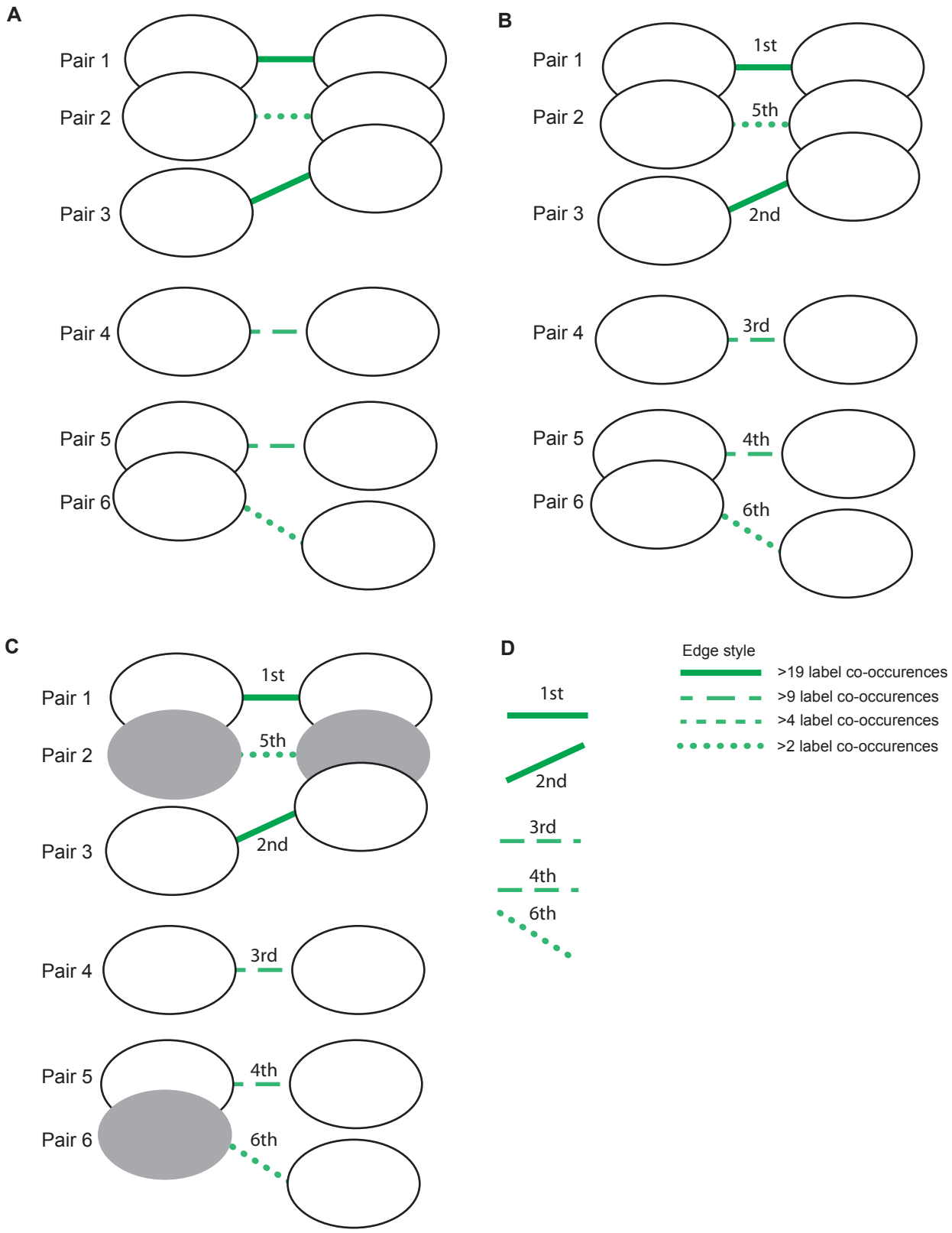
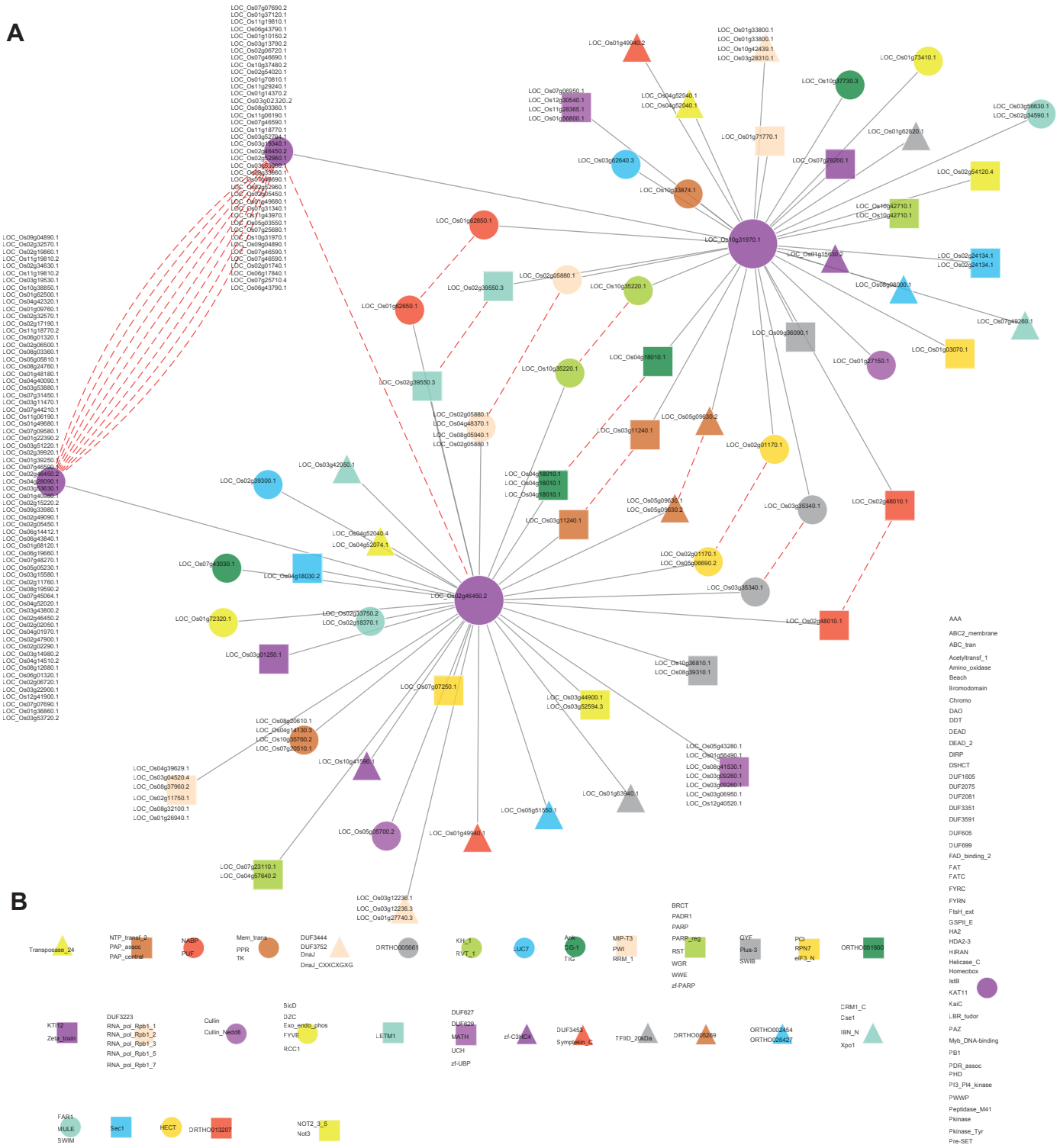
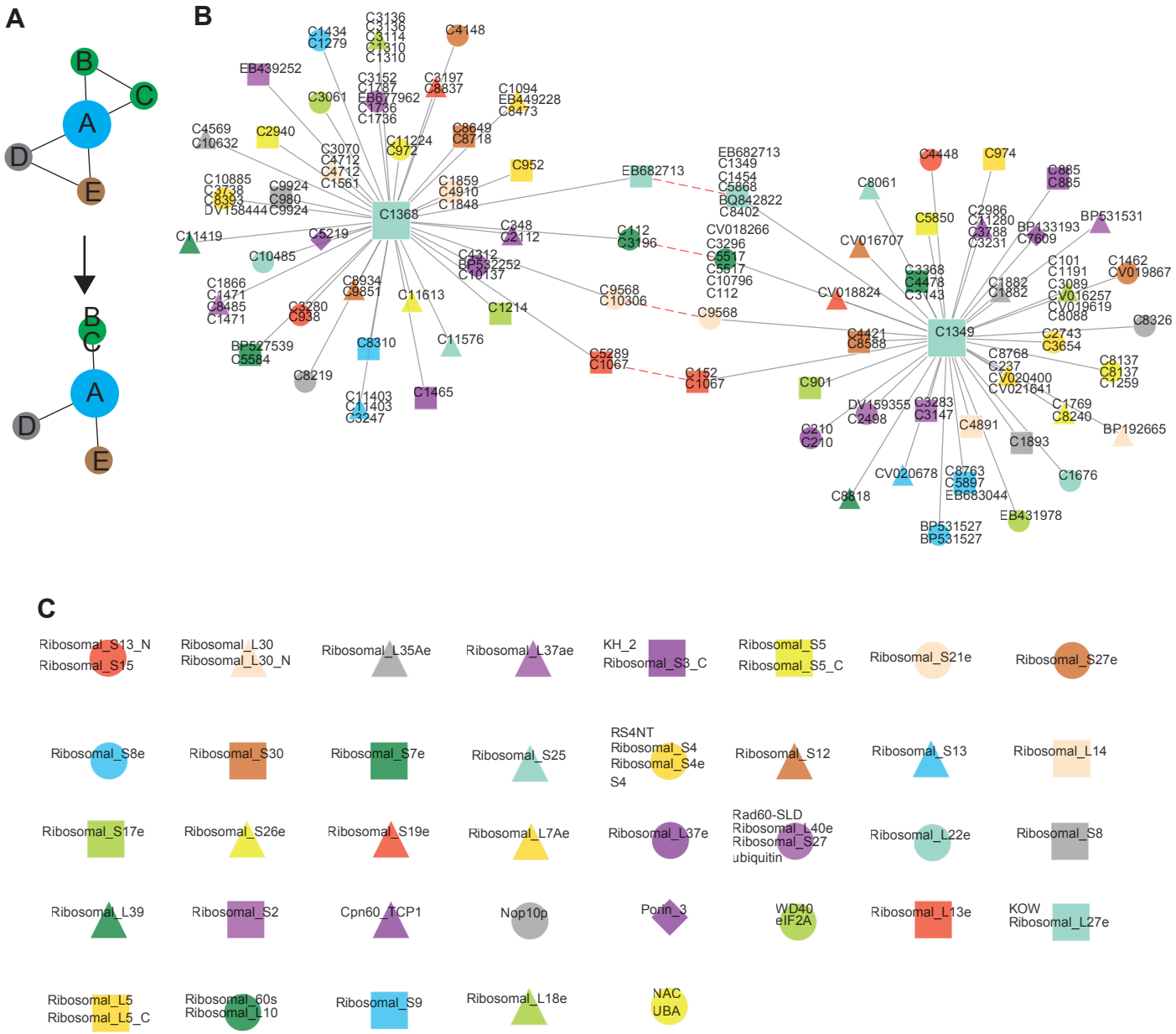


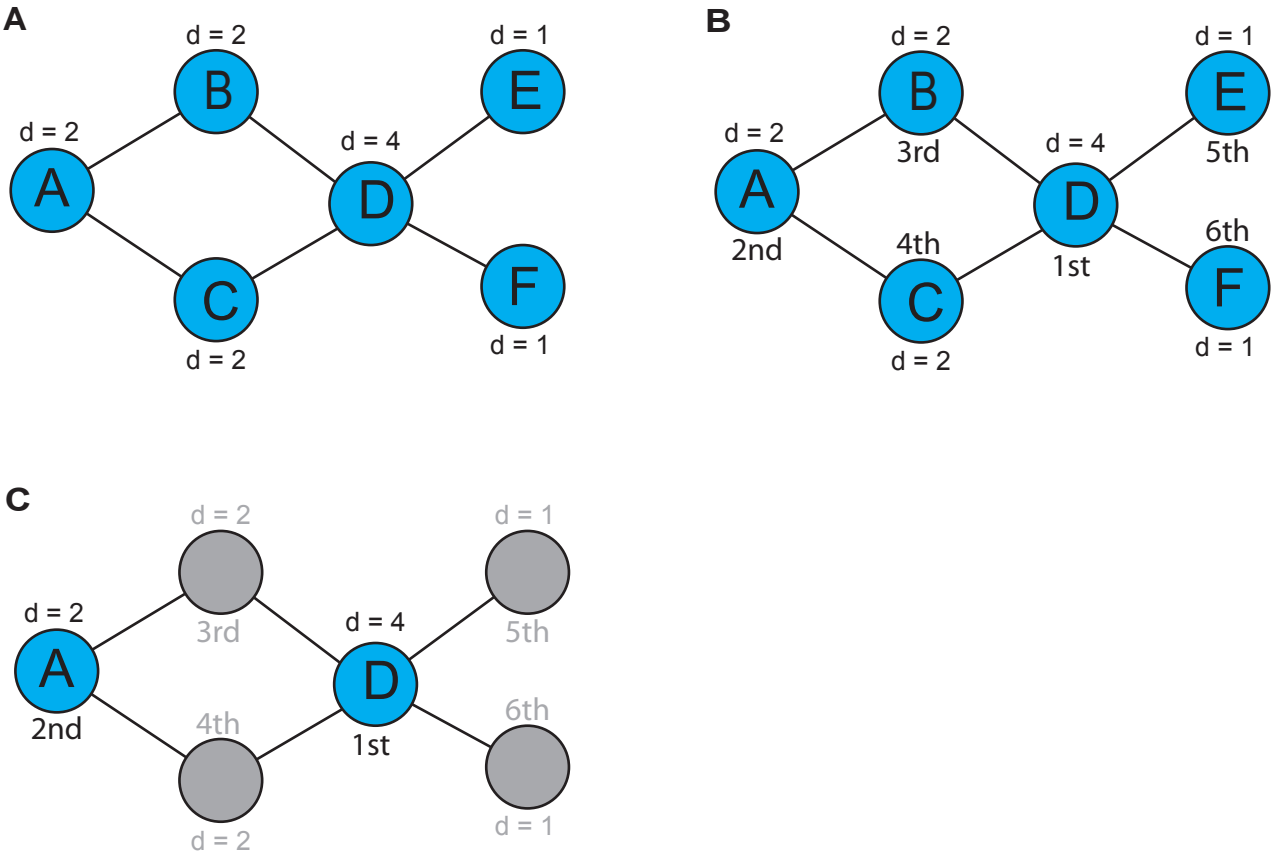
Supplementary Figure 1. Estimating genome-wide distribution of label co-occurrences between gene modules. (**A**) Ellipses represent gene modules, while green edges depict significantly similar gene modules. Number of label co-occurrences between the modules are indicated by edge styles. Overlapping ellipses indicate which modules are sharing genes, i.e. overlapping. (**B**) Module-pair collection is sorted according to the number of label co-occurrences, with more similar module-pairs being collected first. (**C**) Here, the heuristic is determining overlapping modules, with modules having more label co-occurrences having higher precedence over modules with less number of label co-occurrences. For example, both modules in module-pair 2, and one module in module pair 6 are overlapping with modules that have higher precedence. If at least one of the modules is not overlapping with modules of higher precedence, the label co-occurrence value is collected. (**D**) In this example, out of six module pairs, 5 label co-occurrence values are collected. Note that the label co-occurrence value from pair 2 is disregarded, as both modules are overlapping with pair 1.



