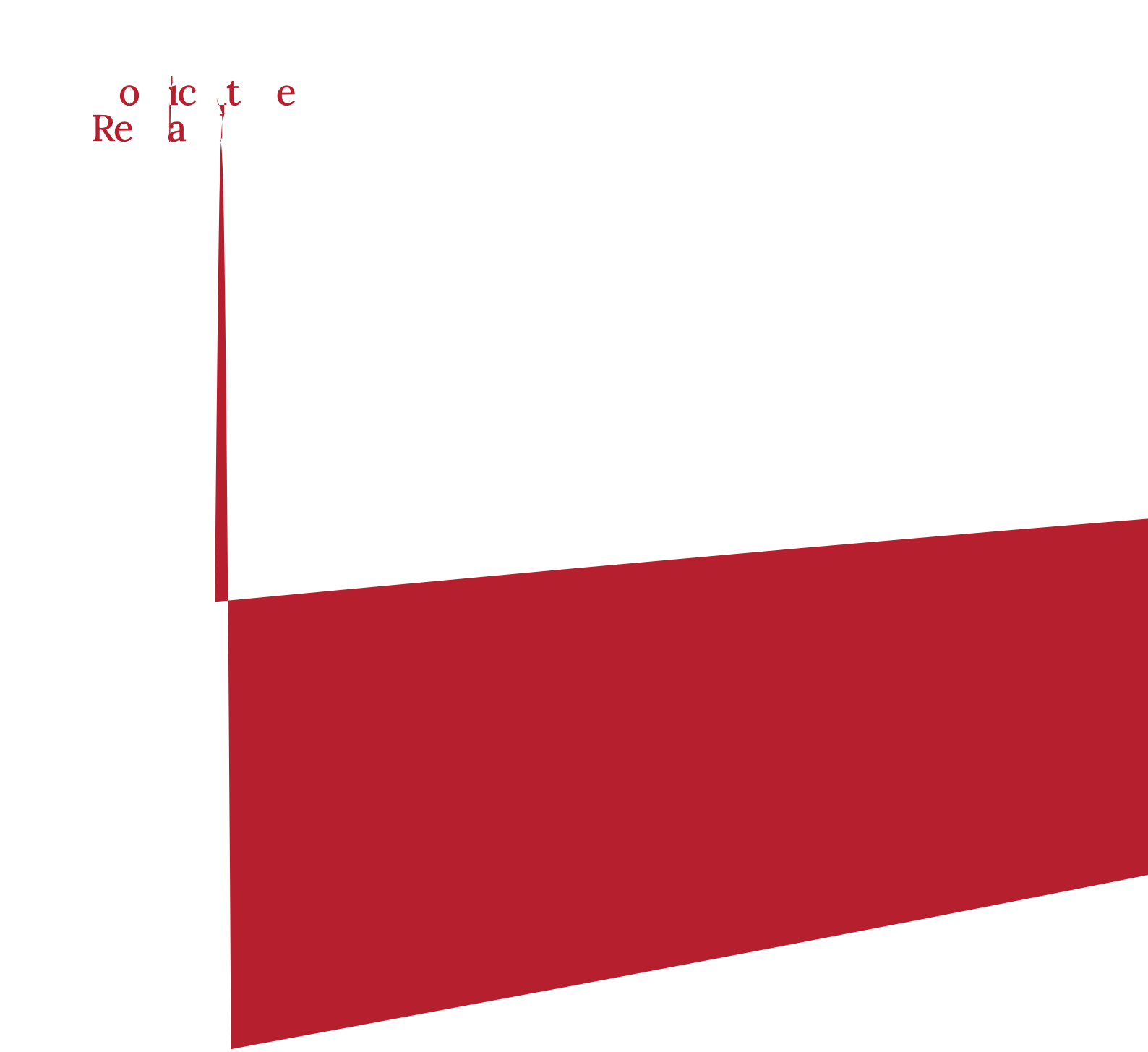

Received: 16 A g 2023; Accepted: 6 N e be 2023; Published: 14 N e be 2023; Corrected and Typeset: 26 Dece be 2023
The A h ( ) 2023. P bli hed b O f d U i e i P e behalf f Na ji g Ag ic l al U i e i Thi i a O e Acce a icle di ib ed de he e f he C ea i e G A ib i Lice e (h : c ea i eo g lice e b 4.0), hich e i e ic ed e e, di ib i , a d e d c i i a edi , ided he igi al k i e l ci ed.

instance, the Genome Database for Rosaceae (GDR) [13] contains a large collection of Rosaceae genomes and integrates various tools for genetic and genomic analyses.

