A genome to celebrate

This week, Science celebrates the impending 20th anniversary of the publication of the draft human genome sequence—a landmark achievement by any measure. The American Association for the Advancement of Science (AAAS, the publisher of Science) also looks forward to next week’s annual meeting, whose theme is “Understanding Dynamic Ecosystems.” At first glance, these two events may seem unrelated. But the successful completion of the human genome sequence ushered in biology’s era of “big science” and created a research ecosystem for tackling complex, technology-driven, and data-intensive multidisciplinary projects that continue to improve our understanding of cancer, the microbiome, the brain, and other areas of biology.

The Human Genome Project (HGP) was an internationally supported public project (Cephalon Genomics was the private effort that simultaneously sequenced the human genome). When the endeavor was launched in 1990, collaboration among a diverse group of scientists was essential because the sequencing was distributed across a number of international research sites. High-throughput technologies for DNA sequencing were critical to the project’s success, and the participation of biotech companies in the effort was instrumental in driving down the cost, speed, and throughput of generating DNA sequence. The ever-increasing amount of sequence data drove the development of mathematical and computational tools for assembling and annotating the data. Neither the laboratory scientists nor the computational scientists could have done this alone, and the convergence of these disciplines has been one of the most important legacies of the early genome efforts. There was also a commitment to train the next generation of genome scientists, and over the past 20 years, many colleges and universities have established new undergraduate and graduate programs in quantitative and systems biology. Life sciences students today graduate with a very different set of skills than they did in 2000.

The topic of ongoing data availability was addressed throughout the sequencing endeavor. The Bermuda accord in 1996 and the Fort Lauderdale agreement in 2003 were crafted with the idea that these types of large, data-intensive projects should be community resources and the data generated should be freely available to all. This notion ran counter to the tradition that scientific data were made available at the time of peer-reviewed publication, and the idea was initially met with some skepticism. Today, through U.S. National Center for Biotechnology Information and affiliated databases, users can query genome, epigenome, transcriptome, metagenome, and metatranscriptome datasets from thousands of species. Despite concerns that the HGP would divert scarce resources away from individual investigators, and a view that the work was monotonous and repetitive, the availability of genome sequences from hundreds of thousands of microbes, plants, and animals has transformed the biological sciences. These public resources provide a robust foundation for biomedical research and discovery, and scientists with internet access can participate in genome analysis work without having to generate sequence data themselves. Indeed, the more these large datasets are interrogated by investigators with diverse perspectives, the greater the number of insights that will emerge.

At its inception, efforts to sequence the human genome effort raised questions about the potential ethical, legal, and social issues that would arise as we deciphered our own genetic blueprint. These concerns ranged from privacy, discrimination, and ownership of data to the concept of “self.” Passage of the Genetic Information Nondiscrimination Act (GINA) by the U.S. Congress in 2008 ensured that genetic information would not be used in health insurance and employment decisions. Senator Edward Kennedy called GINA “the first major new civil rights bill of the new century.”

The completion of the draft sequence laid the foundation for a new precision medicine paradigm that aims to use a person’s unique genetic profile to guide decisions about the treatment and prevention of disease. We have already seen some signs that precision medicine is possible, and although off to a slow start, the promise of this approach may ultimately be realized.

Given the pace at which breakthroughs based on the human genome sequence are happening, when we next commemorate the publication of the draft human genome sequence, be it at 25, 30, or 50 years, we may look back again, realize that this accomplishment was a watershed for the biological sciences, and marvel at how far we have come in such a short period of time.

—Claire M. Fraser
A genome to celebrate
Claire M. Fraser

Science 371 (6529), 545.
DOI: 10.1126/science.abg8615