Plant Biotechnology Journal

WILEY | aab SEB

BRIEF COMMUNICATION OPEN ACCESS

Parental Specific Expression by scRNA-Seq Reveals *HaWRKY40* Contributing to Heterosis of Sunflower

Yuliang Han[1] | Juncheng Zhang[1] | Maohong Cai[1] | Siqi Zhang[1] | Wenting Dai[2] | Zhonghua Lei[3] | Qixiu Huang[3] | Lijun Xiang[3] | Tao Chen[1,4]

[1]College of Life and Environmental Science, Hangzhou Normal University, Hangzhou, China | [2]Genedenovo Biotechnology Co. Ltd, Guangzhou, China | [3]Institute of Economic Crops, Xinjiang Academy of Agricultural Sciences, Urumqi, China | [4]Advanced Computing Center, Hangzhou Normal University, Hangzhou, China

Correspondence: Tao Chen (chentao@hznu.edu.cn)

Received: 14 May 2025 | **Revised:** 21 June 2025 | **Accepted:** 1 July 2025

Funding: This study was supported by Interdisciplinary Research Project of Hangzhou Normal University (Grant 2025JCXK01). The Starting Research Fund from Hangzhou Normal University (2022QDL034).

Keywords: heterosis | metabolomics | single cell RNA sequencing | sunflower

Sunflower (*Helianthus annuus* L.) is the world's second-largest hybrid crop after maize, holding significant agricultural value due to its versatile applications in food production, oil extraction, and ornamental horticulture. While heterosis has substantially enhanced sunflower yield and stress resistance, the molecular mechanisms underlying this phenomenon remain elusive (Hübner et al. 2019). We addressed this gap by integrating single-cell RNA sequencing (scRNA-seq) and metabolomic profiling to construct a high-resolution transcriptomic atlas across two parental lines and their hybrid progeny for the first time in sunflowers. This work reveals the effects of heterosis on the cellular level of sunflower and offers a framework for precision breeding optimisation in hybrid crops.

Our study focused on sunflower inbred lines 2399B, 2399R, and their hybrid descendant 2399F. We observed that 2399F outperformed its parent in all key indicators at both the seedling stage and the mature stage, showing obvious heterosis (Figure S1). A total of 27 719 leaf cells were initially captured following cell wall digestion, with 33 116, 35 802, and 35 110 genes detected across the three leaf samples, respectively (Figure 1a; Table S1). These cells were classified into 14 clusters (Figure 1b–e; Table S2). We identified a total of seven cell types, including epidermal cells, mesophyll cells, meristematic cells, xylem cells, bundle sheath cells, guard cells, and unknown cells (Figure 1f). Following the identification of cell subpopulations, we designated the five genes exhibiting the highest expression levels within each cluster as potential novel marker candidates (Figure S2; Table S3). These genes will serve as valuable references for future cluster identification in sunflower.

Statistical analysis revealed that mesophyll cells were predominant in inbred lines 2399B and 2399R, followed by epidermal cells, whereas the hybrid 2399F exhibited an inverted distribution (Figure S3). Notably, this cellular redistribution was particularly noteworthy given mesophyll cells' central role in photosynthesis (Zhang and Ambrose 2022), suggesting that 2399F may exhibit enhanced photosynthetic efficiency and metabolic activity. Transcriptomic profiling uncovered significant metabolic reprogramming in the hybrid. KEGG enrichment analysis demonstrated that differentially expressed genes in 2399F were primarily enriched in metabolic pathways, including those for carbohydrates, energy, lipids, and amino acids, as well as protein translation pathways (Figure S4A–C). Specifically, critical genes associated with carbon metabolism, such as *MD1*, *ASD*, and *SS* (Chatterjee et al. 2023; Schmölzer et al. 2016; Zhang et al. 2025), showed significantly elevated transcriptional activity in 2399F. Furthermore, we detected upregulated expression of genes involved in ribosome biogenesis, including ribosomal large subunit (RPL) and small subunit

Yuliang Han and Juncheng Zhang contributed equally to this work.

