Measuring natural selection on the transcriptome *

John R. Stinchcombe[1,2] and John K. Kelly[3]

1. Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada
2. Koffler Scientific Reserve at Joker's Hill, University of Toronto, King, ON
3. Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, United States of America

Emails: john.stinchcombe@utoronto.ca
jkk@ku.edu

ORCIDs: https://orcid.org/0000-0003-3349-2964
https://orcid.org/0000-0001-9480-1252

*** We dedicate this paper to Mark Rausher for his fundamental contributions to the study of natural selection in wild plant populations.**

Summary

The level and pattern of gene expression is increasingly recognized as a principal determinant of plant phenotypes and thus of fitness. The estimation of natural selection on the transcriptome is an emerging research discipline. We here review recent progress and consider the challenges posed by the high dimensionality of the transcriptome for the multiple regression methods routinely used to characterize selection in field experiments. We consider several different methods, including classical multivariate statistical approaches, regularized regression, latent factor models, and machine learning, that address the fact that the number of traits potentially affecting fitness (each expressed gene) can greatly exceed the number of plants that researchers can reasonably monitor in a field study. While such studies are currently few, extant data are sufficient to illustrate several of these approaches. With additional methodological development coupled with applications to a broader range of species, we believe prospects are favorable for directly characterizing selection on gene expression within natural plant populations.

Introduction

One of the fundamental goals of evolutionary biology is to understand how natural selection acts on phenotypes. Understanding the form, strength, and direction of selection is crucial to making predictions about the evolutionary trajectory of traits, understanding adaptation, and quantitatively testing alternative hypotheses about the extent to which organismal features evolve by adaptive or non-adaptive mechanisms. For this reason, evolutionary biologists have devoted considerable effort to measuring natural selection in field, experimental, and common garden environments (Kingsolver et al. 2001, 2012; Siepielski et al. 2013). While the rapid progress in molecular biology and genomics continually offers the promise of characterizing the genetic basis of complex traits (Hill 2010), there is a growing realization that these techniques and approaches yield a suite of molecular phenotypes that are themselves amenable to evolutionary (and genetic) analysis. Here we outline the prospects and challenges for characterizing natural selection on one particularly relevant– and increasingly attainable– set of molecular phenotypes, gene expression.

Several lines of evidence suggest that gene expression is an important determinant of organismal fitness, and thus likely to experience selection. Early experimental results, from mutation accumulation experiments in which the strength of selection has been minimized or reduced, suggested that stabilizing selection was acting on gene expression (Rifkin et al. 2005, Gilad et al. 2006). Likewise, observations from the microarray-era indicated that populations experiencing different environmental conditions can diverge in gene expression, even in the face of substantial gene flow (Oleksiak et al. 2002), potentially indicating the past action of selection. Collectively, these and more recent studies reveal that gene expression can and does evolve on a wide array of time-scales, including in the laboratory (Rifkin et al. 2005), between adjacent populations of the same species (Oleksiak et al. 2002), in response to severe weather events (Campbell-Staton et al. 2017; Hamann et al. 2021), and in ecologically realistic, complex communities within a handful of generations (Ghalambor et al. 2015, 2018).

Despite prominent examples of gene expression evolution on microevolutionary timescales, as well as theorizing on its relevance on macroevolutionary time scales (e.g., King and Wilson 1975), few researchers have directly estimated natural selection on gene expression. In contemporary populations, is gene expression subject to stabilizing selection as first predicted, or is it frequently subject to directional selection as might be deduced from these studies of evolutionary divergence on short time scales? How does the strength of selection on gene expression compare to that on ‘macroscopic’ traits such as life history, morphology, or behavior? Are the levels of transcription among multiple genes in the transcriptome sufficiently correlated as to

require distinguishing between direct and indirect selection? Is there a relationship between the level of expression and the strength of phenotypic selection, analogous to the relationship between the level of expression and rates of molecular evolution (Wright et al. 2004, Slotte et al. 2011)? These and a host of other questions require extending the Lande-Arnold revolution (Lande and Arnold 1983; Svensson 2023) from traditional macroscopic phenotypes to include gene expression.

