

When one reference genome is not enough

December 20 2017



A single reference genome is not enough to harness the full genetic variation of a species so pan-genomes of crops would be extremely useful. The phenotypic diversity of *Brachypodium* plants is demonstrated in this image, which is associated with a news release for a Nature Communications paper in which an international team led by DOE Joint Genome Institute researchers gauged the size of a plant pan-genome using the model grass *Brachypodium distachyon*.

Credit: John Vogel

Much of the research in the field of plant functional genomics to date has relied on approaches based on single reference genomes. But by itself, a single reference genome does not capture the full genetic variability of a species. A pan-genome, the non-redundant union of all the sets of genes found in individuals of a species, is a valuable resource

for unlocking natural diversity. However, the computational resources required to produce a large number of high quality genome assemblies has been a limiting factor in creating plant pan-genomes.

Having plant pan-genomes for crops that are important for fuel and food applications would enable breeders to harness natural diversity to improve traits such as yield, disease resistance, and tolerance of marginal growing conditions. In a paper published December 19, 2017 in *Nature Communications*, an international team led by researchers at the U.S. Department of Energy (DOE) Joint Genome Institute (JGI), a DOE Office of Science User Facility at Lawrence Berkeley National Laboratory (Berkeley Lab), gauged the size of a plant pan-genome using *Brachypodium distachyon*, a wild grass widely used as a model for grain and biomass crops. As one of the JGI's Plant Flagship Genomes, *B. distachyon* ranks among the most complete plant reference genomes.

"There are a vast number of genes that are not captured in a single reference genome," added study senior author John Vogel, head of the JGI's Plant Functional Genomics group. "Indeed, about half of the genes in the pan-genome are found in a variable number of lines." Working toward the primary goal of accurately estimating the size of a plant pan-genome, Vogel and his colleagues performed whole-genome de novo assembly and annotation of 54 geographically diverse lines of *B. distachyon*, yielding a pan-genome containing nearly twice the number of genes found in any individual line.

"The genome of a species is a collection of genomes, each with their own unique twist," added JGI bioinformaticist and study first author Sean Gordon. "Now knowing that focusing on a single reference genome leads to incomplete and biased estimates of genetic diversity and ignores genes potentially important for breeding applications, we should better incorporate multiple references in future studies of [natural diversity](#)."

Moreover, genes found in only some lines tend to contribute to biological processes (e.g., disease resistance, development) that may be beneficial under some environmental conditions, whereas genes found in every line usually underpin essential cellular processes (e.g., glycolysis, iron transport).

"This means that the variable genes are being preferentially retained if they are beneficial under some conditions. These are exactly the types of genes that breeders need to improve crops." Vogel said.

In addition, genes found in only a subset of lines displayed faster rates of evolution, lay closer to transposable elements (thought to play a key role in pan-[genome](#) evolution), and were less likely to be found in the same chromosomal location as functionally equivalent [genes](#) in other grasses.

The sequence assemblies, gene annotations and related information can be downloaded from the project website BrachyPan: brachypan.jgi.doe.gov. The [Brachypodium distachyon genome](#) is available on the JGI Plant Portal Phytozome: phytozome.jgi.doe.gov.

More information: Sean P. Gordon et al, Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure, *Nature Communications* (2017). [DOI: 10.1038/s41467-017-02292-8](#)

Provided by DOE/Joint Genome Institute

Citation: When one reference genome is not enough (2017, December 20) retrieved 24 April 2024 from <https://phys.org/news/2017-12-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.