PlantaeViz: a comprehensive platform for the analysis, integration, and visualization of public omics data across plant and related species

Antonio Santiago[1], Luis Orduña[1], José David Fernández[2], Álvaro Vidal[1,3,4], Iñigo de Martín-Agirre[1], Chen Zhang[1], Purificación Lisón[5], David Posé[6], Carmen Martín-Pizarro[6], Ioannis Kandylas[7], Iban Eduardo[7], Elena Monte[8], M. Águila Ruiz-Sola[9], Aurélien Devillars[10], Justin Joseph[10], Alessandro Vannozzi[10], Elena A. Vidal[2], David Navarro-Payá[1+], José Tomás Matus[1+]

[1] Institute for Integrative Systems Biology (I2SYSBIO, UV-CSIC), Valencia, Spain

[2] Center for Genomics and Bioinformatics, Universidad Mayor (Chile)

[3] Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy

[4] Center Agriculture Food Environment (C3A), University of Trento, San Michele all'Adige, Italy

[5] Instituto de Biología Molecular y Celular de Plantas (IBMCP), Valencia, Spain

[6] Instituto de Hortofruticultura Subtropical y Mediterránea (IHSM), Universidad de Málaga-CSIC, Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, UMA, Málaga 29071, Spain

[7] IRTA (Institut de Recerca i Tecnologia Agroalimentàries), Barcelona, Spain.

[8] Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra 08193 Barcelona, Spain

[9] Institute of Plant Biochemistry and Photosynthesis (IBVF), CSIC, Universidad de Sevilla, C. Américo Vespucio, 49, 41092 Sevilla, Spain.

[10] Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), Campus of Agripolis, University of Padova, Viale dell'Università, Legnaro, Italy

+ **Correspondence: David Navarro-Payá (david.navarro.paya@csic.es), Tomás Matus Picero (tomas.matus@uv.es)**

Abstract

The increasing availability of omics data from non-model plant species has created a pressing need for centralized, user-friendly platforms that maximize the utility of these datasets in a FAIR manner. Here, we introduce PlantaeViz, a web-based tool designed

for the integration, visualization, and analysis of multi-omics data across a wide range of plant species. PlantaeViz offers advanced functionalities, including gene catalogues built from curated literature, transcriptomic meta-analyses presented as gene expression atlases, gene co-expression and regulatory networks with on-the-fly ontology analyses, cistrome visualization and metabolomics-transcriptomics integration, among other tools, providing a robust framework for hypothesis generation and biological interpretation. *Gene Cards* applications, tailored to each plant species, provide both community-curated and automatically generated functional annotation information. One of the platform's core features is its big data approach: over 58,000 publicly available SRA transcriptomic samples have been processed and visualized to date. Significant efforts have been made in orthology assessment using multiple layers of evidence, as well as in the automatic classification and standardization of omics metadata through regular expressions and data mining. As a result, around 90% of transcriptomic runs have been successfully classified according to sample tissue. These data have been used to construct gene networks via computationally intensive methods based on diverse algorithms. We present a study case to illustrate the platform's integration and exploratory capabilities. PlantaeViz bridges genomics and functional knowledge between model and non-model plant species and aims to expand its species catalogue of species in the future, democratizing access to large-scale plant omics data. Further developments will include the incorporation of additional data types, and the implementation of new tools to further support plant research across diverse biological contexts.

Introduction

The rapid accumulation of omics data in non-model plant species has reached levels comparable to those of well-studied model organisms in the recent past. For example, Illumina-based transcriptomic data in grapevine has exceeded 10,000 runs in the NCBI Sequence Read Archive (SRA) (Sayers et al., 2022), a milestone that the Arabidopsis community only reached in 2017. This growth calls for up-to-date, centralized web-based platforms dedicated to these emerging systems. Such platforms are essential for enabling the efficient reuse, integration, and analysis of public datasets. In their absence, the full potential of the knowledge generated by the research community remains underexploited.

While the volume of omics data in non-model crop species now rivals that of model organisms from just a few years ago, the depth of biological insight remains significantly limited. This disparity is largely due to the smaller and more fragmented research communities involved in these species. *In silico* approaches, such as sequence-based functional annotation or gene co-expression analyses, offer powerful strategies to bridge this gap by enabling the prediction of gene functions. Gene co-expression networks, which operate on the principle of 'guilt-by-association', have been widely applied in non-model crops (Childs et al., 2011; Liesecke et al., 2019; Orduña et al., 2023). In these networks, genes that are consistently co-expressed across diverse conditions are more likely to share functional relationships, even if their specific roles remain unknown. As such, co-expression networks provide a valuable framework for inferring the functions of poorly annotated genes based on their association with better-characterized counterparts.

The integration of multi-omic datasets, spanning genomics, transcriptomics, and metabolomics, is central to systems biology, offering a holistic understanding of organismal function. Yet, despite their potential, these datasets are often underexploited due to the absence of accessible, intuitive tools for their visualization, exploration, and reanalysis. Publicly available omics data can reveal complex gene expression patterns across tissues, developmental stages, and environmental conditions, providing insights into regulatory mechanisms and phenotypic traits that frequently transcend the original scope of the study. Unlocking this latent value requires platforms that empower both experts and non-specialists to mine and reinterpret existing datasets in meaningful ways.

Model organisms such as *Arabidopsis thaliana* benefit from comprehensive, well-maintained databases, notably The Arabidopsis Information Resource (TAIR) (Huala et al., 2001). TAIR integrates omics resources through standardized gene nomenclature, high-quality genome assemblies, and curated structural and functional annotations. It also offers a suite of tools for exploring large-scale datasets, including access to mutant collections, literature-based gene function curation, and gene expression profiles. In contrast, existing databases for non-model plants are often limited in scope and functionality. They frequently lack advanced features such as gene co-expression network analysis or tissue- and condition-specific expression exploration. These are essential tools for effective hypothesis generation and systems-

level research. Moreover, broad cross-species platforms like STRING (Szklarczyk et al., 2023), while valuable, tend to underperform for less-characterized plant species, as they rely on outdated annotations, legacy gene identifiers from obsolete older genome assemblies or annotations, or microarray probe-specific identifiers, which limits their utility for more modern, RNA-seq-based studies.

In this study, we present PlantaeViz (https://plantaeviz.tomsbiolab.com/), a versatile and centralized platform for the visualization and analysis of functional, and genomic data across a range of plant species. The platform is organized into modular suites, which currently include grapevine (*Vitis vinifera*), tomato (*Solanum lycopersicum*), mulberry (*Morus alba*), *Cannabis sativa,* and *Arabidopsis thaliana* in its initial release. Unlike model organism-centric resources, PlantaeViz incorporates the latest genome assemblies and annotations for these species. Its development has involved extensive manual curation, including literature-based compilation of gene catalogues and the creation of hierarchical ontologies to support consistent metadata classification across datasets. Notably, the platform integrates *Arabidopsis* data as a functional gateway, employing its well-annotated genome to facilitate comparative insights into less-characterized species. Future releases are planned to expand the platform with additional species suites, in response to evolving research priorities. PlantaeViz features an intuitive, user-friendly interface equipped with built-in tools to generate heatmaps, line plots, and co-expression network diagrams, designed to support data interpretation and hypothesis generation. By addressing the lack of dedicated and integrative resources for non-model plants, this platform empowers plant researchers to maximize the utility of omics datasets and accelerate functional gene discovery in key agricultural and medicinal crops.

Results

A modular framework for expanding species and tools in plant omics

PlantaeViz (https://plantaeviz.tomsbiolab.com/) consolidates a broad and diverse set of transcriptomic resources across several plant species of agronomic and medicinal relevance: grapevine, tomato, mulberry, and cannabis. Arabidopsis is included as a reference, to allow for easier transfer of functional knowledge from model to non-model species. The platform is organized into different suites that can be accessed from the

homepage; *Vitis spp.* (VitViz), *Solanum lycopersicum* (TomViz), *Morus spp.* (MulViz), *Cannabis spp.* (CannaViz), and *Arabidopsis thaliana* (AraViz). The suites follow a modular structure designed for continuous expansion, facilitating the adaptation of the platform for additional species, several of which are currently under development, including *Chlamydomonas reinhardtii* (ChlamyViz), *Prunus spp.* (PrunViz), *Fragaria spp.* (StrawViz), among others.

Each suite offers a homepage that features an informative index with menus linking to various applications (apps) and tools, including gene network visualization, gene expression atlases, differential gene expression analysis, gene set enrichment analysis, genome browsers, multiple-sequence alignment, gene function annotation, and an updated BLAST server (Figure 1). Interoperability between the different suites has been enabled through comprehensive orthology analyses between species, cultivars and/or genome versions. These multi-tool analyses include a summary output and orthologs have been assigned high, medium or lower confidence based on a custom scale.

The platform integrates a vast collection of Illumina-based transcriptomic data, over 57,000 SRA runs, which have been systematically categorized by metadata categories such as tissue type, developmental stage, treatment condition, cultivar, and organism. A custom set of hierarchical ontologies, coupled to regular expressions, has been designed to achieve this. These transcriptomic datasets are the basis for multiple functional modules within PlantaeViz, enabling both gene-level exploration and systems-level analyses.

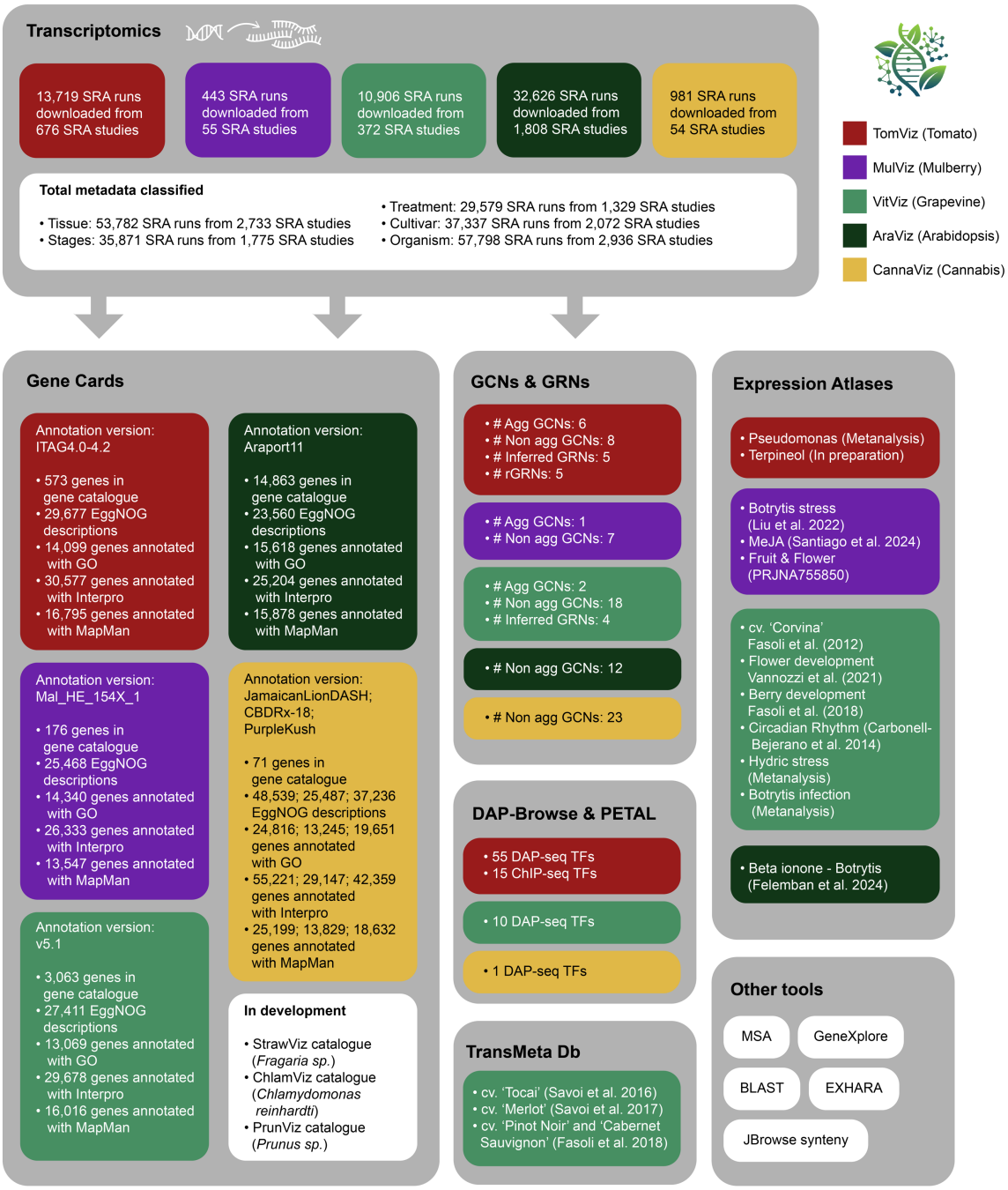


Figure 1. Overview of the content and functionalities in the different suites of the PlantaeViz platform. Publicly available, reannotated transcriptomic datasets are used by the *Gene Cards* module to generate tissue-specific expression profiles for each species, and by the Gene Co-expression Networks (GCNs) and Gene Regulatory Networks (GRNs) modules. PlantæViz also includes: *DAP-Browse* and *PETAL* (Plant Expression and Transcription factor Association Lookup) apps, for exploring genome-wide DNA-binding peaks; *TransMetaDb*, for integrated transcriptome-metabolome analyses; Expression Atlases providing curated, study-specific and meta-analysis overviews; and a collection of auxiliary tools for sequence alignment, functional annotation and comparative genomics.