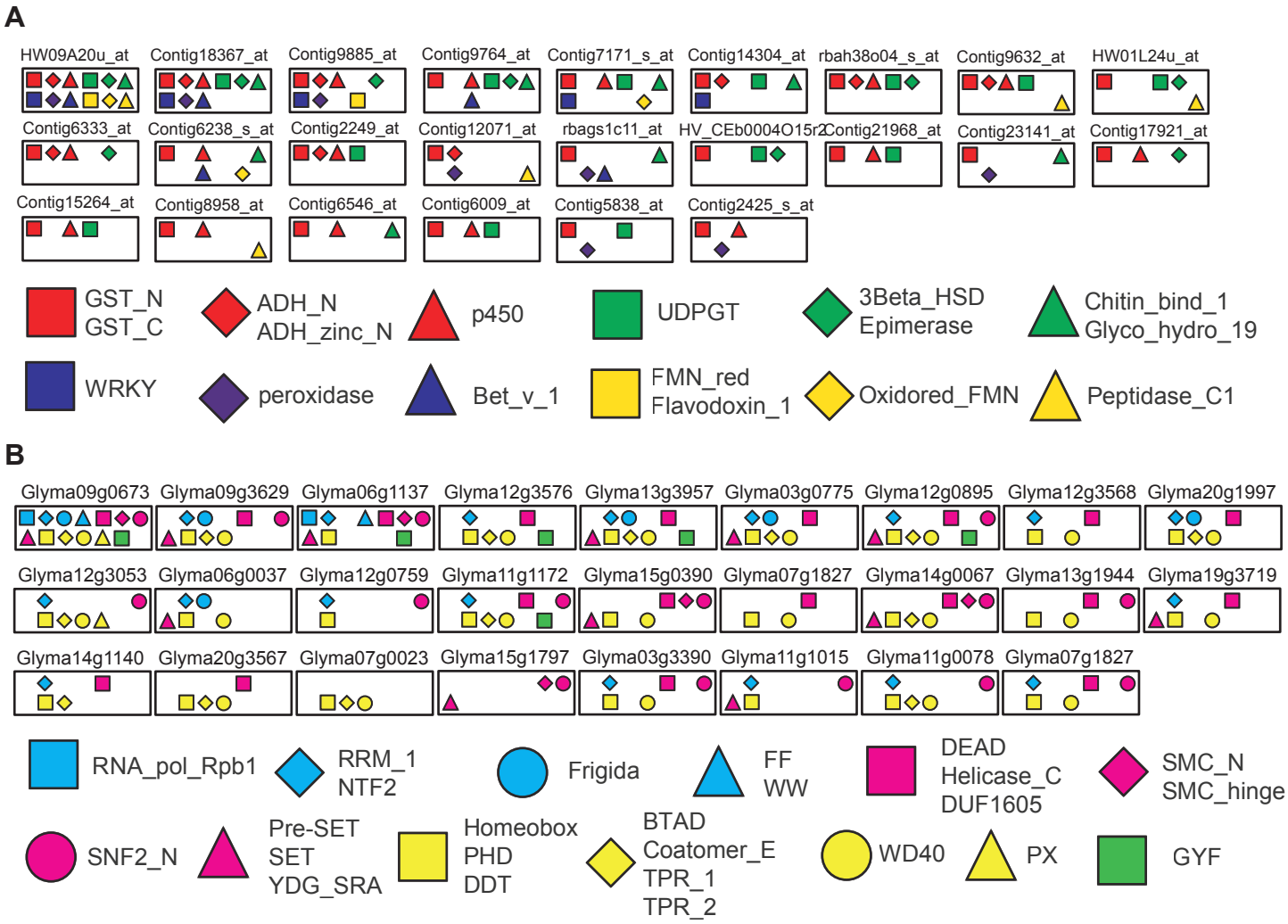
Supplementary Figure 2. An example of large gene modules involved in chromatin remodeling in rice. (**A**) Two gene modules from rice with loc_os10g31970 and loc_os02g46450 used as module centers (large nodes). The nodes represent label co-occurrences, while node labels represent genes assigned to the label co-occurrences. Gray edges represent associations of the label co-occurrences to the module centers. The two modules are overlapping to some degree and consequently share genes, shown by red dashed edges. The number of dashed edges is equal to the number of genes shared between the label co-occurrences. (**B**) Labels found in the label co-occurrences. For simplicity, only pfam labels are shown. The two modules show enrichment in ontologies representing transcription factors, chromatin remodeling/structure factors, signaling and cell division. The ontology analysis for both modules can be viewed at http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!osa!13835 and http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!osa!8427



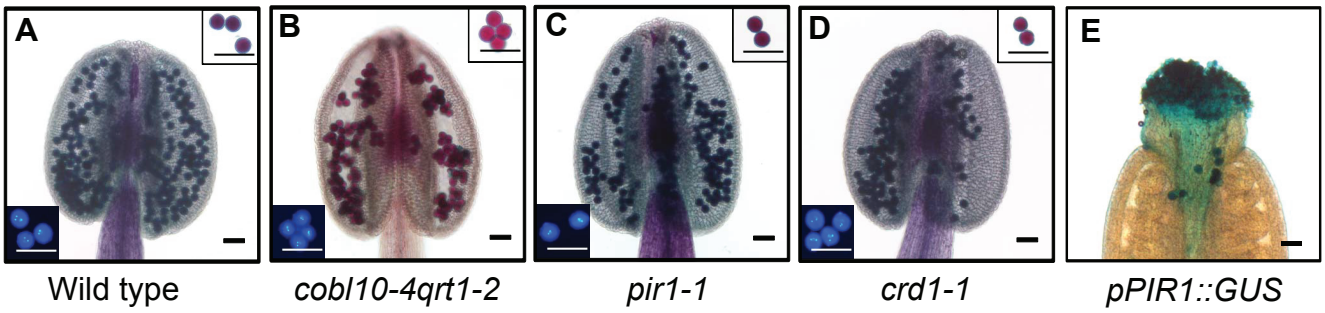
