



OPEN ACCESS

EDITED BY
Vikram Singh,
Amrapali University, India

REVIEWED BY
Asmamaw Menelih,
Mekdela Amba University, Ethiopia
Alenka Hafner,
The Pennsylvania State University,
United States

*CORRESPONDENCE
Anna V. Klepikova
✉ annklepikova@gmail.com

RECEIVED 02 April 2026
REVISED 10 July 2026
ACCEPTED 13 July 2026
PUBLISHED 17 August 2026

CITATION
Klepikova AV, Kasianova AM,
Kasianov AS, Penin AA and
Logacheva MD (2026) Calm in the midst
of the storm: inferring gene expression
stability in flowering plants.
Front. Plant Sci. 17:1846139.
doi: 10.3389/fpls.2026.1846139

COPYRIGHT
© 2026 Klepikova, Kasianova, Kasianov,
Penin and Logacheva. This is an open-
access article distributed under the terms
of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

Calm in the midst of the storm: inferring gene expression stability in flowering plants

Anna V. Klepikova^{1,2*}, Alexandra M. Kasianova^{1,2},
Artem S. Kasianov^{3,4}, Aleksey A. Penin²
and Maria D. Logacheva^{1,2}

¹Center of Bio- and Medical Technology, Moscow, Russia, ²Laboratory of Plant Genomics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia, ³CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal, ⁴Biopolis Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão, Portugal

Introduction: Gene expression in plants is inherently dynamic, shifting rapidly during development and in response to environmental stimuli. However, many essential cellular processes require remarkable transcriptomic stability. The identity, genomic features, and evolutionary conservation of stably expressed genes (SEGs) in plants across broad evolutionary scales remain poorly understood.

Results: We analyzed comprehensive transcriptome atlases from five species representing major clades of angiosperms (*Arabidopsis thaliana*, *Fagopyrum esculentum*, *Solanum lycopersicum*, *Phalaenopsis equestris*, and *Zea mays*) to identify genes with consistently low expression variation across organs, developmental stages, and environmental conditions. We identified 672 orthogroups that are universally stable across angiosperms; these are predominantly single-copy genes enriched for fundamental cellular functions including DNA repair, mRNA splicing, and translation. Validation using independent datasets and additional species confirmed that stability is maintained across diverse genotypes, conditions, and species. SEGs tend to be longer, have more exons, and exhibit lower GC content than other expressed genes. Using multiomics information available for *Arabidopsis*, we also explored epigenetic and functional characteristics associated with stability. Loss-of-function mutations in SEGs result in lethality far more frequently than in ubiquitously expressed genes, and protein abundance is similarly stable. SEGs have a significantly higher level of gene body methylation and H3K4me1 marks.

Conclusions: Our results reveal several genomic and epigenetic features that distinguish SEGs from ubiquitously expressed genes. They also suggest that SEG stability is not governed by specific regulatory motifs, but is rather “by default” state driven by a distinctive chromatin architecture.

KEYWORDS

angiosperms, flowering plants, housekeeping genes, multiomics, orthology, stably expressed genes, transcriptomics, ubiquitous genes

1 Introduction

Changes in gene expression are the primary drivers of biological processes in all living organisms. Drastic transcriptomic shifts are essential for development and for responses to external and internal stimuli. At the evolutionary scale, the changes in gene expression level and pattern are associated with the emergence of morphological and physiological innovations (Álvarez-Buylla et al., 2010; Mika et al., 2021; Martín-Zamora et al., 2023). At the same time, many cellular processes (such as DNA repair, protein synthesis, and basic metabolism) are universal across a range of developmental stages and conditions and require remarkable stability of the transcriptome. However, the structure and regulation of the fraction of the transcriptome that mediates this stability is not comprehensively explored. Stably expressed genes (SEGs) (or “housekeeping genes”—these terms are often used interchangeably, though strictly, housekeeping genes represent a broader class of genes, where SEGs are a subset) have been characterized the most extensively in animal model species (especially in human and mouse; Hounkpe et al., 2021). In animals, SEGs are typically associated with specific genomic signatures, including higher gene body methylation (Sarda et al., 2012), lower conservation of promoter sequences (Farré et al., 2007), and shorter gene lengths and fewer introns, which may minimize the metabolic cost and noise of transcription or reflect simpler protein structure (Eisenberg and Levanon, 2003; Vinogradov, 2004; Nie et al., 2010).

In plants, there are several studies focused on within-species gene stability, mostly motivated by practical applications, such as identifying reliable reference genes for qPCR normalization (Czechowski et al., 2005; Yim et al., 2015; Zhuo et al., 2016). In an evolutionary context, there are few studies aimed at understanding the promoter architecture of stable/housekeeping genes versus tissue-specific genes (Das and Bansal, 2019; Yang et al., 2023). Also, a link between expression stability and gene body methylation found in animals (see above) persists in plants as well (Zastapilo et al., 2026) and presumably has a causative relationship. However, several fundamental research questions persist. First, it remains unclear whether there is a truly universal set of SEGs shared across the deeply divergent lineages of angiosperms. Second, while animal SEGs often exhibit specific structural features, it is unknown if plants (given their unique genome architecture dominated by frequent whole-genome duplications) rely on the same genomic or epigenomic hallmarks to maintain expression stability. Finally, a significant challenge is the distinction between “ubiquitously expressed” genes (present in all tissues) and truly “stably expressed” genes (maintaining constant transcript levels). To address these questions, it is necessary to analyze high-resolution transcriptome maps. Low-resolution or tissue-limited datasets often conflate ubiquity with stability, failing to capture the subtle variations in expression that occur during stress or rare developmental transitions. By using comprehensive RNA-seq datasets that provide full coverage of organs, developmental stages, and environmental conditions, we can isolate the most robustly stable fraction of the genome.

In this study, we utilize high-resolution data from five representative flowering plant species (*Arabidopsis thaliana*, *Fagopyrum esculentum*, *Solanum lycopersicum*, *Phalaenopsis equestris*, and *Zea mays*) to: 1) identify the fraction of genes that maintain stability across a vast range of conditions within a species, 2) assess the evolutionary conservation of these SEGs across the angiosperm phylogeny, and 3) characterize the genomic and epigenomic features associated with expression stability. Flowering plants serve as an ideal model for this inquiry due to their deep morphological divergence and the vast amount of available transcriptomic data covering many distinct lineages (Klepikova and Penin, 2019).

2 Materials and methods

2.1 Orthogroup inference

Protein sequences of 109 angiosperm species were downloaded from Phytozome 14 (Goodstein et al., 2012) (for the full list of species and names of downloaded files, see [Supplementary File 1](#)). For the five “core” species, proteins derived from the longest isoform of each protein-coding gene were obtained using the following genome assemblies and annotations: *A. thaliana*—GCA_978657495.1_TAIR12 (Reiser et al., 2026), *F. esculentum*—GCA_965199375.1 (Gunko et al., 2025), *S. lycopersicum*—Slycopersicum_796_ITAG5.0.gene (Phytozome 14), *Z. mays*—Zm-B73-REFERENCE-NAM-5.0 (Hufford et al., 2021), and *P. equestris*—PLAZA database v.4.5 (Cai et al., 2015). OrthoFinder v3.1.4 (Emms and Kelly, 2019) with default settings was used to construct gene orthogroups.

2.2 RNA-seq read mapping and counting

Transcriptomic data were analyzed for the five “core” species: *A. thaliana*, *F. esculentum*, *S. lycopersicum*, *Z. mays*, and *P. equestris*. The full list of NCBI SRR IDs of transcriptome samples is presented in [Supplementary Table 1](#). Reads were trimmed using Trimmomatic v0.36 (Bolger et al., 2014) with the following parameters “ILLUMINACLIP:adapters.all:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:30” and mapped on the respective genomes (listed above) with STAR v2.7.10b (Dobin et al., 2013). Only uniquely mapped reads were considered. The read number per gene was counted together with mapping by the “--quantMode GeneCounts” STAR option.

2.3 Definition of ubiquitously expressed genes

Read counts were normalized on library size using the DESeq median method (Anders and Huber, 2010) for each species separately. A gene was counted as expressed in a sample if it had 16 or more normalized read counts in each biological replicate of the sample (a “strict” threshold; Su et al., 2014). Genes expressed in every sample of a species were considered ubiquitously expressed.

