Title: Strength of purifying 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 purifying selection than genes with high additive variance. This study demonstrates that the potential of standing variation for future evolution is shaped the history of purifying selection.

Main

Natural populations harbor substantial 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*). The higher this fraction, the more genetically variable traits are limited in their potential to respond to future selective pressures. 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 (*3–7*). In *Drosophila serrata*, for example, no link was found between the amount of non-additive variance in wing traits and their co-variation with standing genetic variance in male competitive fitness (*7*). Furthermore, other factors may determine the relative amount of non-additive genetic variance present in a population (*8*). 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 (*9–13*). 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 should 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, depending on the frequency of the variants, may generate non-additive genetic variance (*14*).

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 (*15–17*). When quantified over the appropriate pedigree of full- and half-sib families (*1*, *14*, *18*), transcriptome studies can reveal the substantial levels of non-additive genetic variance that characterize expressed genes (*19*). Reconstructing how variation in these genes impacts fitness in natural conditions is challenging (*17*, *20*). Quantitative genetics analyses of expression have shown that natural selection considerably decreases the amount of standing genetic variance in

natural populations, therefore an absence of association between non-additive genetic variance and fitness may not be conclusive (*21*). 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 selection, or to quantify the strength of purifying selection on variation affecting the amino-acid sequence of transcribed genes (*22–24*).

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

Results

Predominance of non-additive genetic variance in gene expression

We took advantage of a set of 10 full- and half-sib families of the outcrossing plant species *Arabidopsis lyrata* ssp. *petraea*, which were obtained by crossing individuals from two locally adapted populations (*25*, *26*) (Supplementary Table S1). These individuals, which all had the same inter-population genomic background, have the same expected level of inbreeding (Figure 1a). We collected leaf material during the exponential rosette growth phase of the plants and sequenced their transcriptome. Full- and half-sib relationships allowed us to partition the total phenotypic variance (V_P) of gene expression into additive variance (V_A), non-additive (V_{NA}), maternal (V_M), and residual (V_R) components of variance in the expression for each transcript in the genome, while controlling for inbreeding (Figure 1a, Supplementary Table S1). The main non-additive component includes genetic variance due to allelic interactions both within locus and between loci (*28*). Indeed, the statistical variance due to intra- and inter-locus interactions, i.e. dominance and epistatic variance, cannot easily be disentangled in pedigrees without

assuming specific molecular models (*18*, *29*). In the following, we thus report on the amount of phenotypic variance (V_P) that is explained by non-additive genetic variance (V_{NA}). In 67.7% of the 17,656 genes analyzed, more than 50% of total expression variance (V_P) was explained by the significant contribution of the two genetic components of variance, V_{NA} (mean = 0.375) and V_A (mean = 0.215). The maternal variance (V_M) was close to zero for most transcripts (mean = 0.088). Finally, the origin of the maternal plant did not impact gene expression (all $p > 2.486e-06$, the Bonferroni-adjusted significance threshold).

Delineating overall genetic variance showed that V_{NA} contributed the most (25.4% of 17,656 genes had more than 50% of gene expression variance V_P due to V_{NA}), in contrast to the additive genetic component V_A (6.3% of these genes had more than 50% of their variance in expression, Figure 1b-c). Hence, gene expression variance V_P is due largely to genetics, but only a limited fraction of that genetic variance is inherited additively. Although amounts of non-additive genetic variance are rarely quantified in natural populations, the amount we detected is not unexpected. Similar fractions of non-additive variance have been reported in phenotypic traits (*30*), and non-additive variance was as high if not higher than additive variance for most transcripts expressed in threespine stickleback (*19*). In a control dataset generated by permuting individuals across families, V_R explained more than half of the V_P for more than 73% of transcripts (mean = 0.717), and V_{NA} was not significant (mean = 0.135), confirming that the levels of genetic variance we report are not random and that, despite the relatively small number of individuals, we have the power to quantify non-additive variance.

