Biological Sciences: Evolution

Topological features of gene regulatory networks predict patterns of natural diversity in environmental response

David L. Des Marais[1]
Rafael F. Guerrero[2]
Jesse R. Lasky[3]
Samuel V. Scarpino[4,5]

1. Arnold Arboretum and Department of Organismic and Evolutionary Biology, Harvard University. 1300 Centre St. Boston, MA. 20131.
2. Department of Biology, Indiana University. Jordan Hall 142. Bloomington, IN. 47405
3. Department of Biology, Pennsylvania State University. 408 Life Sciences Building. University Park, PA. 16802.
4. Department of Mathematics & Statistics and Vermont Complex Systems Center, University of Vermont. 210 Colchester Ave, Burlington, VT. 05405
5. Santa Fe Institute. 1399 Hyde Park Rd, Santa Fe, NM. 87501.

Corresponding authors:
David L. Des Marais (desmarais@fas.harvard.edu)
Samuel V. Scarpino (samuel.v.scarpino@uvm.edu)

Keywords: gene by environment interaction, gene regulatory network, *Arabidopsis thaliana*, abiotic stress, drought, cold

Abstract

Molecular interactions affect the evolution of complex traits. For instance, adaptation may be constrained by pleiotropic or epistatic effects, both of which will be reflected in the structure of molecular interaction networks. To date, empirical studies investigating the role of molecular interactions in phenotypic evolution have been idiosyncratic, offering no clear patterns. Here, we investigated the network topology of genes putatively involved in local adaptation to two abiotic stressors—drought and cold—in *Arabidopsis thaliana*. Our findings suggest that the gene-interaction topologies for both cold and drought stress response are non-random, with genes that show genetic variation in drought response (GxE) being significantly more peripheral and cold response genes being significantly more central than genes not involved in either response. We suggest that the observed topologies reflect different constraints on the genetic pathways involved in the assayed phenotypes. The approach presented here may inform predictive models linking genetic variation in molecular signaling networks with phenotypic variation, specifically traits involved in environmental response.

Significance Statement

Our study focuses on genes whose transcriptional activity exhibits genetic variation in response to the environment, or "GxE." GxE is a widely observed phenomenon of critical importance to understanding the genotype-to-phenotype map, the evolution of natural populations, medical genetics, population response to climate change, and agricultural improvement. We investigated expression GxE in plant responses to two abiotic cues: cold and drought. We found that genes showing genetically variable response to cold stress are centrally located in regulatory networks whereas genes showing genetically variable response to drought stress are peripherally located in regulatory networks. This result suggests that selection is presented with vastly different mutational landscapes for shaping evolutionary or breeding response to these two important climatic factors

Introduction

Genes do not function nor evolve in isolation. The transcriptional activities of genes in a genome are often highly correlated with one another, forming hierarchical regulatory

58 networks comprised of functionally related modules (1). Within such networks, some
59 genes -- “nodes” – have stronger or more interactions -- “edges” – with one another than do
60 other genes. Because the effect size of mutations is strongly associated with their
61 evolutionary fate (2), the structural properties of genetic regulatory networks will likely
62 affect selection acting on individual component genes (3-5). Advances in high-throughput
63 molecular phenotyping and systems analysis have improved our ability to characterize
64 molecular interaction networks, providing the opportunity to address classic questions
65 about the evolution of genetic interactions.
66 Two related features of gene regulatory networks might affect the evolution of
67 individual genes within those networks. The first is the widespread observation that genes
68 vary in their number of interacting neighbor genes, perhaps even by orders of magnitude
69 (6). This feature -- the centrality or connectivity of a gene -- can be measured in many
70 different ways, including the number of directly interacting genes or the number of paths to
71 other genes that pass through a given gene (7). The second feature of networks that can
72 impact gene evolution is modularity, i.e. the degree to which the network is composed of
73 functionally related sub-networks of genes, or modules. Modules are often under the
74 transcriptional control of core proteins, high-level switches that regulate the module’s
75 activity (8) using shared regulatory motifs among genes within it (9). Evidence for the
76 pleiotropic nature of core genes has been found through decades of developmental genetics
77 research, which identified putative master regulators of the level, timing, and location of
78 expression of tens to thousands of other genes (3, 8, 10).
79 Transcriptional regulation by core genes plays an important role in adaptive
80 responses to the environment (11, 12). Environmentally responsive transcripts are often
81 co-regulated as functional modules (13, 14), and in some instances the response of
82 particular genes to environmental cues may be characteristic of entire species or kingdoms
83 (15, 16). Considerable genetic variation in transcriptional response to environment --
84 expression Genotype by Environment interaction, eGxE -- has also been identified within
85 species (17). At the molecular level, eGxE may be controlled by genetic variants acting in
86 cis, e.g. by SNP or presence-absence variants in promoter motifs, or by genetic variants
87 acting in trans, such as transcription factors, small RNA species, or a number of other
88 regulatory factors upstream of genes showing eGxE. Genetic variants affecting eGxE are of

