BioScience Advance Access published January 22, 2015

It Is Time for IsoBank

JONATHAN N. PAULI, SHAWN A. STEFFAN, AND SETH D. NEWSOME

It was back in 1982, when the United States was mired in the Cold War and a recession, that the National Institutes of Health awarded a 5-year, $3.2 million grant to a group of scientists at the Los Alamos National Laboratory to develop GenBank. It currently houses nearly 200 billion bases from 178 million sequences, representing more than 300,0000 species, and is the largest and most frequently accessed collection of experimental data in the world. Although it is now difficult to imagine a world without GenBank and the opportunities it provides, its creation did not come without questions—about data collection and distribution, attribution of credit and authorship, and the design of infrastructure and access (Strasser 2011). Still, the need for a unifying database of genetic sequences commanded wide agreement.

Now is the time to invest in a parallel special-purpose database for another burgeoning field of research with enormous promise: the use of stable isotopes. In some ways, the history of stable isotopes mirrors that of DNA sequencing. From the 1920s through the early 1950s, chemists were busy identifying these isotopes—nondecaying variants of an element in which the atoms differ in mass because of different numbers of neutrons—and determining the ratio of heavier to lighter isotopes in nature (e.g., the ratio of ^{18}O to ^{17}O). Isotopic ratios are often idiosyncratic, encoding information about the origin and history of particular matter, whether it be an ancient ice core, a Martian rock, or a fossil. In biological systems, elements such as hydrogen, carbon, nitrogen, and oxygen are light enough for isotopes to be differentially affected by biochemical processes such as photosynthesis and respiration, creating isotopic variation. This is extraordinarily useful for studying biological processes. For example, stable isotopes are routinely used for studies of animal and plant physiology, community and ecosystem ecology, biomedicine, paleobiology, evolution, and climatology. The diversity of questions that can be answered with stable isotopes, coupled with advancements in instrumentation that now enable the analysis of hundreds of samples per week on a single mass spectrometer, is what makes the growth of this field parallel to that of DNA sequencing (figure 1).

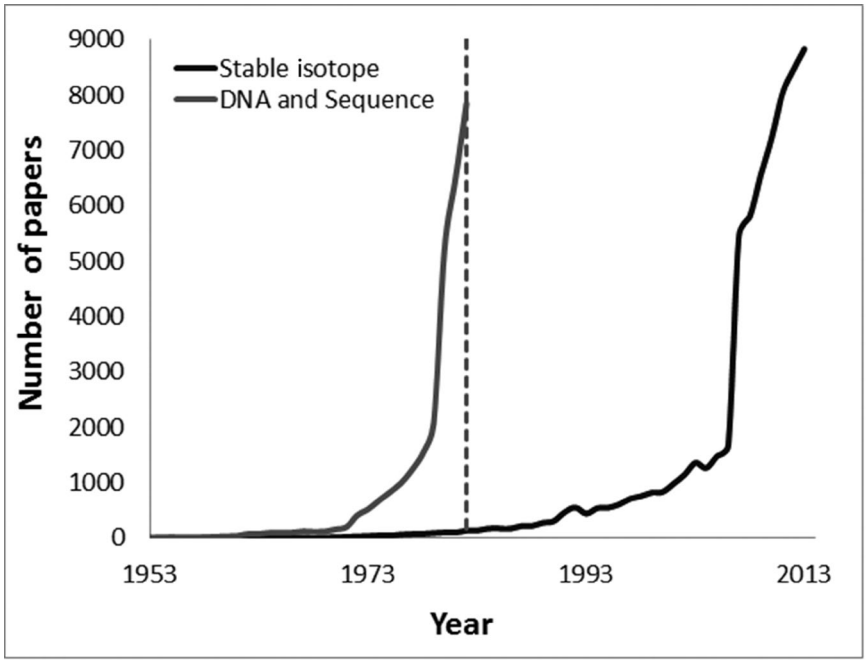

Figure 1. The number of papers published each year (obtained from the Thomson Reuters Web of Science in October 2014) using the search phrases "DNA" AND "sequence" (the first search phrase) and "stable isotope" (the second) beginning in 1953, with the discovery the structure of DNA, through 2013. The vertical reference line denotes the year 1982, when GenBank was created.

As with gene sequence data in 1981, stable isotopic data do not yet have a centralized database. Arguments analogous to those made for the creation of GenBank apply. It was true of gene sequence data in 1982—and is true of stable isotope data now—that the limits to progress are not in data acquisition but in data exchange. Such inefficiencies have led to unnecessary duplication of effort and redundancy in data creation. They have stymied progress as researchers have spent more time than is necessary generating pilot isotopic data to explore the feasibility of techniques. We have now reached a point at which, as in 1982, the first centralized database for the data—a so-called *IsoBank*—is needed. It will allow researchers to spend more time generating and testing hypotheses.

In the current era of "big data," many scientists are organizing to make the flood of information, which

is being generated at an increasing rate, more accessible. These researchers recognize that a well-developed database is not a silo or a museum but can be harnessed as a powerful motor to generate new knowledge. So what is holding back the development of IsoBank? Foremost are the challenges of constructing and staffing of a facility to organize, store, and vet data. All databases are faced with these tasks, however, and they can usually be dealt with by adequate funding. This would be used to equip IsoBank with space, equipment, and personnel. Alternatively, the work could be accomplished through partnerships with existing museum databases.

The second major challenge relates to data acquisition. Again, however, this is not a novel concern. GenBank provides a successful model. It relies on individual labs to furnish data and on journals to reinforce good behavior by requiring authors to submit data following publication. Finally, and arguably more important for isotopic data than for sequence data, are the necessary metadata. Unlike DNA sequences, which require limited metadata (typically, information such as the species or type and date of specimen), isotope data generally require, in addition, georeferencing, details on the analytical facility where the isotopes were measured—including standards and instrumentation used—the date of collection, sample preparation information (e.g., lipid extraction, decalcification), and finer-scale environmental information. This might include its position in the water column (for aquatic samples); its height in canopy (for plants); or the sex, age and size class (for animals).

The challenges are all important but surmountable, and most are not novel or unique to isotopic data. They do need to be discussed. With this Viewpoint, we hope to launch the necessary discussion for IsoBank, and we echo an anonymous lament at the slow pace of the creation of GenBank nearly 25 years ago, "It matters little by whom, or where, a bank is started. What is important is that it should start soon" (Anonymous 1980).

References cited

Anonymous. 1980. Banking DNA sequences. Nature 285: 59.

Strasser BJ. 2011. The experimenter's museum: GenBank, natural history, and the moral economies of biomedicine. Isis 102: 60–96.

Jonathan N. Pauli (jnpauli@wisc.edu) and Shawn A. Steffan are assistant professors at the University of Wisconsin–Madison. Seth D. Newsome is an assistant professor at the University of New Mexico, in Albuquerque.

doi:10.1093/biosci/biu230