Genomes of Begonia Reveals the Mechanisms of Diversity Formation and Adaptive Evolution

MGI's stLFR Technology Combined with DNBSEQ Sequencing Platform Facilitate High-quality Assembly of Genomes of Complex Species

The Shenzhen Fairy Lake Botanical Garden and the BGI-Research have jointly completed the study of genomes of Begonia genus and published the results in the journal *New Phytologist*, titled "Genomes shed light on the evolution of Begonia, a mega-diverse genus"[1].

This study, based on multiple long-read sequencing technologies, completed the sequencing and assembly of the whole genome of four species of Begonia genus, better explaining the origin, evolution and shade adaptation of the genus. The genomes of Begonia masoniana and Begonia darthvaderiana were high-quality genomes obtained by sequencing with the library constructed using the stLFR technology of MGI, and were additionally assembled into genomes with relatively high heterozygosity. Some of the sequencing work in this study was completed on the DNBSEQ sequencing platform of MGI.

Recommended application: Long-read DNA sequencing

Recommended models: DNBSEQ-G400RS, DNBSEQ-T7RS

- **More comprehensive long-read DNA information reading**

  The MGIEasy stLFR kit can detect and analyze DNA molecules ranging from 10kb to 300kb in length, while also taking advantage of the benefits of second and third generation sequencing technologies.

- **Excellent performance in various variants calling**

  Various types of variant information can be obtained through a single calling, and the overall variant detection performance reaches the best level compared with conventional whole-genome sequencing with the same depth.

- **Coverage uniformity**

  Good genome coverage uniformity can be obtained with stLFR libraries starting with only 1 ng DNA.

- **High-quality homologous genome alignment**

  More information on highly homologous genomes can be obtained, resulting in better genome alignment results.

- **Providing combinational products for stLFR application**

  The combination of MGI's stLFR technology, automated sample preparation system, and stLFR de novo assembly software can provide customers with a complete product solution.

Background

The Begonia genus is a large genus of plants, with up to 2000 species. The mechanisms of diversification of its clades remain one of the mysteries in plant biology[2]. Among the 10 largest angiosperm genera, only 3 representative ones have completely assembled nuclear genomes published, namely Solanum[3], Dendrobium[4] and Begonia[5]. The Begonia masoniana (Iron Cross Begonia) is a pan-tropical genus that includes over 2000 currently recognized species of herbaceous plants and occasional sub-shrubs. Therefore, it stands for a superb evolutionary study system for processes that give rise to a great number of closely related species.

The Begonia has a high diversity of species in America and Asia, while the number of species in Africa, its speculated origin, is relatively low. Most begonias are shade-tolerant and easily harmed by exposure to sunlight. However, the ability of Begonia species to adapt to varying levels of light, from deep shade to full sun, provides us with the chance to understand the mechanisms underlying these adaptations. Understanding how crops can adapt to shade to optimize photosynthesis and physical defense by utilizing this mechanism of Begonia is of great value. A study compared the genomes of four species of Begonia and reconstructed the paleo-genome of Begonia based on this.

This study analyzed the evolutionary significance of the whole genome duplication event of Begonia. In addition, Brennan et al. analyzed the molecular basis of shade adaptation in Begonia based on cyto-nuclear incongruences[6]. These research sources offer insightful references for the diversity formation and adaptive evolution of the genus Begonia.

Through long-read sequencing and genome assembly, this study provides a more comprehensive evolutionary explanation for the genomic diversity of Begonia. The high-quality genome assembly of the Begonia is a challenging task, but the MGI-developed stLFR (single tube Long Fragment Read) technology offers clear benefits that can aid in completing this challenging task. stLFR is a segmentation-free co-barcoded long fragment reading technology (Figure 1), which combined with the DNBSEQ sequencing technology and can obtain long fragment DNA information through short read long sequencing. With stLFR, researchers could achieve high-accuracy SNP/InDel/CNV/SV variant calling results and haplotype typing results in one sequencing[7]. The stLFR technology provides a more thorough and uniform genome sequencing based on long-read sequencing, which assists in the assembly of Begonia genomes and provides a potent technique for sequencing analysis of the adaptation and evolution mechanisms of the genus.