Supplementary Figure 10. An example of large gene modules involved in ribosome biosynthesis in tobacco. (**A**) To make comparisons of gene module content easier, the co-expression networks are simplified by collecting all genes belonging to one label co-occurrence and representing it as one node. In this example, genes B and C belong to same label co-occurrence (green node) and are assigned to the same node in simplified network. (**B**) Two gene modules from tobacco with C1368 and C1349 used as module centers (large nodes). The nodes represent label co-occurrences, while node labels represent genes assigned to the label co-occurrences. Gray edges represent associations of the label co-occurrences to the module centers. The two modules are weakly overlapping and consequently sharing genes, which is shown by connecting the overlapping label co-occurrences by red dashed edges. (**C**) Labels found in the label co-occurrences. For simplicity, only pfam labels are shown. The two modules show enrichment in ontologies representing ribosome structural components.



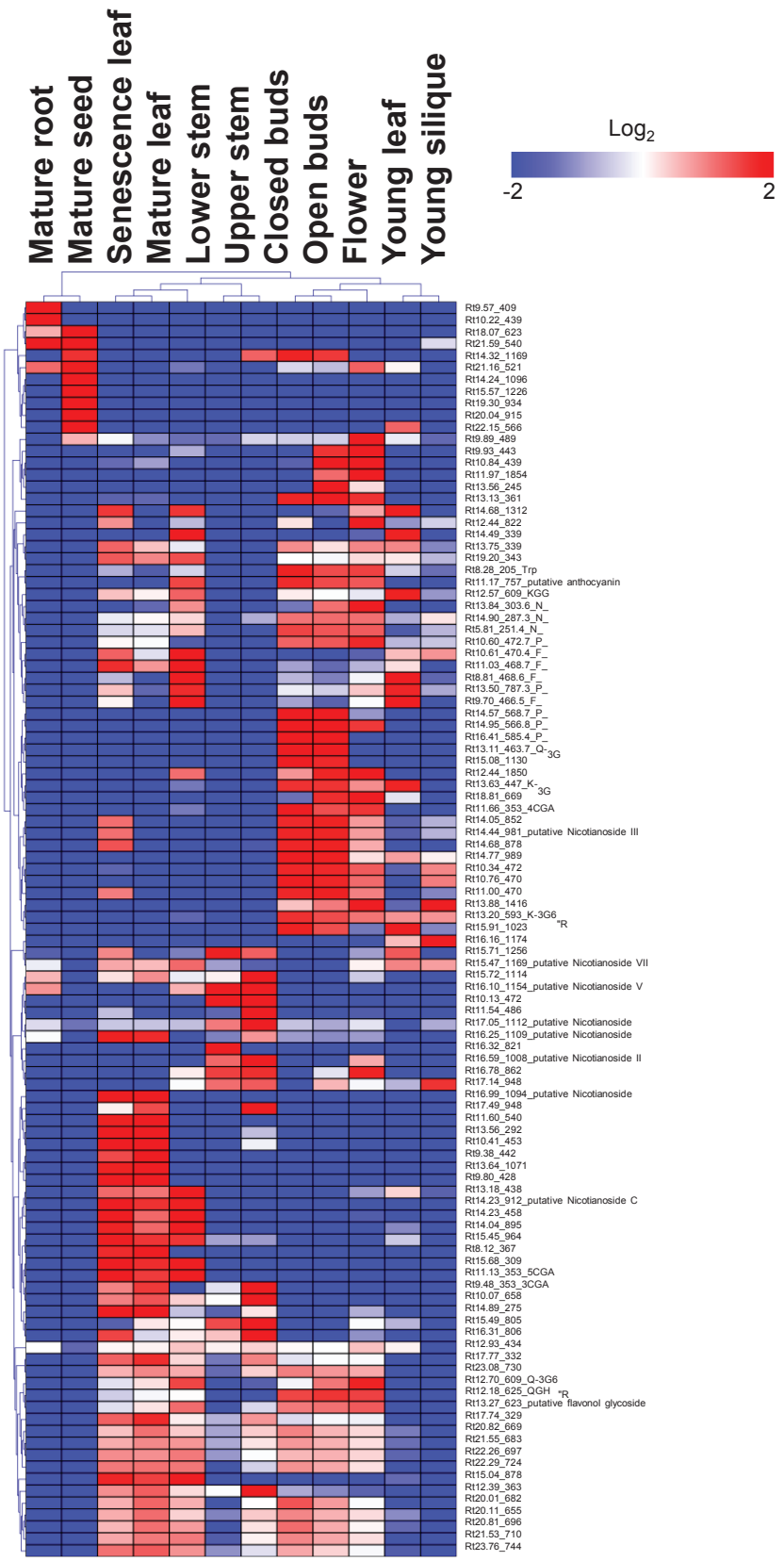
Supplementary Figure 4. Estimating the distribution of representative module degrees.
