



OPEN ACCESS

EDITED AND REVIEWED BY
Sarah Veronica Schiessl,
University of Giessen, Germany

*CORRESPONDENCE
Agnieszka Zmienko
✉ akisiel@ibch.poznan.pl

RECEIVED 23 October 2023
ACCEPTED 10 November 2023
PUBLISHED 16 November 2023

CITATION
Zmienko A, Wojciechowski P and
Figlerowicz M (2023) Editorial: Resolving
the complexity of plant genomes and
transcriptomes with long reads, volume II.
Front. Plant Sci. 14:1326500.
doi: 10.3389/fpls.2023.1326500

COPYRIGHT
© 2023 Zmienko, Wojciechowski and
Figlerowicz. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Editorial: Resolving the complexity of plant genomes and transcriptomes with long reads, volume II

Agnieszka Zmienko^{1,2*}, Paweł Wojciechowski^{1,2}
and Marek Figlerowicz¹

¹Institute of Bioorganic Chemistry of Polish Academy of Sciences, Poznan, Poland, ²Faculty of Computing and Telecommunications, Institute of Computing Science, Poznan University of Technology, Poznan, Poland

KEYWORDS

long read DNA sequencing, assembly, transcriptomics, transposable element (TE), PacBio, nanopore

Editorial on the Research Topic

Resolving the complexity of plant genomes and transcriptomes with long reads, volume II

It has been over 20 years since the genome assembly of the first land plant – *Arabidopsis thaliana* – has been published ([Arabidopsis Genome Initiative, 2000](#)). During this time, numerous plant genomes have been sequenced and large-scale studies of intra- and interspecific genetic variation were conducted. Many plant genomes are large and complex, with varying levels of ploidy and/or high abundance of the transposable elements (TEs), which has been a strongly limiting factor in genomic studies. However, within just a decade since the emergence of long read sequencing (a.k.a third generation sequencing) we have observed a dynamic increase in the number of *de novo* assembled genomes in the public databases. This could have not been achieved without the impressive improvement of the PacBio HiFi and Nanopore sequencing, which are currently the most popular third generation sequencing methods. Both techniques deliver reads with >99% base accuracy and the accuracy of PacBio HiFi reads, which are generated by circular consensus sequencing, is comparable to that of the classical Sanger sequencing. On the other hand, Nanopore sequencing offers unrestricted read length, with >4 million base already reported. Accordingly, the range of analyzed species quickly expanded beyond the most common or economically important ones, promising unprecedented insight into the genetic biodiversity on Earth ([Gupta, 2022](#)). In line with this trend, within this Research Topic we present studies which utilized long read sequencing to investigate plant genomes, with the special emphasis on producing high quality *de novo* assemblies of species unique to specific geographical localizations or habitats. This included two trees from Fagaceae family, Chinese cork oak (*Quercus variabilis*) and Japanese chestnut (*Castanea crenata* Sieb. et Zucc), which are ecologically and economically important species native to East Asia ([Han et al., Wang et al.](#)), as well as four tree species representing *Syzygium* genus (Myrtaceae), three of them being autotetraploid ([Ouadi et al.](#)). The assemblies were

compared with the available genomes of related plants, which provided valuable insight into the genome evolution and the history of gene family expansions/contractions in the species of interest.

Another interesting study reported assembling the genome of Tibetan sea-buckhorn (*Hippophae tibetana*). This perennial dense thorny shrub can be found in Tibetan Plateau, which is the world's highest and largest plateau. The extreme environmental conditions in this region impose strong selective pressures and drive genome evolution, leading to unique adaptations of the local species. Wang et al. sequenced the Tibetan sea-buckhorn genome to study its organization and search for the genetic features that might contribute to its ability of growing at high altitudes – even 5000 meters above sea level. They found that TE amplification largely accounted for its genome-size expansion. Moreover, based on the comparison of the relative position of genes and intact or fragmentary Gypsy/Copia elements, they suggested that these retrotransposons might specifically play a role in adaptation to high-altitude habitats. Interestingly, Ouadi et al. also observed species-specific evolutionary dynamics of Gypsy and Copia retrotransposons in the *Syzygium* species. Repetitive elements also constituted a substantial fraction of the Chinese cork oak and Japanese chestnut assemblies (67.6% and 58.78%, respectively), which highlights the important role of TE amplification in shaping plant genomes.