Transcriptomes as Quantitative Traits

Progress on these questions starts with the recognition that gene expression is itself a quantitative trait. The expression levels of genes across the genome are quantitative traits with strong environmental influences combined with multi-locus genetic effects (Liu et al. 2019). In fact, given that modern RNA-seq experiments often obtain expression estimates for many genes simultaneously (N in the 1,000s), the transcriptome is really a collection of vectors (or a matrix). Considering the transcriptome as a set of correlated characters within a quantitative genetic perspective offers several insights. Perhaps most importantly, there is a well-developed machinery to analyze selection on correlated quantitative traits (Lande & Arnold, 1983; Rausher, 1992).

The transcriptome is hugely multivariate and thus offers investigators a chance to measure many phenotypes simultaneously. While it is true that these phenotypes are "snapshots" – measured at a particular time, life stage, or tissue – that is also usually true of measures of morphology, physiology, and life history. For example, scoring the expression of genes from plants harvested at the expansion of the 2nd leaf pair is not necessarily more restrictive than measuring morphology on the day of anthesis of the first flower. In a fundamental way, transcriptome studies are more inclusive because the set of traits considered in the final analysis is not driven by the inclinations of the investigator. Quite understandably, biologists focus on traits they hypothesize to be important determinants of fitness (drought stress, pollinator recruitment, deterrence of herbivores, etc). Unfortunately, the more accurate the intuition of biologists in choosing critical traits, the more biased our estimates of selection based on these macroscopic phenotypes will be. It is entirely possible that the strength and relative frequencies of directional, stabilizing, or disruptive selection will be systematically different between chosen traits and the rest of the phenotype.

Despite clear advantages, the volume of data produced by transcriptome studies forces quantitative genetics to confront a serious challenge of scale. Most studies of phenotypic selection utilize a regression framework. In the simplest implementation of this approach, an estimate of relative fitness (e.g., individual seed set divided by mean

seed set for the population) is regressed on a single phenotype in a univariate regression. In the context of gene expression, this would involve regressing an estimate of relative fitness on the expression of an individual gene, for all the individuals in the experimental population or sample. If expression has been standardized (i.e., $\bar{x} = 0$ and $\sigma = 1$), the resulting parameter estimate is the standardized selection differential for the expression of that gene; positive values would indicate that greater expression of the gene was associated with increased relative fitness. Groen et al. (2020) applied this approach to populations of rice growing under field and drought environments. They found that selection differentials for gene expression were generally weak, but stronger under drought than well-watered conditions.

The scale of the transcriptome introduces two key problems with the univariate approach. First, RNA-seq experiments estimate the expression of thousands of genes at a time. Simply repeating a univariate analysis for all the genes for which one has data introduces several inter-related problems. First, it is unlikely that the expression of each gene is independent of the expression of other genes, in the same way that a single macroscopic phenotype is often correlated with other phenotypes. Selection differentials measure total selection on a phenotype, which is the sum of direct selection on the trait and indirect selection through correlated traits (Lande and Arnold 1983). Because the expression of any individual gene is likely to be correlated with the expression of other genes (and other traits), a selection differential alone cannot tell whether it is the expression of a focal gene that is directly important for relative fitness, or whether the expression of that gene is simply correlated with other traits that are under selection. Second, testing the relationship between each gene's expression and fitness independently ignores the fact that it is impossible for these estimates to be independent when there are more 'traits' than there are observations (there are simply insufficient degrees of freedom). Lastly, analyzing the relationship between expression and fitness for each gene in succession introduces multiple testing problems in a hypothesis-testing framework: A large number of genes associated with relative fitness will undoubtedly be false positives. Addressing the number of tests thus requires multiple testing or false discovery rate corrections. We caution against obsession with significance testing and some of the methods we describe do not use it at all. However, it is important for investigators to realize that performing many thousands of tests at a time will incur false positives.

