Horticulture Research

Horticulture Research, 2025, **12**: uhaf076

https://doi.org/10.1093/hr/uhaf076

Letter to the Editor

Water lily pond: a multiomics database for water lilies

Chengjun Zhao[1,2,‡], Ji Zhang[1,2,‡], Yayu Chen[1,2,‡], Lishuang Yang[1,2,‡], Hongliang Chen[1,2], Yufan Liang[1,2], Wenquan Wang[1,2], Shuang He[1,2], Yunqing Luo[1,2], Junyu Zhang[1,2], Hongbin Zhang[1,2], Shuting Yang[1,2], Guilian Guo[1,2], Wenbai Dai[1,2], Zhijuan Yang[1], Junhao Chen[3], Yuhan Zhou[4], Wasi Ullah Khan[1,2], Guanhua Liu[5], Yifan Jiang[6], Tianlong Zhu[7], Yingchun Xu[6], Pedro García-Caparros[8], Yves Van de Peer[9,10,11,12], Jia-yu Xue[6], Chengjie Chen[13], Liangsheng Zhang[4] and Fei Chen[1,2,*]

[1]National Key Laboratory for Tropical Crop Breeding, College of breeding and multiplication, Sanya Institute of Breeding and Multiplication, Hainan University, Sanya 572025, China
[2]College of Tropical Agriculture and Forestry, Hainan University, Danzhou 571700, China
[3]Department of Biology, Saint Louis University, St. Louis, MO 63103, USA
[4]College of Agricultural & Biotechnology, Zhejiang University, Hangzhou 310085, China
[5]Tea Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310008, China
[6]College of Horticulture, Nanjing agricultural University, Nanjing 210095, China
[7]Hainan Fodu Lianyuan Ecological Agriculture Co. Ltd., Haikou 570105, China
[8]Agronomy Department of Superior School Engineering, University of Almería, 04120 Almeria, Spain
[9]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent 9052, Belgium
[10]Center for Plant Systems Biology, VIB, Ghent 9052, Belgium
[11]Department of Biochemistry, Genetics and Microbiology, Centre for Microbial Ecology and Genomics, University of Pretoria, Pretoria 0028, South Africa
[12]College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing 210095, Nanjing, China
[13]National Key Laboratory for Tropical Crop Breeding, Laboratory of Crop Gene Resources and Germplasm Enhancement in South China, Ministry of Agriculture and Rural Affairs, Key Laboratory of Tropical Crops Germplasm Resources Genetic Improvement and Innovation of Hainan Province, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China
*Corresponding author. E-mail: feichen@hainanu.edu.cn
‡Co-first authors

Dear Editor,

As one of the earliest diverging lineages of angiosperms, water lilies hold unique value for evolutionary studies [1]. Water lilies also hold significant cultural and economic value. For instance, the seeds of *Euryale ferox* are highly valued in traditional Chinese medicine as a starch-rich tonic with major health benefits [2]. Previously, we successfully sequenced and analyzed the first genome water lily (*Nymphaea colorata*), to provide multiple and valuable insights into the early evolution of angiosperms [3]. In recent years, multiomics data on water lilies have been steadily accumulating. Whole-genome sequencing [3–5], transcriptomics, and metabolomics analyses have provided a wealth of foundational resources to support water lily breeding research. However, these advances have also introduced new challenges—chiefly, how to effectively integrate and systematically analyze these diverse datasets in order to develop efficient genetic breeding strategies.

We started in 2018 to construct a database for water lilies, which we termed as Water Lily Pond (WLP). The WLP platform includes 11.14 Gb of genomic data from nine water lily species, encompassing a remarkable 409 321 genes. The transcriptomic data are equally extensive, with 1.2 Tb of *de novo* sequences and 1.235 Tb of reference-guided sequences. Rich annotation datasets—spanning gene families, transcription factors, KEGG pathways, GO terms, and SignalP predictions—comprise an extraordinary 3 034 356 entries. Moreover, the platform offers 141 932 expression profiles, 13 467 proteomic entries, 1841 metabolite data points, and phenotypic image data for 187 distinct species (Fig. 1A). Together, these resources establish WLP as an indispensable database for advancing water lily genomics and beyond. Website of the database: https://bioinformatics.hainanu.edu.cn/waterlily

WLP consists of seven primary modules each targeting a distinct area of research and community engagement: 'Taxonomy', 'Genomics', 'Transcriptomics', 'Proteomics', 'Metabolomics', 'Phenomics', 'Breeding' (Fig. 1A).

