RNA-seq analysis and gene regulatory network construction

After removing adapter sequences and low-quality reads using fastp (v.0.23.0) (1), clean reads were aligned to the reference genome of each species using Hisat2 (v.2.1.2) (2) with default parameters. Expression quantification was performed using StringTie (v.1.3.5) (3) with default settings, and gene expression levels were normalized using transcripts per kilobase million reads (TPM). Raw gene expression counts were generated using featureCounts (v1.6.4) (4). Differential gene expression analysis was conducted using the Bioconductor package DESeq2 (5) in R (v4.3.1). Genes with a log2-tranformed fold change ≥ 1 or ≤ −1 and a False Discovery Rate (FDR) ≤ 0.05 were considered differential expressed genes (DEGs). Tissue-specific gene expression was assessed using transcriptomes from various developmental. Batch effects were removed using the ComBat (6) package in R (v4.3.1). To measure tissue specificity, expression specificity metric (Tau) (7) and specificity measure (SPM) (8) were calculated. A co-expression network was constructed by calculating the Pearson correlation coefficient (PCC) of pairwise gene expression levels with gene pairs having a PCC > 0.8 retained to form the network. Furthermore, transcription factor regulatory networks for 50 species were predicted using the regulatory prediction function in PlantRegMap (9) , based on the regulatory network of *Arabidopsis*.

Genome variation calling

Due to the complexity and large size of *Asteraceae* genomes, population-level resequencing studies are limited (10,11). Nonetheless, transcriptome-based variant identification has emerged as a cost-effective alternative and has been successful applied in several species, such as lettuce, maize, and rice (12-14). In this study, we adopted a workflow previously used for variant identification using RNA-seq data (12-14). Initially, high quality clean reads were aligned to the reference genome using STAR with default parameters (15). The alignment outputs were subsequently converted and sorted using Samtools (v1.13) (16). The mapped reads were further processed for sorting, read group addition, and duplicate marking with Sambamba (v0.8.2) (17). After that, GATK’s SplitNCigarReads was utilized to split “N” cigar reads (18). GATK HaplotypeCaller was then employed to generate GVCF files for each sample. These individual GVCF files were combined using CombineGVCFs and GenotypeGVCFs to produce a unified VCF file. The resulting raw VCF file was filtered using GATK’s VariantFiltration to filtered out low mapping quality SNPs with parameters “QUAL < 30.0 || MQ < 50.0 || QD < 2”. All SNPs and InDels with minor allele frequencies (MAF) < 0.01 or missing rate > 0.1 were discarded by VCFtools (v.0.1.16) (19). Finally, the genetic variant annotation and effect prediction was carried out using SnpEff (5.0d) (20).

1. Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884-i890.

2. Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, **12**, 357-360.

3. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. and Salzberg, S.L. (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols*, **11**, 1650-1667.

4. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923-930.

5. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 1-21.

6. Zhang, Y., Parmigiani, G. and Johnson, W.E. (2020) ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics*, **2**.

7. Kryuchkova-Mostacci, N. and Robinson-Rechavi, M. (2017) A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform*, **18**, 205-214.

8. Xiao, S.-J., Zhang, C., Zou, Q. and Ji, Z.-L. (2010) TiSGeD: a database for tissue-specific genes. *Bioinformatics*, **26**, 1273-1275.

9. Tian, F., Yang, D.C., Meng, Y.Q., Jin, J. and Gao, G. (2020) PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res*, **48**, D1104-D1113.

10. 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.

11. Song, C., Liu, Y., Song, A., Dong, G., Zhao, H., Sun, W., Ramakrishnan, S., Wang, Y., Wang, S., Li, T. *et al.* (2018) The Chrysanthemum nankingense Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits. *Molecular Plant*, **11**, 1482-1491.

12. Zhang, L., Su, W., Tao, R., Zhang, W., Chen, J., Wu, P., Yan, C., Jia, Y., Larkin, R.M., Lavelle, D. *et al.* (2017) RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nat Commun*, **8**, 1-12.

13. Liu, S., Li, C., Wang, H., Wang, S., Yang, S., Liu, X., Yan, J., Li, B., Beatty, M., Zastrow-Hayes, G. *et al.* (2020) Mapping regulatory variants controlling gene expression in drought response and tolerance in maize. *Genome Biology*, **21**, 163.

14. Liu, C., Zhu, X., Zhang, J., Shen, M., Chen, K., Fu, X., Ma, L., Liu, X., Zhou, C., Zhou, D.X. *et al.* (2022) eQTLs play critical roles in regulating gene expression and identifying key regulators in rice. *Plant biotechnology journal*, **20**, 2357-2371.

15. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.

16. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

17. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. and Prins, P. (2015) Sambamba: fast processing of NGS alignment formats. *Bioinformatics*, **31**, 2032-2034.

18. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, **20**, 1297-1303.

19. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

20. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*, **6**, 80-92.