Co-expression clustering across flower development identifies modules and candidate genes for diverse floral forms in *Achimenes* (Gesneriaceae)

Wade R Roberts[1,2], Eric H Roalson[1]

[1] School of Biological Sciences, Washington State University, Pullman, WA, United States
[2] Current address: Department of Biological Sciences, University of Arkansas, Fayetteville, AR, United States

Corresponding Author:
Wade Roberts[1]
Department of Biological Sciences, University of Arkansas, SCEN 601, Fayetteville, AR, 72701, United States
Email address: wader@uark.edu

Abstract

Background. Genetic pathways involved with flower color and shape are thought to play an important role in the development of floral forms associated with different pollination syndromes, such as those associated with bee, butterfly, or hummingbird pollination. Because pollination syndromes are complex traits that are orchestrated by multiple genes and pathways, the gene regulatory networks have not been explored. Gene co-expression networks provide a systems level approach to identify important contributors to floral diversification.

Methods. RNA-sequencing was used to assay gene expression across two stages of flower development (an early bud and an intermediate stage) in 10 species of *Achimenes* (Gesneriaceae). Two stage-specific co-expression networks were created from 9503 orthologs and analyzed to identify module hubs and the network periphery. Module association with bee, butterfly, and hummingbird pollination syndromes was tested using phylogenetic mixed models. The relationship between network connectivity and evolutionary rates (d_N/d_S) was tested using linear models.

Results. Networks contained 65 and 62 modules that were largely preserved between developmental stages and contained few stage-specific modules. Over a third of the modules in both networks were associated with floral form and pollination syndrome. Within these modules, several hub nodes were identified that related to the production of anthocyanin and carotenoid pigments and the development of flower shape. Evolutionary rates were decreased in highly connected genes and elevated in peripheral genes.

Discussion. This study aids in the understanding of the genetic architecture and network properties underlying the development of floral form and provides valuable candidate modules and genes for future studies.

Introduction

Flowers display an extraordinary range of forms, shapes, and colors. Understanding the evolution and development of diverse flowers across angiosperms has long been a major goal in biology. This variation has often been attributed to pollinator-mediated selection and shifts in floral form are often observed over time (Stebbins 1970; O'Meara et al. 2016). The evolution of flowers and the relationship between plant and pollinator is considered one of the key innovations in angiosperm diversification (Faegri and van der Pijl 1979; Fenster et al. 2004; Van der Niet and Johnson 2012; Barrett 2013; Sauquet and Magallón 2018). Characters of the floral phenotype are tightly linked to pollinator efficiency and overall plant reproductive success (Rosas-Guerrero et al. 2014). Pollinators may positively select specific floral traits across evolutionary time (Gervasi and Schiestle 2017) and flowers may be under constant and strong selective pressure. The genetic basis of floral and pollination syndrome divergence has been examined in model systems extensively, particularly in *Mimulus* (Bradshaw et al. 1998; Yuan et al. 2016) and *Petunia* (Hoballah et al. 2007; Hermann et al. 2015), and more recently in many non-model systems (e.g., Wessinger et al. 2014; Alexandre et al. 2015).

Among the most widely used large-scale experimental approaches to investigate genome function are transcriptome analyses (e.g., Pickrell et al. 2010; Raherison et al. 2015), particularly in non-model organisms where no reference genome or functional genomics data exists. The transcriptome is highly dynamic as a consequence of quantitative variation over spatial and temporal scales. Changes in expression often result from the combinatorial action of genetic regulatory pathways orchestrating development and responses to environmental stimuli. Therefore, the transcriptome can be viewed as a link between the genotype and phenotype and may be acted upon through selection (Romero et al. 2012; Prasad et al. 2013). As changes in gene expression may underlie many of the phenotypic changes between species (Brawand et al. 2011; Romero et al. 2012; Uebbing et al. 2016), studying the transcriptome may shed light on important pathways and targets of selection. Phenotypic changes frequently arise through functional changes in conserved developmental pathways among closely related species. For example, changes in enzyme function and transcriptional regulation of the anthocyanin pathway has been implicated frequently across angiosperms in the evolution of red flowers (e.g., Des Marais and Rausher 2010; Smith and Rausher 2011). It is now recognized that most genes act as members of biological pathways or of co-regulated modules (Stuart et al. 2003).

Here, we undertake a comparative study of gene expression during flower development in the genus *Achimenes*. This group is a member of the diverse African violet family (Gesneriaceae) and distributed throughout Mexico and Central America. *Achimenes* is a young lineage (*c*. 7-12 Mya; Roalson and Roberts 2016) known for its floral diversity, a feature thought

to be associated with speciation (Ramírez Roa 1987; Roalson et al. 2003). Four pollination syndromes are found in *Achimenes,* including bee (melittophily), butterfly (psychophily), female euglossine bee (euglossophily), and hummingbird (ornithophily) (Figure 1). These syndromes have traditionally been defined on the basis of flower color and flower shape (Ramírez Roa 1987) and recently through pollinator observation (Martén-Rodríguez et al. 2015; Ramírez-Aguirre et al. 2019). Repeated origins and transitions between these pollination syndromes have been hypothesized in *Achimenes*, making this small lineage an attractive system to understand the genetic and ecological factors underlying floral diversification.

In this study, we aimed to characterize and compare patterns of gene co-expression during flower development across 10 *Achimenes* species. So far, few studies have investigated the genetic basis of floral form in the Gesneriaceae family (e.g., Alexandre et al. 2015; Roberts and Roalson 2017). Given that bee, butterfly, and hummingbird pollination syndromes evolved in parallel across *Achimenes,* we hypothesized that: (1) distinct sets of co-expressed genes would correlate to each syndrome during flower development, and (2) genes in pathways involved with flower color and shape may be central in each network during flower development. Using a comparative transcriptome approach, we clustered a shared set of 9503 orthologs into two different co-expression networks, each corresponding to two stages of flower development: an early bud stage and an intermediate stage. This strategy allowed us to compare gene co-expression patterns temporally. We compared these networks and tested whether gene co-expression clusters correlated with different floral traits (flower color, flower shape, spurs). We also tested whether network properties, such as connectivity and expression, influence evolutionary constraints. Our results allowed us to quantify the extent of shared gene co-expression among closely related species and to identify candidate genes and pathways that may be associated with the repeated evolution of floral forms underlying pollinator specificity.

Materials & Methods

Plant materials

Ten of the 26 currently recognized species in *Achimenes* were sampled: five from Clade 1 and five from Clade 2 (Figure 1). These species were chosen based on their hypothesized close relationships and their diversity in floral form (Roalson et al. 2003). On the basis of previous molecular studies (Roalson and Roberts 2016), two species were chosen as outgroups to represent related lineages: *Eucodonia verticillata* and *Gesneria cuneifolia* (Figure 1). All selected species were grown in standard greenhouse conditions at Washington State University, under 16-hour days, 24-27°C, and 80-85% humidity. Whole flower buds from two stages of flower development were sampled for each species: an immature bud stage (Bud stage; Figure 1) and an intermediate stage before anthesis (D stage; Figure 1). Each stage was sampled with two or three biological replicates, contributing four or six samples for each species. The bud stage was chosen to represent a baseline level of gene expression when floral phenotypes are largely similar that could then be compared to the intermediate stage where many important pathways

118 involved in flower development and pigmentation show increased expression levels (Roberts and
119 Roalson 2017). All tissues were immediately frozen into liquid nitrogen and stored at -80°C.
120
121 **RNA extraction and RNA-seq library construction**
122 Total RNA was extracted from frozen tissue using the RNEasy Plant Kit (Qiagen) and
123 ribosomal depleted RNA samples were prepared using the RiboMinus Plant Kit (Thermo Fisher
124 Scientific). Stranded libraries were prepared using the NEBNext Ultra Directional RNA Library
125 Kit (New England Biolabs), barcoded, and pooled based on nanomolar concentrations. Library
126 quality and quantity were assessed using the Qubit (dsDNA HS assay; Thermo Fisher Scientific)
127 and BioAnalyzer 2100 (Agilent). The library pool was sequenced across 4 lanes of an Illumina
128 HiSeq2500 for paired-end 101 bp reads at the Genomics Core Lab at Washington State
129 University, Spokane. All new sequence data were deposited in the NCBI Sequence Read Archive
130 (BioProject: PRJNA435759).
131 Single replicate libraries for four of the species in the current study (*A. cettoana, A.*
132 *erecta, A. misera,* and *A. patens*) were previously sequenced for the same two timepoints (Bud
133 and D stages) (Roberts and Roalson 2017). These samples were combined with the newly
134 generated data to assembly de novo transcriptomes and perform co-expression clustering. This
135 data was deposited in the NCBI Sequence Read Archive (BioProject: PRJNA340450).
136
137 **Transcriptome assembly**
138 Per base quality, adapter contamination, and per base sequence content of raw reads was
139 assessed with FastQC tools (http://bioinformatics.babraham.ac.uk/projects/fastqc). One library
140 (AT6) returned poor yields and low-quality reads and was excluded from further analyses.
141 Trimmomatic v.0.36 (Bolger et al. 2014) was then used to remove contaminating adapter
142 sequences and dynamically trim low-quality bases (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10
143 HEADCROP:13 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50). Trimmed
144 reads derived from an individual species were combined and de novo assembled into contigs
145 using Trinity (--SS_lib_type RF; Haas et al. 2013; Grabherr et al. 2011), generating a reference
146 transcriptome for each species. Sequence redundancy was reduced with CDHIT-EST (-c 0.99 -n
147 5; Li and Godzik 2006). Open reading frames were then predicted for each transcriptome
148 assembly with TransDecoder (Haas et al. 2013) using a BLASTp search against the SwissProt
149 database (www.uniprot.org) to maximize sensitivity. Finally, transcriptome assembly
150 completeness was assessed using BUSCO (Simão et al. 2015) against a set of single-copy plant
151 orthologs (embryophyte_odb10 dataset).
152
153 **Orthogroup inference**
154 To help facilitate orthogroup inference, transcriptome assemblies from flowers in
155 *Sinningia eumorpha* and *S. magnifica* were downloaded and combined with our assemblies
156 (https://doi.org/10.5061/dryad.4r5p1; Serrano-Serrano et al. 2017a). *Sinningia* is closely related
157 to our focal lineage in the Neotropical Gesneriaceae (Roalson and Roberts 2016). To identify

158 orthogroup sequences across all 14 species, we used a modified version of the approach
159 described in Yang and Smith (2014). All peptide sequences were clustered using an all-by-all
160 BLASTp search performed using DIAMOND (--evalue 1e-6; Buchfink et al. 2014) and results
161 were clustered using MCL v14-137 (-I 1.4; Enright et al. 2002). Homolog clusters were aligned
162 using MAFFT v7.271 (--genafpair --maxiterate 1000; Katoh and Standley 2013) and alignments
163 were trimmed using Phyutility v2.7.1 (-clean 0.1; Smith and Dunn 2008). Codon alignments
164 were then produced for each cluster using PAL2NAL v14 (Suyama et al. 2006) and maximum-
165 likelihood trees were constructed using FastTree v2.1.8 (Price et al. 2010) using the GTR model.
166 Trees were then trimmed to exclude branches >10 times longer than its sister or longer than 0.2
167 substitutions per site. Final homolog group alignments were created from the trimmed trees using
168 MAFFT and used for final homolog tree inference using RAxML v8.2.9 (Stamatakis 2014) and
169 the GTRGAMMA model. Orthogroups were inferred from the final homolog trees using the
170 rooted ingroup (RT) method of Yang and Smith (2014), using *S. eumorpha* and *S. magnifica* as
171 outgroups. Homolog trees were trimmed to include at least 2 ingroup species, resulting in 83595
172 orthogroups.
173
174 **Quantifying gene expression**
175 Trimmed reads for each sample were aligned to the corresponding transcriptome
176 assembly to quantify expression using default parameters in Kallisto v0.43.0 (Bray et al. 2016).
177 The resulting counts for each contig were matched with the inferred orthogroups derived above,
178 creating two separate matrices for the Bud and D stage samples. Rows corresponded to
179 orthogroups and columns corresponded to individual samples. Counts for the Bud and D stage
180 samples were separately transformed using variance-stabilizing transformation implemented in
181 the Bioconductor package DESeq2 (Love et al. 2014) and quantile normalized using the
182 Bioconductor package preprocessCore (Bolstad 2018). Lastly, counts for both stages were
183 separately corrected for confounding effects, such as batch or species effects, using principal
184 component regression with the 'sva_network' function in the Bioconductor package sva (Parsana
185 et al. 2019).
186
187 **Co-expression networks and identification of modules**
188 The Bud and D stage RNAseq data was clustered into gene co-expression networks using
189 the R package WGCNA (Langfelder and Horvath 2008). Prior to network construction,
190 orthogroups were filtered to remove any with >50% missing values or <0.3 expression variance.
191 In total, 9503 orthogroups common to both Bud and D stage data were used to construct separate
192 networks for each stage. Parameters for the Bud stage network were as follows: power = 18,
193 networkType = "signed", corType = "bicor", maxPOutliers = 0.05, TOMType = "signed",
194 deepSplit = 2, mergeCutHeight = 0.25. For the D stage network, the following parameters were
195 used: power = 16, networkType = "signed", corType = "bicor", maxPOutliers = 0.05, TOMType
196 = "signed", deepSplit = 2, mergeCutHeight = 0.25.

Module eigengenes and orthogroup connectivity were calculated in each network separately using the 'moduleEigengenes' and 'intramodularConnectivity' functions in WGCNA, respectively. The module eigengene is defined as the first principal component of a module and represents the gene expression profile of a sample within a module (Langfelder and Horvath 2008). Connectivity refers to the sum of connection strengths a node has to other network nodes.

Module hub genes (defined as the most highly connected nodes within a module) were identified in both networks based on module membership (kME) scores >0.9, calculated using the 'signedKME' function of WGCNA. kME is defined as the correlation between the expression levels of a gene and the module eigengene (Horvath and Dong 2008). Additionally, using connectivity measures, we defined peripheral genes in each network as the lowest 10% of connected nodes.

We tested whether Bud stage modules were preserved in the D stage and vice versa using module preservation statistics in WGCNA (Langfelder et al. 2011). Median Rank and Zsummary statistics were computed using the 'modulePreservation' function of WGCNA, using 200 permutations (Langfelder et al. 2011). Median rank statistics were compared to the "gold" module, which consists of 1000 randomly selected orthogroups. Modules with a lower median rank exhibit stronger preservation than modules with a higher median rank. Zsummary scores > 10 indicate strong evidence of preservation and scores < 2 indicate no evidence of preservation.

Phylogeny

We previously inferred a phylogeny for the 12 sampled species using 1306 single-copy orthologs identified from the same transcriptome dataset used here (Roberts and Roalson 2018). For comparative analyses of module-trait relationships, we randomly sampled 50 single-copy ortholog gene trees and rescaled branch lengths to be proportional to time (ultrametric) using the 'chronos' function in the R package ape (Paradis et al. 2004). Bootstrap support was 100 for nearly every branch in the Roberts and Roalson (2018) phylogeny, therefore we chose to use randomly sampled ortholog trees to account for phylogenetic uncertainty.

Module and trait relationship

We correlated external traits (e.g., red flower color) with the module eigengenes using a phylogenetic Markov Chain Monte Carlo method, implemented in the R packages MCMCglmm (Hadfield 2010) and mulTree (Guillerme and Healy 2014). These analyses were performed individually for each module in both the Bud and D stage networks. Three floral traits and pollination syndrome were coded: primary flower color (purple, red, white, yellow), flower shape (funnelform, salverform, tubular), corolla spur (absent, present), and syndrome (bee, butterfly, hummingbird). We fit a multivariate mixed model with a non-informative prior, where the shape and scale parameters equal 0.001, and residual variance was fixed. Analyses were performed over the 50 randomly selected ortholog trees to account for phylogenetic uncertainty. For each tree, two MCMC chains were run for 250k generations and discarding the first 50k as burn-in. This process was run individually over each module in both the Bud and D stages. We

checked for convergence between model chains using the Gelman-Rubin statistic (Gelman and Rubin 1992). Potential scale reduction values were all less than 1.1 and effective sample sizes for all fixed effects were greater than 400. We considered fixed effects to be statistically significant when the probabilities in the 95% credible region did not include zero.