As expected for a transcriptome dataset, many genes formed co-expressed clusters of variable size (Supplementary Figure S2). Yet, major genetic incompatibilities or large-effect variants were unlikely to explain the data (*31*). Indeed, a genome-wide association showed that most of the genetic variation in expression could not be tracked down to a handful of large effect loci (see Supplementary Information).

Strength of purifying selection associates with inheritance of regulatory variation

Having determined levels of non-additive genetic variance in gene expression, we then used a population genetics method to determine whether the mode of gene expression inheritance was associated with past selective pressures. We first examined rates of amino-acid evolution in the transcripts that were expressed in our dataset. We partitioned expressed transcripts in three non-

overlapping sets based on their relative excess of V_A, V_{NA}, or V_R (Figure 2a). As a control, we included a fourth set of 3,500 genes sampled at random (Fig. 2a). We then estimated the rate of amino acid divergence that evolved between *A. lyrata* and *A. thaliana* across these four groups of genes with different inheritance. The four sets of transcripts differed significantly in their rate of amino-acid divergence. Genes with high V_R had the rate of amino-acid divergence, whereas genes with highest V_A had the fastest rate (Figure 2b-c). Interestingly, genes with the largest V_{NA} displayed a rate of amino-acid divergence that was intermediate between those shown by genes with large V_A and genes with large V_R.

The rate of amino-acid divergence is expected to increase if positive selection is acting on the encoded protein, but increases may also occur if the evolutionary constraints on the amino-acid sequence are weak. In order to differentiate these alternatives, the segregation of molecular variation within species can be used to infer the fitness effect of new mutations and quantify the strength of the evolutionary constraints imposed on the amino-acid sequence. We took advantage of a previous study, which determined the demographic history of the same populations of *A. lyrata* (*26*, *32*). We inferred the distribution of fitness effects (DFE) of new non-synonymous variants in these gene sets, following a method described in (*26*) that incorporates the specific demographic parameters of each population (Supplementary Figure S3). Both the inferred demographic model and the DFE fitted the site-frequency spectra (SFS) of synonymous and non-synonymous variants well (Supplementary Figure S3a), indicating that the model provides a good representation of the evolutionary and selective history of these populations.

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

Our analysis further revealed that selective constraints on genes with large V_A (threshold 0.5) are comparatively lower than on genes with high V_{NA} (Figure 2c, Supplementary Figure S3b-c). The DFE analysis showed that, compared to other gene sets, genes with large V_A in transcript expression hosted a significantly larger fraction of new non-synonymous mutations that are nearly neutral ($0 < N_{anc}s < 1$, $p = 0.03$). At the same time, a significantly smaller fraction of new mutations was under strong purifying selection in genes with large V_A ($N_{anc}s > 10$, $p = 0.01$). Furthermore, genes with high V_A were less likely to map to genomic regions harboring footprints of selective sweeps. Although regions with signatures of selective sweeps extended over 27% of the genome, only 17% of the high V_A genes mapped to these regions (pairwise t-test, $p = 2.115e-12$), further confirming that these genes are unlikely to contain variants that have been under directional selection. Therefore, both the high rate of amino-acid divergence observed in these genes and their inferred DFE indicate that genes with additive variance in gene expression tend to be exposed to lower levels of purifying selection than genes from all other sets. In fact, genes with high V_A also tend to show a higher mutation rate (Supplementary Information, Supplementary Figure S3).

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

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

In order to better understand whether genomic factors modulate the relative fractions of non-additive and additive variance, we compiled a total of 200 descriptors of genetic variation at each gene; these included the number of known transcription-factor binding sites in the upstream non-coding region, the length of the genes, and various descriptors of nucleotide diversity (see Methods). We trained a random forest model to identify the genomic architecture or population genetic parameters that best explain variation in the prevalence of V_A or V_{NA} in gene expression

174 and tested the significance of their measured importance by permutation (see methods). We first
175 performed the analysis using the 200 descriptors to predict variation in V_A, but the model's
176 ability to classify high V_A genes was poor (R^2 = 0.009 and mean square error of 0.021).