157 Automated classification of transcriptomic data using custom species ontologies

158 Automatic classification of SRA runs was performed for five key metadata categories
159 (cultivar, organism, stage, tissue, and treatment type), across five analyzed species
160 (grapevine, tomato, mulberry, cannabis, and arabidopsis). This classification was
161 based on species-specific ontology frameworks developed for each organism
162 (Supplemental Data 1-10, in both JSON and pdf format). The resulting classifications
163 are presented in Supplemental Tables 1-5, displaying all terms automatically identified
164 by the pipeline (in the "*_curated_details*" columns) alongside the single, most specific
165 assigned term (in the "*_curated*" columns). In cases where terms from distinct ontology
166 branches (e.g., 'fruit' and 'root'), were simultaneously detected, a single-term
167 assignment was not feasible and such entries were labeled as "mix". Assignment
168 confidence for each term was also provided on a scale of 0-10.

169 Overall, assignment success varied considerably across categories (Figure 2A). The
170 "tissue" category showed one of the highest success rates, with nearly complete
171 annotation coverage in most species. This high performance can be attributed not only
172 to the common practice of submitting tissue-related metadata, even when not
173 standardized, but also to the robustness of our ontology-driven approach. The tissue
174 ontologies were carefully adapted for each species, accounting for anatomical and
175 terminological differences, which contributed to more accurate classification results.

176 Analysis of the specific tissue types assigned revealed distinct patterns reflecting the
177 predominant research focus for each species (Figure 2B). In the model organism *A.*
178 *thaliana*, the most frequently represented tissues were leaf, root, and seedling, with
179 notable contributions from vascular tissues and cell cultures, indicative of its broad
180 experimental use. In cannabis, flower and trichome tissues dominated, consistent with
181 a research focus on specialized metabolite biosynthesis. For grapevine, tomato, and
182 mulberry, the majority of samples were classified as leaf or fruit, aligning with studies
183 on fruit development and metabolic regulation. Notably, root tissues were markedly
184 underrepresented in grapevine, comprising only 3.4% of runs, compared to 14.3% in
185 tomato. This discrepancy may stem from the fact that *Vitis vinifera* is commonly studied
186 for its aerial organs, while rootstocks used in viticulture often belong to other *Vitis*
187 species, which are less frequently included in transcriptomic analyses.

188 Similar to the "tissue" category, the "organism" category achieved near-complete
189 assignment success, reaffirming that species-level information is generally well
190 represented in public repositories. In contrast, assignment rates for the "cultivar" and
191 "stage" categories were more variable (Figure 2A). For instance, mulberry cultivars
192 appear to be more sparsely annotated as the percentage of assigned runs falls below
193 40% whilst the other species surpass 50%. Similarly, developmental stage information
194 was assigned to over 50% of runs in tomato, cannabis and arabidopsis, but remained
195 lower for grapevine and mulberry, potentially reflecting greater heterogeneity or
196 incompleteness in the available metadata. The "treatment type" category, which is still
197 under refinement, reached approximately 50% assignment success in some species.
198 This is encouraging, particularly given that not all transcriptomic studies involve
199 treatments, and associated metadata is often inconsistently reported.

200 These results underscore the inherent challenges in curating and standardizing
201 unstructured metadata, particularly for non-model species where such information is
202 often poorly annotated. While our ontology-based approach (Figure 2C) significantly
203 improved the reannotation of SRA runs, its performance was ultimately contingent on
204 the quality and completeness of the metadata available for each species. Figure 2C
205 illustrates the hierarchical structure of tissue terms in *Cannabis sativa*, reflecting
206 species-specific anatomical and developmental distinctions, such as separate
207 categories for male and female flowers, and a refined classification of globular
208 trichomes based on their reported developmental stages and morphological forms
209 (Livingston et al., 2020). As a further example, cannabis fruits are technically achenes;
210 however, since the use of "fruit" was more frequently used in the original metadata
211 and both terms are correct, we adopted “fruit” as the primary ontology term, with
212 regular expressions configured to recognize both "achene" and "fruit" as valid
213 synonyms.

214 In total, we downloaded and processed 32,626 transcriptomic samples for *A. thaliana*,
215 13,719 for *S. lycopersicum*, 10,631 for *V. vinifera*, 981 for *C. sativa*, and 442 for *M.*
216 *alba* (Supplemental Tables 1-5). Expression data from samples with successfully
217 assigned tissue annotations were integrated into the *Gene Cards* and *GeneXplore*
218 applications, enabling intuitive visualization and exploration. A curated subset of these
219 annotated datasets was further used in downstream analyses, including co-expression
220 network generation or regulatory network inference.

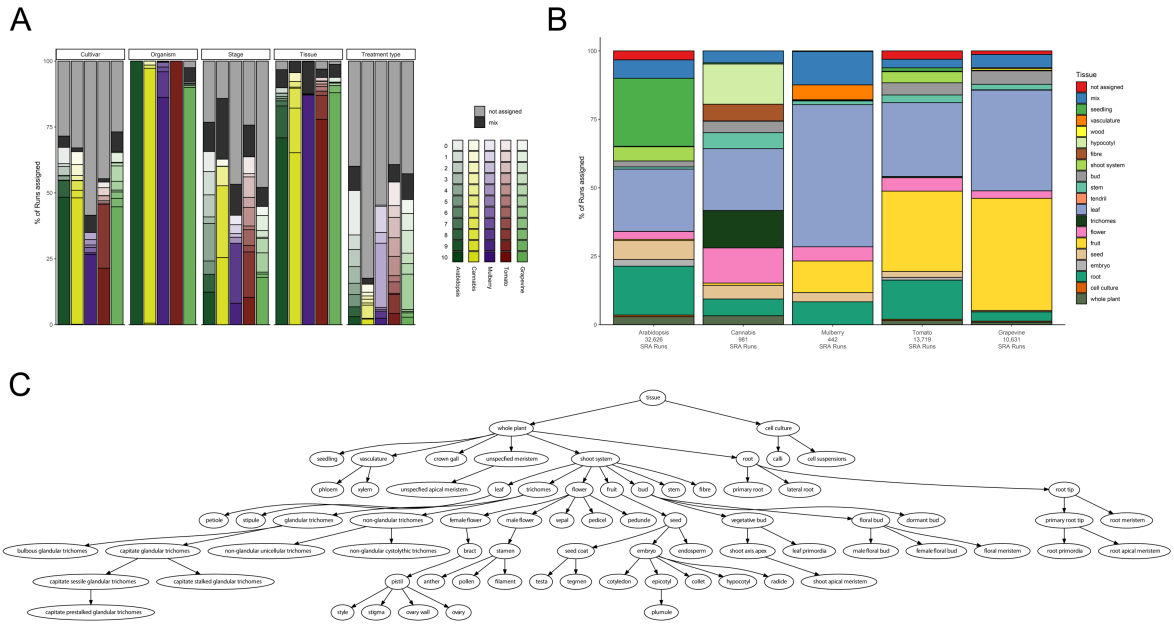


Figure 2. Ontology-driven reannotation of SRA runs across metadata categories and plant species. (A) Percentage of SRA runs successfully assigned to metadata categories across different species, alongside their corresponding confidence level on a 0-10 scale. Lighter tones indicate assignments made with lower confidence according to a custom scale based on a combination of metadata column and regular expression reliability. The *tissue* and *organism* categories show the highest assignment success across all species. (B) The distribution of tissue-type assignments varied substantially between species, reflecting differences in tissue representation that align with species-specific research priorities. (C) Example of the tissue ontology for *Cannabis sativa*, illustrating hierarchical relationships between terms. This structure enables flexible grouping of SRA runs into broader categories, even when the sampled tissue corresponds to a specific subcomponent. Full graphical representations of all ontologies can be found in Supplemental Data 6-10.

A unified resource for gene-centric insights

The *Gene Cards* application provides gene-centric information across all species, incorporating automatic functional annotation for queried genes, structural models integrated via *JBrowse2*, ortholog inference, TPM-based expression profiles across all metadata-annotated RNA-seq tissue samples from SRA, and access to gene, transcript, CDS, and protein sequences (Figure 3).

Each *Gene Cards* app functions as the main access point to a curated gene catalogue, allowing users to retrieve detailed and curated information for individual genes. Grapevine is currently the best-curated species within the platform, and similar curation efforts are underway for other plant species to broaden the platform's utility.

In the case of this species, it is tightly linked to the grapevine gene reference catalogue (Navarro-Payá et al., 2022) (available at https://grapedia.org/genes/). This curated resource, currently in its third version, includes functional data for over 3,000 genes and is open to community submissions through the Grapedia portal. The gene catalogue is built upon the foundation of the telomere-to-telomere (T2T) v5 cv. "PN40024" genome assembly, where v5.1 gene IDs serve as the primary identifiers for genes. Provided information includes gene symbols and synonyms, descriptive gene names, pathway and gene family classifications, and custom-defined validation levels for functional data. To facilitate the research community's transition to updated genome annotations, the platform includes a cross-annotation tool that enables users to input one or more gene IDs and retrieve the corresponding identifiers across all supported annotation versions, ensuring backward compatibility and smooth interoperability.

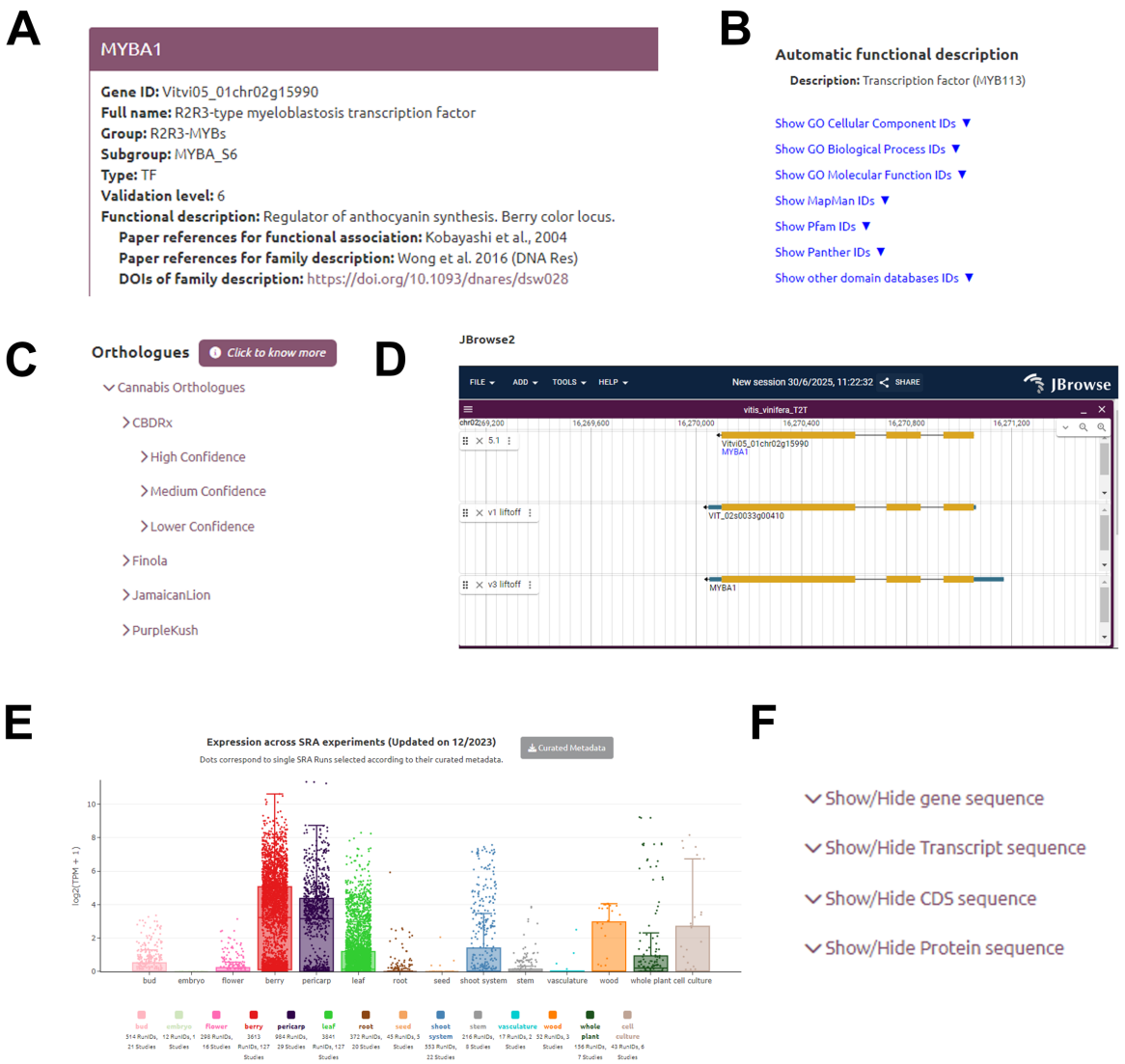


Figure 3. The grapevine gene *VviMYBA1* as an example in VitViz *Gene Cards* app. (A) Manually curated grapevine gene catalogue information. (B) Automated functional annotation with structured, category-specific dropdown menus. (C) Predicted orthologs across other species and cultivars included in PlantaeViz split by confidence level. (D) Embedded *JBrowse2* instance displaying the genomic context of the selected gene. (E) Tissue-specific expression profile of the gene across more than 10,000 SRA runs, shown in $\log_2$(TPM+1). The interactive boxplots allows users to identify the corresponding SRA studies by hovering over individual data points. (F) FASTA-format sequences for the gene, transcript, coding sequence (CDS), or protein are also available.

