

Gene expression divergence following gene and genome duplications in spatially resolved plant transcriptomes

Fabricio Almeida-Silva^{1,2*} and Yves Van de Peer^{1,2,3,4*}

¹ Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium

² VIB Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium

³ Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0028, South Africa.

⁴ College of Horticulture, Academy for Advanced Interdisciplinary Studies, Nanjing Agricultural University, Nanjing, China.

* To whom correspondence should be addressed (fabricio.almeidasilva@psb.vib-ugent.be; yves.vandepeer@psb.vib-ugent.be)

Abstract

Gene and genome duplications are key drivers of plant genome evolution, expanding genetic repertoires and facilitating functional innovation. However, genes originating from different duplication mechanisms undergo different outcomes, especially at the expression level. Previous studies have demonstrated that segmental or whole-genome duplications generate duplicates with similar and somewhat redundant expression profiles across multiple tissues, while other duplicates display increased divergence, ultimately leading to functional innovations. However, little is known about how duplicates diverge in expression across cell types in a single tissue. Here, we used high-resolution spatial transcriptomic data from five species (*Arabidopsis thaliana*, *Glycine max*, *Phalaenopsis aphrodite*, *Zea mays*, and *Hordeum vulgare*) to investigate the evolution of gene expression following gene duplications. We found that genes originating from segmental or whole-genome duplications display increased expression levels, expression breadths, spatial variability, and number of coexpression partners. Duplication mechanisms that preserve cis-regulatory landscapes typically generate paralogs with more preserved expression profiles, but such differences by duplication mode disappear over time. Expression divergence also depends on gene functions, with dosage-sensitive gene families displaying highly preserved expression profiles, while families involved in more specialized processes (e.g., flowering and phytohormone biosynthesis) display increased divergence. Paralogs originating from large-scale (including whole-genome) duplications display redundant and/or overlapping expression profiles, indicating functional redundancy and/or subfunctionalization, while small-scale duplicates diverge asymmetrically, indicating neofunctionalization. Collectively, our findings provide new insights into the tempo and mode of gene expression evolution, helping understand how gene and genome duplications shape cell identities.

Keywords: polyploidy, spatial transcriptomics, transcriptional regulation, regulatory genomics.

Introduction

Gene duplications have long been recognized as a major source of novel genetic material upon which evolution can act (Ohno 1970). Individual genes can be duplicated locally (though tandem and proximal duplications), and/or through transposon-mediated duplications (Panchy et al. 2016; Almeida-Silva and Van de Peer 2025). In plants, single-gene duplications have fueled the evolution of key traits, including secondary metabolism and stress response (Baumgarten et al. 2003; Rizzon et al. 2006; Chae et al. 2014; Almeida-Silva and Van de Peer 2023; Ji et al. 2024). Whole-genome duplications (WGDs) are an extreme duplication mechanism whereby all chromosome sets double in copy number, and they have occurred multiple times in all eukaryotic kingdoms, but are particularly pervasive in plants (Van De Peer et al. 2017; Ren et al. 2018; One Thousand Plant Transcriptomes Initiative 2019; Almeida-Silva and Van de Peer 2025). Despite the detrimental effects of doubling the entire genome, such as reduced fertility and genomic instability (Comai 2005; Woodhouse et al. 2010), evidence for the evolutionary and ecological potential of WGD has accumulated steeply, with WGDs being associated with increased diversity (Tank et al. 2015; Landis et al. 2018; Ren et al. 2018), origin of novel traits (Soltis and Soltis 2016), increased robustness in gene regulatory networks (Almeida-Silva and Van de Peer 2023), and survival during environmental catastrophes (Chen et al. 2024).

Duplications of genes and entire genomes immediately result in redundancy, allowing for the accumulation of mutations in coding and *cis*-regulatory sequences through genetic drift. While redundancy can be selected for in special cases (e.g., to increase mutational robustness), duplicated genes are thought to undergo one of three major outcomes in the long term: gene loss (pseudogenization), partitioning of ancestral functions (subfunctionalization), or gain of novel functions (neofunctionalization) (Panchy et al. 2016; Birchler and Yang 2022; Iohannes and Jackson 2023). Importantly, the fate of a duplicated gene largely depends on its duplication mechanism and functions. For instance, dosage-sensitive gene families (e.g. regulatory proteins and multiprotein complexes) are typically lost after single-gene duplications, but not after whole-genome duplications, what is likely due to selection to preserve stoichiometric balance in a cell (Blanc and Wolfe 2004; Maere et al. 2005; Veitia et al. 2008; Freeling 2009; Birchler and Veitia 2010). Likewise, retained defense-related genes often undergo neofunctionalization after tandem and/or proximal duplications (Birchler and Yang 2022; Ugalde and Straube 2023).

Over the past decades, several studies have explored large compendia of expression data to investigate how paralogous genes diverge in expression following gene and genome duplications. For instance, WGD-derived duplicates in *Arabidopsis thaliana* display highly redundant and overlapping expression profiles, while duplicates created by small-scale duplications display asymmetric divergence (Casneuf et al. 2006). Also in *A. thaliana*, WGD-derived genes display greater expression levels and are expressed in more tissues (Ganko et al. 2007). In rice (*Oryza sativa*), segmental and tandem duplicates display higher expression correlations compared with duplicates originating from other modes (Li et al. 2009). In soybean (*Glycine max*), WGD-derived paralogs display greater levels of expression preservation compared to other duplication modes, with particularly larger differences for pairs originating at the *Glycine*-specific WGD at around 13 million years ago (Almeida-Silva et al. 2020a).

Although analyses of large-scale, multi-tissue expression compendia have provided important insights into expression-level divergence following different modes of duplications, they are limited in resolution. As bulk RNA-seq samples aggregate gene expression across multiple (often heterogeneous) cell populations, cell type- and cell state-specific changes in expression following duplications are not captured. While single-cell RNA sequencing offers high-resolution expression profiling, it also comes with limitations, such as loss of original spatial information (important for accurate cell type annotation), and biased proportions for cell types that are rare and/or difficult to protoplast (Kiselev et al. 2019; Xu et al. 2021; Cox Jr et al. 2022; Yin et al. 2023). Spatial transcriptomics addresses such issues, allowing for *in situ* quantification of transcript abundances. Investigating how paralogous genes diverge in expression within tissue compartments using spatially resolved transcriptomes could help understand how gene and genome duplications shape cell identities.

Here, we analyzed spatially resolved plant transcriptomes from five species to investigate how duplicated genes diverge in expression across tissue compartments. We found that genes originating from different duplication modes display differences in expression levels, expression breadth, and spatial expression variability. Expression divergence between paralogous gene pairs increases exponentially with time, and varies in an age- and duplication-mode specific manner. Expression divergence varies across gene families, with dosage-dependent gene families displaying greater preservation. Segmental duplicates are highly connected in coexpression networks and typically display expression-level redundancy or symmetric divergence (suggestive of

subfunctionalization), while small-scale duplicates (tandem, proximal, and transposed) typically diverge asymmetrically (suggestive of neofunctionalization). Collectively, our findings provide important insights into the impact of gene and genome duplications on plant cell identities.

Results

Data set overview and sample quality statistics

We manually curated the literature to find plant spatial transcriptomics data sets for which count matrices and spot metadata were publicly available (see Materials and Methods for details). We obtained five independent data sets, which represent *Arabidopsis* cauline leaves (Xia et al. 2022), soybean (*Glycine max*) nodules infected with rhizobia (Liu et al. 2023), developing maize (*Zea mays*) ears (Wang et al. 2024), *Phalaenopsis* Big Chili inflorescences (buds and shoot apical meristem) at successive developmental stages (hereafter referred to as *Phalaenopsis aphrodite*, whose genome was used as a reference in the original publication) (Liu et al. 2022), and germinating barley (*Hordeum vulgare*) seeds (Peirats-Llobet et al. 2023) (Table 1, Supplementary Fig. S1a).

We performed minimal preprocessing to remove genes with zero counts across all spots, spots with zero counts, and genes with less than 1 count in $\geq 99.9\%$ of the spots. After preprocessing, all samples had $>10,000$ genes, and the median number of spots per sample in each species ranged from 1545 in *H. vulgare* to 8062 in *Zea mays* (Supplementary Fig. S1b). The number of spots per sample in *A. thaliana* was exceptionally smaller than that for other species (median = 631) because of the considerably smaller tissue area in transversal leaf sections (Supplementary Fig. S1a). The median number of distinct spatial domains per sample ranged from 6 in *A. thaliana* and *G. max* to 14 in *H. vulgare*. Collectively, summary gene- and spot-level statistics indicate that samples display high quality, as all samples have had their transcriptomes profiled in a high-throughput manner (over 10,000 genes) across hundreds to thousands of spots.

Gene duplication mode impacts expression levels, expression breadths, and spatial variability

To investigate whether genes originating from different duplication modes display differences in global expression profiles, we first classified genes as originating from segmental (SD), tandem (TD), proximal (PD), retrotransposed (rTRD), DNA transposed (dTRD), or dispersed duplications (DD) (see Materials and Methods for details). Importantly, SD gene sets include both genes created by WGD and by duplications of large genomic regions (e.g., entire chromosomes, entire chromosome arms, etc), so we chose to broadly use ‘SD’. Then, we tested for differences in mean expression levels, expression breadths (i.e., number of distinct cell types in which genes are expressed), and spatial variability (i.e., variability in expression across spatial domains).

We observed that duplication modes that preserve *cis*-regulatory landscapes (i.e., SD and TD) often generate genes with greater expression levels and expression breadth, especially compared to DD genes (Kruskal-Wallis test, $P < 0.05$; Fig. 1a and 1b). However, dTRD genes in *H. vulgare* and *P. aphrodite* are exceptions, with expression levels and breadths comparable to SD and TD genes. In all species, SD genes are overrepresented in spatially variable gene (SVG) sets, indicating that genes originating from this duplication mode are more likely to display variable expression profiles across tissue compartments (Fisher’s exact test, $P < 0.05$; Fig. 1c). However, dTRD genes are also overrepresented in SVGs in *H. vulgare*, *P. aphrodite*, *Z. maysa*, and some samples from *A. thaliana*. Likewise, rTRD genes are overrepresented in SVGs in *H. vulgare* and *P. aphrodite*. While some associations between duplication modes and expression profiles are data set (or species)-specific, we found consistent associations across species, indicating that duplication modes play an important role in determining a gene’s fate at the expression level.