89 particular interest because GxE represents the mutational substrate for the evolution
90 environmental response (18, 19) and because GxE for fitness is required for local
91 adaptation (20).
92 In the present study, we explore two hypotheses for how environmentally
93 responsive regulatory networks evolve and might thereby be involved in local adaptation
94 to environment. The first hypothesis, that eGxE is driven by genetic variation in core
95 transcriptional regulatory proteins, arises from the observation that suites of traits often
96 show high genetic correlation ((21); in this context, “traits" could be either individual
97 transcripts or higher-level physiological or developmental phenotypes). Genetic variants
98 in one or a small number of regulatory genes could therefore have considerable
99 downstream consequences, both positive and negative with respect to transcription level,
100 trait expression, and fitness. This model predicts that eGxE genes would have relatively
101 high network connectivity and, by extension, be clustered in relatively discrete functional
102 modules. The second hypothesis posits that eGxE could be primarily driven by variation in
103 genes located peripherally in transcriptional networks, which are expected to have smaller
104 effect sizes and reduced deleterious pleiotropy. Variation in peripheral genes could
105 therefore allow natural selection to “fine-tune" environmental response by changing only a
106 small number of expression or higher-order traits. While these are not mutually exclusive
107 hypotheses, their relative importance in nature has not been established.
108 Here, we extend earlier work assessing genetic variation in transcriptional activity
109 during acclimation to cold (22) and soil drying (23) in *Arabidopsis thaliana*. We predict
110 different patterns for the genes associated with each environmental response based on our
111 previous observations. The sequence conservation in environmentally responsive
112 promoter motifs among cold eGxE genes in Arabidopsis (24) leads us to predict that cold
113 acclimation eGxE genes will be clustered and highly connected. The conserved patterns of
114 nucleotide diversity imply that *cis*-regulatory elements of eGxE genes for cold are
115 evolutionarily conserved and that expression diversity is controlled by genetic diversity in
116 upstream regulators such as transcription factors. These transcription factors, acting as
117 core regulators, would cause the coexpression of large sets of eGxE genes that would be
118 observed as a highly connected network. Conversely, we predict that drought eGxE genes
119 will be located peripherally in networks because these genes show evidence of adaptive *cis*-

120 regulatory variation (24). This predominant role of *cis* variants suggests that most genes
121 involved in drought local adaptation have small *trans* effects and are therefore likely to
122 have low connectivities in the coexpression network. By extension, we predict that the
123 relative contribution of cis- and trans- associated expression variation indicates the
124 structure of response to environment across a molecular network.
125
126 **Results and Discussion**
127 *Stress-responsive genes are non-randomly distributed in a transcriptional co-expression*
128 *network*
129 We first tested the hypothesis that genes whose expression responds to environmental
130 gradients are non-randomly distributed in a transcriptional regulatory network.
131 Specifically, we assessed the positions of stress-responsive genes in a co-expression
132 network estimated for *A. thaliana* by Feltus et al. ((25); hereafter "Feltus network"). We
133 take stress-responsive genes from a previous analyses of two experimental datasets: our
134 "cold" dataset comes from Lasky et al ((24) a re-analysis of Hannah et al. (22)), while our
135 "drought" dataset comes from Des Marais et al. (23). Briefly, the approach taken by Lasky
136 et al. (24) and Des Marais et al. (23) was to partition variance in gene expression level
137 among the effects of genotype (inbred natural accession), environment (response to
138 experimental treatment), and their interaction ("eGxE"). In the context of this paper, stress-
139 responsive genes are those classified as "eGxE", which show a genetically variable response
140 to environmental changes, as well as genes that show a similar pattern of response across
141 natural accessions ("cold-response" or "drought-response" genes).
142 The Feltus network was reconstructed by aggregating data from 7105 published
143 microarray experiments differing in environment, tissue, and genotype; this network
144 hypothesis is therefore a meaningful summary of the transcriptional relationships among
145 genes in diverse environmental settings and genomic backgrounds. For each gene (node) in
146 the network, we calculated the degree of the node, which measures the number of
147 neighboring nodes in the network, as well as the centrality if the node. Centrality of a node
148 estimates the number of paths between other nodes that pass through this node; here, we
149 present node centrality as eigenvector centrality. A gene will have a high eigenvector
150 centrality if it is both well connected itself and if its neighbors are also well connected (see

151 Methods and Newman (26) for more details). We obtained qualitatively similar results
152 with alternative centrality metrics, e.g., degree, betweeness, and closeness centrality (see
153 Supplement).
154 For both data sets, genes showing significant eGxE were non-randomly distributed
155 with respect to network degree and centrality. Drought eGxE genes had lower degrees
156 (Figure 1a) [median for drought: eGxE = 4; non- eGxE = 13], i.e. had fewer connections to
157 other genes, and were less central (Figure 1b) when compared to non- eGxE genomic
158 controls. Cold eGxE genes exhibited the opposite effect, having higher degree (Figure 1c)
159 [median cold: eGxE = 38; non- eGxE = 11] and being more centrally located (Figure 1d)
160 compared to genomic controls. Statistical significance was determined by selecting a
161 random subset of genes equal in size to the genes showing eGxE and then calculating their
162 degree and centrality (see Supplement). Out of 10,000 permutations, we did not observe a
163 single set of genes with more extreme low (drought) or high (cold) distributions of degree
164 and eigenvector centrality, corresponding to a p-value of 10^{-4}. Interestingly, cold eGxE
165 genes had higher average connectivity than that of cold or drought response (environment)
166 genes, of which we have hypothesized represent the conserved environmental response of
167 a species (23, 24) (see Supplement).
168 To verify the robustness of our results, we confirmed the non-random distribution
169 of genes using community detection. The above analysis using the Feltus network makes
170 the assumption that all genes in the network are equally likely to participate in
171 environmental response. However, because certain pathways *must* be involved in the
172 phenotypic response to drought or cold, it is possible that the appropriate null distribution
173 should be constructed using nearby genes (e.g. those with strong co-expression in the
174 Feltus network). To explore this possible statistical artifact, we performed standard
175 community detection using the leading eigenvector method (27) as implemented in the R
176 package igraph v.1.0.1 (28) on the Feltus network and subdivided the graph into sets of
177 genes that are densely connected among themselves and loosely connected to other parts
178 of the gene co-expression network. Averaging the results across all detected communities,
179 with significance again determined by permutation test, we recover the same pattern seen
180 in the global network: Cold eGxE genes have higher degree (cold eGxE genes had a median
181 of 34 more connections than non eGxE genes) and a significantly higher median

