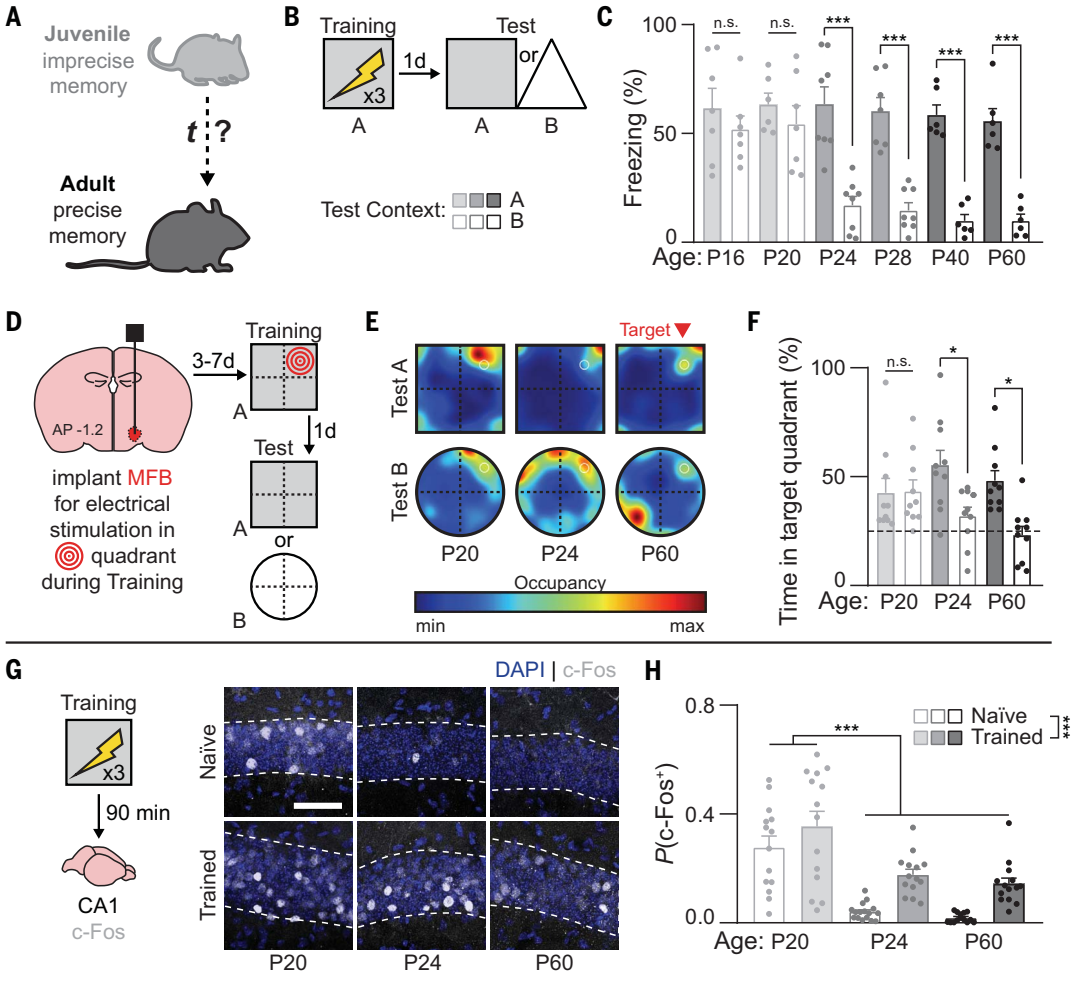


Fig. 1. Memory precision and sparse engrams develop in the dorsal CA1 during the fourth postnatal week. (**A**) The development of memory precision was assessed in mice. (**B**) Schematic of contextual fear conditioning protocol. (**C**) Juvenile mice (P16 to P20) formed imprecise contextual fear memories, whereas older (>P24) mice formed precise memories [analysis of variance (ANOVA), age × test context interaction: $F_{1,70}$ = 4.91, P < 0.001]. (**D**) Schematic of spatial foraging task. (**E**) Heatmaps depicting the average search pattern of P20, P24, and P60 mice during the test session. (**F**) Juvenile mice (P20) formed imprecise spatial memories, whereas older (>P24) mice formed precise memories (ANOVA, age × test context interaction: $F_{2,54}$ = 3.65, P < 0.05). (**G**) c-Fos expression in dorsal CA1 90 min after contextual fear conditioning. Images show c-Fos expression in a segment of the dorsal CA1 pyramidal layer. (**H**) Approximately twice as many CA1 pyramidal layer cells expressed c-Fos after conditioning (or in home cage) in P20 mice compared with conditioned P24 and P60 mice (ANOVA, main effect of age: $F_{2,80}$ = 34.81, $P < 10^{-6}$; main effect of experience: $F_{1,80}$ = 20.50, P < 0.0001). Data points are individual mice with mean ± SEM. Scale bar, 50 μm. *P < 0.05; ***P < 0.001.

Fig. 2, G and H), and these adult mice exhibited juvenile-like memory imprecision (equal freezing in contexts A and B) (Fig. 2, I and J, and fig. S5, E and F). Thus, the delayed onset of adult-like memory functions by hyperactivity within developing memory circuits may be a core feature of ontogeny across animal species (*24*).

Mature neuronal allocation develops during the fourth postnatal week in CA1

Such differences in engram sparsity between young and old mice suggest that the mechanisms of memory formation differ across development. In adult animals, eligible neurons are allocated to a sparse engram based on relative neuronal excitability or activity at the time of memory formation. To maintain engram sparsity, not only are neurons with relatively higher excitability included in the engram, but neurons with relatively lower excitability are excluded from the engram through lateral inhibition (*25*). We probed how neuronal allocation changes across development. We injected a replication-defective herpes simplex virus (HSV) into dorsal CA1 to infect a sparse random population of neurons with both a blue light (BL)–sensitive excitatory opsin, ChR2, and red light (RL)–sensitive inhibitory opsin, eNpHR3.0 (HSV-NpACY; Fig. 3, A and B, and fig. S6A). This approach allowed us to bidirectionally modulate the activity of the same population of infected neurons with different wavelengths of light (*26*). We briefly excited NpACY$^+$ neurons with BL immediately before conditioning to bias their allocation into the engram. Control mice were treated similarly but received no BL. After training, c-Fos was preferentially expressed in NpACY$^+$ neurons in P20, P24, and P60 mice in the BL$^+$ (allocated) but not BL$^-$ (control, nonallocated) groups (fig. S6, B to E), suggesting that optogenetically mediated allocation was effective regardless of mouse age.

In a second cohort of mice, we repeated the same allocation procedure and then determined whether infected neurons were necessary for subsequent memory expression by using RL to silence NpACY$^+$-expressing neurons during a memory test. Silencing NpACY$^+$ neurons impaired fear recall in P24 and P60 mice in the allocated group, but not in the control, nonallocated group (Fig. 3, C and D), indicating that the CA1 engram was localized to the

sparse NpACY$^+$ population of neurons. By contrast, silencing a similar number of NpACY$^+$ neurons did not impair fear memory recall in P20 mice. This suggests that information is more broadly distributed in densely encoded engrams in juvenile mice, such that silencing only a fraction of these neurons is not sufficient to disrupt memory recall. Consistent with this interpretation, a greater proportion of the dense juvenile engram was localized to NpACY$^-$ neurons (fig. S6F).

In juvenile mice, chemogenetic shrinking of the engram promoted adult-like neuronal allocation. Silencing allocated NpACY$^+$ neurons impaired fear recall in the juvenile C21 group (Fig. 3, E to G), indicating that artificially shrinking the CA1 engram in P20 mice localized the memory to the sparse NpACY$^+$ population of neurons. In adult mice, chemogenetic expansion of the engram induced juvenile-like neuronal allocation. Silencing allocated NpACY$^+$ neurons no longer impaired fear memory recall in the adult C21 group (Fig. 3, H to J), consistent with the idea that information was not localized to the sparse population of NpACY$^+$ neurons, but rather was more broadly distributed within the artificially expanded engram.