265 Interactive visualization of transcriptomic data

266 The platform incorporates expression atlas applications, which are built upon manually
267 curated datasets specifically designed to address distinct research needs. These
268 applications consolidate transcriptomic data from various studies, encompassing both
269 microarray and RNA-seq datasets and allowing to investigate gene expression
270 patterns across a wide array of biological contexts, including specific tissues,
271 developmental stages, and experimental treatments. Examples include atlases for
272 specific flower developmental stages (Vannozzi et al., 2021), hormone elicitation
273 (Santiago et al., 2024) or the circadian rhythm (Carbonell-Bejerano et al., 2014), as
274 well as different meta-analysis atlases combining multiple experiments regarding
275 treatments, such as hydric stress, and *Botrytis* infection (Felemban et al., 2024; Liu et
276 al., 2022). Among their key features, users can generate interactive heatmaps based
277 on custom gene input lists, apply data transformations such as logarithmic or Z-score
278 scaling, and perform clustering to group data by genes or conditions, thereby
279 highlighting co-expression patterns. Expression atlases support the visualization of
280 normalized transcript levels (e.g. TPM, transcripts per kilobase of exon model per
281 million mapped reads) (Figure 4A) as well as differential expression data (Figure 4B).
282 In differential expression views, heatmap cell color represents fold-change values,
283 while dot size conveys adjusted p-values, facilitating intuitive interpretation of
284 significance and magnitude. These features collectively enable robust comparative
285 analyses, allowing researchers to detect expression trends and infer potential
286 functional associations under varying biological conditions.

287 For further data exploration, we developed two complementary web applications:
288 *EXHARA* (EXpression Heatmaps Across all SRA Runs Available) and *GeneXplore*.
289 *EXHARA* is designed for the broad visualization of gene expression from a user-
290 provided list of genes across all curated runs (Supplemental Figure 1A). Its primary
291 output is an interactive heatmap that can be dynamically filtered by tissue or by specific
292 user-selected runs, providing a versatile resource for customized analysis.
293 Alternatively, *GeneXplore* was developed to integrate automatic metadata
294 classification results with gene expression data in a user-friendly web application that
295 allows visualization and comparison of gene expression levels across a custom range
296 of experimental categories (Supplemental Figure 1B). The tool integrates the ontology-
297 based annotations described above and displays normalized TPM expression values

enabling cross-sample comparability. *GeneXplore* allows users to examine expression differences for specific genes across tissues, developmental stages, and treatment conditions, individually or in a combinatorial manner, based on the curated and available metadata. Together, these tools provide a powerful resource for uncovering patterns and gaining insights into gene behavior across diverse biological contexts.

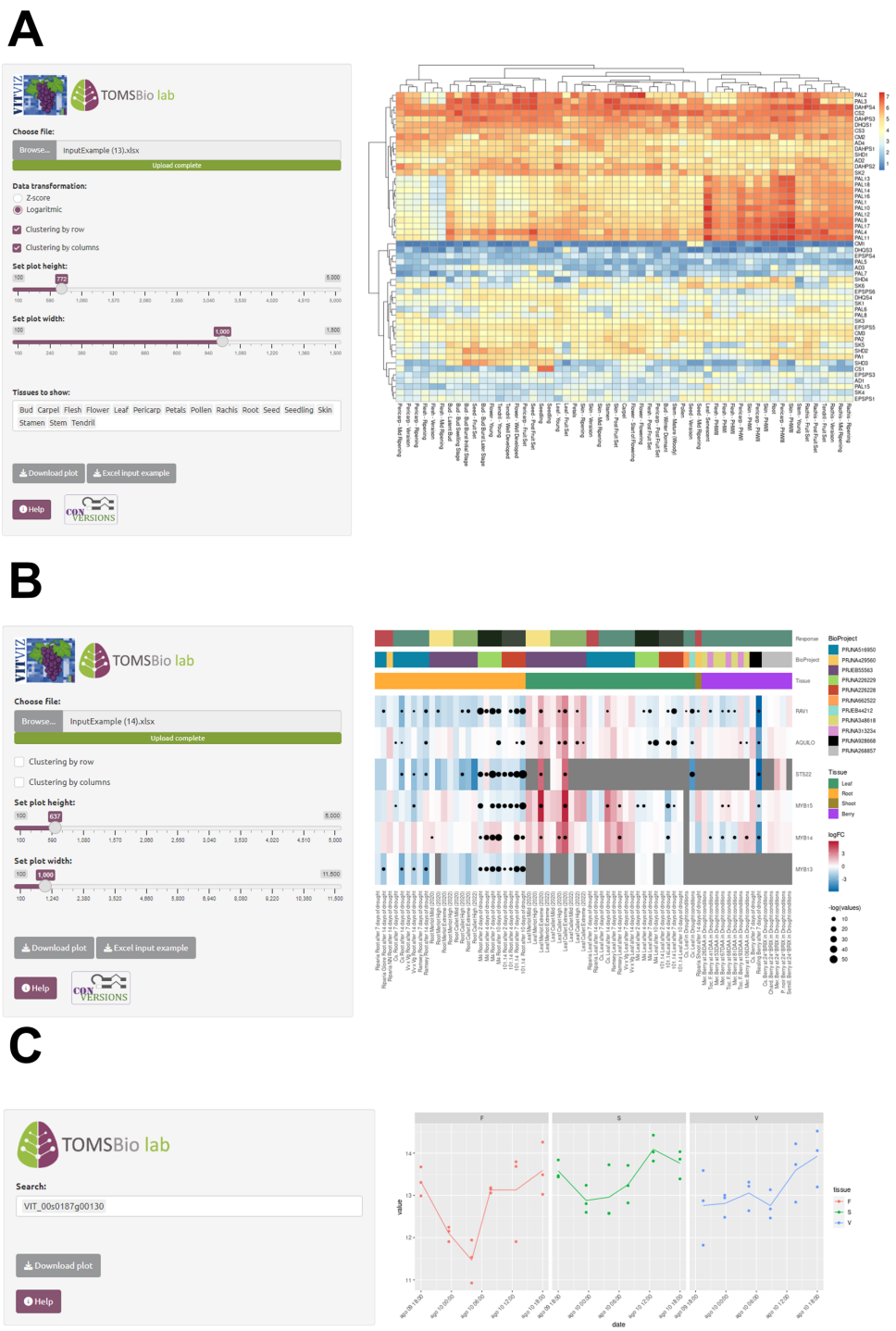


Figure 4. Examples of gene expression atlases integrated into PlantaeViz. (A) A detailed view of the *Vitis vinifera* cv. 'Corvina' expression atlas based on normalized microarray data, showcasing gene expression patterns across 54 developmental stages and tissue types (Fasoli et al., 2012). (B) Differential gene expression results compiled from various independent transcriptomic studies focused on hydric stress conditions. This meta-analysis allows users to explore gene-specific responses to drought and water deficit treatments across genotypes, organs, and experimental setups. (C) Normalized microarray gene expression at different timepoints for grapevine cv. 'Tempranillo' flesh (F) and skin (S).

313 Exploring gene-to-gene relationships through network-based analyses

314 PlantaeViz hosts a comprehensive collection of over 90 whole-genome networks,
315 including aggregated and non-aggregated gene co-expression networks (GCNs), as
316 well as gene regulatory networks (GRNs), inferred either through co-expression or
317 supported by experimental evidence. Among these, GCNs are particularly amenable
318 to AUROC-based performance evaluation, as the 'guilt-by-association' principle
319 underpins their structure and predictive utility (Supplemental Table 6). In this study,
320 we applied the GCN construction methods previously developed by Orduña et al.
321 (2023), ensuring methodological consistency with prior benchmarking efforts. Analysis
322 reveals that network performance improves with increasing number of studies and/or
323 sequencing runs, regardless of species or GCN construction method (Supplemental
324 Figure 2A). This trend supports the benefit of dataset breadth and diversity in co-
325 expression network reliability. Notably, aggregated GCNs almost always outperform
326 non-aggregated networks, showing marked improvements in predictive power across
327 species (Supplemental Figure 2B).

328 We have developed interactive visualization tools for exploring gene co-expression
329 networks (GCNs) (Figure 5A) and gene regulatory networks (GRNs) (Figure 5B), each
330 combining multiple functionalities to investigate gene relationships and regulatory
331 interactions. The applications are organized into two main tabs, each tailored to
332 facilitate different aspects of network analysis while presenting advanced visualization
333 features for in-depth exploration. By combining co-expression data with regulatory
334 predictions, including transcription factor (TF) binding events at open chromatin
335 regions, the platform enables construction and interrogation of regulatory networks.
336 This dual approach enables users to retrieve TF-target gene pairs, visualize specific
337 subnetworks based on custom gene sets, and explore co-expression and regulatory
338 interactions concurrently. Notably, the underlying network data are not static; they are
339 periodically updated to incorporate new experimental datasets.

340 The application interface includes a co-expression table with chromosomal mapping,
341 as well as a 2D force-directed network visualization module. In the first tab of GCN
342 apps, users can query a gene of interest to retrieve a list of co-expressed genes that
343 is by default filtered to the top 1% of interactions, this is otherwise known as a gene-
344 centered network. This filter can be optionally skipped to display all co-expression

relationships present in the whole network, to better reflect the scale-free topology of biological networks. While most nodes are expected to have few connections, a few nodes will be highly interconnected, i.e. not all nodes will have the same number of edges within the whole network. This table provides detailed information such as gene identifiers and names (sourced from curated gene catalogues), along with co-expression metrics. Users can perform on-the-fly gene set enrichment analysis (GSEA) on the retrieved list to identify enriched biological pathways or functional categories. Additionally, the chromosomal locations of the co-expressed genes are displayed in an integrated genomic map, enabling the exploration of their genomic distribution. This feature facilitates the identification of potential chromosomal clustering or co-localization of functionally related genes, bridging expression-level analysis with genomic context. GRN apps offer similar functionality, however, gene search can be focused on source, target genes, or both. This reflects the fact that source genes in a GRN are always TFs and target genes may be TFs or otherwise, thereby allowing to query potential regulators as well as regulatory targets.

The second tab offers an interactive 2D force-directed graph visualization of gene co-expression or regulatory networks based on the network D3 R package. A key feature is the ability to input custom gene lists, enabling users to explore specific gene sets and their interconnections based on co-expression patterns or regulatory interactions. The network is rendered as a dynamic, two-dimensional graph in which nodes represent genes and edges denote co-expression or regulatory relationships. This visualization is especially valuable for exploring pathway- or family-oriented gene networks, helping to uncover potential co-regulation and connectivity among genes of interest. Users can identify hub genes (i.e., potential key regulators), and detect tightly connected clusters within their input set. The topology of the D3 subnetwork is influenced by node connectivity: node positions are a compromise of push/pull forces, they repel each other by default and attract each other based on number edges and their weights. The layout is also interactive, allowing the user to identify gene names and edge weights by simply hovering over. Moreover, some visual rearrangement of nodes is allowed, after which forces are recalculated, always maintaining a force-directed topology. Users can adjust repulsion parameters to optimize layout visualization according to the characteristics of their subnetwork. To ensure analytical reproducibility in light of these periodic updates, each downloaded metadata includes

a timestamp, allowing users to reference the specific network version used for their visualization and analysis. In compliance with FAIR principles, the full network can be downloaded at any time as well as a comprehensive set of metadata that describe each of the transcriptomic studies used to build it and any relevant network building parameters. NCBI data download timestamps and transcriptomic query dates are provided for traceability, accounting for the fact that further aggregation of SRA runs to these networks will occur in future updates, thereby changing the networks topology.