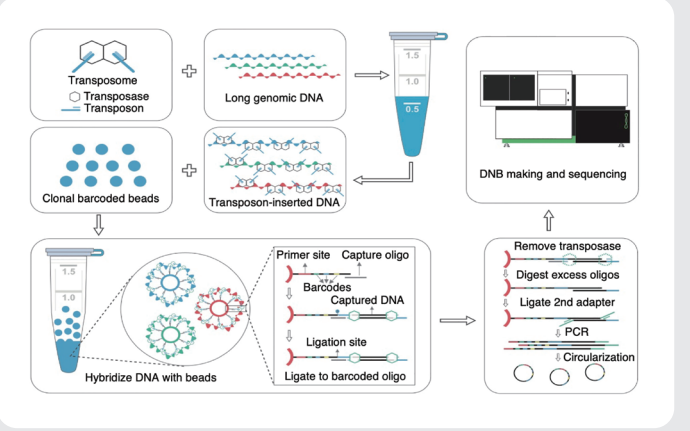

Figure 1. The diagram of stLFR library construction process.

Materials and Methods

Sample Collection and DNA/RNA Extraction

All samples of Begonia were collected from the greenhouse of Shenzhen Fairy Lake Botanical Garden. The genomic DNA of young leaves was extracted using the Cetyltrimethylammonium Bromide (CTAB) method and wait for the following whole-genome sequencing (WGS). The high molecular weight genomic DNA was obtained using a plant tissue DNA isolation kit and wait for the following stLFR sequencing. The mRNA from the root, stem, leaf, peduncle and flower of the four Begonias were extracted and wait for the following transcriptome sequencing.

Library Preparation and Sequencing

The MGIEasy FS DNA Library Prep Set was used to construct the DNA libraries for whole genome sequencing with 300-500 bp fragment sizes, followed by paired-end 150 (PE150) sequencing. RNA library preparation kit was used to construct the transcriptome libraries with inserts of 200-400 bp, followed by paired-end 100 (PE100) sequencing. Each library generated over 5 Gb of sequence data.

The stLFR library was prepared using the MGIEasy stLFR library prep kit, followed by paired-end sequencing with reads of 100bp+ 42bp. Each library generated >150 Gb of raw sequence data (Figure 2). The insert fragment size of the 10x Genomics Chromium genome library is 350-500bp, which was constructed using a related reagent kit and then sequenced with PE150 sequencing on the DNBSEQ platform by introducing appropriate sequencing primers. In addition, some of the data was sequenced based on PacBio. For details of related SMART library construction and sequencing, please refer to the original supplementary materials of this research.

Bioinformatics Analysis

Accurate reads were obtained using a self-developed script for the assembly of sequencing reads from the 10xGenomics and stLFR libraries, and de novo assembly was performed using SPERNOVA. A minimum fast record size of 100 bp was specified during the "output" stage in order to output the assembly in "pseudohap" style. The data obtained from PacBio library sequencing reads was de novo assembled using CANU, and two rounds of iterative correction of PacBio long reads were performed using RACON software, followed by PILON correction. For stLFR-related sequence assembly, stLFR de novo genome assembly plays an important role. After assembly of the genome, the study subsequently conducted variant calling analysis, phylogenetic analysis, genome collinearity analysis, chlorophyll fluorescence analysis, and identification and phylogenetic analysis of light-harvesting Chl a/b-binding proteins superfamily.

Results

Genome Sequencing and Genome Characteristics

This team has assembled the genomes of four species of Begonia, namely Begonia loranthoides, Begonia masoniana, Begonia darthvaderiana, and Begonia peltatifolia. Besides, the shallow genome sequencing of 74 globally representative species had also been finished (Figure 2). According to the results of genome assembly, the genome size of four species ranges from 331 Mb to 799 Mb, and their gene counts range from 22,059 to 23,444. The BUSCO evaluation scores are all above 91%. The high-quality genome assembly is used for functional genomics research on the Begonia, and the advantages of adopting stLFR for long-read de novo assembly is clear.

