Horticulture Research

Horticulture Research, 2024, **11**: uhae342

https://doi.org/10.1093/hr/uhae342

Letter to the Editor

plantGIR: a genomic database of plants

Zhuo Liu[1,‡], Chenhao Zhang[1,‡], Jinghua He[1], Chunjin Li[1], Yanhong Fu[1], Yongfeng Zhou[2,3], Rui Cao[1,*], Haibin Liu[1,*] and Xiaoming Song[1,*]

[1]School of Life Sciences/School of Basic Medical Sciences/ Key Laboratory For Quality of Salt Alkali Resistant TCM of Hebei Administration of TCM, North China University of Science and Technology, Tangshan, Hebei 063210, China
[2]National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518000, China
[3]National Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China
*Corresponding authors. E-mail: songxm@ncst.edu.cn; liuhaibinhappy1013@126.com; caoruivv@163.com
‡The authors contributed equally to the work.

Downloaded from https://academic.oup.com/hr/article/11/12/uhae342/7916742 by guest on 11 February 2026

Dear Editor,

With the advancement of genomic sequencing technology and the continuous reduction in costs, an increasing number of species have undergone whole-genome sequencing, which provides a wealth of resources for comparative genomics and functional genomics research [1–3]. In the field of plant genomics and bioinformatics research, the standardization, uniformity, and accessibility of genomic data are extremely important. However, the diversity of data submission pathways has resulted in numerous data download channels. Consequently, it is challenging to ensure the accuracy and standardized consistency of the genomic data. Furthermore, some databases are no longer accessible for use. Typically, genomic data for individual species are scattered across a multitude of databases, with sequence IDs being prone to change during the transfer process between these repositories. Although numerous databases exist, each focusing on specific families, genera, or taxonomic categories [4–7], there remains a void for a comprehensive repository that includes genomic data for all sequenced plant species.

Here, we have meticulously curated the original download links from 797 publications, amassing a comprehensive dataset from 110 diverse databases, encompassing 1117 plant species. To facilitate the use of these genomic data resources by a wider range of relevant researchers, we have developed the plant genomic information resources (plantGIR, http: //plantgir.cn/) data-sharing platform, which is a treasure trove of genomic insights into these 1117 species from lower to higher plants. The database mainly included three principal genomic file formats, including coding sequences, protein sequences, and general feature format files.

Moreover, we have conducted a thorough identification of genes associated with flowering, auxin, and anthocyanin biosynthesis, along with transcription factors. This dataset deposited in plantGIR serves as a vital foundation for advancing genomic studies and enhancing our understanding of plant genetics. Transcription factor (TF) families were identified using PlantTFDB website (https://planttfdb.gao-lab.org). The flowering, anthocyanin, and auxin genes were identified by BLASTP (*E*-value <1e−5; identity >60%; score > 150) with *Arabidopsis thaliana* according to the previous reports, with manual verification for accuracy [8–10]. Finally, a total of 200 906 flowering, 60 438 auxin, and 42 626 anthocyanin biosynthesis genes were identified in 1117 plant genomes. The plantGIR database contains 45 792 677 gene annotations obtained using the Gene Ontology, nonredundant, Swiss-Prot, and TrEMBL databases. Moreover, 2 206 412 transcription factors were detected in all 1117 plants, which was further divided into 59 gene families. Among these transcription factors, there was the most number for basic helix-loop-helix (bHLH) (191862) gene family, followed by myeloblastosis (MYB) (167682) and ethylene-responsive factor (ERF) (166880) gene families.