(RPS) genes in 2399F (Figure S4D–G). Quantification of soluble protein content revealed that 2399F leaves exhibited 8 mg/g of soluble proteins, significantly higher than those in 2399B and 2399R (Figure S5). Pseudo-time analysis delineated nine distinct cell states across the developmental trajectory of mesophyll cells. Specifically, parental lines 2399B mesophyll cells were predominantly localised to branch node 1, and 2399R mesophyll cells mainly at branch node 2–4. In contrast, 2399F mesophyll cells spanned all pseudo-time branches, indicating a broader developmental plasticity in the hybrid (Figure 1g–i). Given the critical role of transcription factors (TFs) in orchestrating cell development and differentiation (Deng et al. 2024), we systematically identified TFs governing mesophyll cell differentiation dynamics. A total of 121 core transcription factors were identified from differentially expressed genes (DEGs) associated with pseudo-time trajectories, differentiation status, and cell fate (Figure 1j; Tables S4–S6). These TFs were assigned into 22 gene families, with the ERF ($n = 30$) and WRKY ($n = 18$) families being most represented (Figure S6; Table S7). Functional annotation revealed that these genes were predominantly involved in

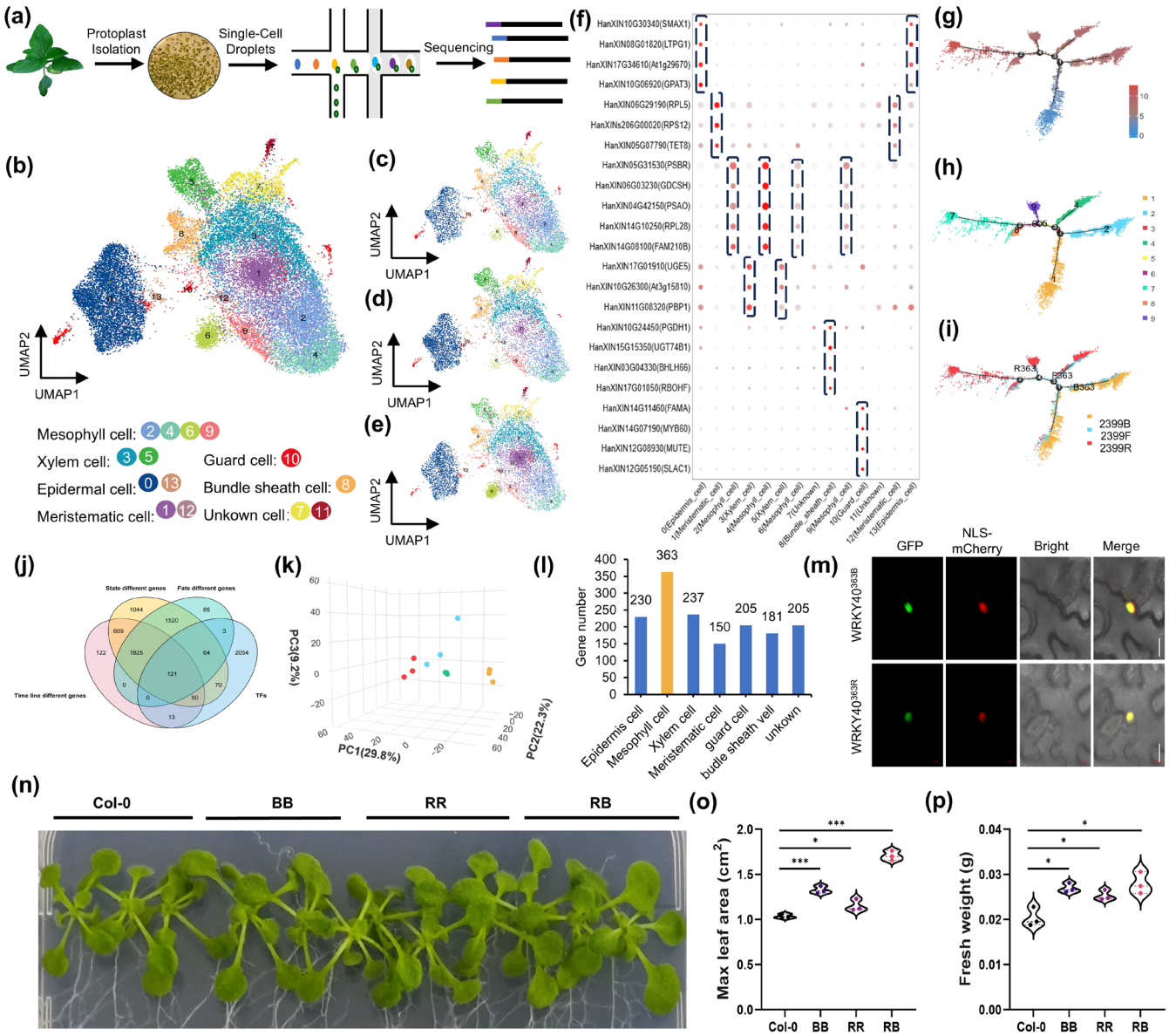

FIGURE 1 | Single-cell atlas and heterosis mechanisms in sunflower. (a) ScRNA-seq workflow of sunflower leaves. (b–e) UMAP visualisation of 14 cell clusters across all samples: (b) all samples combined, (c) 2399B, (d) 2399F and (e) 2399R. (f), The expression pattern of known marker genes in the identified cluster. (e–g), Pseudo-time analysis illustrating (e) the developmental trajectory, (f) differentiation state and (g) sample information along the progression of mesophyll cell development. (h) Identification of core TFs based on the profiles of developmental trajectory, cell states and cell fate. (k) Three-dimensional PCA cluster analysis. Red, blue and yellow represent 2399B, 2399F and 2399R respectively. The QC samples (green dot) was used to visualise the systematic error in the inspection analysis. Three replicates were set for each sample. (l) The number of highly heterozygous genes in different cell types. (m) Subcellular location of WRKY40^{2399B} and WRKY40^{2399R} in tobacco leaves with NLS-mCherry serving as a nuclear marker. Scale bars = 20 μm. (n) Growth phenotype of transgenic *Arabidopsis* after 2 weeks on 1/2 MS culture medium. (o, p) Max leaf area (o) and fresh weight (p) Max leaf area and fresh weight statistics of Col-0 and *HaWRKY40* overexpressed lines. Student's *t*-test was used to determine *p* values. $^{*}p < 0.05$; $^{***}p < 0.001$.