Gene duplication mode impacts expression divergence in young paralogs

To investigate the impact of gene duplication modes on expression divergence between paralogs, we aggregated spots into metaspots and compared distributions of mean gene-gene Spearman’s correlations across samples (see Materials and Methods for details). We observed that expression divergence increases exponentially with time (represented by synonymous substitution rates, K_s), as log-log regressions display a better fit compared to simple linear regressions (smaller BIC and AIC; Supplementary Fig. S2). To

account for a potential confounding effect of duplication age, paralogous gene pairs were divided into age groups defined based on peaks in K_s distributions. Thus, SD duplicates here are likely derived from whole-genome duplications (WGDs), as they are close to K_s peaks. Comparisons of gene-gene correlations by duplication mode were only made for paralogs from the same age group. Gene-gene correlations were used as a measure of expression divergence, with higher correlations indicating less divergence.

We observed no differences in expression divergence for older duplicates (i.e. generated tens of million years ago, MYA) (Fig. 2a and 2b). However, younger duplicates (generated <20 MYA) display decreased expression divergence when duplicated by mechanisms that preserve *cis*-regulatory landscapes (SD, TD, and PD). For instance, younger SD and PD duplicates in soybean (originated ~13 MYA) display less expression divergence, but no differences by duplication mode are observed for duplicates originated ~62 MYA (Fig. 2a and 2b). Likewise, SD, TD, and PD duplicates in maize (originated ~16 MYA) display less expression divergence, while no differences are observed for barley duplicates (originated ~53 MYA). Collectively, our findings suggest that duplication modes impact expression divergence in an age-dependent manner. An exception exists for *A. thaliana*, whose SD duplicates display less expression divergence despite their age (~51-72 MYA), which could be due to slower evolutionary rates, or tissue-specific differences in expression divergence rates.

Expression divergence varies across gene families and functions

We hypothesized that different gene families display different rates of expression divergence, probably as a result of variability of selective constraints acting upon *cis*-regulatory elements of some functional groups. To test this hypothesis, we assigned paralogous pairs to gene families using delineated gene families from PLAZA (Van Bel et al. 2022), and calculated the interquartile range (IQR) of gene-gene correlations for each family. In line with our hypothesis, we observed large variations in IQRs across families, with some gene families comprising highly similar paralogs (HSPs), and some families comprising highly dissimilar paralogs (HDPs) (Fig. 2c). This finding suggests that strong selection pressures act on *cis*-regulatory elements of some gene families to preserve expression similarity, while more relaxed selection pressures allow genes in most other families to diverge in expression.

To better understand the functional profiles of genes in families with HSPs and HDPs, we selected the top 10% gene families with highest and lowest median gene-gene correlations, and performed an enrichment analysis for GO terms. Families with HSPs are enriched in genes associated with basic and dosage-sensitive cellular processes, including chromatin assembly, rRNA processing, regulation of transcription, splicing, and photosystem assembly. Families with HDPs are enriched in genes associated with more specialized biological processes, including flower development, auxin and abscisic acid biosynthesis, detoxification, sucrose metabolism and transport, and cell wall biogenesis and remodeling (Supplementary Table S1). Collectively, our findings highlight biological processes under stronger selection pressures, leading to increased preservation and conservation across species, and biological processes that are more likely to diversify through the subfunctionalization and neofunctionalization of genes thereof.

Network-based expression divergence reveals potential selection pressures constraining divergence

To investigate how expression divergence between paralogs impacts sets of genes in a coordinated way, we inferred a gene coexpression network (GCN) for each species by pseudobulking and combining expression matrices (see Materials and Methods for details). In GCNs, genes with similar profiles are clustered together in so-called coexpression modules. Thus, we used co-occurrence of paralogous pairs in coexpression modules as an indication of preserved expression profiles, while paralogous pairs with genes in different modules were interpreted as having distinct (and, hence, diverged) expression profiles.

In line with our findings based on correlation distributions (Fig. 2b), we observed that most paralogous pairs display expression divergence, with relative frequencies of diverged pairs >90% in nearly all species and duplication modes (Fig. 3a). However, in most species and duplication modes, observed frequencies of diverged pairs are still smaller than the expected by chance in degree-preserving simulated networks, suggesting that selection pressures may be acting to constrain expression divergence (Fig. 3a). We found no consistent differences in network-based expression divergence by duplication mode. Module dissimilarities for diverged pairs were mostly small, indicating that, despite being in different modules (and thus displaying different expression profiles), most diverged pairs are still somewhat similar (Supplementary Fig. S3).

Gene duplication mode impacts gene connectivities in coexpression networks

We identified hubs in coexpression networks and tested for potential associations between duplication modes and gene connectivities (or node degree). We found that SD genes are enriched in hubs in *G. max*, *P. aphrodite*, and *Z. mays*, and dTRD genes are also enriched in hubs in *P. aphrodite* (Fig. 3b). This finding suggests that such duplication mechanisms tend to create genes with more central roles in transcriptional networks. However, the lack of association between particular duplication modes and hub status in *A. thaliana* and *H. vulgare* could indicate that such associations are species- or tissue-specific. Nevertheless, segmental duplications were significantly associated with hub status in three out of five data sets, providing stronger evidence for a potentially broader association between this duplication mode and higher connectivities.

Pervasive asymmetric divergence among small-scale, but not segmental duplicates

To understand the evolutionary fates of paralogous pairs at the expression level, we classified paralogous gene pairs in divergence classes based on the relative expression breadths (REBs) of genes therein (see Materials and Methods for details). In short, paralog pairs are classified as displaying i. redundancy, if both genes are expressed in (nearly) the same cell types; ii. asymmetric divergence, if a gene specializes in one or a few cell types while the other remains broadly expressed; and iii. symmetric divergence, if genes are expressed in complementary, (mostly) non-overlapping cell types.

Overall, we observed that most paralogous gene pairs diverge asymmetrically, with one gene in a pair displaying high REB while the other displays low REB (Fig. 4a). This divergence pattern indicates that, in high resolution, specialization in particular cell types is the most common outcome of duplicated genes at the expression level. However, a large fraction of SD pairs display expression-level redundancy in *Z. mays* and *P. aphrodite*, as indicated by high REBs for both genes in a pair (Fig. 4a). We tested for associations between duplication modes and divergence classes, and found that paralog pairs derived from small-scale duplications (TD, PD, TRD, and DD) are overrepresented in pairs displaying asymmetric divergence in most species (Fig. 4b). Conversely, segmental duplicates are typically overrepresented in pairs displaying redundancy (Fig.

4b). Collectively, our findings indicate the duplication modes play a role in expression divergence outcomes.

Discussion

To the best of our knowledge, this is the first study on the evolution of high-resolution expression divergence (i.e. spatially resolved, *quasi* single-cell level) following gene duplications in plants. Similarly to previous findings based on (low-resolution) bulk RNA-seq data, we demonstrate that gene duplication modes and gene functions play a major role in determining the expression-level outcome of a gene following duplication. Despite the diversity of gene duplication mechanisms, some of our findings seem to group duplication modes into two major classes based on whether or not cis-regulatory landscapes are preserved. This is highlighted by the greater expression levels and expression breadth for SD- and TD-derived genes, and the consistent association between segmental duplications and spatial variability in gene expression across species.

In lotus (*Nelumbo nucifera*), greater expression levels and breadth among SD-derived genes was associated with gene essentiality, as essential genes are typically highly expressed in all plant tissues (Lloyd et al. 2015; Shi et al. 2020). However, this seems to be contradicted by the observed association between segmental duplications and spatial expression variability, because expression levels of essential genes should be stable across cell types. We hypothesize that greater and broader expression for SD- and TD-derived genes is intrinsically linked with the preservation of ancestral transcription factor binding sites (TFBS) in promoters and enhancers. Duplication mechanisms that dramatically change cis-regulatory landscapes (e.g., transposon-mediated duplications) could generate gene copies in genomic regions that have, for instance, less TFBS, or more silencing, ultimately leading to reduced and more restricted expression levels. For example, segmental duplicates in *Arabidopsis* sp. displayed less DNA methylation in promoter regions, resulting in greater expression levels (Ha et al. 2009). In *Drosophila melanogaster*, expression levels of duplicates are often greater than the expected two-fold (i.e., additive expression levels) for genes arranged in tandem, but not when the same genes are in *trans*, indicating that the tandem organization is important for non-additive increases in expression (Loehlin and Carroll 2016).

Besides the effect of shared cis-regulatory landscapes, the scale of the duplication (i.e., segmental or large-scale *versus* small-scale) plays an important role. For instance,

we observed that hubs in most of the coexpression networks are enriched in SD-derived genes. Hubs are typically regarded as essential nodes for the functioning and survival of biological networks. In yeast protein-protein interaction networks, hubs were found to be enriched in essential genes (i.e., genes that lead to cell lethality when not expressed) (Jeong et al. 2001). This association was later reported in several other studies using different data sets, species, and types of biological networks, culminating in the so-called 'lethality-centrality rule' (Yu et al. 2004; Hahn and Kern 2005; Resendis-Antonio et al. 2005; Batada et al. 2006; Raman et al. 2014; Song et al. 2015; Almeida-Silva et al. 2020a). Essential hubs in a large-scale soybean coexpression network are involved in typically dosage-sensitive biological processes, such as nucleic acids metabolism, translation, and regulation of cell division (Tasdighian et al. 2017; Almeida-Silva et al. 2020a). As dosage-sensitive gene families are typically retained after whole-genome duplications, the association between SD-derived genes and hubs is likely due to the increased frequency of essential genes following SD (including WGD).

The scale of the duplication also clearly separates paralogs in terms of divergence classes. While duplicates originating from large-scale duplications typically undergo redundancy or subfunctionalization, small-scale duplicates often undergo neofunctionalization. Similar findings have been reported in other organisms, including *Saccharomyces cerevisiae* (Fares et al. 2013), *Arabidopsis thaliana* (Casneuf et al. 2006), *Drosophila melanogaster* (Assis and Bachtrog 2013), and grasses (Jiang and Assis 2019). In grasses, associations between small-scale duplications and neofunctionalization was hypothesized to be related to retrotransposon-derived duplications, which generate copies lacking introns and regulatory elements of their ancestral genes (Jiang and Assis 2019). However, we also observed associations between neofunctionalization and other small-scale duplication modes, such as tandem and proximal duplications, which typically preserve (although to a lesser extent) cis-regulatory landscapes. Hence, we hypothesize that differences in distal regulatory elements (e.g., enhancers) play a considerable role in the functional divergence of duplicate pairs. Alternatively, as large-scale duplications preserve genome organization, greater levels of redundancy and overlapping expression patterns among SD-derived pairs could be due to structural constraints imposed by the preservation of chromatin structure.