Evolutionary rate analysis

Codon alignments of the filtered orthogroups (n=9503) were produced using PAL2NAL v14 (Suyama et al. 2006). Subsequent alignments and corresponding gene trees were used as input to the codeml package in PAML v4.9 (Yang 2007) to estimate d_N/d_S (omega, ω). d_N/d_S for each orthogroup was estimated using a one-ratio model (model = 0, NSsites = 0), providing a single estimate for each orthogroup to match other single orthogroup metrics, such as connectivity. If d_N/d_S values were exceptionally large (= 999) because of zero synonymous differences, those values were removed from the dataset.

First, we performed linear regression (LM) to test the effect of orthogroup connectivity, average expression levels, and their interactions (explanatory variables) on the estimated d_N/d_S values (main effect), LM = d_N/d_S ~ connectivity * expression. This was performed using the lm function in R with 10000 bootstrap pseudoreplicates. Second, we ran two-sample t-tests with 10000 permutations to test whether hub nodes and periphery nodes in each network had lower (for hubs) or higher (for periphery) d_N/d_S values than all background nodes. Third, we fit a LM with 10000 bootstrap pseudoreplicates to test whether modules associated with a pollination syndrome (bee, butterfly, or hummingbird; see Results below) had increased d_N/d_S values, LM = d_N/d_S ~ syndrome. For all LM and permutation analyses, d_N/d_S, connectivity, and expression values were log transformed to reach a normal distribution before being processed. Analyses were performed separately for each network using the separately estimated connectivity and expression levels.

Functional annotation and GO enrichment analysis

Filtered orthogroups (n=9503) were annotated based on a search against all green plant proteins in the SwissProt database (release 2018_07; The UniProt Consortium 2017) using Diamond BLASTp (--evalue 1e-6). Results were used to assign annotations and gene ontology (GO) terms. GO enrichment analyses were performed using the R package TopGO with the default "weight01" algorithm with the recommended cutoff of $p < 0.05$ (Alexa et al. 2006). Enriched GO term redundancy was removed and summarized using the semantic similarity measures implemented in the REVIGO web server (Supek et al. 2011).

Results

Transcriptome assembly

Twelve transcriptomes were sequenced (72 libraries) and assembled in 12 species after sampling across two floral developmental stages (Supplemental Table 1). The transcriptomes had

277 an average assembly length of 244.9 Mb (± 33.9 Mb), an average of 217,510 (± 47k) transcripts,
278 and an average N50 length of 1988 bp (± 306 bp). Between 22,942 and 33,501 putative genes
279 were detected with 2.34 (± 0.35) putative isoforms. BUSCO identified an average of 87% of the
280 complete, conserved single-copy plant orthologs (Supplemental Table 1). De novo assembly
281 enables functional genomic studies in non-model systems without a reference genome, but has
282 the potential for misassemblies (Hornett and Wheat 2012) and subsequent downstream biases.
283 Our focus in our comparative study was on the presence of conserved orthologs, and not on the
284 presence of taxonomically restricted genes, which are not likely preserved across species. We
285 subsequently identified 83595 orthologous clusters (orthogroups), each containing at least 2
286 species.
287
288 **Constructing the co-expression network for 12 species**
289 After removing 74092 orthogroups with too many missing values (>50% missing) or low
290 expression variance (<0.3 variance) and keeping common orthogroups shared between the Bud
291 and D stages, 9503 orthogroups were used to construct two co-expression networks using the
292 weighted gene correlation network analysis (WGCNA) approach. Modules of co-expressed
293 orthogroups are inferred using the expression profiles of each sample regardless of species.
294 Clustering identified 65 and 62 co-expression modules in the Bud and D stage networks,
295 respectively (Figure 2). Module size in the Bud stage ranged from 22 ('mediumorchid') to 592
296 ('turquoise') orthogroups (mean = 146) and ranged from 21 ('antiquewhite4') to 704
297 ('turquoise') orthogroups (mean = 153) in the D stage. Out of the 9503 orthogroups, 7845 (83%)
298 and 8735 (92%) were assigned to modules in each network.
299 The remaining unassigned genes from each network that were not placed into a well-
300 defined module were assigned to the 'grey' module by WGCNA. Despite being unplaced, these
301 orthogroups may still be functionally relevant within each network. Unassigned genes in the
302 'grey' module were enriched for many processes, including transcription (Bud, n = 173; D, n =
303 84) and organ development (D, n = 55), among other functions (Supplemental Tables 2-3).
304 We tested whether modules in the Bud stage were preserved in the D stage, and vice
305 versa. Overall, preservation of modules between the Bud and D stage networks was high. Eight
306 of the 65 Bud stage modules ('brown4', 'coral2', 'floralwhite', 'lavenderblush3', 'lightcyan1',
307 'lightsteelblue1', 'maroon', 'mediumpurple3') were not strongly preserved in the D stage, while
308 only 1 of the 62 D stage modules ('darkseagreen4') was not strongly preserved in the Bud stage
309 (Supplemental Table 4). These patterns suggest that these non-preserved modules may play
310 unique roles within their network. Among the 8 Bud stage-specific modules were many
311 orthogroups related to primary cell growth, meristem identity and differentiation, stamen
312 development, and flowering (Supplemental Table 2). In the single D stage-specific module were
313 many orthogroups related to methylation, flowering, and carotenoid biosynthesis (Supplemental
314 Table 3). Despite the relatively high preservation, correspondence between module membership
315 in the Bud and D stages was moderate (Figure 2). Orthogroup module membership tended to

316 shift between each network. For example, most members of module 'midnightblue' in the Bud
317 stage are transferred to modules 'brown' and 'black' in the D stage (Figure 2).
318 We defined network hubs (the most highly connected nodes within a module) with
319 module membership (kME) scores > 0.9. Using this definition, all modules in each network
320 contained at least one hub, with the percentage of hubs ranging from 1% to 24% of orthogroups
321 in a module (Figures 3 and 4; Supplemental Table 5). Of the 632 and 691 hubs identified in the
322 Bud and D stages, a significant portion (110) were shared between networks (Fisher's Exact
323 Test, odds ratio = 3.02, P < 0.001). Functional enrichment of hubs in the Bud stage revealed that
324 many were involved in the regulation of development and metabolic processes (n = 40), transport
325 and protein localization (n = 39), biosynthetic processes (n = 19), and DNA methylation (n = 7),
326 consistent with the notion that hubs usually play important roles in integrating other genes during
327 network evolution (Ravasz et al. 2002) (Table 1). In the D stage, hubs were enriched for
328 functions related to translation (n = 33), transport (n = 9), ethylene and abscisic acid signaling
329 pathways (n = 17), carotenoid and anthocyanin biosynthetic processes (n = 11), and the
330 regulation of gene expression (n = 20), among others (Table 1).
331 Nodes in the periphery of each network were loosely defined by considering the lowest
332 10% connected ($connectivity_{Bud}$ < 2.824, $connectivity_{D}$ < 4.651). These definitions identified 951
333 periphery nodes (Supplemental Table 6), of which 301 were peripheral in both networks
334 (Fisher's Exact Test, odds ratio = 5.67, P < 0.001). These periphery nodes in the Bud stage were
335 enriched for functions related to reproductive structure development (n = 75), transport and
336 localization (n = 107), and signaling (n = 389), among others (Table 2). In the D stage, the
337 periphery was enriched for functions related to the regulation of flower development (n = 64),
338 the regulation of flowering time (n = 18), the regulation of flavonoid biosynthesis (n = 7), and
339 cell fate specification (n = 12), among others (Table 2). Greater than expected numbers of nodes
340 in the D stage were related to regulation (Fisher's exact test, odds ratio = 1.191, p = 0.02752) and
341 transcription (Fisher's exact test, odds ratio = 1.370, p = 0.001549). In contrast, there were no
342 differences in the expected number of nodes related to regulation (Fisher's exact test, odds ratio
343 = 0.955, p = 0.7044) or transcription (Fisher's exact test, odds ratio = 1.047, p = 0.3549) in the
344 Bud stage.
345
346 **Modules associated with floral traits**
347 We calculated module eigengenes, single values which represent the gene expression
348 profiles of a sample within a module. The extent of module involvement in various biological
349 processes can be tested by correlating eigengenes with external traits, such as flower color. We
350 tested the relationship between the module eigengenes and the four floral traits (pollination
351 syndrome [bee, butterfly, hummingbird], flower color [purple, red, yellow, white], flower shape
352 [funnelform, salverform, tubular], and corolla spurs [presence]), while controlling for any
353 phylogenetic bias in the dataset using a phylogenetic mixed model (Figures 3-4; Supplemental
354 Figures 3-6). We also examined whether eigengene values (> 0.1 or < -0.1) in single species
355 contributed to a strong correlation or whether eigengene values (> 0.1 or < -0.1) in multiple

species contributed to a strong correlation (Figures 3-4; Supplemental Figures 3-6). We considered strong evidence for trait association when a significant correlation from MCMCglmm was coupled with elevated eigengene expression in multiple species. WGCNA does not use trait information when inferring modules, so modules that correlated strongly with any floral trait were inferred without a priori knowledge.

Pollination syndrome—We included 3 species with the bee pollination syndrome (*A. candida, A. misera,* and *E. verticillata*), 4 species with the butterfly pollination syndrome (*A. cettoana, A. grandiflora, A. longiflora,* and *A. patens* 'Major'), and 5 species with the hummingbird pollination syndrome (*A. admirabilis, A. antirrhina, A. erecta, A. pedunculata,* and *G. cuneifolia*) (Figure 1). These pollination syndromes are traditionally defined in *Achimenes* and other gesneriads based on primary flower color, flower shape, and the presence of corolla spurs (Ramírez Roa 1987). Expression of 22 and 28 modules were positively associated with the pollination syndromes in the Bud and D stage networks, respectively (Figures 3-4). There were 3, 7, and 13 modules associated with bee, butterfly, and hummingbird pollination syndrome in the Bud stage network, while 2, 8, and 12 modules were associated with the same syndromes in the D stage. Modules were never associated with more than one syndrome. Butterfly and hummingbird syndromes were never correlated to the same modules and were often correlated in opposite directions (i.e., butterfly was positive correlated and hummingbird was negative correlated).

Flower color—Primary flower color is closely associated to pollination syndrome in *Achimenes,* and we identified 26 and 14 modules associated to any flower color in the Bud and D stage networks, respectively (Figures 3-4). Red and yellow are associated with hummingbird pollination, purple with butterfly pollination, and white with bee pollination. In the Bud stage, there were 14, 11, and 1 module associated with red, purple, or yellow flower color. In the D stage, no modules were associated with yellow, but 8 and 6 modules were associated with red or purple flower color. In both networks, no modules were corelated to white flower color. Similar to syndrome, purple and red flower color were never associated with the same modules and were always correlated in opposite directions.

Flower shape—Flower shape is also important but not as closely associated to pollination syndromes as primary flower color in *Achimenes*. Flowers with bee pollination tend to have funnelform flowers, butterfly pollinated flowers all have salverform flowers, and hummingbird pollinated flowers have either salverform or tubular flowers. Here, there were 1, 8, and 5 modules whose expression correlated to funnelform, salverform, or tubular flowers in the Bud stage (Figure 3). Fewer modules were correlated in the D stage network, but expression in 5 and 1 module were associated with salverform and tubular shapes (Figure 4). No modules were positively associated with more than one flower shape.

Corolla spurs—The presence of corolla spurs is found only on butterfly pollinated species, two of which were included: *A. grandiflora* and *A. patens*. There was expression in 5 modules associated with spurs in the Bud stage network, while only 1 module was associated in the D stage network (Figures 3-4).

Evolutionary rates correlate to network location

We ran several analyses to explore how several network variables may affect evolutionary rates of the orthogroups within each network (as measured by d_N/d_S). First, we tested the effects of orthogroup connectivity, expression levels, and their interaction on the d_N/d_S values (Supplemental Table 7). Results indicated a collective effect between connectivity and expression on d_N/d_S in both the Bud stage (LM, $F(3,9548) = 92.58$, p = 0, $\eta^2 = 0.03$) and the D stage networks (LM, $F(3,9548) = 78.36$, p = 0, $\eta^2 = 0.02$). We found that orthogroup connectivity was negatively correlated with d_N/d_S in the Bud stage network (LM, t = -16.31, p = 0) and the D stage network (LM, t = -15.13, p = 0) (Figure 5), while expression was positively correlated with d_N/d_S in the Bud stage (LM, t = 2.12, p = 0.0344) and not in the D stage (LM, t = 1.51, p = 0.1324) (Supplemental Figure 7). The interaction between connectivity and expression also indicated a marginal dependency of both predictors on d_N/d_S in the Bud stage (LM, t = -2.022, p = 0.0433) but not in the D stage network (LM, t = -1.823, p = 0.0683). These patterns have been observed in model species and may be universal to all eukaryotes.

Second, we tested whether the nodes we defined as hubs in each network had lower d_N/d_S values, suggestive of increased evolutionary constraint (Supplemental Table 7). Hubs are considered to functionally significant and may control large portions of the network. We found that the hubs in the Bud stage network had lower d_N/d_S than all background genes (Two-sample t-test, $t(679.31) = 7.0934$, p = 1.65e-12, Cohen's d = 0.383; Permutation t-test. mean diff. = 0.266, p = 0.0001) (Supplemental Figure 8). Similarly, hubs in the D stage network also had lower d_N/d_S than all background genes (Two-sample t-test, $t(755.91) = 4.1918$, p = 1.55e-05, Cohen's d = 0.206; Permutation t-test, mean diff. = 0.143, p = 0.0001) (Supplemental Figure 8).

Third, we tested whether the most peripheral nodes (the lowest 10% connected nodes) had higher d_N/d_S values, which may suggest relaxed evolutionary constraint (Supplemental Table 7). These nodes are loosely connected within the network and their roles may be more apt to fluctuate during evolution and development. These peripheral genes had higher d_N/d_S than all background genes in the Bud stage network (Two-sample t-test, $t(1487.70) = -10.272$, p = 0, Cohen's d = -0.253; Permutation t-test, mean diff. = -0.176, p = 0.0001) and the D stage network (Two-sample, $t(1512.30) = -10.387$, p = 0, Cohen's d = -0.252; Permutation t-test, mean diff. = -0.175, p = 0.0001) (Supplemental Figure 8).

Lastly, we tested whether modules correlated with bee, butterfly, or hummingbird pollination syndromes showed increased evolutionary rates over the non-associated modules. Results indicated a very weak effect of syndrome association on d_N/d_S in the Bud network (LM, $F(3,9548) = 5.186$, p = 0.001406, $\eta^2 = 0.0016$) and the D network (LM, $F(3,9548) = 4.277$, p = 0.005038, $\eta^2 = 0.0013$) (Supplemental Table 7). Only bee syndrome associated modules had marginally lower d_N/d_S values in both networks (Bud stage, LM, t = -3.58, p = 0.000346; D stage, LM, t = -2.508, p = 0.0122). Butterfly associated modules had marginally lower d_N/d_S only in the D stage (LM, t = -2.548, p = 0.0108), while hummingbird associated modules did not have any effect in either network.

436

437 ## Discussion

438 Here we report on patterns of gene co-expression across two stages of flower
439 development in 12 species of *Achimenes, Eucodonia,* and *Gesneria*. Our comparative and
440 phylogenetic context allowed us to reveal gene co-expression patterns across a group of closely
441 related plant species that correlated to diverse pollination syndromes and floral traits, including
442 flower color, flower shape, and corolla spurs. Our results suggest floral forms that are associated
443 with bee, butterfly, and hummingbird pollinator groups are correlated with many co-expression
444 modules. We found that nearly a third of modules in each network (23 and 21 modules in the
445 Bud and D stages, respectively) had evidence for association with the three pollination
446 syndromes (Figures 3-4). In our analysis, large fractions of our 9503 orthogroups set could be
447 partitioned into modules, network hubs and peripheral nodes identified, and candidate
448 orthogroups were identified for their potential importance in floral phenotypic differentiation.
449 Lastly, orthogroup position in each network (measured by connectivity) had an association with
450 evolutionary rate (d_N/d_S) (Figure 5). These analyses provide a first overview of floral
451 transcriptional architecture across members of the same genus that display diverse floral forms.
452

453 **Hub nodes contain important candidates for floral form**

454 To date, a large number of studies conducted in diverse angiosperm systems, have found
455 that relatively few genes are associated with floral form differentiation (e.g., Quattrocchio et al.
456 1999; Howarth et al. 2011). Despite evidence for extensive gene/genome duplications that could
457 expand the genetic repertoire (e.g., Tang et al. 2008), many recent studies have suggested that the
458 generation of novel genes is rare (Patterson et al. 2009; De Smet et al. 2013). Thus, the evolution
459 and development of diverse floral forms may predominantly occur through the co-option of
460 existing pathways (Preston et al. 2011). To our knowledge, gene co-expression network analyses
461 have not been previously used across multiple species in a non-model lineage to understand
462 flower development and evolution.

