

One species, many genomes



Arabidopsis plants from different geographical origins differ in many traits (the background shows schematically sequence variation in the DNA of these plants). Credit: MPI for Developmental Biology

Faster growth, darker leaves, a different way of branching - wild varieties of the plant *Arabidopsis thaliana* are often substantially different from the laboratory strain of this small mustard plant, a favorite of many plant biologists.

Which detailed differences distinguish the genomes of strains from the polar circle or the subtropics, from America, Africa or Asia has been investigated for the first time by research teams from Tübingen, Germany, and California led by Detlef Weigel from the Max Planck Institute for Developmental Biology. The results were surprising: The extent of the genetic differences far exceeds the expectations for such a streamlined genome, as the scientists write in this week's edition of *Science* magazine.

To track down the variation in the genome of the different *Arabidopsis* strains, the researchers compared the genetic material of 19 wild strains with that of the genome of the lab strain, which was sequenced in the year 2000. Using a very elaborate procedure, they examined every one of the roughly 120 million building blocks of the genome. For their molecular sleuthing they used almost one billion specially designed DNA probes. "All together, these probes would have seven times the length of human genome," illustrates Weigel the extent of the project. The data were evaluated with several specially designed statistical methods, including a variant of machine learning.

The result of this painstaking analysis: on average, every 180th DNA building block is variable. And about four percent of the reference genome either looks very different in the wild varieties, or cannot be found at all. Almost every tenth gene was so defective that it could not fulfill its normal function anymore!

Results such as these raise fundamental questions. For one, they qualify the value of the model genomes sequenced so far. "There isn't such a thing as the genome of a species," says Weigel. He adds "The insight that the DNA sequence of a single individual is by far not sufficient to understand the genetic potential of a species also fuels current efforts in human genetics."

Still, it is surprising that *Arabidopsis* has such a plastic genome. In contrast to the genome of humans or many crop plants such as corn, that of *Arabidopsis* is very much streamlined, and its size is less than a twentieth of that of humans or corn—even though it has about the same number of genes. In contrast to these other genomes, there are few repeats or seemingly irrelevant filler sequences. "That even in a minimal genome every tenth gene is dispensable, has been a great surprise," admits Weigel.

Detailed analyses showed that genes for basic cellular functions such as protein production or gene

regulation rarely suffer knockout hits. Genes that are important for the interaction with other organisms, on the other hand, such as those responsible for defense against pathogens or infections, are much more variable than the average gene. "The genetic variability appears to reflect adaptation of local circumstances," says Weigel. It is likely that such variable genes allow plants to withstand dry or wet, hot or cold conditions, or make use of short and long growing seasons.

Such genome analyses of unprecedented details will allow a much better understanding of local adaptation, and this was indeed one of the main reasons for conducting the study. "By extending these types of studies to other species we hope to help breeders to produce varieties that are optimally adapted to rapidly changing environmental conditions," explains Weigel. He is already collaborating with the International Rice Research Institute (IRRI) in the Philippines to apply the methods and experience gathered with *Arabidopsis* to twenty different rice varieties.

How environment and genome interact is also the goal of new, even more powerful methods. While the technology used so far can only identify genes that have changed or are lost relative to the reference genome, direct sequencing of the genome of wild strains will allow the detection of new genes. The plan is to decipher the genomes of at least 1001 *Arabidopsis* varieties. A new instrument, with which the entire genome of a plant can be read in just a few days, is already available. Still missing are the computational algorithms to interpret the anticipated flood of data.

Citation: Richard M. Clark, et al., Common Sequence Polymorphisms Shaping Genetic Diversity in *Arabidopsis thaliana*, *Science*, July 20, 2007

Source: Max-Planck-Gesellschaft

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.