(**A**) Nodes represent modules, and edges indicate similar modules. Numbers adjacent to a module indicate the degree (d) of a module. (**B**) Module collection is determined by module degree, with modules with higher degree having higher precedence. (**C**) The first module, with highest degree is collected (module D), together with its neighbors (modules B, C,E and F). Modules can only be collected once. In this example, out of six modules, two module degrees were collected (d=2, d=4 for modules A and D).



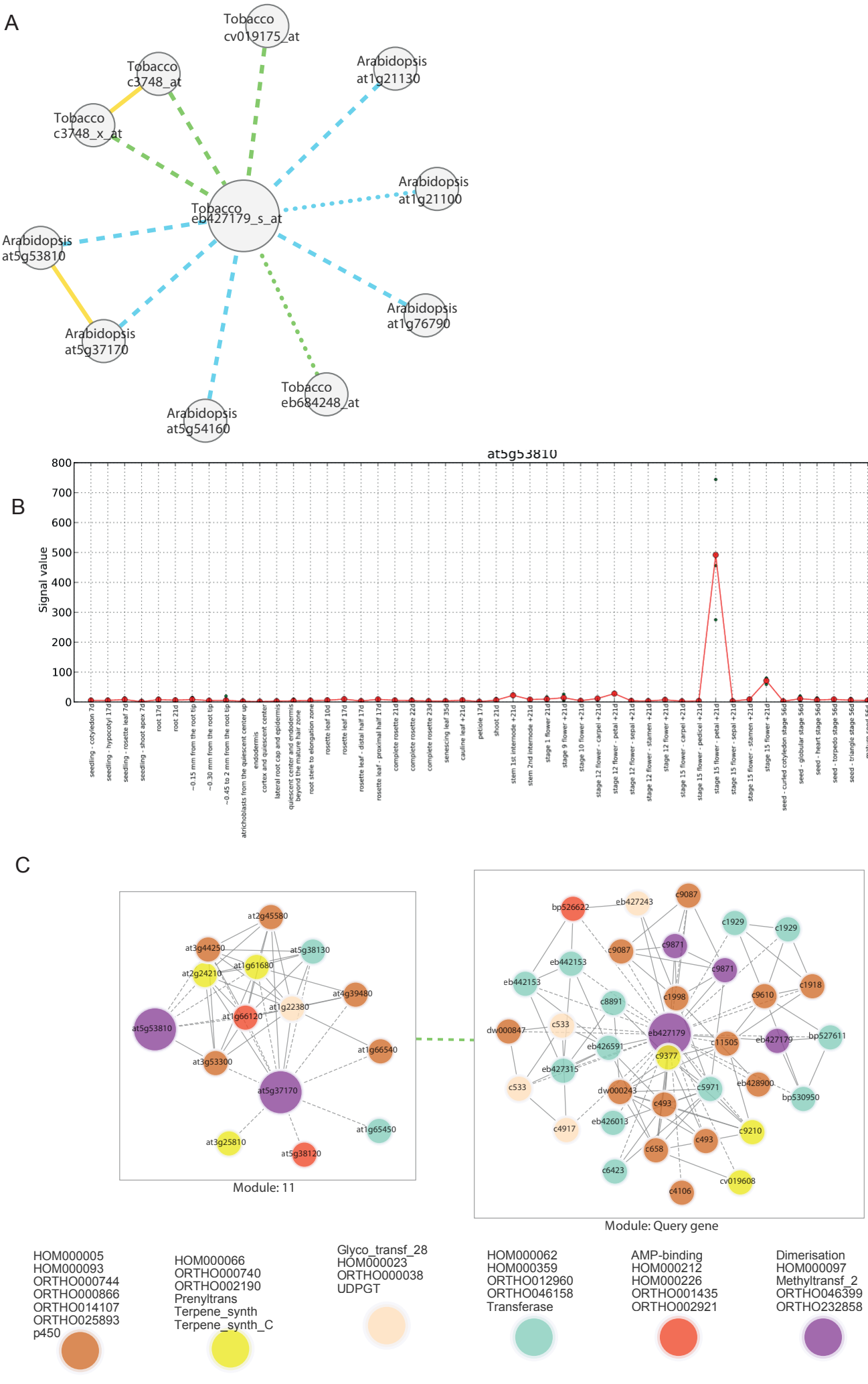
Supplementary Figure 5. Examples of frequently multiplied modules in plants.