Maturation of parvalbumin-mediated inhibition is necessary for mature neuronal allocation and memory precision in CA1

The denser engrams in P20 mice raised the possibility that the second component of the mature neuronal allocation process, the exclusion of less-active or inactive neurons from engrams by interneurons, is not yet fully developed in juvenile mice. In the lateral amygdala, parvalbumin-expressing (PV$^+$) basket cells provide strong somatic inhibition onto excitatory neurons at the time of memory formation to exclude less excitable neurons from the engram (*26*, *27*). CA1 PV$^+$ interneurons are born embryonically in the medial ganglionic eminence, resulting in adult-like levels of these cells by the third postnatal week in mice (*6*, *28*) (Fig. 4, A and B). Despite their early birthdate, morphological and functional development of CA1 PV$^+$ interneurons continues into the fourth postnatal week in mice (*29*, *30*). Therefore, in P20 mice, developing PV$^+$ interneurons in the CA1 may provide only weak lateral inhibition, precluding the formation of sparse engrams at the time of memory encoding. Adult-like levels of PV$^+$ neurites and Syt2$^+$ puncta (labeling PV$^+$ interneuron presynaptic terminals; fig. S7A) emerged in the CA1 pyramidal layer at P24 (Fig. 4, C and D). Moreover, contextual fear conditioning increased the selectivity of perisomatic PV staining surrounding c-Fos$^-$ (compared with c-Fos$^+$) pyramidal layer cells at P24 and P60, but not at P20, consistent with the idea that experience-dependent lateral inhibition does not occur

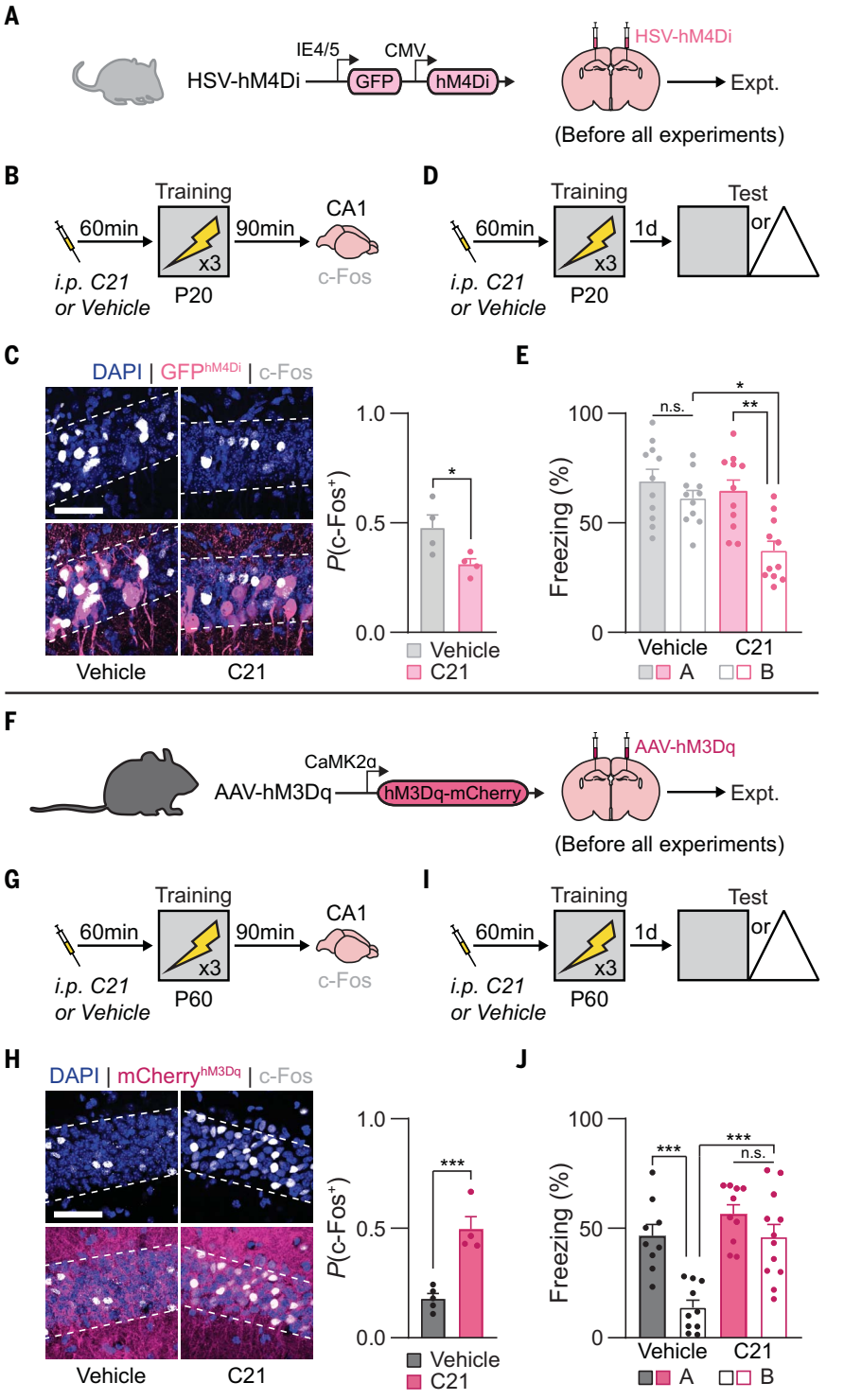


Fig. 2. Engram sparsity in the dorsal CA1 controls memory precision in juvenile and adult mice. (**A**) Engram size in dorsal CA1 of P20 mice was artificially decreased with HSV-hM4Di. (**B** and **C**) C21 was injected 1 hour before training (B) and reduced c-Fos expression in the juvenile CA1 (C) (unpaired *t* test: $t_6 = 2.57$, $P < 0.05$). (**D** and **E**) P20 mice administered C21 before training (D) formed precise contextual fear memories, and vehicle-treated P20 mice formed imprecise memories (E) (ANOVA, drug × test context interaction: $F_{1,26} = 4.48$, $P < 0.05$). (**F**) Engram size in dorsal CA1 of P60 mice was artificially increased with AAV-hM3Dq. (**G** and **H**) C21 was injected 1 hour before training (G) and increased c-Fos expression in the adult CA1 (H) (unpaired *t* test: $t_7 = 5.53$, $P < 0.001$). (**I** and **J**) P60 mice administered C21 before training (I) formed imprecise contextual fear memories, and vehicle-treated P60 mice formed precise memories (J) (ANOVA, drug × test context interaction: $F_{1,37} = 5.14$, $P < 0.05$). Data points are individual mice with mean ± SEM. Scale bars, 50 μm. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

before the fourth postnatal week (Fig. 4, E and F, and fig. S7, B to E).

Next, we expressed hM4Di in adult PV-Cre mice to inhibit PV$^+$ interneurons during memory formation (Fig. 4G and fig. S8, A and B). Administering C21 before fear conditioning disinhibited local excitatory neurons, resulting in high levels of c-Fos in the CA1 after training (Fig. 4, H to J). Moreover, inhibiting CA1 PV$^+$ interneurons in adult mice promoted juvenile-like allocation and memory imprecision (Fig. 4, K to O, and fig. S8, C and D). Reinstatement of these juvenile-like mnemonic phenotypes in adult PV-Cre mice required the presence of both hM4Di and C21 at the time of memory encoding (and not memory retrieval) (fig. S8, E to H). Thus, acute PV$^+$ interneuron inhibition at the time of memory formation was sufficient to phenocopy the allocation and memory specificity profile of juvenile mice while also maintaining the critical role of CA1 in encoding and retrieving these modified memories (fig. S8, I to L). These experiments identify maturation of CA1 inhibitory PV$^+$ circuitry as the key driver of memory development during the fourth postnatal week.

