Title: Strength of stabilizing selection on the amino-acid sequence is associated with the amount of non-additive variance in gene expression

Authors: Margarita Takou[1,2], Daniel J. Balick[3], Kim A. Steige[1,4], Hannes Dittberner[1], Ulrike Göbel[1], Holger Schielzeth[5], Juliette de Meaux[1]

Addresses:

1 Institute of Botany, University of Cologne, Cologne, Germany

2 Department of Biology, Pennsylvania State University, PA, USA

3 Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America

4 State Plant Breeding Institute, University of Hohenheim, Stuttgart, Germany

5 Institute of Ecology and Evolution, Friedrich Schiller University Jena, Jena, Germany

Abstract:

Contemporary populations are unlikely to respond to natural selection if much of their genetic variance is non-additive. Understanding the evolutionary and genomic factors that drive amounts of non-additive variance in natural populations is therefore of paramount importance. Here, we use a quantitative genetic breeding design to delineate the additive and non-additive components of expression variance of 17,656 gene transcripts segregating in the outcrossing plant *Arabidopsis lyrata*. We find that most of the genetic variance in gene expression represents non-additive variance, especially among long genes or genes involved in epigenetic gene regulation. Genes with the most non-additive variation not only display a markedly lower rate of synonymous variation, they have also been exposed to stronger stabilizing selection than genes with high additive variance. This study demonstrates that a gene's potential for future evolution is shaped by its past history of stabilizing selection.

Main

Natural populations harbor considerable levels of genetic variance. However, owing to complex interactions between alleles, a fraction of this variance is not available to selection because it is not additively transferred to the next generation (*1*). This can limit the potential of genetically variable traits to respond to selection in the future. On the other hand, past selection has long been suspected to shape the amount of non-additive variance. Theoretical studies predict that selection will deplete additive genetic variance, so that non-additive variance will form an increasing fraction of the remaining genetic variance (*2*). To this date, evidence that the amount of non-additive genetic variance is shaped by past selective pressures remains elusive (*3*). Studies that estimate the amount of non-additive variance and the strength of selection simultaneously are sparse (*3–6*). Furthermore, other factors may determine the relative amount of non-additive genetic variance present in a population. A recent bottleneck, for example, will increase the proportion of additive genetic variance, because recessive alleles tend to increase in frequency, and variation at interacting loci tends to become lost (*7–11*). Besides, the amount of non-additive variance present in populations may also depend on the genetic architecture of phenotypic variation. As the number of loci controlling a phenotype increases, so will the likelihood that new allelic variants interact with other variants of the same or other loci to define a new phenotypic value; this new value deviates from the sum of effects of the individual variants and generates non-additive genetic variance (*12*).

Because analyses of additive and non-additive variance in natural populations focus on a handful of traits, they necessarily lack the statistical power. Power can be improved by the analysis of transcript expression, because it allows screening thousands of genes simultaneously (*13–15*). When quantified over the appropriate pedigree of full- and half-sib families (*1*, *12*, *16*), transcriptome studies can reveal the considerable levels of non-additive genetic variance that characterize expressed genes (*17*). Reconstructing how variation in these genes impacts fitness in natural conditions is challenging (*15*, *18*), but population genomics methods can be leveraged to determine how the components of genetic variance in expression were shaped by selection over evolutionary time. Indeed, with the advent of population genomics datasets, sophisticated approaches have been developed to identify genomic regions targeted by recent positive

54 selection, or to quantify the strength of stabilizing selection on variation affecting the amino-acid
55 sequence of transcribed genes (*19–21*).

56 Here, we analyzed expression variance of 17,656 genes expressed during the exponential
57 growth phase of the outcrossing plant species *Arabidopsis lyrata*. In this system, we can
58 rigorously quantify the additive and non-additive components of genetic variance because it
59 allows controlling for environmental and maternal sources of variance. Combining this
60 information with population genomics estimates of natural selection and functional genomics
61 data, we tested whether the amount of non-additive genetic variance in present-day natural
62 populations associates with the history of selection in the population.

63

64 **Results**

65 **Predominance of non-additive genetic variance in gene expression**

66 We took advantage of a set of 10 full- and half-sib families of the outcrossing plant species
67 *Arabidopsis lyrata* ssp. *petraea*, which were obtained by crossing individuals from two locally
68 adapted populations (*22*, *23*) (Supplementary Table S1). Given the amount of shared variation
69 between the populations and their similar frequency distribution, this set of inter-population
70 hybrids is expected to show levels of non-additive variance similar to those within populations
71 (Supplementary Figure S1)(*24*). We collected leaf material during the exponential rosette
72 growth phase of the plants and sequenced their transcriptome. Full- and half-sib relationships
73 allowed us to partition the total phenotypic variance (V_P) of gene expression into additive
74 variance (V_A), non-additive (V_D), maternal (V_M), and residual (V_R) components of variance in
75 the expression for each transcript in the genome (Figure 1a, Supplementary Table S1). The main
76 non-additive component is often termed dominance variance (V_D) in quantitative genetics
77 studies, and although strictly defined for intra-locus interactions, the use of the term typically
78 also includes inter-locus allelic interactions that give rise to epistatic genetic variance. Indeed,
79 the statistical variance due to intra- and inter-locus interactions, i.e. dominance and epistatic
80 variance, cannot easily be disentangled. In 67.7% of the 17,656 genes analyzed, more than 50%
81 of total expression variance (V_P) was explained by the significant contribution of the two genetic
82 components of variance, V_D (mean = 0.375) and V_A (mean = 0.215). The maternal variance (V_M)

83 was close to zero for most transcripts (mean = 0.088). Finally, the origin of the maternal plant
84 did not impact gene expression (all p > 2.486e-06, the Bonferroni corrected significance
85 threshold).