182 eigenvector centrality (median 0.14 above non eGxE), while drought eGxE genes have
183 lower degree (cold eGxE genes had a median of 16 fewer connections than non eGxE genes)
184 and had a significantly lower eigenvector centrality (median 0.04 below non eGxE). A
185 similar result was found using the subnetworks (their "GILs") previously constructed by
186 Feltus (see Supplement).
187 We further validated this result by performing iterative out-of-sample model
188 validation. Briefly, we randomly selected 80% of genes in the network and constructed a
189 generalized linear model with a binomial error distribution (i.e. a logistic regression) to
190 predict genes as eGxE based solely on their degree and eigenvector centrality. We then
191 predicted the eGxE state for the remaining 20% of genes and recorded the error. We
192 repeated this procedure 1,000 times for both cold and drought. Assuming a threshold for
193 accurate classification of 5%, we were able to correctly classify 95.4% of genes for cold and
194 77.0% of genes for drought. These results accommodate classification errors for both eGxE
195 and non- eGxE genes, which means that for cold we were able to correctly classify nearly
196 every gene included in the co-expression network as being eGxE based solely on its degree
197 and eigenvector centrality.
198
199 *eGxE genes show modular distribution that differs between environments*
200 We next asked whether the non-random distribution of node connectivity of eGxE
201 genes reflects their membership in particular sub-communities, or modules, of interacting
202 genes as identified using our community detection approach, described above.
203 Interestingly, both the cold and drought eGxE genes were non-randomly distributed with
204 respect to the sub-communities defined using our community detection approach. For
205 cold, 32.5% of all eGxE genes exist within a single, large sub-community containing 605
206 genes (Figure 2a) and an additional 26.5% of cold eGxE genes are found in a second large
207 sub-community containing 425 genes. In contrast, for drought, the two sub-communities
208 with highest accumulation of eGxE genes together contain only 18% of the total number of
209 eGxE genes (Figure 2b shows the larger of these two). Moreover, drought eGxE are
210 statistically over-represented in five small sub-communities comprised of between 10 and
211 100 members (Figure 3b), while cold eGxE genes are clustered in a few large communities
212 (Figure 3a). The membership of eGxE genes in sub-communities of differing size

213 recapitulates our earlier result: cold eGxE genes tend to be functionally connected to many
214 other genes, while genes involved in drought response tend to be in peripheral network
215 positions.
216 To test the hypothesis that the sub-communities with diverging patterns of
217 expression eGxE reflect natural variation in function, we took the genes in the two sub-
218 communities with the most overrepresentation in cold and in drought response and tested
219 for enrichment of gene ontology (GO) annotations. We found 116 significant GO terms
220 enriched in the most over-represented cold eGxE sub-community. The top seven terms
221 were all related to photosynthesis and related processes, and the next two terms were for
222 response to abiotic stimulus and response to cold (Table S3). Altered primary metabolism
223 is frequently observed during cold acclimation, in part due to the accumulation of sugars as
224 cryoprotectants (29, 30), so this eGxE may reflect that some of the sampled accessions
225 modify metabolism during cold response to a different degree than do other accessions. We
226 found 89 significant GO terms enriched in the most over-represented drought eGxE sub-
227 community. The top term and many of the subsequent terms were for immune and defense
228 responses (Table S4; many genes, particularly kinases, annotated as immune and defense
229 responses also show responses to abiotic stress (31)). Previously, we found that drought
230 eGxE genes showed very few significant functional enrichments using a genome-wide test
231 for statistical enrichment (23), suggesting that the network-informed approach used here
232 may afford additional statistical power to detect functional patterns in these high-
233 dimensional datasets.
234
235 *The evolution of gene expression response to the environment*
236 Previously, we demonstrated an important role of cis-regulatory variants
237 underlying diversity of environmental response among natural genotypes of Arabidopsis
238 (24). Natural variation in response to drought showed different genomic patterns than did
239 natural variation in response to cold, suggesting that natural selection may affect different
240 parts of the transcriptional regulatory networks for these two complex traits. Specifically,
241 the proximal promoters of genes showing eGxE for drought had significantly higher
242 nucleotide diversity and significantly higher among-genotype variation in key drought-
243 responsive promoter motifs (abscisic acid responsive elements, ABREs) when compared to

244 genome averages (24). These earlier observations for eGxE drought genes are consistent
245 with the results presented here: drought eGxE genes are in smaller modules and are
246 relatively lowly connected to other genes, suggesting that genetic variation in expression
247 response to the environment is controlled locally, possibly by a large number of cis-acting
248 variants. This architecture may permit functionally diverse modules to act independently
249 from one another, i.e. showing environmental response in only some genotypes (32). Our
250 results may also explain why expression QTL (eQTL) studies of drought response identify a
251 preponderance of cis- acting eQTL and few trans- eQTL (33, 34).
252 Collectively, these observations suggest two, non-exclusive, hypotheses for how
253 natural populations of plants adapt to variation in soil water content. First, diverse
254 populations and species acclimate to transient soil drying stress in diverse ways -- via
255 changes in growth, transpiration, leaf area-volume ratios, timing of reproduction, cell wall
256 composition, and synthesis of various osmoprotectants and chaperonins, to name but a few
257 (35). In this model, the modular regulatory architecture observed here reflects functionally
258 different transcriptional modules driving these different physiological responses. The
259 extent to which such physiological alterations are under independent or common genetic
260 control is presently unknown. Second, because most plants likely experience fluctuations in
261 water availability on a daily or seasonal basis (36-38), plants must maintain some degree of
262 physiological response to drying. Conditionally-active regulatory alleles may therefore
263 allow for fine-tuning of response because they could provide fitness benefits in one
264 environment without deleterious consequences in second environment ("yield-drag";
265 (39)). Indeed, cis-acting eGxE variants for drought response are predominantly variance-
266 changing (i.e. conditionally neutral or differentially sensitive to conditions) rather than
267 direction-changing (33). Moreover, the vast majority of QTL showing GxE in studies of
268 abiotic stress are also conditionally neutral, or show differential sensitivity to
269 environmental cues (17).
270 Nucleotide diversity in the proximal promoters of cold eGxE genes is also elevated
271 but, in contrast to drought genes, not to a degree that is statistically significant compared to
272 genome averages (24). The promoters of cold eGxE genes exhibit lower among-genotype
273 turnover of known cold-responsive motifs compared to genome averages (specifically, the
274 c-repeat binding factor/dehydration responsive elements, CRT/DREs). Along with our