2.4 Measurement of expression stability

For each ubiquitously expressed gene, the coefficient of variation (CV) was calculated as the mean expression divided by the standard deviation of expression across all samples of a species.

2.5 Gene Ontology enrichment

Gene Ontology enrichment of stably expressed *A. thaliana* genes was calculated using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) tool (Huang et al., 2009; Sherman et al., 2022). Genes expressed in all samples of *A. thaliana* (i.e., ubiquitous genes) were used as a background for the analysis. Terms with fold enrichment ≥ 2 and FDR < 0.05 were considered overrepresented in stable genes compared with ubiquitous ones.

2.6 Proteome analysis

A proteome map of *A. thaliana* was used to assess stability of SEGs at the proteome level (Mergner et al., 2020). Intensity-based absolute quantification (iBAQ) values were extracted for stable and ubiquitous genes from supplementary materials 41586_2020_2094_MOESM4_ESM.xlsx. The CV of iBAQ values was calculated on a proteome map as described above for transcriptome maps.

2.7 Codon usage analysis

DAMBE7 software (Xia, 2018) was used to calculate the effective number of codons for stable and ubiquitous genes of each species. CDS sequences for the studied genes were used as input for the software (in case of multiple isoforms, the longest one was taken in the analysis). The expected effective number of codons was calculated using DAMBE7 as well. The R package “coRdon” (Bioconductor, 2026) was used to calculate codon frequencies for stable and ubiquitous genes of each species. Then, to generate PR2 bias plots, the frequencies of each nucleotide at the third position of the codon were calculated as the sum of respective codon frequencies. The frequency of the A nucleotide was divided by the sum of A and T nucleotide frequencies, and the frequency of the G nucleotide was divided by the sum of G and C nucleotide frequencies, resulting in $A3/(A3+T3)$ and $G3/(G3+C3)$ values in the PR2 bias plot. For neutrality plots, frequencies of GC content were calculated at each position of the codon, naming GC1, GC2, and GC3, respectively. GC12 was calculated as the mean of GC1 and GC2. Relative synonymous codon usage (RSCU) was calculated as follows: for each group of synonymous codons, codon frequencies were summed and then divided by the number of synonymous codons in the group, resulting in uniform codon frequencies. RSCU was calculated as the ratio of observed codon frequencies to expected uniform codon frequency.

2.8 Cytosine methylation analysis

To explore cytosine methylation, several datasets were used: 20 ecotypes were randomly selected from 1001 Epigenomes data on

Arabidopsis (Kawakatsu et al., 2016), 24 samples of developing and germinating seeds (Kawakatsu et al., 2017), 6 samples of ovules and anthers (Wu et al., 2025), 12 samples treated with salicylic acid together with controls (Margaritopoulou et al., 2026), 12 samples subjected to nutritional stress, and 4 samples treated with heat condition; 91 samples were analyzed in total. For the full list of samples, see Supplementary Table 2. Reads were trimmed with Trim Galore v2.2.0 with default settings (Babraham Bioinformatics, 2026). Read mapping, read deduplication, and methylation calling were performed with Bismark v0.25.1 (Krueger and Andrews, 2011). Only fully methylated positions (with $>90\%$ of reads supporting methylated C) were considered for the subsequent analysis. For each sample and stable and ubiquitous gene, the number of methylated C in three contexts (CG, CHH, and CHG) was calculated together with the total number of C in these three contexts in 5'-UTR, CDS, introns, and 3'-UTR. For each context, the ratio of methylated C was calculated as the number of methylated C to the total number of C in the context.

2.9 Histone modifications analysis

BigWig files describing the level of histone modifications at certain chromosome positions were downloaded from the Plant Chromatin State Database (Liu et al., 2018). BigWig files were converted to bedGraph using the bigWigToBedGraph tool (Anaconda, 2026). The levels of histone modifications were counted and averaged for the 5'-UTR, gene body, and 3'-UTR of each protein-coding ubiquitous or stable gene.

2.10 ATAC-seq analysis

BED files with ATAC-seq peak calling results were downloaded from the PlantCADB database (Ding et al., 2023) (for the list of samples used, see Supplementary Table 3). For each gene, the number of ATAC-seq peaks located in the 1,000-bp region upstream of the TSS was calculated in each sample. If a gene had at least one ATAC-seq peak in both replicates of any sample, the gene is considered to have a promoter region with open chromatin.

2.11 Overrepresentation of cis-regulatory motifs in gene promoters

Promoter regions, i.e., 1,000 bp sequence upstream of the annotated start of the longest protein-coding isoform of stable and ubiquitous genes, were extracted. Promoter coordinates were generated using BEDTools flank v2.31.1 with the following options “-l 1000 -r 0 -s” (Quinlan and Hall, 2010). The sequences were extracted from the reference genomes using the BEDTools getfasta with parameters “-s -name.” Overrepresentation of transcription factor binding motifs was identified using motif models from the CIS-BP database (Weirauch et al., 2014). Motif enrichment analysis was performed using the findMotifs.pl module of HOMER v5.1 with the parameters “-len 8,10,12 -S 10 -basic” (Heinz et al., 2010). Analyses were performed separately for each of the “core” species and on a dataset combining promoter sequences from all five species.

2.12 Data visualization

The R package “ggplot2” (Wickham, 2016) was used to create histograms, violin plots, pie charts, box plots, scatter plots, and distribution plots. Expression level was normalized on GeTMM to remove the influence of both library size and gene length (Smid et al., 2018). For heatmap visualization, the R package “complexHeatmap” was used (Gu et al., 2016). Venn diagram was drawn with the UGent Bioinformatics & Evolutionary Genomics Venn tool (Bioinformatics & Evolutionary Genomics, 2026). The R package “chromoMap” (Anand and Rodriguez Lopez, 2022) was used to analyze the distribution of stable genes across *A. thaliana* chromosomes. Two input files were used: the first one with coordinates of chromosomes (i.e., chromosome length) and chromosome centromeres, and the second file with gene coordinates.

3 Results

3.1 The inference of stably expressed orthogroups

We selected 114 angiosperm species with high-quality genome assemblies and annotations available; these species represent main evolutionary lineages within flowering plants. At the first step, we inferred gene orthogroups using OrthoFinder (Emms and Kelly, 2019) and excluded single-gene and single-species orthogroups from the analysis. The 45,900 remaining orthogroups comprised 3,964,165 genes with 2–10,647 genes and 2–114 species per orthogroup (for the description of orthogroups, see Supplementary File 1).

Out of the species used for orthology inference, we selected five species: *F. esculentum* (buckwheat, caryophyllids, eudicots), *A. thaliana* (thale cress, rosids, eudicots), *S. lycopersicum* (tomato, asterids, eudicots), *Z. mays* (maize, grass monocots), and *P. equestris* (horse phalaenopsis, Orchidaceae, non-grass monocots) (Klepikova et al., 2016; Klepikova et al., 2021; Stelpflug et al., 2016; Penin et al., 2019; Penin et al., 2021) for which highly detailed transcriptome maps are available (for the list of samples, see Supplementary Table 1; for the description of expression dataset quality, see Supplementary File 2). A total of 13,876 out of 45,900 orthogroups included at least two genes of two or more aforementioned species (these orthogroups covered 86.7% of all genes in the analysis). Based on the gene expression data for the five well-studied “core” species, we retained for further analysis only those genes that are expressed in all replicates of every sample—ubiquitously expressed genes (see Materials and methods). This filtering step retained only ~10% of the orthogroups: 1,539 out of 13,876 orthogroups had all genes of five species expressed in all samples.

As a measure of gene expression stability, we used CV (Klepikova et al., 2016). To identify SEGs, we applied a relaxed threshold (CV lower than 0.75) due to the large evolutionary distances between species. Under this definition, 672 orthogroups consisted of SEGs (see example of a stable orthogroup in Figure 1A). A striking feature of the fraction of genes that are ubiquitously expressed (either stably or not) is that they belong prevalently to the

orthogroups that have five genes (considering only five “core” species, Figure 1B). Within such orthogroups, the largest fraction is orthopentades—orthogroups with one gene from each of the five species, i.e., the groups of single-copy genes. This trend toward single-copy is more pronounced in stable orthogroups: they have a significantly higher number of orthopentades than ubiquitously expressed or other genes (Figure 1B, Fisher’s exact test p -value = $2.924\text{e-}06$ and $<2.2\text{e-}16$, respectively). Stable orthogroups comprised 3,410 genes, which were almost equally distributed between species (from 648 *Phalaenopsis* to 707 *Arabidopsis* genes); the majority of stable orthogroups contained all five species, congruent with the high number of orthopentades.