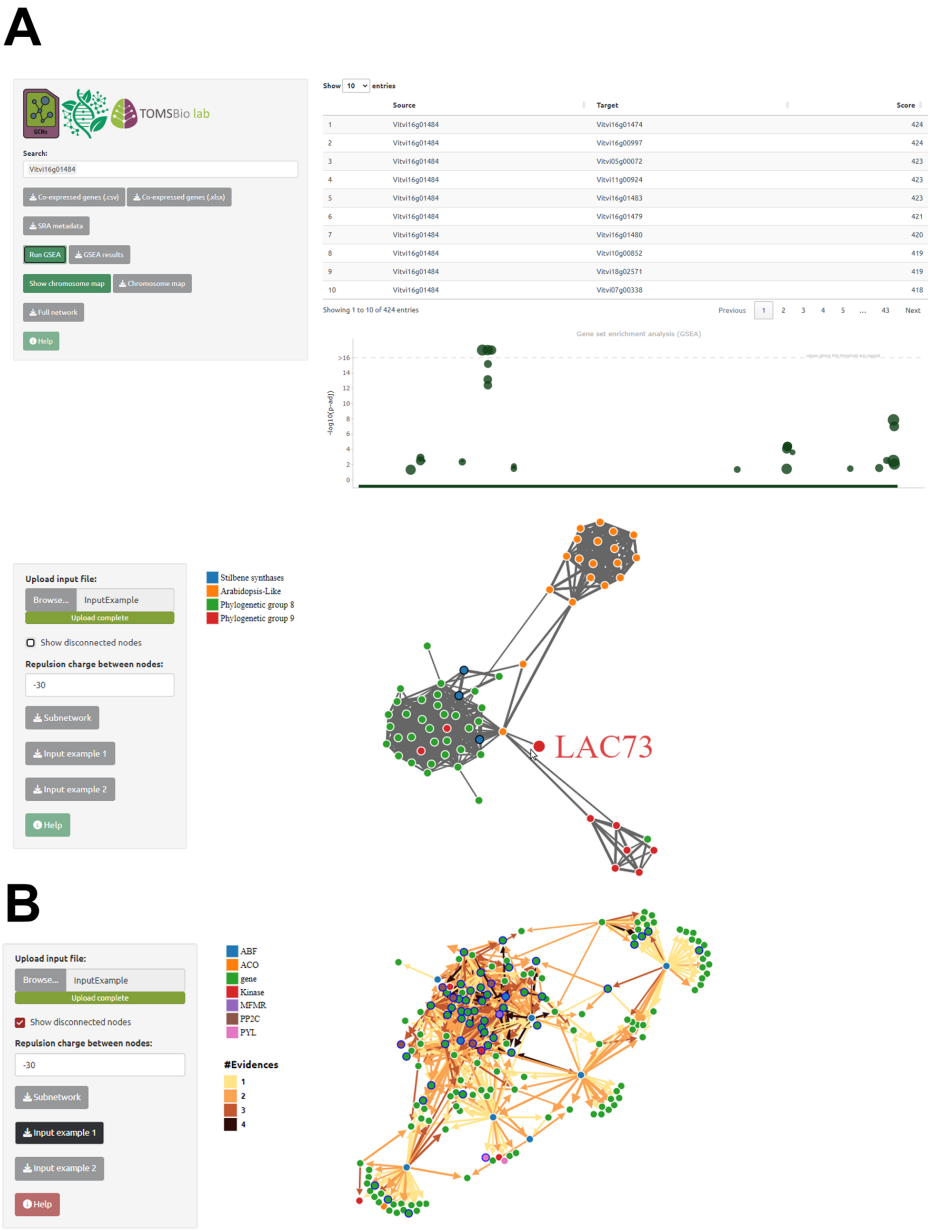


Figure 5. Examples of gene co-expression network (GCN) and gene regulatory network (GRN) applications in PlantaeViz. (A) The GCN app provides a list of genes co-expressed with a selected gene of interest based on large-scale transcriptomic data. Users can input a gene identifier to retrieve co-expression partners and visualize their relationships using an interactive network plot. This plot highlights the strength and direction of co-expression correlations, allowing users to identify potential functional associations or gene modules. Additionally, users can submit a custom list of gene IDs to generate interaction plots, facilitating the visualization of known or hypothesized networks. The figure displays the AggGCN App interface within the VitViz module of PlantaeViz. (B) The GRN app enables the visualization of predicted regulatory interactions between transcription factors (TFs) and their target genes. The example shows an interaction plot where edges represent putative regulatory relationships inferred from expression data and/or experimental binding evidence. Nodes are color-coded by gene

398 function or regulatory role, providing an intuitive overview of gene network topology and regulatory hubs.
399 The figure displays the GRN App interface within the TomViz module of PlantæViz.

400 Exploratory applications and genome browsers for ChIP-seq and DAP-seq data

401 Understanding transcriptional regulation is essential for dissecting traits such as
402 development, stress response, and fruit quality. Chromatin immunoprecipitation
403 (ChIP)-seq and DNA-affinity purification (DAP)-seq resources facilitate the exploration
404 of transcription factor (TF)-DNA interactions across plant species and represent a very
405 good, yet partially complete, view of the the regulatory landscape. This data is
406 presented in the form of JBrowse adaptations, here named *ChIP-/DAP-Browse*s,
407 which integrate community-contributed ChIP- and DAP-seq datasets, acting as an
408 interactive platform for visualizing cistromes and DNA-binding profiles (Figure 6A). A
409 summary of currently available transcription factors is provided in Supplemental Table
410 7, including both our own and publicly available data. The collection is expected to
411 grow as further contributions are added by the research community.

412 Genome annotation tracks follow a standardized color scheme: coding exons are
413 represented as yellow boxes, untranslated regions (UTRs) as blue boxes, and introns
414 as connecting lines. For each transcription factor, *ChIP-/DAP-Browse* provides three
415 types of visualization tracks: 1) Individual replicate tracks, showing raw alignment
416 distributions for both TF and input libraries, 2) Merged replicate tracks, allowing
417 comparison of aggregated TF versus input signal, and 3) GEM peak-caller tracks,
418 highlighting predicted binding peaks specific to each TF. Unlike static tabular peak
419 data, which may be prone to false negatives or positives due to algorithmic limitations,
420 the genome browser enables manual inspection of read alignments, allowing users to
421 visually assess peak quality and binding signal consistency. GEM peak-caller tracks
422 also associate detected peaks with their most likely target genes, using a nearest-
423 genomic-feature assignment method. In addition to TF binding data, the genome
424 browser includes dot plot visualizations for inter-genome comparisons (Figure 6B),
425 and interactive synteny maps displaying conserved gene relationships across species
426 (Figure 6C).

427 In order to complement *JBrowse2* capabilities and to delve deeper into cistrome data,
428 we also introduce *PETAL* (Plant Expression and Transcription factor Association

429 Lookup), an interactive app that allows users to explore DAP-seq data from both gene-
430 centric and protein-centric perspectives.

431 The gene-centered tool (Supplemental Figure 3) allows to examine DAP-seq peaks
432 from multiple TFs associated with multiple genes in a single view. Peaks are
433 represented within a user-defined distance from the transcription start site (TSS) of
434 their corresponding genes. Users can also optionally filter peaks based on fold-change
435 thresholds. The shape of each peak is represented by plotting the centers of reads
436 from the bigWig files (x-axis) against their scores (y-axis). To standardize the display,
437 maxima and minima are set equal, and other points are scaled accordingly. Due to
438 this, absolute peak heights are not comparable between different peaks, but fold-
439 change, derived from the annotation, is conveyed through peak color intensity.

440 On the other hand, the protein-centered tool (Figure 6D) integrates DAP-seq data with
441 the previously presented GCNs, to draw gene regulatory networks and identify
442 putative targets for TFs of interest. This combined dataset, derived from both
443 ChIP/DAP-seq binding profiles and co-expression patterns, allows users to prioritize
444 tissue-specific regulatory links by filtering targets based on TF binding peak quality,
445 proximity to transcription start site (TSS), and tissue expression. The interface
446 provides a drop-down menu of TFs, accompanied by sliders and numeric filters to
447 adjust q-value thresholds and maximum distance to TSS, enabling users to focus on
448 high-quality peaks and gene targets relevant to the TF's specific role. Results are
449 visualized as a dynamic bar plot summarizing the number of co-expressed genes per
450 tissue type. The bar plot updates interactively as the user adjusts filters, providing
451 insight into the reliability and tissue specificity of the associations. Detailed results are
452 displayed in a dynamic, filter-responsive table listing gene targets with their IDs, gene
453 symbols (where available), ChIP/DAP-seq peak information (including q-value and
454 distance to TSS), and co-expressed tissues. Gene IDs in the table are hyperlinked to
455 detailed gene information within the gene catalogue and the *Gene Cards* app. For
456 gene targets that are also co-expressed with the TF, the table lists the tissues in which
457 co-expression is observed, providing a comprehensive view of regulatory interactions.
458 All of the above tools offer a robust and user-friendly framework for investigating
459 transcription factor binding dynamics and their regulatory roles across diverse plant
460 genomes.

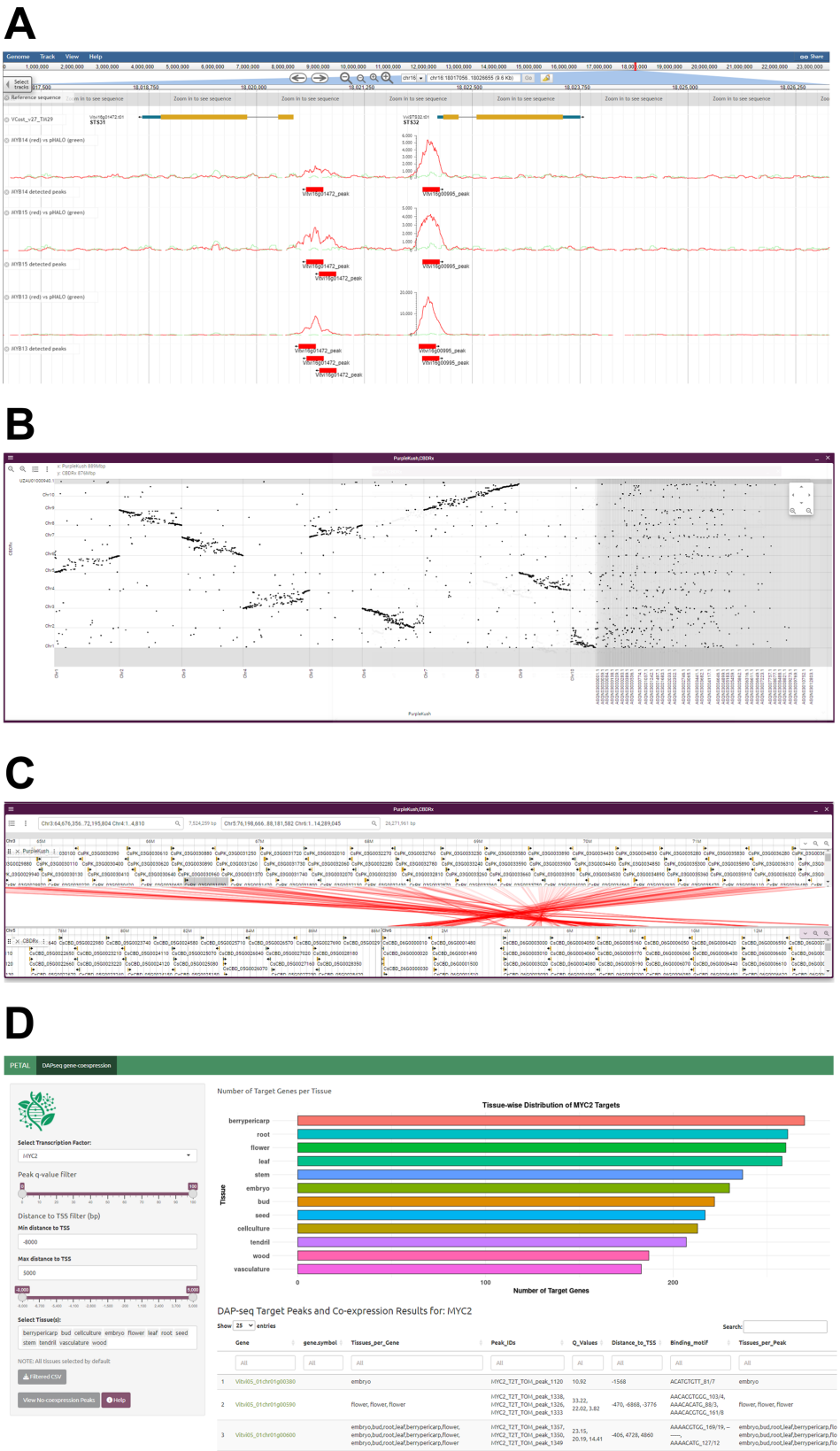


Figure 6. Examples of *ChIP/DAP Browse* and gene- and protein-centered module of *PETAL* app. (A) Genome-wide DNA-binding data displayed in the *DAP-Browse* module, showing DAP-seq peaks associated with stilbene synthase genes targeted by the transcription factors *VviMYB13, VviMYB14*

465 *and VviMYB15* in *Vitis vinifera*. The viewer allows inspection of binding site locations relative to gene
466 features, enabling hypothesis generation on transcriptional regulation. (B) Dotplot comparison between
467 the *Cannabis sativa* cultivars cv. 'CBDRx' (y-axis) and cv. 'Purple Kush' (x-axis), illustrating sequence
468 similarity and large-scale structural variation across the two genomes. (C) Detailed synteny view
469 between a region in chromosome 06 of CBDRx and its homologous region in chromosome 03 of Purple
470 Kush. Gene models within these regions are linked based on conserved collinearity, providing insight
471 into genome evolution and gene conservation across cultivars. (D) The protein-centric tool within
472 *PETAL* integrates DAP-seq binding data with tissue-dependent and independent GCN information,
473 MYC2 is shown as an example query. The interface displays a dynamic bar plot summarizing the tissue-
474 wide distribution of co-expressed gene targets for a selected TF, based on user-defined filters such as
475 peak quality (q-value) and distance to TSS. A responsive table provides detailed target information,
476 including gene identifiers (hyperlinked to catalogue information within the *Gene Cards* apps),
477 ChIP/DAP-seq peak metrics, and a list of tissues where TF–target co-expression is detected. Together,
478 these components enable tissue-specific prioritization of regulatory interactions.

479 Integration of gene-metabolite data across experiments

480 *TransMetaDb* is an interactive tool designed for the integrated visualization and
481 analysis of transcriptomic and metabolomic datasets (Supplemental Figure 4). It
482 enables users to explore correlations between gene expression levels and metabolite
483 abundances across diverse experimental conditions. By facilitating the identification
484 of functional associations, *TransMetaDb* helps uncover potential gene-to-metabolite
485 relationships that may underlie specific physiological or biochemical traits. The primary
486 output is a downloadable heatmap or table, where rows represent user-input genes
487 and columns correspond to different metabolites. Each cell reports the Pearson's
488 correlation coefficient (PCC) along with the associated p-value (shown in brackets),
489 calculated using the Student's asymptotic method. This matrix allows rapid visual and
490 statistical assessment of correlation strength and significance. Additionally, the tool
491 features a module for exploring gene co-expression clusters derived from weighted
492 gene co-expression network analysis (WGCNA), performed locally. Users can
493 download a complete table assigning all genes to their respective modules, enabling
494 further exploration of gene-metabolite associations within co-regulated gene sets. This
495 dual functionality provides a powerful framework for systems-level analysis of
496 metabolic and transcriptional correlation.