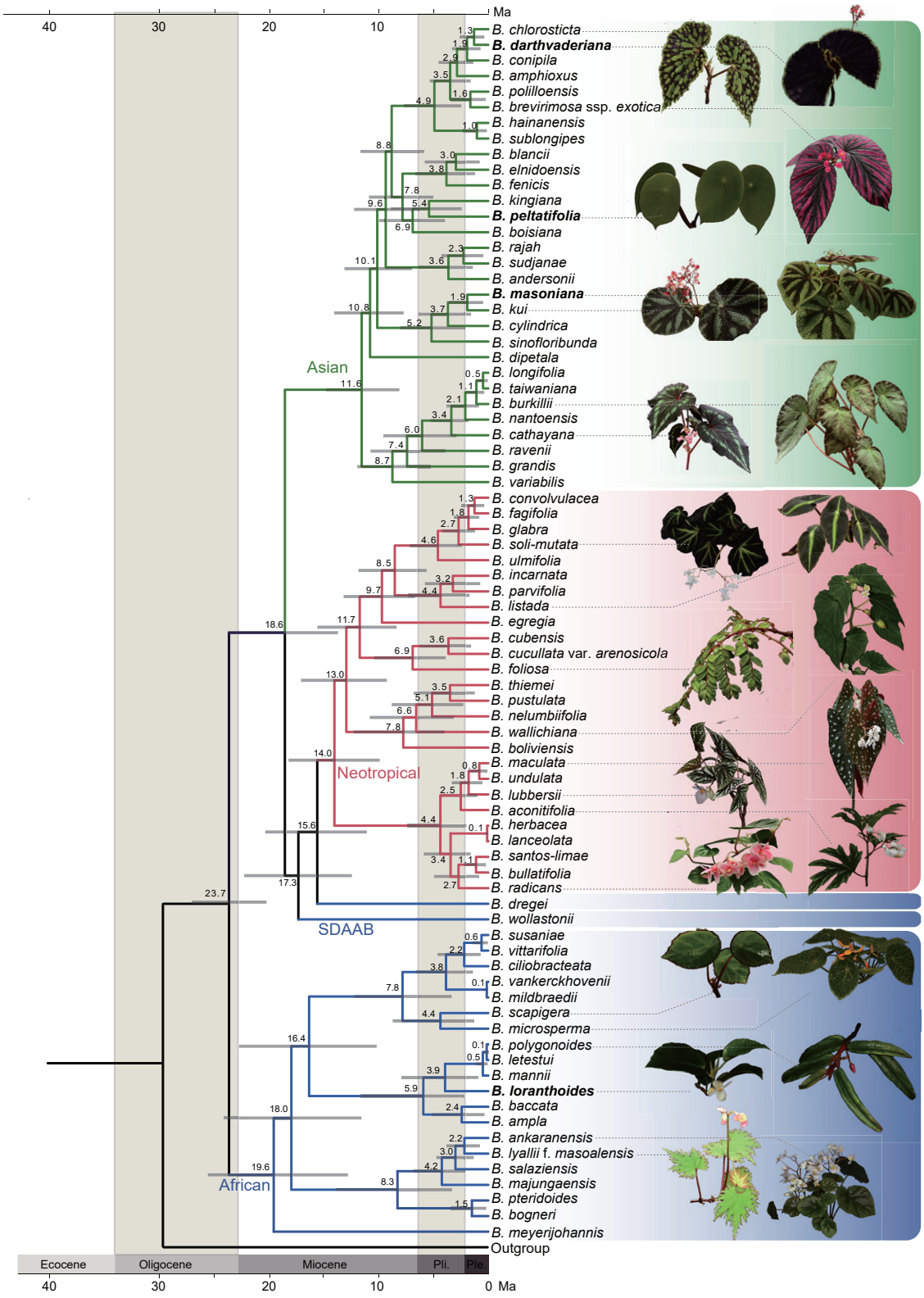

Figure 2. The phylogenetic tree of Begonia plants.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Whole-Genome Duplication Event

Begonia experienced a whole genome duplication (WGD) event about 35 million years ago, which played an important role in the evolution of plant diversity in Begonia, and led to the retention and enrichment of genes in Begonia, retaining multiple copies (Figure 3), while the adaptability of Begonia to the environment also changed during this period.