The module 'Taxonomy' compiles fundamental information on the order Nymphaeales, providing detailed biological descriptions. This includes comprehensive information on Nymphaeales, as well as the families Hydatellaceae, Cabombaceae, Nymphaeaceae, and fossils. For each genus within these families, an introductory content is provided.

The module 'Genomics' (Fig. 1B) integrates 10 water lily genomes and three main tools. Gene Search tool provides detailed information on gene sequences. BLAST tool suite not only provides the sequence alignment for genomes, genes, proteins, and also for the transcriptome sequences are also available. A Gene Family Search tool was provided for the easy access and comparison across water lilies. A chromosomal Synteny Search tool for checking syntenic relationships across water lilies are provided. A genome browser for visualizing genomic data and various

Received: 16 December 2024; **Accepted:** 24 February 2025; **Published:** 11 March 2025; **Corrected and Typeset:** 1 June 2025

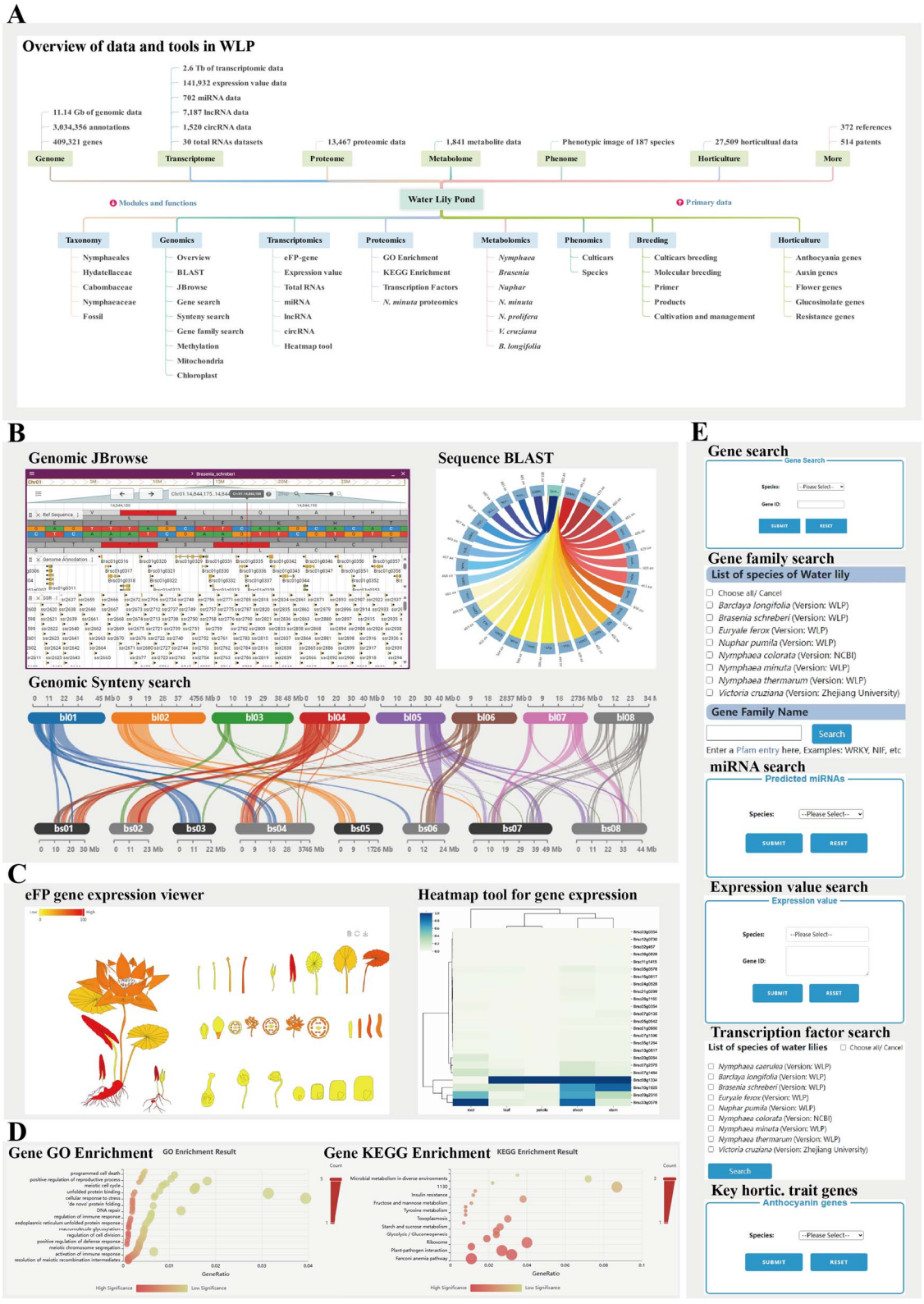

Figure 1. Highlighted data and tools integrated in WLP. (A) The datasets stored in WLP and the integrated tools in this database. (B) The Genomics module includes BLAST, JBrowse, and Synteny Search. (C) The transcriptomics module includes two important tools: eFP-gene tool, and Heatmap tool. (D) The proteomics module includes two critical tools: GO Enrichment, and KEGG Enrichment. (E) The main search tools, including gene search, gene family search, miRNA, expression value, transcription factor, and the Anthocyanin genes tool in the horticulture module.

annotations is integrated in WLP. Furthermore, WLP also provides specialized sections for methylation data under 'Methylation', mitochondrial genome information under 'Mitochondria', and chloroplast genome information under 'Chloroplast', providing a holistic perspective of genomic characteristics across multiple organelles.

WLP collected and sequenced the transcriptome data from 213 water lilies. The module 'Transcriptomics' (Fig. 1C) includes several tools for the visualization and the respective analysis of gene expression data. Notably, this module includes eFP-gene tool for visualizing expression patterns and the Heatmap tool for generating expression heatmaps. The module also features the expression value function for querying expression data across different conditions. Additionally, this module includes statistical analysis of miRNA, lncRNA, and circRNA data from the complete transcriptome of *N. minuta*, enabling in-depth exploration of non-coding RNA profiles within the species.

