Sequencing thousand and one genomes

September 29 2008

Different strains of thale cress, with the background schematically indicating differences in the genetic code. Photo: Max Planck Institute for Developmental Biology

(PhysOrg.com) -- Researchers at the Max Planck Institute for Developmental Biology in Tuebingen, Germany, reported the completion of the first genomes of wild strains of the flowering plant Arabidopsis thaliana. The entire genomes of two individuals of this species, one from Ireland, the other from Japan, have now been compared in great detail. They were found to be astonishingly different from each other, as Detlef Weigel and his colleagues write in *Genome Research*.

This study marks the starting point of the 1001 Genomes Project in which a total of thousand and one individuals of the same species will be sequenced. The scientists aim at correlating the genetic differences between the different strains with variation in the speed of plant growth and their resistance against infectious germs. These strategies could then also be applied to crop plants or trees.

Every genome is different. Everybody knows that the genome of apes must be different from our human genome and that both are different from the genome of a sunflower. It is only a few years ago that a huge research community produced at great cost a single human genome sequence. The assumption was that it would unlock all the essential features of our species, since any differences between us were thought to be very minor, on the order of 0.1 percent of the entire genome. Similar views prevailed for other species, including thale cress Arabidopsis thaliana, a model organism in plant science. It is one of the best-understood organisms on earth, however, the genetic differences that allow different strains of this plant to thrive in very different places all over the Northern hemisphere are largely unknown.

Until very recently, it was assumed that the similarity in appearance of different individuals of thale cress is matched by a similar degree of similarity in the genetic material. “But is it really true that such subtle differences in our DNA or in that of thale cress can account for the great variation in individual traits? Is there indeed something like ‘the’ genome of a species, or do have to change our point of view and focus on the genome of an individual?” asks Detlef Weigel, director at the Max Planck Institute for Development Biology.

Recent advances in the technology of DNA sequencing have reduced the price for reading a single genome by several orders of magnitude, and this can now be accomplished within a week, rather than months or years. However, there are still few analytical tools for the torrent of data

produced by the new generation of sequencing machines, such as the one sold by the San Diego based company Illumina. The Max Planck Institute group had to overcome a series of technical challenges to reconstruct the genome sequences of the two strains it analyzed from the rather short snippets of sequences that the Illumina instrument delivers. But the first feasibility study has now been finished, demonstrating that even with these very short sequence reads not only point differences can be identified, but also missing or extra genetic material can be tracked down. “We are confident that our method is robust, and we have begun to sequence the genomes of 80 thale cress strains”, says Weigel. The project should be finished by January 2009.

The study marks the start of a project on a much larger scale. Within the next two years the 1001 Genomes project, spearheaded by Weigel, plans to sequence at least 1001 different thale cress individuals from around the world. The hope is that armed with this information, it will be possible to correlate genetic differences with variation in the speed with which plants grow, how much they branch, or how well they resist infectious germs. This project, in turn, will inform similar projects on crop plants, which have much larger genomes and are therefore more difficult to analyze. While this is very exciting, the task will not be done once every individual genome is sequenced. In every cell, the genomes are packaged in different ways, allowing for different activities of the same genetic material. With the next sequencing techniques, these subtle differences can be studied as well. Thus, the 1001 Genomes project will peel away only the first layer of variation.

Citation: Stephan Ossowski, Korbinian Schneeberger, Richard M. Clark, Christa Lanz, Norman Warthmann, Detlef Weigel: Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Res.* Published online September 25, 2008, 10.1101/gr.080200.108

Provided by Max Planck Institute