Self-Similarity in the Distribution and Abundance of Species

John Harte,[1] Ann Kinzig,[2] Jessica Green[3]

If the fraction of species in area *A* that are also found in one-half of that area is independent of *A*, the distribution of species is self-similar and a number of observed patterns in ecology, including the widely cited species-area relationship connecting species richness to censused area, follow. Self-similarity also leads to a species-abundance distribution, which deviates considerably from the commonly assumed lognormal distribution and predicts considerably more rare species than the latter. Because the abundance distribution is derived under the condition of self-similarity, it may be widely applicable beyond ecology.

Patterns in the distribution and abundance of species within a biome are central concerns in ecology, providing important information about total species richness, the likelihood of species extinction under habitat loss, the design of reserves, and the processes that allow species to coexist and partition resources (*1*). A number of mathematical functions have been suggested as useful for characterizing observed patterns, with perhaps the most widely cited, but by no means the only plausible, ones being the power law form of the species-area relationship (SAR) (*2–4*) and the lognormal species-abundance distribution (*4–6*). The former states that the number of species found in a census patch of area A is a constant power of A: $S = cA^z$; the latter states that the fraction of species with n individuals is a gaussian function of $\log(n)$.

Although available data sets suggest that the lognormal abundance distribution may underestimate the number of rare species in an ecosystem or biome (*1, 4, 7–10*), in general the use of existing data sets to distinguish among candidate functions describing patterns, and therefore among underlying theories that generate these functions, is quite limited by inadequacies in existing data sets stemming from incomplete censusing and other sources of bias (*1, 9, 11*). Because of these empirical limitations, because an effort (*4, 5*) to demonstrate a theoretical connection between the lognormal abundance distribution and the species-area relationship has been questioned on theoretical grounds (*12*), and because establishing mathematical linkages and incompatibilities among patterns may help us understand the mechanisms generating observed patterns, an overarching theoretical framework that unifies our understanding of patterns of species abundance and distribution in ecology is desirable.

Consider area A_0 where there are S_0 species. The number of individuals in each species is described by probability distribution $P_0(n)$, where $P_0(n)S_0$ is the expected number of species with n individuals. For convenience we take A_0 to be a rectangle with a length-to-width ratio of $\sqrt{2}$ so that repetitive bisections perpendicular to the long dimension yield at each stage rectangles of shape similar to the original. We denote by A_i the area of each of the rectangles that are formed at the ith bisection, so that $A_i = A_0/2^i$, and we denote by S_i the number of species found on average in an A_i rectangle (Fig. 1).

We define self-similarity in conformity with the fractal literature (*13*): a pattern is self-similar if it does not vary with spatial scale. We impose self-similarity in the distribution of species by assuming that if a species is known to be in an A_i rectangle, and nothing else about that species (such as its abundance) is known, then the probability that under bisection it will be found in at least a specific one of the two resulting A_{i+1} rectangles is a constant, **a**, that is independent of i. This implies that the fraction of those species found in A_i that are also found in a specific one of the two A_{i+1} is the same constant **a**. The resulting spatial distribution of species is self-similar in the sense that the likelihood of occurrence in a half-patch under bisection is independent of spatial scale (*14*).

If a species is known to exist in patch A_i, there are three mutually exclusive possibilities for its presence or absence in the two A_{i+1} patches that comprise A_i: it is found only in the left half, it is found only in the right half, or it is found in both halves. From the above definition of **a**, the probability associated with each of these options is easily worked out:

$$\text{probability (only in left half)} = \text{probability (not in right half)} = 1 - \mathbf{a} \quad (1)$$

$$\text{probability (only in right half)} = \text{probability (not in left half)} = 1 - \mathbf{a} \quad (2)$$

$$\text{probability (in both halves)} = 1 - \text{probability (not in left half)} - \text{probability (not in right half)} = 1 - 2(1 - \mathbf{a}) = 2\mathbf{a} - 1 \quad (3)$$

Note that the probabilities of these three options sum to 1, as they must. Because the probability a species in A_i is at least in a specific bisection of A_i must be at least 0.5, it follows that $0.5 \leq \mathbf{a} \leq 1$. The extreme values of **a** correspond to the case in which one species is found everywhere ($\mathbf{a} = 1$) and every individual belongs to a unique species ($\mathbf{a} = 0.5$).