86 Delineating overall genetic variance showed that V_D contributed the most (25.4% of 17,656
87 genes had more than 50% of gene expression variance V_P due to V_D), in contrast to the additive
88 genetic component V_A (6.3% of these genes had more than 50% of their variance in expression,
89 Figure 1b-c). Hence, gene expression variance V_P is due largely to genetics, but only a limited
90 fraction of that genetic variance is inherited additively. Although amounts of non-additive
91 genetic variance are rarely quantified in natural populations, the amount we detected is not
92 unexpected. Similar fractions of non-additive variance have been reported in phenotypic traits
93 (*25*), and non-additive variance was as high if not higher than additive variance for most
94 transcripts expressed in threespine stickleback (*17*). In a control dataset generated by permuting
95 individuals across families, V_R explained more than half of the V_P for more than 73% of
96 transcripts (mean = 0.717), and V_D was not significant (mean = 0.135), confirming that the levels
97 of genetic variance we report are not random.

98 As expected for a transcriptome dataset, many genes formed co-expressed clusters of variable
99 size (Supplementary Figure S2). Yet, major genetic incompatibilities or large-effect variants
100 were unlikely to explain the data (*26*). Indeed, a genome-wide association showed that most of
101 the genetic variation in expression had a complex genetic basis (see Supplementary Information).

102 **Strength of stabilizing selection associates with inheritance of regulatory variation**

103 Having determined levels of non-additive genetic variance in gene expression, we then used a
104 population genetics method to determine whether the mode of gene expression inheritance was
105 associated with past selective pressures. We first examined rates of amino-acid evolution in the
106 transcripts that were expressed in our dataset. We partitioned expressed transcripts in three non-
107 overlapping sets based on their relative excess of V_D, V_A, or V_R (Figure 2a). As a control, we
108 included a fourth set of 3,500 genes sampled at random (Fig. 2a). We then estimated the rate of
109 amino acid divergence that evolved between *A. lyrata* and *A. thaliana* across these four groups of
110 genes with different inheritance. The four sets of transcripts differed significantly in their rate of
111 amino-acid divergence. Genes with high V_R had the lowest rate of amino-acid divergence,
112 whereas genes with highest V_A had the fastest rate (Figure 2b-c). Interestingly, genes with the

113 largest V_D displayed a rate of amino-acid divergence that was intermediate between those shown
114 by genes with large V_A and genes with large V_R.

115 The rate of amino-acid divergence is expected to increase if positive selection is acting on the
116 encoded protein, but increases may also occur if the evolutionary constraints on the amino-acid
117 sequence are weak. In order to differentiate these alternatives, the segregation of molecular
118 variation within species can be used to infer the fitness effect of new mutations and quantify the
119 strength of the evolutionary constraints imposed on the amino-acid sequence. We took advantage
120 of a previous study, which determined the demographic history of the same populations of *A.*
121 *lyrata* (*23*, *27*). We inferred the distribution of fitness effects (DFE) of new non-synonymous
122 variants in these gene sets, following a method described in (*23*) that incorporates the specific
123 demographic parameters of each population (Supplementary Figure S3). Both the inferred
124 demographic model and the DFE fitted the site-frequency spectra (SFS) of synonymous and non-
125 synonymous variants well (Supplementary Figure S3a), indicating that the model provides a
126 good representation of the evolutionary and selective history of these populations.

127 The results showed that the inferred strength of evolutionary constraint (Figure 2d;
128 Supplementary Table S2) co-varies clearly with differences in the mode of inheritance that
129 characterizes the expression variance of the transcript. Compared to the other sets of genes, the
130 set of genes with low genetic variance in expression (i.e., high residual variance, V_R) showed a
131 significantly lower proportion of nearly neutral non-synonymous mutations ($0 < N_{anc}s < 1$, $p =$
132 0.01), as well as a significantly higher proportion of genetic variants under strong purifying
133 selection ($N_{anc}s > 10$, $p = 0.01$). This observation indicates that genes evolving under strong
134 constraint at the amino-acid level also tended to show low levels of genetic variation at the
135 transcript expression level.

136 Our analysis further revealed that selective constraints on genes with large V_A (threshold 0.5) are
137 comparatively lower than on genes with high V_D (Figure 2c, Supplementary Figure S3b-c). The
138 DFE analysis showed that, compared to other gene sets, genes with large V_A in transcript
139 expression hosted a significantly larger fraction of new non-synonymous mutations that are
140 nearly neutral ($0 < N_{anc}s < 1$, $p= 0.03$). At the same time, a significantly smaller fraction of new
141 mutations was under strong purifying selection in genes with large V_A ($N_{anc}s > 10$, $p = 0.01$).
142 Furthermore, genes with high V_A were less likely to map to genomic regions harboring footprints

143 of selective sweeps. Although regions with signatures of selective sweeps extended over 27% of
144 the genome, only 17% of the high V_A genes mapped to these regions (pairwise t-test, p= 2.115e-
145 12), further confirming that these genes are unlikely to contain variants that have been under
146 directional selection. Therefore, both the high rate of amino-acid divergence observed in these
147 genes and their inferred DFE indicate that genes with additive variance in gene expression tend
148 to be exposed to lower levels of purifying selection than genes from all other sets. In fact, genes
149 with high V_A also tend to show a higher mutation rate (Supplementary Information,
150 Supplementary Figure S3).