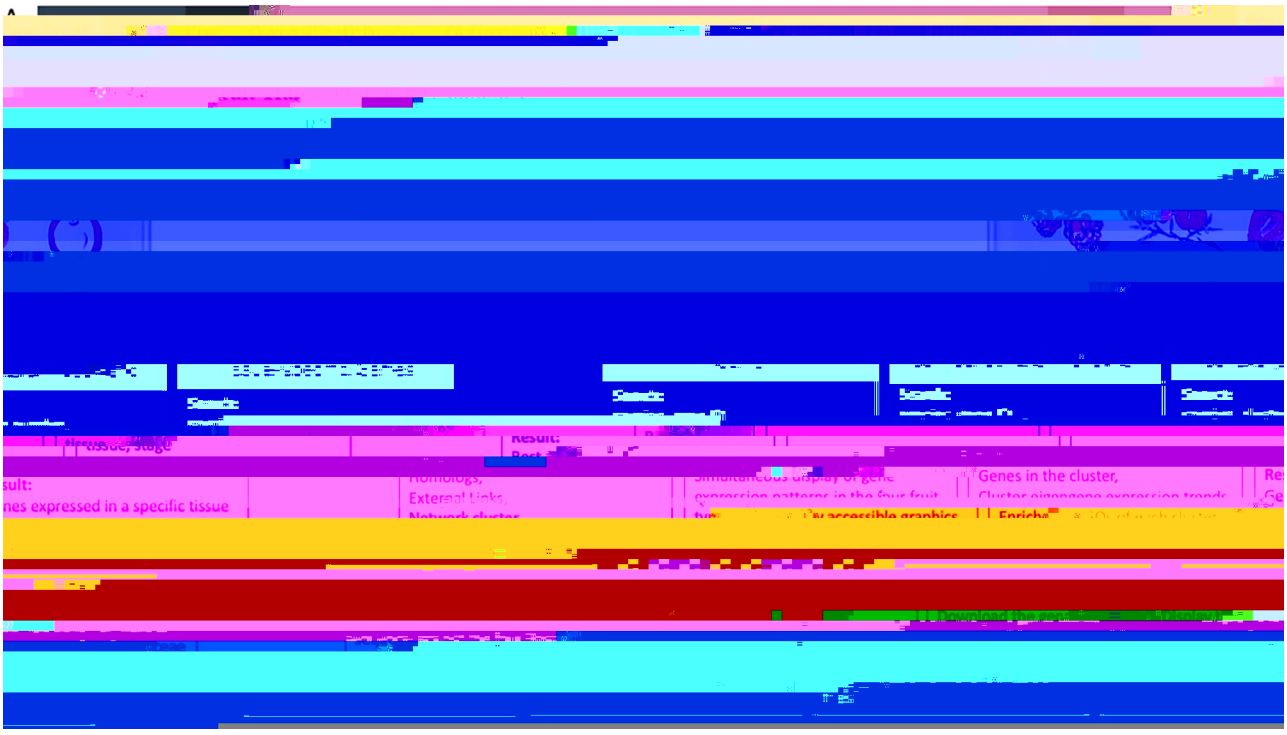

Figure 1. Summary of the ROFT Database (A) A screen shot of the 'Home' page tab. Note the tab bar on top showing different modules as tabs. (B) Summary of the key functions provided by each of the eight tabs.

might be of interest to users looking for fertilization induced gene expression changes specifically in the hypanthium.

In the 'network' subtab (Figure 5), one can select the species and then enter the cluster number in the search box. The search returns with a list of genes (gene ID, its best BLAST hit in Arabidopsis, and the description of the Arabidopsis gene) in the cluster of interest (see Figure 5A). The gene ID and its best Arabidopsis BLAST hit are clickable and linked to the external databases (GDR and TAIR). In addition, the subtab provides the cluster eigengene values across different fruit-related tissues and stages in a boxplot (Figure 5B), as well as the top 20 enriched GO terms for the cluster of interest (Figure 5C). The information will be helpful in identifying co-expressed genes in similar regulatory pathways and conducting comparative analyses between similar or contrasting clusters in different species.