By application under repeated bisections of our probability rule, it follows that the average number of species found in any particular A_i rectangle is

$$S_i = \mathbf{a}^i S_0 \quad (4)$$

From Eq. 4 it follows that $S_i/S_j = a^{i-j}$. Now define z by letting

$$\mathbf{a} = 2^{-z} \quad (5)$$

Then $S_i/S_j = 2^{-iz}/2^{-jz}$. However, $A_i/A_j = 2^{-i}/2^{-j}$, so we can write $S_i/S_j = (A_i/A_j)^z$. This is equivalent to $S_i = cA_i^z$, which is just the power law form of the SAR. Thus, we have shown that our self-similarity condition leads to the power law form of the SAR. Elsewhere (*15*) we have shown that the power law form of the SAR implies Eq. 4 and thus self-similarity. Note from Eq. 5 that $0.5 \leq \mathbf{a} \leq 1$ implies $1 \geq z \geq 0$.

Consider, next, the consequence of Eqs. 1 and 2 above, which can be reexpressed as

$$\text{probability (a species found in } A_i \text{ is found only in a particular half of } A_i) = 1 - \mathbf{a} \quad (6)$$

Using Eq. 6 combined with the same reasoning that led to Eq. 4, the average number, E_i, of species found only in a specified A_i rectangle is given by

$$E_i = (1 - \mathbf{a})^i S_0 \quad (7)$$

Defining $z' = -\ln_2(1 - \mathbf{a})$, Eq. 7 is equivalent to $E(A_i)/E(A_j) = (A_i/A_j)^{z'}$ or $E(A) = c'A^{z'}$. This is just the "endemics-area relationship" previously derived by us from the SAR (*15*). We note that $0.5 \leq \mathbf{a} \leq 1$ implies $z' \geq 1$ and that, for the commonly observed value $z = 0.25$, we have $z' = 2.65$.

To derive the distribution $P_0(n)$ of abundances of individuals within species, we introduce the notion of a smallest patch size or unit rectangle within A_0. This area, A_m, contains on average one individual, so that $A_m = A_0/2^m$, where the mean total number of individuals in A_0 is $N_0 = 2^m$. Because the unit rectangle contains one species as well as one individual, $\mathbf{a}^m S_0 = S_m = 1$, or $S_0 = \mathbf{a}^{-m}$. Moreover, using Eq. 5, $S_0 = N_0^z$.

We generalize our definition of P so that

[1]Energy and Resources Group, University of California, Berkeley, CA 94720, USA. [2]Department of Biology, Arizona State University, Tempe, AZ 85287, USA. [3]Department of Nuclear Engineering, University of California, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: jharte@socrates.berkeley.edu

$P_i(n)$ is the probability that if a species is found in a patch of area A_i, then it contains n individuals. Our interest ultimately is in $P_0(n)$, the fraction of species in the entire surface that have n individuals, but to obtain that distribution we derive it recursively from the $P_i(n)$ for $0 < i \leq m$. Note that $P_m(1) = 1$ (there will be on average one individual of whatever species is present in a unit rectangle) and, for each i, $P_i(n) = 0$ for $n > 2^{m-i}$ (on average, one cannot fit more individuals into an area than there are unit patches in that area) and $\Sigma_n P_i(n) = 1$ (the sum of probabilities of all possible occurrences is 1).

Using Eqs. 1 to 3, and letting $2(1 - \mathbf{a}) = \mathbf{x}$, where $0 \leq \mathbf{x} \leq 1$, it can readily be shown (Fig. 1) that the $P_i(n)$ satisfy the following double recursion relation (*16*):

$$P_i(n) = \mathbf{x}P_{i+1}(n) + (1 - \mathbf{x}) \times \sum_{k=1}^{n-1} P_{i+1}(n - k)P_{i+1}(k) \qquad (8)$$