151 The evolutionary constraints detected on the amino-acid sequence of genes with high levels of
152 V_D were high for genes with high V_A but low for genes with high V_R, indicating that non-
153 additive genetic variation was highest in genes experiencing moderate stabilizing selection
154 pressure. On average, the constraints detected in genes with high V_D were comparable to those
155 experienced throughout the genome (Fig. 2c, Supplementary Figure S3d). Interestingly, the rate
156 of nucleotide divergence at synonymous sites, Ks, which reflects the mutation rate, was
157 markedly lower in genes with high V_D, compared to all other gene groups (Supplementary Figure
158 S3).

159 **Gene architecture associates with the level of non-additive variance in transcripts**

160 In order to better understand whether genomic factors modulate the relative fractions of non-
161 additive and additive variance, we compiled a total of 200 descriptors of genetic variation at each
162 gene; these included the number of known transcription-factor binding sites in the upstream non-
163 coding region, the length of the genes, and various descriptors of nucleotide diversity (see
164 Methods). We trained a random forest model to identify the genomic architecture or population
165 genetic parameters that best explain variation in the prevalence of V_A or V_D in gene expression.
166 We first performed the analysis using the 200 descriptors to predict variation in V_A, but the
167 model's ability to classify high V_A genes was poor (R^2 = 0.009 and mean square error of 0.021).

168 For variation in V_D, however, the model was accurate at classifying genes with high V_D (R^2 =
169 0.18 and mean square error of 0.026), highlighting that the most important factors for classifying
170 genes according to their level of V_D are descriptors of gene structure (Figure 3a; Supplementary
171 Table S3). Specifically, the number of exons per gene (importance: 53.021; p = 0.0164), total
172 gene length (importance: 26.92; p = 0.0164), and transcript length (importance: 14.27; p =

173 0.0164) were the three factors that best predicted levels of non-additive variance. The importance
174 of exon number persisted even when gene length was corrected for. Transcript length was
175 significantly correlated with V_D (Spearman coefficient $\rho = 0.258$, $p = 1.22e-266$, Supplementary
176 Table S4; Figure 3b). We used a linear model to confirm that the effect of transcript length on V_D
177 ($p < 2.2e-16$) remained significant even after correcting for median gene expression ($p = 0.0176$)
178 and V_G ($p < 2.2e-16$). The model did not suggest a significant interaction between transcript
179 length and median gene expression ($p = 0.224$). The number of exons per gene correlated
180 significantly with the level of non-additive variance ($\rho = 0.358$, $p < 2.2e-16$; Figure 3c). The
181 random forest analysis further revealed that the density of SNPs within a gene was a predictor of
182 high V_D. Strikingly, all significant descriptors tend to increase the mutational target size of the
183 transcript. We thus conclude that non-additive variance is likely to be observed as the mutational
184 target size increases.

185 Interestingly, the analysis further revealed that the total number of transcription-factor binding
186 sites explained V_D well ($p = 0.0163$; importance 4.34; Figure 3d). When we corrected for gene
187 length, the total number of transcription-factor binding sites becomes the most important
188 predictor in the random forest analysis (Importance 28.72; $p = 0.0163$), which explains the level
189 of V_D in the dataset. In particular, the transcription factor binding sites associated with high V_D
190 were targets of a DOF zinc finger protein expressed in young rosettes (DOF 5.3; $p = 0.0163$);
191 this protein was recently shown to determine the stem cell division that is necessary for plant
192 thickening (*28*) (AT5G60200 TMO6; $p = 0.01639$). This regulator was likely active in the tissue
193 samples we collected, because we sampled during the exponential phase of rosette growth.

194 We further quantified the distribution of genes with high V_D across gene ontology (GO)
195 categories of expressed genes. We found that genes with high V_D were enriched for GO
196 categories reflecting cell differentiation. The top enriched GO terms for the highest V_D estimates
197 were "chromatin silencing" ($p = 4.5e-16$) and "cytokinesis by cell plate formation" ($p = 2e-11$),
198 "DNA methylation" ($p = 4.3e-11$; Supplementary Table S5). In addition, genes with high V_A
199 were not randomly distributed among molecular functions; but, unlike V_D genes, the functional
200 enrichment of high V_A genes was less striking and not related to cell differentiation. Among the
201 top GO categories, the most enriched among genes with high V_A values were several functions
202 linked to central metabolism such as "glucose catabolic process" ($p = 8.3-05$) and "glycolytic

203 process" (p = 1.6e-05), but we also detected enrichments among genes involved in the "response
204 to cadmium ion" (p = 0.00015) and "proteasome core complex assembly" (p = 0.00016;
205 Supplementary Table S6).

206 **Non-additive variance increases with the size of the co-regulatory network**

207 Our results indicate that two processes may increase the opportunities for the inter-allelic
208 interactions that favor the emergence of non-additive variance: the mutational target size of a
209 transcript (its length and number of exon-intron boundaries) and the complexity of the regulatory
210 network. The observations that transcription-factor binding sites as well as the DOF zinc finger
211 protein TMO6 contribute significantly to the amount of non-additive variance in gene expression
212 confirm this observation. More generally, we found that the median V_D in a cluster of co-
213 regulated genes increased with the number of genes contained in a cluster (ρ = 0.49, p = 1.253e-
214 13, Supplementary Table S7); additionally, the correlation between the gene number per cluster
215 and median V_A was negative (ρ = -0.277, p = 7.101e-05). The median V_D per cluster was
216 positively correlated with the median transcript length (ρ = 0.201, p = 0.0042) and median gene
217 exon number (ρ = 0.439, p = 7.752e-11). Accordingly, our data suggest that both inter-allelic and
218 inter-genic interactions increase the amount of non-additive variance in gene expression.