To simplify the network for visualization, correlations among genes in a cluster with a correlation co-efficient equal or below 0.7 were removed. Consequently, not all genes are connected into a single network, leading to multiple subnetworks (components) each with subsets of the genes in a cluster. The subnetwork with the highest number of genes (nodes) in a cluster was saved and its corresponding correlation co-efficient file (containing both TFs and non-TFs) is downloadable by clicking the grey box 'download the file for cytoscape visualization'. Users can open the file in cytoscape [21] for visualization and exploration of the network. In addition, we further filtered the file to save only transcription factors (TFs) with a correlation co-efficient >0.7, and the resulting file was used to construct an interactive network with the R package networkD3 [20] and displayed visually in the 'network' subtab (Figure 5D). Here, users not only can see connections among the nodes (TF genes) in the largest subnetwork (component) of a cluster but also can drag and rotate the nodes as well as zoom in or out of the network to gain understanding of these connections.

Tissue-specific genes

The 'Tissue-specific Genes' tab, available for all four species, allows one to search for genes that are specifically expressed in a tissue at certain stage (select Tissue&Stage-Enriched in the pull-down menu) or expressed in a specific tissue (select the Tissue-Enriched in the pull-down menu) at multiple stages. In this second option, the output table lists the tissue-specific expression in various combinations of stages. Further, one can select 2-fold or 5-fold enrichment under 'Minimum Fold Change', which identifies genes with at least 2-fold or 5-fold higher expression in the selected tissue-stage than all other tissues and stages. Hence, 2-fold is less stringent than the 5-fold criterium and yields more genes. Finally, users can hit the 'Download the entire table' beneath the table to obtain an excel file which provides expression value of tissue-specific genes in all samples.

As illustrated in Figure 6, users first select strawberry under 'species' and then tissue&stage-enriched under 'gene group' from respective pull-down menus. Then, users select 2-fold enrichment under 'minimum fold change'. Finally, users select Ghost under 'tissue' and Stage 3 under 'stage'. Upon hitting 'search', it returns a table containing 314 genes that are at least 2-fold higher in expression in stage 3 ghost than all other tissues and stages. In the ghost tissue (seedcoat and endosperm) at stage 3 (6–7 DPA), which is a stage soon after fertilization, the identified genes are likely induced by fertilization specifically in the endosperm/seedcoat. Among these genes are several MADS-box genes including *AGL62* and *AGL80*. By clicking 'Download the entire table' grey box beneath the table, users will obtain the entire list of 314 genes and their expression levels, providing candidate genes for further functional studies.

Blast

A BLAST function is included in the ROFT database. The users should select a particular Rosaceae species, and the blast will

Downloaded from https://academic.oup.com/hr/article/10/12/uhad240/7420483 by guest on 23 October 2024

Figure 2. Illustration of the 'Gene' tab function. (A) Searching using the peach gene ID (left) or the Arabidopsis gene ID (right). *Prupe.6G127700* is the peach ortholog of the Arabidopsis gene *AT1G08290*. (B) The result from searching the peach gene shown in A. It provides information about the peach gene *Prupe.6G127700*, including its best BLAST hit in Arabidopsis, its Arabidopsis, apple, strawberry, and raspberry orthologs identified by OrthoFinder (https://github.com/davidemms/OrthoFinder), and external links to GDR regarding the specific genome assembly and gene of interest. Also shown are the consensus network cluster that this gene belongs to and the expression pattern for the gene of interest across the three fruit-related tissues at four early fruit developmental stages (Days Post Anthesis, DPA). Hyperlinks lead to additional information on the gene.

return the top blast hits of the specified species. Four BLAST programs, BLASTP, BLASTN, BLASTX, and TBLASTN, are available