275 observation that cold eGxE genes are highly connected in the transcriptional regulatory
276 network, these patterns of sequence diversity suggest that the transcriptional control of
277 eGxE for cold acclimation is driven by genetic variants in upstream regulatory features,
278 such as transcription factors. Natural variants in transcription factors are expected to have
279 larger mutational effect size due to the regulatory influence of these proteins on
280 downstream genes.

281 It is currently unknown what drives the apparent difference in the genetic
282 architecture of natural variation in response to drought compared to response to cold.
283 While daily or seasonal drying stress is likely experienced by *A. thaliana* plants to some
284 degree across the species range, severe cold stress is likely only experienced by
285 populations at higher latitudes or altitudes. A recent study demonstrated that multiple,
286 apparently independent, loss of function mutations in key transcriptional regulators of
287 cold-responsive genes are associated with geographic variation in winter temperature
288 across the range of *A. thaliana* (40). The activity of these CBF transcription factors shows a
289 strong positive correlation with the capacity of *A. thaliana* natural accessions to acclimate
290 to cold (22, 41, 42). Perhaps cold-associated selective gradients involve sharp transitions
291 (along spatial and climatic gradients) in optimal phenotypes, with only a few fitness optima
292 for cold tolerance – manifested as cold-tolerant, functionally CBF versus cold-intolerant,
293 non-functional CBF accessions. By contrast, drought tolerance may be subject to smoother,
294 more gradual selective gradients, for which many smaller-effect variants allow fine-tuning
295 of response.
296
297 *Conclusions*
298 Our results suggest that topological relationships among genes in transcriptional
299 regulatory networks affect how natural populations adapt to the multivariate environment.
300 A promising extension of our approach is to link information regarding the topological
301 features of a given gene – its connectivity, in the case presented here, as well as its
302 membership in particular functional modules – with information associating genetic
303 variants with phenotype from genetic mapping. Such a combined analysis could clarify how
304 putatively functional variants identified via association mapping result in phenotypic
305 variation via cellular and physiological mechanisms (43).

306 From an applied perspective, understanding how natural variation affects
307 transcriptional regulatory networks may inform decisions about how to improve
308 agricultural performance in challenging environments. We note that breeding for improved
309 performance under soil drying has been quite challenging (44); our results suggest that
310 targeting specific physiological mechanisms by manipulating genes at the "tips" of
311 regulatory networks, shown herein to exhibit drought eGxE, may be a more fruitful strategy
312 than targeting central regulatory molecules which may exhibit undesirable pleiotropic
313 effects ("yield drag;" e.g. (45)).
314
315 **Methods**
316 *Co-expression and regulatory networks*
317 We used two published datasets on gene expression interactions in in Arabidopsis thaliana.
318 The first dataset was global (i.e. genome-wide and not restricted to certain functions or
319 pathways). We used the global co-expression network and 86 subcomponent genome-wide
320 gene coexpression networks created by (25). The authors first obtained 7,105 publicly
321 available ATH1 Affymetrix microarray samples and applied a thresholding algorithm
322 (random matrix theory) to generate a global network containing 3,297 nodes and 129,134
323 edges. These nodes represent 16% of Arabidopsis genes on the ATH1.
324
325 A complication from this approach arises because of interactions between genotype,
326 expression networks and environment (including ontogeny, tissue, or cell type; (46)). Thus
327 a prior step of partitioning expression data may help to account for some of this
328 heterogeneity and better reveal co-expression networks. Through k-means partitioning,
329 Feltus et al. (25) generated 86 gene interaction layers (GILs), i.e. 86 smaller, non-global
330 networks, that together included 19,588 genes, which represents 95% of Arabidopsis genes
331 on the ATH1.
332
333 *Data: transcriptomic responses to cold and drought stress*
334 We used two published studies on natural variation in transcriptomic response to cold (22)
335 and drought (23). Both studies used the ATH1 microarray to estimate genome-wide
336 transcript abundance. Each study subjected a diverse panel of 9 (22) or 17 (23) natural

accessions to a cold or drought treatment, respectively. Lasky et al. (2014) re-analyzed the Hannah et al. dataset to match the analyses by Des Marais et al. In brief, those authors modeled transcript abundance using factorial ANOVA including Accession (i.e. Genotype), Treatment, and their interaction and identified significantly differentially expressed gene models at pFDR of 0.05.