The effect of duplication modes and gene functions on expression divergence are mostly consistent across species, and our findings recapitulate previous findings based

on bulk RNA-seq data. For instance, we demonstrate that segmental duplicates display higher correlations compared to duplicates originating from most single-gene duplication mechanisms, which is in line with previous works using large compendia of microarray and bulk RNA-seq data (Casneuf et al. 2006; Li et al. 2009; Qiao et al. 2019). Likewise, we demonstrate that duplicate pairs in dosage-dependent gene families display highly preserved expression profiles, which is likely due to stronger selection pressures to preserve stoichiometric balance, as previously suggested (Freeling and Thomas 2006; Freeling 2009; Birchler and Veitia 2012).

The consistency of findings across different species and resolutions (i.e., bulk and spatial) raises the question of whether and to what extent (duplicated) gene expression evolution is predictable. From an evolutionary lens, predicting expression divergence following duplications can help understand how the rewiring of gene networks shaped the origin and diversification of traits, similarly to previous works on flowering (Calderwood et al. 2021), nodulation (Libourel et al. 2023), organogenesis (Julca et al. 2021), and response to changing environments (Hamann et al. 2021). From a biotechnological perspective, predicting gene expression evolution could help overcome barriers in crop breeding resulting from paralog diversification. For instance, gene duplications and subsequent paralog diversification within key domestication-related gene families in *Solanum* sp. hinders genotype-to-phenotype predictability, limiting knowledge transfer between related crops (Benoit et al. 2025).

While gene expression evolution is greatly shaped by a gene's duplication mode and functions, our findings highlight that time is an important (and often overlooked) variable. We demonstrate that paralog pairs originating from different duplication modes display differences in expression divergence, but such differences only exist for young duplicates, indicating that the effect of duplication mode is age-dependent. This is in line with our previous finding that WGDs have a significant impact on plant gene regulatory networks (GRNs), increasing the frequency of genetic circuits called GRN motifs, but WGD-derived motifs are lost over time (Almeida-Silva and Van de Peer 2023). Likewise, WGD-derived duplicates in soybean display greater levels of expression preservation, but only for those arising from the most recent WGD at ~13 million years ago (Almeida-Silva et al. 2020b).

By using high-resolution, spatially resolved transcriptomic data, this study advances our understanding of the tempo and mode of gene expression evolution following gene and genome duplications. Since our findings here are similar to previous

findings based on bulk RNA-seq data, a question that remains open is whether patterns of expression divergence are universal and independent in both low and high resolution (i.e., across tissues *versus* across cell types). Specifically, it remains unclear whether expression divergence follows an order in terms of resolution, with duplicates first diverging across cell types in a tissue and then across multiple tissues in an entire plant (or vice versa), or if duplicates diverge across cell types and across tissues independently. Addressing this question requires comparing bulk and spatial transcriptomic data from the same species under the same experimental conditions, which we believe will be possible in the near future as new data become available and technologies advance.

Despite the high resolution and throughput of spatial transcriptomics technologies, tissue capture areas are small, with slides of 6.5 mm x 6.5 mm for the 10x Visium, for instance. This limitation might raise questions as to whether our observations are replicable in other tissues and conditions. Since our findings are consistent across independent data sets, we believe they broadly represent the biology of gene expression divergence within tissue compartments. While only a few spatially resolved transcriptomes exist for plants, we expect a steep increase in the number of spatial transcriptomics studies in the coming years. As more data become available, an expanded version of this study will be valuable to assess whether our findings here are broadly valid, or if exceptions exist for particular tissues or species.

Conclusions

In this paper, we analyzed five spatially resolved plant transcriptomes to investigate the evolution of gene expression following gene and genome duplications. We found that genes originating from different duplication modes display differences in expression levels, expression breadth, spatial variability, and number of coexpression partners. Duplication mechanisms that preserve cis-regulatory landscapes tend to lead to more preserved expression profiles, but differences by duplication mode disappear over time. We also observed that genes originating from large-scale (or segmental) and small-scale duplications undergo different functional divergence outcomes. Our findings align with previous studies based on low-resolution (bulk) expression data, suggesting that patterns of expression divergence following gene duplications are universal within the same tissue and across different tissues.

Materials and Methods

Obtaining and preprocessing spatial transcriptomic data

We manually curated the literature to find spatially resolved plant transcriptomes obtained with high-throughput, sequencing-based technologies. Data sets representing *Arabidopsis* leaves (Xia et al. 2022), *Phalaenopsis* Big Chili inflorescences (Liu et al. 2022), and maize ear primordia (Wang et al. 2024) were obtained from STOmicsDB (Xu et al. 2024), while data sets representing barley seeds (Peirats-Llobet et al. 2023) and soybean nodules (Liu et al. 2023) were obtained from original publications. Data from H5AD files were imported with the R package *zellkonverter* (Luke Zappia, Aaron Lun), and stored as *SpatialExperiment* objects (Righelli et al. 2022). Data were filtered to remove genes with zero counts across all spots, spots with zero counts, and genes with less than 1 count in $\geq 99.9\%$ of the spots. Filtered data were preprocessed to correct for library size differences and compute log-transformed normalized counts with the R package *scater* (McCarthy et al. 2017).

Obtaining classified duplicated gene pairs and genes

Duplicated genes and gene pairs classified by duplication mode were obtained from *doubletroubledb* (Almeida-Silva and Van de Peer 2025). As *Phalaenopsis aphrodite* was not included in this database, we identified and classified its duplicated genes and gene pairs with *doubletrouble* (Almeida-Silva and Van de Peer 2025) using *Amborella trichopoda* as an outgroup. Duplicated gene pairs were classified as originating from segmental duplications (SD, duplication of large chromosomal regions, whole chromosome(s), or the whole genome), tandem duplications (TD, duplicated copy adjacent to the ancestral gene), proximal duplications (PD, duplicated copy near the ancestral gene, but separated by a few genes), retrotransposed (rTRD, insertion of retrotransposed gene), DNA transposed (dTRD, duplicated copy generated by the activity of DNA transposons), and dispersed duplications (DD, none of the above). Functional annotation and gene family assignments were obtained from PLAZA 5.0 (Van Bel et al. 2022).

Calculation of substitution rates and age-based grouping of paralogous gene pairs

Synonymous and nonsynonymous substitution rates (K_s and K_a , respectively) were calculated with the γ -MYN model (Wang et al. 2009) implemented in the R package doubletrouble (Almeida-Silva and Van de Peer 2025). Coding sequences (CDS) for the longest isoform of each gene were obtained from Ensembl Plants release 59 (Yates et al. 2022) and Orchidstra 2.0 (Chao et al. 2017). After removing pairs with $K_s > 2$ to avoid saturation at greater K_s values, we used doubletrouble (Almeida-Silva and Van de Peer 2025) to infer peaks in K_s distributions with Gaussian Mixture Models (GMMs), with pre-defined number of components based on literature evidence (Chen et al. 2024). Parameters of Gaussian components (i.e., K_s peaks) were used to group paralogous gene pairs into age groups, defined based on the mean ± 2 standard deviations of each component.

Inference of spatially variable genes and overrepresentation analyses

Spatially variable genes were inferred with DESpace (Cai et al. 2024) using annotated cell types as spatial clusters. Curated spatial domain annotations were obtained from original publications. Genes were considered spatially variable if FDR < 0.05 . Overrepresentation analyses were performed with the R package HybridExpress (Almeida-Silva et al. 2024).

Measuring expression divergence between paralogous gene pairs

We measured expression divergence between paralogous pairs using three metrics implemented in the R package exdiva (Almeida-Silva and Van de Peer 2024), namely i. Spearman's correlations coefficients between gene pairs, ii. co-occurrence in coexpression modules, and iii. dissimilarities between module eigengenes. To avoid zero-inflated correlation distributions due to the inherent sparsity of spatial transcriptomics data, we aggregated counts for spots into metaspots, as previously suggested (Morabito et al. 2023). To select a proper number of spots in each metaspot considering the dimensions of plant cells, we simulated increasingly large metaspots ($N = 10, 20, 30, 40$, and 50 spots), and inspected distributions of gene-gene correlations (Supplementary Fig. S2). Based on a trade-off between the reduction of noise arising from sparsity and preservation of expression resolution, we used metaspots of size 30. Following previous

papers (Gu et al. 2002; Makova and Li 2003; Ganko et al. 2007), Spearman's correlation coefficients were transformed using Fisher's z-transformation, defined as:

$$\frac{1}{2} \log\left(\frac{1 + \rho}{1 - \rho}\right)$$

Inference and analysis of gene coexpression networks

For each sample, we performed pseudobulk aggregation of counts ('sum' method) by spatial domain using the R package scuttle (McCarthy et al. 2017). Pseudobulk count matrices from the same species were combined, resulting in a single multi-sample pseudobulk matrix per species. Pseudobulk matrices were used to infer a gene coexpression network for each species using the R package BioNERO (Almeida-Silva and Venancio 2022). Signed hybrid networks (i.e., negative correlations set to zero) were inferred using Spearman's correlation coefficients as a similarity metric. The R package exdiva (Almeida-Silva and Van de Peer 2024) was used to classify paralogous pairs based co-occurrence in coexpression modules. Paralogous pairs were classified as 'preserved' if both genes were in the same module, and 'diverged' otherwise. To assess the significance of the frequencies of diverged pairs, we generated 10,000 degree-preserving simulated networks through node label permutation, and compared observed frequencies to null distributions. Dissimilarities between module eigengenes (i.e., first principal component of the expression matrix for a module, summarizing its expression profile) were used as module distance measures. Hub genes were identified with the function *get_hubs_gcn()* from the R package BioNERO (Almeida-Silva and Venancio 2022).

Classification of paralogous gene pairs into divergence classes

Paralogous gene pairs were classified into divergence classes based on the relative expression breadths (REBs) of their genes, similarly to a previous work (Casneuf et al. 2006). For each paralogous gene pair P with genes A and B, the relative expression breadth for gene A is defined as:

$$REB_A = \frac{S_A}{S_A \cup S_B}$$

where S_A represents the set of cell types in which gene A is expressed, and S_B represents the set of cell types in which gene B is expressed. Paralogous gene pairs were classified as displaying i. redundancy (i.e., both genes with REB >0.7), ii. asymmetric divergence (i.e., one gene with REB >0.7, and the other gene with REB <0.3), or iii. symmetric divergence (i.e., both genes with REB between 0.3 and 0.7).