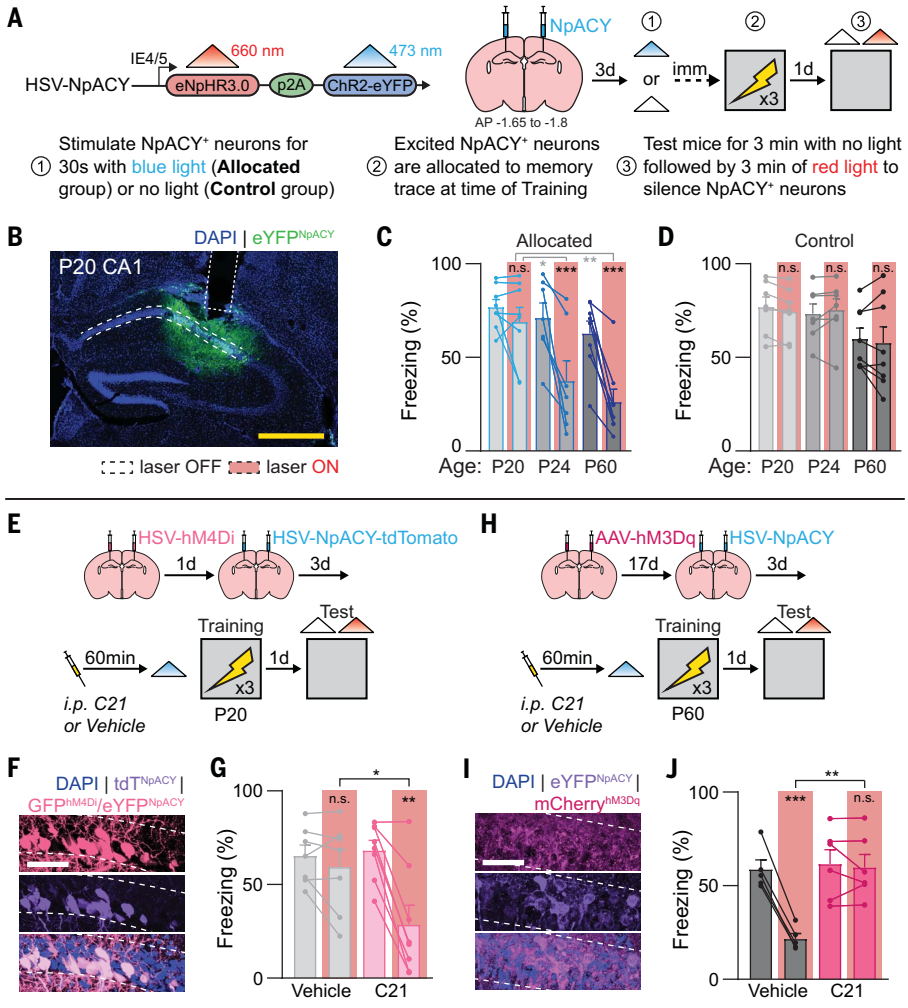


Fig. 3. Immature neuronal allocation mechanisms in the dorsal CA1 preclude localization of memories to sparse engrams during early development. (**A**) Schematic of the "allocate-and-silence" contextual fear conditioning protocol. HSV-NpACY was used to optogenetically excite (ChR2) and inhibit (NpHR3.0) the same neurons. (**B**) HSV-NpACY expression in dorsal CA1 of a P20 mouse. (**C**) Silencing a sparse group of NpACY$^+$ neurons previously allocated to the contextual fear memory engram impaired freezing during the test in P24 and P60 mice, but not P20 mice [repeated-measures (RM)–ANOVA, age × light interaction: $F_{2,20} = 4.96$, $P < 0.05$]. (**D**) Silencing a sparse group of random NpACY$^+$ neurons not allocated to the engram did not impair freezing during the test in any mice (RM-ANOVA, $P > 0.05$). (**E**) HSV-hM4Di and HSV-NpACY were both expressed in dorsal CA1 before contextual fear conditioning. C21 was administered 1 hour before training to promote sparse engram formation, and HSV-NpACY was used to allocate a sparse group of neurons to the engram at the time of training and to silence the same neurons during the test. (**F**) Expression of NpACY and hM4Di in the dorsal CA1 of a P20 mouse. (**G**) Silencing a sparse group of NpACY$^+$ neurons previously allocated to the contextual fear memory engram impaired freezing during the test in P20 mice that received C21 to shrink their engrams (RM-ANOVA, drug × light interaction: $F_{1,13} = 14.25$, $P < 0.01$). (**H**) AAV-hM3Dq and HSV-NpACY were both expressed in dorsal CA1 before contextual fear conditioning. C21 was administered 1 hour before training to promote dense engram formation, and HSV-NpACY was used to allocate a sparse group of neurons to the engram at the time of training and to silence the same neurons during the test. (**I**) Expression of NpACY and hM3Dq in the dorsal CA1 of a P60 mouse. (**J**) Silencing a sparse group of NpACY$^+$ neurons previously allocated to the contextual fear memory engram did not impair freezing during the test in P60 mice that received C21 to expand their engrams (RM-ANOVA, drug × light interaction: $F_{1,9} = 63.22$, $P < 0.0001$). Data points are individual mice with mean ± SEM. Scale bars: white, 50 μm; yellow, 500 μm. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Perineuronal nets regulate maturation of PV$^+$ interneurons in CA1

PV$^+$ interneurons play a pivotal role in the development of cortical sensory systems by closing transient windows of critical period plasticity (*31*). For example, PV$^+$ interneuron maturation closes the critical period for ocular dominance plasticity in the binocular primary visual cortex (V1b) in mice and rats (*32*). The maturation of perineuronal nets (PNNs), extracellular matrix (ECM) structures primarily ensheathing the soma and proximal dendrites of PV$^+$ interneurons, helps to drive PV$^+$ cell maturation in the cortex and hippocampus. By stabilizing excitatory synapses onto PV$^+$ interneurons and inhibitory synapses originating from PV$^+$ interneurons, mature PNNs increase PV$^+$ interneuron–mediated inhibition (*33*–*35*). Because PNN formation in V1b is necessary for the emergence of adult-like visual acuity, we reasoned that PNN formation in the hippocampus may similarly regulate the development of adult-like mnemonic specificity.

Mature PNNs represent a form of dense ECM composed of polymer chains of hyaluronan (HA), chondroitin sulfate proteoglycans (CSPGs) of the lectican family, tenascin-R, and link proteins (*36*, *37*). To establish a detailed developmental profile of PNNs in the hippocampus, we visualized *Wisteria floribunda* agglutinin (WFA), CSPG brevican (BCAN), and hyaluronan and proteoglycan link protein 1 (HAPLN1) in situ. WFA$^+$ PNNs were found throughout the adult hippocampus (fig. S9, A to C). Hippocampal PNNs primarily surrounded PV$^+$ interneurons (except for fasciola cinereum and CA2 PNNs) (fig. S9, D to I) and developed postnatally. In CA1, adult-like levels of PNNs surrounding PV$^+$ interneurons were achieved by P24 in both male and female mice (Fig. 5, A to D, and fig. S10, A to P). CA1 PNN density was not altered by contextual fear conditioning (fig. S10, Q and R), suggesting that only microscale PNN alterations are induced by memory formation (*38*). In contrast to CA1, PV$^+$ interneuron–associated PNNs reached adult levels before P16 in the dorsal dentate