219 **Discussion**

220 Delineating the genomic and evolutionary factors that increase the non-additive components of
221 genetic variance is crucial for understanding the adaptive potential of a population. Non-additive
222 variance, a statistical component of genetic variance, contributes to similarity between siblings
223 but is not inherited in their offspring and thus cannot be selected for (*1*). As such, the only way to
224 quantify non-additive variance is by applying quantitative genetics methods (*12*, *16*, *29*). Plants
225 are especially suited for dissecting genetic variance because both genetic crosses and
226 environmental variance can be completely controlled; reliable estimates of non-additive
227 components of variance depend on such control (*12*, *29*). Here, the estimates of non-additive
228 variance in the 17,657 gene transcripts we report, reveal that non-additive variance forms a
229 substantial fraction of regulatory variation.

230 Gene regulation, which is central to the formation of phenotypes, is subject to natural selection
231 (*15*, *30*) and thus ideal to test whether the strength of selection shapes the amount of non-additive

variance. Our data compellingly show that stabilizing selection that has operated over long evolutionary time scales in the past, impact levels of non-additive variance today, and thus the potential of a gene to adapt in the future.

Genes whose genetic variance at the transcript level is predominantly non-additive were significantly more constrained than genes showing a large fraction of additive variance. The most strongly constrained genes, in turn, were found among the genes that were almost completely depleted of any genetic variance in transcript level, additive and non-additive variance alike. Although it has long been known that levels of genetic variation are low in genes with constrained expression (*13*, *14*, *31*, *32*), as far as we know, ours is the first report that non-additive transcript variance increases when the strength of stabilizing selection increases but ultimately decreases in genes exposed to the strongest stabilizing selective forces. These differences are revealed by a skew in the frequency spectrum of non-synonymous mutations as well as in the number of amino-acid substitutions that ultimately fix in a population. Our data further reveal an important additional feature: the rate of synonymous divergence and polymorphism is significantly lower in genes with high non-additive variance compared to in all other genes, suggesting that variation in the mutation rate may also play a role on the composition of genetic variance for expression.

We further identify genomic features that favor non-additive variation in gene expression. Non-additive variance in gene expression was promoted by transcript length, suggesting that the mutational target size of the expressed locus favors the emergence of such variance. Our study of transcript-level variance further indicates that non-additive variance increases as the network of co-regulated genes increases, which confirms the observation that proteins that are more central tend to be more constrained at the amino-acid level (*33–36*). In addition, this work shows that what promotes non-additive phenotypic variance is distinct from the molecular interaction between alleles. Indeed, contrary to our findings, additive molecular effects on fitness (e.g. co-dominant alleles) were inferred to be more frequent in highly connected genes of the same species (*37*).

Finally, we report that non-additive variance is particularly predominant among genes influencing the epigenetic remodeling of chromatin in differentiating cells. We associated non-additive variance in expression with the presence in the promoter of DNA motifs bound by a

regulator of stem thickening during growth (*38*). Although epigenetic variation has been often hypothesized to facilitate adaptation (*39*), our results indicate that given the mode of inheritance of epigenetic regulators, they are unlikely to produce the additive variance needed to fuel adaptation in *A. lyrata*. This work therefore demonstrates that, to understand the role of gene expression variation for past and future responses to selection, it is essential to differentiate additive and non-additive sources of variance.

Acknowledgments

We would like to thank Kirsten Bell for help with the laboratory work, the Cologne Genome Center for the transcriptome sequences and Emily Wheeler, Harwich, USA, for editorial help. Furthermore, we would like to thank F. He, S. Sunyaev, O. Savolainen, T. Hämälä, and M. Stetter for feedback, and F. Hospital for inspiring discussions ahead of this project. This research was funded by ERC Grant No. 648617 Adaptoscope.

Data availability

The whole-genome and transcriptome sequences are available via ENA. The whole-genome sequences of the parental lines are under the number PRJEB33206, and the transcriptome sequences are under the number PRJEB49153. The gene counts per individual and the sql database compiling descriptors of genomic variation will be deposited in dryad upon acceptance of the manuscript. All the original scripts are in git repository https://github.com/mtakou/Quantitative-analysis-of-gene-expression.

References

1. R. A. Fisher, The correlation between relatives on the supposition of mendelian inheritance. *Trans. R. Soc. Edinb*. **53**, 399–433 (1918).

2. M. Turelli, Phenotypic Evolution, Constant Covariances, and the Maintenance of Additive Variance. *Evolution*. **42**, 1342–1347 (1988).

3. P. Crnokrak, D. A. Roff, Dominance variance: associations with selection and fitness. *Heredity*. **75**, 530–540 (1995).

4. T. A. Mousseau, D. A. Roff, Natural selection and the heritability of fitness components. *Heredity*. **59**, 181–197 (1987).

5. N. H. Barton, P. D. Keightley, Understanding quantitative genetic variation. *Nat Rev Genet*. **3**, 11–21 (2002).

292 6. J. Hunt, M. W. Blows, F. Zajitschek, M. D. Jennions, R. Brooks, Reconciling Strong Stabilizing
293 Selection with the Maintenance of Genetic Variation in a Natural Population of Black Field Crickets
294 (Teleogryllus commodus). *Genetics*. **177**, 875–880 (2007).

295 7. J. H. Willis, H. A. Orr, Increased Heritable Variation Following Population Bottlenecks: The Role of
296 Dominance. *Evolution*. **47**, 949–957 (1993).

297 8. J. Wang, A. Caballero, P. D. Keightley, W. G. Hill, Bottleneck effect on genetic variance. A theoretical
298 investigation of the role of dominance. *Genetics*. **150**, 435–447 (1998).