497 Other more common tools: multiple sequence alignment

498 To facilitate the exploration of sequence conservation and functional relationships
499 among homologous genes, we included a multiple sequence alignment (MSA)
500 application, enabling users to align nucleotide (mRNA or CDS) or protein sequences,
501 helping identify conserved domains, motifs, and sequence variations that provide
502 insights into gene function and evolutionary dynamics (Supplemental Figure 5). The
503 application supports a list of gene IDs for the species of interest and offers multiple
504 alignment algorithms, including ClustalW, Clustal Omega, and MUSCLE. These
505 supported methods allow customization of the alignment process to meet specific
506 analytical needs. Alignment results are displayed in an interactive, color-coded viewer.
507 Conserved residues or nucleotides are highlighted, and a corresponding sequence
508 logo graphic summarizes conservation patterns across the alignment. Users can zoom
509 in on specific regions for detailed inspection and export results in FASTA format for
510 downstream analyses. By integrating flexible alignment options with intuitive
511 visualization, the MSA application offers a streamlined solution for studying sequence
512 conservation, functional domains, and genetic variation across related gene sets.

513 Case study: Interspecies functional association of MYB24 in terpene biosynthesis

514 This case study demonstrates the utility of the GCNs applications, cistrome browser,
515 and ortholog inference tools available in the platform to perform integrative cross-
516 species regulatory analyses. Specifically, we investigated the functional conservation
517 of the MYB24 transcription factor in regulating terpene biosynthesis in *Vitis vinifera*
518 and *Cannabis sativa*.

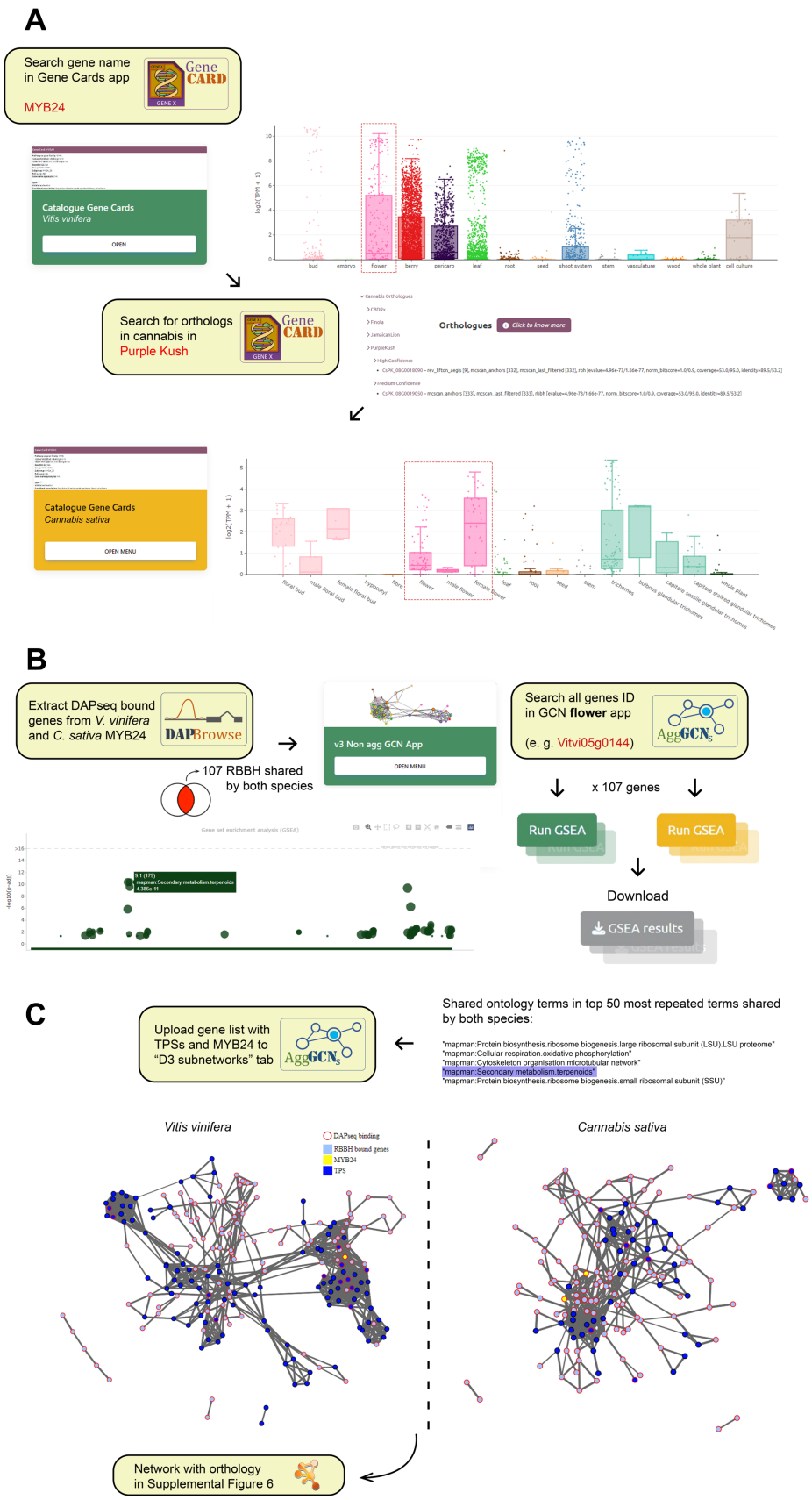
519 We first used DAP-seq data to identify direct TF-DNA interactions. In *V. vinifera*, the
520 DAP-seq assay for *VviMYB24* (Zhang et al., 2023) revealed 3,557 high-confidence
521 binding events located within 5 kb upstream of transcription start sites (TSSs), while
522 *CsMYB24G* showed 1,926 such binding events (Zhang et al., unpublished results).
523 Consistent with their potential regulatory roles in reproductive tissues, both genes
524 showed high expression levels in flower samples, as displayed in the *Gene Cards*
525 application (Figure 7A).

526 To compare targets across species, we used the platform's built-in reciprocal best
527 BLAST hit (RBBH) ortholog mapping between *C. sativa* and *V. vinifera*. We identified

528 107 genes bound by *MYB24*/*MYB24G* with 1-to-1 orthology across the two species.
529 These shared DAP-seq targets were further analyzed for functional context by
530 performing gene set enrichment analysis (GSEA) on their top co-expressed partners,
531 using the "Co-expressed Genes" tab of the GCN app (Figure 7B). To improve
532 specificity and exclude overly generic terms, GO categories with more than 200
533 annotated genes were filtered out.

534 Remarkably, in both species, the fourth most frequently enriched term among the co-
535 expressed gene sets, after more general categories like LSU proteasome, oxidative
536 phosphorylation, and microtubule network, was associated with terpene biosynthesis.
537 This was the top-ranked term specific to specialized metabolism, indicating a
538 potentially conserved regulatory function of *MYB24* in controlling terpene production.

539 To further investigate the network connectivity among *MYB24* targets and terpene
540 biosynthesis genes, we curated a list of terpene synthase (*TPS*) genes from both
541 genomes. These TPSs were combined with the DAP-seq derived targets and
542 submitted to the “D3 Subnetworks” visualization tool in the platform (Figure 7C). This
543 allowed us to reconstruct co-expression subnetworks and inspect the architectural
544 organization of *MYB24*-associated modules highlighting direct and indirect
545 associations between MYB24 targets and *TPS* genes. The graphical outputs revealed
546 both conserved and species-specific co-expression structures. The resulting networks
547 for *V. vinifera* and *C. sativa* were also exported as CSV files and visualized using
548 Cytoscape (Supplemental Figure 6). Together, these findings support the hypothesis
549 of a conserved MYB24-centered regulatory module involved in floral terpene
550 biosynthesis, while also reflecting evolutionary divergence in network topology and
551 regulatory wiring between these two species.



552

553 **Figure 7. Cross-species validation of MYB24 regulatory function using gene coexpression**
554 **networks (GCNs) and DAP-seq data using PlantaeViz.** (A) Tissue-specific expression profiles of

MYB24 in *Vitis vinifera* and its ortholog *MYB24G* in *Cannabis sativa* based on publicly available SRA transcriptomes visualized through the *Gene Cards* app, highlight conserved high expression in floral tissues. (B) Gene set enrichment analysis (GSEA) of a DAP-seq-bound gene, conducted using the "Co-expressed Genes" tab in the GCN app. Each dot represents an enriched Mapman term linked to orthologous co-expressed genes (via Reciprocal Best BLAST Hits, RBBH) in both species, illustrating conserved regulatory pathways. (C) Cross-species co-expression subnetworks were generated via the "D3 Subnetworks" tab, using manually curated terpene synthase (*TPS*) genes and DAP-seq targets as inputs. Networks show MYB24 targets and their co-expression relationships with *TPS* genes in *V. vinifera* (left) and *C. sativa* (right). Nodes with red borders indicate direct DAP-seq binding events, supporting a degree of functional conservation across species.

Discussion

PlantaeViz distinguishes itself from existing omics databases through its unique integration of diverse functionalities, its emphasis on non-model plants, scalability, and a user-centered design. The platform unifies transcriptomic, genomic, and metabolic datasets into a single coherent framework, enabling multi-omics exploration across a growing number of plant species. It has been engineered to handle large-scale datasets efficiently, with improved webpage responsiveness and robust data-handling capabilities. Extensive manual curation and intuitive analytical interfaces allow users to perform species-specific or comparative analyses. Applications such as *Gene Cards* and Expression Atlases present a detailed view of gene expression dynamics, while tools like *TransMetaDb* and interactive gene network visualizations provide insights into functional associations between genes, metabolites, and regulatory elements.

Compared to existing platforms, PlantaeViz offers a more integrative, up-to-date, and FAIR-compliant approach to plant genomics and transcriptomics. Many widely used resources remain static, siloed, or limited in scope and accessibility. For example, TAIR, while historically invaluable for Arabidopsis research, has transitioned to a restricted-access model, limiting open data availability; something fully addressed in the open-access AraViz suite of PlantaeViz (Huala et al., 2001; Reiser et al., 2024). Similarly, Sol Genomics and TomExpress provide rich resources for Solanaceae (Fernandez-Pozo et al., 2015; Zouine et al., 2017), but TomViz offers a curated re-annotated gene models and improved catalogues that are regularly updated for accuracy and completeness. VitExpress, in contrast, is built on a non-reference

588 grapevine genome and lacks curated gene catalogues or gene network tools (Djari et
589 al., 2024), whereas VitViz utilizes the latest T2T reference genome and integrates co-
590 expression and regulatory networks using ChiP- and DAP-seq data. Even large-scale
591 platforms such as Phytozome, despite their broad range of genomes available, do not
592 support transcriptomic analyses or functional network visualizations (Goodstein et al.,
593 2012). PlantaeViz surpasses these limitations by offering a dynamic, integrative
594 environment that unifies up-to-date genomic and transcriptomic data with advanced
595 analytical tools, making it not just a complement but a necessary next-generation
596 platform for plant systems biology.

597 Another key distinguishing feature of PlantaeViz is its extensive transcriptomic data
598 coverage across multiple plant species. In its initial release, the platform integrates a
599 total of 58,351 accessions from the NCBI SRA database, which has been instrumental
600 in developing advanced tools such as expression profiling and regulatory network
601 visualizations. This achievement was made possible through the use of high-
602 performance computational resources, ensuring that the platform remains efficient and
603 scalable as dataset complexity grows. The processed transcriptomic data has been
604 used to build tissue-specific co-expression networks for all included species, and high-
605 performing aggregated gene co-expression networks (AggGCNs) were generated
606 wherever data volume permitted. While network aggregation has been demonstrated
607 to improve network performance, it requires substantial datasets and computational
608 power (Wong, 2020).

609 A further critical advantage of PlantaeViz lies in its modular and species-agnostic
610 design. This flexible architecture ensures broad compatibility with a wide range of plant
611 species and facilitates the integration of additional organisms and emerging
612 computational methods over time. By supporting both well-established and non-model
613 plants, interoperability across knowledge bases is enhanced, improving data
614 accessibility and analytical capabilities for underrepresented species while
615 maintaining essential links to well-characterized model organisms. The platform's
616 capacity to integrate multi-omic datasets, including transcriptomics and metabolomics,
617 further highlights its value for functional genomics and hypothesis generation. Its suite
618 of analytical tools enables the exploration of functional relationships, identification of
619 conserved gene regulatory motifs, and analysis of gene-metabolite correlations,
620 providing researchers with a systems-level understanding of plant biology. This is

621 particularly advantageous for advancing studies in plant development, specialized
622 metabolism, and environmental responses, where interest in non-model species is
623 rapidly growing.

624 Looking ahead, the ongoing incorporation of updated datasets, additional species, and
625 new analytical methods will continue to expand PlantaeViz's utility. By embracing the
626 latest multi-omics technologies and broadening its species coverage, the platform is
627 poised to remain a central resource for systems biology research, facilitating novel
628 discoveries and accelerating translational applications in agriculture and
629 biotechnology.

630 **Materials and methods**

631 Web server implementation

632 PlantaeViz was developed as a Shiny server (v1.5.17.973) running R version 3.6.3.
633 The index page is based on React as the JavaScript library. The platform is freely
634 available for use on the web and can be found at https://plantaeviz.tomsbiolab.com/.

635 Download of public transcriptomic data and metadata classification

636 Raw RNA-Seq data were retrieved from the Sequence Read Archive (SRA), using the
637 SRA Toolkit with a specific focus on Illumina-based sequencing runs. Only
638 transcriptomic datasets were included, excluding those related to microRNAs, small
639 RNAs, or non-coding RNAs. Species-specific queries were designed to identify
640 relevant RNA-Seq experiments, with the full list provided in Supplemental Table 8. In
641 cases where species-level annotations were limited, genus-level queries were used to
642 capture data from closely related accessions. For *Arabidopsis thaliana,* an additional
643 filter was applied to include only paired-end sequencing studies given the high volume
644 of available data.