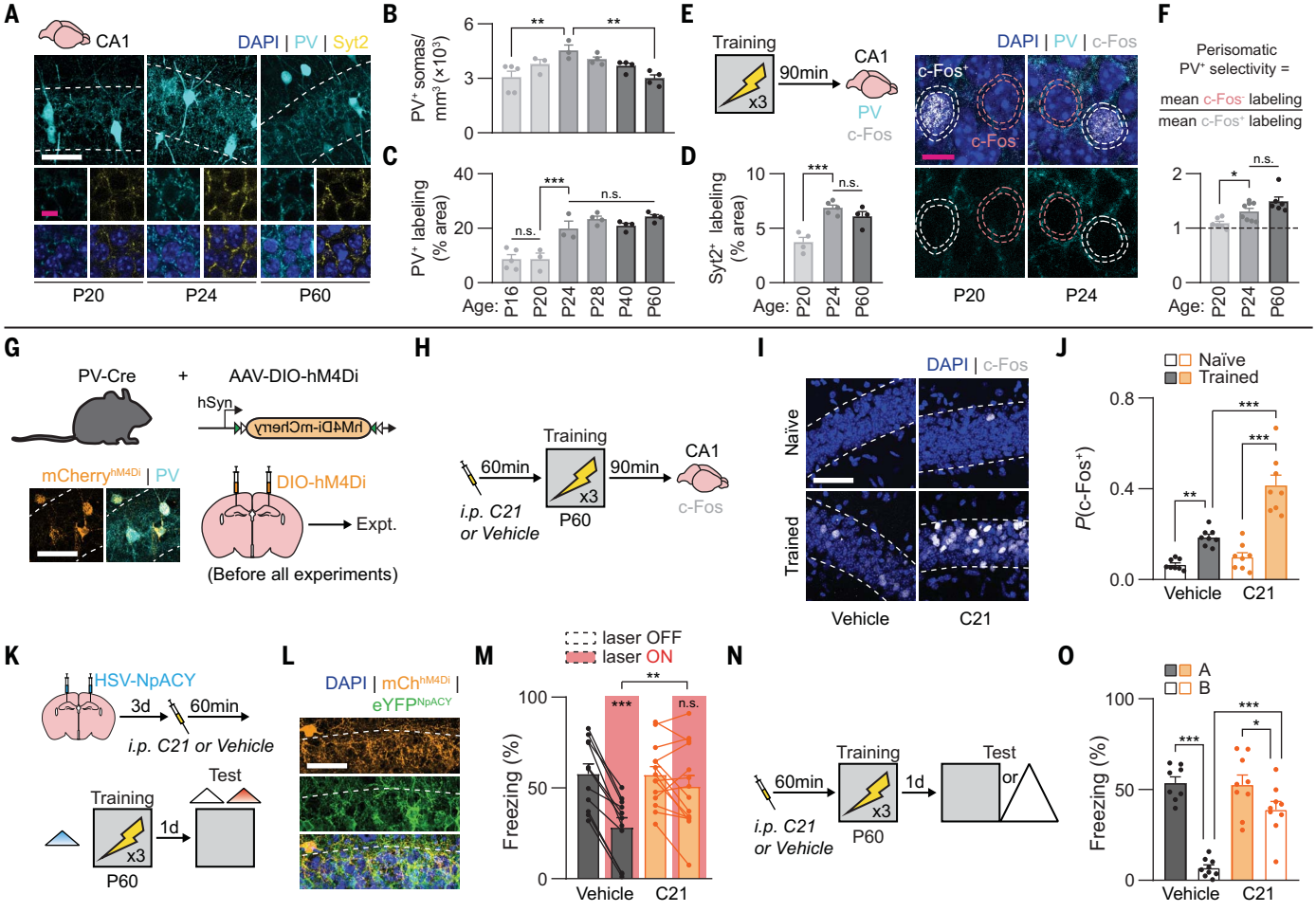


Fig. 4. Mature dorsal CA1 PV$^+$ interneuron function is required for competitive neuronal allocation, sparse engram formation, and memory precision. (**A**) PV$^+$ interneurons, PV$^+$ neurites, and Syt2$^+$ terminals in dorsal CA1 across development. (**B**) The number of PV$^+$ interneurons peaks transiently at P24 (B) (ANOVA, effect of age: $F_{5,17} = 5.02$, $P < 0.01$). (**C** and **D**) The density of PV$^+$ neurites (C) (ANOVA, effect of age: $F_{5,17} = 21.78$, $P < 10^{-5}$) and Syt2$^+$ synaptic terminals (D) (ANOVA, effect of age: $F_{2,10} = 18.01$, $P < 0.001$) in the pyramidal layer reach adult-like levels by P24. (**E**) c-Fos and PV expression were examined in dorsal CA1 90 min after contextual fear conditioning. Localization of PV$^+$ neurites around c-Fos$^-$ and c-Fos$^+$ pyramidal layer cells is shown. PV$^+$ labeling was quantified in a 3-μm ring surrounding the nuclei. (**F**) The selectivity of PV$^+$ neurites around c-Fos$^-$ compared with c-Fos$^+$ cells after training reached adult-like levels by P24 (ANOVA, effect of age: $F_{2,17} = 9.04$, $P < 0.01$). (**G**) PV$^+$ interneurons were inhibited by expressing AAV-DIO-hM4Di in the dorsal CA1 of adult PV-Cre mice. (**H**) c-Fos expression in dorsal CA1 was examined 90 min after contextual fear conditioning. C21 was administered 1 hour before training to inhibit PV$^+$ interneurons. (**I**) c-Fos expression in the dorsal CA1 pyramidal layer. (**J**) Inhibiting PV$^+$ interneurons before training resulted in a twofold increase in c-Fos expression after training (ANOVA, drug × experience interaction: $F_{1,28} = 13.82$, $P < 0.001$). (**K**) Schematic of the contextual fear conditioning protocol. C21 was administered 1 hour before training to inhibit PV$^+$ interneurons, and HSV-NpACY was used to excite (ChR2) and inhibit (NpHR3.0) the same neurons. (**L**) Expression of NpACY and hM4Di in the dorsal CA1 of a P60 mouse. (**M**) Silencing a sparse group of NpACY$^+$ neurons previously allocated to the contextual fear memory engram did not impair freezing during the test in P60 mice that received C21 to inhibit their PV$^+$ interneurons (RM-ANOVA, drug × light interaction: $F_{1,23} = 21.12$, $P < 0.001$). (**N**) Schematic of the contextual fear conditioning protocol. (**O**) P60 mice administered C21 to inhibit PV$^+$ interneurons formed imprecise contextual fear memories, and vehicle-treated mice formed precise memories (ANOVA, drug × test context interaction: $F_{1,31} = 16.56$, $P < 0.001$). Data points are individual mice with mean ± SEM. Scale bars: magenta, 10 μm; white, 50 μm. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

gyrus and CA3 (fig. S11, A to K), consistent with the stepwise maturation of hippocampal trisynaptic circuitry (*21*).

HAPLN1 plays a key role in cross-linking and stabilizing CSPG-hyaluronan interactions, and *Hapln1* transcription corresponds with mature PNN formation in the cortex (*39*). To promote or interfere with PNN integrity, we designed viral vectors to overexpress wild-type full-length HAPLN1 (AAV-Hapln1) or mutant dominant-negative ΔHAPLN1 (AAV-ΔHapln1) proteins tagged with cysteine-free green fluorescent protein (cfGFP) or cfGFP only (AAV-cfGFP) (Fig. 5E). In contrast to full HAPLN1, ΔHAPLN1 did not promote binding of HA to lecticans (*40*), as we showed here for the first time (fig. S12, A to L). Expression of the constructs did not affect total endogenous expression of lecticans in the dorsal hippocampus of adult mice (fig. S13, A to L). However, AAV-ΔHapln1 expression destabilized CA1 PNNs in P60 mice and consequently reduced the density of PV$^+$ neurites in the CA1 pyramidal layer (Fig. 5, F to J). By contrast, AAV-Hapln1 expression accelerated the maturation of CA1 PNNs and PV$^+$ neurites in P20 mice (Fig. 5, K to O).

PNN maturation regulates the emergence of neuronal allocation and memory precision in CA1

First, we tested whether destabilizing CA1 PNNs in adult mice with AAV-ΔHapln1 would phenocopy the formation of a dense engrams and