299 9. P. Waldmann, Additive and non-additive genetic architecture of two different-sized populations of
300 Scabiosa canescens. *Heredity*. **86**, 648–657 (2001).

301 10. W. G. Hill, M. E. Goddard, P. M. Visscher, Data and Theory Point to Mainly Additive Genetic Variance
302 for Complex Traits. *PLoS Genet*. **4**, e1000008 (2008).

303 11. J. Clo, J. Ronfort, D. Abu Awad, Hidden genetic variance contributes to increase the short-term
304 adaptive potential of selfing populations. *Journal of Evolutionary Biology*. **33**, 1203–1215 (2020).

305 12. M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland,
306 Massachusetts, 1998).

307 13. J. F. Ayroles, M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman, M. M. Magwire, S. M. Rollmann,
308 L. H. Duncan, F. Lawrence, R. R. H. Anholt, T. F. C. Mackay, Systems genetics of complex traits in
309 Drosophila melanogaster. *Nat Genet*. **41**, 299–307 (2009).

310 14. I. G. Romero, I. Ruvinsky, Y. Gilad, Comparative studies of gene expression and the evolution of gene
311 regulation. *Nat Rev Genet*. **13**, 505–516 (2012).

312 15. S. C. Groen, I. Ćalić, Z. Joly-Lopez, A. E. Platts, J. Y. Choi, M. Natividad, K. Dorph, W. M. Mauck, B.
313 Bracken, C. L. U. Cabral, A. Kumar, R. O. Torres, R. Satija, G. Vergara, A. Henry, S. J. Franks, M. D.
314 Purugganan, The strength and pattern of natural selection on gene expression in rice. *Nature*, 1–5
315 (2020).

316 16. W. Huang, T. F. C. Mackay, The Genetic Architecture of Quantitative Traits Cannot Be Inferred from
317 Variance Component Analysis. *PLOS Genetics*. **12**, e1006421 (2016).

318 17. E. H. Leder, R. J. S. McCairns, T. Leinonen, J. M. Cano, H. M. Viitaniemi, M. Nikinmaa, C. R. Primmer,
319 J. Merilä, The Evolution and Adaptive Potential of Transcriptional Variation in Sticklebacks—
320 Signatures of Selection and Widespread Heritability. *Mol Biol Evol*. **32**, 674–689 (2015).

321 18. J. G. Kingsolver, S. E. Diamond, Phenotypic Selection in Natural Populations: What Limits Directional
322 Selection? *Am Nat*. **177**, 346–357 (2011).

323 19. A. Eyre-Walker, P. D. Keightley, The distribution of fitness effects of new mutations. *Nat Rev Genet*.
324 **8**, 610–618 (2007).

325 20. W. Stephan, Signatures of positive selection: from selective sweeps at individual loci to subtle allele
326 frequency changes in polygenic adaptation. *Mol Ecol*. **25**, 79–88 (2016).

21. B. Y. Kim, C. D. Huber, K. E. Lohmueller, Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples. *Genetics*. **206**, 345–361 (2017).

22. T. M. Mattila, J. Tyrmi, T. Pyhäjärvi, O. Savolainen, Genome-Wide Analysis of Colonization History and Concomitant Selection in Arabidopsis lyrata. *Mol Biol Evol*. **34**, 2665–2677 (2017).

23. M. Takou, T. Hämälä, E. M. Koch, K. A. Steige, H. Dittberner, L. Yant, M. Genete, S. Sunyaev, V. Castric, X. Vekemans, O. Savolainen, J. de Meaux, Maintenance of Adaptive Dynamics and No Detectable Load in a Range-Edge Outcrossing Plant Population. *Molecular Biology and Evolution*. **38**, 1820–1836 (2021).

24. J. C. Reif, F.-M. Gumpert, S. Fischer, A. E. Melchinger, Impact of Interpopulation Divergence on Additive and Dominance Variance in Hybrid Populations. *Genetics*. **176**, 1931–1934 (2007).

25. P. Crnokrak, D. A. Roff, Dominance variance: associations with selection and fitness. *Heredity*. **75**, 530–540 (1995).

26. K. Bomblies, D. Weigel, Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat Rev Genet*. **8**, 382–393 (2007).

27. F. He, K. A. Steige, V. Kovacova, U. Göbel, M. Bouzid, P. D. Keightley, A. Beyer, J. de Meaux, Cis-regulatory evolution spotlights species differences in the adaptive potential of gene expression plasticity. *Nat Commun*. **12**, 3376 (2021).

28. S. Miyashima, P. Roszak, I. Sevilem, K. Toyokura, B. Blob, J. Heo, N. Mellor, H. Help-Rinta-Rahko, S. Otero, W. Smet, M. Boekschoten, G. Hooiveld, K. Hashimoto, O. Smetana, R. Siligato, E.-S. Wallner, A. P. Mähönen, Y. Kondo, C. W. Melnyk, T. Greb, K. Nakajima, R. Sozzani, A. Bishopp, B. De Rybel, Y. Helariutta, Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature*. **565**, 490–494 (2019).

29. C. J. Yang, L. F. Samayoa, P. J. Bradbury, B. A. Olukolu, W. Xue, A. M. York, M. R. Tuholski, W. Wang, L. L. Daskalska, M. A. Neumeyer, J. de J. Sanchez-Gonzalez, M. C. Romay, J. C. Glaubitz, Q. Sun, E. S. Buckler, J. B. Holland, J. F. Doebley, The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proc Natl Acad Sci USA*. **116**, 5643–5652 (2019).