The module 'Proteomics' (Fig. 1D) provides users with a range of powerful analysis tools and informational resources, including two visualization tools for GO enrichment analysis (GO Enrichment) and KEGG enrichment analysis (KEGG Enrichment). These tools help users visually assess the enrichment of gene and protein functions, thereby revealing key pathways in biological processes. Additionally, the module includes proteomics data for *Nymphaea minuta* (*N. minuta* Proteomics), offering detailed data support for the study of protein characteristics and functions in this species.

The module 'Metabolomics' systematically compiles a rich dataset of metabolites, covering various species within the *Nymphaea*, *Brasenia*, and *Nuphar* genera, and further detailing important species such as *N. minuta*, *N. prolifera*, *Victoria cruziana*, and *Barclaya longifolia*. The module organizes metabolic data from different species into distinct functional categories, allowing efficient querying and navigation. Users can easily browse and analyze the unique metabolic profiles of each species, allowing comparative studies and the identification of species-specific metabolic features.

The module 'Phenomics' encompasses a wealth of visual resources, including flower and leaf phenotypic images for 180 different cultivars, showcasing the morphological diversity within these cultivars. Additionally, the module features a variety of tissue-specific phenotypic images from five wild species, covering different plant parts such as petals, leaves, and stems, providing a comprehensive display of the unique phenotypic characteristics of the original species across various tissue levels. This rich morphological information offers valuable insights into research and breeding applications supporting targeted breeding strategies.

The module 'Breeding' provides users with illustrated resources for common pests and diseases affecting water lilies and also offers comprehensive molecular breeding information to help in the identification and management of plant health issues. The module includes a molecular marker tool, Primer, for designing specific primers to more accurately select target genes in breeding programs.

The module 'Horticulture' systematically integrates detailed data on anthocyanin content, flowering time, floral scent composition, and flower color changes, enabling users to explore these traits in depth.

WLP also includes transcription factor data for nine species, providing critical resources for molecular mechanism studies. Additionally, the Library and Community modules aggregate a wide spectrum of academic resources and communication platforms, including literature, patents, open data, and detailed tutorials, thereby fostering knowledge exchange and supporting research collaboration.