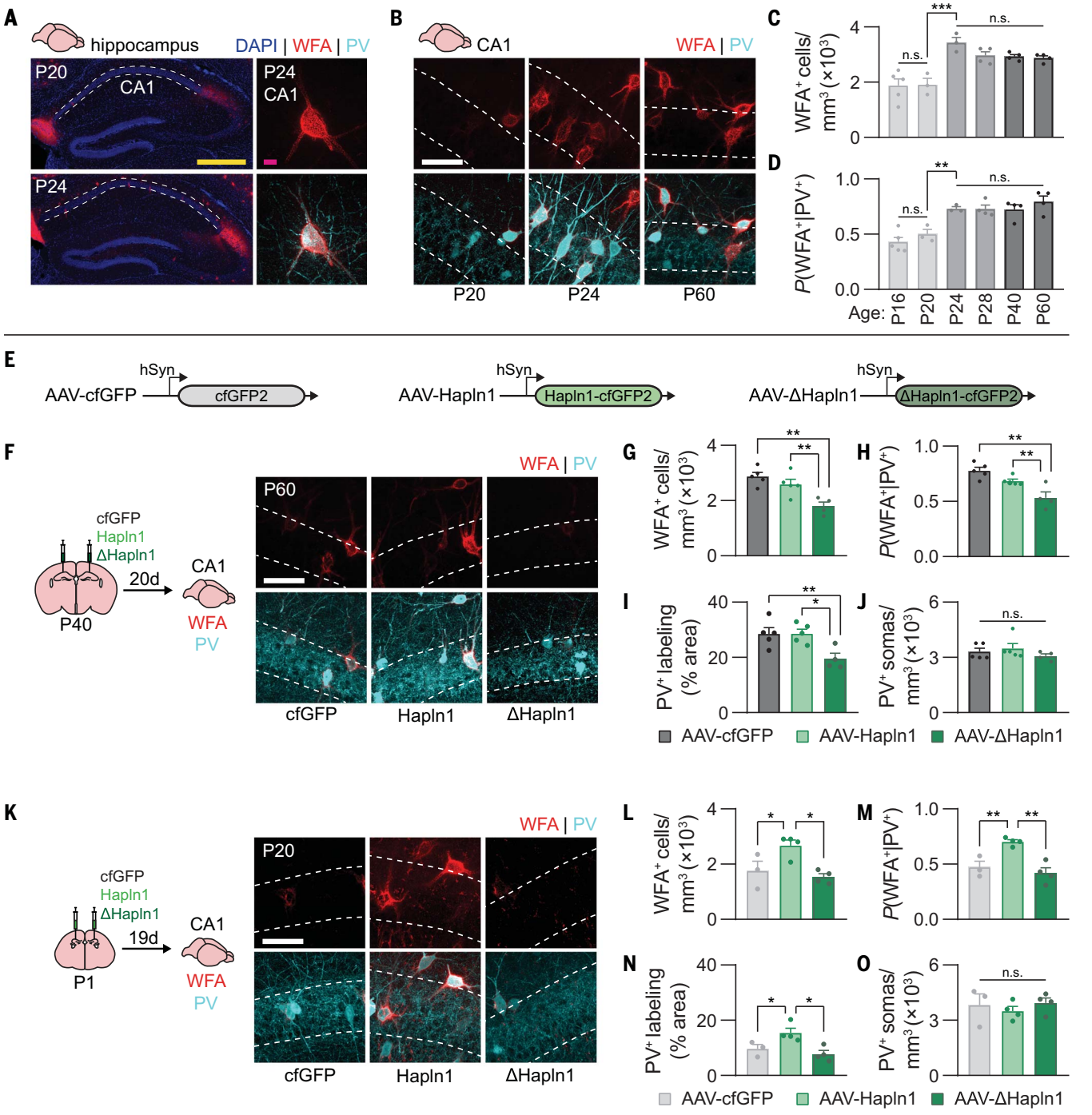


Fig. 5. Maturation of dorsal CA1 PNNs and PV$^+$ interneurons requires HAPLN1. (**A**) WFA$^+$ PNNs in the hippocampus of a P20 and P24 mouse. High-magnification image of a WFA$^+$ PNN on P24 surrounding the soma and proximal dendrites of a PV$^+$ interneuron is shown. (**B**) WFA$^+$ PNNs surrounding PV$^+$ interneurons in dorsal CA1 across development. (**C** and **D**) The density of WFA$^+$ PNNs in dorsal CA1 (ANOVA, effect of age: $F_{5,17} = 11.88$, $P < 0.0001$) surrounding PV$^+$ interneurons (ANOVA, effect of age: $F_{5,17} = 13.51$, $P < 0.0001$) reached adult-like levels by P24. (**E**) To bidirectionally manipulate PNN integrity in vivo, we used AAV-Hapln1, AAV-ΔHapln1, and AAV-cfGFP to overexpress wild-type mouse HAPLN1 or a mutated dominant-negative ΔHAPLN1 or cfGFP as a negative control, respectively. (**F**) Mice were microinjected with AAVs, and P60 brains were stained for WFA$^+$ PNNs and PV$^+$ interneurons in the dorsal CA1. (**G** to **J**) Expression of AAV-ΔHapln1 in P60 CA1 decreased the number of WFA$^+$ PNNs (G) (ANOVA, effect of virus: $F_{2,11} = 10.88$, $P < 0.01$) surrounding PV$^+$ interneurons (H (ANOVA, effect of virus: $F_{2,11} = 12.52$, $P < 0.01$), and density of PV$^+$ neurites in the pyramidal layer (I) (ANOVA, effect of virus: $F_{2,11} = 6.88$, $P < 0.05$), without affecting the number of PV$^+$ interneurons (J) (ANOVA, no effect of virus: $F_{2,11} = 0.89$, $P = 0.43$). (**K**) Mice were microinjected with AAVs, and P20 brains were stained for WFA$^+$ PNNs and PV$^+$ interneurons in the dorsal CA1. (**L** to **O**) Expression of AAV-Hapln1 in P20 CA1 increased the number of WFA$^+$ PNNs (L) (ANOVA, effect of virus: $F_{2,8} = 8.02$, $P < 0.05$) surrounding PV$^+$ interneurons (M) (ANOVA, effect of virus: $F_{2,8} = 14.43$, $P < 0.01$) and density of PV$^+$ neurites in the pyramidal layer (N) (ANOVA, effect of virus: $F_{2,8} = 7.47$, $P < 0.05$) without affecting the number of PV$^+$ interneurons (O) (ANOVA, no effect of virus: $F_{2,8} = 0.42$, $P = 0.66$). Data points are individual mice with mean ± SEM. Scale bars: magenta, 10 μm; white, 50 μm; yellow, 500 μm. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

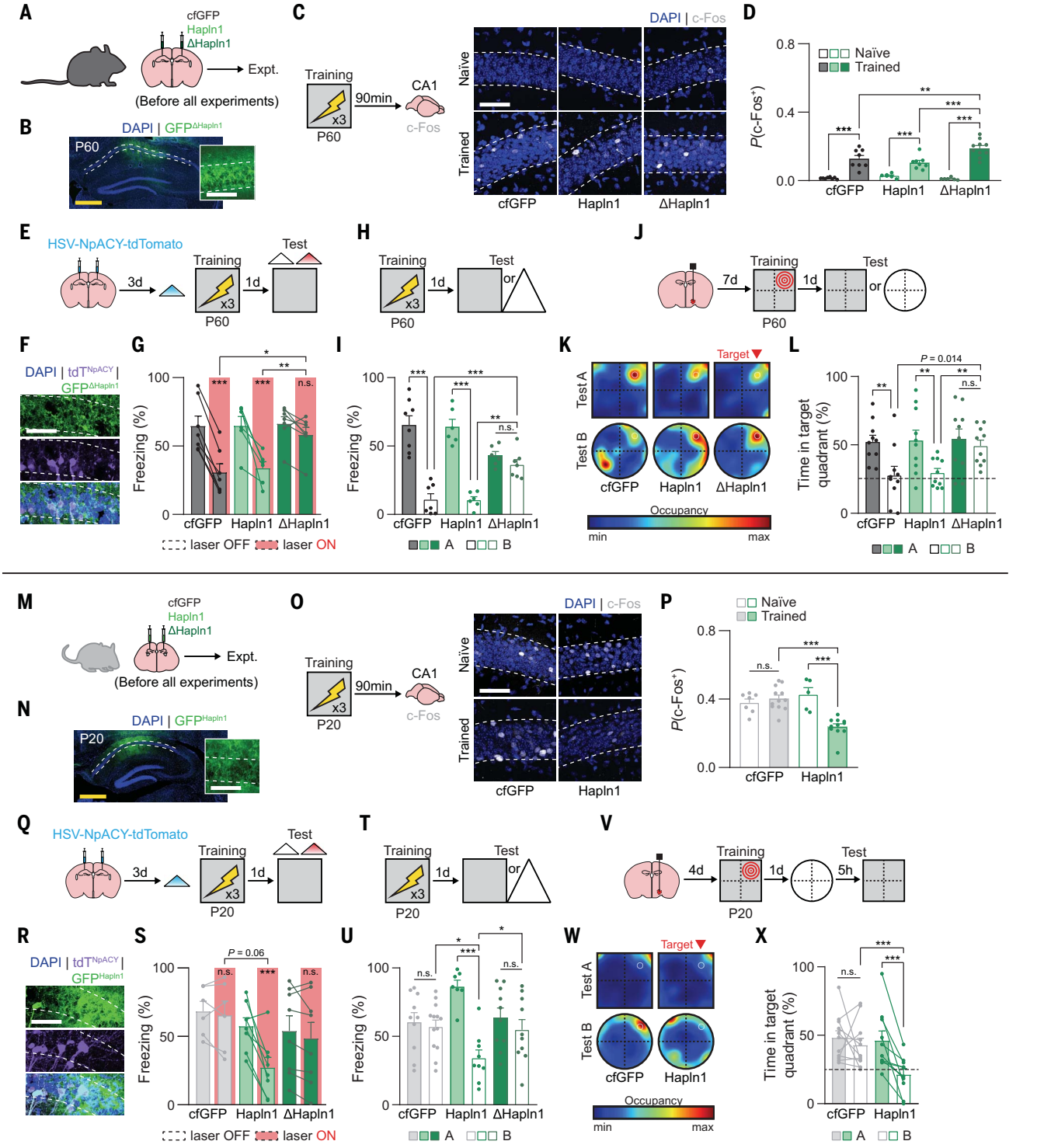


Fig. 6. Maturation of dorsal CA1 PNNs is required for competitive neuronal allocation, sparse engram formation, and memory precision. (**A**) PNNs were manipulated in the adult dorsal CA1 by microinjecting AAVs before each experiment. (**B**) AAV-ΔHapln1 expression in the P60 dorsal CA1. (**C**) c-Fos expression in dorsal CA1 was examined 90 min after contextual fear conditioning. (**D**) Expression of AAV-ΔHapln1 in P60 dorsal CA1 before training resulted in a twofold increase in c-Fos expression after training (ANOVA, virus × experience interaction: $F_{2,40} = 6.89$, $P < 0.01$). (**E**) Schematic of the contextual fear conditioning protocol. HSV-NpACY was used to optogenetically excite (ChR2) and inhibit (NpHR3.0) the same neurons. (**F**) Expression of NpACY and AAV-ΔHapln1 in dorsal CA1 of P60 mouse. (**G**) Silencing a sparse group of NpACY$^+$ neurons previously allocated to the contextual fear memory engram did not impair freezing during the test in P60 mice expressing AAV-ΔHapln1 in dorsal CA1 (RM-ANOVA, virus × light interaction: $F_{2,17} = 9.58$, $P < 0.01$). (**H**) Schematic of the contextual fear conditioning protocol. (**I**) P60 mice expressing AAV-ΔHapln1 in dorsal CA1 formed imprecise contextual fear memories, and P60 mice expressing AAV-cfGFP or AAV-Hapln1 formed precise memories (ANOVA, virus × test context interaction: $F_{2,34} = 16.40$, $P < 0.0001$). (**J**) Schematic of the spatial foraging task. (**K**) Heatmaps depicting example search patterns during the test. (**L**) P60 mice expressing AAV-ΔHapln1 in

