

Charting ever-changing genomes

Instead of immutable proprietary software, any species' genetic information resembles open source code that is constantly tweaked and optimized to meet the users' specific needs. But which parts of the code have withstood the test of time and which parts have undergone rapid evolutionary change has been difficult to assess.

An international collaboration by researchers at the Salk Institute for Biological Studies, the University of Chicago, and the Max-Planck Institute for Developmental Biology developed a simple method to comb whole genomes for all the software fixes and security patches accumulated over time. In a first trial run, the scientists catalogued the genetic variations in 23 strains of the mustard weed *Arabidopsis thaliana* that were collected from the wild all over the world.

“Our study represents one of the first whole genome scans for levels and patterns of genetic variation within a species,” says Joseph R. Ecker, Ph.D., professor in the Plant Biology Laboratory and director of the Salk Institute Genomic Analysis Laboratory, who led the current study published in last week's online edition of the *Proceedings of the National Academy of Science*. “It reveals the regions that are currently targeted by natural selection or have been so during the evolutionary past.”

In an independent study the collaborators -- this time led by Detlef Weigel, Ph.D., director of the Max Planck Institute for Developmental Biology in Tübingen, Germany, and an adjunct professor at the Salk Institute -- went through the genomes of 20 different strains of *Arabidopsis thaliana* with an even finer-toothed comb, allowing them to determine the exact nature of the changes. The findings of the second study are published in the July 20 issue of the journal *Science*.

“We found that one out of 10 genes is very different,” says Weigel. “This plasticity is truly surprising for a genome that's very streamlined and unlike bigger genomes doesn't contain a lot of junk DNA,” he adds.

A decade ago, *Arabidopsis* was widely adopted by plant scientists as an easily manipulated model for other plants because it is simple to grow in the laboratory, has a short life cycle and a small genome. Compared to corn, which might have as many as 2.5 billion base pairs of DNA and the human genome with roughly 3 billion pairs, *Arabidopsis* only has about 120 million base pairs of DNA.

With nowhere to run, plants are under constant threat from heat, cold, high acidity or salinity, or pathogens such as viruses and leaf-munching insects. In response, plants mobilize physiological and biochemical defenses that help them survive. “We expected certain classes of genes to be highly variable due to natural selection in different environments. Both studies revealed precisely which gene family members indeed were shaped by evolution,” says Justin Borevitz, Ph.D., a former post-doctoral researcher in the Ecker lab and now an assistant professor in the Department of Ecology and Evolution at the University of Chicago.

As a general rule, genes that don't change over time are under strong negative selection because they perform important housekeeping functions, while genes that vary widely such as disease resistance genes are under strong positive selection. “We covered both ends of the spectrum and ended up with a top list of no changes and a top list of a lot of changes,” explains Borevitz. “All the data have been placed in a publicly accessible database and now researchers everywhere can look up their favorite genes.”

To assemble their lists, the Ecker team poured over data derived from old-fashioned gene-chip technology, in which 25 nucleotide-long samples of every gene expressed in an *Arabidopsis* cell are spotted onto a tiny glass slide known as a microarray. The chopped up genomes of the different strains were then allowed to bind to their immobilized counterparts. Reduced hybridization resulted in a signal telling the researchers which regions the genomes differed from the fully sequenced reference strain.

“This method is simple and relatively inexpensive and can be applied to any organism whose whole genome has been sequenced and for which a gene array is available or can be easily made,” explains Ecker. “For these reasons it is attractive to a wide audience practicing evolutionary genomics.”

Weigel’s team went a step further and effectively re-sequenced whole genomes with the help of nearly a billion 25-mers tiled on 5 large arrays that cover every possible nucleotide exchange on both strands of DNA. The high-resolution approach revealed a high number of specific changes in genes belonging to the so-called F-box superfamily, whose members plays a crucial part in flagging proteins for degradation.

“As highlighted by both studies, many genes that harbor major-effect changes in wild populations are likely to mediate interactions with the environment,” says Weigel. “Ultimately, experiments under more natural conditions will be required to fully appreciate the functional relevance of such sequence variation.”

Source: Salk Institute

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