463 Hub nodes (genes) are considered to be conserved, highly connected nodes that are
464 central to network architecture and may directly orchestrate numerous biological processes
465 (Ravasz et al. 2002; Hu et al. 2016; Mähler et al. 2017). Our analyses identified over 600 gene
466 nodes as hubs within each network, the majority of which were only hubs within a specific floral
467 stage. These hub nodes tended to have lower d_N/d_S estimates, suggesting that they not prone to
468 rapid evolutionary change (Supplemental Table 7). Unsurprisingly, hub nodes in the Bud stage
469 were enriched for many processes related to early flower development, including cell growth,
470 leaf formation, and photosystem assembly (Table 1). Similarly, hub nodes in the D stage were
471 enriched for genes that may be related the development of distinct floral forms, including
472 carotenoid biosynthesis, anthocyanin biosynthesis, and cell morphogenesis (Table 1). Within this
473 context of flower development, we identified a number of hubs that we consider important
474 candidates for involvement in the production of flower color, the development of floral form, and
475 the production of corolla spurs. Many of these candidates were considered previously (Roberts

476 and Roalson 2017) as potentially important for the diversification of floral form in *Achimenes*.
477 Undoubtedly, this list of hubs contains many genes that may important for other biological traits
478 that we do not consider here, such as scent, nectar, and pollen production.
479 Flower color in *Achimenes* is primarily due to the accumulation of anthocyanin and
480 carotenoid pigments (Roberts and Roalson 2017). We identified two members of the core
481 anthocyanin pathway as hubs. Homologs of flavonoid 3'-hydroxylase (*F3'H*;
482 cluster8703_1rr.include1.ortho3; module 'tan') and dihydroflavonol 4-reductase *(DFR*;
483 cluster5082_1rr.include1.ortho6; module 'darkslateblue'*)* were identified as hubs in the Bud and
484 D stages, respectively (Supplemental Table 8). *F3'H* is involved in the shifting metabolic flux of
485 the pathway between production of red pelargonidins and purple cyanidins while *DFR* acts
486 downstream of *F3'H* and directs the production of anthocyanins. Shifting expression of *F3'H*
487 and shifting substrate specificity of *DFR* have been identified in many systems as important
488 contributors to transitions between red and blue flowers (e.g., Des Marais and Rausher 2010;
489 Smith and Rausher 2011). We also identified two homologs of carotenoid biosynthetic pathway
490 genes as hubs in the D stage: homologs of geranylgeranyl pyrophosphate synthase (*GGPPS*;
491 cluster10945_2rr.include1.ortho1; module 'sienna3') and lycopene beta-cyclase (*LCYB*;
492 cluster13774_1rr.include1.ortho1; module 'black'). *GGPPS* is involved in the production of
493 precursor molecules while *LCYB* directs production of either alpha-carotenes or beta-carotenes,
494 the latter of which is thought to be the predominant carotenoids found in *Achimenes* flowers
495 (Roberts and Roalson 2017).
496 Less is known about the genetic pathways giving rise to the diverse flower shapes in
497 *Achimenes*. Among the hubs found in the Bud stage were homologs of many proteins involved in
498 floral development and cell differentiation, including the floral homeotic protein agamous (*AG*;
499 cluster7311_1rr.include1.ortho3; module 'darkturquoise'), phantastica (*PHAN* or rough sheath 2
500 [*RS2*]; cluster11220_1rr.include1.ortho3; module 'yellow'), and tcp4 (*TCP4*;
501 cluster12380_1rr.include1.ortho2; module 'darkgreen'). *AG* plays a role in maintaining floral
502 meristem determinancy and is required for normal development of stamens and carpels (Gómez-
503 Mena et al. 2005). *PHAN* is involved in cell differentiation and adaxial identity by repressing
504 numerous knox homeobox genes (Waites et al. 1998). *TCP4* is a cell division regulator that was
505 identified as a candidate for corolla spur development in columbine (Yant et al. 2015) and
506 showed increased expression during later stages of floral development in *Achimenes patens*
507 (Roberts and Roalson 2017). Finally, among the hubs in the D stage network was a homolog of
508 the floral homeotic protein globosa (*GLO*; cluster2750_1rr.include1.ortho3; module 'skyblue3'),
509 which acts in conjunction with deficiens (*DEFA*) to specify petal identity (Tröbner et al. 1990).
510
511 **Peripheral genes are enriched for transcriptional regulators in later stages of development**
512 Peripheral nodes have low connectivity within co-expression networks (Supplemental
513 Table 7), tend to have higher rates of molecular evolution (e.g., Masalia et al. 2017), and have
514 been hypothesized to contribute more to evolutionary innovations than network hubs (e.g.,
515 Ichihashi et al. 2014). Of the 951 nodes we defined as peripheral in both Bud and D stage

516 networks, roughly a third were consistently peripheral under our definitions (lowest 10%
517 connected; Supplemental Table 6). After identifying the nodes with homologies to transcriptional
518 regulators, we found there was no difference in the expected number in the early Bud stage,
519 while there were greater than expected numbers of transcriptional regulators in the periphery of
520 the D stage (see Results). No such patterns existed when we tested these transcriptional
521 regulators as hub genes. While Mähler et al. (2017) found an enrichment for transcription factors
522 among hub genes in a single species network of *Populus tremula,* our results in a multi-species
523 network find the opposite pattern. Shifting the expression patterns of various genes and pathways
524 can lead to phenotypic divergence among closely related species (West-Eberhard 1989). This can
525 be seen during later stages of flower development when distinct traits, such as red or purple
526 flower color, can appear in different species due to shifting expression patterns within the
527 anthocyanin biosynthetic pathway (Smith and Rausher 2011). The connection between the
528 apparent overabundance of transcriptional regulators in later stages of flower development and
529 the rapid phenotypic divergence in floral form will need to be explored further.
530 Some notable peripheral nodes included homologs of divaricata (*DIV*), dichotoma
531 (*DICH*), and wuschel (*WUS*) (Supplemental Table 6). Both *DIV*
532 (cluster13684_1rr.include1.ortho1, module 'grey) and *DICH* (cluster13245_1rr.include1.ortho1;
533 module 'grey') are involved in the dorsoventral asymmetry of flowers (Almeida et al. 1997; Luo
534 et al. 1999) and were only peripheral in the Bud stage. Flowers at this developmental stage are
535 lacking or only just beginning to establish asymmetry and the relegation of both *DIV* and *DICH*
536 to the periphery may reflect this small role. Meanwhile, the homolog of *WUS*
537 (cluster9273_1rr.include1.ortho1; module 'grey') was only peripheral during the D stage. *WUS* is
538 an important transcription factor that plays an important role to maintain meristem identity and
539 participates in the vegetative to reproductive transition (Lohmann et al. 2001; Zuo et al. 2002).
540 The relegation of *WUS* to the periphery at later developmental stages might be expected given
541 that it plays such a prominent role during early floral development.
542
543 **Evolutionary rates are correlated to network connectivity**
544 Increased protein evolutionary rates during rapid diversification has been suggested in
545 other radiations of animals and plants (e.g., Kapralov et al. 2013; Brawand et al. 2014; Pease et
546 al. 2016). Our study is among the first to examine patterns of network evolution across multiple
547 closely-related species of plants with multiple transitions in floral form and pollination
548 syndrome. Hummingbirds are hypothesized to have had an important influence on diversification
549 in Neotropical Gesneriaceae (Roalson and Roberts 2016; Serrano-Serrano et al. 2017b), while
550 the role of butterflies has not been explored. Modules correlated with hummingbird or butterfly
551 pollination syndromes did not exhibit clear increases or decreases in evolutionary rates compared
552 to other modules (Supplemental Table 7). Detecting selection is subject to many factors, such as
553 a gene's transcriptional abundance and its importance within the protein interaction network
554 (Lemos et al. 2005). For instance, we found that higher network connectivity was strongly
555 associated with lower evolutionary rates across both networks in our study (Figure 5;

556 Supplemental Table 7). This corroborates the recent results found in other eukaryotic systems
557 (Morandin et al. 2016; Masalia et al. 2017; Josephs et al. 2017; Mähler et al. 2017) while also
558 showing the apparent strength of correlation between connectivity and evolutionary rate differs
559 between studies and systems.
560 Network effects (such as the type of network analysis employed) and the expression
561 levels of highly abundant genes are known to affect evolutionary rates, potentially confounding
562 analysis of selection pressure (Krylov et al. 2003). We found that d_N/d_S in early development may
563 be dependent on both connectivity and expression levels, while this interactive effect was gone
564 during later development (Supplemental Table 7). As there are few highly connected genes
565 (hubs, which are important determinants of the observed network structure), a random mutation
566 would be less likely to affect such a gene. Changing the amino acid sequence of a hub gene
567 could have multiple associated effects, some deleterious (Hahn and Kern 2004; Luisi et al.
568 2015). In comparison, a random mutation would be more likely to affect a gene with lower
569 connectivity, while also resulting in variants with fewer negative consequences. The hub genes
570 have more direct interactions and are more likely to be essential and involved in more biological
571 processes, placing them under greater constraint.
572
573 **Numerous network modules correlated to floral form**
574 One primary use for gene co-expression network analyses is to identify modules of co-
575 expressed genes whose overall expression (as measured by the eigengene) may correlate with
576 different traits. These traits can be both qualitative or quantitative, such as social caste (Morandin
577 et al. 2016), seed oil production (Hu et al. 2016), or anther development (Hollender et al. 2014).
578 Since our dataset was composed of samples from 12 species, we performed correlation analyses
579 using a phylogenetic model to test whether module expression correlated to different flower
580 colors, flower shapes, corolla spurs, and pollination syndromes. Nearly half of our modules in
581 both networks had evidence for an association with these floral traits (Figures 3-4). There were
582 no modules whose correlation to a pollination syndrome had increased eigengene expression in
583 all species sharing that trait (Figures 3-4). Instead many modules with correlation to a trait
584 showed a phylogenetic pattern with a shared increased eigengene expression among the most
585 closely related species. For example, module 'darkturquoise' in the Bud stage network had a
586 correlation to butterfly pollination and increased eigengene expression in *A. cettoana* and *A.*
587 *longiflora,* both members of Clade 1 (Figure 1). Our aim was to identify sets of co-expressed
588 genes that would provide candidates for involvement in the development of these traits.
589 For instance, many of the candidate hubs we identified (see above) were found in
590 modules that were correlated to one or multiple traits to which they might contribute, such as
591 *PHAN* (module 'yellow', correlated to corolla spurs) and *DFR* (module 'darkslateblue',
592 correlated to hummingbird pollination and purple flowers). Additionally, modules correlated to
593 various floral traits were enriched for multiple genes that might be involved in the development
594 of those traits. Module 'palevioletred3' was correlated to corolla spurs in the D stage network
595 and was enriched for many genes involved in cell division and cell growth (Figure 4;

596 Supplemental Table 3). Other times, modules were correlated to a specific trait but the enriched
597 GO terms did not provide a clear idea of how these co-expressed orthogroups may contribute
598 (Supplemental Tables 2-3). In our analyses, we considered relatively few traits and only focused
599 on those that have traditionally been considered the most important for the differentiation of
600 floral form in *Achimenes*, namely flower color, flower shape, and corolla spurs. Undoubtedly
601 there are a myriad number of molecular and biochemical pathways that we have not considered
602 here that underlie the development of these complex floral phenotypes. With the idea that
603 networks represent complex systems of gene and protein interactions, it may be that these
604 seemingly simple and generalized floral phenotypes are produced through much more complex
605 means than many have previously considered.
606
607 **Strengths and limitations of network analyses in non-model organisms**
608 Our analysis covers genes with orthologs found in most species, representing a set of
609 evolutionarily conserved genes. While no reference genome currently exists for Neotropical
610 Gesneriaceae, we constructed, annotated, and presented de novo transcriptomes for 10
611 *Achimenes* species as well as 2 outgroup species, *E. verticillata* and *G. cuneifolia*. Our approach
612 enables functional genomic studies in this non-model system, but the potential for misassemblies
613 may bias our downstream results (Hornett and Wheat 2012). Using a gene tree-based approach to
614 carefully identify orthogroups, we focused our attention on conserved gene regulatory machinery
615 and not on taxonomically restricted genes. The latter may be involved in species-specific
616 functions. Nevertheless, it is worth noting that many recent studies have indicated that transitions
617 in floral form may be due to the shifting expression of genes found in conserved genetic
618 pathways (e.g., Wessinger and Rausher 2014). Taxonomically restricted genes are likely
619 interacting with existing regulatory pathways, and the manner in which they integrate into the
620 co-expression modules identified here will be a fascinating topic for future research. With
621 extended sampling within *Achimenes* we could infer species-specific co-expression networks and
622 perform more detailed cross-species comparisons. We expect that these taxonomically restricted
623 genes may be poorly connected with the network, which may allow for increased diversification.
624 We characterized gene expression patterns by sequencing RNA across two timepoints
625 during flower development. Previous gene expression experiments across three timepoints (early
626 Bud, intermediate D, and late Pre-anthesis stages) in four *Achimenes* species suggested that the
627 majority of differential gene expression occurred between the early Bud stage and the
628 intermediate D stage (Roberts and Roalson 2017). We believe our approach allows us to capture
629 dynamic transcriptional patterns during flower development for comparison across multiple
630 species. We found that the co-expression networks for the Bud and D stages were largely
631 preserved, despite there being a few stage specific modules (Supplemental Table 4). However,
632 this approach may greatly over-simplify the nature of transcriptional regulation in these species.
633 Organ- and structure-specific expression studies may have greater power to detect floral form-
634 specific differences, both in terms of the number of differentially expressed genes, and co-
635 expression network structure (Hollender et al. 2014; Suzuki et al. 2017; Shahan et al. 2018). As a

result, our data likely represents an underestimate of the true number of modules. Future studies could focus on the comparative analysis of expression patterns in specific organs, such as the petals, stigma, and stamens, and additional timepoints, which would provide greater functional insight into the evolution, development, and diversification of these organs.

Network analysis represents a more complex approach than differential expression analyses because it can capture system-level properties (Langfelder and Horvath 2008). Most phenotypes involve interactions of proteins from diverse biochemical pathways. Although inferred modules do not necessarily correspond to entire biochemical pathways, or other components of cellular organization, the approach performs well in reconstructing the complexity of protein-protein interaction networks (Allen et al. 2012). In many systems this approach has succeeded in identifying candidate genes for various biological traits, for example floral organ development in strawberry (Hollender et al. 2014; Suzuki et al. 2018) or carotenoid accumulation in carrots (Iorizzo et al. 2016). We were able to identify numerous candidate orthogroups involved in the development of floral form in our non-model system. While no functional genomic data exists yet for *Achimenes,* our annotations are based on the manually curated SwissProt database. Applying this network-based approach to the diversification of floral form across multiple species in *Achimenes*, we recovered distinct modules of co-expressed genes that may underlie a range of floral phenotypic traits tied to different pollination syndromes. Although our data come from whole flowers during development, and do not allow us to test specific hypotheses regarding whether particular genes have shifted expression location in the developing flower, they do suggest that there are conserved regulatory modules that may have been co-opted in convergent floral forms multiple times. It will be interesting to apply network analyses to study floral diversification in additional lineages and examine how conserved or divergent the patterns are across angiosperms.