metabolism regulation and biological processes, with a subset linked to stimulus response pathways (Figure S7). Collectively, these findings suggest that metabolic reprogramming critically influences developmental trajectory establishment, potentially amplifying the plasticity of cell differentiation processes.

To elucidate the metabolic basis of heterosis, we performed comprehensive metabolic profiling of parental and hybrid lines. Principal component analysis (PCA) of leaf metabolites revealed clear separation among 2399B, 2399R, and 2399F, demonstrating distinct metabolic signatures in the hybrid (Figure 1k). Metabolite set enrichment analysis (MSEA) identified that these differential metabolites are mainly involved in a variety of important plant metabolic pathways including galactose metabolism, pyruvate metabolism, fatty acid synthesis metabolism, and trehalose metabolism (Figure S8). These data suggested that hybridization induces extensive metabolic rewiring, providing a biochemical foundation for heterosis phenotypes.

Consistent with established effects of hybridization on transcriptional regulation (Hochholdinger and Yu 2024), we systematically analysed allele-specific expression patterns across cell types in 2399F. Through comprehensive evaluation of 2399F heterozygous SNP and indel sites, we identified 6585 genes exhibiting significant allelic expression bias in all cell types (Table S8). Notably, 1571 genes demonstrated strong heterozygosity maintenance during cell replication and division, rather than becoming homozygous for either the paternal or maternal allele. These were termed high heterozygous genes (HHGs). Mesophyll cells contained the highest HHG burden (Figure 1l), with functional enrichment analysis revealing their predominant involvement in processes related to ribosome biogenesis, protein translation, energy metabolism, carbohydrate processing, and others (Figure S9A,B). This included 43 ribosomal genes (17 RPS, 26 RPL) and 13 core carbon metabolism enzymes (Figure S9C–E). Strikingly, six key transcription factors (*ATHB-7*, *NAC081*, *WRKY40*, *AZF3*, *C2H2-1*, and *ZAT7*) emerged as HHGs, suggesting their heterozygous state may regulate broader metabolic networks (Figure S9F). Collectively, these findings supported a model where heterozygosity maintenance in ribosomal and metabolic genes enhances functional diversity, potentially driving hybrid vigour through proteomic and energetic optimization.