Network Methods

We quantified the degree to which a gene was central using four standard network metrics, 1.) degree (raw number of connections), 2.) closeness centrality (the inverse of the average shortest path between the focal gene and all other genes in the network), 3.) betweenness centrality (number of shortest paths between all pairs of nodes in the network, which pass through the focal node), and 4.) eigenvector centrality. Eigenvector centrality, which is closely related to Google's PageRank algorithm (47) measures a node's centrality based on both the node's own position in the network and the position of that node's neighbors in the network. More specifically, a node's eigenvector centrality will be proportional to the average centralities of its neighbors (26)

To identify community structure and assign genes to communities, we used the leading eigenvalue algorithm. Our goal was to determine whether there are groups of genes, which are more connected to each other than they are to other genes, referred to as community detection (48). A variety of methods exist for performing community detection, but we selected the leading eigenvalue approach because it is computationally efficient on large networks. Briefly, the adjacency matrix of the network is corrected based on the expected number of edges in a random graph, using the configuration model, then the distribution of eigenvalues and the loading of nodes onto eigenvectors can be used to 1.) determine whether evidence exists for the presence of modular communities and 2.) assuming such structure exists, assign genes to communities, see Newman 2006 for more details. All of the analyses, i.e. calculation of centrality measures and community detection, was performed using the R package igraph v.1.0.1 (28).

Enrichment analyses

368 We tested whether genes with multiple types of expression response to abiotic stress
369 exhibited non-random network metrics. We also tested whether genes exhibiting high Fst
370 exhibited non-random network metrics. Determining whether a gene has a higher value for
371 any of these metrics is not appropriate for parametric stats. Therefore, we conducted
372 permutations to generate null expectations. The natural accessions used in this study all
373 show varying degrees of sequence divergence and gene gain/loss as compared to the
374 reference Col-0 genome, which was used to generate the microarrays used in these
375 experiments. This variance could generate spurious "gene-by-environment interaction" for
376 gene expression. We therefore used a strict filtering scheme to exclude genes that had
377 polymorphisms in ATH1 probe sites (23). In order to assess the biological function of
378 regulatory communities, we first identified communities containing the greatest
379 proportion of GxE genes for each abiotic stressor. We then tested gene ontology (GO) term
380 enrichment in each of these communities (AgriGO).
381

382 **Acknowledgements**

383 We thank R. Hopkins and E. Josephs for comments on the manuscript. T. Juenger and J. G.
384 Monroe provided stimulating discussion on the genetic architecture of abiotic stress
385 response and expression GxE. This work is a product of the "Molecular Networks and
386 Evolution Across Biological Scales" workshop supported by the Santa Fe Institute.
387

388 **References**

389

390 1. Wagner GP & Zhang J (2011) The pleiotropic structure of the genotype–phenotype
391 map: the evolvability of complex organisms. *Nature Reviews Genetics* 12(3):204-213.
392 2. Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon Press,
393 Oxford).
394 3. Erwin DH & Davidson EH (2009) The evolution of hierarchical gene regulatory
395 networks. *Nature Reviews Genetics* 10:141-148.
396 4. Jeong H, Mason SP, Barabasi AL, & Oltvai ZN (2001) Lethality and centrality in
397 protein networks. *Nature* 411:41.

398 5. Rausher MD, Miller RE, & Tiffin P (1999) Patterns of Evolutionary Rate Variation
399 Among Genes of the Anthocyanin Biosynthetic Pathway. *Molecular Biology and*
400 *Evolution* 16(2):266-274.
401 6. Guelzim N, Bottani S, Bourgine P, & Kepes F (2002) Topological and causal structure
402 of the yeast transcriptional regulatory network. *Nat Genet* 31(1):60-63.
403 7. Lin W-h, Liu W-c, & Hwang M-j (2009) Topological and organizational properties of
404 the products of house-keeping and tissue-specific genes in protein-protein
405 interaction networks. *BMC systems biology* 3:32.
406 8. Wilkins AS (2002) *The evolution of developmental pathways* (Sinauer, Sunderland,
407 MA).
408 9. Wray GA*, et al.* (2003) The evolution of transcriptional regulation in eukaryotes.
409 *Molecular Biology and Evolution* 20(9).
410 10. Wagner GP, Pavlicev M, & Cheverud JM (2007) The road to modularity. *Nat Rev*
411 *Genet* 8(12):921-931.
412 11. Gibson G (2008) The environmental contribution to gene expression profiles.
413 *Nature Reviews Genetics* 9:575-581.
414 12. Weake VM & Workman JL (2010) Inducible gene expression: diverse regulatory
415 mechanisms. *Nature Reviews Genetics* 11:426-437.
416 13. Fowler S, Cook D, & Thomashow MF (2007) The CBF cold-response pathway. *Plant*
417 *Abiotic Stress*, eds Jenks MA & Hasegawa PM (Wiley-Blackwell, Chichester), pp 71-
418 99.
419 14. Luscombe NM*, et al.* (2004) Genomic analysis of regulatory network dynamics
420 reveals large topological changes. *Nature* 431:308-312.
421 15. McCarroll SA*, et al.* (2004) Comparing genomic expression patterns across species
422 identifies shared transcriptional profile in aging. *Nat Genet* 36(2):197-204.
423 16. Carvallo MA*, et al.* (2011) A comparison of the low temperature transcriptomes and
424 CBF regulons of three plant species that differ in freezing tolerance: *Solanum*
425 *commersonii*, *Solanum tuberosum*, and *Arabidopsis thaliana*. *J Exp Bot* 62(11):3807-
426 3819.

