

[Sign In](#) or [Register](#)

[News](#) [Magazine](#) [Multimedia](#) [Subjects](#) [Surveys](#) [Careers](#)

Search



[The Scientist](#) » [News & Opinion](#) » [News Analysis](#)

Sequencing the Tree of Life

Charting the progress of the various large-scale genome-sequencing projects as researchers working separately on their chosen species begin to pool analytical resources

By Anna Azvolinsky | April 24, 2014

7 Comments

Like 89

[g+](#) 9

[Link this](#)

[Stumble](#)

[Tweet this](#)



Gustav Klimt's "The Tree of Life," 1909
WIKIMEDIA

Scientists working to sequence all manner of bacteria, Archaea, plants, and animals and to make these genomes publicly available hope to use the data to inform health, industrial, and environmental issues. 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 are still in the developmental stage, where every consortium focuses on a specific domain and is building up their own data and making sure it's in good enough shape," said Igor Grigoriev, head of the fungal genomics program at the US Department of Energy (DOE) Joint Genome Institute (JGI) in Walnut Creek, California, and part of the 1,000 Fungal Genomes project. "Some dialog between the consortia is happening but grand-scale data integration remains to happen."

Although there is still relatively little crosstalk among consortia, some of their data are being collected in central repositories. Aside from the National Center for Biotechnology Information's genome database, there is the JGI-funded Genomes Online Database (GOLD), which functions as a hub for completed and ongoing genome sequencing initiatives and metagenome projects. GOLD is mainly focused on microbial genomes, but includes some eukaryotic genomes. Data from many of these projects are integrated in to JGI's databases and can also be uploaded into newly developed KnowledgeBase tools funded by the DOE.

"Both talking across communities and coming up with creative tools to ask broader scientific questions across the domains of life is important," said Grigoriev. "As a scientific community, I think we are just now at the moment . . . moving towards this."



The genome of cultivated sorghum (*Sorghum bicolor*) was published in August 2013.
FLICKR, KAY LEDBETTER

Amidst this early collective momentum, however, some groups are still working to sequence critical species within their own domains. In 2011, members of the microbial genomics community were ready to publish a manuscript, rallying the scientific community to fund a large-scale genomic sequencing project covering important bacterial and archaea strains. The goal was to fill gaps in the microbial tree of life. The decree proved unnecessary, however, as progress in genome sequencing initiatives—including those on thousands of bacterial and archaea genomes, such as the Genomic Encyclopedia of Bacteria and Archaea (GEBA) pilot project—gave the geneticists and microbial biologists reason to believe that the sequencing would be completed.

"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."

Follow The Scientist



Subscribe!

Print or Digital

- iPad
- Kindle
- Tablet

Stay Connected with The Scientist

- [The Scientist Magazine](#)
- [The Scientist Careers](#)
- [Neuroscience Research Techniques](#)
- [Genetic Research Techniques](#)
- [Cell Culture Techniques](#)
- [Microbiology and Immunology](#)
- [Cancer Research and Technology](#)
- [Stem Cell and Regenerative Science](#)

Popular Posts

- [1. The Surprising Evolution of Sex Determination](#)
- [2. Does Brain Training Work?](#)
- [3. Sequencing the Tree of Life](#)
- [4. Is Cannabis Really That Bad?](#)
- [5. Updated Review: Tamiflu Is a Bust](#)

Current Issue



[View the April 2014 contents.](#)

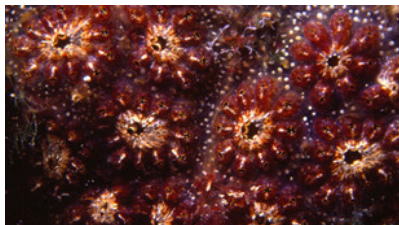
Subscribe to RSS feed

- [All](#)
- [The Nutshell](#)
- [News & Opinion](#)
- [Careers](#)

Investigators at the JGI and several international institutions have since sequenced the full genomes of 3,000 additional microbes, but coverage of the bacterial and archaea domains remains fairly sparse. So Kyrpides and his colleagues are now submitting an updated manuscript to raise awareness of the importance of their Microbial Earth Project, which aims to sequence 7,830 representative type strains from the 11,000 species available in culture collections over the next three years. "Only about 10 percent to 15 percent of the diversity of cultured Archaea and bacterial species has been captured by sequencing so far," said Kyrpides. "That's enormously small."