177 For variation in V_{NA}, however, the model was accurate at classifying genes with high V_D (R^2 =
178 0.18 and mean square error of 0.026), highlighting that the most important factors for classifying
179 genes according to their level of V_{NA} are descriptors of gene structure (Figure 3a; Supplementary
180 Table S3). Specifically, the number of exons per gene (importance: 53.021; p = 0.0164), total
181 gene length (importance: 26.92; p = 0.0164), and transcript length (importance: 14.27; p =
182 0.0164) were the three factors that best predicted levels of non-additive variance. The importance
183 of exon number persisted even when gene length was corrected for. Transcript length was
184 significantly correlated with V_{NA} (Spearman coefficient ρ = 0.258, p = 1.22e-266,
185 Supplementary Table S4; Figure 3b). We used a linear model to confirm that the effect of
186 transcript length on V_{NA} (p < 2.2e-16, see methods) remained significant even after correcting
187 for median gene expression (p = 0.0176) and V_G (p < 2.2e-16). The model did not suggest a
188 significant interaction between transcript length and median gene expression (p = 0.224). The
189 number of exons per gene correlated significantly with the level of non-additive variance
190 (Spearman coefficient ρ = 0.358, p < 2.2e-16; Figure 3c). The random forest analysis further
191 revealed that the density of SNPs within a gene was a predictor of high V_{NA}. Strikingly, all
192 significant descriptors tend to increase the mutational target size of the transcript. We thus
193 conclude that non-additive variance is likely to be observed as the mutational target size
194 increases.

195 Interestingly, the analysis further revealed that the total number of transcription-factor binding
196 sites explained V_{NA} well (p = 0.0163; importance 4.34; Figure 3d). When we corrected for gene
197 length, the total number of transcription-factor binding sites becomes the most important
198 predictor in the random forest analysis (Importance 28.72; permutation based p =0.0163, see
199 Supplementary Information), which explains the level of V_{NA} in the dataset. In particular, the
200 transcription factor binding sites associated with high V_{NA} were predicted targets of a DOF zinc
201 finger protein expressed in young rosettes (DOF 5.3; Importance 4.722; permutation based p =
202 0.0163); this protein (AT5G60200 TMO6) was recently shown to determine the stem cell

division that is necessary for plant thickening (*33*). This regulator was likely active in the tissue samples we collected, because we sampled during the exponential phase of rosette growth.

We further quantified the distribution of genes with high V_{NA} across gene ontology (GO) categories of expressed genes and tested for specific functional enrichment by permutation (see methods). We found that genes with high V_{NA} were enriched for GO categories reflecting cell differentiation. The top enriched GO terms for the highest V_{NA} estimates were "chromatin silencing" (p = 4.5e-16) and "cytokinesis by cell plate formation" (p = 2e-11), "DNA methylation" (p = 4.3e-11; Supplementary Table S5). In addition, genes with high V_A were not randomly distributed among molecular functions; but, unlike V_{NA} genes, the functional enrichment of high V_A genes was less striking and not related to cell differentiation. Among the top GO categories, the most enriched among genes with high V_A values were several functions linked to central metabolism such as "glucose catabolic process" (p = 8.3-05) and "glycolytic process" (p = 1.6e-05), but we also detected enrichments among genes involved in the "response to cadmium ion" (p = 0.00015) and "proteasome core complex assembly" (p = 0.00016; Supplementary Table S6).