Acknowledgements

YVdP acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (No. 833522). YVdP and FA-S acknowledge funding from Ghent University (Methusalem funding, BOF.MET.2021.0005.01).

Author contributions

FA-S and YVdP conceived the study; FA-S collected and analyzed the data; FA-S and YVdP wrote the manuscript. All authors read and approved the manuscript.

Competing interests

None declared.

Data availability statement

To ensure full reproducibility, all code and data used in this manuscript are available in a GitHub repository at <https://github.com/almeidasilvaf/PlantSpatialDiv>, and archived at Zenodo (Almeida-Silva 2025).

REFERENCES

- Almeida-Silva F.** almeidasilvaf/PlantSpatialDiv: Paper submission. 2025.
<https://doi.org/10.5281/zenodo.15518522>
- Almeida-Silva F, Moharana KC, Machado FB, and Venancio TM.** Exploring the complexity of soybean (*Glycine max*) transcriptional regulation using global gene co-expression networks. *Planta*. 2020a;**252**(6):104. <https://doi.org/10.1007/s00425-020-03499-8>
- Almeida-Silva F, Moharana KC, Machado FB, and Venancio TM.** Exploring the complexity of soybean (*Glycine max*) transcriptional regulation using global gene co-expression networks. *Planta*. 2020b;**252**(6):1–12.
- Almeida-Silva F, Prost-Boxoen L, and Van de Peer Y.** HYBRIDEXPRESS : an R/Bioconductor package for comparative transcriptomic analyses of hybrids and their progenitors. *New Phytol*. 2024;**243**(2):811–819. <https://doi.org/10.1111/nph.19862>
- Almeida-Silva F and Van de Peer Y.** Whole-genome Duplications and the Long-term Evolution of Gene Regulatory Networks in Angiosperms. *Mol Biol Evol*. 2023;**40**(7):msad141. <https://doi.org/10.1093/molbev/msad141>
- Almeida-Silva F and Van de Peer Y.** exdiva: A metric set to compute expression divergence between gene pairs. 2024.
- Almeida-Silva F and Van de Peer Y.** *doubletrouble* : an R/Bioconductor package for the identification, classification, and analysis of gene and genome duplications. *Bioinformatics*. 2025;**41**(2):btaf043. <https://doi.org/10.1093/bioinformatics/btaf043>
- Almeida-Silva F and Venancio TM.** BioNERO: an all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Funct Integr Genomics*. 2022;**22**(1):131–136.
- Assis R and Bachtrog D.** Neofunctionalization of young duplicate genes in *Drosophila*. *Proc Natl Acad Sci*. 2013;**110**(43):17409–17414. <https://doi.org/10.1073/pnas.1313759110>
- Batada NN, Hurst LD, and Tyers M.** Evolutionary and Physiological Importance of Hub Proteins. *PLoS Comput Biol*. 2006;**2**(7):e88. <https://doi.org/10.1371/journal.pcbi.0020088>
- Baumgarten A, Cannon S, Spangler R, and May G.** Genome-Level Evolution of Resistance Genes in *Arabidopsis thaliana*. *Genetics*. 2003;**165**(1):309–319. <https://doi.org/10.1093/genetics/165.1.309>
- Benoit M, Jenike KM, Satterlee JW, Ramakrishnan S, Gentile I, Hendelman A, Passalacqua MJ, Suresh H, Shohat H, Robitaille GM, et al.** Solanum pan-genetics reveals paralogues as contingencies in crop engineering. *Nature*. 2025;**640**(8057):135–145. <https://doi.org/10.1038/s41586-025-08619-6>
- Birchler JA and Veitia RA.** The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol*. 2010;**186**(1):54–62. <https://doi.org/10.1111/j.1469-8137.2009.03087.x>

- 579 **Birchler JA and Veitia RA.** Gene balance hypothesis: Connecting issues of dosage
580 sensitivity across biological disciplines. *Proc Natl Acad Sci.* 2012;**109**(37):14746–
581 14753. <https://doi.org/10.1073/pnas.1207726109>
- 582 **Birchler JA and Yang H.** The multiple fates of gene duplications: Deletion,
583 hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance
584 constraints, and neutral variation. *Plant Cell.* 2022;**34**(7):2466–2474.
585 <https://doi.org/10.1093/plcell/koac076>
- 586 **Blanc G and Wolfe KH.** Functional Divergence of Duplicated Genes Formed by Polyploidy
587 during Arabidopsis Evolution[W]. *Plant Cell.* 2004;**16**(7):1679–1691.
588 <https://doi.org/10.1105/tpc.021410>
- 589 **Cai P, Robinson MD, and Tiberi S.** *DESpace* : spatially variable gene detection via
590 differential expression testing of spatial clusters. *Bioinformatics.* 2024;**40**(2):btac027.
591 <https://doi.org/10.1093/bioinformatics/btac027>
- 592 **Calderwood A, Hepworth J, Woodhouse S, Bilham L, Jones DM, Tudor E, Ali M, Dean**
593 **C, Wells R, Irwin JA, et al.** Comparative transcriptomics reveals desynchronisation
594 of gene expression during the floral transition between *Arabidopsis* and *Brassica*
595 *rapa* cultivars. *Quant Plant Biol.* 2021;**2**:e4. <https://doi.org/10.1017/qpb.2021.6>
- 596 **Casneuf T, De Bodt S, Raes J, Maere S, and Van de Peer Y.** Nonrandom divergence of
597 gene expression following gene and genome duplications in the flowering plant
598 *Arabidopsis thaliana*. *Genome Biol.* 2006;**7**(2):R13. [https://doi.org/10.1186/gb-2006-](https://doi.org/10.1186/gb-2006-7-2-r13)
599 [7-2-r13](https://doi.org/10.1186/gb-2006-7-2-r13)
- 600 **Chae L, Kim T, Nilo-Poyanco R, and Rhee SY.** Genomic Signatures of Specialized
601 Metabolism in Plants. *Science.* 2014;**344**(6183):510–513.
602 <https://doi.org/10.1126/science.1252076>
- 603 **Chao Y-T, Yen S-H, Yeh J-H, Chen W-C, and Shih M-C.** Orchidstra 2.0—A
604 Transcriptomics Resource for the Orchid Family. *Plant Cell Physiol.* 2017:pcw220.
605 <https://doi.org/10.1093/pcp/pcw220>
- 606 **Chen H, Almeida-Silva F, Logghe G, Bonte D, and Van de Peer Y.** The Rise of Polyploids
607 During Environmental Catastrophes. 2024.
608 <https://doi.org/10.1101/2024.11.22.624806>
- 609 **Comai L.** The advantages and disadvantages of being polyploid. *Nat Rev Genet.*
610 2005;**6**(11):836–846. <https://doi.org/10.1038/nrg1711>
- 611 **Cox Jr KL, Gurazada SGR, Duncan KE, Czymmek KJ, Topp CN, and Meyers BC.**
612 Organizing your space: The potential for integrating spatial transcriptomics and 3D
613 imaging data in plants. *Plant Physiol.* 2022;**188**(2):703–712.
614 <https://doi.org/10.1093/plphys/kiab508>
- 615 **Fares MA, Keane OM, Toft C, Carretero-Paulet L, and Jones GW.** The Roles of Whole-
616 Genome and Small-Scale Duplications in the Functional Specialization of
617 *Saccharomyces cerevisiae* Genes. *PLoS Genet.* 2013;**9**(1):e1003176.
618 <https://doi.org/10.1371/journal.pgen.1003176>
- 619 **Freeling M.** Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem,
620 Whole-Genome, Segmental, or by Transposition. *Annu Rev Plant Biol.*
621 2009;**60**(1):433–453. <https://doi.org/10.1146/annurev.arplant.043008.092122>

- 622 **Freeling M and Thomas BC.** Gene-balanced duplications, like tetraploidy, provide
623 predictable drive to increase morphological complexity. *Genome Res.*
624 2006;**16**(7):805–814. <https://doi.org/10.1101/gr.3681406>
- 625 **Ganko EW, Meyers BC, and Vision TJ.** Divergence in Expression between Duplicated
626 Genes in Arabidopsis. *Mol Biol Evol.* 2007;**24**(10):2298–2309.
627 <https://doi.org/10.1093/molbev/msm158>
- 628 **Gu Z, Nicolae D, Lu HH-S, and Li W-H.** Rapid divergence in expression between duplicate
629 genes inferred from microarray data. *Trends Genet.* 2002;**18**(12):609–613.
630 [https://doi.org/10.1016/S0168-9525\(02\)02837-8](https://doi.org/10.1016/S0168-9525(02)02837-8)
- 631 **Ha M, Kim E-D, and Chen ZJ.** Duplicate genes increase expression diversity in closely
632 related species and allopolyploids. *Proc Natl Acad Sci.* 2009;**106**(7):2295–2300.
633 <https://doi.org/10.1073/pnas.0807350106>
- 634 **Hahn MW and Kern AD.** Comparative Genomics of Centrality and Essentiality in Three
635 Eukaryotic Protein-Interaction Networks. *Mol Biol Evol.* 2005;**22**(4):803–806.
636 <https://doi.org/10.1093/molbev/msi072>
- 637 **Hamann E, Pauli CS, Joly-Lopez Z, Groen SC, Rest JS, Kane NC, Purugganan MD, and**
638 **Franks SJ.** Rapid evolutionary changes in gene expression in response to climate
639 fluctuations. *Mol Ecol.* 2021;**30**(1):193–206. <https://doi.org/10.1111/mec.15583>
- 640 **Iohannes SD and Jackson D.** Tackling redundancy: genetic mechanisms underlying
641 paralog compensation in plants. *New Phytol.* 2023;**240**(4):1381–1389.
642 <https://doi.org/10.1111/nph.19267>
- 643 **Jeong H, Mason SP, Barabási A-L, and Oltvai ZN.** Lethality and centrality in protein
644 networks. *Nature.* 2001;**411**(6833):41–42. <https://doi.org/10.1038/35075138>
- 645 **Ji W, Osbourn A, and Liu Z.** Understanding metabolic diversification in plants: branchpoints
646 in the evolution of specialized metabolism. *Philos Trans R Soc B Biol Sci.*
647 2024;**379**(1914):20230359. <https://doi.org/10.1098/rstb.2023.0359>
- 648 **Jiang X and Assis R.** Rapid functional divergence after small-scale gene duplication in
649 grasses. *BMC Evol Biol.* 2019;**19**(1):97. <https://doi.org/10.1186/s12862-019-1415-2>
- 650 **Julca I, Ferrari C, Flores-Tornero M, Proost S, Lindner A-C, Hackenberg D,**
651 **Steinbachová L, Michaelidis C, Gomes Pereira S, Misra CS, et al.** Comparative
652 transcriptomic analysis reveals conserved programmes underpinning organogenesis
653 and reproduction in land plants. *Nat Plants.* 2021;**7**(8):1143–1159.
654 <https://doi.org/10.1038/s41477-021-00958-2>
- 655 **Kiselev VY, Andrews TS, and Hemberg M.** Challenges in unsupervised clustering of
656 single-cell RNA-seq data. *Nat Rev Genet.* 2019;**20**(5):273–282.
657 <https://doi.org/10.1038/s41576-018-0088-9>
- 658 **Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, and Soltis PS.** Impact of
659 whole-genome duplication events on diversification rates in angiosperms. *Am J Bot.*
660 2018;**105**(3):348–363. <https://doi.org/10.1002/ajb2.1060>
- 661 **Li Z, Zhang H, Ge S, Gu X, Gao G, and Luo J.** Expression pattern divergence of duplicated
662 genes in rice. *BMC Bioinformatics.* 2009;**10**(S6):S8. <https://doi.org/10.1186/1471-2105-10-S6-S8>
- 663