dorsal CA1 formed imprecise spatial memories, and P60 mice expressing AAV-cfGFP or AAV-Hapln1 formed precise memories (unpaired *t* tests with Bonferroni correction, $\alpha = 0.01$, AAV-cfGFP context A versus context B: $t_{18} = 2.99$, $P < 0.01$; AAV-Hapln1 context A versus context B: $t_{18} = 2.90$, $P < 0.01$; AAV-ΔHapln1 context A versus context B: $t_{18} = 0.59$, $P = 0.56$; AAV-cfGFP context B versus AAV-ΔHapln1 context B: $t_{18} = 2.70$, $P = 0.014$; AAV-Hapln1 context B versus AAV-ΔHapln1 context B: $t_{18} = 3.61$, $P < 0.01$). (**M**) PNNs were manipulated in the juvenile dorsal CA1 by microinjecting AAVs before each experiment. (**N**) AAV-Hapln1 expression in the P20 dorsal CA1. (**O**) c-Fos expression in dorsal CA1 was examined 90 min after contextual fear conditioning. (**P**) Expression of AAV-Hapln1 in P20 dorsal CA1 before training resulted in a decrease in c-Fos expression after training (ANOVA, virus × experience interaction: $F_{1,31} = 19.07$, $P < 0.001$). (**Q**) Schematic of the contextual fear conditioning protocol. HSV-NpACY was used to optogenetically excite (ChR2) and inhibit (NpHR3.0) the same neurons. (**R**) Expression of NpACY and AAV-Hapln1 in the dorsal CA1 of P20 mouse. (**S**) Silencing a sparse group of NpACY$^+$ neurons previously allocated to the contextual fear memory engram impaired freezing during the test in P20 mice expressing AAV-Hapln1 in dorsal CA1 (RM-ANOVA, virus × light interaction: $F_{2,19} = 8.10$, $P < 0.01$). (**T**) Schematic of the contextual fear conditioning protocol. (**U**) P20 mice expressing AAV-Hapln1 in CA1 formed precise contextual fear memories, and P20 mice expressing AAV-cfGFP or AAV-ΔHapln1 formed imprecise memories (ANOVA, virus × test context interaction: $F_{2,53} = 7.83$, $P < 0.01$). (**V**) Schematic of the spatial foraging task. (**W**) Heatmaps depicting example search patterns during the test. (**X**) P20 mice expressing AAV-Hapln1 in dorsal CA1 formed precise spatial memories, and P20 mice expressing AAV-cfGFP formed imprecise memories (RM-ANOVA, virus × test context interaction: $F_{1,19} = 4.09$, $P = 0.05$). Data points are individual mice with mean ± SEM. Scale bars: white, 50 μm; yellow, 500 μm. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

memory imprecision typical of juvenile mice (Fig. 6, A and B). In CA1 of adult mice overexpressing ΔHAPLN1 (but not cfGFP or wild-type HAPLN1), training-induced PV-mediated lateral inhibition was precluded (fig. S14, A to C) and dense c-Fos$^+$ engrams formed (Fig. 6, C and D), resulting in intact memory when sparse populations of artificially allocated NpACY$^+$ neurons were silenced during testing (Fig. 6, E to G). CA1 PNN destabilization in separate groups of adult mice produced imprecise contextual fear memory (Fig. 6, H and I, and fig. S15, A and B) and appetitive spatial memory (Fig. 6, J to L, and fig. S16, A and B). The effects of ΔHAPLN1 in adult mice were specific to episodic-like memory formation. Destabilization of CA1 PNNs with AAV-ΔHapln1 did not decrease precision for auditory fear memories or increase anxiety-like behavior or locomotion in an open field (fig. S15, C to G). Changes in contextual memory precision were largely restricted to manipulation of PNNs in CA1 (and not cortex) (fig. S15, H to K), which we also observed using the enzyme ChABC (*35*) to transiently digest PNNs (figs. S15, L to O, and S17, A to R).

Finally, we tested whether accelerating PNN maturation with AAV-Hapln1 was sufficient to produce adult-like memory phenotypes of sparse engram formation and precise contextual memories (Fig. 6, M and N). We observed adult-like neuronal allocation in CA1 of juvenile mice overexpressing wild-type HAPLN1, evidenced by the early onset of PV-mediated lateral inhibition (fig. S14, D to F) and sparse c-Fos$^+$ engram formation (Fig. 6, O and P). Consistent with sparse engram formation, silencing sparse populations of artificially allocated NpACY$^+$ neurons impaired memory recall in juveniles expressing AAV-Hapln1, but not AAV-cfGFP or AAV-ΔHapln1 (Fig. 6, Q to S), and these mice formed precise fear memories (Fig. 6, T and U, and fig. S15, P and Q) and reward memories (Fig. 6, V to X, and fig. S16, D and E). Similarly, direct infusion of recombinant brain-derived neurotrophic factor, a treatment that accelerates maturation of neural circuits (*32*), into CA1 on P17 also resulted in the precocial maturation of PNNs, PV$^+$ interneurons, and memory precision in juvenile mice (fig. S18, A to I).

Discussion

The specific neurobiological mechanisms regulating age-dependent increases in the precision of episodic and episodic-like memory have long remained elusive (*5*, *7*, *8*). We found that the developing CA1 is actively engaged in early memory formation despite the immaturity of some memory formation mechanisms. Our findings identify engram sparsity as a key determinant of memory precision. Specifically, we identified maturation of a competitive neuronal allocation process, supported by developing PV$^+$ interneurons, as a prerequisite for encoding precise, episodic-like memories in sparse engrams. These findings support the recent proposal that the hippocampus transitions through distinct stages of functional development, with the stepwise emergence of CA1 phenomena important for memory consolidation (*41*, *42*). These processes disproportionately affect engram neurons (*43*, *44*), providing potential mechanisms for post-encoding stabilization of sparse engrams during the fourth postnatal week and onward.

We discovered that maturation of PV$^+$ inhibitory circuitry required for the onset of adult-like neuronal allocation and episodic-like memory precision in CA1 are dependent on the extracellular matrix PNNs. In mammalian and avian brains, the accumulation of PNNs around PV$^+$ interneurons in defined cortical regions shifts local excitatory-inhibitory balance, dampens critical period plasticity, and controls the development of corresponding sensory processes or behaviors (*31*, *35*, *37*, *39*). Our data suggest that maturation of the hippocampal memory system is regulated by the same cellular and molecular mechanisms as cortical sensory systems. They also indicate that memory formation changes the relative levels of perisomatic inhibition of engram neurons through PNN-dependent mechanisms, likely mediated in part by the PNN glycoprotein tenascin-R (*45*). Therefore, ECM-dependent maturation of inhibitory neural circuits may be a brainwide mechanism for not only sensory development, but also cognitive and emotional development (*34*, *46*–*48*).

Why are there multiple stages of functional maturation in the development of the hippocampal memory system? One possibility is that hippocampal development involves progression of the episodic memory system from an "incomplete" (child-like) to a "complete" (adult-like) state. An alternative possibility is that the mnemonic functions of children are perfectly adapted for their stage of development (*49*). At the cognitive level, the encoding of schemas and other forms of broad or imprecise semantic knowledge during early life may be favored over encoding specific episodes, given that young children have comparatively few life experiences from which to draw and typically are not without parents or alloparents (with fully fledged episodic memory systems) for the first years of life (*50*). The immature hippocampus may have evolved to fulfill this purpose, exploiting the protracted development of inhibition and co-opting the same activity-dependent mechanisms required for structural and functional development of hippocampal circuitry (*21*–*23*) for early memory formation and storage in dense memory engrams.