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

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

Discussion

Delineating the genomic and evolutionary factors that increase the non-additive components of genetic variance is crucial for understanding the adaptive potential in a population. Non-additive variance, a statistical component of genetic variance, contributes to similarity between siblings but is not inherited in their offspring and thus cannot be selected for (*1*). As such, the only way to quantify non-additive variance is by applying quantitative genetics methods (*14*, *18*, *34*). Plants are especially suited for dissecting genetic variance because both genetic crosses and environmental variance can be completely controlled; reliable estimates of non-additive components of variance depend on such control (*14*, *34*). Here, the estimates of non-additive variance in the 17,657 gene transcripts we report, reveal that non-additive variance forms a substantial fraction of regulatory variation. Although initially assumed to be very low, empirical data reveal that the non-additive fraction of genetic variation is not negligible in natural populations (*8*, *35*). A meta-analysis concluded that this fraction may reach 38% of the genetic variance, on average (*8*, *36*). For life-history traits in non-domesticated species, this fraction even reached 60% of the genetic variance (*8*). Furthermore, a previous study in stickleback also determined that non-additive genetic variance forms about 50% of the total genetic variance in transcript levels (*19*). The amount we report is therefore not unexpected. It is to be noted, however, that the pedigree used here was derived from inter-population crosses. This design does not affect our ability to estimate dominance variance, because all our indvidiuals have the same inter-population genomic background and the animal model corrects for inbreeding (*28*). But we cannot rule out that intra-population crosses may yield different estimates. Indeed, molecular models of genetic dominance effects indicate that only shared variants can contribute to the component of non-additive variance that is due to dominant/recessive interactions between alleles (*27*). Their frequency in the two parental populations is sufficiently similar so as not to change very much the amount of non-additive variance that is expected in our breeding design compared to a design based on intra-population crosses (Supplementary Figure S1, details in Suppl. Material). Variants that are not shared between populations, however, can contribute to the additive or epistatic components of genetic variation (*27*). Furthermore, the effect of linkage and epistasis in such crossing schemes has, to our knowledge, not yet been modelled. So within population, the fraction of non-additive genetic variance may be different from what we report in this study.

Inspite of this limitation, the data allows us to examine whether the relative amount of non-additive variance we detect in the genome associates with population genomic footprints of natural selection. Indeed, there is no doubt that gene expression levels are subject to natural selection (*15*, *17*, *21*, *37*). Our data compellingly show that purifying selection that has operated over long evolutionary time scales in the past, impact levels of non-additive variance today, and thus the potential of standing variation in gene expression to contribute to future adaptation.

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. Both levels of constraints measured within populations and between species confirmed this pattern. Although it has long been known that levels of genetic variation in expression are low in genes with constrained amino-acid evolution (*15*, *16*, *21*, *38*, *39*), as far as we know, ours is the first report that non-additive transcript variance increases when the strength of purifying selection increases but ultimately decreases in genes exposed to the strongest purifying 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. Indeed, selection against recessive deleterious variants will tend to maintain their frequency very low, in the realm where it can make for a majority of genetic variance(*28*). 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 (*40–43*). 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 (*44*).

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 (*45*). Although epigenetic variation has been often hypothesized to facilitate adaptation (*46*), 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).

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

7. J. L. Sztepanacz, M. W. Blows, Dominance Genetic Variance for Traits Under Directional Selection in Drosophila serrata. *Genetics*. **200**, 371–384 (2015).

8. M. E. Wolak, L. F. Keller, "Dominance genetic variance and inbreeding in natural populations" in *Quantitative Genetics in the Wild*, A. Charmantier, D. Garant, L. E. B. Kruuk, Eds. (Oxford University Press, 2014; https://doi.org/10.1093/acprof:oso/9780199674237.003.0007), p. 0.

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

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

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

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

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

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

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

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

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

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

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

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

21. A. Hodgins-Davis, D. P. Rice, J. P. Townsend, Gene Expression Evolves under a House-of-Cards Model of Stabilizing Selection. *Molecular Biology and Evolution*. **32**, 2130–2140 (2015).

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

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

24. 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).

25. 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).

26. 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).

27. A. Caballero, Quantitative Genetics. *Higher Education from Cambridge University Press* (2020), , doi:10.1017/9781108630542.

28. Z. G. Vitezica, A. Reverter, W. Herring, A. Legarra, Dominance and epistatic genetic variances for litter size in pigs using genomic models. *Genetics Selection Evolution*. **50**, 71 (2018).

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

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

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

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

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

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

400 35. P. Crnokrak, D. A. Roff, Dominance variance: associations with selection and fitness. *Heredity*. **75**,
401 530–540 (1995).

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

404 37. 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.
405 Bradbury, E. S. Buckler, Dysregulation of expression correlates with rare-allele burden and fitness
406 loss in maize. *Nature*. **555**, 520–523 (2018).