Conclusions

Floral forms corresponding to bee, butterfly, and hummingbird pollination have evolved multiple times across the Neotropical plant genus *Achimenes*. Genome-wide gene expression estimates were taken from flowers in 12 species across two development stages in order to construct, analyze, and compare stage-specific gene co-expression networks. We hypothesized that numerous modules in each network would correlate to these pollination syndromes and that central genes in the network may be candidates for involvement in the development of important floral traits, such as flower color and shape. We found that nearly a third of modules were correlated to the pollination syndromes and many more were correlated to different flower colors and shapes. Several of the hub genes in the networks were homologs of the anthocyanin and carotenoid biosynthetic pathways, important for the production of floral pigments that attract different pollinators. A negative relationship between network connectivity and d_N/d_S corroborates the findings in model systems that more centrally located nodes (likely hubs) are under increased evolutionary constraint. We found that the less connected genes (peripheral) are under more relaxed constraints and contained numerous transcriptional regulators. Our results demonstrate

676 the utility of applying co-expression network analyses in non-model plant lineages to begin
677 identifying important modules and pathways that will be useful starting points for additional
678 analyses into the evolution and development of floral form.
679

680 ## Acknowledgements

681 We thank Joanna Kelley, Amit Dhingra, and Andrew McCubbin for helpful comments that
682 improved this work. Sequencing was performed at the Genomics Core Lab at Washington State
683 University, Spokane, and the Genomics Sequencing and Analysis Facility at the University of
684 Texas, Austin.
685

686 ## References

687 Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene
688 expression data by decorrelating GO graph structure. Bioinformatics 22:1600-1607.
689 Alexandre H, Vrignaud J, Mangin B, Joly S. 2015. Genetic architecture of pollination syndrome
690 transition between hummingbird-specialist and generalist species in the genus
691 Rhytidophyllum (Gesneriaceae). PeerJ 3:e1028.
692 Allen JD, Xie Y, Chen M, Girard L, Xiao G. 2012. Computing statistical methods for
693 constructing large scale gene networks. PLOS One 7:e29348.
694 Almeida J, Rocheta M, Galego L. 1997. Genetic control of flower shape in Antirrhinum majus.
695 Development 124:1387-1392.
696 Barrett SC. 2013. The evolution of plant reproductive systems: how often are transitions
697 irreversible? P Roy Soc B-Biol Sci 280: 20130913.
698 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
699 data. Bioinformatics 30:2114-2120.
700 Bolstad B. 2018. preprocessCore: A collection of pre-processing functions. R package version
701 1.46.0, https://github.com/bmbolstad/preprocessCore.
702 Bradshaw HD, Otto KG, Frewen BE, McKay JK, Schemske DW. 1998. Quantitative trait loci
703 affecting differences in floral morphology between two species of monkeyflower
704 (Mimulus). Genetics 149:367-382.
705 Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A,
706 Aximu-Petri A, Kircher M, Albert FW. 2011. The evolution of gene expression levels in
707 mammalian organs. Nature 478:343.
708 Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW,
709 Bezault E, Turner-Maier J. 2014. The genomic substrate for adaptive radiation in African
710 cichlid fish. Nature 513:375.
711 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq
712 quantification. Nat Biotechnol 34:525.
713 Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND.
714 Nat Methods 12:59-60.

715 Des Marais DL, Rausher MD. 2010. Parallel evolution at multiple levels in the origin of
716 hummingbird pollination flowers in Ipomoea. Evolution 64:2044-2054.
717 De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y. 2013.
718 Convergent gene loss following gene and genome duplications creates single-copy
719 families in flowering plants. Proc Natl Acad Sci USA 110:2898-2903.
720 Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection
721 of protein families. Nucleic Acids Res 30:1575-1584.
722 Faegri K, van der Pijl L. 1979. *Principles of Pollination Ecology*. Oxford: Pergamon Press.
723 Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD. 2004. Pollination syndromes
724 and floral specialization. Annu Rev Ecol Evol S 35:375-403.
725 Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. Stat
726 Sci 7:457-472.
727 Gervasi DD, Schiestl FP. 2017. Real-time divergent evolution in plants driven by pollinators.
728 Nat Commun 8: 14691.
729 Gómez-Mena C, De Folter S, Costa MMR, Angenent GC, Sablowski R. 2005. Transcriptional
730 program controlled by the floral homeotic gene AGAMOUS during early organogenesis.
731 Development 132:429-438.
732 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
733 Raychowdhury R, Zeng Q, Chen Z. 2011. Trinity: reconstructing a full-length
734 transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29:644.
735 Guillerme T, Healy K. 2014. mulTree: a package for running MCMCglmm analysis on multiple
736 trees. ZENODO. 10.5281/zenodo.12902.
737 Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic
738 protein-interaction networks. Mol Biol Evol 22:803-806.
739 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,
740 Li B, Lieber M, MacManes MD. 2013. De novo transcript sequence reconstruction from
741 RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc
742 8:1494.
743 Hadfield JD. 2010. MCMC methods for multi-response generalized linear mixed models: the
744 MCMCglmm R pckage. J Stat Softw 33:1-22.
745 Hermann K, Klahre U, Venail J, Brandenburg A, Kuhlemeier C. 2015. The genetics of
746 reproductive organ morphology in two Petunia species with contrasting pollination
747 syndromes. Planta 241:1241-1254.
748 Hoballah ME, Gübitz T, Stuurman J, Broger L, Barone M, Mandel T, Dell'Olivo A, Arnold M,
749 Kuhlemeier C. 2007. Single gene-mediated shift in pollinator attraction in Petunia. Plant
750 Cell 19: 779-790.
751 Hollender CA, Kang C, Darwish O, Geretz A, Matthews BF, Slovin J, Alkharouf N, Liu Z. 2014.
752 Floral transcriptomes in woodland strawberry uncover developing receptacle and anther
753 gene networks. Plant Physiol 165:1062-1075.

754 Hornett EA, Wheat CW. 2012. Quantitative RNA-Seq analysis in non-model species: assessing
755 transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic
756 reference species. BMC Genomics 13:361.
757 Horvath S, Dong J. 2008. Geometric interpretation of gene co-expression network analysis.
758 PLOS Comput Biol 4:e1000117.
759 Howarth DG, Martins T, Chimney E, Donoghue MJ. 2011. Diversification of CYCLOIDEA
760 expression in the evolution of bilateral flower symmetry in Caprifoliaceae and *Lonicera*
761 (Dipsacales). Ann Bot-London 107:1521-1532.
762 Hu G, Hovav R, Grover CE, Faigenboim-Doron A, Kadmon N, Page JT, Udall JA, Wendel JF.
763 2016. Evolutionary conservation and divergence of gene coexpression networks in
764 Gossypium (cotton) seeds. Genome Biol Evol 8:3765-3783.
765 Ichihashi Y, Aguilar-Martínez JA, Farhi M, Chitwood DH, Kumar R, Millon LV, Peng J, Maloof
766 JN, Sinha NR. 2014. Evolutionary developmental transcriptomics reveals a gene network
767 module regulating interspecific diversity in plant leaf shape. Proc Natl Acad Sci-USA
768 111:E2616-E2621.
769 Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang J, Bowman M, Iovene M,
770 Sanseverino W, Cavagnaro P, Yildiz M. 2016. A high-quality carrot genome assembly
771 provides new insights into carotenoid accumulation and asterid genome evolution. Nat
772 Genet 48:657.
773 Josephs EB, Wright SI, Stinchcombe JR, Schoen DJ. 2017. The relationship between selection,
774 network connectivity, and regulatory variation within a population of *Capsella*
775 *grandiflora*. Genome Biol Evol 9:1099-1109.
776 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
777 improvements in performance and usability. Mol Biol Evol 30:772-780.
778 Kapralov MV, Votintseva AA, Filatov DA. 2013. Molecular adaptation during a rapid adaptive
779 radiation. Mol Biol Evol 30:1051-1059.
780 Krylov DM, Wolf YI, Rogozin IB, Koonin EV. 2003. Gene loss, protein sequence divergence,
781 gene dispensability, expression level, and interactivity are correlated in eukaryotic
782 evolution. Genome Res 13:2229-2235.
783 Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network
784 analysis. BMC Bioinformatics 9:559.
785 Langfelder P, Luo R, Oldham MC, Horvath S. 2011. Is my network module preserved and
786 reproducible? PLOS Comput Biol 7:e1001057.
787 Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. Evolution of proteins and gene
788 expression levels are coupled in *Drosophila* and are independently associated with
789 mRNA abundance, protein length, and number of protein-protein interactions. Mol Biol
790 Evol 22:1345-1354.
791 Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein
792 or nucleotide sequences. Bioinformatics 22:1658-1659.

793 Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, Simon R, Weigel D. 2001. A molecular
794 link between stem cell regulation and floral patterning in Arabidopsis. Cell 105:793-803.
795 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
796 RNA-seq data with DESeq2. Genome Biol 15:550.
797 Luisi P, Alvarez-Ponce D, Pybus M, Fares MA, Bertranpetit J, Laayouni H. 2015. Recent
798 positive selection has acted on genes encoding proteins with more interactions within the
799 whole human interactome. Genome Biol Evol 7:1141-1154.
800 Luo D, Carpenter R, Copsey L, Vincent C, Clark J, Coen E. 1999. Control of organ asymmetry
801 in flowers of Antirrhinum. Cell 99:367-376.
802 Mähler N, Wang J, Terebieniec BK, Ingvarsson PK, Street NR, Hvidsten TR. 2017. Gene co-
803 expression network connectivity is an important determinant of selective constraint.
804 PLOS Genet 13:e1006402.
805 Martén-Rodríguez S, Quesada M, Castrao AA, Lopezaraiza-Mikel M, Fenster CB. 2015. A
806 comparison of reproductive strategies between island and mainland Caribbean
807 Gesneriaceae. J Ecol 103:1190-1204.
808 Masalia RR, Bewick AJ, Burke JM. 2017. Connectivity in gene coexpression networks
809 negatively correlates with rates of molecular evolution in flowering plants. PLOS One
810 12:e182289.
811 Morandin C, Tin MM, Abril S, Gómez C, Pontieri L, Schiøtt M, Sundström L, Tsuji K, Pedersen
812 JS, Helanterä H, Mikheyev AS. 2016. Comparative transcriptomics reveals the conserved
813 building blocks involved in parallel evolution of diverse phenotypic traits in ants.
814 Genome Biol 17:43.
815 O'Meara BC, Smith SD, Armbruster WS, Harder LD, Hardy CR, Hileman LC, Hufford L, Litt
816 A, Magallón S, Smith SA, Stevens PF. 2016. Non-equilibrium dynamics and floral trait
817 interactions shape extant angiosperm diversity. P Roy Soc B-Biol Sci 283:20152304.
818 Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R
819 language. Bioinformatics 20:289-290.
820 Parsana P, Ruberman C, Jaffe AE, Schatz MC, Battle A, Leek JT. 2019. Addressing confounding
821 artifacts in reconstruction of gene co-expression networks. Genome Biol 20:94.
822 Patterson AH, Bowers JE, Ruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G,
823 Hellsten U, Mitros T, Poliakov A, Schmutz J. 2009. The *Sorghum bicolor* genome and
824 the diversification of grasses. Nature 457:551.
825 Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of
826 adaptive variation during a rapid radiation. PLOS Biol 14:e1002379.
827 Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens
828 M, Gilad Y, Pritchard JK. 2010. Understanding mechanisms underlying human gene
829 expression variation with RNA sequencing. Nature 464:768.
830 Prasad A, Kumar SS, Dessimoz C, Bleuler S, Laule O, Hruz T, Gruissem W, Zimmerman P.
831 2013. Global regulatory architecture of human, mouse and rat tissue transcriptomes.
832 BMC Genomics 14:716.

833 Preston JC, Hileman LC, Cubas P. 2011. Reduce, reuse, and recycle: developmental evolution of
834 trait diversification. Am J Bot 98:397-403.
835 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for
836 large alignments. PLOS One 5:e9490.
837 Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R. 1999.
838 Molecular analysis of the *anthocyanin2* gene of petunia and its role in the evolution of
839 flower color. Plant Cell 11:1433-1444.
840 Raherison ESM, Giguère I, Caron S, Lamara M, MacKay JJ. 2015. Modular organization of the
841 white spruce (*Picea glauca*) transcriptome reveals functional organization and
842 evolutionary signatures. New Phytol 207:172-187.
843 Ramírez-Aguirre E, Martén-Rodríguez S, Quesada-Avila G, Quesada M, Martínez-Díaz Y,
844 Oyama K, Espinosa-García FJ. 2019. Reproductive isolation among three sympatric
845 *Achimenes* species: pre- and post-pollination components. Am J Bot 106:1021-1031.
846 Ramírez Roa MA. 1987. Revision de *Achimenes* (Gesneriaceae). PhD dissertation, Universidad
847 Nacional Autonoma de Mexico, Mexico City, Mexico.
848 Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barbási AL. 2002. Hierarchical organization of
849 modularity in metabolic networks. Science 297:1551-1555.
850 Roalson EH, Roberts WR. 2016. Distinct processes drive diversification in different clades of
851 Gesneriaceae. Syst Biol 65:662-684.
852 Roalson EH, Skog LE, Zimmer EA. 2003. Phylogenetic relationships and the diversification of
853 floral form in *Achimenes* (Gesneriaceae). Syst Bot 28:593-608.
854 Roberts WR, Roalson EH, 2017. Comparative transcriptome analyses of flower development in
855 four species of *Achimenes* (Gesneriaceae). BMC Genomics 18:240.
856 Roberts WR, Roalson EH. 2018. Phylogenomic analyses reveal extensive gene flow within the
857 magic flowers (*Achimenes*). Am J Bot 105:726-740.
858 Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the
859 evolution of gene regulation. Nat Rev Genet 13:505-516.
860 Rosas-Guerrero V, Aguilar R, Martén-Rodriguez S, Ashworth L, Lopezaraiza-Mikel M, Bastida
861 JM, Quesada M. 2014. A quantitative review of pollination syndromes: do floral traits
862 predict effective pollinators? Ecol Lett 17:388-400.
863 Sauquet H, Magallón S. 2018. Key questions and challenges in angiosperm macroevolution.
864 New Phytol 219:1170-1187.
865 Serrano-Serrano ML, Marcionetti A, Perret M, Salamin N. 2017a. Transcriptome resources for
866 an endemic Neotropical plant lineage (Gesneriaceae). Appl Plant Sci 5:1600135.
867 Serrano-Serrano ML, Rolland J, Clark JL, Salamin N, Perret M. 2017b. Hummingbird
868 pollination and the diversification of angiosperms: an old and successful association in
869 Gesneriaceae. P Roy Soc B-Biol Sci 284:20162816.
870 Shahan R, Zawora C, Wight H, Sittmann J, Wang W, Mount SM, Liu Z. 2018. Consensus
871 coexpression network analysis identifies key regulators of flower and fruit development
872 in wild strawberry. Plant Physiol 178:202-216.

873 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zbodnov EM. 2015. BUSCO:
874 assessing genome assembly and annotation completeness with single-copy orthologs.
875 Bioinformatics 31:3210-3212.
876 Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and
877 molecular data. Bioinformatics 24:715-716.
878 Smith SD, Rausher MD. 2011. Gene loss and parallel evolution contribute to species difference
879 in flower color. Mol Biol Evol 28:2799-2810.
880 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
881 large phylogenies. Bioinformatics 30:1312-1313.
882 Stebbins GL. 1970. Adaptive radiation of reproductive characteristics in angiosperms, I:
883 pollination mechanisms. Annu Rev Ecol Syst 1:307-326.
884 Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene co-expression network for global discovery
885 of conserved genetic modules. Science 302:249-255.
886 Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of
887 gene ontology terms. PLOS One 6:e21800.
888 Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence
889 alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609-W612.
890 Suzuki JY, Amore TD, Calla B, Palmer NA, Scully ED, Sattler SE, Sarath G, Lichty JS, Myers
891 RY, Keith LM, Matsumoto TK. 2017. Organ-specific transcriptome profiling of
892 metabolic and pigment biosynthetic pathways in the floral ornamental progenitor species
893 *Anthurium amnicola* Dressler. Sci Rep-UK 7:1596.
894 Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. 2008. Unraveling ancient
895 hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 18:1944-1954.
896 Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Löning WE, Saedler H, Sommer H, Schwarz-
897 Sommer Z. 1990. GLOBOSA: a homeotic gene which interacts with DEFICIENS in the
898 control of Antirrhinum floral organogenesis. EMBO J 11:4693-4704.
899 Uebbing S, Künster A, Mäkinen H, Backström N, Bolivar P, Burri R, Dutoit L, Mugal CF, Nater
900 A, Aken B, Flicek P. 2016. Divergence in gene expression within and between two
901 closely related flycatcher species. Mol Ecol 25:2015-2028.
902 The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids
903 Res 45:D158-D169.
904 Van der Niet T, Johnson SD. 2012. Phylogenetic evidence for pollinator-driven diversification of
905 angiosperms. Trends Ecol Evol 27: 353-361.
906 Waites R, Selvadurai HR, Oliver IR, Hudson A. 1998. The PHANTASTICA gene encodes a
907 MYB transcription factor involved in growth and dorsoventrality of lateral organs in
908 Antirrhinum. Cell 93:779-789.
909 Wessinger CA, Rausher MD. 2014. Ecological transition predictably associated with gene
910 degeneration. Mol Biol Evol 32:347-354.
911 West-Eberhard MJ. 1989. Phenotypic plasticity and the origins of diversity. Ann Rev Ecol Syst
912 20:249-278.