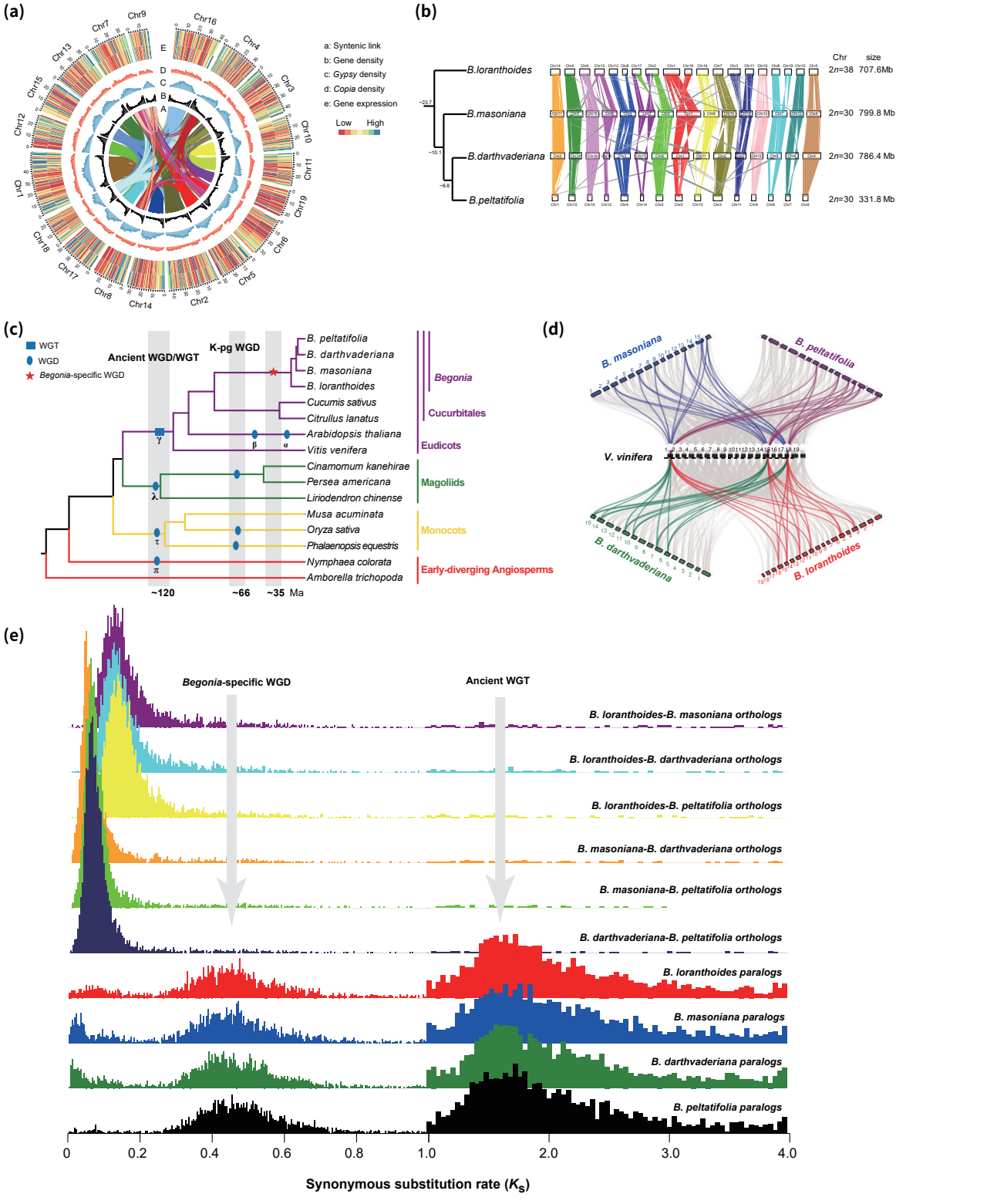

Figure 3. Whole-genome duplication event.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Transposable Elements Evolution and Distribution

Transposable elements account for a large proportion in the genome of Begonia and have high interspecific specificity. The insertion patterns and numbers of transposable elements at different positions of functional gene promoters and introns are significantly different among different species, and hybridization and gene introgression promote the diversification of Begonia. Studies have found that the ancestors of Begonia in the Americas may have had multiple hybridization events, revealing that transposable elements are closely related to the formation and adaptability of Begonia species (Figure 4).

