

# Researchers predict infinite genomes

22 September 2005

In a new study, TIGR scientists conclude that researchers might never fully describe some bacteria and viruses--because their genomes are infinite.

Ever since the genomics revolution took off, scientists have been busily deciphering vast numbers of genomes. Cataloging. Analyzing. Comparing. Public databases hold 239 complete bacterial genomes alone.

But scientists at The Institute for Genomic Research (TIGR) have come to a startling conclusion. Armed with the powerful tools of comparative genomics and mathematics, TIGR scientists have concluded that researchers might never fully describe some bacteria and viruses--because their genomes are infinite. Sequence one strain of the species, and scientists will find significant new genes. Sequence another strain, and they will find more. And so on, infinitely.

"Many scientists study multiple strains of an organism," says TIGR President Claire Fraser. "But at TIGR, we're now going a step further, to actually quantify how many genes are associated with a given species. How many genomes do you need to fully describe a bacterial species?"

In pursuit of that question, TIGR scientist Hervé Tettelin and colleagues published a study in this week's (September 19-23) early online edition of the Proceedings of the National Academy of Sciences (PNAS). In the study, TIGR scientists, with collaborators at Chiron Corporation, Harvard Medical School and Seattle Children's Hospital, compared the genomic sequence of eight isolates of the same bacterial species: *Streptococcus agalactiae*, also known as Group B Strep (GBS), which can cause infection in newborns and immuno-compromised individuals.

Analyzing the eight GBS genomes, the researchers discovered a surprisingly continual stream of diversity. Each GBS strain contained an

average of 1806 genes present in every strain (thus constituting the GBS core genome) plus 439 genes absent in one or more strains. Moreover, mathematical modeling showed that unique genes will continue to emerge, even after thousands of genomes are sequenced. The GBS pan-genome is expected to grow by an average of 33 new genes every time a new strain is sequenced.

"We were surprised to find that we haven't cornered this species yet," says Tettelin, lead author of the PNAS paper. "We still don't know--and apparently, we'll never know--the extent of its diversity."

To interpret this infinite view of microbial genomes, Tettelin and colleagues propose describing a species by its "pan-genome": the sum of a core genome, containing genes present in all strains, and a dispensable genome, with genes absent from one or more strains and genes unique to each strain.

The pan-genome is more than mere syntax. The concept has real implications for molecular biology. Many important pathogens--including those responsible for influenza, Chlamydia, and gastrointestinal infections, all under study at TIGR--contain multiple strains with specific genomes. By bringing a pan-genome perspective to the study of these organisms, scientists may better learn how new pathogens emerge and better target therapies to specific conditions. One approach is to spotlight a species's core genome. On the flip side, scientists may eliminate a core genome, hunting instead for fringe genes that explain a specific strain's unique activity.

TIGR researchers say the pan-genome concept also underscores the limits of traditional known genomes. Researchers often refer to a "type" genome to describe a given species. That singular, representative genome is often simply the strain easiest to acquire from nature or grow in the lab. Yet scientists worldwide routinely tap these known genomes in public databases to hunt for drug targets, explain ecological niches, and chart

evolution. How well do these microbial genomes reflect reality?

As comparative genomics itself evolves, Fraser expects TIGR to increasingly focus on pan-genomes. Many questions remain. Although some microbial species, such as GBS, have infinite pan-genomes, for instance, others are more limited. Comparing eight independent isolates of *Bacillus anthracis* (the bacterium that causes anthrax), for instance, Tettelin and colleagues found that just four genomes were sufficient to characterize its pan-genome. That raises interesting questions about rates of evolution, notes Fraser. "We're intrigued to learn more about the diversity within a given species, and how it happens," she says.

Source: The Institute for Genomic Research

APA citation: Researchers predict infinite genomes (2005, September 22) retrieved 24 September 2020 from <https://phys.org/news/2005-09-infinite-genomes.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*