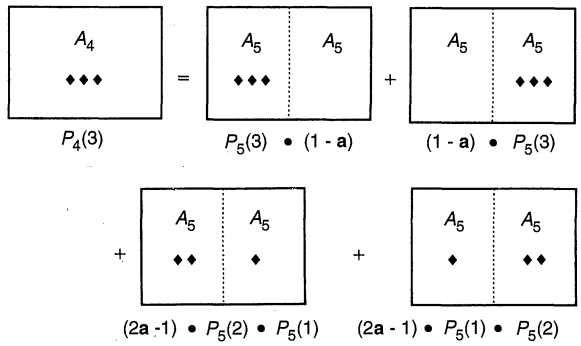

Fig. 1. Origin of the recursion relation for $Pi(n)$. Consider the case $i = 4$ and $n = 3$. Diamond symbols correspond to individuals of a particular species found in a patch. On the left side of the "picture equation" there are three individuals of a particular species in an A_4 rectangle. $P_4(3)$ is the probability that the particular species has exactly three individuals here. The right side of the equation sums the probabilities for all possible ways species can have three individuals in patch A_4. Algebraically, using Eqs. 1 to 3, this equation can be written $P_4(3) = 2(1 - \mathbf{a})P_5(3) + (2\mathbf{a} - 1)[P_5(2)P_5(1) + P_5(1)P_5(2)]$. Denoting $2(1 - \mathbf{a})$ by x, and noting that then $2\mathbf{a} - 1 = 1 - x$, this expression becomes $P_4(3) = xP_5(3) + (1 - x)[2P_5(2)P_5(1)]$. This particular case of Eq. 8 readily generalizes to all i and n. Note that we are assuming, parsimoniously, that the $P_i(n)$ distributions are constructed from independent draws from the ensemble distribution for the $P_{i+1}(n)$. One could, perhaps, construct landscapes based on suitable constrained dependent draws, while still imposing the species-area relationship through the $(1 - \mathbf{a})$ and $2(1 - \mathbf{a})$ probability constraints for pairing empty and occupied halves. Whether abundance distributions that emerge from dependent draws are biologically reasonable and are shaped independent of scale, as are ours (see Figs. 2 and 3), is unclear.

Analytical solutions to Eq. 8 can be derived for the first few values of n (*17*), but we have not been able to derive the general analytical solution for all i, n, and x. Nevertheless, numerical solutions for $P_0(n)$ are revealing. With P plotted against $\log(n)$, these species-abundance distributions are seen to deviate considerably from lognormal, being skewed more toward rarity (more species with low abundance) (Fig. 2).

Plotted on a linear abundance scale, the distributions are more skewed toward commonness than the gaussian but less so than the lognormal. Because the lognormal distribution results from a product of random variables and the normal from a sum, it is not surprising that the distribution resulting from the sum of products in Eq. 8 exhibits intermediate features. Plotted on log-log scales, the $P_0(n)$ are seen to be of the form $P_0(n) \sim n^{c(\mathbf{x})}$ (*18*) for n values sufficiently below the modal abundance, with $c \sim 3/2$, 1, and 3/4 for $\mathbf{x} = 0.26$, 0.376, and 0.484; the exponents $c(\mathbf{x})$ are independent of m, as expected from self-similarity (Fig. 3). The parameter $\mathbf{x}$ in the species-abundance distribution can be related to the SAR parameter z; using the relations $\mathbf{x} = 2(1 - \mathbf{a})$ and $\mathbf{a} = 2^{-z}$, we get $z = -\ln_2[1 - (\mathbf{x}/2)]$. Corresponding to the values $\mathbf{x} = 0.260$, 0.376, and 0.484 in Fig. 2 are the values for the SAR power $z = 0.2$, 0.3, and 0.4.

There is considerable observational support for our self-similarity condition and the abundance distribution it predicts. First, numerous census data sets are compatible with the power-law form of the SAR, as reviewed by Rosenzweig (*3*). Second, the few tests carried out on the endemics-area relationship (Eq. 7) show good agreement (*15, 19*), although considerably more testing is needed. Third, there is considerable evidence (*7, 19, 20*) that the fraction of species in common to two spatially separated censused patches is a decreasing function of interpatch distance ($\propto d^{-2z}$), in conformity with self-similarity (*15*). Fourth, measurements of the dependence of species richness on the shape as well as area of censused patches agree with predictions (*19, 21*). Fifth, Kunin (*22*) presented empirical evidence that the amount of habitat occupied by a given plant species exhibits an approximate scale independence when viewed at different scales of resolution through "censusing windows" of various sizes. Our theory not only predicts this result but also provides an explicit relation between the box-counting fractal dimension implied by Kunin's finding and the abundance of the given species (*23*).