On plantGIR databse, we present an intuitive interface for showcasing functional genes and transcription factors in the introduction section on the homepage (Fig. 1A), with clickable links that lead to detailed display pages (Fig. 1C-E). Our advanced fuzzy search enables users to input just the initial part of a term, prompting the database to reveal corresponding species, complete with taxonomic details and a statistic of species within each category (Fig. 1A). By clicking on these numbers, users can comprehensively view of all species within that category, including literature references, while clicking on a species name transports them to an elaborate information display, offering rich data and analytical insights specific to that species (Fig. 1B). We also employ a dynamic word cloud to spotlight the top ten most-accessed species (Fig. 1A), selecting these leads to a data-monitoring hub that illustrates the top 20 species in terms of search and download frequency (Fig. 1K). We display the provenance of species data, detailing the publication year and the original databases obtaining the species information. Additionally, we provide easy access to all associated publication articles, streamlining the user's quest

Received: 13 August 2024; **Accepted:** 27 November 2024; **Published:** 5 December 2024; **Corrected and Typeset:** 20 December 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of Nanjing Agricultural University. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Acknowledgments

This work was supported by the Natural Science Foundation for Distinguished Young Scholars of Hebei (C2022209010), National Natural Science Foundation of China (32172583), Tangshan Science and Technology Plan Project (24130219C), the National Key Research and Development Program of China (2023YFF1002000), the Medical-engineering Integration Project of North China University of Science and Technology (ZD-YG-202411), and the S&T Program of Hebei (23372505D).

Author Contributions

X.S. conceived the project and was responsible for the initiation of the project. X.S., H.L., R.C., and Z.L. supervised and managed the project and research. Data collection and bioinformatics analysis were performed by X.S., Z.L., C.L., H.L., R.C., and Y.F. Database construction was led by X.S., Z.L., and J.H. The manuscript was organized, written, and revised by X.S., Z.L., H.L., R.C., Y.Z., and C.Z All the authors read and revised the manuscript.

Data Availability

All materials and data related to this study are available in our plantGIR database (http://plantgir.cn).

Conflict of interest statement

The authors declare no competing interests.

References

1. Xie L, Gong X, Yang K. *et al.* Technology-enabled great leap in deciphering plant genomes. *Nat Plants*. 2024;**10**:551–66
2. Song X. The Orychophragmus violaceus genome. *Vegetable Research*. 2023;**3**:1–2
3. Yue J, Chen Q, Wang Y. *et al.* Telomere-to-telomere and gap-free reference genome assembly of the kiwifruit *Actinidia chinensis*. *Hortic Res*. 2023;**10**:uhac264
4. Liu Z, Li N, Yu T. *et al.* The Brassicaceae genome resource (TBGR) A comprehensive genome platform for Brassicaceae plants. *Plant Physiol*. 2022;**190**:226–37
5. Yu T, Ma X, Liu Z. *et al.* TVIR: a comprehensive vegetable information resource database for comparative and functional genomic studies. *Hortic Res*. 2022;**9**:uhac213
6. Yu J, Wu S, Sun H. *et al.* CuGenDBv2: an updated database for cucurbit genomics. *Nucleic Acids Res*. 2023;**51**: D1457–64
7. Liu Z, Shen S, Li C. *et al.* SoIR: a comprehensive Solanaceae information resource for comparative and functional genomic study. *Nucleic Acids Res*. 2024;gkae1040
8. Wu T, Liu Z, Yu T. *et al.* Flowering genes identification, network analysis and database construction for 837 plants. *Hortic Res*. 2024;**11**:uhae013
9. Feng S, Liu Z, Chen H. *et al.* PHGD: an integrative and user-friendly database for plant hormone-related genes. *iMeta*. 2024;**3**:e164
10. Chen H, Wang T, He X. *et al.* BRAD V3.0: an upgraded Brassicaceae database. *Nucleic Acids Res*. 2022;**50**:D1432–41

Downloaded from https://academic.oup.com/hr/article/11/12/uhae342/7916742 by guest on 11 February 2026

© The Author(s) 2024. Published by Oxford University Press on behalf of Nanjing Agricultural University. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Horticulture Research, 2024, **11**: uhae342
https://doi.org/10.1093/hr/uhae342
Letter to the Editor