913 Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and
914 low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics.
915 Mol Biol Evol 31:3081-3092.
916 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-
917 1591.
918 Yant L, Collani S, Puzey J, Levy C, Kramer EM. 2015. Molecular basis for three-dimensional
919 elaboration of the Aquilegia petal spur. P Roy Soc B-Biol Sci 282:20142778.
920 Yuan YW, Rebocho AB, Sagawa JM, Stanley LE, Bradshaw HD. 2016. Competition between
921 anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral
922 pigments between Mimulus species. Proc Natl Acad Sci USA 113: 2448-2453.
923 Zuo J, Niu Q-W, Frugis G, Chua N-H. 2002. The WUSCHEL gene promotes vegetative-to-
924 embryonic transition in Arabidopsis. Plant J 30:349-359.
925

Tables

Table 1. Gene Ontology enrichment of network hubs.

Table 2. Gene Ontology enrichment of network periphery.

Figures

Figure 1. Phylogenetic relationships in *Achimenes* and floral form. (A) Phylogeny of *Achimenes* adapted from Roberts and Roalson (2018). Flowers of each species are shown alongside the pollination syndrome (bees, butterflies, or hummingbirds). Clades 1 and Clades 2 (sensu Roalson et al. 2003) are indicated on the phylogeny with a circle and box, respectively. (B) Two stages of flower development were sampled for each of the 12 species, an early Bud stage and an intermediate D stage.

Figure 2. Co-expression clustering and the correspondence between the Bud and D stage networks. (A) Hierarchical clustering by WGCNA showing co-expressed modules for the Bud stage and (B) the D stage. Each of the 9503 orthogroups were assigned to 65 and 62 modules, respectively, and the color rows underneath show the module assignment. The 'grey' module is for unassigned orthogroups. (C) Correspondence of Bud stage modules and D stage modules. Each row corresponds to one Bud stage module (labeled by color and module size), and each column corresponds to one D stage module (labeled by color and module size). Numbers in the table indicate gene counts in the intersection of the corresponding modules. Coloring of the table encodes -log(p), with p being the Fisher's exact test p-value for the overlap of the two modules. The stronger the red color, the more significant the overlap.

Figure 3. Summary of Bud stage network module association to floral traits and pollination syndrome. Each row represents an individual module in the Bud stage network and shows the total module size (column All) and the number of hubs (column Hubs). Each column in the central table represents the floral traits (Pollination Syndrome, Flower color, Flower shape, and Spurs) that were tested for an association to module eigengenes. The eigengene expression is indicated for each module in which all three replicates from a species either had > 1.0 or < -1.0 expression. An association was considered 'Strongly positive' when both the eigengene expression was > 1.0 in species that had the trait and MCMCglmm results were significant. An association was considered 'Strongly negative' when both the eigengene expression was < -1.0 in species that had the trait and MCMCglmm results were significant. Associations were considered 'Weakly positive' or 'Weakly negative' when only MCMCglmm results were significant.

Figure 4. Summary of D stage network module association to floral traits and pollination syndrome. Each row represents an individual module in the D stage network and shows the total module size (column All) and the number of hubs (column Hubs). Each column in the central table represents the floral traits (Pollination Syndrome, Flower color, Flower shape, and Spurs) that were tested for an association to module eigengenes. The eigengene expression is indicated for each module in which all three replicates from a species either had > 1.0 or < -1.0 expression.

An association was considered 'Strongly positive' when both the eigengene expression was > 1.0 in species that had the trait and MCMCglmm results were significant. An association was considered 'Strongly negative' when both the eigengene expression was < -1.0 in species that had the trait and MCMCglmm results were significant. Associations were considered 'Weakly positive' or 'Weakly negative' when only MCMCglmm results were significant.

Figure 5. Relationship between orthogroup dN/dS (omega) and orthogroup connectivity. (A) Bud stage network. (B) D stage network. Each point indicates an individual orthogroup. The red line indicates the linear regression line for the relationship between dN/dS and connectivity. The dotted line indicates dN/dS = 1 with orthogroups above this line considered to be under relaxed or positive selection.

Supplemental Files

Suppl. Table 1. Sequencing and transcriptome assembly results.

Suppl. Table 2. Gene Ontology enrichment for the Bud stage modules.

Suppl. Table 3. Gene Ontology enrichment for the D stage modules.

Suppl. Table 4. MedianRank and Zsummary module preservation scores.

Suppl. Table 5. Orthogroups identified as hubs, including module membership and annotation.

Suppl. Table 6. Orthogroups identified as periphery, including module membership and annotation.

Suppl. Table 7. Statistical results of the evolutionary rates analyses.

Suppl. Figure 1. Module eigengenes in the Bud stage network. Plots show the module eigengenes for each module in the Bud stage network. Each bar represents the eigengene expression (y-axis) for each sample in the Bud stage dataset (x-axis). Eigengene expression represents the expression of a sample within a module.

Suppl. Figure 2. Module eigengenes in the D stage network. Plots show the module eigengenes for each module in the D stage network. Each bar represents the eigengene expression (y-axis) for each sample in the D stage dataset (x-axis). Eigengene expression represents the expression of a sample within a module.

Suppl. Figure 3. Effects of floral traits on the eigengene expression in the Bud stage modules summarized over 50 phylogenetic trees. Each plot shows a summary of the important floral predictors (flower color, flower shape, and corolla spur) for each module in the Bud stage network. The posterior distribution of an independent variable with negligible effect on

eigengene expression is centered around zero (dashed line). The distribution of an influential variable is expected to be substantially different from zero. Black dots indicate the mode of the posterior distribution, and boxes indicate the 50 and 95% credible intervals.

Suppl. Figure 4. Effects of pollination syndrome on the eigengene expression in the Bud stage modules summarized over 50 phylogenetic trees. Each plot shows a summary of the important pollination syndrome predictors for each module in the Bud stage network. The posterior distribution of an independent variable with negligible effect on eigengene expression is centered around zero (dashed line). The distribution of an influential variable is expected to be substantially different from zero. Black dots indicate the mode of the posterior distribution, and boxes indicate the 50 and 95% credible intervals.

Suppl. Figure 5. Effects of floral traits on the eigengene expression in the D stage modules summarized over 50 phylogenetic trees. Each plot shows a summary of the important floral predictors (flower color, flower shape, and corolla spur) for each module in the D stage network. The posterior distribution of an independent variable with negligible effect on eigengene expression is centered around zero (dashed line). The distribution of an influential variable is expected to be substantially different from zero. Black dots indicate the mode of the posterior distribution, and boxes indicate the 50 and 95% credible intervals.

Suppl. Figure 6. Effects of pollination syndrome on the eigengene expression in the D stage modules summarized over 50 phylogenetic trees. Each plot shows a summary of the important pollination syndrome predictors for each module in the D stage network. The posterior distribution of an independent variable with negligible effect on eigengene expression is centered around zero (dashed line). The distribution of an influential variable is expected to be substantially different from zero. Black dots indicate the mode of the posterior distribution, and boxes indicate the 50 and 95% credible intervals.

Suppl. Figure 7. Relationship between orthogroup dN/dS (omega) and average expression. (A) Bud stage. (B) D stage. Each point indicates an individual orthogroup. The red line indicates the linear regression line between fit for dN/dS ~ expression. The dotted line indicates dN/dS = 1, where orthogroups with dN/dS > 1 may be under stronger relaxed selection.

Suppl. Figure 8. Distribution of dN/dS (omega) among hubs and periphery nodes. (A) dN/dS for hub and non-hub nodes in the Bud stage network. (B) dN/dS for hub and non-hub nodes in the D stage network. (C) dN/dS for periphery and non-periphery nodes in the Bud stage network. (D) dN/dS for periphery and non-periphery nodes in the D stage network. Violin plots show the distribution of dN/dS values for each category. Horizontal lines in each distribution indicate the median dN/dS value.

Table 1. Gene Ontology enrichment of network hubs.

	GO.ID	Term	Annotated	Significant	Expected	P
A. Bud stage hubs	GO:0031325	positive regulation of cellular metabolic process	224	16	14.52	0.000
	GO:0006607	NLS-bearing protein import into nucleus	7	4	0.45	0.001
	GO:0015991	ATP hydrolysis coupled proton transport	9	4	0.58	0.002
	GO:0046470	phosphatidylcholine metabolic process	9	4	0.58	0.002
	GO:0018298	protein-chromophore linkage	32	7	2.07	0.004
	GO:0070507	regulation of microtubule cytoskeleton organization	6	3	0.39	0.005
	GO:1990778	protein localization to cell periphery	6	3	0.39	0.005
	GO:0051592	response to calcium ion	7	3	0.45	0.008
	GO:0006412	translation	233	24	15.10	0.009
	GO:0042594	response to starvation	60	10	3.89	0.012
	GO:0007035	vacuolar acidification	8	3	0.52	0.012
	GO:0006534	cysteine metabolic process	9	4	0.58	0.012
	GO:0009833	plant-type primary cell wall biogenesis	15	4	0.97	0.013
	GO:0048831	regulation of shoot system development	123	12	7.97	0.017
	GO:0032509	endosome transport via multivesicular body sorting pathway	9	3	0.58	0.017
	GO:0051568	histone H3-K4 methylation	9	3	0.58	0.017
	GO:0000271	polysaccharide biosynthetic process	123	11	7.97	0.023
	GO:0006732	coenzyme metabolic process	116	8	7.52	0.023
	GO:0010338	leaf formation	5	2	0.32	0.037
	GO:0048564	photosystem I assembly	5	2	0.32	0.037
	GO:0017001	antibiotic catabolic process	26	3	1.69	0.037
	GO:0044282	small molecule catabolic process	75	10	4.86	0.038
	GO:0006833	water transport	13	3	0.84	0.048
	GO.ID	Term	Annotated	Significant	Expected	P
B. D stage hubs	GO:0006412	translation	233	30	16.03	0.001
	GO:0042939	tripeptide transport	8	4	0.55	0.001
	GO:0016117	carotenoid biosynthetic process	20	6	1.38	0.002
	GO:0006744	ubiquinone biosynthetic process	5	3	0.34	0.003
	GO:0032259	methylation	88	15	6.05	0.004
	GO:0045814	negative regulation of gene expression, epigenetic	37	5	2.55	0.005
	GO:0009206	purine ribonucleoside triphosphate biosynthetic process	44	6	3.03	0.013
	GO:0000271	polysaccharide biosynthetic process	123	11	8.46	0.013
	GO:0009873	ethylene-activated signaling pathway	78	12	5.37	0.014
	GO:0015991	ATP hydrolysis coupled proton transport	9	3	0.62	0.020
	GO:0009718	anthocyanin-containing compound biosynthetic process	26	5	1.79	0.030
	GO:0000018	regulation of DNA recombination	5	2	0.34	0.041
	GO:0015976	carbon utilization	5	2	0.34	0.041
	GO:0034314	Arp2/3 complex-mediated actin nucleation	5	2	0.34	0.041
	GO:0080144	amino acid homeostasis	5	2	0.34	0.041
	GO:1901615	organic hydroxy compound metabolic process	121	9	8.32	0.041
	GO:0000902	cell morphogenesis	174	9	11.97	0.042

Table 2. Gene Ontology enrichment of network periphery.

	GO.ID	Term	Annotated	Significant	Expected	P
A. Bud stage periphery	GO:0016579	protein deubiquitination	24	16	8.70	0.002
	GO:0048825	cotyledon development	21	12	7.61	0.005
	GO:0015846	polyamine transport	5	5	1.81	0.006
	GO:0034314	Arp2/3 complex-mediated actin nucleation	5	5	1.81	0.006
	GO:0006020	inositol metabolic process	12	7	4.35	0.011
	GO:0010229	inflorescence development	9	7	3.26	0.014
	GO:0009744	response to sucrose	31	16	11.24	0.021
	GO:1903338	regulation of cell wall organization or biogenesis	11	7	3.99	0.026
	GO:0006809	nitric oxide biosynthetic process	6	5	2.18	0.026
	GO:0009773	photosynthetic electron transport in photosystem I	8	6	2.90	0.030
	GO:0034613	cellular protein localization	183	75	66.34	0.031
	GO:0006004	fucose metabolic process	12	8	4.35	0.032
	GO:0009926	auxin polar transport	42	22	15.23	0.037
	GO:0010090	trichome morphogenesis	38	19	13.78	0.041
	GO:0009414	response to water deprivation	182	69	65.98	0.045
	GO:0009873	ethylene-activated signaling pathway	78	38	28.28	0.046
	GO:0007166	cell surface receptor signaling pathway	93	42	33.71	0.046
	GO:0006163	purine nucleotide metabolic process	96	38	34.80	0.047
	GO:0035195	gene silencing by miRNA	15	7	5.44	0.048
	GO:0016236	macroautophagy	21	9	7.61	0.048
	GO:0055083	monovalent inorganic anion homeostasis	13	5	4.71	0.048
	GO:0051338	regulation of transferase activity	38	13	13.78	0.048
	GO:0006473	protein acetylation	23	7	8.34	0.048
	GO.ID	Term	Annotated	Significant	Expected	weight01
B. D stage periphery	GO:0045490	pectin catabolic process	20	19	12.88	0.002
	GO:0007166	cell surface receptor signaling pathway	93	71	59.90	0.006
	GO:0000398	mRNA splicing, via spliceosome	46	38	29.63	0.007
	GO:0009740	gibberellic acid mediated signaling pathway	31	24	19.97	0.015
	GO:0006310	DNA recombination	54	41	34.78	0.015
	GO:0009909	regulation of flower development	80	57	51.53	0.017
	GO:0009058	biosynthetic process	2121	1360	1366.19	0.019
	GO:0042545	cell wall modification	34	26	21.90	0.021
	GO:0001708	cell fate specification	13	12	8.37	0.027
	GO:0042538	hyperosmotic salinity response	25	21	16.10	0.027
	GO:0002181	cytoplasmic translation	21	18	13.53	0.029
	GO:0006897	endocytosis	40	27	25.76	0.029
	GO:0010197	polar nucleus fusion	20	17	12.88	0.039
	GO:0071702	organic substance transport	492	319	316.91	0.046
	GO:0006949	syncytium formation	7	7	4.51	0.046
	GO:0008295	spermidine biosynthetic process	7	7	4.51	0.046
	GO:0009963	positive regulation of flavonoid biosynthetic process	7	7	4.51	0.046

