

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

How many genomes are enough?

ArticleInfo		
ArticleID	:	4884
ArticleDOI	:	10.1186/gb-spotlight-20031118-01
ArticleCitationID	:	spotlight-20031118-01
ArticleSequenceNumber	:	236
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	5
ArticleHistory	:	RegistrationDate : 2003-11-18 OnlineDate : 2003-11-18
ArticleCopyright	:	BioMed Central Ltd2003
ArticleGrants	:	
ArticleContext	:	130594411

Tabitha M Powledge

Email: tam@nasw.org

How many sequenced genomes are enough? The minimum number for [comparative genomics](#), researchers say, depends on what you want to learn. The optimum number is still a mystery.

For identifying cis-regulatory regions such as enhancers and promoters, the genomes of three species that are roughly equidistant evolutionarily is the bare minimum, and more is better, according to Lincoln Stein. Stein, who is at [Cold Spring Harbor Laboratory](#), is first author of a paper on the *Caenorhabditis briggsae* genome in the first issue of [PLoS Biology](#), an open-access online journal published by the Public Library of Science. He pointed out that having the genomes of at least three species permits distinguishing between the signal-bases conserved because they actually do something - and noise-bases conserved simply because they haven't mutated yet. "As you add more species to the alignment, the signal remains the same while the noise decreases because there are fewer and fewer bases that are the same because of an accident of history," he told us.

For finding conserved sequences in mammals, with their more complicated genomes and long stretches of repetitive DNA, some half-dozen species might be enough, said Eric Green, scientific director at the [National Human Genome Research Institute](#) (NHGRI). His estimate comes from his lab's unpublished work and from a paper in the August 14 [Nature](#) comparing a single large genomic region in 13 vertebrate species.

"By the time you get to about five or six mammals, you start to plateau in the detection of these highly conserved sequences," Green said. "But," he cautioned, "we don't know if we've found everything we need to find, and we don't know if our algorithms are properly developed to find everything we're trying to find. It just means that with our algorithms, you plateau. In the absence of knowing what it is that you're really trying to find, it's hard to assess the methods and the datasets that you're using to find it."

Studies of evolutionary change require another comparative approach. *Caenorhabditis elegans* and *C. briggsae* differ in their ability to respond to RNAi, Stein pointed out. Did *C. elegans* acquire the ability or did *C. briggsae* lose it? To determine the direction of change, worm researchers need the genome of a third, somewhat distant species. If it has the *elegans* form of the hypothetical systemic RNAi gene, then the common ancestor of *C. elegans* and *C. briggsae* probably possessed that form, and it was lost in the *C. briggsae* lineage, Stein noted. If the third species has the *briggsae* form, then *elegans* probably learned a new trick in the hundred million years or so that separates them.

That's why the *C. briggsae* paper urges sequencing of *Caenorhabditis remanei*, a genome project that NHGRI has already accorded [moderate priority](#). John Spieth, of the [Genome Sequencing Center](#) at Washington University School of Medicine in St. Louis, who is first author of the *remanei* 'white paper' (proposal), says he and his coauthors are thinking about rewriting the paper in an attempt to persuade NHGRI to bump *remanei* up to high priority. That strategy has already worked for partisans of the rhesus macaque, which made the A list at the [second attempt](#).

The *C. briggsae* paper also suggests sequencing the more distant nematodes *Caenorhabditis japonica*, *Caenorhabditis drosophila*, or *Brugia malayi*. A *B. malayi* project is underway at [The Institute for Genomic Research](#), said Avril Coghlan, a comparative genomicist at the [Smurfit Institute](#) of Trinity College Dublin. "It will be possible to compare *B. malayi* to *C. elegans* and *C. briggsae* across their

entire genomes. This three-genome data set will be a treasure trove to nematode geneticists and molecular evolutionists."

Will three genomes really be enough for evolution studies? "If you're trying to understand details about genome evolution, there's probably never enough genomes. Every genome you get more data from, you're getting insight about the evolution of that genome relative to all other genomes," Green said. Green has put his resources where his mouth is. The *Nature* paper presented comparisons of a single swath of DNA from a mere 13 far-flung vertebrates. But his lab is beavering away on more than 30 - a list that does not yet include beavers, although there is a hedgehog, along with marsupials, bats, and several primates.

"I'd say we're just scraping the surface right now," noted Hugh Robertson, who studies the evolution of insect transposons at the [University of Illinois at Urbana-Champaign](#). He forecasts eventual genome projects on several insect orders, including beetles, moths, and bugs (Hemiptera). To say nothing of fruit flies: *Drosophila simulans* and *Drosophila yakuba* have already been added to NHGRI's high-priority roster, and a white paper in preparation urges an additional eight *Drosophila* species. "The bacterial people are already way ahead of this. They're doing tons of comparative genomics," Robertson points out. "Look at what the yeast people learned when they [sequenced five yeast species](#) closely related to *Saccharomyces cerevisiae*. And that was a relatively cheap project."