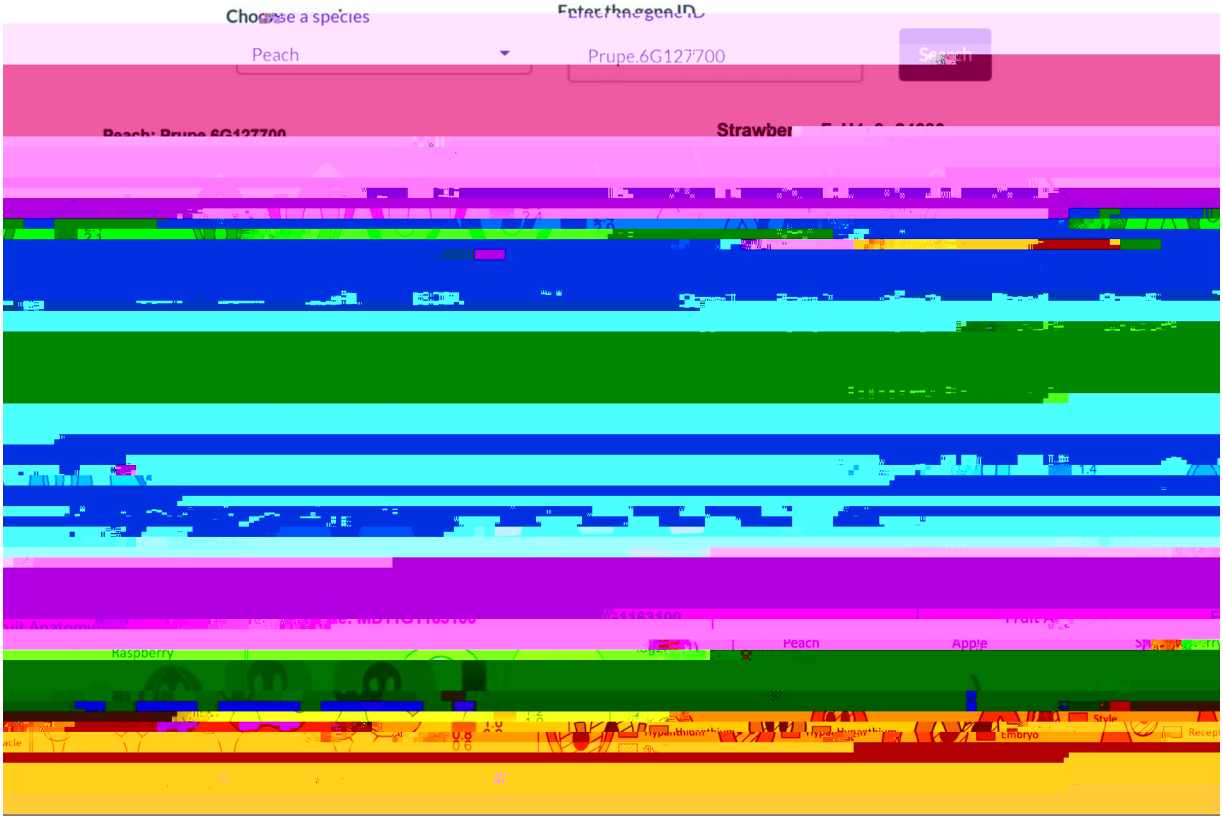

Figure 3. Comparative eFP browser showing the expression patterns of *WIP3* in fruit tissues of all four Rosaceae species. The peach *WIP3* gene (gene ID *Prupe.6G127700*

Figure 4. Illustration of the 'summary' subtab of the consensus 'co-expression network' tab. (A) Summary of consensus co-expression network analysis results showing the number of clusters in each species. (B) Heatmaps of cluster eigengene values, which provide the general expression trend of each cluster. An arrow points to cluster 45 of raspberry.

Case study 2: Metal ion transport appears active in raspberry receptacle and post-fertilization seeds

Previously, we showed that iron can travel from the receptacle to the ghost (seed coat and endosperm) after fertilization in strawberry [14]. The iron transported to the ghost may serve as the cofactor for GA biosynthetic enzymes, GA20ox and GA3ox, which lead to GA synthesis required for strawberry receptacle fruit enlargement [17]. Therefore, we explored the red raspberry consensus co-expression network in ROFT to determine if such iron transport activity may be conserved in the red raspberry. First, through the co-expression network's 'summary' page, we identified raspberry cluster 45 that exhibits receptacle-enriched expression as well as fertilization-induced expression in seeds (see red arrow in Figure 4B). Further exploration of cluster 45 in the 'network' page revealed that the top-ranking enriched GO terms of cluster 45 are associated with metal ion transport and homeostasis (Figure 5C). Hence, similar to strawberry, red raspberry receptacle also appears to experience active iron transport.

Case study 3: Multiple MADS-box transcription factors encoded by *AGL*s (*AGAMOUS-LIKE*) may