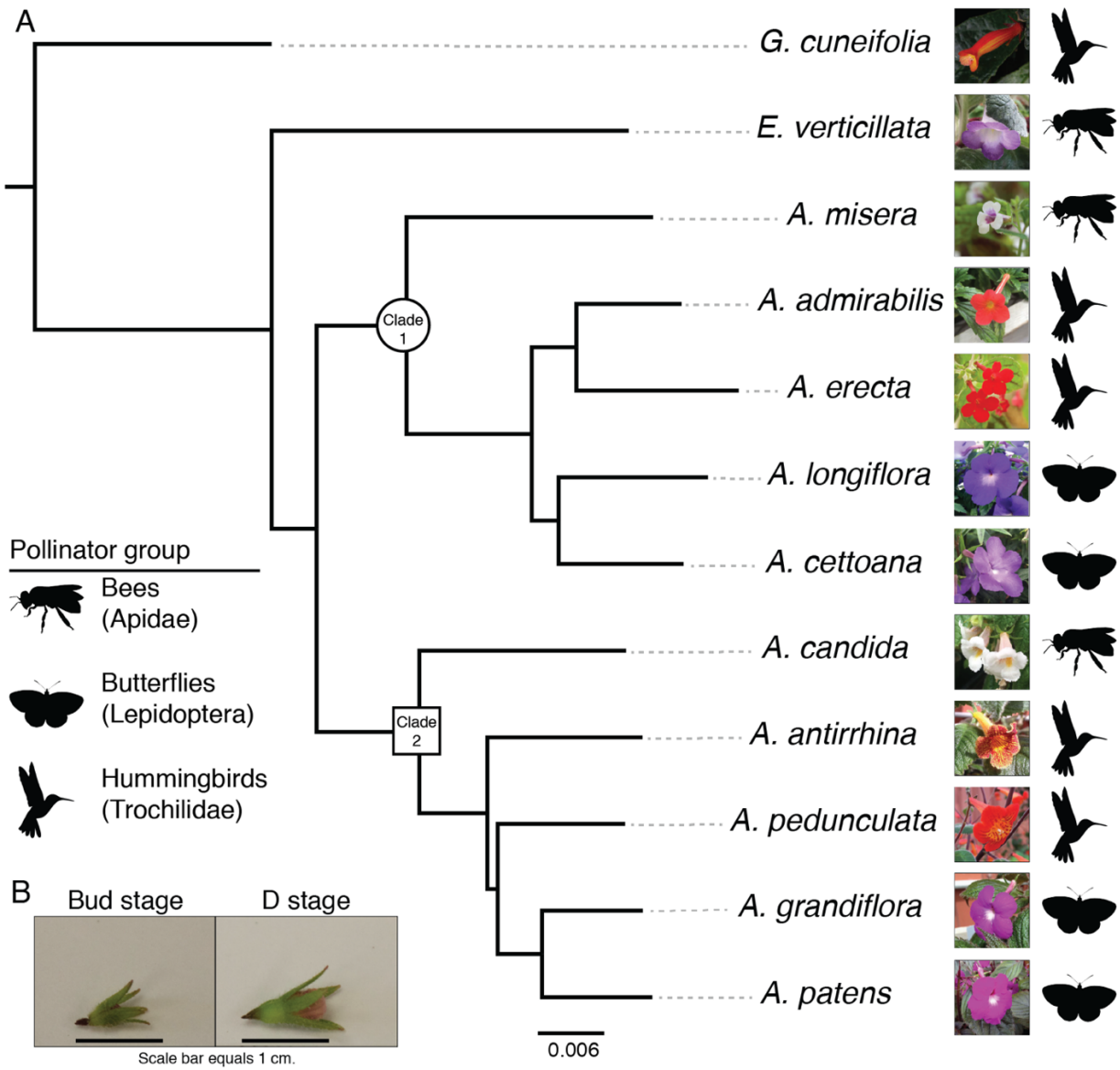


Figure 1. Phylogenetic relationships in *Achimenes* and floral form.

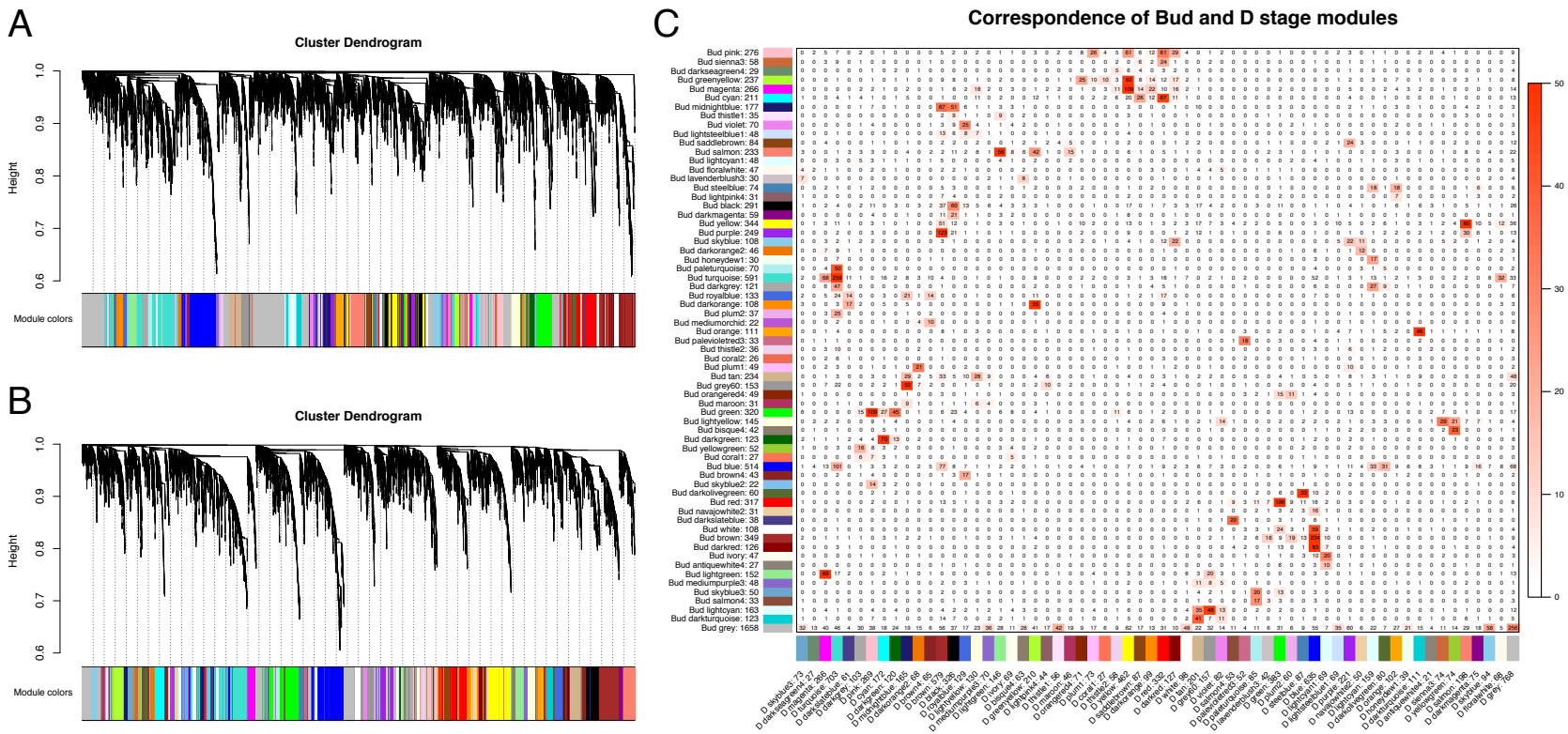
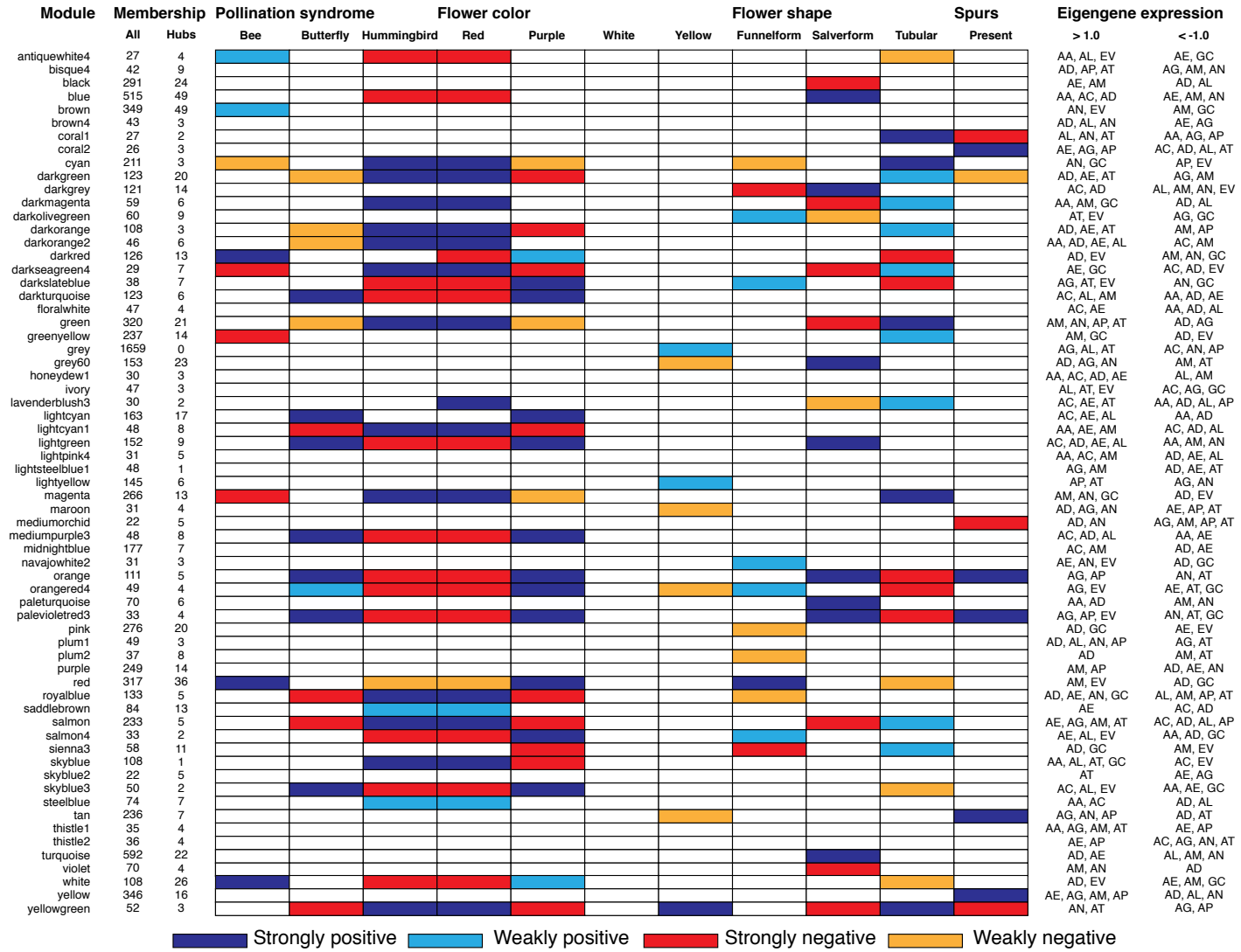
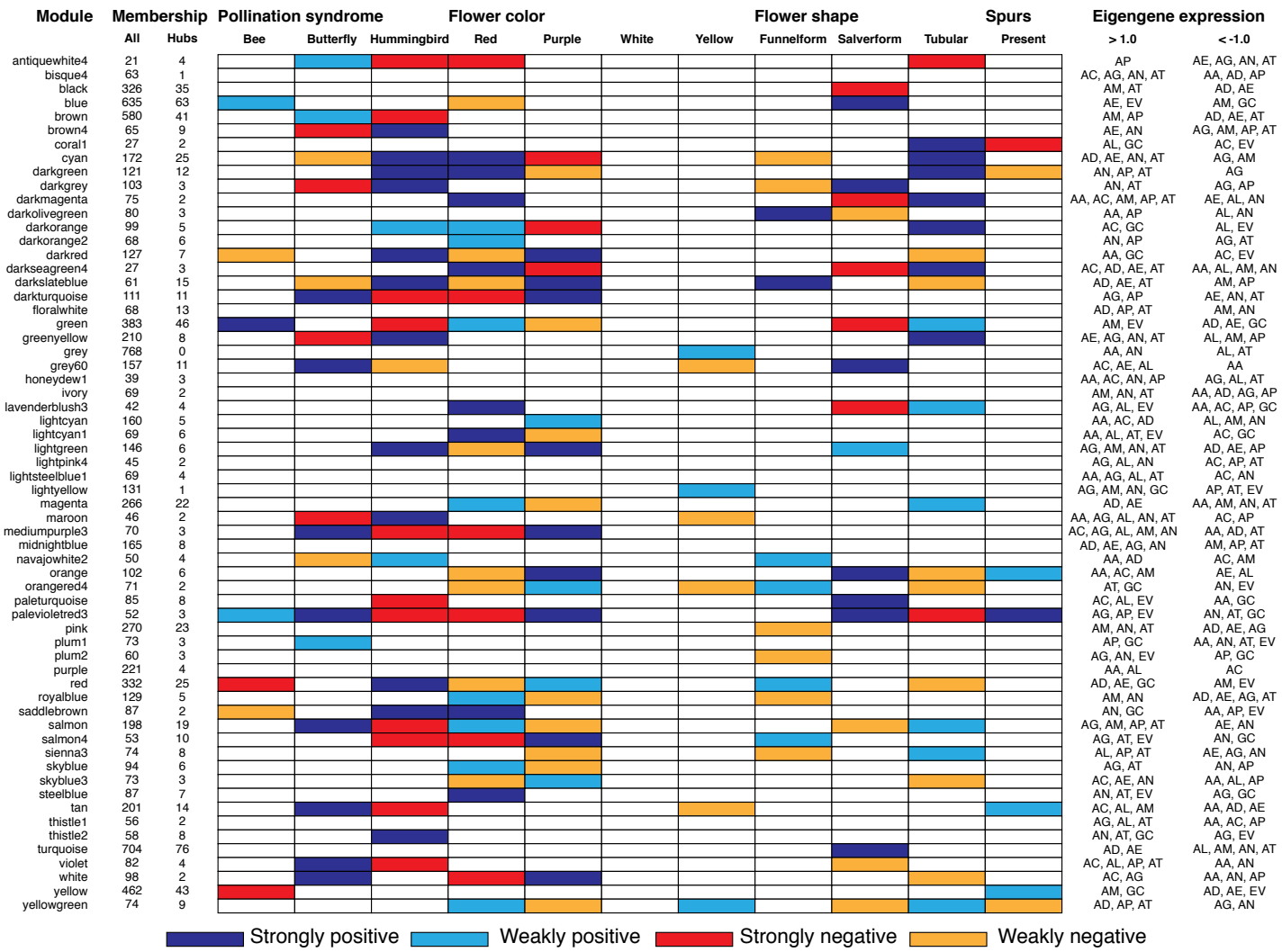


Figure 2. Co-expression clustering and the correspondence between the Bud and D stage networks.



Abbreviations: AA, *A. admirabilis*; AC, *A. cettoana*; AD, *A. candida*; AE, *A. erecta*; AG, *A. grandiflora*; AL, *A. longiflora*; AM, *A. misera*; AN, *A. pedunculata*; AP, *A. patens*; AT, *A. antirrhina*; EV, *E. verticillata*; GC, *G. cuneifolia*.

Figure 3. Summary of Bud stage network module association to floral traits and pollination syndrome.



Abbreviations: AA, *A. admirabilis*; AC, *A. cettoana*; AD, *A. candida*; AE, *A. erecta*; AG, *A. grandiflora*; AL, *A. longiflora*; AM, *A. misera*; AN, *A. pedunculata*; AP, *A. patens*; AT, *A. antirrhina*; EV, *E. verticillata*; GC, *G. cuneifolia*.

Figure 4. Summary of D stage network module association to floral traits and pollination syndrome.

