Plant genomes: toward goals of decoding both complex and complete sequences

Fei Chen[1,2,3*]

[1] *Sanya Nanfan Research Institute of Hainan University, Sanya 572025, China*
[2] *School of Tropical Crops, Hainan University, Haikou 570228, China*
[3] *Hainan Yazhou Bay Seed Laboratory, Sanya 572024, China*
* Corresponding author, E-mail: feichen@hainanu.edu.cn

Over the last 20 or so years, since the first release of the Arabidopsis genome in 2000, scientists have accomplished more than 1,000 *de novo* genome assemblies (https://plabipd.de/portal/plant-genomes) as well as a series of pan-genomes. The real challenges facing us are the decoding at the highly ploidy level and highly heterozygous plant genomes, as well as giga chromosomes and numerous chromosomes within a cell.

Luckily, relying on the (i) ultra-long read sequencing technologies such as Oxford Nanopore Technologies and PacBio HiFi (www.pacb.com/technology/hifi-sequencing/) and Revio technology (www.pacb.com/revio/), and (ii) novel assembly algorithms such as wtdbg2 (https://github.com/ruanjue/wtdbg2) and nextdenovo (https://github.com/Nextomics/NextDenovo), researchers are able to generate chromosome-level assemblies and even complete chromosomes (www.maxapress.com/article/doi/10.48130/TP-2022-0007), thereby perform chromosomal level discoveries and comparisons. Previous studies have reported the hexaploidy sugarcane genome, and giga genomes of garlic, onion, and the Chinese pine.

Recently, scientists have used a combined sequencing strategy, i.e., PacBio long-read sequencing coupled with Illuminar short-read sequencing and Hi-C sequencing, as well as the combined assembly technology of SOAPdenovo and wtdbg. They have eventually decoded the genome of paeonia (*Paeonia ostii*), which is a plant with both giga genome (12.28 Gb) and giga chromosomes (up to 2.6 Gb of chromosome 1, see Fig. 1). Because of the rapid and tremendous LTR insertions, but not due to ploidy duplication, both the genome size grow to an enormous size and the protein-coding gene reached 73,177 in number (www.nature.com/articles/s41467-022-35063-1).

In the near future, we are confident that more complex and giga sized genomes will be decoded, such as okra (*Abelmoschus esculentus*, allopolyploid with 2n = 56–196), lilies (*Lilium* spp., genome size around 35–60 Gb), tulips (*Tulipa* spp, genome size around 34 Gb), and Australian water lily (*Nymphaea gigantea* with 2n = 16x = 224). One of the 125 large questions proposed by *Science* journal in 2021 asked: 'Why are some genomes so big and others very small?' (www.science.org/content/resource/125-questions-exploration-and-discovery). We are confident that these giga and complex genomes will eventually provide a very clear clue to this question.

Conflict of interest

Fei Chen is an Editorial Board member of *Ornamental Plant*
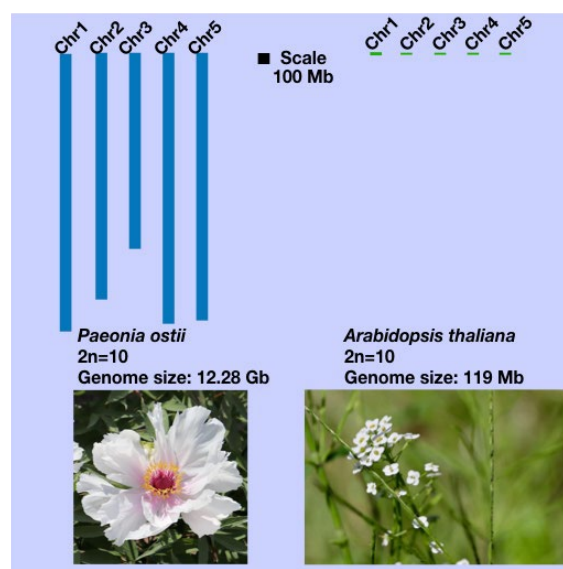
Fig. 1 The genome size and chromosome comparison between the ornamental plant *Paeonia ostii* and the model plant *Arabidopsis thaliana*.

Research. He was blinded from reviewing or making decisions on the manuscript. The article was subject to the journal's standard procedures, with peer-review handled independently of these Editorial Board members and their research groups.