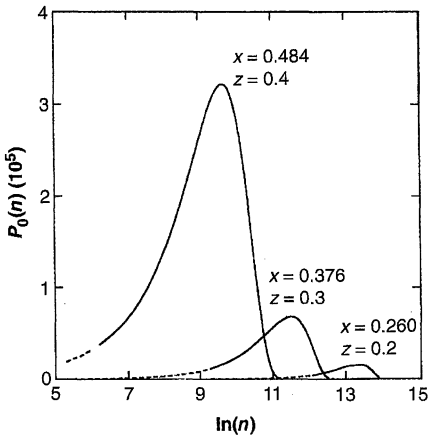

Fig. 2. Species-abundance distributions [$P_0(n)$ versus $\ln(n)$] from Eq. 8 for $m = 24$ and $\mathbf{x} = 0.484$, 0.376, and 0.260 (corresponding to the species-area exponents $z = 0.4$, 0.3, and 0.2, and total species richnesses $S_0 = 772$, 148, and 28, respectively). The total number of individuals in each case is $2^{24} \sim 1.7 \times 10^7$. Dashed portions of distributions correspond to the rarest and most abundant single species.

Sixth, available abundance data, while often qualitative at best because of sampling problems (*9, 11*), generally resemble our predicted distribution more than they do the lognormal, exhibiting considerably more rarity than is predicted by the latter distribution (*1, 4, 8–10*). Important exceptions to this occur, however, indicating that self-similarity and the SAR do not always describe species abundance and distributions (*24*).

Two caveats are in order. It is extremely unlikely that a strictly constant value of z in the SAR holds across an entire accessible scale range (*3*). If, however, z is a nonconstant function of scale area, so that $z = z(i)$, then that dependence can be inserted into Eq. 8 and an abundance distribution can still be derived. The nature of the breakdown of strict self-similarity in small patches—for example, strong attraction ($\mathbf{x} \approx 0$) or repulsion

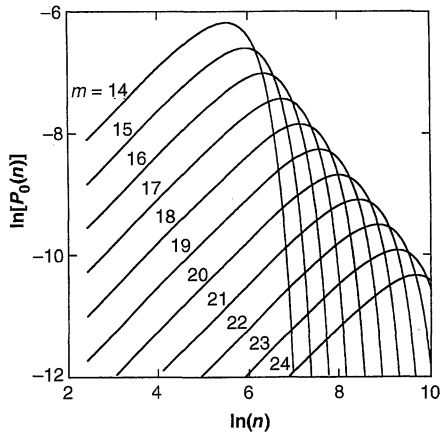

Fig. 3. Species-abundance distributions [$\ln P_0(n)$ versus $\ln(n)$] from Eq. 8 for $\mathbf{x} = 0.484$ and $m = 14$ through 24 (corresponding to a total number of individuals ranging from $\sim 1.6 \times 10^4$ to 1.7×10^7).

($\mathbf{x} \approx 1$) between nearby individuals of the same species—will then influence the shape of the abundance distribution in larger patches in a testable manner. Nevertheless, an abundance distribution skewed toward rarity relative to the lognormal still results as long as the curvature in $z(i)$ is not extreme. Secondly, ecosystems are heterogeneous with respect to habitat quality, and thus quantities like $S(A_i)$ and $P_i(n)$ depend on which patch of area A_i is censused. Moreover, the minimum area per individual (A_m) will differ among species and among individuals in a species and thus can be defined only statistically (especially for motile organisms). Thus all statements we have made about the number of species, or the number of individuals within a particular species, in a patch of area A refer to the average over all the nonoverlapping patches of area A that comprise the system.

We have demonstrated that self-similarity theory provides an overarching framework within which empirically supported patterns in ecology are unified, new and plausible results are derived, and the connection between the SAR and the lognormal abundance distribution is questioned. Because our recursion relation for the species-abundance distribution is derived under the assumption of self-similarity, it may be more widely applicable to other spatial arrays of types of objects or to the distribution of energy fluctuations in turbulent media (*25*).