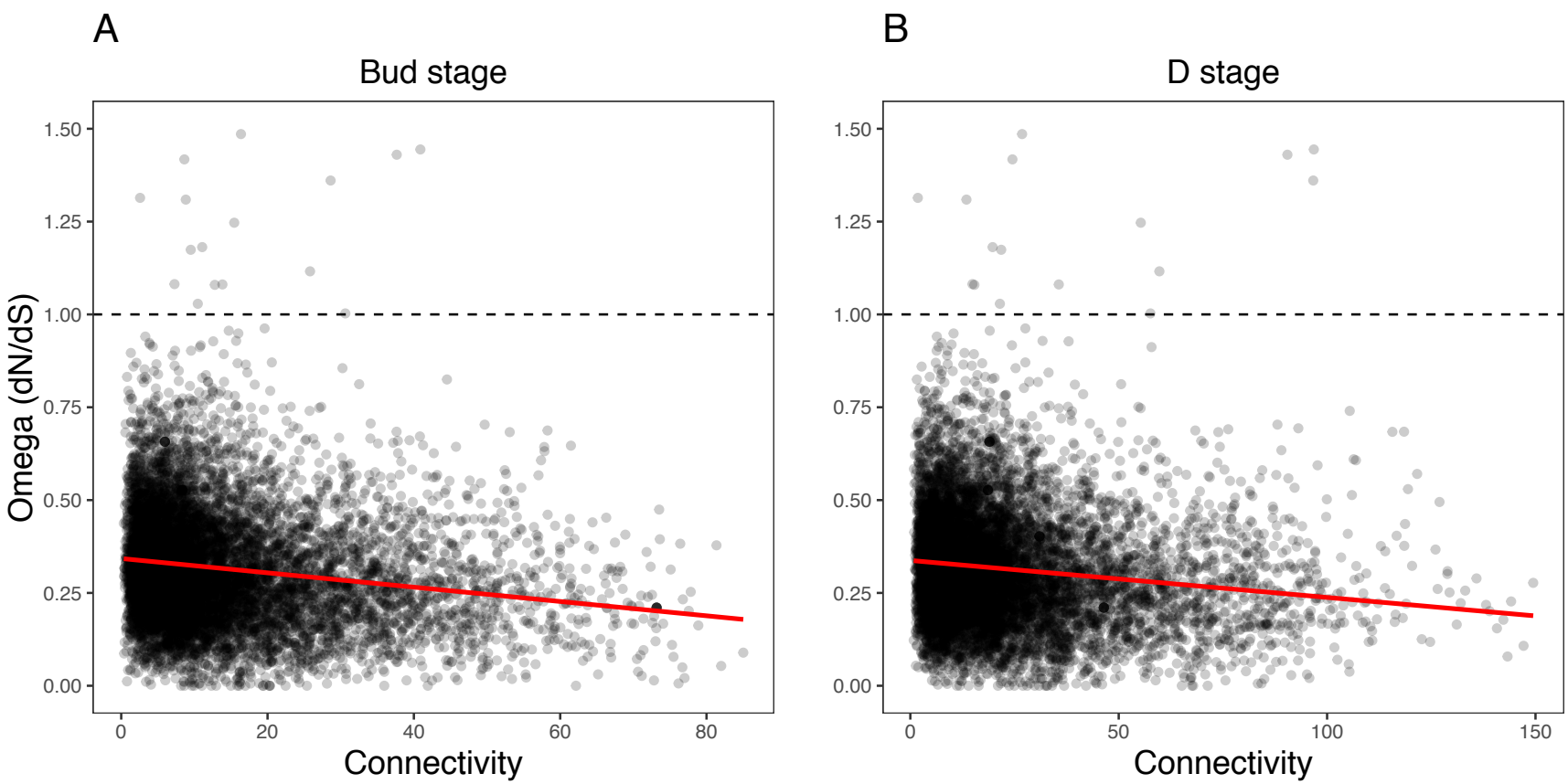
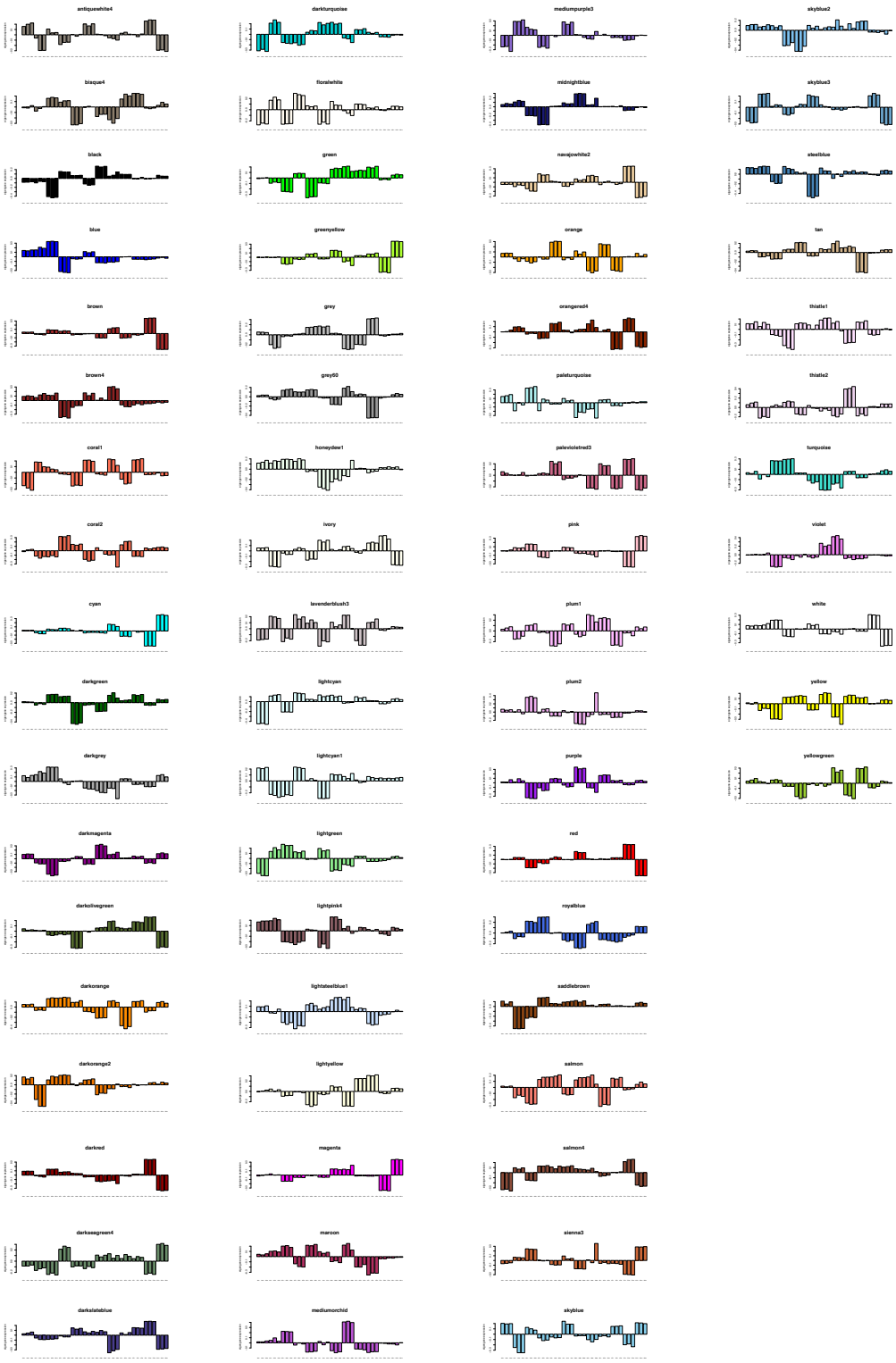
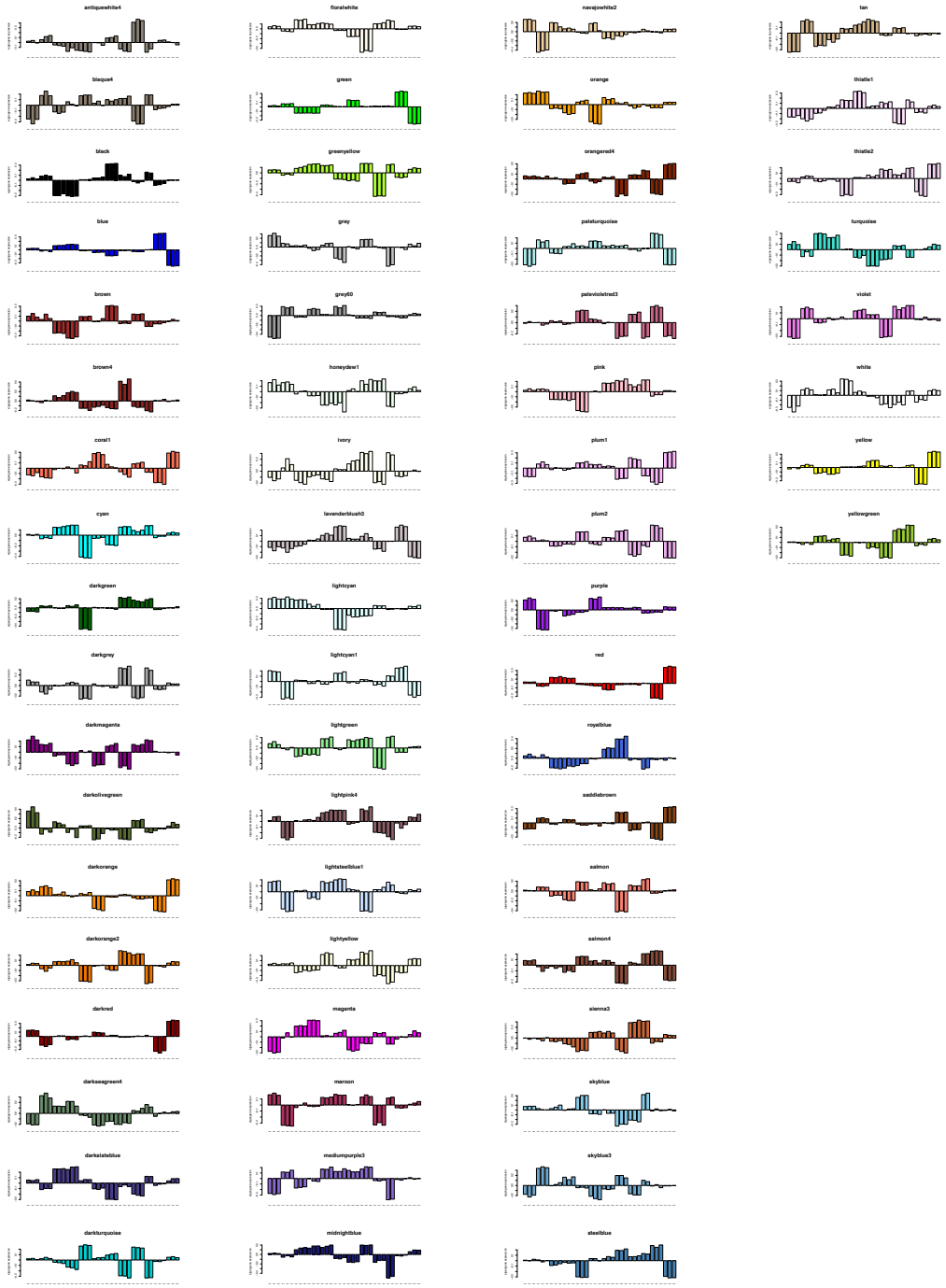


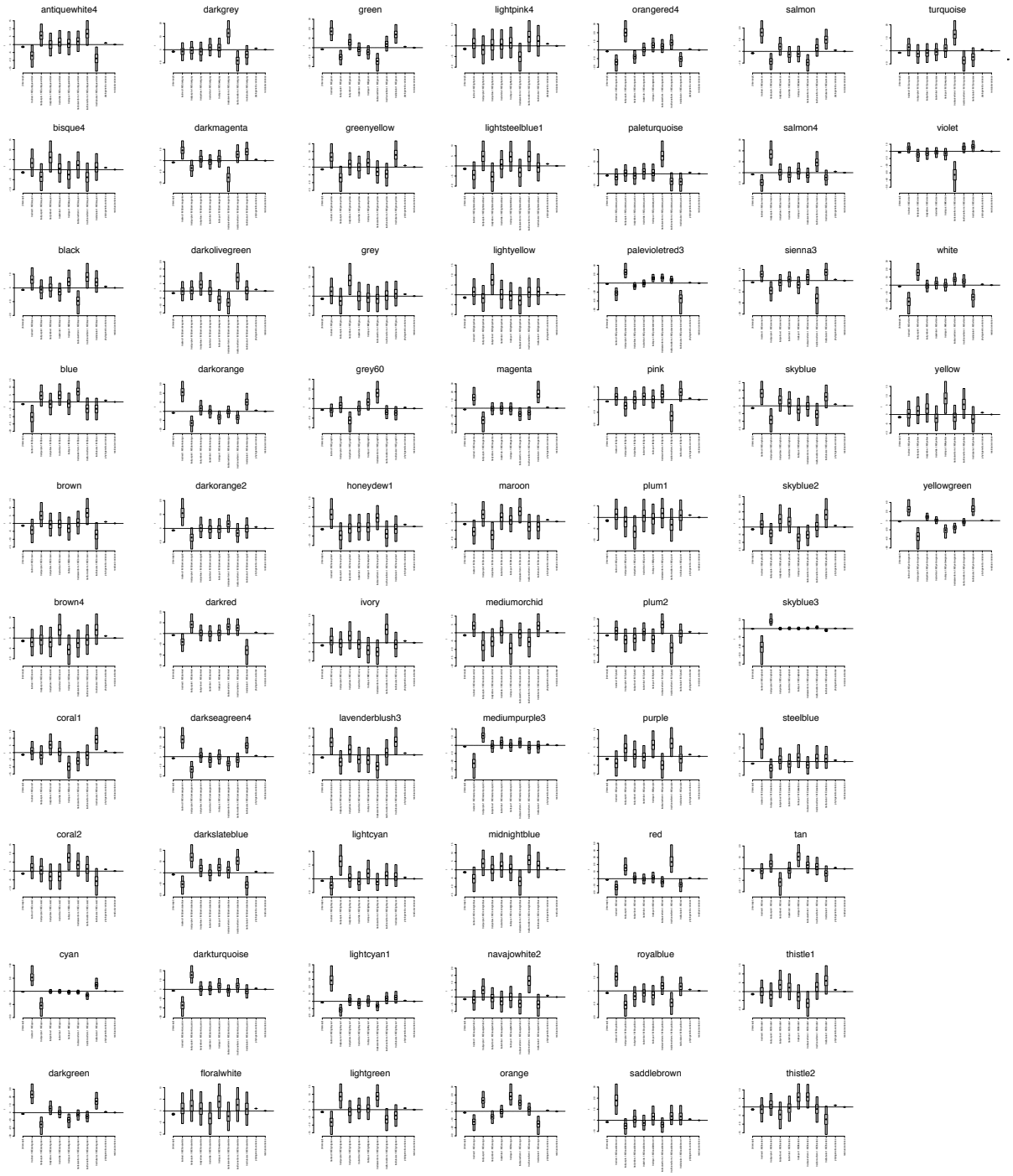
Figure 5. Relationship between orthogroup dN/dS (omega) and orthogroup connectivity.



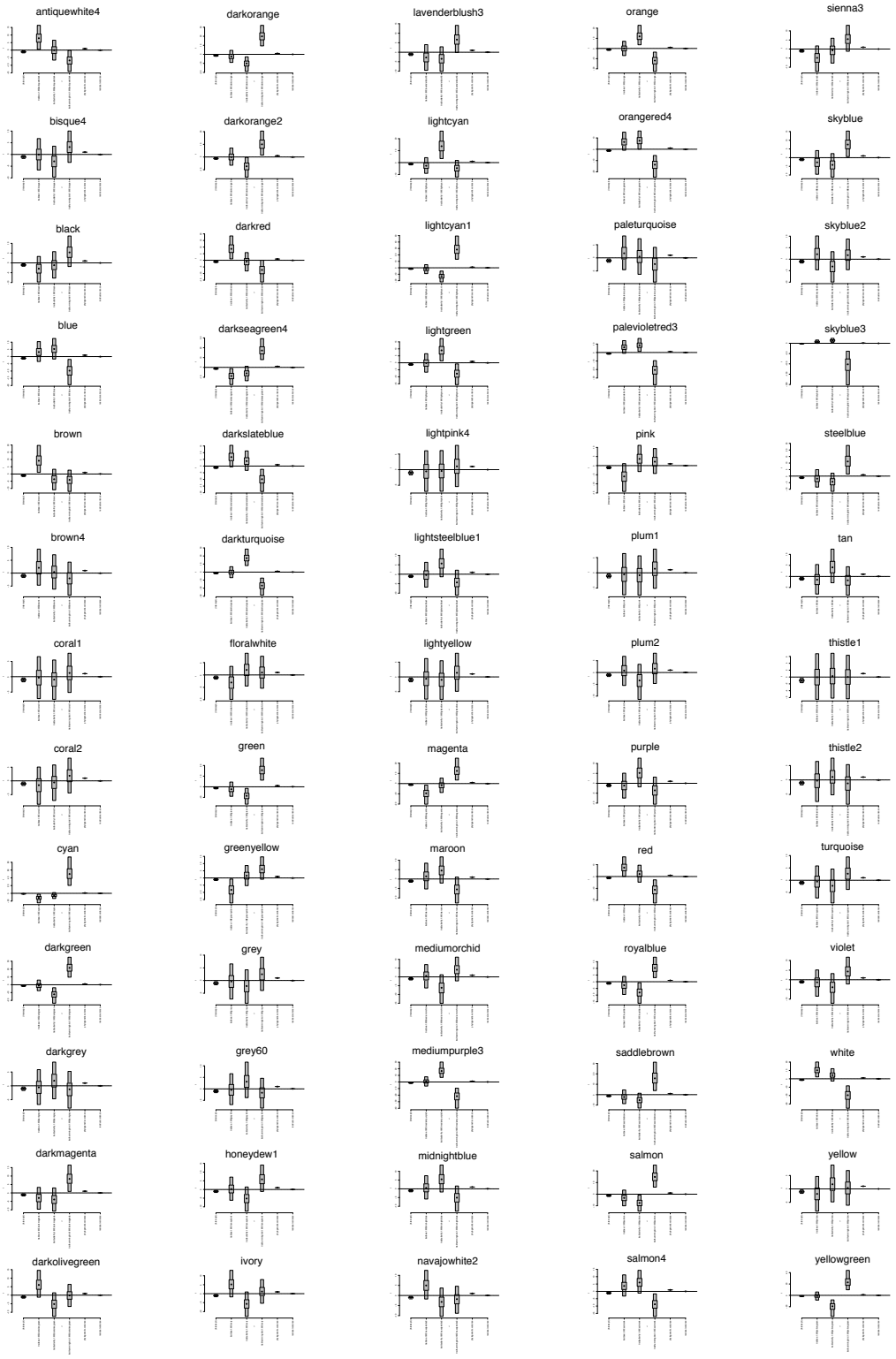
Supplemental Figure 1. Module eigengenes in the Bud stage network.