664 **Libourel C, Keller J, Brichet L, Cazalé A-C, Carrère S, Vernié T, Couzigou J-M, Callot**
665 **C, Dufau I, Cauet S, et al.** Comparative phylotranscriptomics reveals ancestral and
666 derived root nodule symbiosis programmes. *Nat Plants*. 2023;**9**(7):1067–1080.
667 <https://doi.org/10.1038/s41477-023-01441-w>

668 **Liu C, Leng J, Li Y, Ge T, Li J, Chen Y, Guo C, and Qi J.** A spatiotemporal atlas of
669 organogenesis in the development of orchid flowers. *Nucleic Acids Res*.
670 2022;**50**(17):9724–9737.

671 **Liu Z, Kong X, Long Y, Liu S, Zhang H, Jia J, Cui W, Zhang Z, Song X, and Qiu L.**
672 Integrated single-nucleus and spatial transcriptomics captures transitional states in
673 soybean nodule maturation. *Nat Plants*. 2023;**9**(4):515–524.

674 **Lloyd JP, Seddon AE, Moghe GD, Simenc MC, and Shiu S-H.** Characteristics of Plant
675 Essential Genes Allow for within- and between-Species Prediction of Lethal Mutant
676 Phenotypes. *Plant Cell*. 2015;**27**(8):2133–2147. <https://doi.org/10.1105/tpc.15.00051>

677 **Loehlin DW and Carroll SB.** Expression of tandem gene duplicates is often greater than
678 twofold. *Proc Natl Acad Sci*. 2016;**113**(21):5988–5992.
679 <https://doi.org/10.1073/pnas.1605886113>

680 **Luke Zappia, Aaron Lun.** zellkonverter.
681 <https://doi.org/10.18129/B9.BIOC.ZELLKONVERTER>

682 **Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, and Van De Peer Y.**
683 Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci*.
684 2005;**102**(15):5454–5459. <https://doi.org/10.1073/pnas.0501102102>

685 **Makova KD and Li W-H.** Divergence in the Spatial Pattern of Gene Expression Between
686 Human Duplicate Genes. *Genome Res*. 2003;**13**(7):1638–1645.
687 <https://doi.org/10.1101/gr.1133803>

688 **McCarthy DJ, Campbell KR, Lun ATL, and Wills QF.** Scater: pre-processing, quality
689 control, normalization and visualization of single-cell RNA-seq data in R.
690 *Bioinformatics*. 2017;**33**(8):1179–1186. <https://doi.org/10.1093/bioinformatics/btw777>

691 **Morabito S, Reese F, Rahimzadeh N, Miyoshi E, and Swarup V.** hdWGCNA identifies co-
692 expression networks in high-dimensional transcriptomics data. *Cell Rep Methods*.
693 2023;**3**(6):100498. <https://doi.org/10.1016/j.crmeth.2023.100498>

694 **Ohno S.** *Evolution by Gene Duplication* (Springer Berlin Heidelberg: Berlin, Heidelberg).
695 <https://doi.org/10.1007/978-3-642-86659-3>

696 **One Thousand Plant Transcriptomes Initiative.** One thousand plant transcriptomes and
697 the phylogenomics of green plants. *Nature*. 2019;**574**(7780):679–685.
698 <https://doi.org/10.1038/s41586-019-1693-2>

699 **Panchy N, Lehti-Shiu M, and Shiu S-H.** Evolution of Gene Duplication in Plants. *Plant*
700 *Physiol*. 2016;**171**(4):2294–2316. <https://doi.org/10.1104/pp.16.00523>

701 **Peirats-Llobet M, Yi C, Liew LC, Berkowitz O, Narsai R, Lewsey MG, and Whelan J.**
702 Spatially resolved transcriptomic analysis of the germinating barley grain. *Nucleic*
703 *Acids Res*. 2023;**51**(15):7798–7819.

704 **Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, and Paterson AH.** Gene duplication

705 and evolution in recurring polyploidization–diploidization cycles in plants. *Genome*
706 *Biol.* 2019;**20**(1):38. <https://doi.org/10.1186/s13059-019-1650-2>

707 **Raman K, Damaraju N, and Joshi GK.** The organisational structure of protein networks:
708 revisiting the centrality–lethality hypothesis. *Syst Synth Biol.* 2014;**8**(1):73–81.
709 <https://doi.org/10.1007/s11693-013-9123-5>

710 **Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, and Qi J.** Widespread Whole
711 Genome Duplications Contribute to Genome Complexity and Species Diversity in
712 Angiosperms. *Mol Plant.* 2018;**11**(3):414–428.
713 <https://doi.org/10.1016/j.molp.2018.01.002>

714 **Resendis-Antonio O, Freyre-González JA, Menchaca-Méndez R, Gutiérrez-Ríos RM,**
715 **Martínez-Antonio A, Ávila-Sánchez C, and Collado-Vides J.** Modular analysis of
716 the transcriptional regulatory network of *E. coli*. *Trends Genet.* 2005;**21**(1):16–20.
717 <https://doi.org/10.1016/j.tig.2004.11.010>

718 **Righelli D, Weber LM, Crowell HL, Pardo B, Collado-Torres L, Ghazanfar S, Lun ATL,**
719 **Hicks SC, and Risso D.** SpatialExperiment: infrastructure for spatially-resolved
720 transcriptomics data in R using Bioconductor. *Bioinformatics.* 2022;**38**(11):3128–
721 3131. <https://doi.org/10.1093/bioinformatics/btac299>

722 **Rizzon C, Ponger L, and Gaut BS.** Striking Similarities in the Genomic Distribution of
723 Tandemly Arrayed Genes in *Arabidopsis* and Rice. *PLoS Comput Biol.*
724 2006;**2**(9):e115. <https://doi.org/10.1371/journal.pcbi.0020115>

725 **Shi T, Rahmani RS, Gugger PF, Wang M, Li H, Zhang Y, Li Z, Wang Q, Van De Peer Y,**
726 **Marchal K, et al.** Distinct Expression and Methylation Patterns for Genes with
727 Different Fates following a Single Whole-Genome Duplication in Flowering Plants.
728 *Mol Biol Evol.* 2020;**37**(8):2394–2413. <https://doi.org/10.1093/molbev/msaa105>

729 **Soltis PS and Soltis DE.** Ancient WGD events as drivers of key innovations in
730 angiosperms. *Curr Opin Plant Biol.* 2016;**30**:159–165.
731 <https://doi.org/10.1016/j.pbi.2016.03.015>

732 **Song H-S, McClure R, Bernstein H, Overall C, Hill E, and Beliaev A.** Integrated in silico
733 Analyses of Regulatory and Metabolic Networks of *Synechococcus* sp. PCC 7002
734 Reveal Relationships between Gene Centrality and Essentiality. *Life.*
735 2015;**5**(2):1127–1140. <https://doi.org/10.3390/life5021127>

736 **Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW,**
737 **Sessa EB, and Harmon LJ.** Nested radiations and the pulse of angiosperm
738 diversification: increased diversification rates often follow whole genome duplications.
739 *New Phytol.* 2015;**207**(2):454–467. <https://doi.org/10.1111/nph.13491>

740 **Tasdighian S, Van Bel M, Li Z, Van De Peer Y, Carretero-Paulet L, and Maere S.**
741 Reciprocally Retained Genes in the Angiosperm Lineage Show the Hallmarks of
742 Dosage Balance Sensitivity. *Plant Cell.* 2017;**29**(11):2766–2785.
743 <https://doi.org/10.1105/tpc.17.00313>

744 **Ugalde JM and Straube H.** New genes on the block: Neofunctionalization of tandem
745 duplicate genes with putative new functions in *Arabidopsis*. *Plant Physiol.*
746 2023;**192**(4):2574–2576. <https://doi.org/10.1093/plphys/kiad271>

747 **Van Bel M, Silvestri F, Weitz EM, Kreft L, Botzki A, Coppens F, and Vandepoele K.**

748 PLAZA 5.0: extending the scope and power of comparative and functional genomics
749 in plants. *Nucleic Acids Res.* 2022;**50**(D1):1468–1474.

750 **Van De Peer Y, Mizrahi E, and Marchal K.** The evolutionary significance of polyploidy.
751 *Nat Rev Genet.* 2017;**18**(7):411–424. <https://doi.org/10.1038/nrg.2017.26>

752 **Veitia RA, Bottani S, and Birchler JA.** Cellular reactions to gene dosage imbalance:
753 genomic, transcriptomic and proteomic effects. *Trends Genet.* 2008;**24**(8):390–397.
754 <https://doi.org/10.1016/j.tig.2008.05.005>

755 **Wang D-P, Wan H-L, Zhang S, and Yu J.** γ -MYN: a new algorithm for estimating Ka and Ks
756 with consideration of variable substitution rates. *Biol Direct.* 2009;**4**(1):20.
757 <https://doi.org/10.1186/1745-6150-4-20>

758 **Wang Y, Luo Y, Guo X, Li Y, Yan J, Shao W, Wei W, Wei X, Yang T, and Chen J.** A
759 spatial transcriptome map of the developing maize ear. *Nat Plants.* 2024;**10**(5):815–
760 827.