30. K. A. G. Kremling, S.-Y. Chen, M.-H. Su, N. K. Lepak, M. C. Romay, K. L. Swarts, F. Lu, A. Lorant, P. J. Bradbury, E. S. Buckler, Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature*. **555**, 520–523 (2018).

31. J. C. Fay, P. J. Wittkopp, Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*. **100**, 191–199 (2008).

32. T. Hämälä, M. J. Guiltinan, J. H. Marden, S. N. Maximova, C. W. dePamphilis, P. Tiffin, Gene Expression Modularity Reveals Footprints of Polygenic Adaptation in Theobroma cacao. *Mol Biol Evol*. **37**, 110–123 (2020).

33. H. A. Orr, Fitness and its role in evolutionary genetics. *Nat Rev Genet*. **10**, 531–539 (2009).

362 34. M. W. Hahn, A. D. Kern, Comparative Genomics of Centrality and Essentiality in Three Eukaryotic
363 Protein-Interaction Networks. *Mol Biol Evol*. **22**, 803–806 (2005).

364 35. E. B. Josephs, S. I. Wright, J. R. Stinchcombe, D. J. Schoen, The Relationship between Selection,
365 Network Connectivity, and Regulatory Variation within a Population of Capsella grandiflora.
366 *Genome Biol Evol*. **9**, 1099–1109 (2017).

367 36. R. R. Masalia, A. J. Bewick, J. M. Burke, Connectivity in gene coexpression networks negatively
368 correlates with rates of molecular evolution in flowering plants. *PLOS ONE*. **12**, e0182289 (2017).

369 37. C. D. Huber, A. Durvasula, A. M. Hancock, K. E. Lohmueller, Gene expression drives the evolution of
370 dominance. *Nature Communications*. **9**, 2750 (2018).

371 38. R. Lister, O. Chew, M.-N. Lee, J. L. Heazlewood, R. Clifton, K. L. Parker, A. H. Millar, J. Whelan, A
372 Transcriptomic and Proteomic Characterization of the Arabidopsis Mitochondrial Protein Import
373 Apparatus and Its Response to Mitochondrial Dysfunction. *Plant Physiol*. **134**, 777–789 (2004).

374 39. M. W. Schmid, C. Heichinger, D. Coman Schmid, D. Guthörl, V. Gagliardini, R. Bruggmann, S. Aluri, C.
375 Aquino, B. Schmid, L. A. Turnbull, U. Grossniklaus, Contribution of epigenetic variation to
376 adaptation in Arabidopsis. *Nat Commun*. **9**, 4446 (2018).