In summary, WLP serves as the first and most comprehensive multiomics database. WLP provides researchers and breeders extensive genomic, proteomic, and transcriptomic data resources. The database not only offers first-hand genomic sequences, annotations, and various omics datasets, but also integrates multiple functional tools, allowing users to easily search and analyze specific genes or proteins, thereby facilitating cross-dimensional data analysis. Currently, the database achieves comprehensive annotation spanning genomics, proteomics, and transcriptomics, offering unique support for the exploration of gene functions, regulatory mechanisms, and protein interaction networks.

With the maturation and widespread adoption of third-generation sequencing technologies, such as PacBio, Oxford Nanopore, along with high-precision data processing tools, such as Hi-C, ATAC-seq, and scRNA-seq, the WLP platform will further offer more refined multidimensional omics data and innovative analytical tools. These new technologies support longer read lengths, lower error rates, and precise sequencing at the single-cell and spatial resolution levels, thereby facilitating research into genomic structural variation, epigenetic modifications, and cell-specific expression profiles. Moving forward, the WLP platform will continue to expand data types in areas such as genome assembly, transcriptome analysis, and epigenome annotation, while integrating efficient analysis tools and intelligent algorithms, enabling users to quickly complete the full analysis pipeline from data acquisition to insights. In addition, by incorporating high-quality visualization and interactive features, WLP will significantly enhance the depth and research value of multidimensional omics data, providing precise and comprehensive data support for plant biology and biodiversity research.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 32172614), Hainan Province Science and Technology Special Fund (no. ZDYF2023XDNY050), Hainan Provincial Natural Science Foundation of China (no. 324RC452), and The Project of National Key Laboratory for Tropical Crop Breeding (no. NKLTCB202337).
No conflict of interest is declared.

Author Contributions

Fei Chen conceived and designed the study, while Chengjie Chen, Liangsheng Zhang, Jiayu Xue, Feng Chen, Junhao Chen, Yves van de Peer participated in database design. Chengjun Zhao lead the data analysis and database construction. Yayu Chen, Lishuang Yang, Shuang He, Yunqin Luo participated in the database construction, code implementation, tool development, and data organization. Ji Zhang, Yufan Liang, Yuhan Zhou, Wenquan Wang, Shuting Yang, Hongliang Chen, Guilian Guo, Wenbao Dai, Guanhua Liu, Yifan Jiang, and others contributed to the provision and organization of genomic data, as well as downstream analysis and server deployment. Tianlong Zhu, Yinchun Xu, Hongbin Zhang, Zhijuan Yang, and Wasi Ullah Khan handled the collection of sequencing materials. Chengjun Zhao, Fei Chen, and Pedro García-Caparros wrote the draft MS. All the authors approved the final MS. We thank Zong-Ming (Max) Cheng for his valuable help on the database initiation, which started in Nanjing Agricultural University.

Data availability

The data generated in this study have been released in the National Genomics Data Center, with GSA numbers: CRA019059 and CRA018961. The small RNA data can be accessed under GSA numbers: CRA019696 and CRA019738.

Conflict of interest statement

None declared.

Supplementary Data

Supplementary data is available at *Horticulture Research* online.

References

1. Chen F, Liu X, Yu C. *et al.* Water lilies as emerging models for Darwin's abominable mystery. *Hortic Res-England*. 2017;**4**:17051
2. Abelti AL, Teka TA, Bultosa G. Review on edible water lilies and lotus: future food, nutrition and their health benefits. *Appl Food Res*. 2023;**3**:100264
3. Zhang L, Chen F, Zhang X. *et al.* The water lily genome and the early evolution of flowering plants. *Nature*. 2020;**577**:79–84
4. Povilus RA, DaCosta JM, Grassa C. *et al.* Water lily (*Nymphaea thermarum*) genome reveals variable genomic signatures of ancient vascular cambium losses. *Proc Natl Acad Sci*. 2020;**117**:8649–56
5. Yang Y, Sun P, Lv L. *et al.* Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat Plants*. 2020;**6**:215–22

Horticulture Research, 2025, **12**: uhaf076
https://doi.org/10.1093/hr/uhaf076
Letter to the Editor