In our investigation of key transcription factors, *HaWRKY40* emerged as a pivotal regulator influencing both the pseudo-time trajectory of cell development and heterozygosity maintenance in mesophyll cell. To elucidate the function of HaWRKY40, we cloned its variants, $HaWRKY40^{2399B}$ and $HaWRKY40^{2399R}$, identifying seven SNPs that resulted in two amino acid changes (Figure S10A). Both $HaWRKY40^{2399B}$ and $HaWRKY40^{2399R}$ variants exhibited canonical nuclear localization, confirming their functional role as transcription factors (Figure 1m). To mimic the intracellular heterozygous state, we constructed dual overexpression vectors and developed the corresponding transgenic *Arabidopsis* lines (Figure S10B). The transgenic lines were then cultured on 1/2 MS medium for 2 weeks, demonstrated superior leaf size, root length, lateral root number, fresh weight and dry weight of all OE lines surpassed those of the Col-0. Notably, the heterozygous dual overexpression line consistently outperformed both homozygous transgenics, demonstrating a clear heterozygote advantage (Figure 1n–p; Figure S10C). These results provide direct evidence that allelic interactions at the HaWRKY40 locus contribute to heterosis phenotypes.

In summary, this study establishes the first comprehensive single-cell transcriptomic atlas of sunflower, providing unprecedented resolution of heterosis mechanisms at cellular and molecular levels. Our multi-omics framework not only advances fundamental understanding of heterosis but also provides a blueprint for targeted crop improvement through precision breeding strategies.

Author Contributions

The idea and overall design were proposed by T.C. and J.Z. Most of the experiments were performed by Y.H. M.C., Z.L., Q.H., S.Z. and L.X. participated in some of the experiments. W.D. assisted in some bioinformatics analyses. The article was written by Y.H. and J.Z., and revised by T.C.

Acknowledgements

This study was supported by the Starting Research Fund from Hangzhou Normal University (Grant 2022QDL034) and the Interdisciplinary Research Project of Hangzhou Normal University (2025JCXK01).

Data Availability Statement

The metabolites spectral data have been deposited to the National Genomics Data Center (NGDC). BioProject:PRJCA036408.

References

Chatterjee, Y., B. Bhowal, K. J. Gupta, A. Pareek, and S. L. Singla-Pareek. 2023. "Lactate Dehydrogenase Superfamily in Rice and *Arabidopsis*: Understanding the Molecular Evolution and Structural Diversity." *International Journal of Molecular Sciences* 24: 5900.

Deng, Q., P. Du, S. S. Gangurde, et al. 2024. "ScRNA-Seq Reveals Dark- and Light-Induced Differentially Expressed Gene Atlases of Seedling Leaves in *Arachis hypogaea* L." *Plant Biotechnology Journal* 22: 1848–1866.

Hochholdinger, F., and P. Yu. 2024. "Molecular Concepts to Explain Heterosis in Crops." *Trends in Plant Science* 24: S1360–S1385.

Hübner, S., N. Bercovich, M. Todesco, et al. 2019. "Sunflower Pan-Genome Analysis Shows That Hybridization Altered Gene Content and Disease Resistance." *Nature Plants* 5: 54–62.

Schmölzer, K., A. Gutmann, M. Diricks, T. Desmet, and B. Nidetzky. 2016. "Sucrose Synthase: A Unique Glycosyltransferase for Biocatalytic Glycosylation Process Development." *Biotechnology Advances* 34: 88–111.

Zhang, L., and C. Ambrose. 2022. "CLASP Balances Two Competing Cell Division Plane Cues During Leaf Development." *Nature Plants* 8: 682–693.

Zhang, Z., A. Xu, Y. Bai, et al. 2025. "A Subcellular Map of Translational Machinery Composition and Regulation at the Single-Molecule Level." *Science* 387: eadn2623.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.