Genes/probesets that were used as module centers are indicated above the boxes. Colored shapes indicate label co-occurrences that were present in the respective modules. For simplicity, only pfam labels are shown.
(**A**) Metabolism related modules in barley. (**B**) Transcription related modules in soybean.



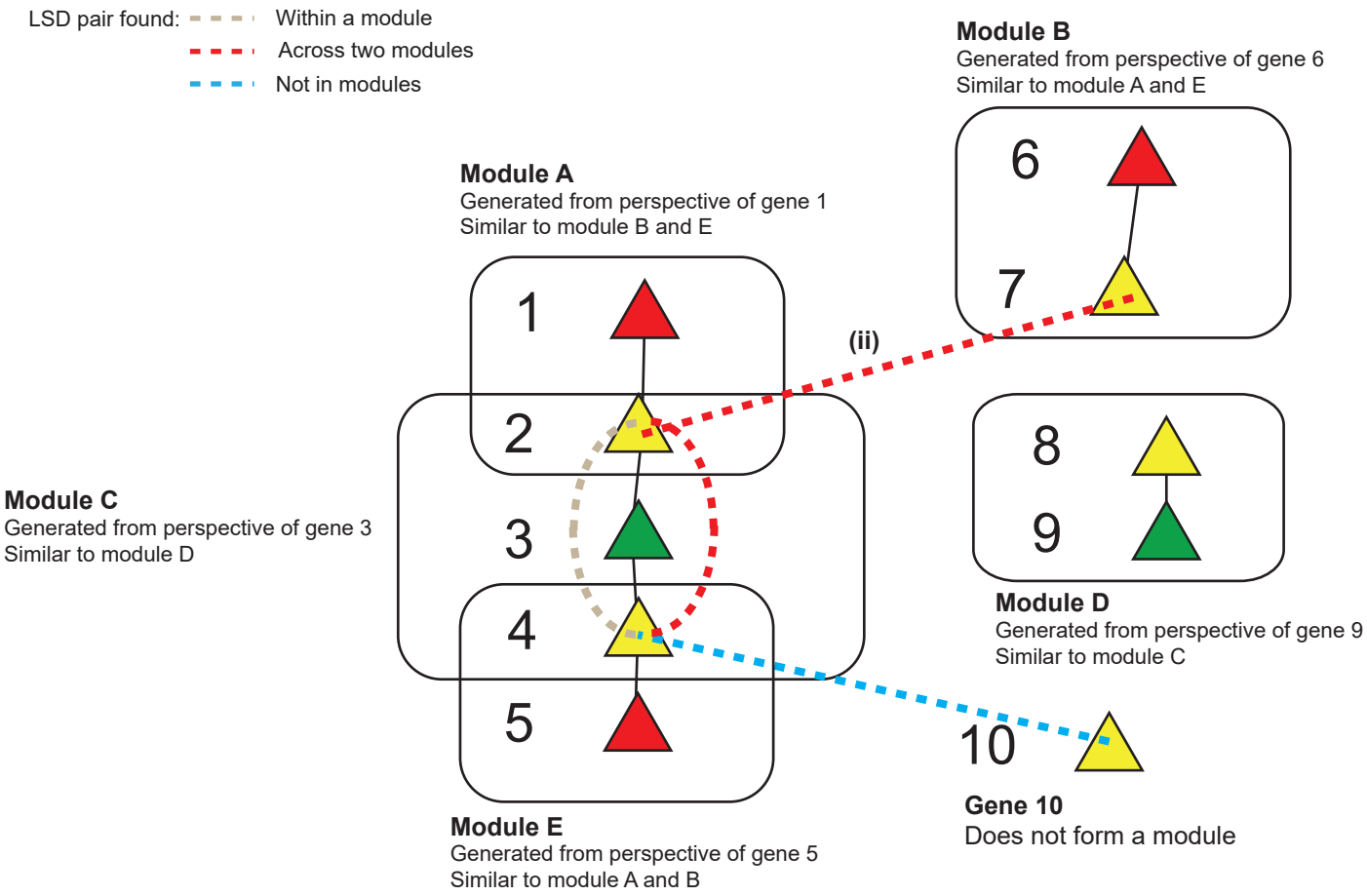
Supplementary Figure 6. Mutants from the pollen cell wall module show normal pollen.
(**A-D**) Whole anthers and mature pollen (inset upper right) stained with Alexander stain and DAPI (insets lower left) indicate that pollen viability is not affected in the mutants. Note that *cobl10-4* was crossed into the *quartet (qrt)1-2* background, which displays tetrads of pollen grains after meiosis (Francis et al., 2006). (**E**) Pollination of wild type pistils with *pPIR::GUS* pollen shows pollen and pollen tube specific expression of *PIR1*.
Scale bars: 50 µm (including insets).