427 17. Des Marais DL, Hernandez KH, & Juenger TE (2013) Genotype-by-environment
428 interaction and plasticity: exploring genomic responses of plants to the abiotic
429 environment. *Annual Review of Ecology, Evolution and Systematics* 44:5-29.
430 18. Schlichting CD & Pigliucci M (1998) *Phenotypic Evolution: A Reaction Norm*
431 *Perspective* (Sinauer Associates, Sunderland, MA).
432 19. Via S & Lande R (1985) Genotype-environment interaction and the evolution of
433 phenotypic plasticity. *Evolution* 39:505-523.
434 20. Kawecki TJ & Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*
435 7:1225-1241.
436 21. Lynch M & Walsh B (1998) *Genetics and Analysis of Quantitative Traits* (Sinauer,
437 Sunderland, MA).
438 22. Hannah MA*, et al.* (2006) Natural genetic variation of freezing tolerance in
439 Arabidopsis. *Plant Physiol* 142(1):98-112.
440 23. Des Marais DL*, et al.* (2012) Physiological genomics of response to soil drying in
441 diverse Arabidopsis accessions. *Plant Cell* 24(3):893-914.
442 24. Lasky JR*, et al.* (2014) Natural variation in abiotic stress responsive gene expression
443 and local adaptation to climate in *Arabidopsis thaliana*. *Molecular Biology and*
444 *Evolution* 31:2283-2296.
445 25. Feltus FA, Ficklin SP, Gibson SM, & Smith MC (2013) Maximizing capture of gene co-
446 expression relationships through pre-clustering of input expression samples: An
447 *Arabidopsis* case study. *BMC systems biology* 7:44.
448 26. Newman ME (2008) The mathematics of networks. *The new palgrave encyclopedia*
449 *of economics* 2(2008):1-12.
450 27. Newman ME (2006) Finding community structure in networks using the
451 eigenvectors of matrices. *Phys Rev E Stat Nonlin Soft Matter Phys* 74(3 Pt 2):036104.
452 28. Csardi G & Nepusz T (2006) The igraph software package for complex network
453 research. *International Journal of Complex Systems* 1695(5):1-9.
454 29. Thomashow M (1999) Plant cold acclimation: freezing tolerance genes and
455 regulatory mechanisms. *Annual Reviews of Plant Physiology and Plant Molecular*
456 *Biology* 50:571-599.

457 30. Colton-Gagnon K, *et al.* (2014) Comparative analysis of the cold acclimation and
458 freezing tolerance capacities of seven diploid *Brachypodium distachyon* accessions.
459 *Ann Bot* 113(4):681-693.
460 31. Marshall A, *et al.* (2012) Tackling drought stress: receptor-like kinases present new
461 approaches. *Plant Cell* 24(6):2262-2278.
462 32. Draghi JA, Parsons TL, Wagner GP, & Plotkin JB (2010) Mutational robustness can
463 facilitate adaptation. *Nature* 463(7279):353-355.
464 33. Cubillos FA, *et al.* (2014) Extensive cis-regulatory variation robust to environmental
465 perturbation in Arabidopsis. *Plant Cell* 26(11):4298-4310.
466 34. Lowry DB, *et al.* (2013) Expression quantitative trait locus mapping across water
467 availability environments reveals contrasting associations with genomic features in
468 Arabidopsis. *Plant Cell* 25(9):3266-3279.
469 35. Chaves MM, Maroco JP, & Pereira J (2003) Understanding plant responses to
470 drought -- from genes to the whole plant. *Functional Plant Biology* 30:239-264.
471 36. Richards CL, Rosas U, Banta J, Bhambhra N, & Purugganan MD (2012) Genome-wide
472 patterns of Arabidopsis gene expression in nature. *PLoS Genet* 8(4):e1002662.
473 37. Chaves MM, *et al.* (2002) How Plants Cope with Water Stress in the Field?
474 Photosynthesis and Growth. *Annals of Botany* 89(7):907-916.
475 38. Plessis A, *et al.* (2015) Multiple abiotic stimuli are integrated in the regulation of rice
476 gene expression under field conditions. *Elife* 4.
477 39. Mickelbart MV, Hasegawa PM, & Bailey-Serres J (2015) Genetic mechanisms of
478 abiotic stress tolerance that translate to crop yield stability. *Nat Rev Genet*
479 16(4):237-251.
480 40. Monroe JG, *et al.* (2016) Adaptation to warmer climates by parallel functional
481 evolution of CBF genes in Arabidopsis thaliana. *Mol Ecol* 25(15):3632-3644.
482 41. Gehan MA, *et al.* (2015) Natural variation in the C-repeat binding factor cold
483 response pathway correlates with local adaptation of Arabidopsis ecotypes. *Plant J*
484 84(4):682-693.
485 42. Gery C, *et al.* (2011) Natural variation in the freezing tolerance of *Arabidopsis*
486 *thaliana*: effects of RNAi-induced CBF depletion and QTL localisation vary among
487 accessions. *Plant Sci* 180(1):12-23.

488 43. Rockman MV (2008) Reverse engineering the genotype-phenotype map with
489 natural genetic variation. *Nature* 456:738-744.
490 44. Passioura J (2010) Scaling up: the essence of effective agricultural research.
491 *Functional Plant Biology* 37:585-591.
492 45. Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, & Shinozaki K (1999) Improving
493 plant drought, salt and freezing tolerance by gene transfer of a single stress-
494 inducible transcription factor. *Nature Biotechnology* 17:287-291.
495 46. Fukushima A*, et al.* (2012) Exploring tomato gene functions based on coexpression
496 modules using graph clustering and differential coexpression approaches. *Plant*
497 *Physiol* 158(4):1487-1502.
498 47. Langville AN & Meyer CD (2011) *Google's PageRank and beyond: The science of*
499 *search engine rankings.* (Princeton University Press, Princeton, NJ).
500 48. Fortunato S & Hric D (2016) Community detection in networks: A user guide.
501 *arXiv*:1608.00163.
502