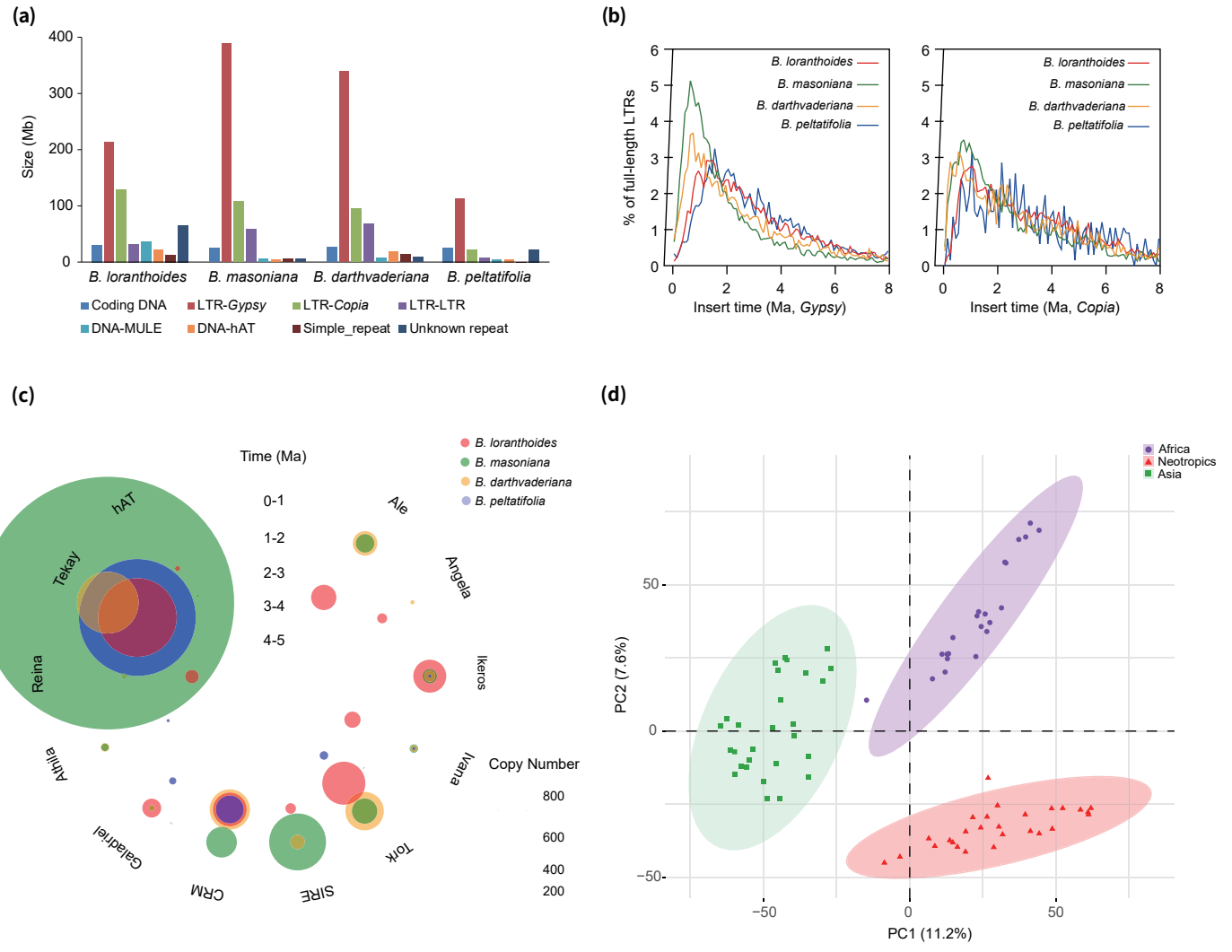

Figure 4. Evolution and distribution of transposable elements.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Evolution of Shade Adaptation

Numerous genes involved in photosynthesis and energy metabolism were retained and enriched by WGD. It was found that many copies of receptor genes (PHOT, CYR1/2, PHY and UVR8) related to the reception of red, blue and ultraviolet light were retained, and some of them had undergone functional differentiation. This is of great significance for Begonia to adapt to shade environment and improve photosynthetic efficiency (Figure 5).