761 **Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, and**
762 **Freeling M.** Following Tetraploidy in Maize, a Short Deletion Mechanism Removed
763 Genes Preferentially from One of the Two Homeologs. *PLoS Biol.*
764 2010;**8**(6):e1000409. <https://doi.org/10.1371/journal.pbio.1000409>

765 **Xia K, Sun H-X, Li J, Li J, Zhao Y, Chen L, Qin C, Chen R, Chen Z, and Liu G.** The
766 single-cell stereo-seq reveals region-specific cell subtypes and transcriptome
767 profiling in Arabidopsis leaves. *Dev Cell.* 2022;**57**(10):1299–1310.

768 **Xu M, Du Q, Tian C, Wang Y, and Jiao Y.** Stochastic gene expression drives mesophyll
769 protoplast regeneration. *Sci Adv.* 2021;**7**(33):eabg8466.
770 <https://doi.org/10.1126/sciadv.abg8466>

771 **Xu Z, Wang W, Yang T, Li L, Ma X, Chen J, Wang J, Huang Y, Gould J, and Lu H.**
772 STOmicsDB: a comprehensive database for spatial transcriptomics data sharing,
773 analysis and visualization. *Nucleic Acids Res.* 2024;**52**(D1):D1053–D1061.

774 **Yates AD, Allen J, Amode RM, Azov AG, Barba M, Becerra A, Bhai J, Campbell LI,**
775 **Carbajo Martinez M, Chakiachvili M, et al.** Ensembl Genomes 2022: an expanding
776 genome resource for non-vertebrates. *Nucleic Acids Res.* 2022;**50**(D1):D996–D1003.
777 <https://doi.org/10.1093/nar/gkab1007>

778 **Yin R, Xia K, and Xu X.** Spatial transcriptomics drives a new era in plant research. *Plant J.*
779 2023;**116**(6):1571–1581. <https://doi.org/10.1111/tpj.16437>

780 **Yu H, Greenbaum D, Xin Lu H, Zhu X, and Gerstein M.** Genomic analysis of essentiality
781 within protein networks. *Trends Genet.* 2004;**20**(6):227–231.
782 <https://doi.org/10.1016/j.tig.2004.04.008>

FIGURES

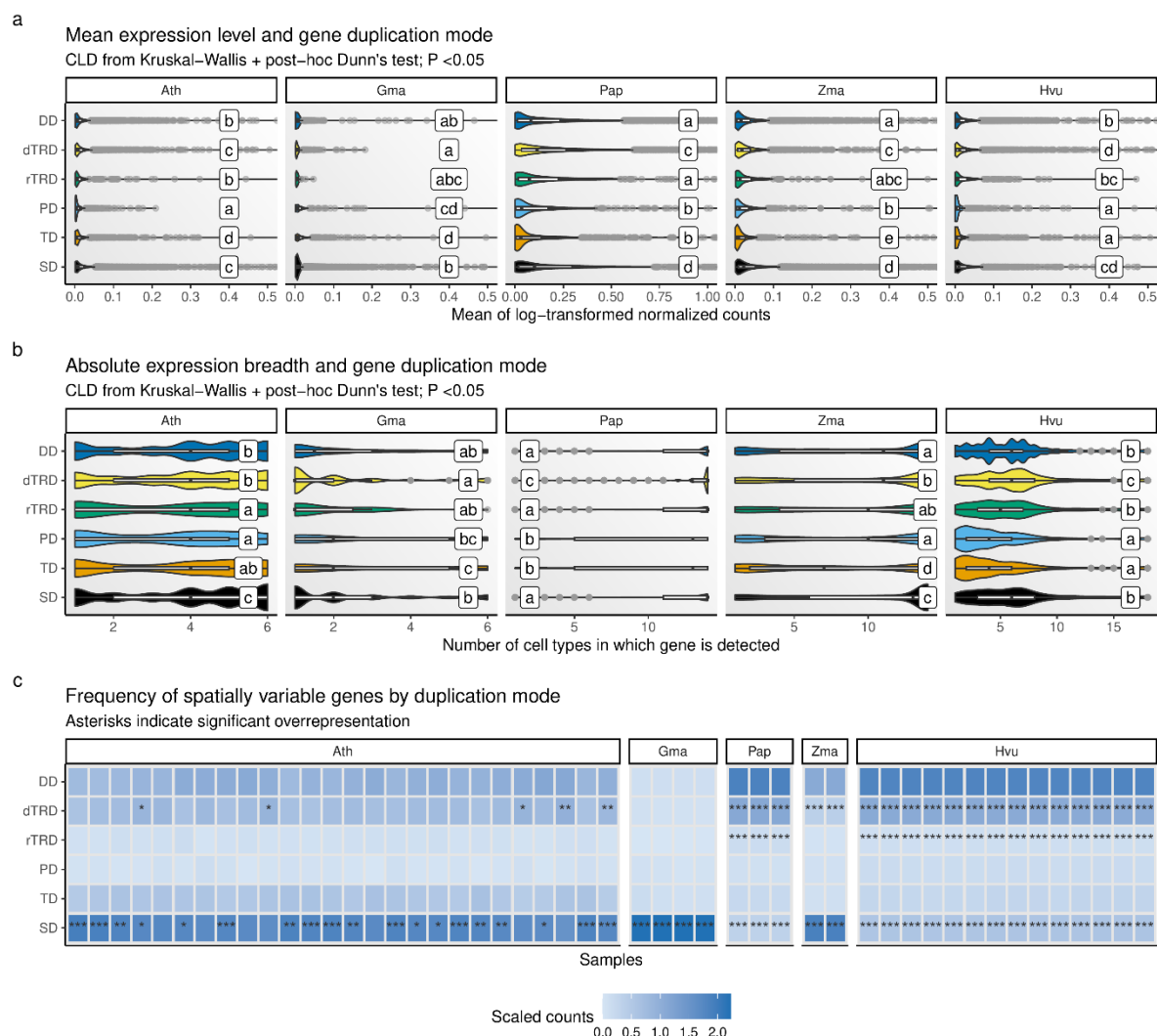


Fig. 1. Impact of gene duplications on gene expression profiles. **a,b.** Mean expression levels and expression breadths of genes originating from different modes. Expression breadths indicate the number of different cell types in which a gene is expressed. Compact letter displays (CLDs) indicate significant differences obtained with a Kruskal-Wallis test followed by a post-hoc Dunn's test ($P < 0.05$). **c.** Frequency of spatially variable genes among genes originating from different duplication modes. Significant associations between duplication modes and spatial expression variability are indicated with asterisks, and were obtained with Fisher's exact tests ($P < 0.05$).

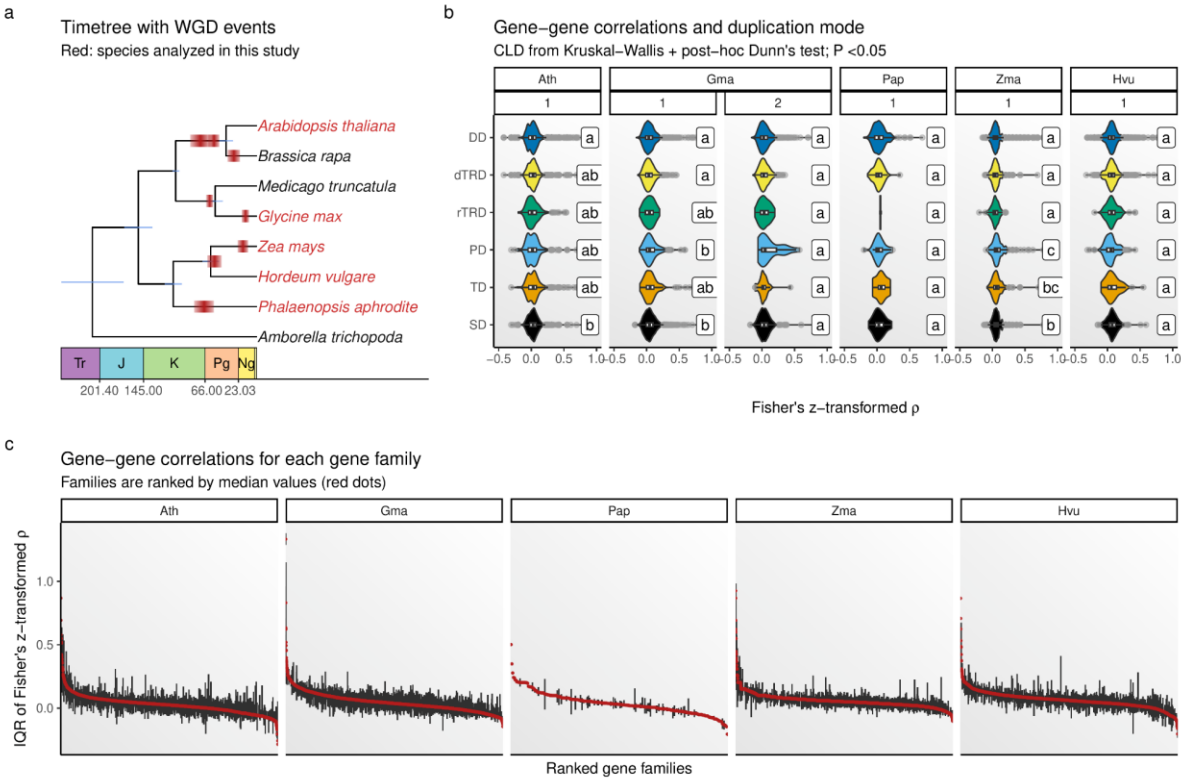


Fig. 2. Expression divergence between paralogous gene pairs. **a.** Ultrametric species tree (a.k.a. ‘timetree’) with rectangles indicating polyploidization events. Phylogenetic positions and dates of polyploidization events were obtained from AngioWGD (Chen et al. 2024). Names of species included in this study are highlighted in red. **b.** Spearman’s correlation coefficients between paralogous gene pairs originating from different duplication modes. Numbers in facet names indicate age groups obtained from parameters of peaks in K_s distributions, with 1 corresponding to younger paralogs and 2 corresponding to older paralogs. Paralogs in each age group have K_s values near peaks in K_s distributions and, hence, originated by different duplication modes at around the same time of polyploidization events indicated in **a**. Compact letter displays (CLDs) indicate significant differences obtained with a Kruskal–Wallis test followed by a post-hoc Dunn’s test ($P < 0.05$). **c.** Interquartile range (IQR) of correlation coefficients between paralogous gene pairs from different gene families. Red dots in the center of IQRs indicate median correlations for each gene family, which were used to arrange families in descending order. Spearman’s correlation coefficients were z-transformed.

