

SOYBEAN GENOMICS

Pan-genome upgrade

Cell **182**, 1–15 (2020).



Credit: IGOR STEVANOVIC / ALAMY STOCK PHOTO

Thanks to the advances of sequencing technologies, pan-genomes can now be assembled to capture the full genetic diversity of a certain species. A soybean pan-genome based on seven wild accessions was assembled with short reads a few years ago. To provide a higher-quality pan-genome that better represents global soybean diversity, Yucheng Liu, from the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, and colleagues have now generated a graph-based high-quality pan-genome using 26 newly assembled genomes together with three previously published ones.

Using short-read sequencing, the researchers first re-sequenced a total of 2,898 representative accessions that include wild soybeans, landraces and cultivars. Based on their phylogeny, geographic distributions and contributions to breeding, 26 representative accessions were chosen for *de novo* assembly, each of which was then sequenced and assembled using multiple technologies. The yielded genomes were of high contiguity and completeness, with protein-coding genes and non-coding RNAs well annotated by transcriptomes.

Core and dispensable genes were identified, with core genes being more functionally conserved and enriched in different biological processes and pathways from dispensable genes.

Sequence variations including small indels, single-nucleotide polymorphisms as well as

large structural variations (SVs) were also identified. An integrated graph-based genome was then built based on the presence and absence variations as well as a ZH13 reference genome. Additional SVs were found by mapping short reads of the 2,898 accessions to the graph-based genome. SVs tend to be enriched in repeat sequences and are present at a higher rate in the wild than in cultivated soybeans.

Sequence variations modulate gene expression, profoundly affect gene structures and cause at least 15 gene fusion events in these accessions. Some variations showed association with domestication or agronomical traits, or differentiation between wild and cultivated soybeans, supporting their importance in domestication and breeding.

A whole genome duplication (WGD) event expanded the soybean genome ~13 million years ago. This WGD appears to restrain the evolutionary rate of DNA sequences because WGD regions contain a higher ratio of core genes and a lower ratio of dispensable genes than the non-WGD regions. Moreover, WGD regions exhibit less SVs and lower nucleotide diversity than non-WGD regions.

This pan-genome provides a valuable resource that should boost soybean research in the post-genomic era.

Jun Lyu

Published online: 6 July 2020
<https://doi.org/10.1038/s41477-020-0731-2>