The main barrier to the immediate sequencing of many more genomes remains cost. "The reason there is so much discussion about which vertebrates to sequence is simply because it still costs between \$50 million and \$100 million to sequence a vertebrate genome, and that's real money," Green pointed out. "There's no question that if the cost of a sequence were ten or a hundred times cheaper, we wouldn't be worried about whether we were going to sequence three mammals, or six, or ten. We would just sequence a lot of them."

Even at present prices, sequencing is a bargain, Robertson argued. "When you think about it relative to the enormous amount of resources that NIH puts into grants to characterize individual genes of different species, sometimes it just seems ludicrous not to be sequencing their genomes." Sequencing an insect genome, he declared, would cost no more than a few National Institutes of Health (NIH) grants.

Sequencing costs have dropped several orders of magnitude, from \$10 per finished base in 1990 to today's cost, which Green estimates at about 5 or 6 cents per base for finished sequence and about 2 to 4 cents for draft sequence. For some comparisons, draft sequence is adequate. Last spring NHGRI projected future cost at about a cent per finished base by 2005.

Although the plummeting price of sequencing is welcome, it is due to incremental improvements on the basic technology. "What we're all praying for is one of those great breakthroughs - a new technology that will allow us to read single-molecule sequences, or whatever the trick is going to be that will give us several orders of magnitude increase in speed and reduced cost," Robertson said. Teams of competitive technology developers around the world are racing toward that goal, cheered on by a lot of casual prophecy about the [\\$1000 genome](#).

Nor is cost the only challenge. "The really big questions about genome dynamics, selection, adaptation, and gene networks await better theory, methods, and clever hypothesis testing," said Cristian I. Castillo-Davis, who studies regulatory sequence evolution at [Harvard University](#). "Currently, the field

is very thin on biological analysis and very heavy on technology and the reporting of numbers for numbers' sake." Castillo-Davis' solution? "We are in great need of biologists who can develop novel analytical tools and theory to make biological sense of comparative genomic data."

But the infrastructural problems of comparative genomics tend to fade in the dazzle of its prospects. With the data that will flood public databases in the next few years, Coghlan expects researchers to take on questions such as: How does regulatory DNA evolve? How is chromosome and protein evolution related to population size and structure? How do differences in meiosis and recombination in different species, such as those with holocentric chromosomes versus those with a true centromere, affect the structure of chromosomes and proteins?

"Research communities are realizing that they're going to wither if they don't have a genome project," Robertson said. "I suppose we're not going to sequence every genome on the planet, and that's certainly true if technology stays the way it is. But if technology changes as radically as some people think it will, then yes, why not sequence most of the species on earth?"

References

1. Holding C: *Caenorhabditis* comparative genomics *Genome Biology*, November 18, 2003., [http://genomebiology.com/researchnews/default.asp?arx_id=gb-spotlight-20031118-02]
2. Cold Spring Harbor Laboratory, [<http://www.cshl.org/>]
3. *PLoS Biology*, [<http://biology.plosjournals.org/>]
4. National Human Genome Research Institute (NHGRI), [<http://www.genome.gov/>]
5. Comparative analyses of multi-species sequences from targeted genomic regions
6. NHGRI Genome Sequencing Proposals, Status of Organisms in the Prioritization Process for Genome Sequencing and their 'White Paper' Proposals, [<http://genome.gov/page.cfm?pageID=10002154>]
7. Genome Sequencing Center, Washington University School of Medicine in St. Louis, [<http://genome.wustl.edu/>]

8. Powledge TM: Macaque advocates seek higher status *Genome Biology*, September 17, 2002., [http://genomebiology.com/researchnews/default.asp?arx_id=gb-spotlight-20020917-01]
9. The Institute for Genomic Research, [<http://www.tigr.org/>]
10. Smurfit Institute, Trinity College Dublin, [<http://www.tcd.ie/Genetics/>]
11. University of Illinois at Urbana-Champaign, [<http://www.uiuc.edu/index.html>]
12. Sequencing and comparison of yeast species to identify genes and regulatory elements
13. Pray L: A cheap personal genome? *Genome Biology*, October 7, 2002., [http://genomebiology.com/researchnews/default.asp?arx_id=gb-spotlight-20021007-02]
14. Harvard University, [<http://www.harvard.edu/>]