377

378

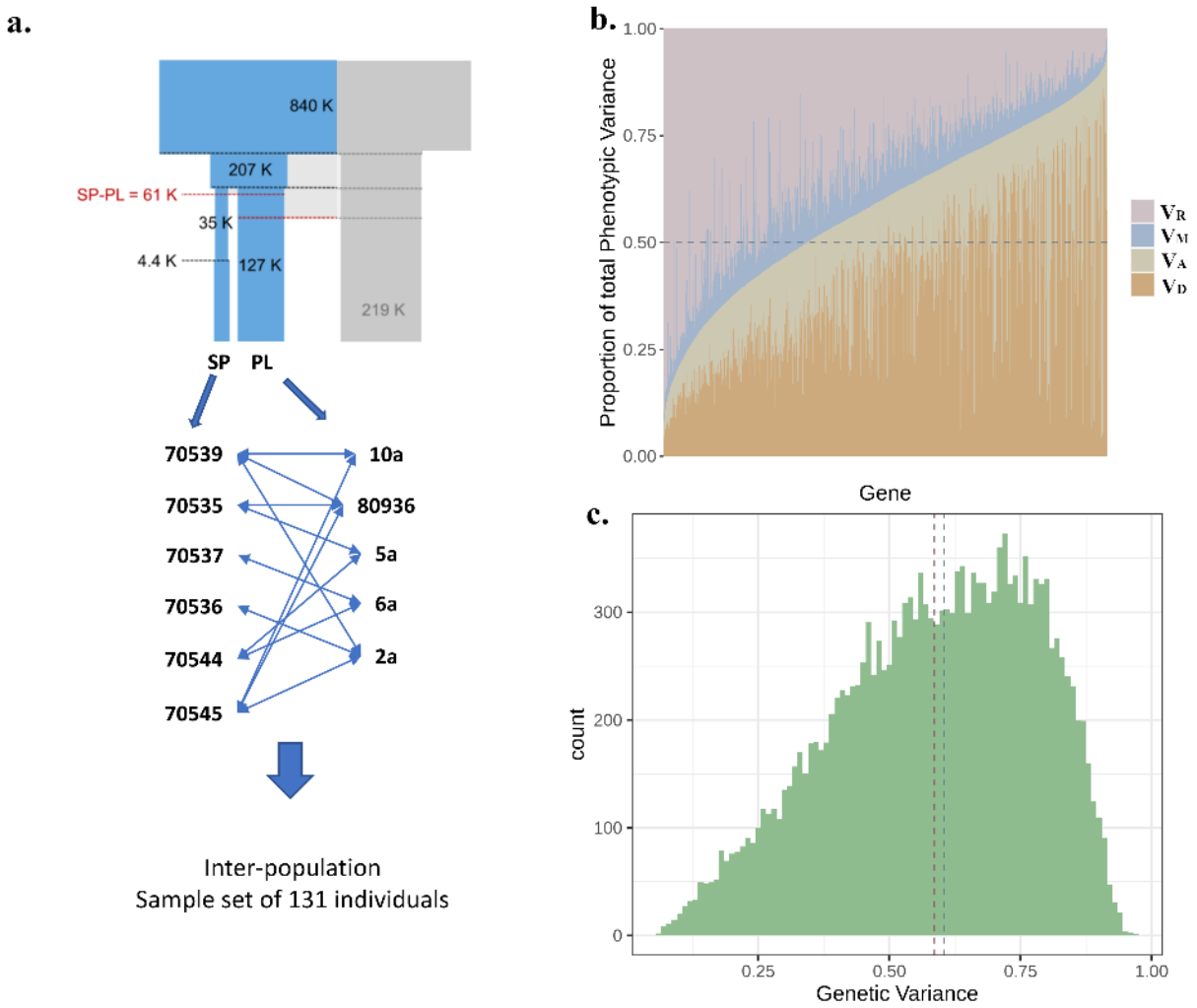

Figure 1: Phenotypic variance of gene expression is mostly composed of non-additive genetic variance. **a.** We took advantage of a set of 131 individuals obtained via crosses between two populations of *A. lyrata*. A previous study had shown that the two populations have diverged approximately 61 kyears ago, with the SP populations being more severely bottlenecked than the PL population (redrawn from (*23*)). In total, 5 individuals sampled in Plech (PL) were reciprocally crossed with 6 individuals from Spiterstulen (SP). The crosses were done in such a way that plants were reciprocally crossed with 2 individuals from the other population in order to produce full- and half-sib families. **b.** The total phenotypic variance of each one of the 17,656 genes included in the analysis was partitioned into non-additive, additive, maternal, and residual components of variance. The genes were ordered by increasing values of the fraction of total phenotypic variance in transcript level that was attributed to genetic variance (V_G). The dashed line marks the 0.5 point where half of the variance is explained. **c.** The distribution of total genetic variance for each transcript (V_G), as a fraction of the total observed variance in expression level. The dashed blue line shows the median value (0.604) of the distribution; the dashed red line indicates the mean value (0.585).