645 Associated metadata was automatically classified using a custom species-specific
646 ontology. This ontology, stored as a structured JSON file, includes hierarchical terms
647 across five main categories: tissue, organism (species), developmental stage, cultivar
648 and treatment type. Each term is linked to a curated set of search terms and regular
649 expressions used to detect their presence within relevant metadata fields
650 (Supplemental Table 9). Only columns relevant to a given category and species were

651 considered during classification. To assess classification confidence, a qualitative
652 scoring scale ranging from 0 to 10 was implemented (Supplemental Table 10). Scores
653 were assigned based on two criteria: (1) the reliability of the metadata field (e.g.,
654 dedicated column vs. abstract) and (2) the specificity of the regular expression used
655 (categorized as 'reliable' or not). A score of 9 represents high-confidence assignments,
656 where a reliable regular expression matches a dedicated reliable metadata field. In
657 contrast, a score of 0 indicates a low-confidence match, typically from general text
658 fields using less specific expressions. A score of 10 was reserved for manually curated
659 annotations.

660 Development of *Gene Cards* applications

661 Functional annotation was generated for each species using both eggNOG-mapper
662 and Mercator4, extracting GO and MapMan terms, Pfam, Panther and other
663 databases domains, and then reformatting GO and MapMan ontologies to GMT format
664 for gprofiler2 GSEA analysis. In the case of grapevine, curated gene functional
665 information was integrated based on the submission forms provided by the research
666 community (https://grapedia.org/genes/). Gene ID correspondences between
667 grapevine annotations were established by using the annotation extraction and gene
668 integration suite (AEGIS), which was developed *in-house* and is based on a weighted
669 multi-feature coordinate overlap algorithm. A *JBrowse2* genome browser was set up
670 for each species considering the latest genome assembly and annotation versions. An
671 integrated *JBrowse2* window was included using the inbuilt language for gene
672 coordinate queries.

673 Construction of gene co-expression (GCN), inferred regulatory (iGRN) and multi-
674 layered regulatory (rGRN) networks

675 The network construction involved a series of steps to collect, process, and analyze
676 RNA-Seq data obtained from the NCBI SRA database, which was independent of the
677 type of network built; non-aggregated gene co-expression network (non-agg GCN),
678 aggregated GCN (agg GCN), or gene regulatory networks (GRNs). All RNA-Seq reads
679 were processed in a series of steps to prepare them for alignment to the reference
680 genome. This involved trimming the reads using the fastp (version 0.20.0) (default
681 parameters + *--cut_front_window_size 1 --cut_front_mean_quality 30 --cut_front*
682 *cut_tail_window_size 1 cut_tail_mean_quality 30 --cut_tail -l 20*) to remove low-quality

683 sequences and adapter contamination. Subsequently, the trimmed reads were aligned
684 against the corresponding genome assembly for each species (Table 1) using the
685 STAR aligner (version 2.7.3a) (default parameters + *--runMode alignReads --*
686 *limitOutSJcollapsed 8000000 --limitIObufferSize 220000000*). Following alignment,
687 raw count matrices were generated for each SRA run using featureCounts (version
688 2.0.0) and the respective annotation version for the species (Table 1). To account for
689 variations in sequencing depth across different runs, the raw count matrices were
690 normalized using the transcripts per kilobase of exon model per million mapped reads
691 (TPM) method. Genes exhibiting consistently low expression levels, defined as less
692 than 0.5 TPM across all runs within a study, were removed from further analysis.

693 The process of constructing aggregated (agg) and non-aggGCNs involved a
694 combination of correlation analysis and network aggregation techniques as described
695 in (Orduña et al., 2023) with some modifications in the pipeline that permit SRA study
696 tracking for aggGCNs. In the case of aggGCNs, the measure of co-expression is
697 based on the number of SRA studies where a co-expression relationship can be
698 ascertained for any pair of genes. Knowing the nature of these studies and not only
699 the total number can be particularly useful to uncover potential condition-specific
700 relationships. Network performance evaluation was carried out for all GCNs using the
701 EGAD R package, whose resulting metric is an AUROC (area under the receiver
702 operator curve) (Ballouz et al., 2017).

703 The construction of inferred GRNs (iGRNs), were based on the GENIE3 algorithm
704 implemented in the GENIE3 R package (Huynh-Thu et al., 2010). Reference GRNs
705 (rGRNs) included additional *in silico* and experimental sources of evidence to qualify
706 regulatory relationships into 4 confidence levels. The D3 network R package was
707 modified and implemented within the different network apps to show interactive 3D
708 plots based on user-defined gene lists, allowing for focused exploration of co-
709 expression relationships within a specific set of genes. A custom slider to modify
710 repulsion forces is provided. The chromosome map utility was custom built with
711 ggplot2. D3 subnetwork visualization for GRNs was modified to allow for directional
712 edges in the form of arrows (suggestive of regulation) and a color-coded system was
713 implemented for rGRNs to indicate the different levels of evidence supporting a
714 particular relationship.

715 Implementation of a genome browser with DAP-seq data: *DAP-Browse*

716 Sequencing reads obtained from DAP-seq and ChIP-seq experiments were aligned to
717 the corresponding species reference genome using Bowtie2 (Langmead & Salzberg,
718 2012) with default settings. Reads with MAPQ scores below 30 were filtered out during
719 post-processing to ensure data quality. Peak detection was carried out with the GEM
720 peak caller (Guo et al., 2012) using the respective genome assembly and specifying
721 the following parameters: -q 1 -t 1 -k_min 6 -kmax 20 -k_seqs 600 -
722 k_neg_dinu_shuffle. Analysis was limited to nuclear chromosomes, and replicates
723 were processed using GEM's multi-replicate mode. Peak summits identified by GEM
724 were mapped to the nearest gene models from a custom annotation file using the
725 BioConductor package ChIPpeakAnno (Zhu et al., 2010) under default parameters
726 (NearestLocation).

727 To facilitate downstream analysis and visualization, the detected peak coordinates
728 were converted to a BED file format. Alongside the alignment-derived bigWig file,
729 these data were uploaded into the *JBrowse2* genome browser for integration and
730 inspection.

731 Utilizing the cross-species synteny visualization features of *JBrowse2*

732 Pairwise ortholog identification was conducted across all species. For each unique
733 species pair, coding sequences (CDS) were extracted and their corresponding GFF3
734 annotation files for each species were processed to extract the gene models. Only
735 primary transcripts were retained. Orthologous relationships between species pairs
736 were inferred using MCscan (jcvi.compara.catalog ortholog module) from the JCVI
737 utility libraries and incorporated into the genome browser.

738 Differential expression analysis and metabolite integration

739 Transcriptomic and metabolomic integration was performed as described for the
740 *TransMetaDB* app (Savoi et al., 2022). Briefly, expression data was normalized to
741 TPM. Differential expression analysis (DEA) was performed using the LIMMA R
742 package. Genes were classified as differentially expressed when the adjusted p-value
743 (Benjamini-Hochberg correction) was below 0.05. Co-expression network analysis
744 was performed using the Weighted Gene Correlation Network Analysis (WGCNA) R
745 package for metabolite integration.

746 BLAST servers for custom cross-species interoperability

747 Each species' datasets were individually formatted into BLAST-compatible databases
748 using the *makeblastdb* utility. We implemented a local installation of SequenceServer
749 (v3.1.3) (Priyam et al., 2019), an intuitive and user-friendly platform for running BLAST
750 (basic local alignment search tool) queries on transcriptomic, CDS, and protein data.
751 The user is able to select parameters based on the BLAST command line tool. Output
752 formats include alignments, detailed hit descriptions, and graphical overviews of query
753 matches.

754 Ortholog identification and confidence classification

755 Orthologous gene relationships were identified using a multi-evidence strategy that
756 integrates synteny, sequence similarity, and annotation lift-over methods to maximize
757 both confidence and genome coverage. For each candidate ortholog pair, evidence
758 from all applied methods was aggregated to support robust classification.

759 Coordinate-based orthology was inferred using two annotation lift-over pipelines:
760 Liftoff (Shumate & Salzberg, 2021), designed for closely related species or alternative
761 assemblies, and LiftOn (Chao et al., 2024), optimized for more distantly related
762 species, which augments Liftoff with Miniprot (Li, 2023) to guide gene model transfer
763 via protein similarity. In both pipelines, annotations were transferred reciprocally
764 between each genome pair, from source to target genome and *vice versa*. Resulting
765 gene models were then analyzed using our custom tool, AEGIS (annotation extraction
766 genomic integration suite), which detects orthologs by comparing coordinate overlaps
767 between lifted and native gene models. Gene pairs that successfully mapped to each
768 other in both directions were considered reciprocal orthologs, while those matching in
769 only one direction were classified as unidirectional. Each pair was assigned an AEGIS
770 score (1-11) to reflect the quality of the lift-over match, and was annotated with
771 additional flags indicating local synteny conservation ('synteny') or the matching
772 through duplicate gene models ('copies') from the Liftoff/LiftOn outputs.

773 Synteny-based orthologs were identified using the MCscan module from the JCVI
774 toolkit, which detects collinear genomic blocks between species. Two output sets were
775 generated: a high-confidence “anchors” set, and a broader “last_filtered” set based on
776 less stringent criteria. Sequence homology was assessed through multiple

777 approaches. First, OrthoFinder (Emms & Kelly, 2019) was used to cluster proteins into
778 orthogroups and infer gene-level orthology using evolutionary gene trees, with the
779 orthogroup identifier serving as supporting evidence. Additionally, all-versus-all
780 BLASTp searches between proteomes were conducted to identify several categories
781 of sequence similarity. Among these, Reciprocal Best BLASTp Hits (RBBH) were
782 defined strictly as protein pairs where each is the other's top hit, without further filtering.
783 A broader category of Filtered Reciprocal BLASTp Hits (RBH) was also established,
784 retaining reciprocal pairs that satisfied thresholds of E-value ≤ 1e-5, percent identity ≥
785 30%, and alignment coverage ≥ 30%. For each BLASTp-based hit, the platform
786 reports additional metrics including E-value, percent identity, percent coverage, and a
787 length-normalized bitscore to support comparative evaluation. Finally, all sources of
788 orthology evidence were integrated and used to assign ortholog pairs into three
789 confidence tiers based on agreement across methods (Table 1).

790 **Table 1. Criteria for ortholog confidence tier classification.** Combinations of evidence required to
791 classify an orthologous gene pair into one of three confidence tiers: High, Medium, or Lower. These
792 tiers integrate multiple lines of evidence, including coordinate-based lift-over (AEGIS score), sequence
793 homology (RBBH, RBH), and synteny (MCScan) to provide a final reliability assessment for each
794 identified ortholog pair. RBBH: Reciprocal Best BLASTp Hit; RBH: Filtered Reciprocal BLASTp Hit.

Confidence Score	Conditions to Achieve this Reliability Level
High confidence	An AEGIS (Liftoff/LiftOn) score of 7 or greater must be present.
Medium confidence	Any of the following conditions are met: (RBBH), (AEGIS score ≥ 6 and RBH), (RBH and MCScan synteny). In essence, this level requires strong evidence like RBBH, or a combination of slightly weaker evidence types (RBH, MCScan, and moderately high AEGIS scores).
Lower confidence	Any of the following conditions are met: RBH is present (but not combined with other required evidence for a higher tier), MCScan synteny is present (but not combined with other required evidence for a higher tier), Other evidences like Orthofinder or lower AEGIS scores exists but does not meet any of the criteria for medium or high confidence.

795

796 Big data processing through high performance computing

797 Processing the large volumes of transcriptomic data and generating gene co-
798 expression networks required access to high-performance computing (HPC)
799 infrastructure. For this purpose, the Garnatxa HPC cluster at I2SysBio was utilized.
800 The cluster comprises 464 CPU cores and 14 TB of RAM, with peak usage reaching
801 up to 200 CPU cores and 1.2 TB of RAM simultaneously. Efficient handling of these
802 tasks was achieved through the parallelization of Sequence Read Archive (SRA)
803 downloads and data processing pipelines using the SLURM workload manager. This
804 computational capacity was essential to ensure timely and scalable processing of the
805 extensive dataset integrated into PlantaeViz.

Data availability

In compliance with FAIR (findable, accessible, interoperable, and reusable) principles, all data are made accessible directly through the platform's applications, both in their raw format and as processed subsets generated via the various online tools, together with any corresponding metadata. Additionally, supplemental files include key resources such as the ontology framework and other metadata used in building the platform, ensuring transparency and facilitating reuse by the community.

Acknowledgments

The generation of the networks and all the bioinformatic analyses were performed on the HPC cluster Garnatxa at the Institute for Integrative Systems Biology (I2SysBio, UV-CSIC, Spain).

Funding

This work was supported by the grant Valinet (PID2021-128865NB-I00) awarded to J.T.M from the Ministerio de Ciencia, Innovación y Universidades (MCIU, Spain), Agencia Estatal de Investigación (AEI, Spain), and Fondo Europeo de Desarrollo Regional (FEDER, European Union), and to the PROMETEO grant (PROMETEO/2021/056) awarded to P.L. and its associated doctoral grant (PROMETEO/2021/056-01) granted to AS, by the Generalitat Valenciana (GVA). This work was also supported by the Agencia Nacional de Investigación y Desarrollo (ANID)-Millennium Science Initiative Program, Millennium Institute for Integrative Biology iBio ICN17_022 to EAV, ANID-Fondo de Desarrollo Científico y Tecnológico (FONDECYT) 1211130 to EAV, ANID-Anillo ACT210007 to EAV, ANID – MILENIO – NCN2024_047 to EAV, ANID-Vinculación Internacional FOVI230159 to EAV, and ANID-Beca Doctoral 21230478 to JDF.