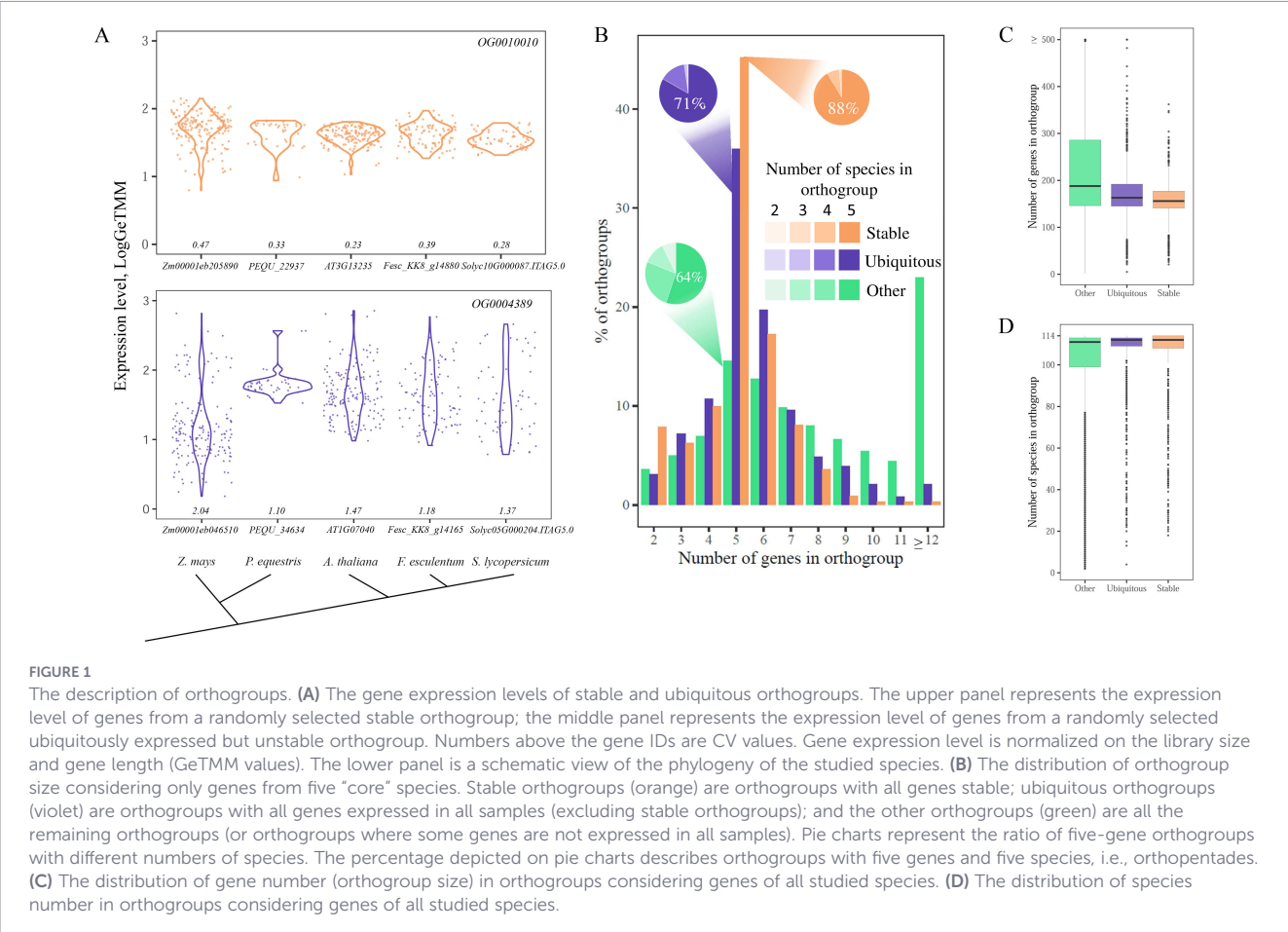
The overall size of stable and ubiquitous orthogroups was close to the number of analyzed species, i.e., 114 genes (median 156 and 163, respectively, Figure 1C), while other orthogroups varied greatly in size (from 2 to 10,647 genes with a median of 188). With respect to species number, stable and ubiquitous orthogroups had almost all species included (72.9% and 72.6% of stable and ubiquitous orthogroups had at least 110 species with a median of 112), and only 61.4% of other orthogroups had 110 species or more (Figure 1D). Thus, SEGs tend to be a single copy at the level of all angiosperm plants.

3.2 The consistency of stable orthogroups

Since the identification of stable orthogroups relies on a threshold in CV, some orthogroups can fall out of the fraction of stable ones because the coefficient of variation for a certain member of the orthogroup is slightly higher than the threshold. Taking this into account, we examined whether all potentially stable orthogroups were found. All but one of the genes of the “core” species were stable in 455 orthogroups with 187 orthopentades among them (Supplementary Figure 1A). The most unstable genes from these orthogroups had high coefficients of variance: the first quartile of the distribution was 0.87/0.84 (for orthopentades/other orthogroups, respectively, Supplementary Figure 1B). Thus, there are no “almost-stable” orthogroups. Though the genes that were unstable within these “all-but-one stable” orthogroups had no specific Gene Ontology (GO) enrichment, the majority of unstable genes had a pattern shifted to anthers (in *A. thaliana*, *P. equestris*, *F. esculentum*, *Z. mays*), proliferating tissues (*P. equestris*, *F. esculentum*, *Z. mays*), or senescent organs (*A. thaliana*) (Supplementary Figures 1C–G).

In order to provide direct testing of orthogroups identified as stable based on five species analysis, we analyzed gene expression of *Cucumis sativus* (cucumber), a species for which a wide sampling of RNA-seq datasets is available but which was not included in our core set used for the identification of SEGs (Supplementary File 3). A total of 672 stable orthogroups included 704 cucumber genes (most of the orthogroups included one cucumber gene, congruent with the tendency of SEGs to be single-copy). A survey of expression stability has shown that 92% of them were stably expressed.

Another important question about SEGs is whether the stability is robust with respect to the sample choice. Transcriptome maps (including the ones used for stability inference in this study) usually represent samples standardized in terms of external conditions and genotypes. In order to explore the robustness of SEGs, we collected



five additional sets of publicly available RNA-seq experiments on *A. thaliana* and *Z. mays* (945 and 1,213 samples, respectively). These experiments include leaf samples of different accessions, a detailed atlas of root, biotic and abiotic stress treatments, and supplementation of plants with vitamins (Supplementary Table 4). Each of the datasets was merged with a transcriptome map, and coefficients of variance were recalculated. The majority of SEGs retain stability in all datasets for both *A. thaliana* and *Z. mays* (Table 1). Although 11% of all *Arabidopsis* SEGs and 16% of maize SEGs became unstable in at least one combination of transcriptome map and public dataset (Table 2), 627 *A. thaliana* and 573 *Z. mays* genes were uniformly expressed across a great variety of plant tissues, organs, environmental conditions, and ecological interactions, confirming their status as universal SEGs.

TABLE 1 Number of SEGs retaining stability in the combination of transcriptome map and additional dataset.

Dataset	<i>A. thaliana</i> (707 SEGs)	<i>Z. mays</i> (685 SEGs)
Leaves of different accessions	702 (99%)	664 (97%)
Detailed root atlas	661 (94%)	654 (95%)
Infection of pathogens	702 (95%)	634 (93%)
Abiotic stress treatments	662 (94%)	633 (92%)
Vitamin supplementation	705 (99%)	655 (96%)

3.3 SEGs at the proteome and organism levels

The increasing knowledge on post-transcriptional regulation and the analysis of proteomics data show that the expression measured at the level of RNA does not always accurately reflect the abundance of proteins (Greenbaum et al., 2003, reviewed in Liu et al., 2016). Proteomic studies usually lag behind transcriptomic studies; this is the case for plants as well, although the situation begins to change. Recently, a proteome map was published for a model plant of *A.thaliana* (Mergner et al., 2020). This allows us to assess whether SEGs are stable at the proteome level in this species. Indeed, we found that the coefficient of variance was significantly lower for SEGs in comparison with ubiquitous genes (Figure 2A Mann–Whitney test p -value < $2.2e-16$). This shows that the stability of SEGs is maintained not only at the level of transcription but also at the post-transcriptional level.