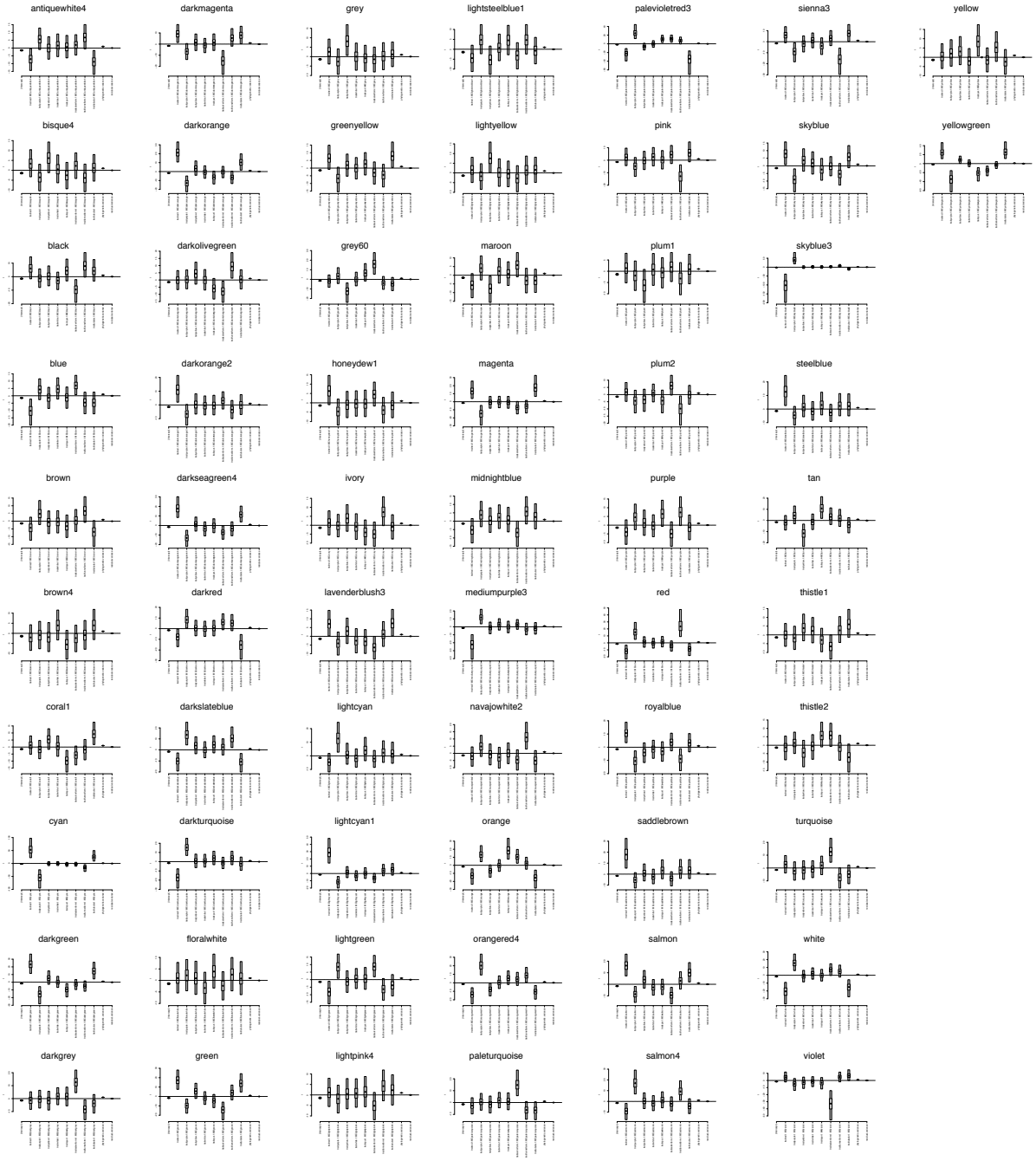
Supplemental Figure 2. Module eigengenes in the D stage network.



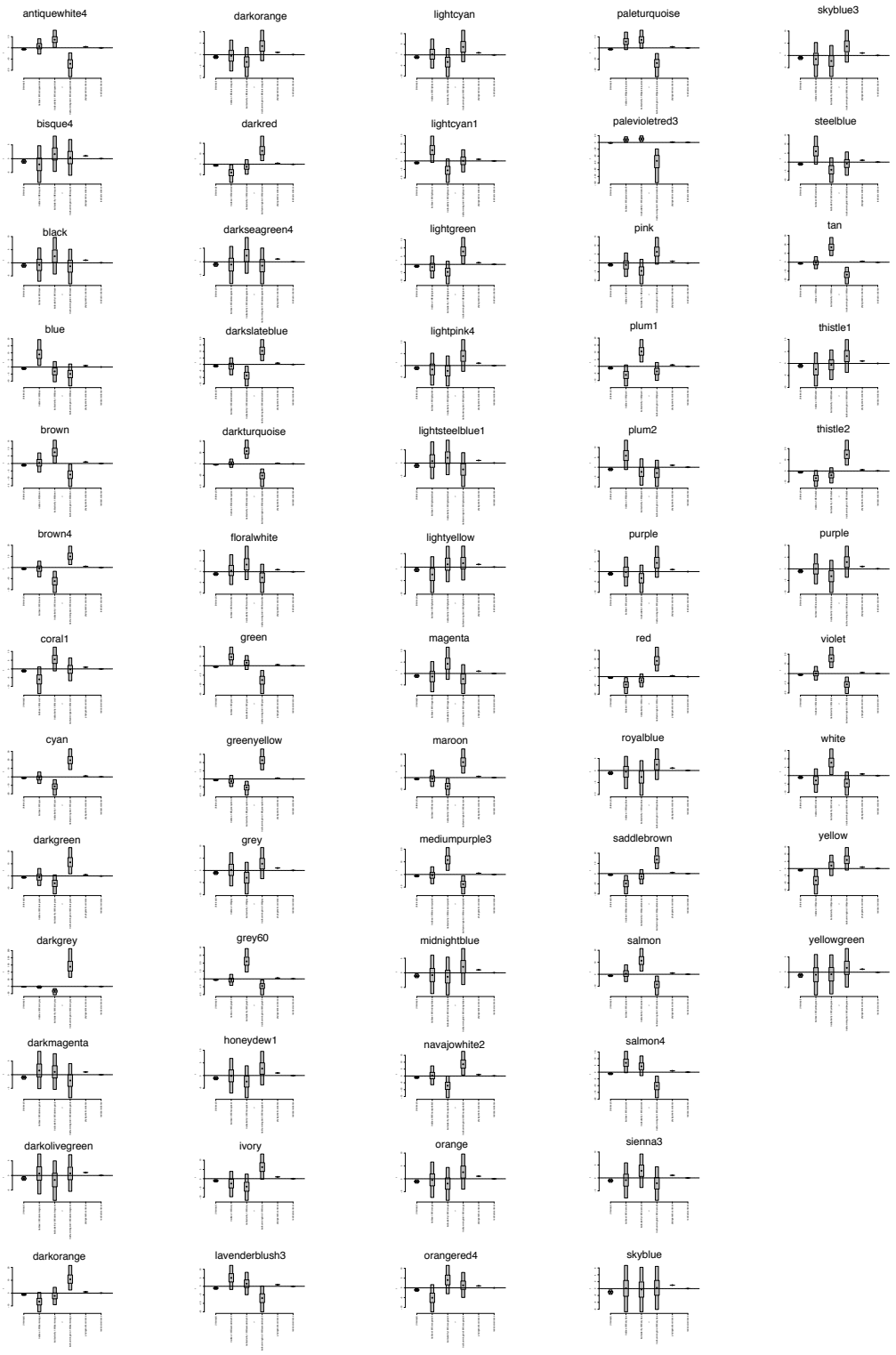
Supplemental Figure 3. Effects of floral traits on the eigengene expression in the Bud stage modules summarized over 50 phylogenetic trees.



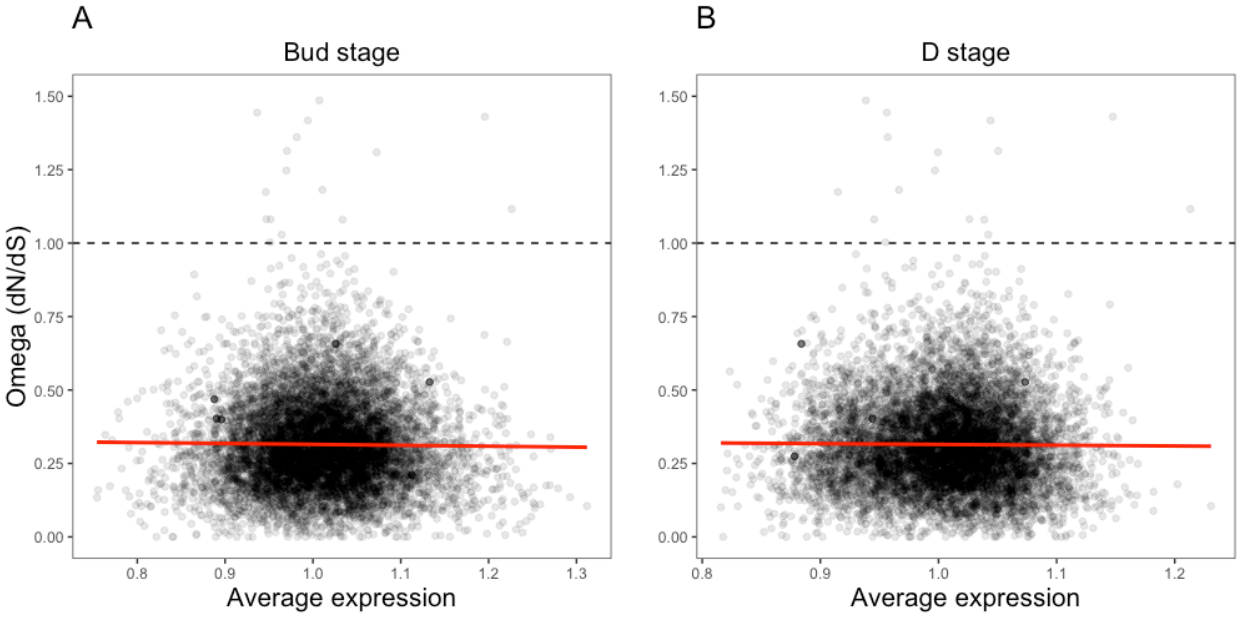
Supplemental Figure 4. Effects of pollination syndrome on the eigengene expression in the Bud stage modules summarized over 50 phylogenetic trees.



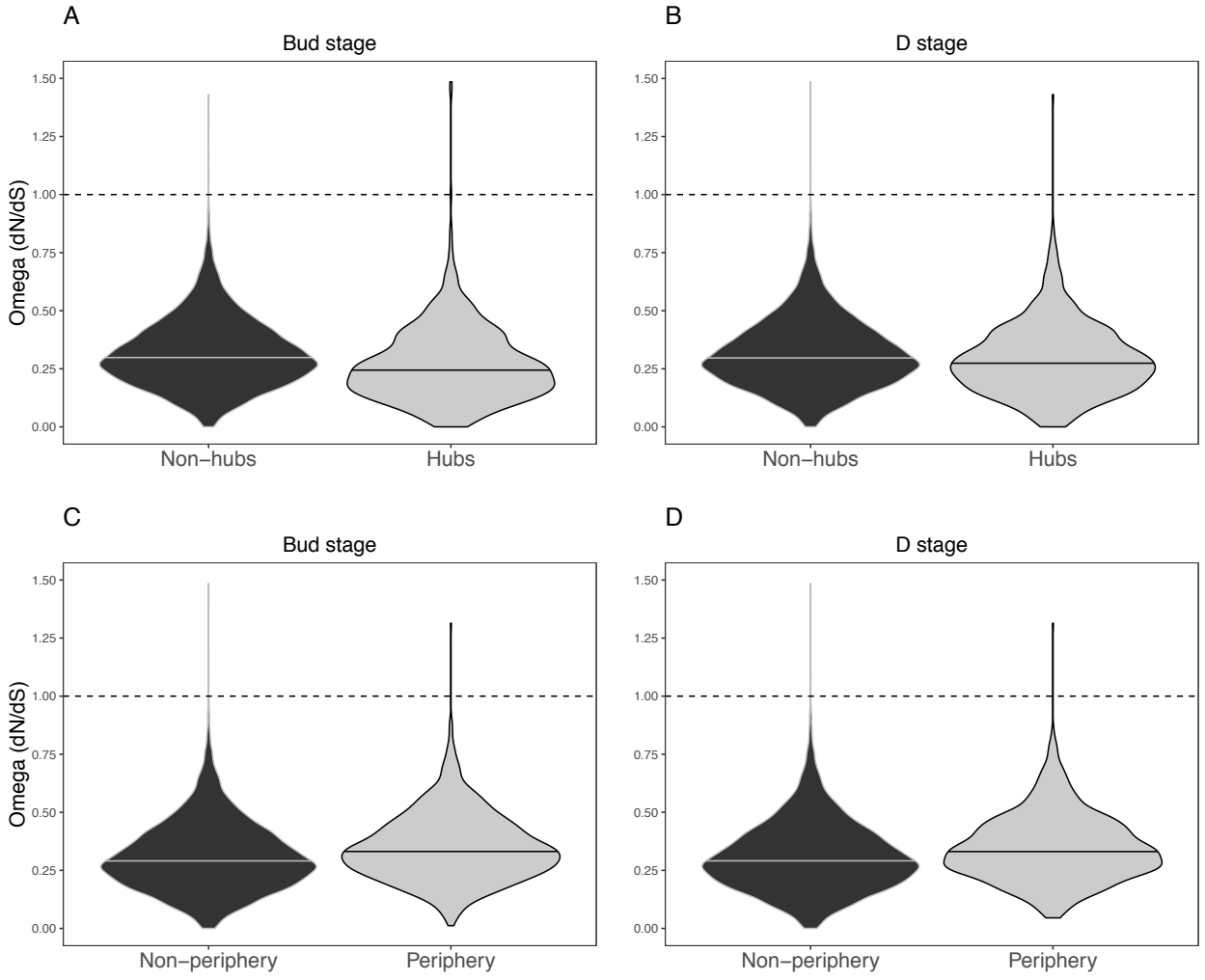
Supplemental Figure 5. Effects of floral traits on the eigengene expression in the D stage modules summarized over 50 phylogenetic trees.



Supplemental Figure 6. Effects of pollination syndrome on the eigengene expression in the D stage modules summarized over 50 phylogenetic trees.



Supplemental Figure 7. Relationship between orthogroup dN/dS (omega) and average expression.



Supplemental Figure 8. Distribution of dN/dS (omega) among hubs and periphery nodes.