REFERENCES AND NOTES

1. C. T. Ngo, Y. Lin, N. S. Newcombe, I. R. Olson, *J. Exp. Psychol. Gen.* **148**, 1463–1479 (2019).
2. L. Picard, S. Cousin, B. Guillery-Girard, F. Eustache, P. Piolino, *Child Dev.* **83**, 1037–1050 (2012).
3. H. Yim, A. F. Osth, V. M. Sloutsky, S. J. Dennis, *Psychol. Sci.* **33**, 1154–1171 (2022).
4. L. Rollins, E. B. Cloude, *Learn. Mem.* **25**, 294–297 (2018).
5. J. Bachevalier, F. Vargha-Khadem, *Curr. Opin. Neurobiol.* **15**, 168–174 (2005).
6. R. Cossart, R. Khazipov, *Physiol. Rev.* **102**, 343–378 (2022).
7. A. Keresztes, C. T. Ngo, U. Lindenberger, M. Werkle-Bergner, N. S. Newcombe, *Trends Cogn. Sci.* **22**, 676–686 (2018).
8. A. I. Ramsaran, M. L. Schlichting, P. W. Frankland, *Dev. Cogn. Neurosci.* **36**, 100591 (2019).
9. T. A. Allen, N. J. Fortin, *Proc. Natl. Acad. Sci. U.S.A.* **110** (Suppl 2), 10379–10386 (2013).
10. M. J. Anderson, D. C. Riccio, *Learn. Behav.* **33**, 444–453 (2005).
11. I. Goshen *et al.*, *Cell* **147**, 678–689 (2011).
12. L. Nadel, S. Zola-Morgan, "Infantile amnesia," in *Infant Memory*, M. Moscovitch, Ed. (Springer, 1984), vol. 9, pp. 145–172.
13. J. Bachevalier, M. Mishkin, *Behav. Neurosci.* **98**, 770–778 (1984).
14. C. T. Ellis *et al.*, *Curr. Biol.* **31**, 3358–3364.e4 (2021).

15. A. C. Schapiro, N. B. Turk-Browne, M. M. Botvinick, K. A. Norman, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20160049 (2017).
16. S. A. Josselyn, S. Köhler, P. W. Frankland, *Nat. Rev. Neurosci.* **16**, 521–534 (2015).
17. J. T. Wixted *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9621–9626 (2014).
18. K. Z. Tanaka *et al*., *Science* **361**, 392–397 (2018).
19. K. K. Tayler, K. Z. Tanaka, L. G. Reijmers, B. J. Wiltgen, *Curr. Biol.* **23**, 99–106 (2013).
20. D. J. Cai *et al*., *Nature* **534**, 115–118 (2016).
21. F. Donato, R. I. Jacobsen, M.-B. Moser, E. I. Moser, *Science* **355**, eaai8178 (2017).
22. A. Travaglia, R. Bisaz, E. Cruz, C. M. Alberini, *Neurobiol. Learn. Mem.* **135**, 125–138 (2016).
23. X. Gao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12531–12536 (2018).
24. S. G. Leinwand, K. Scott, *Neuron* **109**, 1836–1847.e5 (2021).
25. S. A. Josselyn, P. W. Frankland, *Annu. Rev. Neurosci.* **41**, 389–413 (2018).
26. A. J. Rashid *et al*., *Science* **353**, 383–387 (2016).
27. D. J. Morrison *et al*., *Neurobiol. Learn. Mem.* **135**, 91–99 (2016).
28. F. Donato, A. Chowdhury, M. Lahr, P. Caroni, *Neuron* **85**, 770–786 (2015).
29. D. Doischer *et al*., *J. Neurosci.* **28**, 12956–12968 (2008).
30. L. Que, D. Lukacsovich, W. Luo, C. Földy, *Nat. Commun.* **12**, 108 (2021).
31. T. K. Hensch, *Nat. Rev. Neurosci.* **6**, 877–888 (2005).
32. Z. J. Huang *et al*., *Cell* **98**, 739–755 (1999).
33. E. Dzyubenko *et al*., *Cell. Mol. Life Sci.* **78**, 5647–5663 (2021).
34. E. Favuzzi *et al*., *Neuron* **95**, 639–655.e10 (2017).
35. T. Pizzorusso *et al*., *Science* **298**, 1248–1251 (2002).
36. J. W. Fawcett, T. Oohashi, T. Pizzorusso, *Nat. Rev. Neurosci.* **20**, 451–465 (2019).
37. O. Senkov, P. Andjus, L. Radenovic, E. Soriano, A. Dityatev, *Prog. Brain Res.* **214**, 53–80 (2014).
38. V. Nagy *et al*., *J. Neurosci.* **26**, 1923–1934 (2006).
39. D. Carulli *et al*., *Brain* **133**, 2331–2347 (2010).
40. K. Matsumoto *et al*., *J. Biol. Chem.* **278**, 41205–41212 (2003).
41. U. Farooq, G. Dragoi, *Science* **363**, 168–173 (2019).
42. A. Figurov, L. D. Pozzo-Miller, P. Olafsson, T. Wang, B. Lu, *Nature* **381**, 706–709 (1996).
43. Y. Jeong *et al*., *Nat. Commun.* **12**, 3915 (2021).
44. H. Norimoto *et al*., *Science* **359**, 1524–1527 (2018).
45. A. K. Saghatelyan *et al*., *Eur. J. Neurosci.* **12**, 3331–3342 (2000).
46. M. T. Birnie, T. Z. Baram, *Science* **376**, 1055–1056 (2022).
47. E. M. Nabel, H. Morishita, *Front. Psychiatry* **4**, 146 (2013).
48. N. Gogolla, P. Caroni, A. Lüthi, C. Herry, *Science* **325**, 1258–1261 (2009).
49. R. V. Kail, N. E. Spear, in *Comparative Perspectives on the Development of Memory*, R. V. Kail, N. E. Spear, Eds. (Psychology Press, 1984); pp. 325–358.
50. A. Gopnik, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190502 (2020).
51. A. Ramsaran *et al*., Data for: A shift in the mechanisms controlling hippocampal engram formation during brain maturation, Dryad (2023).

ACKNOWLEDGMENTS

We thank A. DeCristofaro, D. Lin, M. Yamamoto, and K. Böhm for excellent technical assistance; C. Seidenbecher and R. Frischknecht for reagents; and members of the Frankland and Josselyn laboratories for general comments. **Funding:** This work was supported by Brain Canada (P.W.F. and S.A.J.); the Canadian Institutes of Health Research (CIHR grant PJT180530 to P.W.F.); a CIHR postdoctoral fellowship (A.Aw.); a CIHR Vanier Canada graduate scholarship (M.L.D.); the German Research Foundation (Deutsche Forschungsgemeinschaft Project-ID 425899996-SFB 1436 TP05 to A.D.); the German Center for Neurodegenerative Diseases (A.D.); the Hilda and William Courtney Clayton Pediatric Research Fund (S.Y.K.); a Hospital for Sick Children Restracomp fellowship (A.Aw. and L.M.T.); the National Institutes of Health (grant R01 MH119421 to P.W.F. and S.A.J. and grant F31 MH120920-01 to A.I.R.); the Natural Sciences and Engineering Research Council of Canada (NSERC Canada Graduate Scholarship to A.I.R. and L.M.T.); an Ontario Graduate Scholarship (Y.W.); an Ontario Trillium Scholarship (Y.W.); and the Vector Institute (L.M.T.). **Author contributions:** Conceptualization: A.I.R., P.W.F.; Development of HAPLN1 viral constructs: R.K., A.D.; Data collection – in vivo experiments: A.I.R., Y.W., A.G., M.L.D., B.A.Y., A.J.R., J.L., L.M.T., S.Y.K., A.Ab., L.C.D., C.M., J.G., M.A.; Data collection – molecular and in vitro binding experiments: S.A., M.L.D., A.Aw.; Funding acquisition: A.D., S.A.J., P.W.F.; Methodology: A.I.R., A.D., P.W.F.; Statistical analyses: A.I.R., S.A.; Writing – original draft: A.I.R., P.W.F.; Writing – review and editing: A.I.R., A.D., S.A.J., P.W.F. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data generated during the present study and complete sequences for the HAPLN1 plasmids are archived in the Dryad repository (*51*). Requests for plasmids used in this study may be made to the corresponding author. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade6530
Materials and Methods
Figs. S1 to S18
References (*52*–*70*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 30 August 2022; accepted 24 March 2023
10.1126/science.ade6530

By **Janaynne Carvalho do Amaral**

Growing apart, staying connected

"When will you stop studying and get married and have children?" My parents have been asking me that question for years. I grew up in a small city in Brazil, in a family where education—especially for women—was not a priority. Neither of my parents finished secondary school; my dad is a truck driver and my mom works as a hotel cleaner and housekeeper. The day I told my mom I got a job offer to do a postdoc in the United States, she said, "Don't forget the man is the leader of the house." I love my family, and they love me. But my education has given me different dreams and created distance between us that we are trying to bridge.