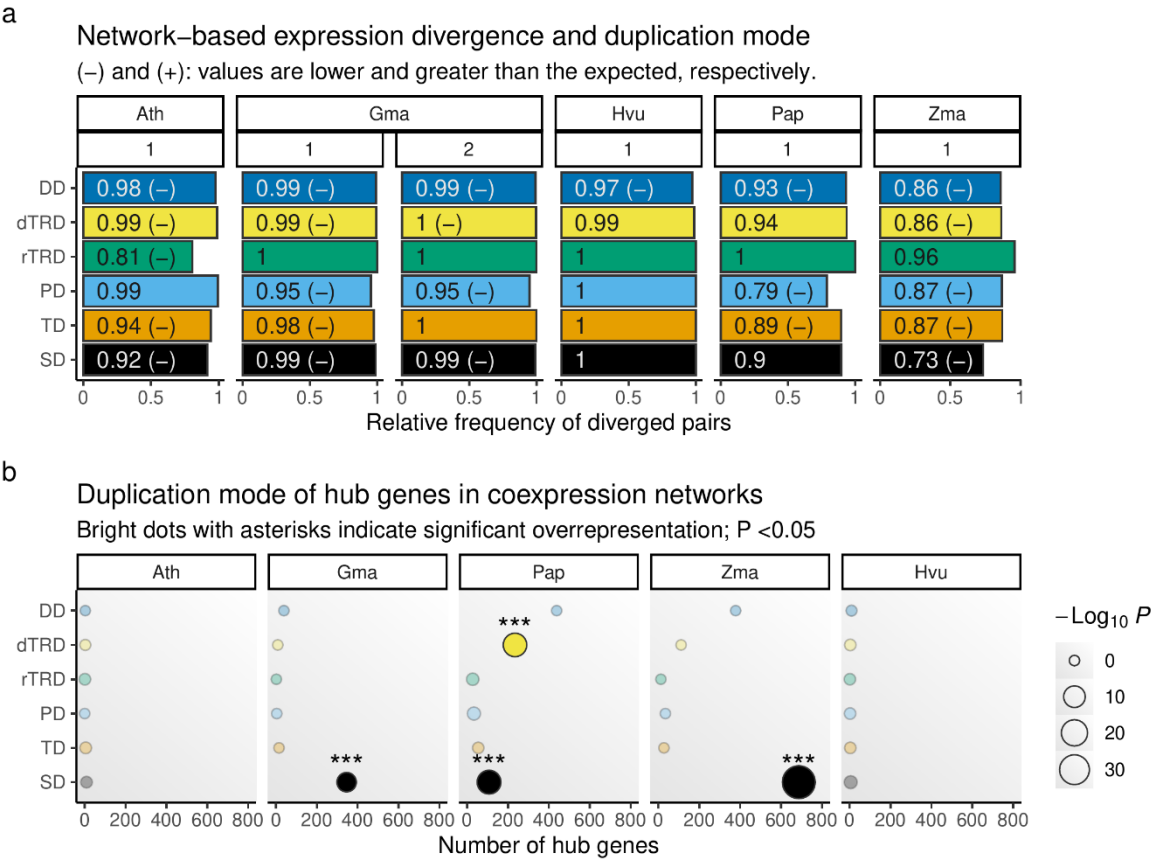


Fig. 3. Impact of gene duplications on gene coexpression networks. **a.** Network-based expression divergence for paralogous gene pairs originating from different duplication modes. Paralog pairs are interpreted as divergent if genes therein occur in different coexpression modules. Observed frequencies of diverged pairs were compared to null distributions generated from 10,000 degree-preserving simulated networks to assess statistical significance. Numbers in facet names indicate age groups obtained from parameters of peaks in K_s distributions, with 1 corresponding to younger paralogs and 2 corresponding to older paralogs. **b.** Frequency of hub genes originating from each duplication mode. Asterisks indicate significant associations between hub status and duplication modes, assessed with Fisher's exact tests ($P < 0.05$).

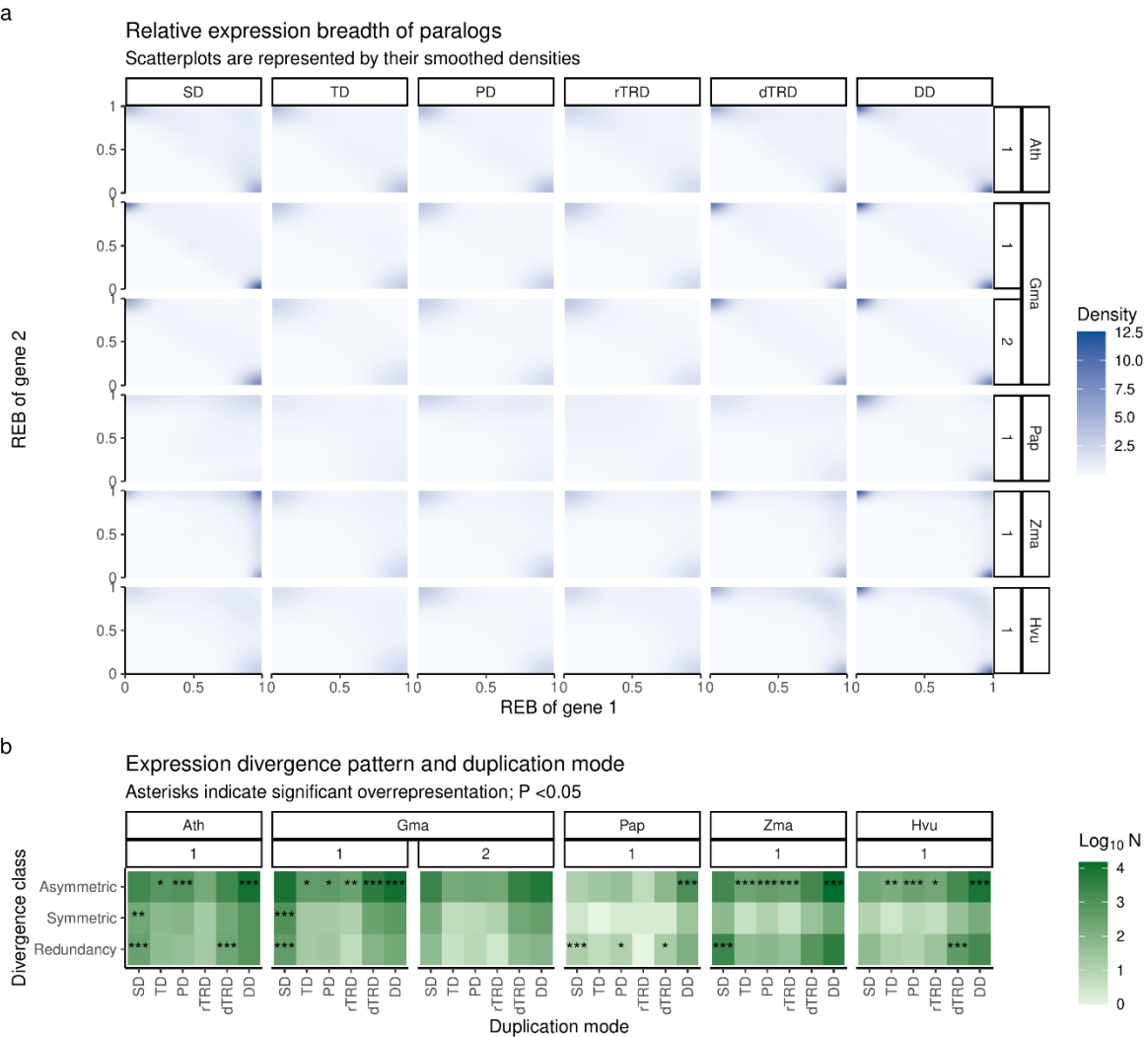


Fig. 4. Divergence classes of paralogous gene pairs originating from different duplication modes. **a.** Relative expression breadth of genes in a paralog pair. For each paralog pair, a gene's relative expression breadth is calculated as the ratio of the number of cell types in which it is expressed and the union of cell types in which both genes in the pair are expressed (see Materials and Methods for details). Scatterplots are represented by their smoothed densities. Numbers in facet names indicate age groups obtained from parameters of peaks in K_s distributions, with 1 corresponding to younger paralogs and 2 corresponding to older paralogs. **b.** Frequency of genes in each divergence class originating from each duplication mode. Asterisks indicate significant associations between divergence classes and duplication modes, assessed with Fisher's exact tests ($P < 0.05$).

846 **TABLES**

847

848 **Table 1. Spatial transcriptomic data sets included in this study.** *N*, number of tissue slides. DPI, days post-infection. HAI, hours after
849 imbibition.

850

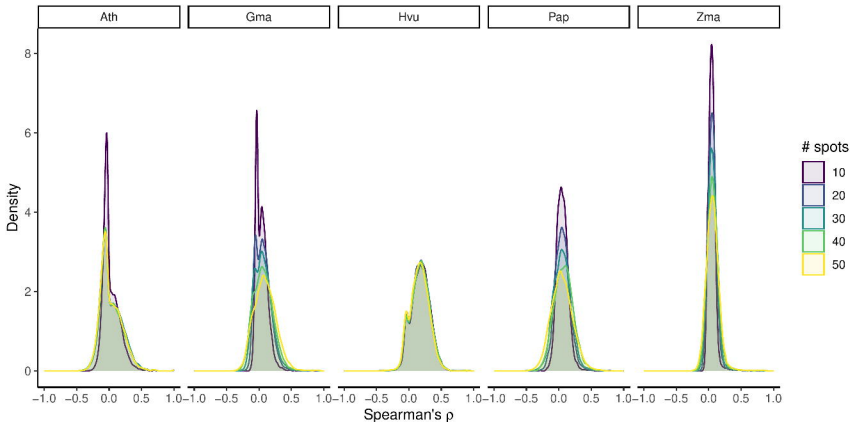
Reference	Species	Tissue	<i>N</i>	Condition	Technology
(Liu et al. 2023)	<i>Glycine max</i>	Infected nodules	4	12 and 21 DPI	Stereo-seq
(Liu et al. 2022)	<i>Phalaenopsis</i> Chili*	Big Inflorescences	3	Buds at successive stages	10X Visium
(Wang et al. 2024)	<i>Zea mays</i>	Developing ear primordia	4	11 or 12-leaf stage	Stereo-seq
(Xia et al. 2022)	<i>Arabidopsis thaliana</i>	Cauline leaves	26	Mature	Stereo-seq
(Peirats-Llobet et al. 2023)	<i>Hordeum vulgare</i>	Germinating seeds	20	0, 1, 3, 6, and 24 HAI	10X Visium

851 * Reads from *Phalaenopsis* Big Chilli were mapped to the genome of *Phalaenopsis aphrodite* in the original publication.