Along with the lowering costs and increased access to the sequencing devices, many labs incorporated genome sequencing as the first step in functional genomic studies. A nice example of such approach has been presented by Hu et al., who produced the genome of *Scutellaria baicalensis* Georgi (Lamiaceae) with third generation sequencing reads and used it to annotate genes. Next, they quantified and compared gene expression among the plants differing by the color of flowers (purple, pink or white). The roots of *S. baicalensis* are rich in flavonoids, therefore the authors focused especially on characterizing and comparing the expression of genes involved in flavonoid biosynthesis pathway and transcription factors related to this process.

The genomes of plants described in this Research Topic's papers have different sizes, from about 380 Mb in *Syzygium* and *S. baicalensis* up to 1.5 Gb in Tibetan sea-buckhorn. It should be stressed that in all presented cases, the draft genomes (either based on the PacBio or Nanopore reads), were scaffolded with the additional information from the chromosome conformation capture (Hi-C) data. Indeed, combining third generation sequencing with the methods allowing to determine chromatin structure or with high density genetic maps has proven to be invaluable in assembling telomere-to-telomere chromosomes of large plant genomes, e.g. watermelon and maize (Deng et al., 2022; Chen et al., 2023). With the fast improvement of the Nanopore reads length and accuracy as well as the development of bioinformatics algorithms, we can expect complex gap-free genomic assemblies resolved solely with long read sequencing data in the near future.

The possibility to obtain full-length cDNA sequences with third generation sequencing creates fantastic opportunities also for transcriptomic studies. Long read-based *de novo* transcriptome assembly and quantification facilitates identification of gene UTRs and distinction between the alternatively spliced transcripts, representing different protein isoforms. It is also much easier to unambiguously map long reads to genes from which they were derived. It seems especially important for plant genomes, which frequently undergo whole genome and segmental duplication events. Hu et al. used long read approach to investigate the transcriptomic complexity of 9 citrus species and their close relatives. Using the Nanopore data they were able to identify both novel isoforms of known genes as well as new expressed genes in all analyzed species. Having full-length transcript sequences also allowed them to analyze the frequency of various types of splicing events. Additionally, they identified long noncoding RNAs in their datasets.

Long read sequencing has been named the Method of the Year 2022 (Marx, 2023). With the growing number of providers and applications, including DNA methylation analyses and single-cell long read sequencing, it seems certain that this technique will irrevocably and positively influence the genomic studies of the entire decade.

Author contributions

AZ: Conceptualization, Writing – original draft, Writing – review & editing. PW: Conceptualization, Writing – review & editing. MF: Conceptualization, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Chen, J., Wang, Z., Tan, K., Huang, W., Shi, J., Li, T., et al. (2023). A complete telomere-to-telomere assembly of the maize genome. *Nat. Genet.* 55, 1221–1231. doi: 10.1038/s41588-023-01419-6
- Deng, Y., Liu, S., Zhang, Y., Tan, J., Li, X., Chu, X., et al. (2022). A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. *Mol. Plant* 15, 1268–1284. doi: 10.1016/j.molp.2022.06.010
- Gupta, P. K. (2022). Earth Biogenome Project: present status and future plans. *Trends Genet.* 38, 811–820. doi: 10.1016/j.tig.2022.04.008
- Marx, V. (2023). Method of the year: long-read sequencing. *Nat. Methods* 20, 6–11. doi: 10.1038/s41592-022-01730-w