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."



The genome of golden star tunicate (*Botryllus schlosseri*) was published in July 2013.

WIKIMEDIA, PARENT GERY

Plenty of consortia dedicated to sequencing specific branches of the tree of life have cropped up as researchers working within the same domains have recognized that pooling resources can boost scientific progress. Among these groups are the Global Invertebrates Alliance (GIGA), the 5,000 Insect Genome Project (i5K), the 1,000 Fungal Genomes Project, the US National Science Foundation (NSF) Plant Genome Research Program, and the Genome 10K Project, which aims to sequence 10,000 vertebrate genomes. There is also the Smithsonian Institution-led Global Genome Initiative (GGI)—a collaborative effort to sequence at least one species from every one of the 9,500 described invertebrate, vertebrate, and plant families.

As more and more long-read sequencing technologies hit the market and the overall costs of decoding genomes drop, an emerging challenge is attracting and coordinating experts to collect, annotate, and place sequencing data in their biological contexts, according to [Kevin Hackett](#), a national program leader at the US Department of Agriculture (USDA) and one of the leaders of the i5K project.

And these analytical efforts are important; rather than having researchers compete for funding, they unite those with common goals, eliminating redundancies and lowering overall costs. According to [Stephen Goff](#), the project director of the iPlant Collaborative, a culture of true cooperation in genomics is just beginning to evolve.

For its part, rather than generating new genomic sequencing data, the iPlant team is making cloud computing, data storage, and genomic analysis tools available to the broader plant community. For example, iPlant is providing the cyberinfrastructure and analysis tools that will help the African Crops Consortium sequence 101 crops important to the continent's agriculture. iPlant has also volunteered to provide infrastructure for the i5K project and other insect sequencing projects, said Goff.



The comb jelly (*Mnemiopsis leidyi*) genome was published in December 2012.

FLICKR, VIDAR-AQUA-PHOTOS

Other consortia are creating their own data storage and analysis tools. Through its Plant Genome Research Program, the NSF aims not only to generate new genome sequences, but to provide a platform to integrate all existing genomic data for evolutionary and species diversity analyses. In Europe, the members of the European Life-Sciences Infrastructure for Biological Information (ELIXIR) group intend to create a resource for scientists to store and share large data sets, such as whole genomes.

Grigoriev's team at JGI has developed a web-based public fungal genomics resource. "MycCosm is an example of integration of fungal genomics data and

computational tools, and the bringing together of the fungal biologist research community," he explained. "From here, we can go to the next step of integrating across multiple domains."

Bioinformatics tools will need to evolve to keep pace as genomic analyses become more complicated—covering complex inter-domain relationships, such as the symbiotic interplay between certain plants, fungi, and endobacteria. But even within a single consortium's database, as the number of genomic sequences increases from tens to many hundreds, scaling the storage and analytical tools has been a challenge.

"Many computational scientists and bioinformaticians are working alongside biologists to analyze and organize the sequencing data. This is a major challenge but I have a lot of optimism because there is plenty of innovation and energy in this field," said [Klaus-Peter Koepfli](#), one of the principle investigators of the Genome 10K project and visiting scientist at the Smithsonian Conservation Biology Institute in Washington, D.C. "There are many obstacles to reconstructing the phylogeny of all living things, but it's a great goal."

How Many Species Have Been Sequenced?

During the last 250 years, 1.2 million eukaryotic species have been identified and taxonomically classified. Number of species estimated to exist on Earth: bacterial and archaea species, from 100,000 to 10 million^{1,2}; eukaryotic species, approximately 8.7 million (including 2.2 million marine