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

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

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

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

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

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

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

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

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

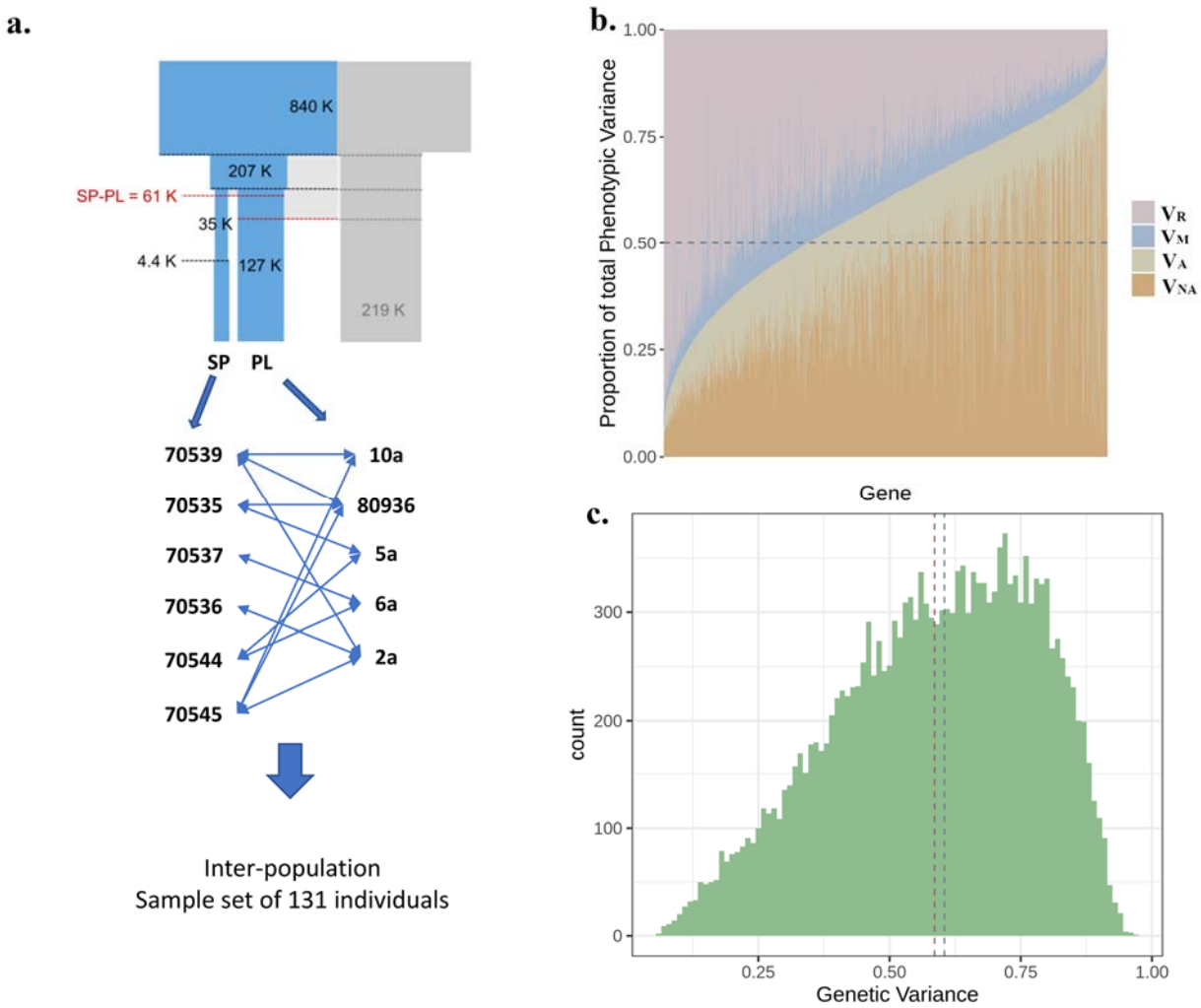

Figure 1: Phenotypic variance of gene expression is mostly composed of non-additive genetic variance. **a.** We took advantage of a set of 131 individuals obtained via crosses between two populations of *A. lyrata*. A previous study had shown that the two populations have diverged approximately 61 kyears ago, with the SP populations being more severely bottlenecked than the PL population (redrawn from (*26*)). 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 (V_{NA}), additive (V_A), maternal (V_M), and residual components of variance (V_R). 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).