Such stability implies an essential role in the life of an organism. Indeed, Gene Ontology enrichment analysis for *Arabidopsis* SEGs shows that they are associated with such basic cellular functions as DNA repair, mRNA splicing, translation initiation, and protein transport and degradation (Supplementary Table 5). The availability of multiple *Arabidopsis* knockout mutants allows us to directly compare the consequences of loss of function of SEGs and other genes. We analyzed the importance of SEGs using two datasets: the list of *Arabidopsis* genes whose loss-of-function mutations lead to

TABLE 2 Preservation of expression stability across plant tissues and environmental conditions.

Number of datasets where SEG was stable	A number of genes remained stable in a certain number of dataset combinations	
	<i>A. thaliana</i> (707 SEGs)	<i>Z. mays</i> (685 SEGs)
5	627 (88.7%)	573 (83.6%)
4	61 (8.6%)	72 (10.5%)
3	16 (2.3%)	17 (2.5%)
2	2 (0.3%)	15 (2.2%)
1	1 (0.1%)	6 (0.9%)
0	0 (0.0%)	2 (0.3%)

lethality (Lloyd et al., 2015) and the classification of *Arabidopsis* genes by the outcome of loss-of-function mutations (Lloyd and Meinke, 2012). In the latter classification, the genes that are classified as essential (ESN) are almost completely nested into the group of genes with lethal mutations from the former classification (675 out of 719) (Supplementary Figure 2).

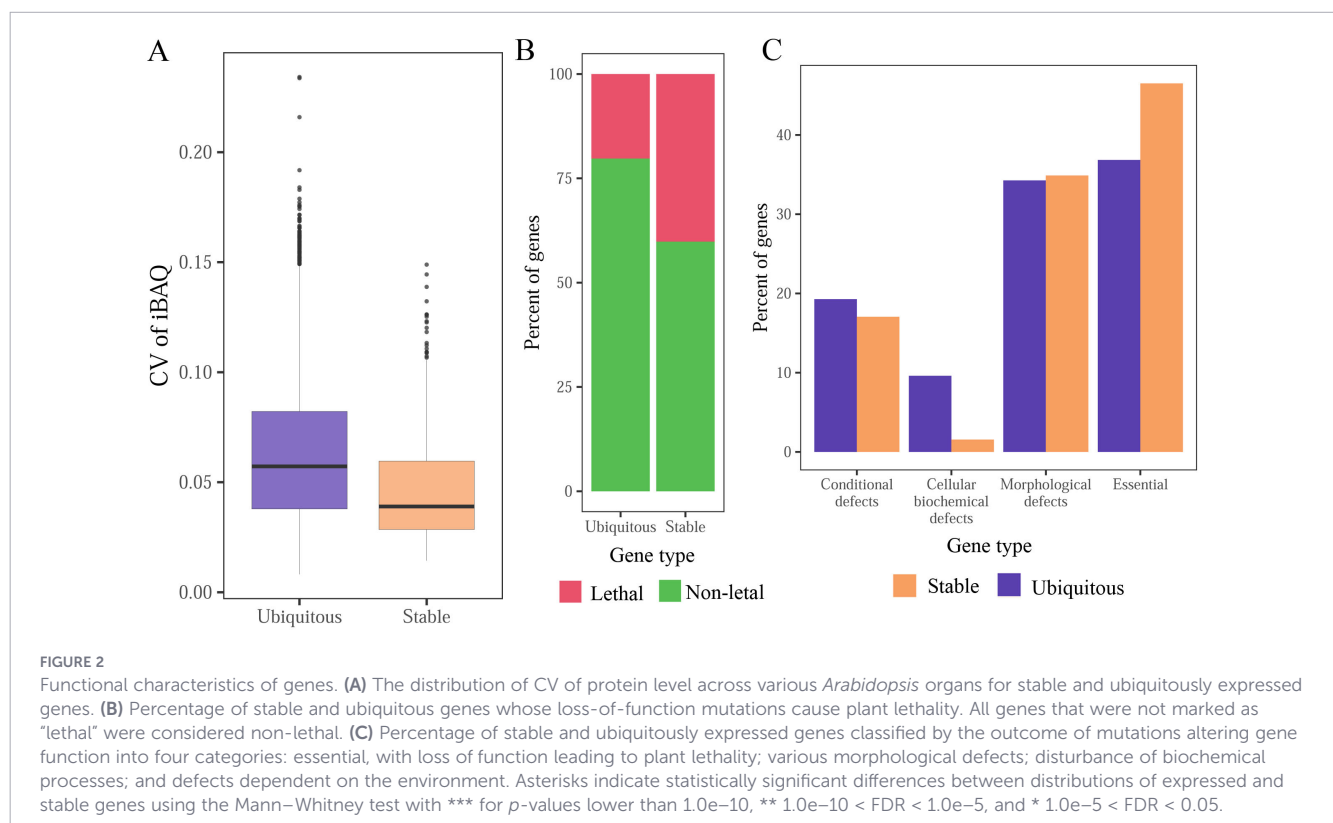
The loss-of-function mutations in SEGs led to plant lethality far more frequently than in other ubiquitous genes (40% compared to 20%, Fisher's exact test p -value $< 2.2 \times 10^{-16}$, Figure 2B). The analysis of the second dataset leads to the same conclusions: mutations in the majority of SEGs resulted in lethality or morphological defects (Figure 2C).

3.4 Expression and structural characteristics of SEGs

In order to gain insight into the mechanisms that support the stability of SEGs at the transcriptional and translational levels, we compared several genomic and epigenomic characteristics of angiosperm-scale stable genes (i.e., genes from 672 orthogroups where all genes are stable) and ubiquitous genes (i.e., genes expressed in all samples excluding SEGs). This allowed us to find features associated with stability. First, SEGs of all species were longer than ubiquitous genes in terms of the median length of CDS (Figure 3A), exon number (Figure 3B), and average intron length (Supplementary Figure 3). The GC content of stable genes was lower, with a significantly smaller standard deviation for four species excluding tomato (Figure 3C).

The median expression levels of SEGs and ubiquitous genes did not statistically differ for *Z. mays*, *A. thaliana*, *F. esculentum*, and *S. lycopersicum*, while for *P. equestris*, SEGs were expressed at a higher level (Figure 3A). Notably, SEGs tend to have more similar expression values: in four species, the standard deviation of median expression level was 2.5–4.9 times lower for SEGs compared with ubiquitous genes (Levene test's p -values were 1.74×10^{-5} for *Z. mays*, 2.06×10^{-3} for *F. esculentum*, and 1.73×10^{-4} for *A. thaliana*. *P. equestris* and *S. lycopersicum* did not show a significant difference).

Another important feature that is known to be correlated with transcript abundance is codon usage. Synonymous codons can be used unequally in different species or certain genes of a species, a feature called codon usage bias (Hanson and Coller, 2018). We explored codon usage bias in stable and ubiquitous genes using