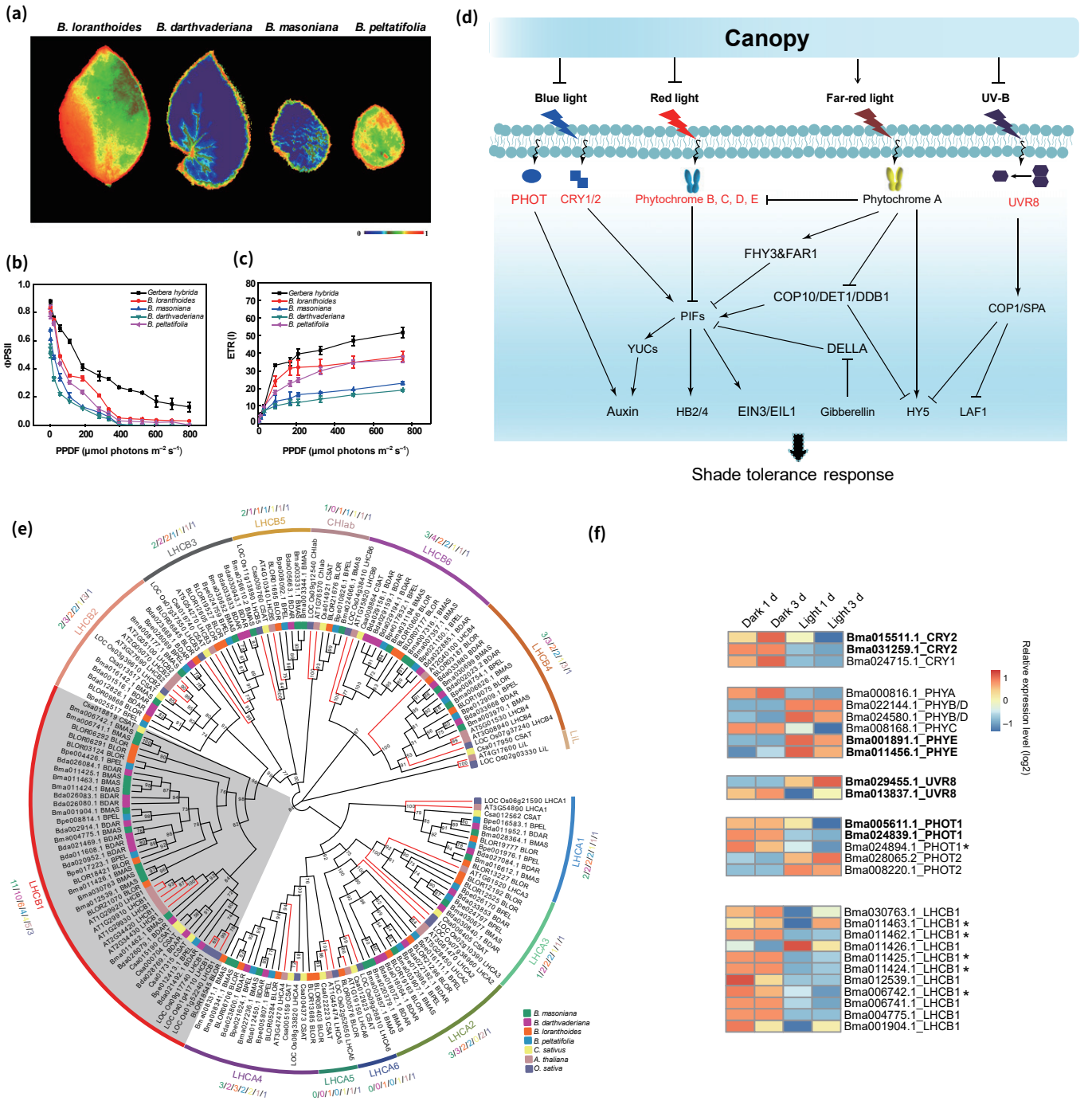

Figure 5. Shade adaptation mechanism in Begonia.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Summary