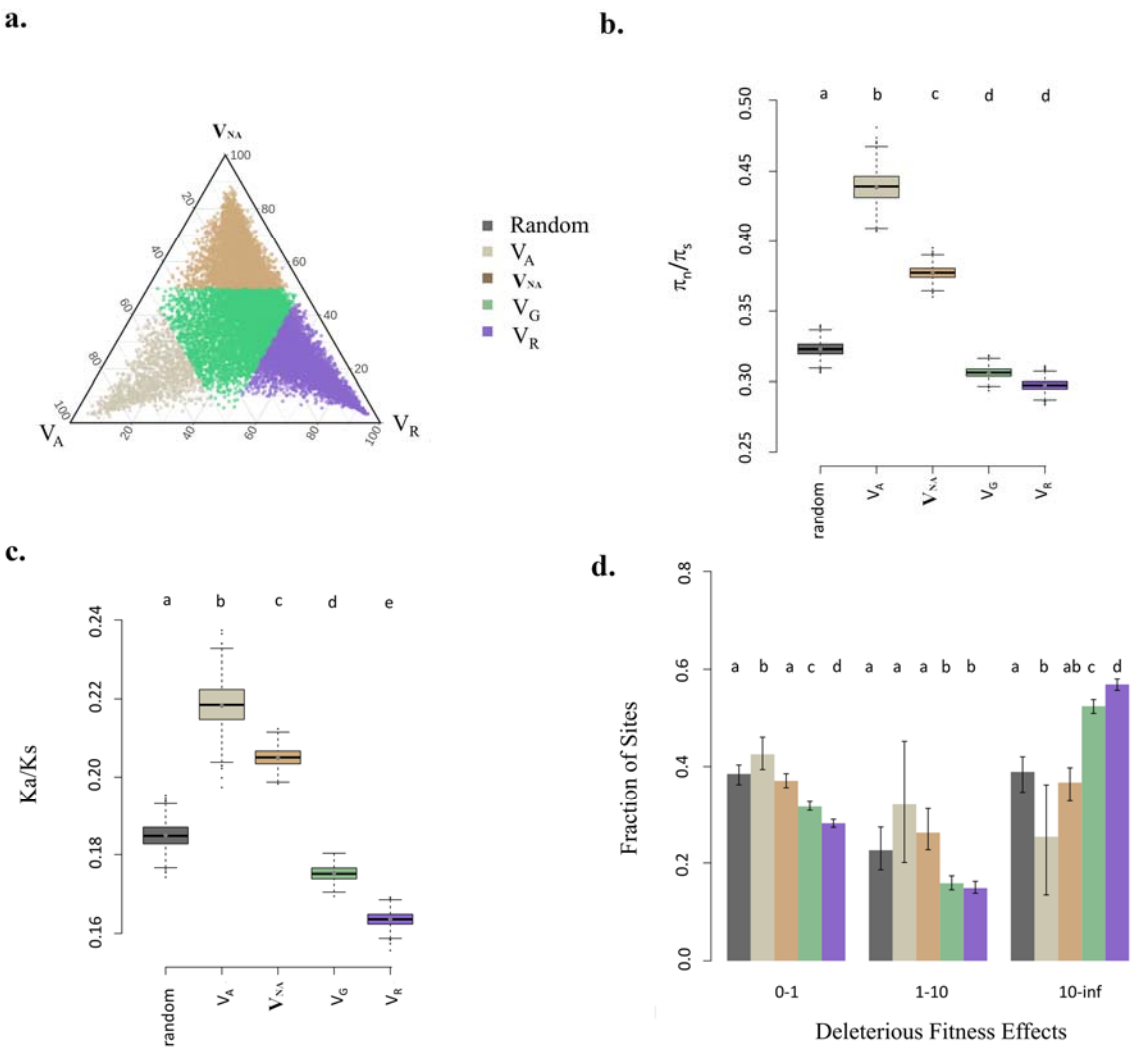

Figure 2: The strength of purifying selection at the amino-acid level is associated with the predominant component of the variance in expression. **a.** We grouped genes based on the partitioning of total phenotypic variance at the transcript level into non-additive, additive, or residual variance. The 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.