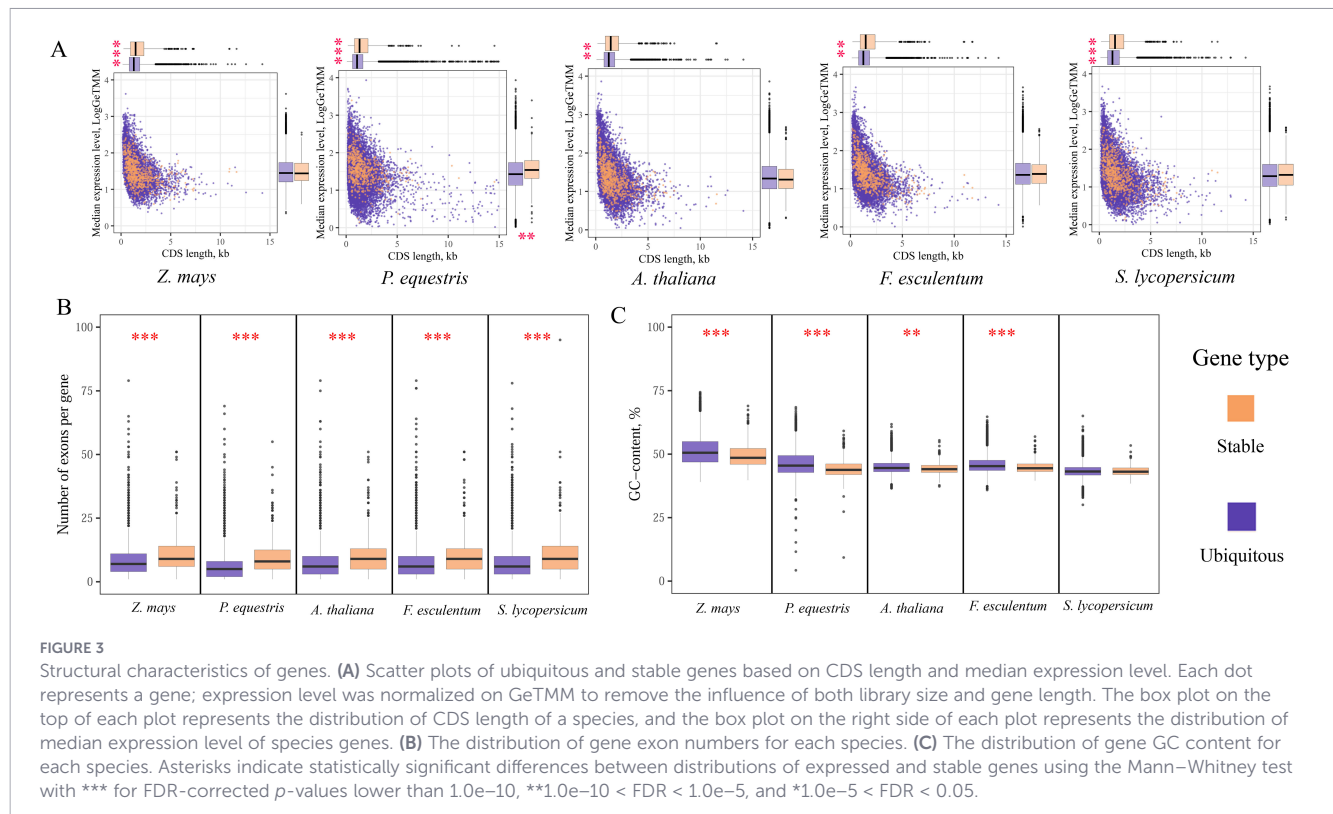


FIGURE 3

Structural characteristics of genes. (A) Scatter plots of ubiquitous and stable genes based on CDS length and median expression level. Each dot represents a gene; expression level was normalized on GeTMM to remove the influence of both library size and gene length. The box plot on the top of each plot represents the distribution of CDS length of a species, and the box plot on the right side of each plot represents the distribution of median expression level of species genes. (B) The distribution of gene exon numbers for each species. (C) The distribution of gene GC content for each species. Asterisks indicate statistically significant differences between distributions of expressed and stable genes using the Mann-Whitney test with *** for FDR-corrected p -values lower than $1.0\text{e-}10$, ** $1.0\text{e-}10 < \text{FDR} < 1.0\text{e-}5$, and * $1.0\text{e-}5 < \text{FDR} < 0.05$.

effective number of gene codons (Supplementary Figure 4A), PR2 bias plots (Supplementary Figure 4B), and neutrality plots (Supplementary Figure 4C). We found mild differences in codon characteristics of stable and ubiquitous genes: codon usage of both gene types is influenced by the mutation process rather than natural selection, and GC content in the third position of the codon was lower for SEGs, matching overall GC content. RSCU is the ratio of estimated codon frequency to the expected frequency if all synonymous codons were used equally. RSCU higher than 1.6 is considered overrepresented. The main trend for the four species excluding *Z. mays* was the increasing codon usage bias in SEGs (Supplementary Table 6).

3.5 Epigenetic profiles of SEGs

Gene expression level and pattern are dependent on the epigenetic state of a gene locus, in particular, on the level of cytosine methylation and histone modifications. In plants, cytosine methylation occurs in three functionally different contexts: CG, CHH, and CHG, where H stands for A, C, or T (Gallego-Bartolomé, 2020). We used a wide variety of epigenomics data, including ecotypes from 1001 Epigenomes data on *Arabidopsis* (Kawakatsu et al., 2016), various plant organs, and treatments with multiple stress conditions (Kawakatsu et al., 2017; Wu et al., 2025; Margaritopoulou et al., 2026), to explore cytosine methylation in angiosperm-scale stable genes in comparison with genes expressed across whole plants, but not universally stable. While cytosine methylation in UTRs was not detected in the majority of genes (only 8%–41% of genes were methylated in at least one sample depending on context), CDS and introns of SEGs showed significantly higher methylation levels in

the CG context (Figure 4A). Notably, the variation of cytosine methylation across the studied samples (assessed with CV) was significantly lower for SEGs in the CG context (Supplementary Figure 5). The level of CHG and CHH methylation was low both in CDS and introns.

We analyzed different kinds of methylation and acetylation of histones 3 and 4 lysines using data from the Plant Chromatin State Database (Liu et al., 2018). We did not find differences in histone mark distribution in both the 5'- and 3'-UTRs of ubiquitous and stable genes, with one exception (Supplementary Figures 6A,B): H3K4me2, the repressive mark in euchromatin, was more frequent in ubiquitous genes compared with stable ones. In contrast, almost all histone modifications in the coding region of genes differed between stable and ubiquitous genes (Figure 4B). Methylation of lysine 4 of histone 3 (H3K4me1) was more pronounced in SEGs, while all the other marks associated with active transcription were more frequent among the ubiquitous genes. As expected, all three histone modifications, which mark heterochromatin and repression of transcription, were found at a very low level in both ubiquitous and stable genes.

Supporting the observation on histone marks, the analysis of ATAC-seq data showed the association of SEGs with open chromatin. We used ATAC-seq peaks for roots, seedlings, and leaves deposited in the database PlantCADB (Ding et al., 2023) (for the list of samples used, see Supplementary Table 3). For each gene, we calculated the number of ATAC-seq peaks located in the 1,000-bp region upstream of the TSS and considered a gene to have a promoter region with open chromatin if a gene had an ATAC-seq peak in both replicates of any sample. We did not find a difference between SEGs and ubiquitously expressed genes (93.5%

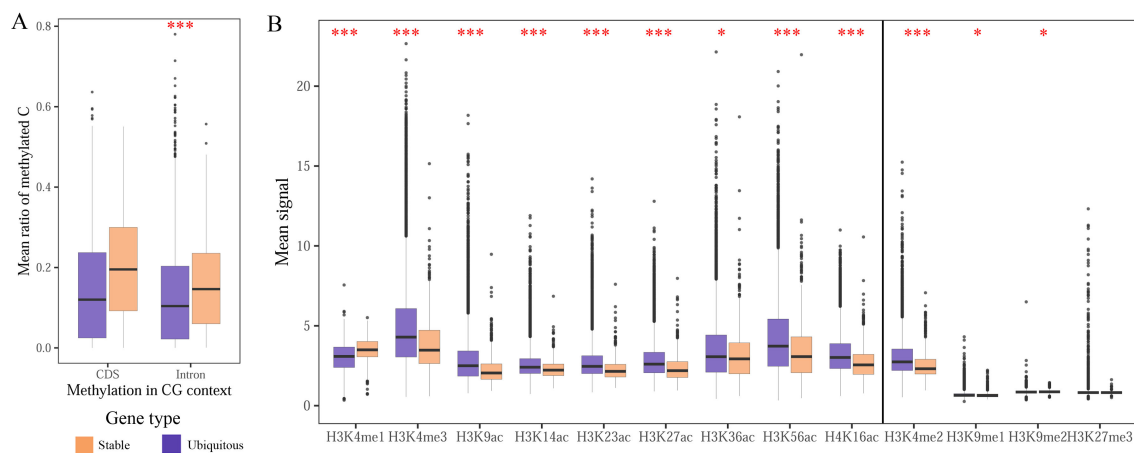


FIGURE 4

Chromatin characteristics of genes. (A) Averaged across 91 samples, the ratio of methylated cytosines to all cytosines in CG contexts of CDS and introns of ubiquitously expressed and stable *A. thaliana* genes. Ratios were averaged across 91 studied samples. (B) The distribution of histone modifications in the gene body of ubiquitously expressed and stable *A. thaliana* genes. The vertical solid line separates active histone marks (left part) from repressive ones (right part). Asterisks indicate statistically significant differences between distributions of expressed and stable genes using the Mann–Whitney test with *** for FDR-corrected p -values lower than $1.0\text{e-}10$, ** $1.0\text{e-}10 < \text{FDR} < 1.0\text{e-}5$, and * $1.0\text{e-}5 < \text{FDR} < 0.05$.

of SEGs had an open chromatin in promoter regions compared with 92.1% of ubiquitous genes; among all the other genes, 64.5% were associated with open chromatin). As open chromatin is distributed unequally along the chromosomes, we analyzed the possibility of SEGs being organized in spatial clusters on chromosomes. However, the analysis of gene distribution across chromosomes with the chromoMap R package showed its uniformity (Supplementary Figure 7).