References and Notes

1. K. J. Gaston, *Rarity* (Chapman & Hall, London, 1994).
2. O. Arhennius, *J. Ecol.* **9**, 95 (1921).
3. M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
4. F. W. Preston, *Ecology* **43**, 185 (1962).
5. R. M. May, in *Ecology and Evolution of Communities*, M. L. Cody and J. M. Diamond, Eds. (Belknap Press, Cambridge, MA, 1975), pp. 81–120.
6. G. Sugihara, *Am. Nat.* **116**, 770 (1980).
7. S. P. Hubbell and R. C. Foster, in *Conservation Biology: The Science of Scarcity and Diversity*, M. Soule, Ed. (Sinauer, Sunderland, MA 1986), pp. 205–231.
8. R. Condit *et al.*, *J. Ecol.* **84**, 549 (1996).
9. F. W. Preston, *Ecology* **29**, 254 (1948); S. Nee, P. H. Harvey, R. M. May, *Proc. R. Soc. London B* **243**, 161 (1991); T. M. Blackburn and K. J. Gaston, *Am. Nat.* **151**, 68 (1998).
10. B. Stone *et al.*, *British Birds* **90**, 1 (1997).
11. R. A. Kempton and L. R. Taylor, *J. Anim. Ecol.* **43**, 381(1974).
12. W. A. Leitner and M. L. Rosenzweig, *Oikos* **79**, 503 (1997).
13. B. Mandlebrot, *The Fractal Geometry of Nature* (Freeman, San Francisco, 1983).
14. Note that we are implicitly assuming symmetry as well as self-similarity, so that the fraction of the species found in A_i that are to be found in at least the left half of A_i and the fraction found in at least the right half are both equal to **a**.
15. J. Harte and A. Kinzig, *Oikos* **80**, 417 (1997).
16. An equivalent expression for the $P_i(n)$ is obtained by writing $Q_i = \Sigma_n q^n P_i(n)$ and noting that Eq. 4 implies the simpler recursion relation $Q_i = \mathbf{x}Q_{i+1} + (1 - \mathbf{x})(Q_{i+1})^2$, where $Q_m = q$. Unfortunately this relation is no more analytically tractable than is Eq. 8.
17. Analytical solutions for the first few values of n include $P_i(1) = \mathbf{x}^{m-i}$, $P_i(2) = \mathbf{x}^{m-i-1}(1 - \mathbf{x}^{m-i})$, and $P_i(3) = 2(\mathbf{x}^{m-i-1})(1 - \mathbf{x}^{m-i})(1 - \mathbf{x}^{m-i-1})/(1 + \mathbf{x})$. Note that when $n = 1$, $i \leq m$; when $n = 2$, $i \leq m - 1$; and when $n = 3$, $i \leq m - 2$. For each value of i, the maximum value of n is $n_{max} = 2^{m-i}$, and for that value the analytical expression $P_i(n_{max}) = (1 - \mathbf{x})^{(n_{max}-1)}$ also follows from Eq. 8.
18. For sufficiently small values of m and of x, the solutions to Eq. 8 exhibit pronounced superimposed oscillations to the left of the modal abundance when $P_0(n)$ is plotted against n, but for $x > 0.3$ and $m > 12$ these oscillations are indiscernible.
19. J. Harte, S. McCarthy, K. Taylor, A. Kinzig, M. Fischer, *Oikos*, in press.
20. M. L. Cody, in *Species Diversity in Ecological Communities*, R. Ricklefs and D. Schluter, Eds. (Univ. of Chicago Press, Chicago, 1993), pp. 147–158.
21. W. E. Kunin, *Biol. Cons.* **82**, 369 (1997).
22. W. E. Kunin, *Science* **281**, 1513 (1998).
23. It can be shown from self-similarity (J. Harte and T. Blackburn, *Am. Nat.*, in press) that a plot of ln(total number of windows of area A occupied by each species) versus ln(abundance of species) yields a straight line result with slope equal to $[\ln(A_0/A)/\ln(N_0)]$, where A is the area of the censusing window, A_0 is the total area censused, and N_0 is the total abundance of all the species plotted within the broader taxonomic category. It also follows that Kunin's ln-ln curves of range versus area of census window have straight line slopes of $1 - [\ln(n_\alpha)/\ln(N_0)]$, where n_α is the abundance of the particular species.
24. I. Hanski, *Oikos* **38**, 210 (1982); I. Hanski, J. Kouki, A. Halkka, in *Historical and Geographical Determinants of Community Diversity*, R. Ricklefs and D. Schluter, Eds. (Univ. of Chicago Press, Chicago, 1993)
25. For turbulent media, it has been shown that a lognormal distribution of energy fluctuations is incompatible with self-similarity [B. Mandelbrot, in *Lecture Notes in Physics 12: Statistical Models and Turbulence*, J. Ehlers K. Munchen, H. Weidenmuller, Eds. (Springer-Verlag, Berlin, 1971), pp. 333–351]. The self-similar distribution in Fig. 3, on the other hand, bears a striking resemblance to empirical distributions of self-similar phenomena [S. T. Bramwell, P. C. W. Holdsworth, J.-F. Pinton, *Nature* **396**, 552 (1998)].