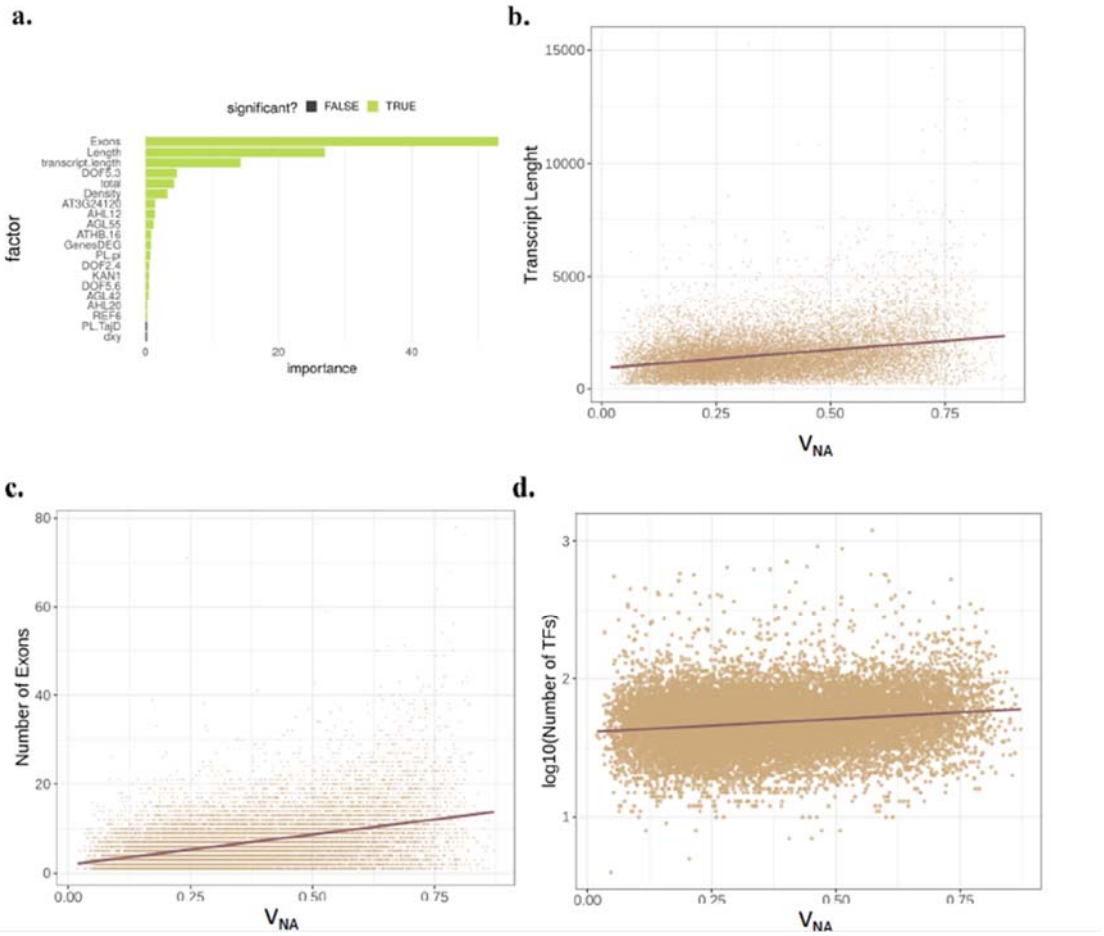

Figure 3: The level of non-additive variance depends on gene architecture. **a.** Importance of tested genomic parameters on the level of non-additive variance each gene. The twenty factors that explain the most variance are shown. The predictive power of the model is R^2 = 0.18. Significant parameters 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).