My grandmother helped spark my education and career ambitions by teaching me to buy and exchange books at secondhand bookshops. Books ignited a hunger to learn about the world around me and become a scientist. Throughout my bachelor's degree, master's degree, and doctorate, I was told that was not an option. Professors and peers said my family background and the poor public schools I had attended put my dream of becoming a researcher out of reach. But I was addicted to learning and wouldn't give up. Thankfully I also had incredible mentors who supported me, and access to programs intended to help students like me succeed.

"I love my family, and they love me. But my education has ... created distance between us."

I had expected the academic hurdles. But I was caught off-guard by the challenges within my family. I felt increasingly distant from them, starting during my bachelor's degree. My parents don't understand academic challenges and rewards—why it takes so long to write and publish my research, why I persist in studying despite many years of poor salary, the happiness of getting a manuscript accepted or getting an award, especially when no money is involved. I can't discuss politics, race, or gender with them without fighting, and whenever I feel sad they still advise me to get married.

Still, there is profound love between us. I celebrate the things they taught me—my mother's lessons in cooking and treating others with empathy, my dad's in driving a car and creatively solving everyday problems. I respect their values, listen to their perspectives, go to church with them, and watch their favorite movies and TV shows without judgment. My mother once said to me, "I don't understand how I am your mother, but you do not know how proud I am of you." After I started my postdoc, she told me, "You are a big star here." My grandmother tells everyone she has a rich granddaughter. My family does not understand my scholarly books, yet they are displayed everywhere in their house like art by a famous painter.

But even this pride creates distance between us. They think I am special, a genius—not that I got where I am thanks to hard work and dedication. Since starting my postdoc, when I come home to visit they feel nervous because they think their food and homes are not good enough for me. Right before I started, my mom bought me a gift—a novel about two scientists falling in love, which I had told her about. I guess she wanted to show her affection for me while also reminding me that I should still get married! I was touched and amused, and I asked her to write a dedication. She asked my aunt to write a message on her behalf. "I don't want embarrass you," she said, referring to her struggles with writing. It broke my heart. I didn't care what she wrote or how she wrote it; I just wanted it to be from her. Yet I had become an intimidating figure to her.

I am still trying to figure out how to talk to my parents about my career and our differences, and our uncomfortable conversations about marriage continue. But I've come to realize that our love for each other transcends any distance. Curiously, my mother has started to say that I am living everything she dreamed for me, and she likes to tell everybody that she is my mother. I guess she never had someone to encourage her. I guess even though I scare her with my life choices, I helped her see that women can pursue our dreams. This article is for her. ■

Janaynne Carvalho do Amaral is a postdoctoral researcher at the University of Illinois Urbana-Champaign. Send your career story to SciCareerEditor@aaas.org.

25-mL Conical Tubes

Eppendorf's last tubes innovation—the conical tube format 25 mL—comes with new versions: Amber, Protein LoBind, and DNA LoBind. The innovative material of the amber-colored 25 mL conical tubes is highly transparent, while at the same time effectively protecting the sample from low-wavelength light. This way, the pipette tip immersion can be safely monitored during liquid transfer, which helps avoid cross-contamination. The Protein LoBind Conical Tubes minimizes the risk of non-specific protein and peptide samples sticking to the walls of the tubes. This ensures the best possible sample recovery, which is especially important when working with samples with low initial concentration. The DNA LoBind Conical Tubes help to optimize DNA recovery when preparing and storing for realtime PCR quanitifications. The choice of vessel can affect the reliability of DNA standard panels preparation used in real-time PCR assays for absolute quantifications. This can be seen, for example, in the retention of nucleic acids in vessels used for serial dilution and storage of standards.

Eppendorf
For info: 800-645-3050
www.eppendorf.com/25mL

Flow Cytometry Yellow and Blue Dyes

Bio-Rad Laboratories announced the launch of both the StarBright Blue-series and StarBright Yellow-series dyes to provide greater flexibility in conventional and full spectrum flow cytometry applications. Offering exceptional brightness with narrow excitation and emission profiles for precise resolution, the dyes expand research capabilities across the blue (488 nm) and yellow (561 nm) lasers. Each dye in the Bio-Rad range of highly stable StarBright Dyes is conjugated to flow-validated antibodies, supporting flow cytometry panel design in immunology research. They can be used in combination with other StarBright Dyes and traditional fluorophores. StarBright Dyes are compatible with most experimental protocols and flow cytometers, including the Bio-Rad ZE5 Cell Analyzer, without the need for special buffers. Minimal lot-to-lot variation helps to ensure reproducible and consistent staining, and the dyes are resistant to photobleaching with no loss of signal in fixation.

Bio-Rad Laboratories
For info: 1-800-424-6723
www.bio-rad-antibodies.com/starbright-test-size.html

UV Spectrophotometer for Batch and Flow Applications

Compact in design, the Flow-UV from Uniqsis allows for real-time monitoring of continuous flow and batch applications. Unlike many UV-visible spectrophotometers, the high-resolution Flow-UV CCD array detector does not require calibration or routine servicing. The Xenon flash lamp source used in the Flow-UV has a lifetime of up to 10 years. In a rapidly growing number of pharma labs, the unit is being used for high-throughput screening of UV-active drug substances and real-time determination of the steady-state concentration of products being produced in drug discovery reaction optimization studies. Researchers are also using the Flow-UV for monitoring algal growth, looking at variations of different metal ion complex solutions in rock cores, and optimizing microporous membrane production. To assure linearity of response, up to five wavelengths can be selected to monitor the progress of a reaction or the concentration changes in a wide variety of media.

Uniqsis
For info: +44-845-864-7747
www.uniqsis.com

HPLC Refractive Index Detector

Biotech Fluidics announces REFRACTiMASTER—a new HPLC refractive index (RI) detector suitable for detecting compounds with little or no UV activity including simple sugars, complex carbohydrates, alcohols, fatty acids, and polymers. Designed with a touch panel screen and a highly intuitive user interface, the REFRACTiMASTER is an easy-to-use refractive index detector truly built for modern HPLC analysis. Operating from room temperature up to an unprecedented 80°C, the design of REFRACTiMASTER offers unmatched baseline stability. Unlike traditional refractive index detectors that may take up to 1 hour to be fully stable, the REFRACTiMASTER is ready to deliver reliable, reproducible results in just a couple of minutes.

Biotech Fluidics AB
For info: + 612-703-5718
www.biotechfluidics.com/products/detectors/refractimaster

Manual Pipettes

SARSTEDT offers a comprehensive range of micro and macro single-channel pipettes, as well as 8- and 12-channel options. Sarpette M pipettes are lightweight and well-balanced for comfortable pipetting. Volume setting is fast and reliable with audible click stops. Sarpette M pipettes are fully autoclavable and easily calibrated. Pipetting buttons are color-coded for volume to coordinate with SARSTEDT pipette tip trays. Sarpette M is also compatible with pipette tips from most other manufacturers.

SARSTEDT
For info:800-257-5101
www.shop.sarstedt.us/sarpette

GPC/SEC Combination Detector

TESTA Analytical has developed a new Multiple Angle Light Scattering / Differential Refractive Index (MALS / DRI) combination detector package for more demanding Gel Permeation Chromatography (GPC/SEC) applications. Delivering absolute molecular weight data from small to large molecules, MALS has established itself as a tried and trusted GPC/SEC detection technique for polymer, protein, and polysaccharide applications. However, the most accurate and sensitive results can be only achieved by pairing MALS with a DRI detector, designed to deliver concentration information detected at the very same wavelength. The new COMBO-TWO MALS / DRI detector package from TESTA Analytical brings together a perfectly matched combination of a high-performance seven-angle MALS and an ultra-stable DRI detector—both designed to optimally operate using an integral 640 nm light source. Operating from room temperature up to 80°C, with high thermal stability, COMBO-TWO is simple to use—just inject your sample into the low-volume flow cell to obtain highly accurate and reliable MALS and DRI data.

Testa Analytical Solutions
For info: +49-30-864-24076
www.testa-analytical.com/gpc-sec-chromatography.html

Electronically submit your new product description or product literature information! Go to www.science.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.