26. Support from the Mellon Foundation and the Class of 1935 Endowed Chair at the University of California, Berkeley, was instrumental in pursuing this work. We thank H. Horn for suggesting the use of golden rectangles; D. Gay for useful discussions about recursion relations; T. Blackburn for numerous helpful conversations and insight into abundance and distribution patterns; and D. Yu, W. Watt, C. Boggs, M. Fischer, W. Kunin, S. Pacala, and S. Saleska for useful comments on the manuscript.

16 October 1998; accepted 2 February 1999

NMDA Receptor–Mediated K^+ Efflux and Neuronal Apoptosis

S. P. Yu,* C.-H. Yeh,* U. Strasser, M. Tian, D. W. Choi†

Neuronal death induced by activating *N*-methyl-D-aspartate (NMDA) receptors has been linked to Ca^{2+} and Na^+ influx through associated channels. Whole-cell recording from cultured mouse cortical neurons revealed a NMDA-evoked outward current, I_{NMDA-K}, carried by K^+ efflux at membrane potentials positive to −86 millivolts. Cortical neurons exposed to NMDA in medium containing reduced Na^+ and Ca^{2+} (as found in ischemic brain tissue) lost substantial intracellular K^+ and underwent apoptosis. Both K^+ loss and apoptosis were attenuated by increasing extracellular K^+, even when voltage-gated Ca^{2+} channels were blocked. Thus NMDA receptor–mediated K^+ efflux may contribute to neuronal apoptosis after brain ischemia.

N-methyl-D-aspartate (NMDA) receptor–gated channels are permeable to Ca^{2+}, Na^+, and K^+ (*1*). NMDA receptor–mediated Na^+ and Ca^{2+} influx participates in synaptic transmission (*2*, *3*) and excitotoxicity (*4*). In contrast, NMDA receptor–mediated K^+ efflux has received little scrutiny, and its functional significance in either normal or abnormal states has not been defined. Stimulating NMDA receptors can induce central neuronal apoptosis (*5*, *6*), and loss of cellular K^+ may be a key step in caspase activation (*7*) and programmed cell death (*8*, *9*). We set out to test the hypothesis that NMDA receptor–mediated K^+ efflux might promote neuronal apoptosis.

To detect K^+ efflux through NMDA receptor channels, the membrane current triggered by NMDA was recorded in mouse cortical neurons by means of patch clamp whole-cell recording in an extracellular solution where Ca^{2+} and Na^+ were replaced by the impermeable cation *N*-methyl-D-glutamine (NMG) (*10*). At the membrane potential of −60 mV, where NMDA normally evokes inward currents in bathing solutions containing physiological concentrations of Na^+ and Ca^{2+}, application of 200 μM NMDA plus 10 μM glycine induced an outward current, designated here as I_{NMDA-K}, of 33 ± 6 pA (mean ± SEM, $n = 11$ cells; Fig. 1). This outward current was enlarged at depolarized membrane potentials, reaching 315 ± 39 pA at 0 mV ($n = 11$). The current-voltage curve of I_{NMDA-K} showed slight outward rectification, and the current reversed at −86 ± 4 mV ($n = 8$), near the calculated K^+ equilibrium potential of −93 mV. The I_{NMDA-K} reversal potential shifted toward more positive potentials when external K^+ was increased from 3 mM to 25 mM (Fig. 1) or when internal K^+ was decreased from 120

Center for the Study of Nervous System Injury and Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: choid@neuro.wustl.edu