503 **Figure Legends**

504 Figure 1 – Expression GxE genes are non-randomly distributed in the Arabidopsis gene
505 regulatory network. As a group, eGxE genes in drought conditions have significantly lower
506 degree (A) and eigenvector centrality (C) than do genomic controls. As a group, eGxE genes
507 in cold condition have significantly higher degree (B) and eigenvector centrality (D) than
508 do genomic controls. The solid line represents the median, colored boxes indicate the inter-
509 quartile range, and whiskers mark the entire range of the data.
510

511 Figure 2 – Graphical representation of two sub-communities of the Arabidopsis gene
512 regulatory network showing the sub-community with the highest over-representation of
513 cold eGxE genes (A) and the sub-community with the highest over-representation of
514 drought eGxE genes (B). Colored circles indicate genes showing significant eGxE at
515 pFDR=0.05 from (23, 24).
516

517 Figure 3 – Cold (A) and drought (B) eGxE genes have different distributions across sub-
518 communities of the gene regulatory network. Cold eGxE genes are overrepresented in two

519 large sub-communities, each containing over 400 total genes. Five smaller sub-
520 communities (<100 genes) are enriched for drought eGxE genes. The vertical axis is the
521 difference between the observed and expected number of eGXE genes in the community,
522 relative to the total number of genes in the community. Each sub-community in the analysis
523 is shown with a circle (only subcommunities with ten or more genes are included). Green
524 circles indicate sub-communities in which eGxE genes are significantly overrepresented,
525 while orange circles represent subcommunities with significant underrepresentation of
526 eGxE genes (Fisher's exact test; FDR=0.05).
527

