

Building a genomic encyclopedia of life

The total number of species on Earth is estimated at 100,000 to 10 million. Despite the wide variation in estimates here, one thing scientists agree on is that it is a Very Big Number. Of the 100,000 to 10 million species that exist, a mere 19,000 or so have complete genomes sequenced.

On the one hand, given the relative novelty of sequencing technology, 19,000 is pretty impressive. On the other, it pales in comparison to the amount of sequencing needed if we're to assemble a comprehensive, genomic version of the "Tree of Life".

[The Scientist's Anna Azvolinsky](#) reports:

Large-scale sequencing consortia have been churning out data at an impressive rate, yet significant gaps remain in the genomic tree of life. And while these groups have largely been working independent of one another, together they might address more far-reaching questions, such as how life has evolved, how it currently functions, and how it might look down the line.

We're at a turning point, Azvolinsky argues, where a truly cooperative approach to genomics will make or break efforts to take the independent work of many consortia and create a global database.

One of the biggest problems here isn't necessarily the independent, competitive approach of genomics groups thus far, but the independent, competitive nature of funding for genomics research. She writes:

"We thought there was a turning point three years ago," said Nikos Kyrpides, who heads up the microbial genomics and metagenomics program at the JGI. "The community believed that more funding agencies would begin to support microbial sequencing studies, not just for public health and industry applications, but to cover the reference genomes of the phylogenetic tree."

Unfortunately, this turning point failed to materialize.

Part of the problem is that many government and private funding agencies are most interested in supporting scientists sequencing the genomes of species that impact human health, industry concerns, and environmental issues.

"Many times it is easier to receive funding to sequence species important for agriculture, for example. These types of projects go faster through the pipeline because there is more funding from governments or companies interesting in funding directed efforts," said [Toni Gabaldon](#), the head of bioinformatics and genomics at the Center for Genomic Regulation in Barcelona, Spain. One issue, said Kyrpides, is that funding agencies don't often work together, and it

typically takes more than a single funding body to support broad, encyclopedic sequencing efforts. “We are pushing for funding agencies to change: to stop delineating projects by application, and to work together.”

Other challenges to a genomic encyclopedia of life include the sheer volume of genetic data that needs to be properly analyzed and sorted. As such, bioinformatics and computational science are increasingly vital disciplines alongside traditional biology, taxonomy, and genetics. “Even within a single consortium’s database,” Azvolinsky writes, “as the number of genomic sequences increases from tens to many hundreds, scaling the storage and analytical tools has been a challenge.”

New groups are appearing whose goal isn’t to add to the grand sequencing efforts but to help coordinate the sequencing that already exists. One such group is iPlant, which is focused on making “cloud computing, data storage, and genomic analysis tools available to the broader plant community.”

One aspect that Azvolinsky doesn’t get into is how [genomics is upending traditional taxonomy](#), in many cases revealing several species where once we saw one. Her article is based on the best understanding of the tree of life currently available, but it’s certain that our map of the tree will change as we incorporate more and more genomic data of ever-increasing sophistication.

Co-operation is going to be necessary, even if — according to scientist behind iPlant — it’s only just starting to evolve within the genomics community. Not only are there at least 81,000 sequences left to collect, analyze, and map against our current understanding of the tree of life, but in doing so biologists will likely need to redraw significant portions of the tree.

Read Anna Azvolinsky’s original article at The Scientist: “[Sequencing the Tree of Life](#)”

Kenrick Vezina is Gene-ius Editor for the Genetic Literacy Project and a freelance science writer, educator, and naturalist based in the Greater Boston area.

Additional Resources:

- “[Modern genetics undermines neat boundaries of species](#),” Kenrick Vezina | Genetic Literacy Project
- “[Mapping the Tree of Life: Progress and Prospects](#),” Norman R. Pace | American Society for Microbiology
- “[It’s time for a cancer genomics revolution](#),” Kenrick Vezina | Genetic Literacy Project