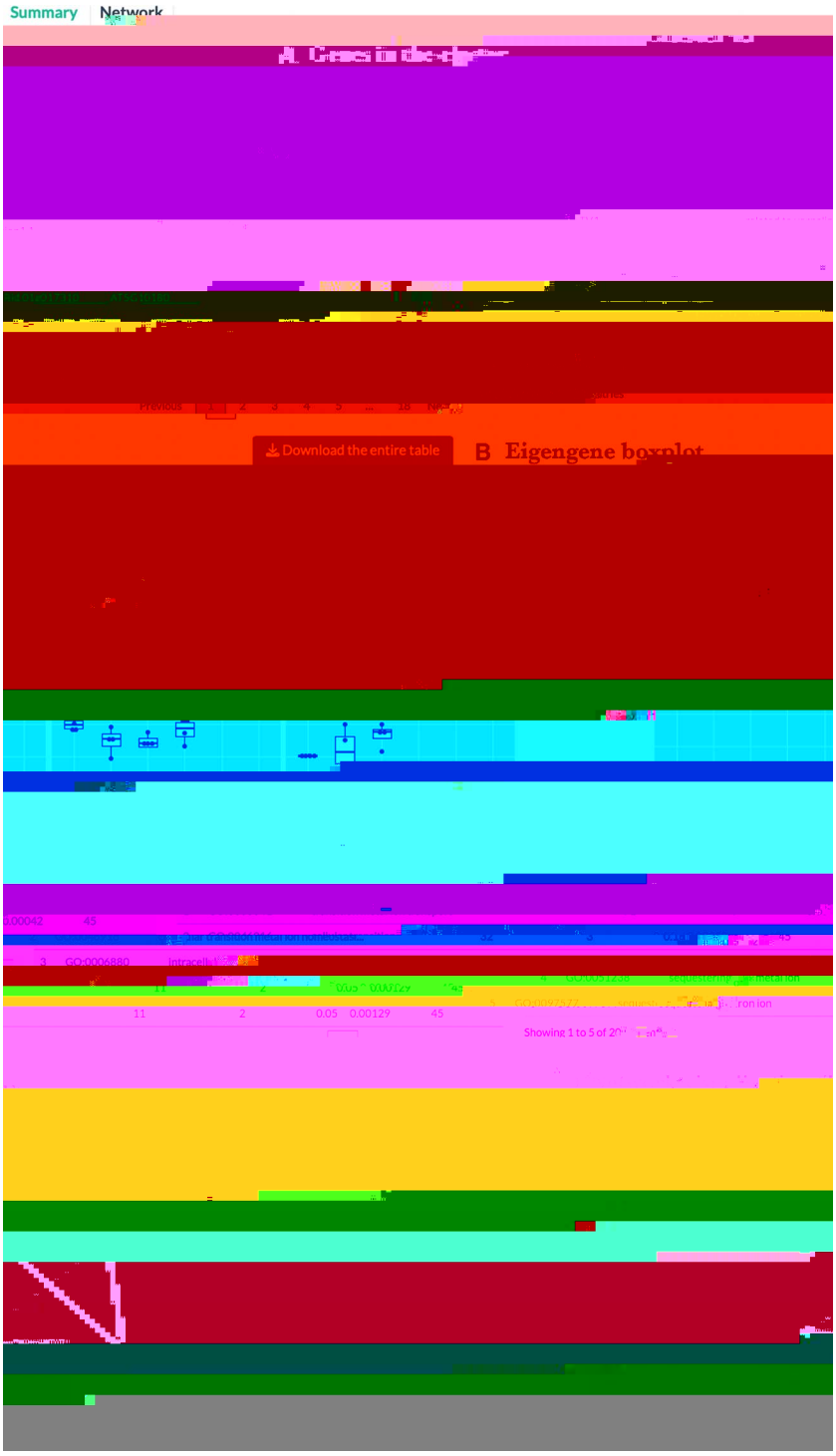

Figure 5. Illustration of the 'network' subtab of the consensus 'co-expression network' tab. By searching for raspberry cluster 45 in the search box, one can obtain a list of 173 raspberry genes belonging to cluster 45. Gene IDs are embedded with hyperlinks leading to information of respective genes in GDR or TAIR. The gene list can be downloaded by clicking the bottom right grey box. (B) Boxplots of cluster 45 eigengene values across raspberry fruit-related tissues and stages. Y-axis indicates the eigengene value, and X-axis indicates the developmental stages (Days Post Anthesis; DPA). (C) Top 20 enriched GO terms for cluster 45 downloadable by clicking the bottom right grey box. (D) Network visualization based on transcription factors (TFs) in the largest subnetwork (component) of cluster 45 with a correlation co-efficient >0.7. The visualization is interactive, where users can drag and rotate nodes (TF genes) and zoom in or out of the network. Edge thickness positively correlates with the correlation co-efficient between nodes; the node size positively correlates with the degree of connectivity (number of connections). For each cluster, a file containing correlations of all genes (TFs and non-TFs) with correlation co-efficient >0.7 in the largest subnetwork (component) of the cluster can be downloaded by clicking the bottom right grey box and explored further with Cytoscape.