Figures

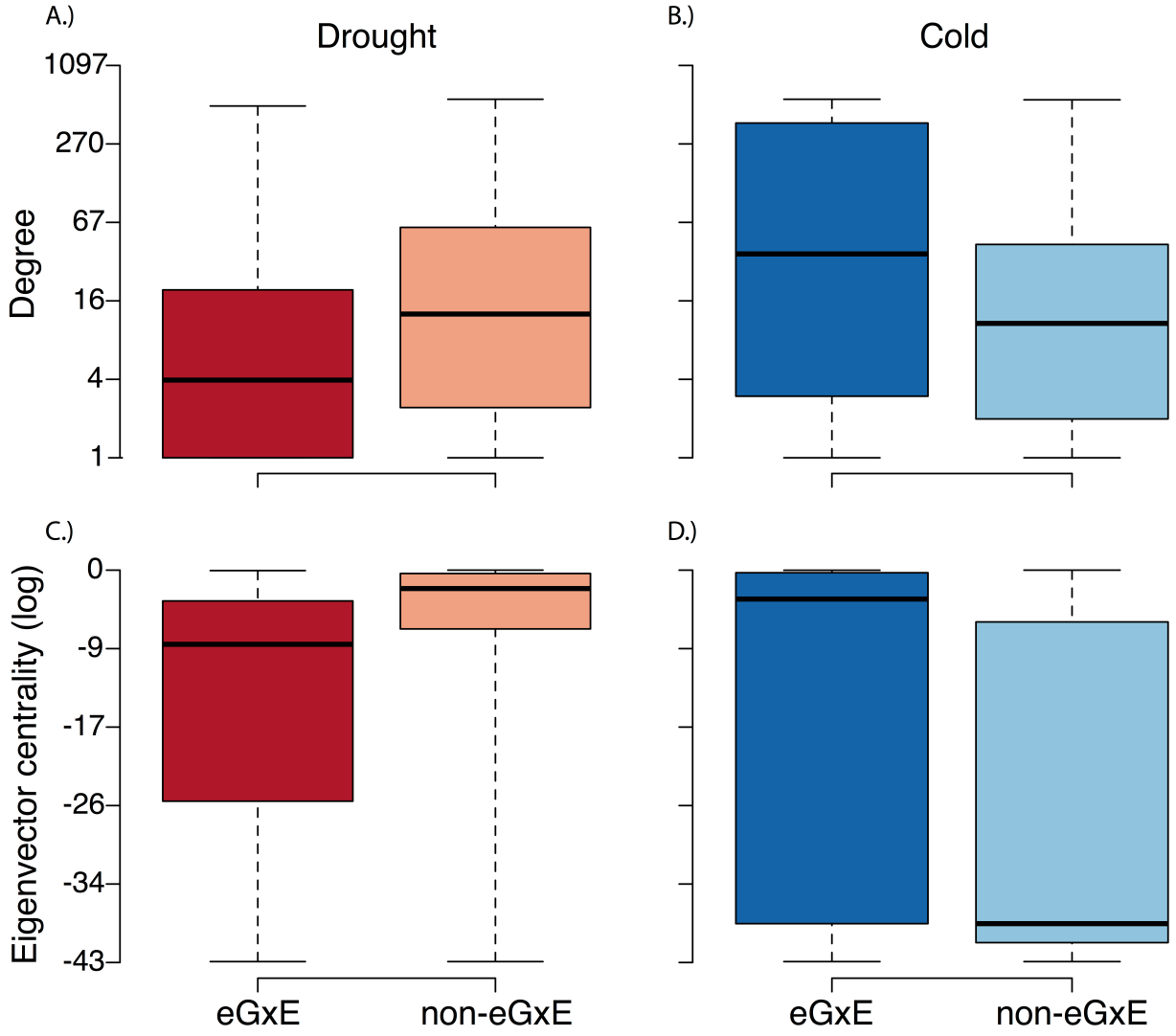

Figure 1

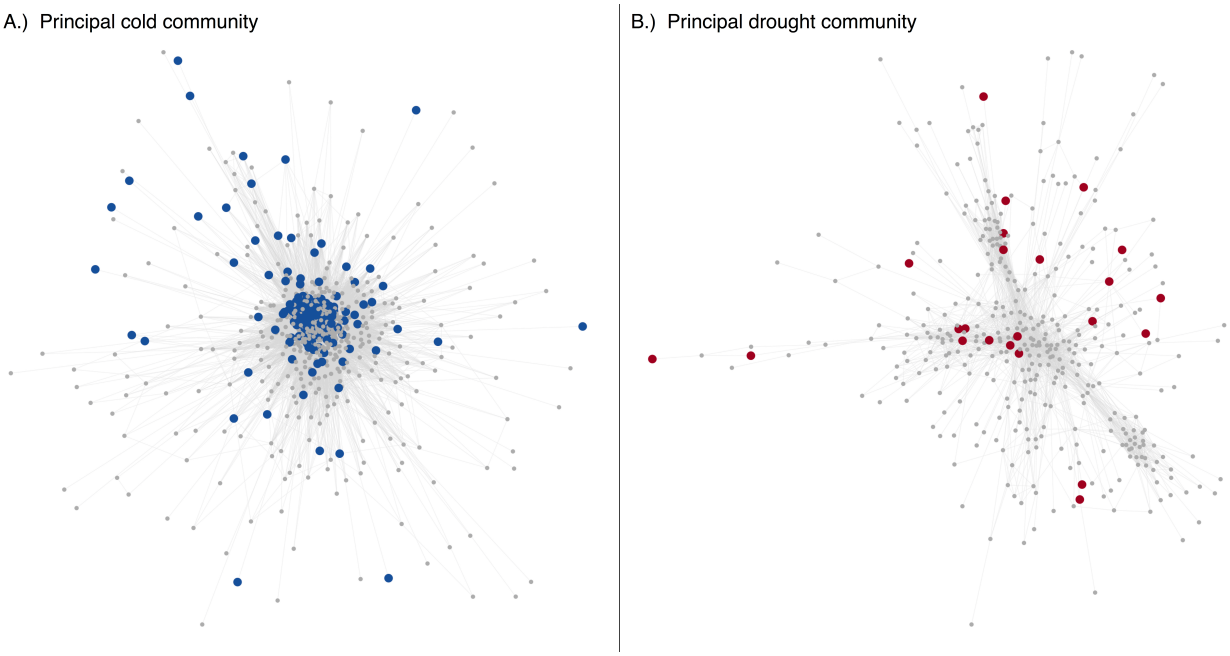

Figure 2

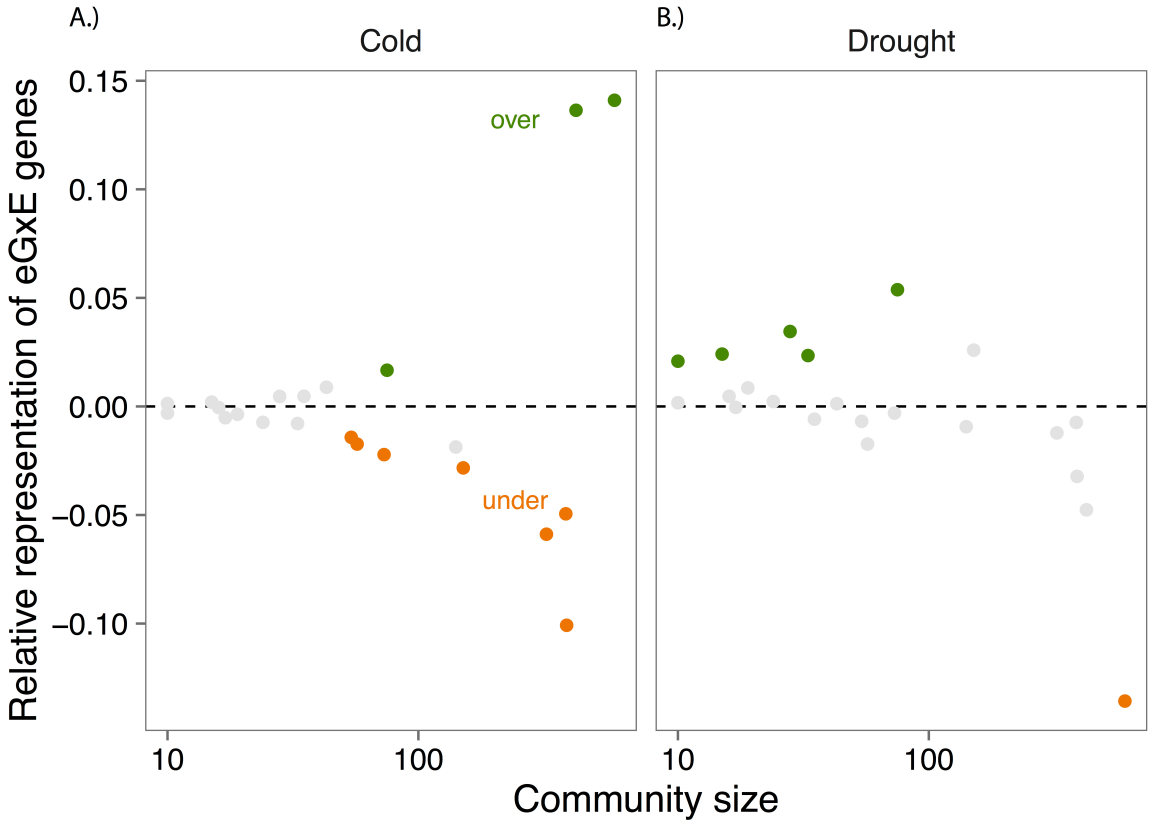

Figure 3