Gene annotation pipeline

Gene annotation was performed for 24 *Asteraceae* species lacking gene annotation data using Braker3 (1). This tool integrates *ab initio* gene prediction, homology-based protein prediction using sequences from *Arabidopsis thaliana* (TAIR10) (https://www.arabidopsis.org/), sunflower (HanXRQr2.0-SUNRISE) (2), cultivated *Chrysanthemum* (Chrysanthemum\_x\_morifolium\_Ramat\_Zszgv0) (3), and lettuce (Lsat\_Salinas\_v11) (4), as well as RNA-seq data with a mapping rate exceeding 70% across diverse tissues and treatments (Supplementary Table S1) (5). The completeness of protein-coding annotations was evaluated using BUSCO v4.0.5 (6) , and results are detailed in Supplementary Table S2. For ease of access, a uniform gene identifier (AMID) was assigned to genes in General Feature Format (GFF) files across 68 collected and predicted genomes.

Functional annotation of protein-coding genes

To enhance genome annotation completeness, we conducted unified functional annotation for each genome. Homologous genes from *Asteraceae* species were identified using a sequence similarity-based approach previously described (7). Specifically, protein sequences from genes in *Arabidopsis* and sunflower (2) were used as queries. BLASTP (2.10.0+) searches were performed against protein sequences of each *Asteraceae* species under the conditions of E-value < 1e-5 and identity > 50% (7). Subsequently, all protein sequences from *Asteraceae* species were reciprocally searched against protein sequences from *Arabidopsis* and sunflower. For protein sequences matching those from *Arabidopsis* and sunflower, only the best alignment sequence was kept. Based on the results from these searches, homologous genes in *Asteraceae* species were extracted.

Transcription factor (TF) in 68 *Asteraceae* genomes were predicted using PlantTFDB v5.0 (8) and iTAK (9). Amino acid sequences were uploaded to PlantTFDB and iTAK database for analysis. Gene families in the *Asteraceae* genomes were identified based on corresponding gene family members in *Arabidopsis* (https://www.arabidopsis.org). To identify conserved motifs, protein sequences from *Asteraceae* species were subjected against HMMER models using HMMER v3.2.1 hmmsearch (10) with an E-value threshold of 1e-5. Protein sequences from each pair of genomes were compared using Diamond v.0.9.14.115 (11). Subsequently, gene collinearity was detected using the McScan (Python version) (12). Visualization of collinearity results for specific species and regions were achieved using ShinySyn (13) to facilitate user access.

Ortholog groups among the *Asteraceae* genomes

The 74 species were used to construct a distance-based phylogenetic tree using JolyTree with default parameters (14). The generated tree was saved in Newick format by default and then visualized using the ggtree R package (15). We selected genomes from 43 *Asteraceae* species to establish a robust pan-genome based on two criteria: i) BUSCO completeness scores greater than 80% (16), and ii) the use of RNA-seq data for predicting gene structures. Drawing from pan-genome construction approaches in Bambusoideae (17), rice (5), wild grape (18), and poplar (19), we developed a gene-based pan-genome for the *Asteraceae* (18). Protein sequences from genes in 43 *Asteraceae* genomes were collected and analyzed using OrthoFinder v2.5.4 (20) with default parameters to identify ortholog groups, which are clusters of genes. These ortholog groups includes three categories: core, dispensable, and species-specific gene clusters. Core gene cluster are genes present in at least 39 genomes (>90% genomes) (21), species-specific gene cluster occur only in one species, and the remaining gene cluster found in 2 to 38 species are classified as dispensable gene cluster.

1. Gabriel, L., Brůna, T., Hoff, K.J., Ebel, M., Lomsadze, A., Borodovsky, M. and Stanke, M. (2024) BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research*, **34**, 769-777.

2. Badouin, H., Gouzy, J., Grassa, C.J., Murat, F., Staton, S.E., Cottret, L., Lelandais-Brière, C., Owens, G.L., Carrère, S., Mayjonade, B. *et al.* (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, **546**, 148-152.

3. Song, A., Su, J., Wang, H., Zhang, Z., Zhang, X., Van de Peer, Y., Chen, F., Fang, W., Guan, Z., Zhang, F. *et al.* (2023) Analyses of a chromosome-scale genome assembly reveal the origin and evolution of cultivated chrysanthemum. *Nature Communications*, **14**, 2021.

4. Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., Xia, L., Froenicke, L., Lavelle, D.O., Truco, M.-J. *et al.* (2017) Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nature Communications*, **8**, 14953.

5. Shang, L., Li, X., He, H., Yuan, Q., Song, Y., Wei, Z., Lin, H., Hu, M., Zhao, F., Zhang, C. *et al.* (2022) A super pan-genomic landscape of rice. *Cell Research*, **32**, 878-896.

6. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210-3212.

7. Jia, Q., Brown, R., Kollner, T.G., Fu, J., Chen, X., Wong, G.K., Gershenzon, J., Peters, R.J. and Chen, F. (2022) Origin and early evolution of the plant terpene synthase family. *Proceedings of the National Academy of Sciences of the United States of America*, **119**, e2100361119.

8. Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J. and Gao, G. (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*, **45**, D1040-D1045.

9. Zheng, Y., Jiao, C., Sun, H., Rosli, Hernan G., Pombo, Marina A., Zhang, P., Banf, M., Dai, X., Martin, Gregory B., Giovannoni, James J. *et al.* (2016) iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Molecular Plant*, **9**, 1667-1670.

10. Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*, **39**, W29-37.

11. Buchfink, B., Xie, C. and Huson, D.H. (2014) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**, 59-60.

12. Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H. *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*, **40**, e49-e49.

13. Xiao, Z., Lam, H.-M. and Marschall, T. (2022) ShinySyn: a Shiny/R application for the interactive visualization and integration of macro- and micro-synteny data. *Bioinformatics*, **38**, 4406-4408.

14. Criscuolo, A. (2019) A fast alignment-free bioinformatics procedure to infer accurate distance-based phylogenetic trees from genome assemblies. *Research Ideas and Outcomes*, **5**, e36178.

15. Yu, G., Lam, T.T., Zhu, H. and Guan, Y. (2018) Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol*, **35**, 3041-3043.

16. Raghavan, V., Kraft, L., Mesny, F. and Rigerte, L. (2022) A simple guide to de novo transcriptome assembly and annotation. *Briefings in Bioinformatics*, **23**.

17. Liu, Y.-L., Gao, S.-Y., Jin, G., Zhou, M.-Y., Gao, Q., Guo, C., Yang, Y.-Z., Niu, L.-Z., Xia, E., Guo, Z.-H. *et al.* (2024) BambooBase: A comprehensive database of bamboo omics and systematics. *Molecular Plant*, **17**, 682-685.

18. Cochetel, N., Minio, A., Guarracino, A., Garcia, J.F., Figueroa-Balderas, R., Massonnet, M., Kasuga, T., Londo, J.P., Garrison, E., Gaut, B.S. *et al.* (2023) A super-pangenome of the North American wild grape species. *Genome Biology*, **24**, 290.

19. Shi, T., Zhang, X., Hou, Y., Jia, C., Dan, X., Zhang, Y., Jiang, Y., Lai, Q., Feng, J., Feng, J. *et al.* (2024) The super-pangenome of Populus unveils genomic facets for its adaptation and diversification in widespread forest trees. *Molecular Plant*, **17**, 725-746.

20. Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**, 238.

21. Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R. *et al.* (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. *Nat Plants*, **6**, 34-45.