The findings of this study offer fresh insights into the origins of diversity and adaptive evolution in Begonia, and the stLFR technique employed in the genome assembly of this genus can serve as a technological benchmark for other genomics investigation. Genome sequencing research could used the benefits of MGI's DNBSEQ sequencing platform as a benchmark for long fragment sequencing and high-quality genome assembly.

DNBSEQ-G400 Genetic Sequencer

References

1. Li, L. et al. Genomes shed light on the evolution of Begonia, a mega-diverse genus. *New Phytol* **234**, 295-310, doi:10.1111/nph.17949 (2022).
2. Frodin, D. G. History and concepts of big plant genera. *Taxon* **53**, 753-776, doi:Doi 10.2307/4135449 (2004).
3. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
4. Yan, L. et al. The Genome of Dendrobium officinale Illuminates the Biology of the Important Traditional Chinese Orchid Herb. *Mol Plant* **8**, 922-934, doi:10.1016/j.molp.2014.12.011 (2015).
5. Griesmann, M. et al. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**, doi:10.1126/science.aat1743 (2018).
6. Gommers, C. M., Visser, E. J., St Onge, K. R., Voesenek, L. A. & Pierik, R. Shade tolerance: when growing tall is not an option. *Trends Plant Sci* **18**, 65-71, doi:10.1016/j.tplants.2012.09.008 (2013).
7. Wang, O. et al. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res* **29**, 798-808, doi:10.1101/gr.245126.118 (2019).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-G400RS	900-000170-00
	Genetic Sequencer DNBSEQ-T7RS	900-000128-00
	MGISP-960RS Automated Sample Preparation System	900-000146-00
Software	MegaBOLT Bioinformatics analysis accelerator	900-000555-00
	Data Center Appliance	900-000444-00
	MGI-tech-bioinformatics/stLFR_v1	https://github.com/MGI-tech-bioinformatics/stLFR_v1
Library Prep	MGIEasy stLFR Library Prep Kit (16 RXN)	940-000193-00
	MGIEasy FS DNA Library Prep Set（16 RXN）	1000006987
	MGIEasy RNA Library Prep Set（16 RXN）	1000006383
Sequencing Reagents	DNBSEQ-G400RS High-throughput Sequencing Set (stLFR FCL PE100)	1000016984
	DNBSEQ-T7RS High-throughput Sequencing Set (stLFR FCL PE100)	1000019251
	DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE100)	1000016950
	DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150)	1000016952

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083

The copyright of this brochure is solely owned by MGI Tech Co. Ltd.. The information included in this brochure or part of, including but not limited to interior design, cover design and icons, is strictly forbidden to be reproduced or transmitted in any form, by any means (e.g. electronic, photocopying, recording, translating or otherwise) without the prior written permission by MGI Tech Co., Ltd.. All the trademarks or icons in the brochure are the intellectual property of MGI Tech Co., Ltd. and their respective producers.

Version: December 2023

+86-4000-688-114
en.mgi-tech.com
MGI-service@mgi-tech.com

Authors: Wu Juye, Zhang Jiale
Editor-in-Charge: Wang Qiwei
Reviewer: Jiang Yao

1. For StandardMPS and CoolMPS: Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, Spain, UK, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland, Portugal, Austria and Romania. Unless otherwise informed, StandardMPS sequencing reagents, and sequencers for use with such reagents are not available in Hong Kong. No purchase orders for StandardMPS products will be accepted in the USA until after January 1, 2023.
2. For HotMPS sequencers: This sequencer is only available in selected countries, and its software has been specially configured to be used in conjunction with MGI's HotMPS sequencing reagents exclusively.
3. For HotMPS reagents: This sequencing reagent is only available in selected countries.