Downloaded from https://academic.oup.com/hr/article/10/12/uhad240/7420483 by guest on 23 October 2024

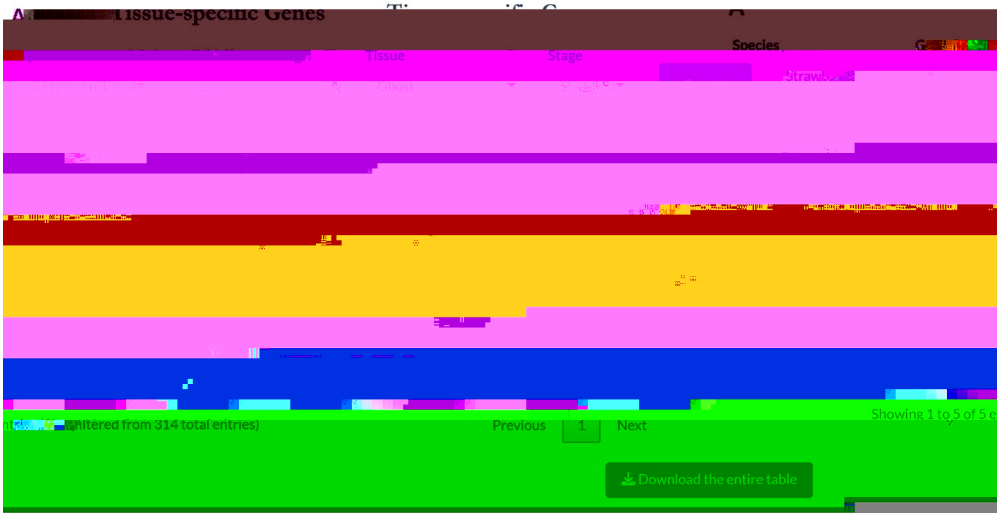

Figure 6. Demonstration of the 'Tissue-Specific Genes' tab. (A) The example search is in strawberry for genes specifically expressed in the 'Ghost' at stage 3 with a minimum of 2-fold enrichment. (B) The resulting gene list from search shown in A, showing a list of strawberry genes enriched specifically in the ghost tissue at stage 3 that includes several MADS-box genes. Note the clickable gene IDs and the 'Download the entire table' button beneath the table.

RNA-Seq data were described previously [11]. The data were deposited at SRA with the accession number PRJNA661345.

The wild strawberry (*F. vesca*) 'yellow wonder' early fruit development was divided into five stages, stage 1 (0 DPA), stage 2 (2–4 DPA), stage 3 (6–7 DPA), stage 4 (8–10 DPA), and stage 5 (10–13 DPA) [26]. The strawberry fruit tissues including style, pith, cortex, ovary wall, and ovule/seed (ghost and embryo) were dissected and harvested at their corresponding stages [17]. Two biological replicates were prepared for RNA-Seq, which was deposited at SRA with accession number PRJNA187983.

Three fruit tissues (receptacle, ovary wall, and ovule/seed) were dissected at six early stages of fruit development (0 DPA, 2 DPA, 4 DPA, 6 DPA, 9 DPA, and 12 DPA) of red raspberry (*R. idaeus*) 'Joan J'. [12]. The RNA-Seq data were collected from four biological replicates and were deposited at SRA with accession number PRJNA869453. Cutadapt (v2.8) [27] was used to trim the low-qualify bases (cutoff: 25) from the 3′ end of the red raspberry reads. Only the reads with a minimum length of 36 bp were retained for the downstream analyses.

Salmon (v0.11.2) [28] was applied to quantify the transcript levels for the four Rosaceae species. The peach, apple, strawberry, and raspberry reference transcripts were retrieved from GDR (*P. persica* Genome v2.0.a1, *Malus domestica* GDDH13 Whole Genome v1.1, *F. vesca* Genome v4.0.a2, and *R. idaeus* Joan J Genome v2.0) [12, 29–32]. For index construction, k-mer size was set to 31, and -keepDuplicates was specified to keep identical sequences in the reference transcripts. And -seqBias was passed to the quantifier to correct the sequence-specific bias. Tximport (v1.10.1) [33] was further utilized to summarize the transcript abundance at gene level.

Ortholog detection

BLAST (v2.5.0) [34] was employed to search the longest protein isoforms of the four Rosaceae species against Arabidopsis protein database generated using the longest peptides in TAIR10_pep_20101214 (https://www.arabidopsis.org/download_files/Proteins/TAIR10_protein_lists/

19.