organisms; ± 1.3 million, total) ¹ .				
	TOTAL NUMBER OF SPECIES (ESTIMATED)	NUMBER OF SPECIES IDENTIFIED/ DESCRIBED (APPROXIMATE)	NUMBER OF SPECIES WITH COMPLETE GENOMES SEQUENCES* (ESTIMATED)	REFERENCES
BACTERIA, ARCHAEA	100,000 to 10 million	12,000 (460 cultured Archaea)	17,420 bacteria, 362 Archaea genomes	GOLD database; <i>Nucleic Acids Research</i> (January 2012); World Data Centre for Microorganisms
FUNGI	1.5 million	100,000	356	JGI
INSECTS	10 million	1 million	98	<i>Pest Management Science</i> (May 2007); NCBI; R.G. Footitt and P.H. Adler, <i>Insect Biodiversity: Science and Society</i> (2009)
PLANTS	435,000 (land plants and green algae)	300,000	150	Botanic Gardens Conservation International; The Plant List ; W.S. Judd et al., <i>Plant systematics—a phylogenetic approach</i> (2008); D. Bramwell, <i>Plant Talk</i> , 28:32–4 (2002)
TERRESTRIAL VERTEBRATES, FISH	80,500 (5,500 mammalian)	62,345 (5,487 mammalian)	235 (80 mammalian)	Genome 10K; Zoological Society of London (2012); <i>Science</i> (December 2010); Australian Biodiversity Information Services (2009)
MARINE INVERTEBRATES	6.5 million	1.3 million	60	<i>Journal of Heredity</i> (May 2014); <i>Current Biology</i> (December 2012); Zoological Society of London (2012); <i>Oceanis</i> 19:5-24 (1993)
OTHER INVERTEBRATES	1 million nematode, several thousand <i>Drosophila</i>	23,000 nematode, 1,300 <i>Drosophila</i>	17 nematode, 21 <i>Drosophila</i>	M. Blaxter et al., "The evolution of parasitism in Nematoda," <i>Parasitology</i> (in press); Texas A&M University ; <i>Molecular Biology and Evolution</i> (1995)
<p>¹ <i>PNAS</i> (August 2002)</p> <p>² Bergy's International Society for Microbial Systematics (2014)</p> <p>³ <i>PLOS Biology</i> (August 2011)</p> <p>*as of April 24, 2014</p>				

Tags

[whole genome sequencing](#), [vertebrates](#), [tree of life](#), [research funding](#), [nonmodel organisms](#),

[invertebrates](#), [genome sequence](#), [genetics & genomics](#), [fungi](#), [data storage](#), [data analysis](#), [collaboration](#), [bacteria](#) and [Archaea](#)

7 Comments

Like 89

g+1 9

Link this

Stumble

Tweet this

Add a Comment



You

Sign In with your LabX Media Group Passport to leave a comment

Not a member? [Register Now!](#)

Comments



Neddy
Posts: 1

April 25, 2014

Doesn't the ToL hypothesis have a lot of conflicting difficulties?

[Sign in to Report](#)



James V. Kohl
Posts: 113

April 25, 2014

Unless there is experimental evidence that links mutations and natural selection to evolution and increasing organismal complexity, I think what we are seeing across genera is that ecological variation results in ecological adaptations via conserved molecular mechanisms in species from microbes to man.

The fact that the conserved molecular mechanisms are nutrient-dependent makes sense in the context of Darwin's 'conditions of life.' The fact that the nutrients metabolize and become species-specific pheromones that control the physiology of reproduction in species from microbes to man also makes sense in the context of nutrient-dependent species diversity.

What doesn't make sense in the light of biology (e.g. what is currently understood about molecular biology), is any theory of evolution that excludes ecological variation and/or substitutes it with constraint-breaking mutations that somehow result in natural selection for something that somehow alters species diversity.

In the real world, as detailed in the linked article below, serious scientists do not simply dismiss physics, chemistry, and molecular biology. They incorporate biophysical constraints on ecological adaptations in their explanations of cause and effect.

[Nutrient-dependent pheromone-controlled ecological adaptations: from atoms to ecosystems](#)

[Sign in to Report](#)



irub
Posts: 1

April 25, 2014

I would suggest reading "Icons of Evolution" by J. Wells.

[Sign in to Report](#)



Alexandru
Posts: 55

April 25, 2014

I am engineer and my friend Dan Tudor is psychologist, and according to a comparative study of genetics and Genesis, we developed a new Theory of Assisted Evolution.

Mitochondrial Adam DNA data transmission theory - ISBN 978-606-92107-1-0:

The origin of species in the third millennium expectation - ISBN 978-606-92107-9-6

The main problem for the natural Evolution is the impossibility to pass from sexless reproduction to bisexual reproduction. It was necessary an external intervention.