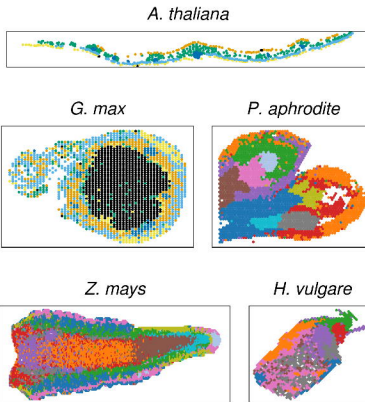
852

Distribution of gene–gene correlations for increasingly large metaspots

Correlations between paralogous gene pairs only



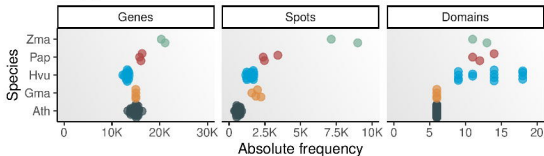
a



b

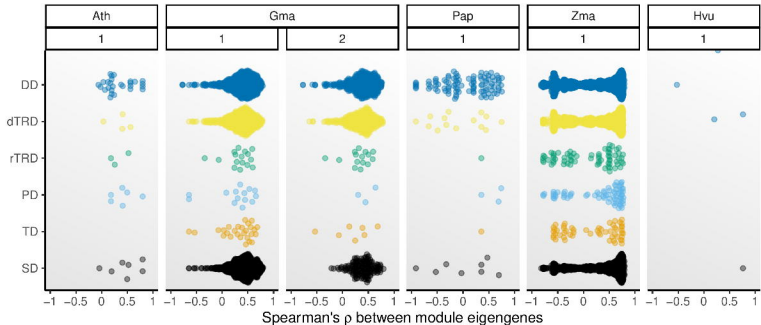
Summary sample statistics across species

Only genes with >1 count in at least 0.1% of the spots



Gene–gene similarities for pairs in different modules

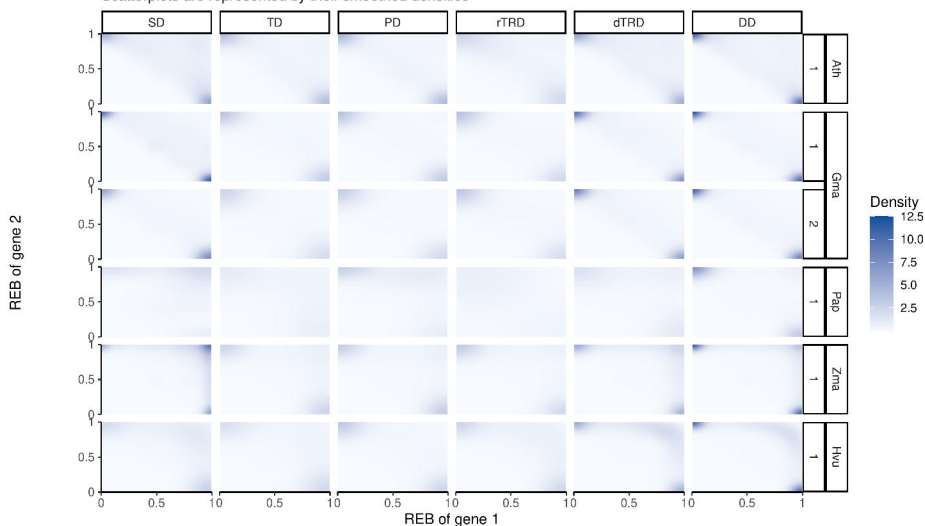
Similarities are measured by correlations between module eigengenes



a

Relative expression breadth of paralogs

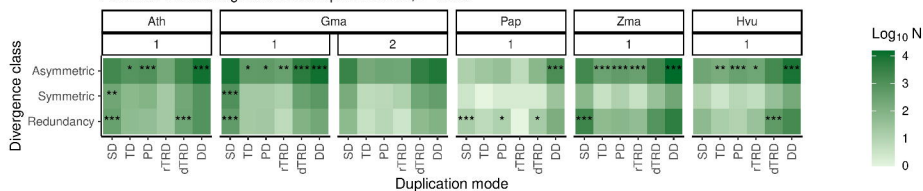
Scatterplots are represented by their smoothed densities



b

Expression divergence pattern and duplication mode

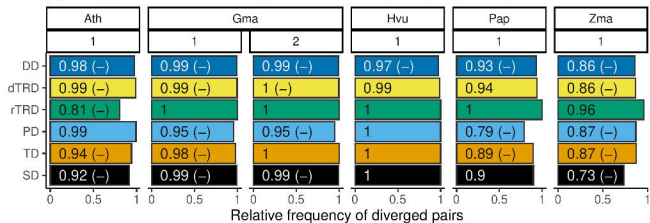
Asterisks indicate significant overrepresentation; $P < 0.05$



a

Network-based expression divergence and duplication mode

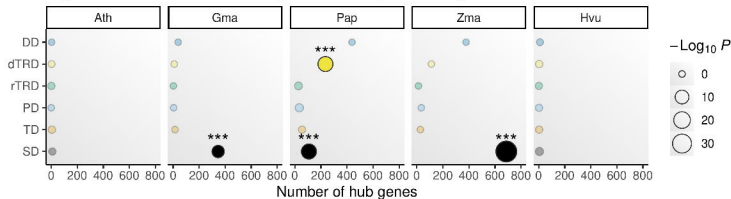
(-) and (+): values are lower and greater than the expected, respectively.



b

Duplication mode of hub genes in coexpression networks

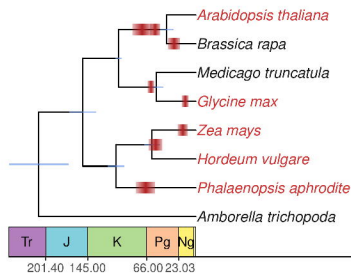
Bright dots with asterisks indicate significant overrepresentation; $P < 0.05$



a

Timetree with WGD events

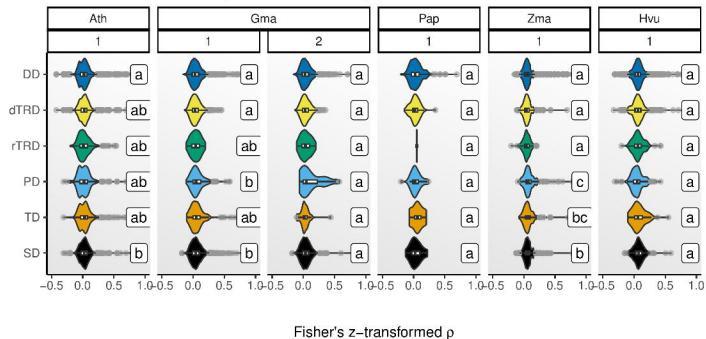
Red: species analyzed in this study



b

Gene–gene correlations and duplication mode

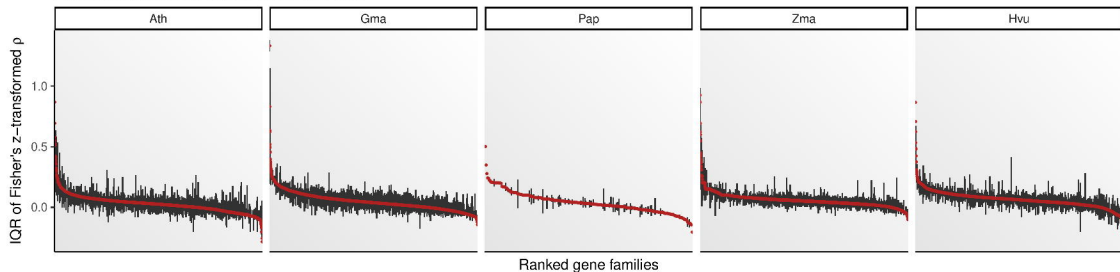
CLD from Kruskal–Wallis + post-hoc Dunn's test; $P < 0.05$



c

Gene–gene correlations for each gene family

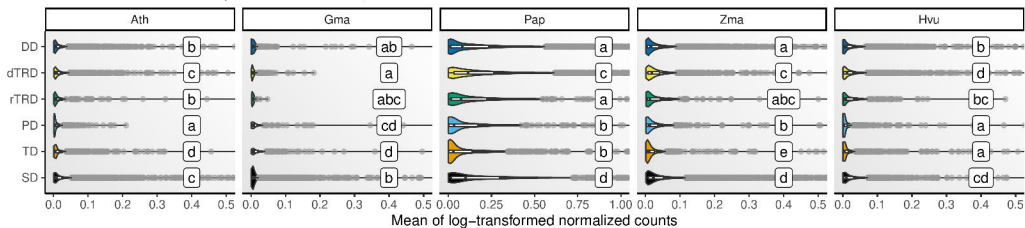
Families are ranked by median values (red dots)



a

Mean expression level and gene duplication mode

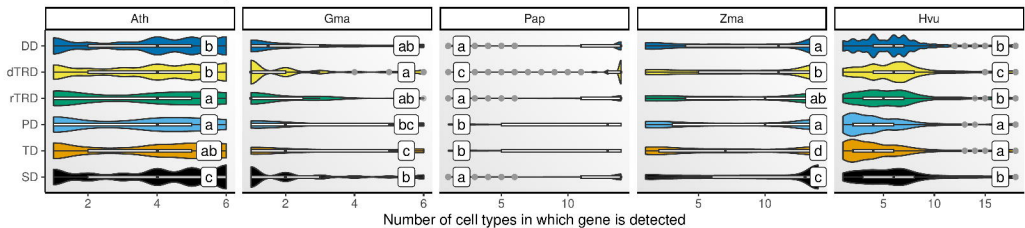
CLD from Kruskal–Wallis + post-hoc Dunn's test; $P < 0.05$



b

Absolute expression breadth and gene duplication mode

CLD from Kruskal–Wallis + post-hoc Dunn's test; $P < 0.05$



c

Frequency of spatially variable genes by duplication mode

Asterisks indicate significant overrepresentation