Finally, we analyzed the overrepresentation of cis-regulatory motifs recognized by various transcription factors in promoter regions of SEGs compared with ubiquitous genes. Surprisingly, we did not find any motif to be more frequent in SEG promoters than in the promoters of ubiquitous genes, indicating the absence of common regulation of expression stability.

4 Discussion

Using a balanced dataset of transcriptome data representing main evolutionary lineages of flowering plants, we found the groups of genes that are stable not only across organs and conditions but also across species. Previous studies on stable genes were mostly focused on within-species stability (Zhuo et al., 2016; Wang et al., 2017; Gu et al., 2023) or paired comparisons between human and model animals (Lin et al., 2019a; Hounkpe et al., 2021). Our study represents a step forward in putting the question of gene expression stability in the evolutionary context and using multiomics data available for a model plant species. First, our analysis highlighted that the stability of expression is strongly associated with single-copy genes present in all species studied. This complements earlier findings on tissue-specific genes: they are, vice versa, associated with duplicated genes—a phenomenon that is presumably common for plants and animals (Ha et al., 2009; Huerta-Cepas et al., 2011;

Mantica et al., 2024). There are several factors that can affect conclusions on the stability of gene expression—in particular, the set of species and samples taken into analysis. We thus corroborated our findings using two lines of evidence. First, we added in the analyses a species outside of our five-species core set (*C. sativus*) and found that more than 96% of orthogroups remained stable. Second, for those species for which a vast amount of transcriptomic data not used for initial inference of stable genes is available—*Arabidopsis* and maize—we added several large datasets in order to see how this will affect gene stability. These datasets include biotic and abiotic stresses, different genotypes, and vitamin supplementation. For most datasets, the decrease in the fraction of stable genes was at the level of 1%–4%; overall, more than 86% of SEGs retained stability. It is important to note that the variation of conditions in these datasets included not only the key experimental condition (e.g., a certain stress) but also a number of other aspects of plant growth (temperature of growth, day length, laboratory or open-ground). This emphasizes that the SEGs found in our study are universally stable. The identification of SEGs is essential for the development of many methodologies, e.g., qPCR (Vandesompele et al., 2002; Andersen et al., 2004) and single-cell transcriptome analysis (Lin et al., 2019b; Deeke and Gagnon-Bartsch, 2020). Also, we suggest that these genes can be used as an analog of BUSCO datasets for plant *de novo* transcriptome assemblies. Indeed, a conventional practice for quality control of genome assemblies is the search for core genes—the ones that are thought to be universally present in the genome (this approach was first adopted in Parra et al., 2007), but now replaced by a newer pipeline, BUSCO (Simão et al., 2015; Tegenfeldt et al., 2025). This approach is hardly applicable to transcriptome assemblies, because the set of transcripts is tissue/condition-specific and may not include the core genes. Our inference of stable genes allows us to circumscribe the set of genes that is expected to be present in any transcriptome assembly, thus contributing to the improvement of assembly quality control.

The availability of multiomics data for *Arabidopsis* allowed us not only to identify stable genes but also to explore a number of genomic and functional characteristics associated with stability. This lays a foundation for the mechanistic exploration of the maintenance of expression stability and its significance for the representation of genotype in phenotype. First, at an organismal level, within SEGs, the fraction of essential genes—the ones that result in a lethal phenotype when disrupted—is significantly higher than for other genes expressed in all samples. Earlier studies established an association between essential genes and single-copy genes (Lloyd and Meinke, 2012; Lloyd et al., 2015). The authors also noted that these genes typically exhibit high and broad expression levels. Their definition of expression breadth was based on the number of samples where a gene is expressed; in the same terms, all the genes listed in our study as ubiquitous are broadly expressed. A significant difference between the fraction of essential genes in SEGs and ubiquitous genes emphasizes the uniqueness and essentiality of SEGs at an organismal level. Expression levels and profiles are known to be affected by a number of epigenetic features—e.g., chromatin accessibility, histone modification marks, and cytosine methylation (Miller and Grant, 2013; Wang et al., 2022). In terms of epigenetic profiles, while SEGs closely resemble genes with detectable yet unstable expression, they also exhibit unique features that likely contribute to their stability. First, stable genes have a significantly higher level of gene body methylation. The association of gene body methylation with actively expressed genes is known for both plants and animals (Sarda et al., 2012; Zilberman, 2017). A recent large-scale study on natural populations of *Arabidopsis* has revealed that gene body methylation definitely contributes to interpopulation expression stability (Shahzad et al., 2025). Our findings are congruent with this. Regarding histone modification, there are also a number of significant differences between SEGs and ubiquitous genes. Most of those differences concern gene body regions. In particular, there is an increase in H3K4me1 levels in SEGs. Other histone modifications either do not show significant differences or have decreased levels. H3K4me1 is a well-known activating mark; it is strongly associated with actively transcribed genes (Zhang et al., 2009). While generally all variants of H3K4 methylation (me1, me2, and me3) are regarded as marks of transcriptionally active chromatin (Le et al., 2025), there are subtle differences in their localization, function, and interaction with other epigenetic mechanisms. In particular, H3K4me2- and H3K4me3-containing regions are depleted in GC gene body methylation, while in contrast, it is highly enriched in H3K4me1-containing regions (Zhang et al., 2009). Congruent with that, we observe a higher level of GBM in SEG and a higher level of H3K4me1, but not other types of H3K4 methylation. For H3K4me2, the situation is even more complex—there is a report on rice showing its function as a repressive mark (Liu et al., 2019). Also, in *Arabidopsis*, there is accumulating evidence that it is in fact repressive (Mori et al., 2023; Noyori et al., 2025). Congruent with this, the level of H3K4me2 in SEG is significantly lower.

At the same time, chromatin accessibility does not differ between stable and expressed genes. When comparing broadly expressed versus tissue- or stage-specific genes, there is a significant

difference (Boyle et al., 2008; Delandre et al., 2022). Indeed, chromatin accessibility determines the possibility of being expressed, but the actual levels of expression are fine-tuned by other mechanisms.

The absence of significantly enriched cis-regulatory motifs in the promoters of SEGs was an unexpected finding. It suggests that despite being conserved at the level of flowering plants, expression stability is not governed by universal regulatory elements. This lack of motif enrichment distinguishes SEGs from tissue-specific or stress-responsive genes, which typically rely on the presence of specialized elements to coordinate dynamic transcriptomic shifts, and many of these elements are evolutionarily conserved (Spensley et al., 2009; Gasch et al., 2016). Instead, we propose that the stability of these core genes may be a fundamental property of their constitutively accessible chromatin and the notable absence of repressive marks like H3K4me2, which ensures continuous recognition by the general transcriptional machinery. These results lead us to the hypothesis that the stability of plant SEGs might be maintained through a decentralized regulatory logic, where constitutive activity is a default state supported by genomic architecture rather than specific promoter motifs. However, currently available chromatin accessibility datasets for plants are quite scarce; further genome-wide analysis and functional studies are necessary to support this hypothesis and to provide a mechanistic explanation for the observed associations.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