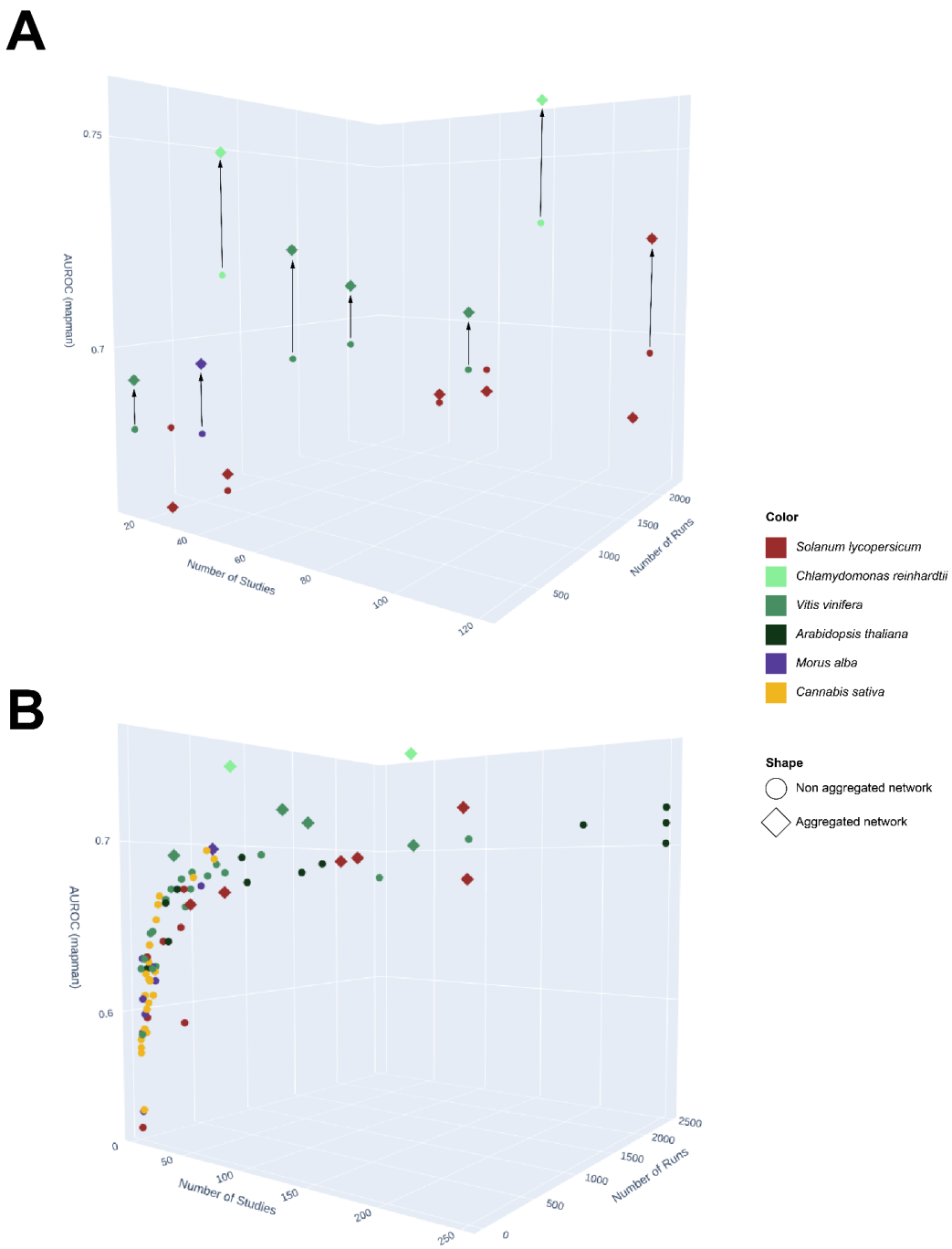
Author contributions

A.S., L.O., D.N.-P., and J.T.M. participated in the design of the platform. A.Vi., I.M.-A., C.Z., P.L., D.P, C.M.-P., I.K., I.E., E.M., M.A.R.-S., A.Va., E.A.V. collected datasets relevant to their species of interest and research expertise. J.J. and A.D. developed the PETAL app. J.D.F. developed the GRN apps. A.S., L.O., and D.N.-P. developed

the GCN apps and the remaining apps and platform structure. A.S., D. N.-P., and J.T.M. wrote the manuscript, and all authors contributed comments and suggestions.

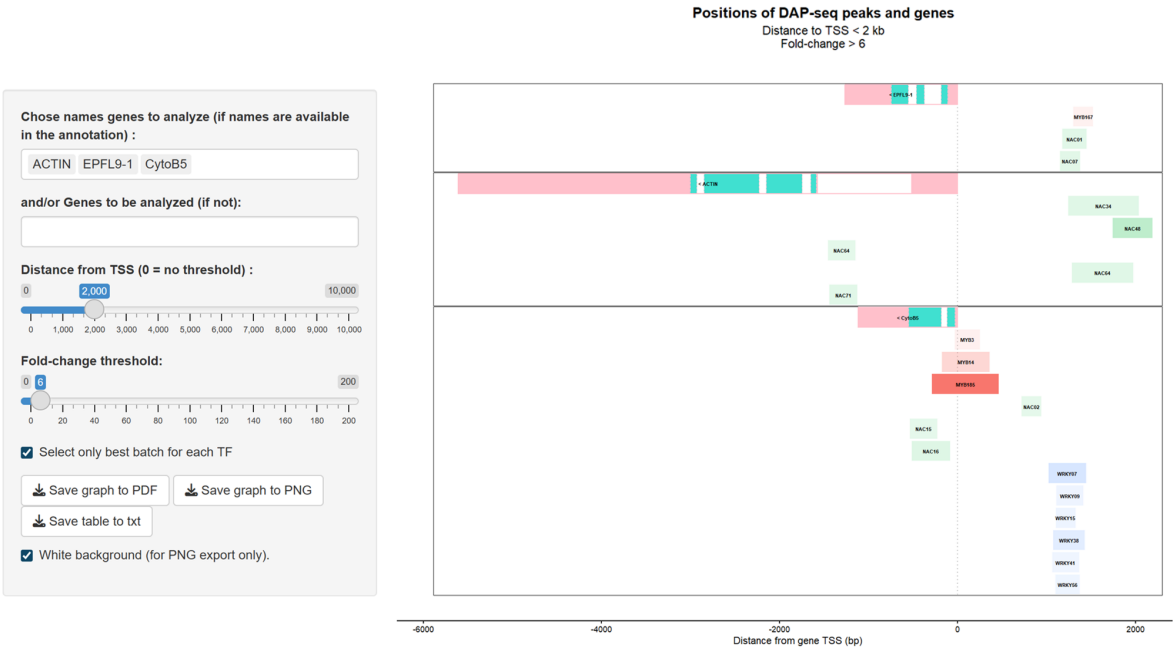
Supplemental Information

Supplemental Figure 1. Example of *EXHARA* and *GeneXplore* expression comparison. (A) Expression heatmap across all SRA runs (*EXHARA*) for Vitis vinifera showing a different pattern for *vviLAC1*, *vviLAC2* and *vviLAC3*. (B) Differential expression profiles of *vviLAC1*, *vviLAC2*, and *vviLAC3* between two grapevine cultivars: cv. Cabernet Sauvignon (red) cv. Pinot Noir (blue). *GeneXplore* enables direct, customizable comparison of expression levels across selected genotypes, tissues, or any other category automatically classified via the platform's ontologies. This feature facilitates the identification of transcriptional patterns and supports the formulation of functional hypotheses regarding gene roles that may only become apparent when looking at particular transcriptomic data subsets. Additional flexibility is provided by allowing filtering by metadata term confidence levels, as well as optionally expanding a query term to include all its child terms within the hierarchy.



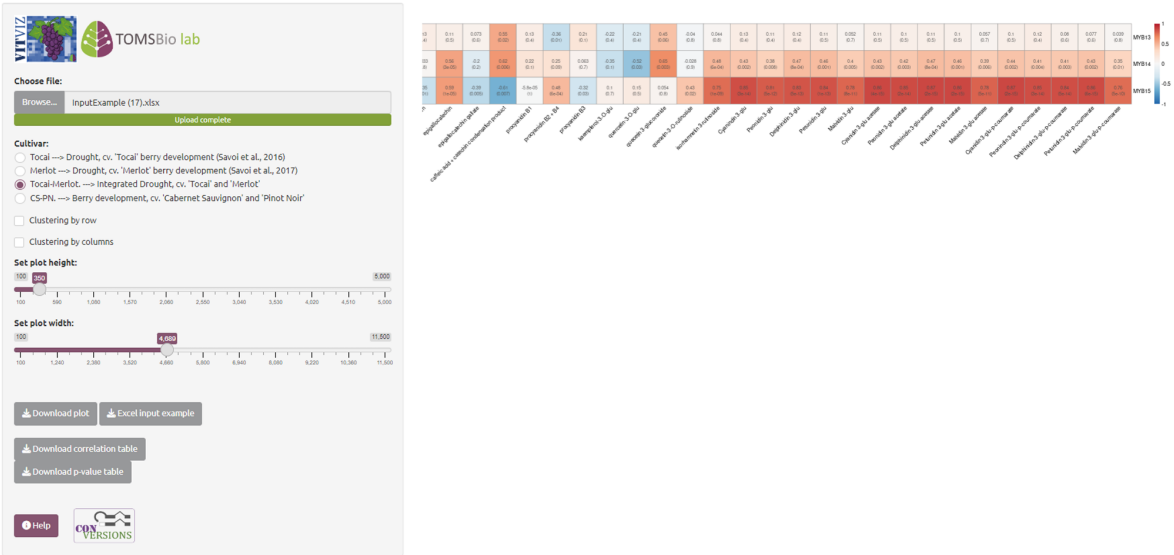
Supplemental Figure 2. Impact of network aggregation and data volume on co-expression network performance. (A) Each vertical arrow connects a non-aggregated network (circle) to its corresponding aggregated network (diamonds) built from the same underlying data. Aggregated networks almost always outperform their non-aggregated counterparts across different species. Exceptions to this rule may be due to the selected tissue or that insufficient runs/studies were available. (B) All generated networks across a wide range of data volumes. This panel illustrates the general trend that network performance improves with an increasing number of studies and/or runs independent of

857 the species within the platform or network method used. Note that lower AUROCs more often
858 correspond to cannabis and mulberry networks, for which lower volumes of data are publicly available.

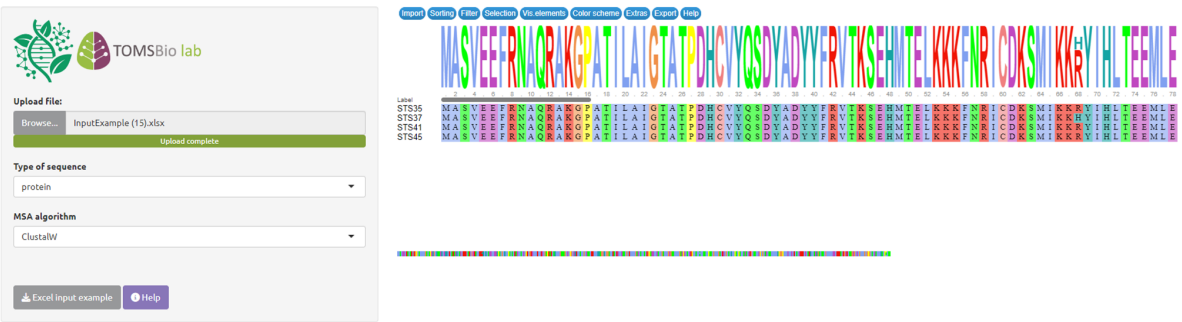


859

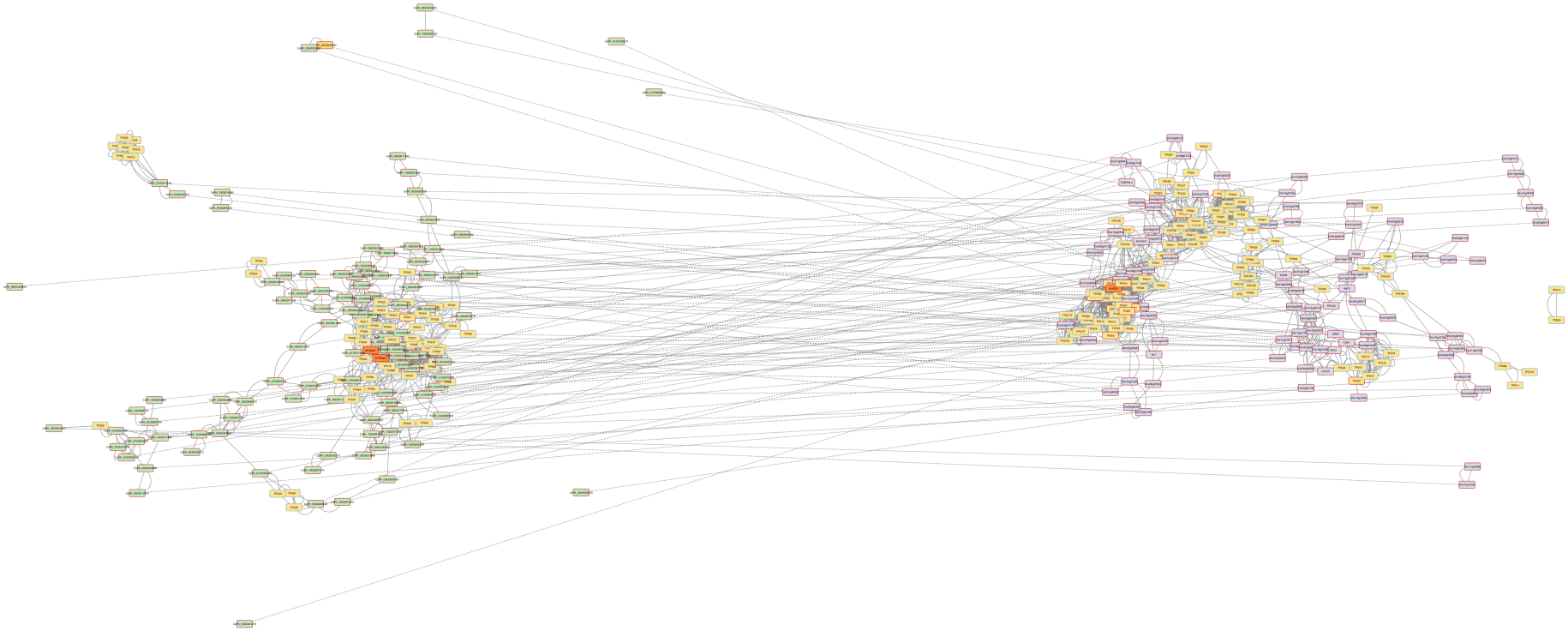
860 **Supplemental Figure 3. ChIP/DAP-seq peaks are visualized for three genes of interest (*vviACTIN*,**
861 ***vviEPFL9-1*, and *vviCytoB5*) in the gene-centered tool of the *PETAL* app.** These are displayed as
862 individual gene plots centered on their respective transcriptional start site (TSS). Users can adjust
863 threshold parameters for the maximum distance from the TSS (set here to 2 kb) and the minimum fold-
864 change of peaks (set to >6). Only the highest-scoring peak batch for each TF is retained. Peaks are
865 color-coded by TF-family, in this example, MYB (red), NAC (green), and WRKY (blue), with color
866 intensity proportional to the fold-change score of the binding event. The promoter region analyzed for
867 each gene is highlighted in pink. The application utilizes custom annotation files to render gene features
868 and overlays peak information derived from DAP-seq data, all within an interactive Shiny interface. This
869 enables intuitive exploration of TF–target gene interactions from a gene-centric view.