Please, just look the shortest Evolution description in the world: "... there before me were the four winds of heaven churning up the great sea. Four great beasts, each different from the others, came up out of the sea. The first was like a lion with the wings of an eagle. As I watched, its wings were pulled off. Then it was lifted to an upright position and made to stand on two feet, just like a human, and it was given a human heart." (Daniel 7.2-4)

As Winston Churchill once said about courage, "*I stood up and expressed my point of view and now I sit down and listen to the opinions of others involved in this dialogue.*"

"*We honour God for what He conceals; we honour kings (GENETICISTS) for what they explain!*" (Proverbs 25.2)

[Sign in to Report](#)



John AB
Posts: 2

April 25, 2014

Several protists (e.g. the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*) have also been sequenced. Where do they fit in this Table??

[Sign in to Report](#)



James V. Kohl
Posts: 113

April 26, 2014

[Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors](#) was reported as [Many genes are switched on by default](#). The revelation in the following quote is a theory-killer.

"Yeast may seem far removed from humans, but its genes are controlled in exactly the same way as in human cells."

I hope to help kill the pseudoscientific theories associated with mutations, natural selection, and evolution. In our 1996 [Hormones and Behavior](#) review: [From Fertilization to Adult Sexual Behavior](#) we linked the conserved molecular mechanisms of cell type differentiation in yeast to sex differences in human morphology and behavior. Since then, the model of nutrient-dependent DNA methylation, alternative splicings of pre-mRNA and de novo creation of olfactory receptor genes has been extended across species to the differentiation of cell types in individuals and species via the metabolism of nutrients to species-specific pheromones that control the physiology of reproduction and nutrient-dependent species diversity.

[Signaling crosstalk: integrating nutrient availability and sex.](#) (yeasts)

[Feedback loops link odor and pheromone signaling with reproduction.](#) (mammals)

It will be interesting to see how much longer it takes for people to look at the extant literature -- even if they only look at the following published works and "Related Citations in PubMed"

[Human pheromones: integrating neuroendocrinology and ethology.](#)

[Human pheromones and food odors: epigenetic influences on the socioaffective nature of evolved behaviors.](#)

[Nutrient-dependent/pheromone-controlled adaptive evolution: a model.](#)

A pattern has emerged that links nutrient-dependent pheromone-controlled cell type differentiation via changes in the microRNA/messenger RNA balance and conserved molecular mechanisms in species from microbes, like yeast, to humans. The pattern includes ecological variation and excludes constraint-breaking mutations. It replaces theory with facts about biophysically-constrained cell type differentiation in accord with what is currently known about physics, chemistry, molecular biology and how ecological variation results in ecological adaptations.

If the pattern is not a theory killer, minimally it may force evolutionary theorists to support their claims with experimental evidence of biologically-based cause and effect that links species from microbes to man via the conserved molecular mechanisms that link the epigenetic landscape to the physical landscape of DNA in organized genomes.

[Sign in to Report](#)



kman23
Posts: 1

April 27, 2014

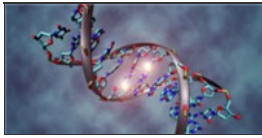
This is interesting and I wonder how fast sequencing like this grow into a more citizen based system. I have already seen people preparing to sequence and annotate genomes on crowdfunding websites such as this one for the Blue Feigning Death Beetle.

<http://rockethub.com/projects/41273-blue-beetle-genome-project>

Apparently given the much lower pricing involved this can truly become a new area of citizen science.

[Sign in to Report](#)

Related Articles



Genetic Brain Disorder Explained

By Rina Shaikh-Lesko

Researchers uncover a mutation responsible for a rare neurological condition.



Week in Review: April 21–25

By Tracy Vence

Evolution of Y chromosome; delivering gene with "bionic ears"; diversity of an important cyanobacterium; charting genome-sequencing progress; blockbuster pharma deals



Genome Digest

By Tracy Vence

What researchers are learning as they sequence, map, and decode species' genomes

TheScientist

[Home](#) [News & Opinion](#) [The Nutshell](#) [Multimedia](#) [Magazine](#) [Advertise](#)
[About & Contact](#) [Privacy Policy](#) [Job Listings](#) [Subscribe](#) [Archive](#)



© 1986-2014 The Scientist

Now Part of the LabX Media Group: [Lab Manager Magazine](#) | [LabX](#) | [LabWrench](#)