AVK: Visualization, Data curation, Conceptualization, Writing – original draft, Writing – review & editing, Investigation, Formal analysis. AMK: Formal analysis, Writing – original draft, Investigation. ASK: Writing – original draft, Formal analysis, Data curation. AP: Methodology, Conceptualization, Project administration, Supervision, Investigation, Writing – original draft, Funding acquisition, Writing – review & editing. ML: Investigation, Writing – review & editing, Conceptualization, Writing – original draft.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the Ministry of Science and Higher Education (the Federal Scientific-Technical Programme for Genetic Technologies Development for 2019–2030, agreement #075-15-2025-476).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- Álvarez-Buylla, E. R., Ambrose, B. A., Flores-Sandoval, E., Englund, M., Garay-Arroyo, A., García-Ponce, B., et al. (2010). B-function expression in the flower center underlies the homeotic phenotype of *Lacandonia schismatica* (Triuridaceae). *Plant Cell* 22, 3543–3559. doi: 10.1105/tpc.109.069153
- Anaconda (2026). *BigwigtoBedGraph*. Available online at: <https://anaconda.org/channels/bioconda/packages/ucsc-bigwigtoBedGraph/overview> (Accessed March 19, 2026).
- Anand, L., and Rodriguez Lopez, C. M. (2022). ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinf.* 23, 33. doi: 10.1186/s12859-021-04556-z
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi: 10.1186/gb-2010-11-10-r106
- Andersen, C. L., Jensen, J. L., and Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64, 5245–5250. doi: 10.1158/0008-5472.CAN-04-0496
- Babraham Bioinformatics (2026). *Trim Galore!*. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (Accessed June 6, 2026).
- Bioconductor (2026). *Cordon: Codon Usage Analysis and Prediction of Gene Expressivity*. Available online at: <https://bioconductor.org/packages/coRdon> (Accessed March 19, 2026).
- Bioinformatics & Evolutionary Genomics (2026). *Venn Diagram*. Available online at: <https://bioinformatics.psb.ugent.be/webtools/Venn/> (Accessed March 19, 2026).
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., et al. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132, 311–322. doi: 10.1016/j.cell.2007.12.014
- Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.-C., Liu, K.-W., et al. (2015). The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* 47, 65–72. doi: 10.1038/ng.3149
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W.-R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* 139, 5–17. doi: 10.1104/pp.105.063743
- Das, S., and Bansal, M. (2019). Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. *PLoS One* 14, e0212678. doi: 10.1371/journal.pone.0212678
- Deeke, J. M., and Gagnon-Bartsch, J. A. (2020). Stably expressed genes in single-cell RNA sequencing. *J. Bioinform. Comput. Biol.* 18, 2040004. doi: 10.1142/S0219720020400041
- Delandre, C., McMullen, J. P. D., Paulsen, J., Collas, P., and Marshall, O. J. (2022). Eight principal chromatin states functionally segregate the fly genome into developmental and housekeeping roles. doi: 10.1101/2022.10.30.514435
- Ding, K., Sun, S., Luo, Y., Long, C., Zhai, J., Zhai, Y., et al. (2023). PlantCADB: a comprehensive plant chromatin accessibility database. *Genomics Proteomics Bioinf.* 21, 311–323. doi: 10.1016/j.gpb.2022.10.005
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Eisenberg, E., and Levanon, E. Y. (2003). Human housekeeping genes are compact. *Trends Genet.* 19, 362–365. doi: 10.1016/S0168-9525(03)00140-9
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi: 10.1186/s13059-019-1832-y
- Farré, D., Bellora, N., Mularoni, L., Messegue, X., and Albà, M. M. (2007). Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol.* 8, R140. doi: 10.1186/gb-2007-8-7-r140
- Gallego-Bartolomé, J. (2020). DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytol.* 227, 38–44. doi: 10.1111/nph.16529
- Gasch, P., Fundinger, M., Müller, J. T., Lee, T., Bailey-Serres, J., and Mustroph, A. (2016). Redundant ERF-VII transcription factors bind to an evolutionarily conserved cis-motif to regulate hypoxia-responsive gene expression in Arabidopsis. *Plant Cell* 28, 160–180. doi: 10.1105/tpc.15.00866
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozone: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* 4, 117. doi: 10.1186/gb-2003-4-9-117
- Gu, J., Dai, J., Lu, H., and Zhao, H. (2023). Comprehensive analysis of ubiquitously expressed genes in humans from a data-driven perspective. *Genomics Proteomics Bioinf.* 21, 164–176. doi: 10.1016/j.gpb.2021.08.017
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi: 10.1093/bioinformatics/btw313
- Gunko, E. M., Schelkunov, M. I., Prokopchuk, S. R., Klepikova, A. V., Omelchenko, D. O., Romanova, O. I., et al. (2025). Seed shattering in common buckwheat (*Fagopyrum esculentum*): insights from RNA-seq and morphological analysis. *BMC Plant Biol.* 25, 1371. doi: 10.1186/s12870-025-07310-2
- Ha, M., Kim, E.-D., and Chen, Z. J. (2009). Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2295–2300. doi: 10.1073/pnas.0807350106
- Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* 19, 20–30. doi: 10.1038/nrm.2017.91
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589. doi: 10.1016/j.molcel.2010.05.004
- Hounkpe, B. W., Chenou, F., de Lima, F., and De Paula, E. V. (2021). HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Res.* 49, D947–D955. doi: 10.1093/nar/gkaa609

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2026.1846139/full#supplementary-material>

- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Huerta-Cepas, J., Dopazo, J., Huynen, M. A., and Gabaldon, T. (2011). Evidence for short-time divergence and long-time conservation of tissue-specific expression after gene duplication. *Briefings Bioinf.* 12, 442–448. doi: 10.1093/bib/bbr022
- Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Ou, S., Liu, J., et al. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* 373, 655–662. doi: 10.1126/science.abg5289
- Kawakatsu, T., Huang, S. C., Jupe, F., Sasaki, E., Schmitz, R. J., Urlich, M. A., et al. (2016). Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* 166, 492–505. doi: 10.1016/j.cell.2016.06.044
- Kawakatsu, T., Nery, J. R., Castanon, R., and Ecker, J. R. (2017). Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biol.* 18, 171. doi: 10.1186/s13059-017-1251-x
- Klepikova, A. V., Kasianov, A. S., Ezhova, M. A., Penin, A. A., and Logacheva, M. D. (2021). Transcriptome atlas of *Phalaenopsis equestris*. *PeerJ* 9, e12600. doi: 10.7717/peerj.13310
- Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., and Penin, A. A. (2016). A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* 88, 1058–1070. doi: 10.1111/tip.13312
- Klepikova, A. V., and Penin, A. A. (2019). Gene expression maps in plants: current state and prospects. *Plants* 8, 309. doi: 10.3390/plants8090309
- Krueger, F., and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. doi: 10.1093/bioinformatics/btr167
- Le, H., Simmons, C. H., and Zhong, X. (2025). Functions and mechanisms of histone modifications in plants. *Annu. Rev. Plant Biol.* 76, 551–578. doi: 10.1146/annurev-arplant-083123-070919
- Lin, Y., Ghazanfar, S., Strbenac, D., Wang, A., Patrick, E., Lin, D. M., et al. (2019a). Evaluating stably expressed genes in single cells. *GigaScience* 8, giz106. doi: 10.1093/gigascience/giz106
- Lin, Y., Ghazanfar, S., Wang, K. Y. X., Gagnon-Bartsch, J. A., Lo, K. K., Su, X., et al. (2019b). scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets. *Proc. Natl. Acad. Sci. U.S.A.* 116, 9775–9784. doi: 10.1073/pnas.1820061116
- Liu, Y., Beyer, A., and Aebersold, R. (2016). On the dependency of cellular protein levels on mRNA abundance. *Cell* 165, 535–550. doi: 10.1016/j.cell.2016.03.014
- Liu, Y., Liu, K., Yin, L., Yu, Y., Qi, J., Shen, W.-H., et al. (2019). H3K4me2 functions as a repressive epigenetic mark in plants. *Epigenet. Chromatin* 12, 40. doi: 10.1186/s13072-019-0285-6
- Liu, Y., Tian, T., Zhang, K., You, Q., Yan, H., Zhao, N., et al. (2018). PCSD: a plant chromatin state database. *Nucleic Acids Res.* 46, D1157–D1167. doi: 10.1093/nar/gkx919
- Lloyd, J., and Meinke, D. (2012). A comprehensive dataset of genes with a loss-of-function mutant phenotype in *Arabidopsis*. *Plant Physiol.* 158, 1115–1129. doi: 10.1104/pp.111.192393
- Lloyd, J. P., Seddon, A. E., Moghe, G. D., Simenc, M. C., and Shiu, S.-H. (2015). Characteristics of plant essential genes allow for within- and between-species prediction of lethal mutant phenotypes. *Plant Cell* 27, 2133–2147. doi: 10.1105/tpc.15.00051
- Mantica, F., Iñiguez, L. P., Marquez, Y., Permanyer, J., Torres-Mendez, A., Cruz, J., et al. (2024). Evolution of tissue-specific expression of ancestral genes across vertebrates and insects. *Nat. Ecol. Evol.* 8, 1140–1153. doi: 10.1101/2022.11.14.516384
- Margaritopoulou, T., Foutadakis, S., Vatsellas, G., Samiotaki, M., and Markellou, E. (2026). A genome-wide DNA methylation survey reveals salicylic acid-induced distinct hypomethylation linked to defense responses against biotrophic pathogens. *IJMS* 27, 1935. doi: 10.3390/ijms27041935
- Martin-Zamora, F. M., Liang, Y., Guynes, K., Carrillo-Baltodano, A. M., Davies, B. E., Donnellan, R. D., et al. (2023). Annelid functional genomics reveal the origins of bilaterian life cycles. *Nature* 615, 105–110. doi: 10.1038/s41586-022-05636-7
- Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., et al. (2020). Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* 579, 409–414. doi: 10.1038/s41586-020-2094-2
- Mika, K., Marinić, M., Singh, M., Muter, J., Brosens, J. J., and Lynch, V. J. (2021). Evolutionary transcriptomics implicates new genes and pathways in human pregnancy and adverse pregnancy outcomes. *eLife* 10, e69584. doi: 10.7554/eLife.69584
- Miller, J. L., and Grant, P. A. (2013). “The role of DNA methylation and histone modifications in transcriptional regulation in humans,” in *Epigenetics: Development and Disease*. Ed. T. K. Kundu (Springer Netherlands, Dordrecht), 289–317. doi: 10.1007/978-94-007-4525-4_13
- Mori, S., Oya, S., Takahashi, M., Takashima, K., Inagaki, S., and Kakutani, T. (2023). Cotranscriptional demethylation induces global loss of H3K4me2 from active genes in *Arabidopsis*. *EMBO J.* 42, e113798. doi: 10.15252/embj.2023113798
- Nie, H., Crooijmans, R. P. M. A., Lammers, A., Van Schothorst, E. M., Keijer, J., Neerincx, P. B. T., et al. (2010). Gene expression in chicken reveals correlation with structural genomic features and conserved patterns of transcription in the terrestrial vertebrates. *PLoS One* 5, e11990. doi: 10.1371/journal.pone.0011990
- Noyori, T., Mori, S., Oya, S., Nishio, H., Kudoh, H., Inagaki, S., et al. (2025). H3K4me2 orchestrates H2A.Z and Polycomb repressive marks in *Arabidopsis*. *Nat. Commun.* 16, 11657. doi: 10.1038/s41467-025-66645-4
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071
- Penin, A. A., Kasianov, A. S., Klepikova, A. V., Kirov, I. V., Gerasimov, E. S., Fesenko, A. N., et al. (2021). High-resolution transcriptome atlas and improved genome assembly of common buckwheat, *Fagopyrum esculentum*. *Front. Plant Sci.* 12, 612382. doi: 10.3389/fpls.2021.612382
- Penin, A. A., Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., and Logacheva, M. D. (2019). Comparative analysis of developmental transcriptome maps of *Arabidopsis thaliana* and *Solanum lycopersicum*. *Genes* 10, 50. doi: 10.3390/genes10010050
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Reiser, L., Proia, A., Bakker, E., Subramaniam, S., Khosa, K., Sawant, S., et al. (2026). Recent major changes to TAIR: updates to the database, website, and *Arabidopsis* genome. *Genetics* 232, iyaf248. doi: 10.1093/genetics/iyaf248
- Sarda, S., Zeng, J., Hunt, B. G., and Yi, S. V. (2012). The evolution of invertebrate gene body methylation. *Mol. Biol. Evol.* 29, 1907–1916. doi: 10.1093/molbev/mss062
- Shahzad, Z., Hollwey, E., Moore, J. D., Choi, J., Cassin-Ross, G., Rouached, H., et al. (2025). Gene body methylation regulates gene expression and mediates phenotypic diversity in natural *Arabidopsis* populations. *Nat. Plants* 11, 2084–2099. doi: 10.1038/s41477-025-02108-4
- Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., et al. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists, (2021 update). *Nucleic Acids Res.* 50, W216–W221. doi: 10.1093/nar/gkac194
- 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. doi: 10.1093/bioinformatics/btv351
- Smid, M., Coebergh Van Den Braak, R. R. J., Van De Werken, H. J. G., Van Riet, J., Van Galen, A., De Weerd, V., et al. (2018). Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinf.* 19, 236. doi: 10.1186/s12859-018-2246-7
- Spensley, M., Kim, J.-Y., Picot, E., Reid, J., Ott, S., Helliwell, C., et al. (2009). Evolutionarily conserved regulatory motifs in the promoter of the *Arabidopsis* clock gene LATE ELONGATED HYPOCOTYL. *Plant Cell* 21, 2606–2623. doi: 10.1105/tpc.109.069898
- Stelpflug, S. C., Sekhon, R. S., Vaillancourt, B., Hirsch, C. N., Buell, C. R., de Leon, N., et al. (2016). An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* 9, plantgenome2015.04.0025. doi: 10.3835/plantgenome2015.04.0025
- Su, Z., Łabaj, P. P., Li, S., Thierry-Mieg, J., Thierry-Mieg, D., Shi, W., et al. (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* 32, 903–914. doi: 10.1038/nbt.2957
- Tegenfeldt, F., Kuznetsov, D., Manni, M., Berkeley, M., Zdobnov, E. M., and Kriventseva, E. V. (2025). OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res.* 53, D516–D522. doi: 10.1093/nar/gkac987
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, research0034.1. doi: 10.1186/gb-2002-3-7-research0034
- Vinogradov, A. E. (2004). Compactness of human housekeeping genes: selection for economy or genomic design? *Trends Genet.* 20, 248–253. doi: 10.1016/j.tig.2004.03.006
- Wang, G., Li, X., Shen, W., Li, M.-W., Huang, M., Zhang, J., et al. (2022). The chromatin accessibility landscape of pistils and anthers in rice. *Plant Physiol.* 190, 2797–2811. doi: 10.1093/plphys/kiac448
- Wang, K., Vijay, V., and Fuscoe, J. C. (2017). Stably expressed genes involved in basic cellular functions. *PLoS One* 12, e0170813. doi: 10.1371/journal.pone.0170813
- Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443. doi: 10.1016/j.cell.2014.08.009
- Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016 (Cham: Springer International Publishing: Imprint: Springer). doi: 10.1007/978-3-319-24277-4
- Wu, Z., Xue, Y., Wang, S., Shih, Y.-H., Zhong, Z., Feng, S., et al. (2025). REM transcription factors and GDE1 shape the DNA methylation landscape through the

- recruitment of RNA polymerase IV transcription complexes. *Nat. Cell Biol.* 27, 1136–1147. doi: 10.1038/s41556-025-01691-0
- Xia, X. (2018). DAMBE7: new and improved tools for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* 35, 1550–1552. doi: 10.1093/molbev/msy073
- Yang, E. J. Y., Maranas, C. J., and Nemhauser, J. L. (2023). A comparative analysis of stably expressed genes across diverse angiosperms exposes flexibility in underlying promoter architecture. *G3: Genes Genomes Genet.* 13, jkad206. doi: 10.1093/g3journal/jkad206
- Yim, A. K.-Y., Wong, J. W.-H., Ku, Y.-S., Qin, H., Chan, T.-F., and Lam, H.-M. (2015). Using RNA-Seq data to evaluate reference genes suitable for gene expression studies in soybean. *PLoS One* 10, e0136343. doi: 10.1371/journal.pone.0136343
- Zastapilo, J., Emmerson, R., Mikheeva, L. A., Catoni, M., Bechtold, U., and Zabet, N. R. (2026). Gene body methylation buffers noise in gene expression in plants. *Nucleic Acids Res.* 54, gkag127. doi: 10.1101/2024.07.01.601483
- Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M., and Jacobsen, S. E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* 10, R62. doi: 10.1186/gb-2009-10-6-r62
- Zhuo, B., Emerson, S., Chang, J. H., and Di, Y. (2016). Identifying stably expressed genes from multiple RNA-Seq data sets. *PeerJ* 4, e2791. doi: 10.7717/peerj.2791
- Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* 18, 87. doi: 10.1186/s13059-017-1230-2