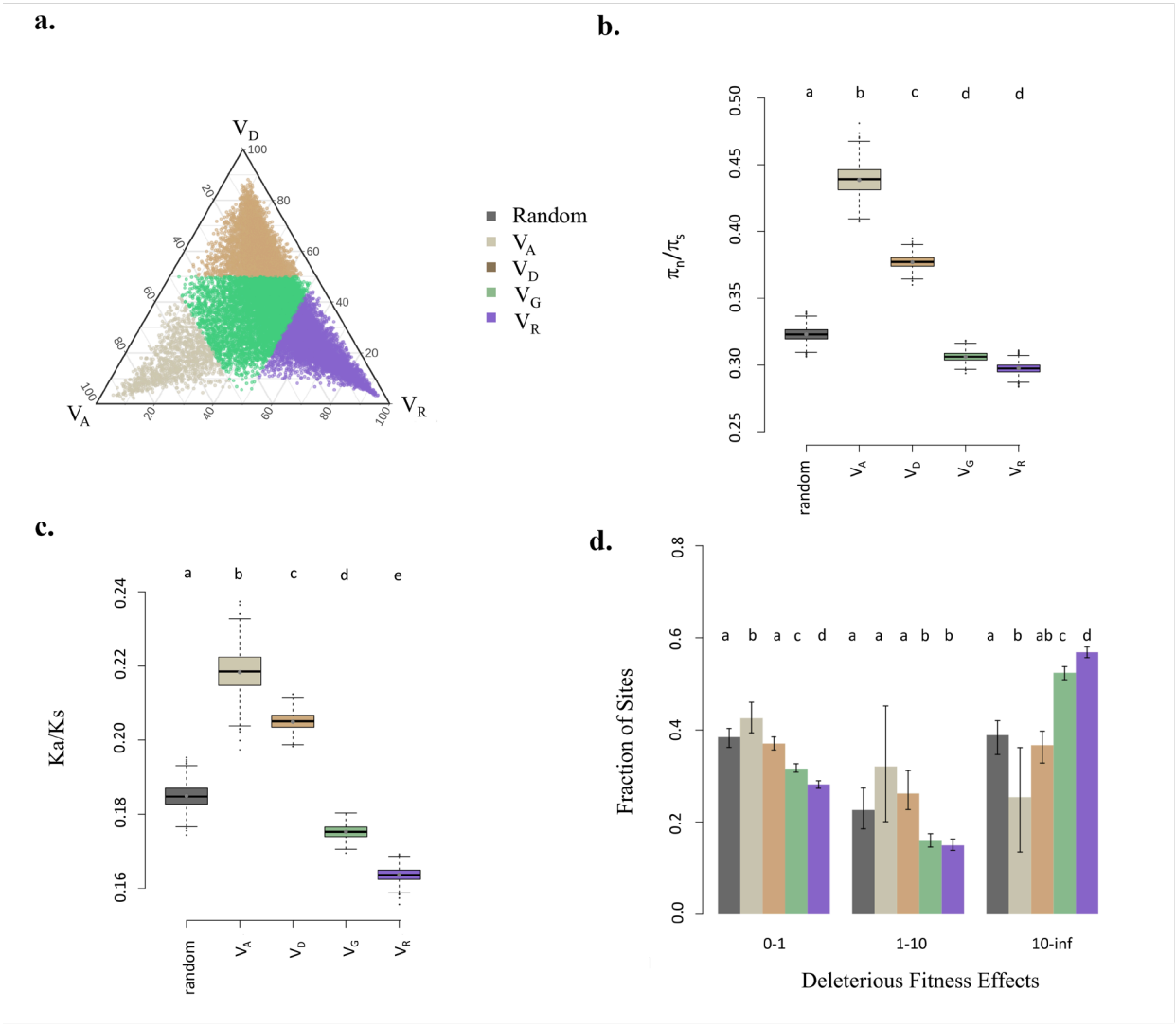

Figure 2: The strength of stabilizing selection at the amino-acid level is associated with the predominant component of the variance in expression. **a.** We grouped genes based on the partitioning of total phenotypic variance at the transcript level into non-additive, additive, or residual variance. The threshold for each case is 50% of the total phenotypic variance. **b.** We computed the mean and variance of average pi/ps values for each group based on population genomics datasets for PL. **c.** The Ka/Ks ratio for the different gene groups. The confidence intervals and *p* values for b through d were estimated based on 100 bootstrap iterations. **d.** Inferred distribution of fitness effects of new mutations in the four gene groups for each of three classes of deleterious fitness effect N_es (nearly neutral, $N_es<1$, weakly deleterious $1<N_es<10$, strongly deleterious $N_es>10$). Gray: control group of 3,500 random genes; beige: high V_A genes; orange: high V_D genes; purple: high V_R genes; green: intermediate levels of V_G. In all plots, the letters indicate whether the values differ significantly from each other.

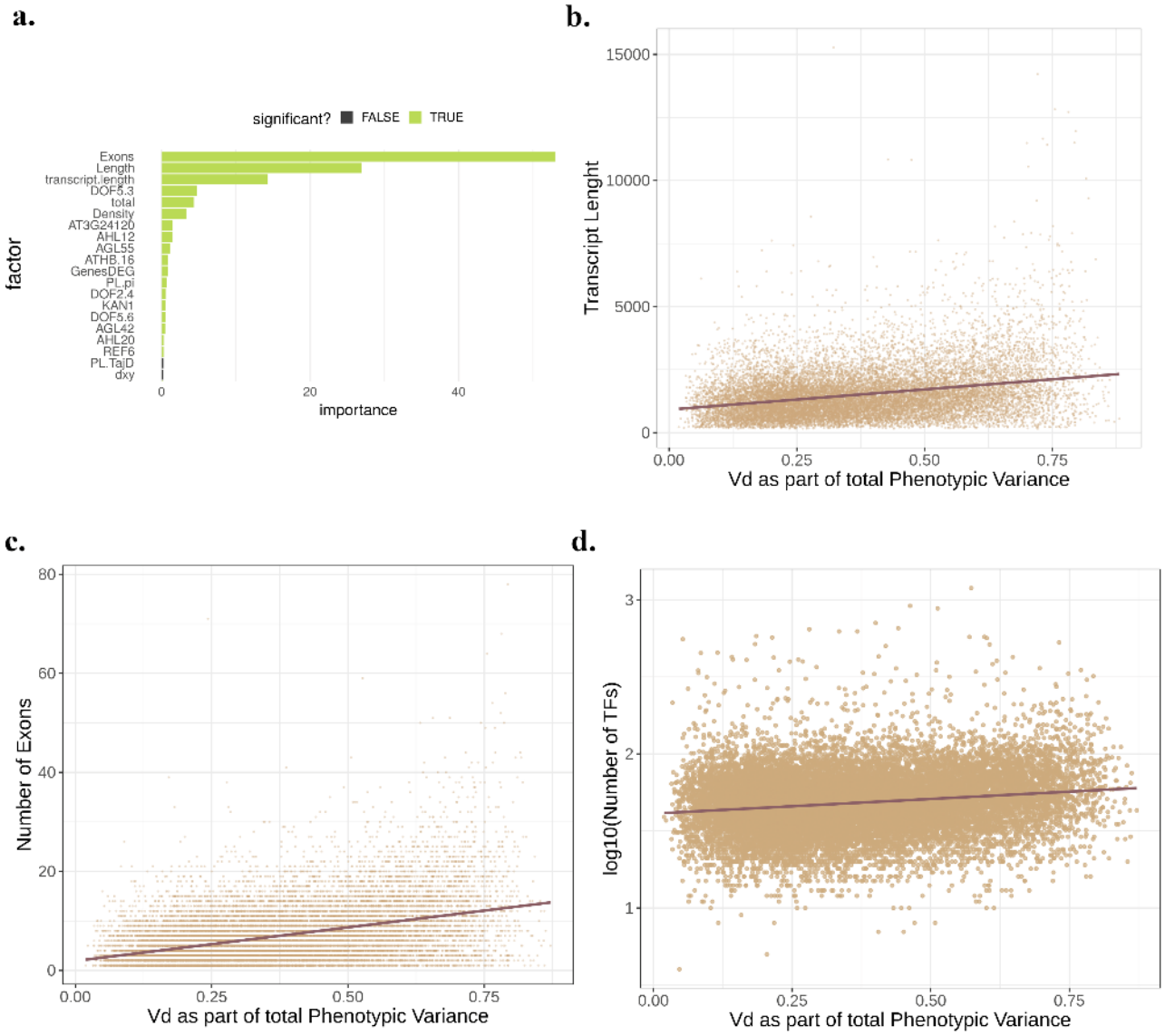

Figure 3: The level of non-additive variance depends on gene architecture. **a.** Importance of tested genomic parameters on the level of non-additive variance each gene. The twenty factors that explain the most variance are shown. The predictive power of the model is R^2 = 0.18. Significant parameters are shown in green. **b.** The fraction of non-additive variance correlates significantly with the transcript length of each gene (p=1.22e-266, ρ=0.258). **c.** The fraction of non-additive variance correlates significantly with the number of exons that a gene has (p < 2.2e-16, ρ=0.385). **d.** The fraction of non-additive variance correlates significantly with the total number of transcription-factor binding sites detected in the upstream vicinity of the gene (p < 2.2e-16, ρ=0.162).