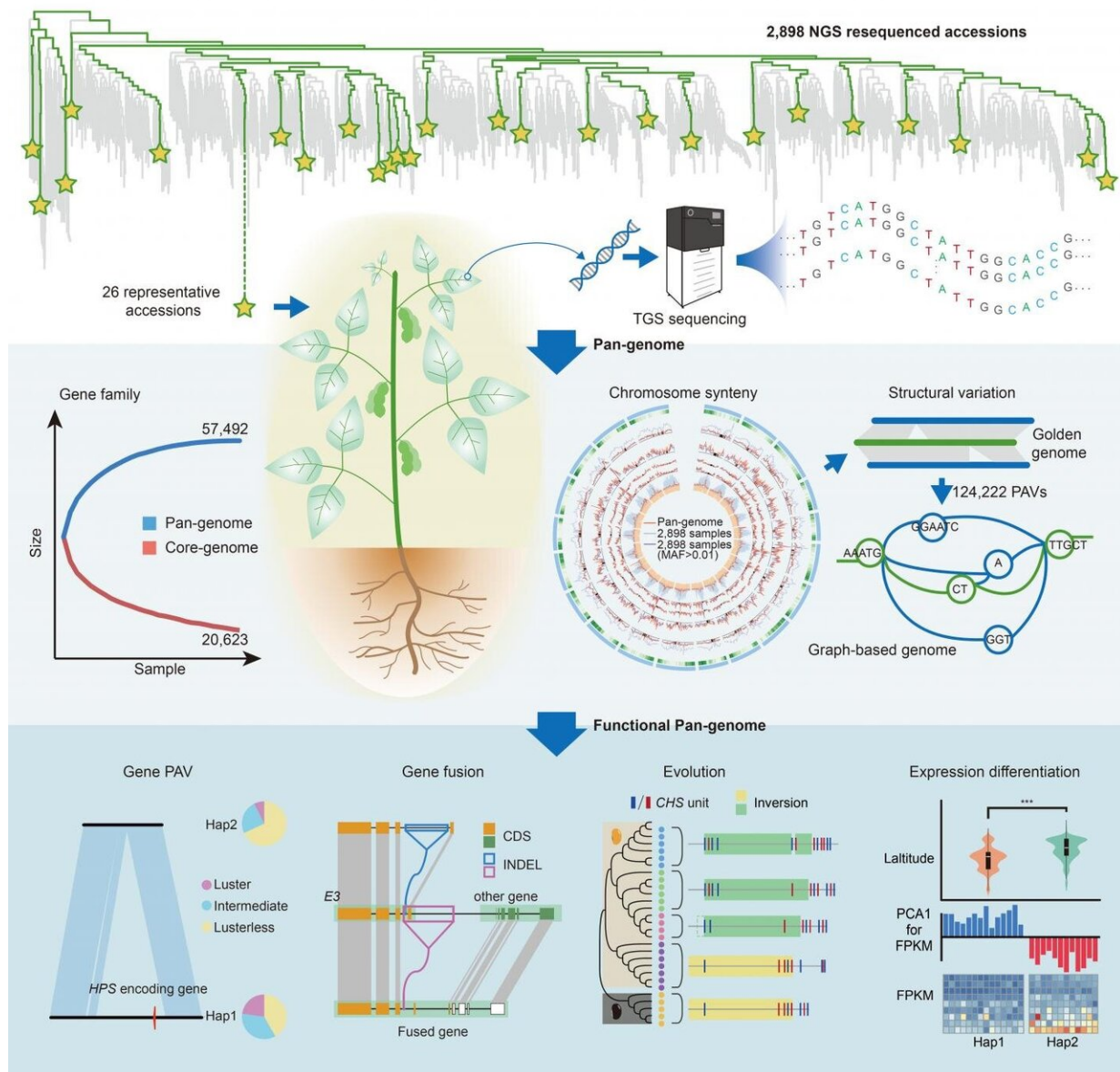


Scientists construct high-quality graph-based soybean genome

June 17 2020



Soybean graph-based genome construction and pan genome analyses. Credit: IGDB

Soybean oil is one of the world's most important vegetable oils and soybeans are a key protein feed crop. Cultivated soybeans were domesticated from wild relatives in China approximately 5,000 years ago. At present, over 60,000 accessions adapted to different ecoregions have been developed. Extensive genetic diversity among soybean germplasms has shown the need for construction of a complete pan-genome from diverse soybean accessions.

Recently, the research group led by Prof. Tian Zhixi from the Institute of Genetics and Developmental Biology (IGDB) of the Chinese Academy of Sciences (CAS), in cooperation with Profs. Liang Chengzhi and Zhu Baoge's team, Prof. Han Bin's team from the Center for Excellence in Molecular Plant Sciences of CAS, Prof. Huang Xuehui's team from Shanghai Normal University, and the Berry Genomics Corporation, individually de novo assembled 26 [soybean](#) genomes and constructed a high quality graph-based soybean pan-[genome](#).

Based on a [phylogenetic analysis](#) of 2,898 soybean accessions, they selected 26 accessions and performed de novo genome assembly for each accession. The contig N50 sizes of the 26 whole-genome assemblies ranged from 18.8 to 26.8 Mb pairs with a mean of 22.6 Mb, and scaffold N50 sizes ranged from 50.3 to 52.3 Mb with a mean of 51.2 Mb.

Through a comparative genome analysis of the 26 genomes plus three previously reported genomes, the scientists identified a total of 14,604,953 SNPs and 12,716,823 small insertions and deletions, 723,862 present and absent variations, 27,531 copy number variations, 21,886

translocation events, and 3,120 inversion events.

Subsequently, by integrating these structural variations, a graph-based genome was built using the ZH13 genome as a standard linear reference genome.

Further investigations illuminated that these structural variations play important roles in driving genome evolution, gene structure [variation](#) and gene functional divergence, which in turn contribute to agronomic trait variations in the soybean population.

Having a reference genome opens the door to functional genomics and molecular design breeding for a species. However, an increasing number of reports has suggested that one or a few reference genomes cannot represent the full range of genetic diversity of a species. Therefore, pan-genome construction is becoming increasingly necessary.

In addition, conventional linear references are limited since they are unable to show the genotypes of different alleles from each locus. How to integrate the genotypes from different alleles into a new form of genome is a challenge.

This is the first reported graph-based genome in a plant. This graph-based genome can be used to reanalyze previously resequenced data, which will generate more comprehensive information than ever and rejuvenate those data. In turn, this will greatly facilitate functional study and breeding. An anonymous reviewer said this work is "a landmark paper for genomics."

More information: Yucheng Liu et al, Pan-Genome of Wild and Cultivated Soybeans, *Cell* (2020). [DOI: 10.1016/j.cell.2020.05.023](https://doi.org/10.1016/j.cell.2020.05.023)

Provided by Chinese Academy of Sciences

Citation: Scientists construct high-quality graph-based soybean genome (2020, June 17) retrieved 21 June 2024 from

<https://phys.org/news/2020-06-scientists-high-quality-graph-based-soybean-genome.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.