Supplemental Figure 4. Example of *TransMetaDB* metabolite-transcript correlation analysis. Pearson correlation coefficients between transcript levels of *VviMYB13*, *VviMYB14*, and *VviMYB15* and multiple quantified metabolites in grape berry skins of *Vitis vinifera* cultivars 'Tocai' and 'Merlot' (Savoi et al., 2016, 2017). Each MYB gene shows distinct correlation profiles with subsets of metabolites, reflecting potential regulatory roles in cultivar-specific metabolic pathways. The app also allows for the use of a berry development dataset with cv. 'Pinot Noir' and 'Cabernet Sauvignon' samples (Fasoli et al., 2018). *TransMetaDB* enables such integrative analyses across paired transcriptomic and metabolomic datasets to support hypothesis generation on transcription factor function in metabolic regulation.



Supplemental Figure 5. Example of MSA app. Alignment of VviSTS35, VviSTS37, VviSTS41 and VviSTS45 proteins.



Supplemental Figure 6. Coexpression networks exported from the GCN app and visualized in Cytoscape. (A) The network topology highlights both conserved and species-specific coexpression patterns. Overall topology similarities support a generally conserved role of MYB24 in terpene biosynthesis regulation in both species, even if some *MYB24-TPS* connections have changed from direct to indirect relationships, or *vice versa*. Node colors: green (C. sativa genes), purple (*V. vinifera* genes), yellow (*TPS* genes), and orange (*MYB24* genes). Red borders indicate DAP-seq binding. Dashed edges connect orthologous genes identified by reciprocal best BLAST hits (RBBH).

References

Ballouz, S., Weber, M., Pavlidis, P., & Gillis, J. (2017). EGAD: Ultra-fast functional analysis of gene networks. *Bioinformatics*, *33*(4), 612–614. https://doi.org/10.1093/bioinformatics/btw695

Carbonell-Bejerano, P., Rodríguez, V., Royo, C., Hernáiz, S., Moro-González, L., Torres-Viñals, M., & Martínez-Zapater, J. (2014). Circadian oscillatory transcriptional programs in grapevine ripening fruits. *BMC Plant Biology*, *14*(1), 78. https://doi.org/10.1186/1471-2229-14-78

Chao, K.-H., Heinz, J. M., Hoh, C., Mao, A., Shumate, A., Pertea, M., & Salzberg, S. L. (2024). Combining DNA and protein alignments to improve genome annotation with LiftOn. *Genome Research*, genome;gr.279620.124v2. https://doi.org/10.1101/gr.279620.124

Childs, K. L., Davidson, R. M., & Buell, C. R. (2011). Gene Coexpression Network Analysis as a Source of Functional Annotation for Rice Genes. *PLoS ONE*, *6*(7), e22196. https://doi.org/10.1371/journal.pone.0022196

Djari, A., Madignier, G., Di Valentin, O., Gillet, T., Frasse, P., Djouhri, A., Hu, G., Julliard, S., Liu, M., Zhang, Y., Regad, F., Pirrello, J., Maza, E., & Bouzayen, M. (2024). Haplotype-resolved genome assembly and implementation of VitExpress, an open interactive transcriptomic platform for grapevine. *Proceedings of the National Academy of Sciences*, *121*(23). https://doi.org/10.1073/pnas.2403750121

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., Porceddu, A., Venturini, L., Bicego, M., Murino, V., Ferrarini, A., Delledonne, M., & Pezzotti, M. (2012). The Grapevine Expression Atlas Reveals a Deep Transcriptome Shift Driving the Entire Plant into a Maturation Program. *The Plant Cell*, *24*(9), 3489–3505. https://doi.org/10.1105/tpc.112.100230

Fasoli, M., Richter, C. L., Zenoni, S., Bertini, E., Vitulo, N., Dal Santo, S., Dokoozlian, N., Pezzotti, M., & Tornielli, G. B. (2018). Timing and Order of the Molecular Events Marking the Onset of Berry Ripening in Grapevine. *Plant Physiology*, *178*(3), 1187–1206. https://doi.org/10.1104/pp.18.00559

Felemban, A., Moreno, J. C., Mi, J., Ali, S., Sham, A., AbuQamar, S. F., & Al-Babili, S. (2024). The apocarotenoid β-ionone regulates the transcriptome of *Arabidopsis thaliana* and increases its resistance against *Botrytis cinerea*. *The Plant Journal*, *117*(2), 541–560. https://doi.org/10.1111/tpj.16510

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A., & Mueller, L. A. (2015). The Sol Genomics Network (SGN)—From genotype to phenotype to breeding. *Nucleic Acids Research*, *43*(D1), D1036–D1041. https://doi.org/10.1093/nar/gku1195

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Guo, Y., Mahony, S., & Gifford, D. K. (2012). High Resolution Genome Wide Binding Event Finding and Motif Discovery Reveals Transcription Factor Spatial Binding Constraints. *PLoS Computational Biology*, *8*(8), e1002638. https://doi.org/10.1371/journal.pcbi.1002638

Huala, E., Dickerman, A. D., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L. A., Bhattacharyya, D., Bhaya, D., Sobral, B. W., Beavis, W., Meinke, D. W., Town, C. D., Somerville, C., & Yon Rhee, S. (2001). The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Research*, *29*(1), 102–105. https://doi.org/10.1093/nar/29.1.102

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE*, *5*(9), Article 9. https://doi.org/10.1371/journal.pone.0012776

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H. (2023). Protein-to-genome alignment with miniprot. *Bioinformatics*, *39*(1), btad014. https://doi.org/10.1093/bioinformatics/btad014

Liesecke, F., De Craene, J.-O., Besseau, S., Courdavault, V., Clastre, M., Vergès, V., Papon, N., Giglioli-Guivarc'h, N., Glévarec, G., Pichon, O., & Dugé De Bernonville, T. (2019). Improved gene co-expression network quality through

expression dataset down-sampling and network aggregation. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-50885-8

Liu, S., Zhong, Z., Sun, Z., Tian, J., Sulaiman, K., Shawky, E., Fu, H., & Zhu, W. (2022). *De novo* Transcriptome Analysis Revealed the Putative Pathway Genes Involved in Biosynthesis of Moracins in *Morus alba* L. *ACS Omega*, *7*(13), 11343–11352. https://doi.org/10.1021/acsomega.2c00409

Livingston, S. J., Quilichini, T. D., Booth, J. K., Wong, D. C. J., Rensing, K. H., Laflamme-Yonkman, J., Castellarin, S. D., Bohlmann, J., Page, J. E., & Samuels, A. L. (2020). Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *The Plant Journal*, *101*(1), 37–56. https://doi.org/10.1111/tpj.14516

Navarro-Payá, D., Santiago, A., Orduña, L., Zhang, C., Amato, A., D'Inca, E., Fattorini, C., Pezzotti, M., Tornielli, G. B., Zenoni, S., Rustenholz, C., & Matus, J. T. (2022). The Grape Gene Reference Catalogue as a Standard Resource for Gene Selection and Genetic Improvement. *Frontiers in Plant Science*, *12*, 803977. https://doi.org/10.3389/fpls.2021.803977

Orduña, L., Santiago, A., Navarro-Payá, D., Zhang, C., Wong, D. C. J., & Matus, J. T. (2023). Aggregated gene co-expression networks predict transcription factor regulatory landscapes in grapevine. *Journal of Experimental Botany*, *74*(21), Article 21. https://doi.org/10.1093/jxb/erad344

Priyam, A., Woodcroft, B. J., Rai, V., Moghul, I., Munagala, A., Ter, F., Chowdhary, H., Pieniak, I., Maynard, L. J., Gibbins, M. A., Moon, H., Davis-Richardson, A., Uludag, M., Watson-Haigh, N. S., Challis, R., Nakamura, H., Favreau, E., Gómez, E. A., Pluskal, T., … Wurm, Y. (2019). Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. *Molecular Biology and Evolution*, *36*(12), 2922–2924. https://doi.org/10.1093/molbev/msz185

Reiser, L., Bakker, E., Subramaniam, S., Chen, X., Sawant, S., Khosa, K., Prithvi, T., & Berardini, T. Z. (2024). The Arabidopsis Information Resource in 2024. *Genetics*, *227*(1), iyae027. https://doi.org/doi.org/10.1093

Santiago, A., Romero, P., Martínez, A., Martínez, M. J., Echeverría, J., Selles, S., Alvarez-Urdiola, R., Zhang, C., Navarro-Payá, D., Pizzio, G. A., Micó, E., Samper, A., Morante, J., Cigliano, R. A., Manzano, D., Bru, R., & Matus, J. T. (2024). *Biosynthesis of oxyresveratrol in mulberry (* Morus alba *L.) is mediated by a group of p-coumaroyl-CoA 2'-hydroxylases acting upstream of*

stilbene synthases. https://doi.org/10.1101/2024.04.04.588114

Savoi, S., Santiago, A., Orduña, L., & Matus, J. T. (2022). Transcriptomic and metabolomic integration as a resource in grapevine to study fruit metabolite quality traits. *Frontiers in Plant Science*, *13*, 937927. https://doi.org/10.3389/fpls.2022.937927

Savoi, S., Wong, D. C. J., Arapitsas, P., Miculan, M., Bucchetti, B., Peterlunger, E., Fait, A., Mattivi, F., & Castellarin, S. D. (2016). Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (Vitis vinifera L.). *BMC Plant Biology*, *16*(1). https://doi.org/10.1186/s12870-016-0760-1

Savoi, S., Wong, D. C. J., Degu, A., Herrera, J. C., Bucchetti, B., Peterlunger, E., Fait, A., Mattivi, F., & Castellarin, S. D. (2017). Multi-Omics and Integrated Network Analyses Reveal New Insights into the Systems Relationships between Metabolites, Structural Genes, and Transcriptional Regulators in Developing Grape Berries (Vitis vinifera L.) Exposed to Water Deficit. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01124

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Shumate, A., & Salzberg, S. L. (2021). Liftoff: Accurate mapping of gene annotations. *Bioinformatics*, *37*(12), 1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(D1), D638–D646. https://doi.org/10.1093/nar/gkac1000

Vannozzi, A., Palumbo, F., Magon, G., Lucchin, M., & Barcaccia, G. (2021). The grapevine (Vitis vinifera L.) floral transcriptome in Pinot noir variety: Identification of tissue-related gene networks and whorl-specific markers in

pre- and post-anthesis phases. *Horticulture Research*, *8*(1). https://doi.org/10.1038/s41438-021-00635-7

Wong, D. C. J. (2020). Network aggregation improves gene function prediction of grapevine gene co-expression networks. *Plant Molecular Biology*, *103*(4–5), 425–441. https://doi.org/10.1007/s11103-020-01001-2

Zhang, C., Dai, Z., Ferrier, T., Orduña, L., Santiago, A., Peris, A., Wong, D. C. J., Kappel, C., Savoi, S., Loyola, R., Amato, A., Kozak, B., Li, M., Liang, A., Carrasco, D., Meyer-Regueiro, C., Espinoza, C., Hilbert, G., Figueroa-Balderas, R., … Matus, J. T. (2023). MYB24 orchestrates terpene and flavonol metabolism as light responses to anthocyanin depletion in variegated grape berries. *The Plant Cell*, *35*(12), 4238–4265. https://doi.org/10.1093/plcell/koad228

Zhu, L. J., Gazin, C., Lawson, N. D., Pagès, H., Lin, S. M., Lapointe, D. S., & Green, M. R. (2010). ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, *11*(1), 237. https://doi.org/10.1186/1471-2105-11-237

Zouine, M., Maza, E., Djari, A., Lauvernier, M., Frasse, P., Smouni, A., Pirrello, J., & Bouzayen, M. (2017). TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. *The Plant Journal*, *92*(4), 727–735. https://doi.org/10.1111/tpj.13711