Surfing the genomic new wave

In the last decade, high-throughput sequencing approaches have revolutionized the field of plant genomics. With the pace of technical improvement showing no sign of slowing what advances could be just around the corner.

Since the beginning of the century, the release of the *Arabidopsis* and rice genomes has benefited plant researchers and enabled the understanding of many aspects of plant biology. Over 300 plant species have been sequenced at different levels of quality, allowing experts to explore general biological questions as well as lineage or species-specific issues. This issue alone, includes three genome papers: one oak genome (*Nat. Plants* <https://doi.org/10.1038/s41477-018-0172-3>; 2018), two fern genomes (*Nat. Plants* <https://doi.org/10.1038/s41477-018-0166-1>; 2018) and a rose genome (*Nat. Plants* <https://doi.org/10.1038/s41477-018-0188-8>; 2018).

Every sequenced genome has value, if only that it adds to our appreciation of the diversity of life. However, there are a variety of reasons that a genome sequence may have particular importance. It may belong to a species that has been extensively used as a model organism, such as *petunia*¹. Or it may be of a plant with a particularly interesting physiology, like the resurrection plant *Xerophyta viscosa*². The genome may belong to a species of agricultural importance, such as goat grass³; industrial value, like the rubber tree⁴; or that has a particular phylogenetic significance, like gnetophytes⁵ and ferns. The plant itself may be of little importance but its genome allows for investigation of a key biological process, as with *Arabidopsis alpina*⁶ — the study of which illuminated how transposons contribute to shaping plant genomes. Lastly, methodological or technological advances can provide a sequencing study with an unusual level of significance, such as the sweet potato genome's tackling of polyploid genome assembly⁷.

Methodological and technological innovations continue to be the main drivers of the genomic revolution. Improvements in throughput, efficiency, read length and accuracy have resulted from new enzymes, reagents and workflows. Simultaneously, cost has dramatically dropped. It is now possible to de novo assemble genomes similar in size to rice at a cost less than US\$10,000. The length of individual reads is also increasing. The PacBio system can now stably generate long-reads at an average length of over 10 kilobases, and has been

combined with short-read sequencing in many genome projects. For example, the *Petunia axillaris* genome¹, rose genome and fern genomes all benefited from this technology. The addition of PacBio long-reads dramatically improved genome contiguity and resulted in contig N50 size approximating or even exceeding the megabase (Mb) level.

Nanopore technology relies on changes in ion current when the DNA strand passes through a protein pore. Oxford Nanopore technologies recently achieved a breakthrough by optimizing the protein nanopore and algorithm of base calling. Read lengths can now average several hundred kilobases, and ultra-long-reads of over 1 Mb have also been reported (<https://go.nature.com/2tgQjCF>). So far, the Nanopore MinION sequencer has generated high-contiguity genomes for *Arabidopsis thaliana*⁸, wild tomato⁹ and teak tree¹⁰. In the near future more plant species, including major crops, will be sequenced by this technology.

Important innovations have also occurred in the development of assembly-assisting methods, such as optical mapping and high-throughput chromosome conformation capture (Hi-C), that aid the assembly of contigs into scaffolds and chromosomes. Optical mapping generates genome-wide high-resolution restriction maps from single, stained DNA molecules. The restriction maps can then be used to order and orient contigs. This approach has been widely applied in the recently published wheat A genome¹¹. It can assemble contigs into chromosome-scale scaffolds, and has been shown to work well in multiple organisms, including human, *Drosophila*, mouse and *Arabidopsis*^{12–15}.

Plant genomes tend to be of high heterozygosity, high repeat rate, large size and are frequently polyploid. These features greatly hinder genome assembly, but the new technologies offer some solutions. Long-read sequencing technologies and Hi-C can partially resolve the problems of high-repeat content and high heterozygosity by increasing the contiguity of reads. The complexity of some plant genomes will still make assembly hard, but with increasingly

efficient tools it is possible to generate reference genomes for most species at an acceptable cost and quality. Recently, an international consortium announced the 10,000 Plant Genome Sequencing Project¹⁶ (*Nat. Plants* 4, 312–313; 2018), aiming to sequence over 10,000 genomes across the plant tree of life. This project will refine our understanding of phylogeny, genome structure and evolution, and provide valuable resources for many plant communities.

Reduced costs and increased throughput will soon allow researchers to investigate the genetics and evolution of species at the population level by massive resequencing. New technologies will also provide opportunities for archaeobotanical research (*Nat. Plants* <https://doi.org/10.1038/s41477-018-0187-9>; 2018); an area where progress is currently slow due, in part, to the large size and repetitiveness of plant genomes. Furthermore, single-cell genome sequencing, should help us understand the details of plant developmental processes in greater detail.

Nature Plants will continue to follow the progress of this genomics revolution and continue to publish genomes with broad interest. Long-read sequencing has already started a new wave of genome sequencing. It will be fascinating to see what further unanticipated knowledge the ever-more detailed interrogation of plant genomes will bring. □

Published online: 5 July 2018
<https://doi.org/10.1038/s41477-018-0209-7>

References

- Bombarely, A. et al. *Nat. Plants* 2, 16074 (2016).
- Costa, M. et al. *Nat. Plants* 3, 17038 (2017).
- Zhao, G. et al. *Nat. Plants* 3, 946–955 (2017).
- Tang, C. et al. *Nat. Plants* 2, 16073 (2016).
- Wan, T. et al. *Nat. Plants* 4, 82–89 (2018).
- Willing, E. et al. *Nat. Plants* 1, 14023 (2015).
- Yang, J. et al. *Nat. Plants* 3, 696–703 (2017).
- Michael, T. et al. *Nat. Commun.* 9, 541 (2018).
- Schmidt, M. et al. *Plant Cell* 29, 2336–2348 (2017).
- Yasodha, R. et al. *DNA Res.* <http://doi.org/gdmf75> (2018).
- Ling, H. et al. *Nature* 557, 424–428 (2018).
- Burton, J. N. et al. *Nat. Biotechnol.* 31, 1119–1125 (2013).
- Selvaraj, S. et al. *Nat. Biotechnol.* 31, 1113–1119 (2013).
- Kaplan, N. & Dekker, J. *Nat. Biotechnol.* 31, 1139–1143 (2013).
- Xie, T. et al. *Mol. Plant* 8, 489–492 (2014).
- Cheng, S. et al. *GigaScience* 7, giy0131 (2018).