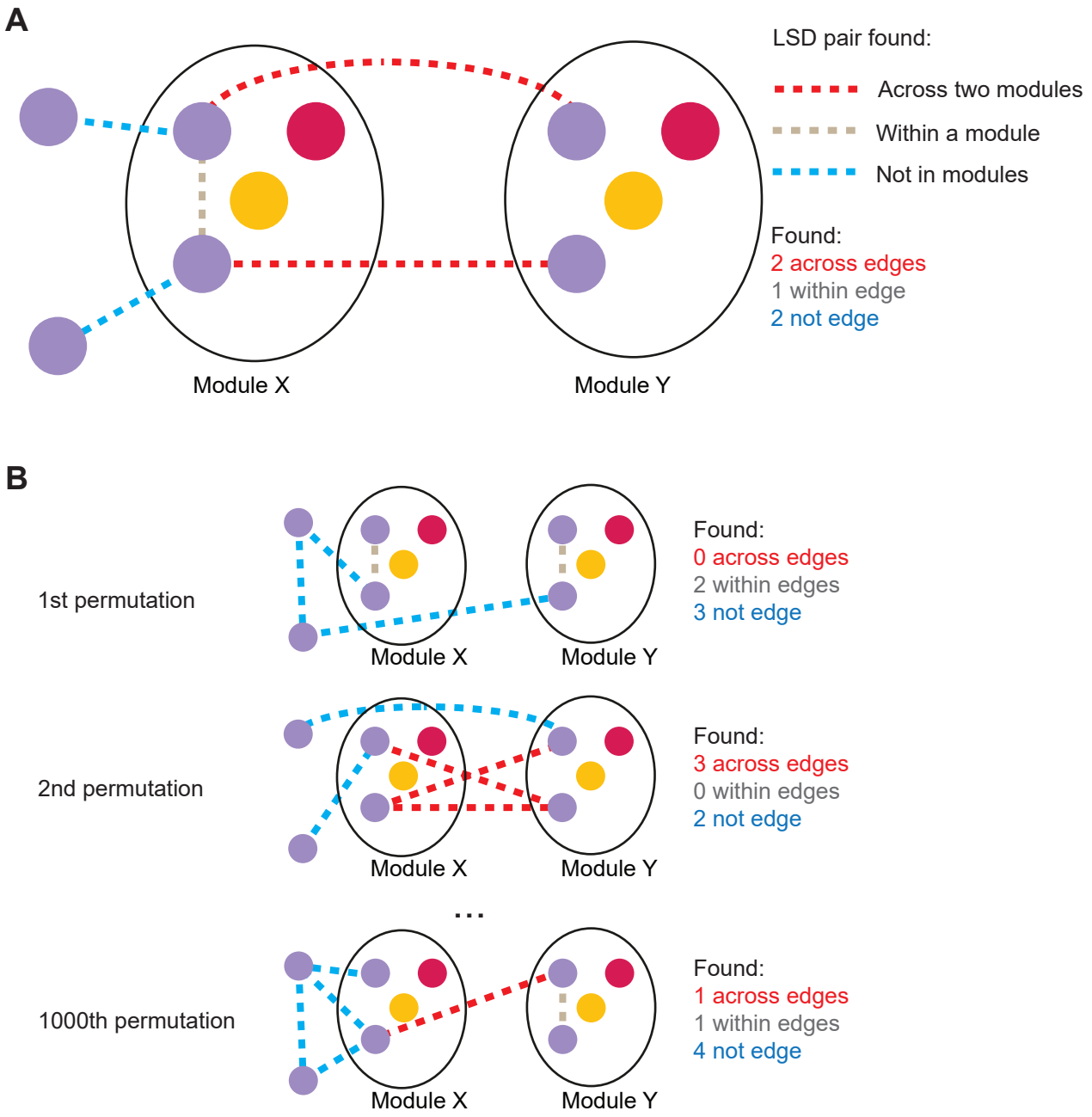
Supplemental Figure 7. Hierarchical clustering analysis of LC-MS metabolite profile of tobacco tissues. Relative peak area was normalized by average value and shown with logarithmic scale (log2). Fold change is visualized by indicating color, red (high) and blue (low), respectively.



Supplementary Figure 8. EB427179-like gene modules in Arabidopsis. A) Gene module network of EB427179 with Arabidopsis modules shown. B) Expression profile of At5g53810. C) Gene module comparison of EB427179 and At5g53810 and At5g37170.



Supplementary Figure 9. Genes can be present in multiple modules and have multiple LSD relationships. Nodes represent genes, while black solid edges represent co-expression relationships. Node colors represent different gene labels. Dashed edges represent the three LSD relationships. In this example, genes 2 and 4 can be in the same module (module C), or in two similar modules (module A and E), depending on the investigated module.



Supplementary Figure 10. Counting and estimating the significance of large-scale duplicated genes (LSD) in modules. (**A**) Consider two similar modules, X and Y, containing four genes each. Nodes and node colors represent genes and labels, respectively. Red edges represent LSD pairs found across the two modules. Gray edges represent LSD pairs found within a module, while blue edges represent LSD pairs not found in two similar modules. In this example, 2 red edges, 1 gray edge and 2 blue edges (5 edges in total) were found. Note that for simplicity, only the violet label is analyzed in this example. (**B**) To estimate the significance of the edge distributions, the 5 LSD edges are distributed randomly among the members of the violet label. The criteria are: the number of edges must stay constant (i.e. 5) and the edges can be only distributed among the violet label. The LSD edges are permuted 1000 times, and the number of red, gray and blue edges is counted for each permutation. The analysis is done for each label, if any LSD gene-pairs are found for the label.