



Sequence Data: Posted vs. Published

LARGE-SCALE DNA SEQUENCING PROJECTS

take a considerable amount of time to complete, including 2 to 3 years for the final or "finishing" stage. This fact is not always appreciated by those who are not directly involved in such efforts, as appears to be the case with Elaine Bell in her letter (1 Dec., p. 1696) about the Policy Forum by Lee Rowen *et al.*, "Publication rights in the era of open data release policies" (15 Sept., p. 1881). In her letter, Bell, as editor of *Immunology Today*, discusses what factors were considered in the decision to publish two articles that contained information from publicly available sequence data that had not been previously published. A major factor, according to her, was the length of time that the primary sequence had been available in the public domain" (emphases added). But the time referred to, about a year, is not adequate for such projects given the nature of the work involved.

Large-scale sequencing projects can be divided into three unequal stages: (1) random (shotgun) sequencing (a relatively fast process); (2) assembly of the shotgun data, done many times during the course of the project; and (3) finishing. During this last stage, physical gaps in the sequence are closed, ambiguities in the sequence are resolved, contaminating sequences are removed, and errors in the sequence are identified and corrected. Finishing is a slow process, often taking 2 to 3 years for large sequencing projects. Thus, the almost complete sequence will be available for an extended length of time while the sequence is finished and published.

Posted sequence (from stages 2 and 3), as well as sequence found in the GenBank database, is easily distinguished from pub-

lished sequence. The posted sequence is often incomplete, might contain errors and contamination, and has not gone through peer review. In fact, the high-throughput genome sequence section of GenBank was established precisely to contain sequences not yet sufficiently complete and secure to be published. Thus, posted sequences are public, but they are not thereby automatically in the public domain.

Unpublished sequence should be treated as are all other unpublished scientific data. Therefore, a third party who wants to publish an analysis of other scientists' unpublished sequence should obtain the written consent of those other scientists. Absent that consent, that third party would be committing a "misappropriation of data" as defined by the National Institutes of Health (NIH) (<http://ori.dhhs.gov/html/misconduct/regulation.asp>). As such, misappropriation of data is one of the NIH definitions of plagiarism: "As a general working definition, [Office of Research Integrity] considers plagiarism to include both the theft or misappropriation of intellectual property and the substantial unattributed textual copying of another's work." Plagiarism is one definition of fraud in science.

RICHARD W. HYMAN

Stanford Genome Technology Center (Department of Biochemistry), Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA. E-mail: hyman@sequence.stanford.edu

Atmospheric Ethics

THE IDEA OF EQUITY PRESENTED BY PAUL Baer and coauthors in their Policy Forum "Equity and greenhouse gas responsibility" (29 Sept., p. 2287) with reference to apportioning the use of the atmospheric commons as a gaseous and aerosol waste dump might sound superficially attractive. They suggest that emissions be allocated "based

on equal rights to the atmospheric commons for every individual." The idea of an equal per capita allocation of greenhouse gases, however, is flawed. It implicitly condones global overpopulation and rewards countries in proportion to their level of transgression of the human carrying capacity of their portion of the global biosphere. Per capita allocations are as ethically indefensible and ecologically counterproductive as allocations on the basis of past arrogations, which the authors properly condemn.

I suggest a more environmentally and socially equitable approach, which entails a two-step process (1). First to be determined is a safe (that is, sustainable) level of total global discharge of greenhouse gas emissions. Second to be determined is a country's allocation of that total on the basis of its air space: to wit, of its areal extent as a fraction of the total global land area under national jurisdiction. Most developed countries are probably discharging more than their fair share on this basis, whereas most of the developing countries are discharging less. The latter countries should be permitted to lease (not sell) a portion of their share to the former countries until such time that their industrial and transportation-sector developments require higher levels of discharge. In the meantime, the developed countries would have the time to institute various approaches to meet their fair allotment.

In short, the 5×10^{18} kilograms of atmosphere circulating around the planet must finally become accepted by all as a common heritage of responsible humankind, perhaps through the vehicle of a comprehensive "Law of the Air" comparable with the 1982 Law of the Sea.

ARTHUR H. WESTING

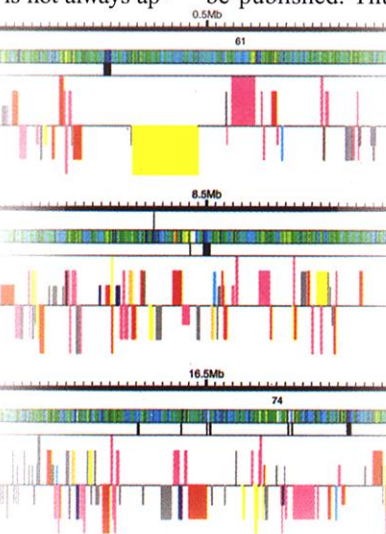
Westing Associates in Environment, Security, and Education, 134 Fred Houghton Road, Putney, VT 05346, USA. E-mail: westing@together.net

References and Notes

1. A. H. Westing, *Environment* 31 (no. 3), 3 (1989); *Sci. World* 34 (no. 4), 5 (1990).

Response

A GLOBAL CAP ON GREENHOUSE GAS EMISSIONS is indeed needed, as Westing says. In our Policy Forum, we advocated an eventual target of under 3 gigatons of carbon per year, consistent with a sustainable oceanic and terrestrial sink and the stabilization of green-



Finishing touches. The final stage required to prepare a large sequence for publication can take several years. A small portion of the *Drosophila* genome is shown here.