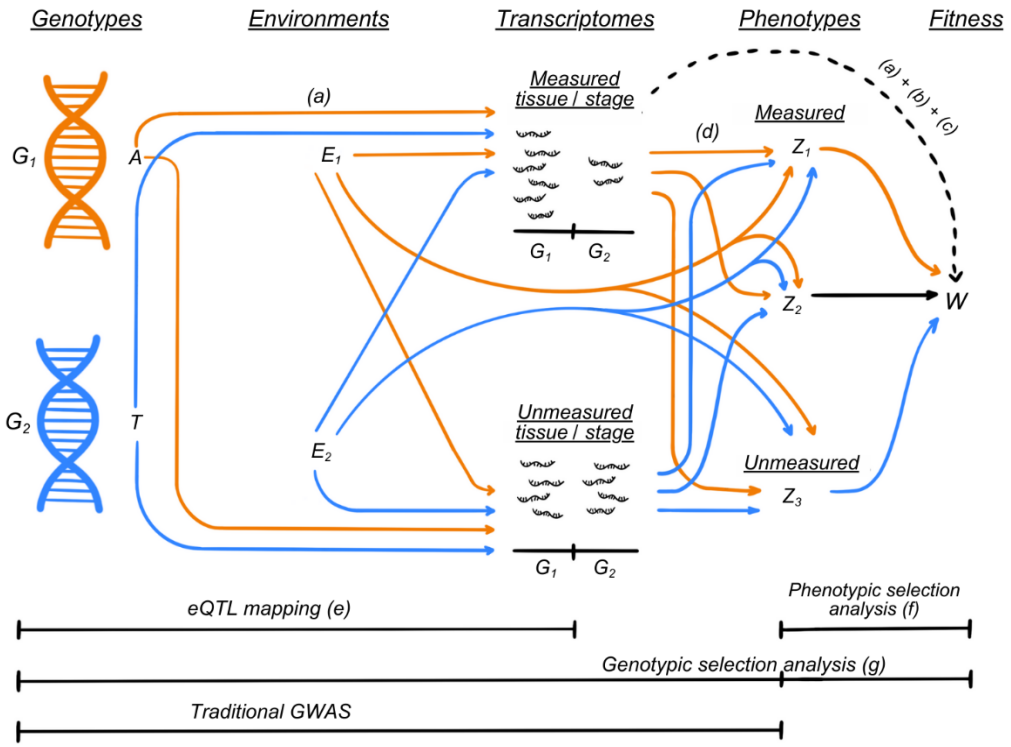

Figure 1. A schematic depicting the mapping from genotype to fitness. From left to right. We present a hypothetical case of two genotypes (G_1 and G_2, differing by a base-pair) and two environments (E_1 and E_2) to illustrate how genetic and environmental variation affect transcriptomes, phenotypes, and fitness. In the middle column, we highlight that genetic and environmental variation may lead to differences in expression in some tissues and stages (measured tissue/stage) but not others (in this case, an unmeasured tissue/stage). Expression and environmental variation, in turn, both affect macroscopic phenotypes (z_1, z_2, z_3). In this case, we highlight that while z_1 and z_2 have been measured, it is likely that unmeasured phenotypes (z_3) are affected by expression and also affect fitness. In the arrows leading to fitness, we note that expression can affect fitness directly (dotted arrow) and via phenotypes (z_1 and z_2). Bars across the bottom are labeled with common analytical approaches to understanding expression, the genetic basis of traits, and selection. Key paths: (a) Groen et al. (2020), (b) Stinchcombe and Henry (2025), (c) Figure 2, this paper, (d) Brown and Kelly (2022), (e) Josephs et al. (2015, 2020), (f) Lande and Arnold (1983), (g) Rausher (1992). Illustration by Martin R. Henry.

The standard approach for measuring selection on correlated traits is multiple regression (Lande and Arnold 1983; Figure 1 path (f)). In this context, a regression of relative fitness on the expression of all the genes in the transcriptome would yield selection gradients for gene expression. These gradients measure the direct effect of expression on relative fitness, accounting for the effects of the other traits (i.e.,

expression of other genes) included in the model. While promising in the abstract, with real data and sample sizes, such a model quickly runs into the *N-p* problem: there are far more parameters to estimate (*p*) than there are total samples (*N*) in even the most heroic of experiments. Consequently, one of the primary advantages of the Lande-Arnold (1983) approach, its ability to distinguish direct and indirect selection on correlated traits, is lost. Many of the questions outlined above—about the strength and form of selection, the prevalence of direct versus indirect selection, and even the fraction of the transcriptome subject to direct selection—remain inaccessible. In the remaining sections, we outline a handful of promising statistical and experimental approaches that can be used to address the *N - p* problem of measuring selection gradients for transcriptomes.

Selection Gradients for the Transcriptome: Statistical Approaches

There are several statistical approaches for measuring selection gradients for gene expression, and here we comment on some variants that appear to be emerging. Our expectation is that there will be continued work and that future developments are likely. At their core, these methods share one fundamental feature: dimensionality reduction, the compression of the data so as to estimate fewer parameters than the sample size. To use a hypothetical example, if an investigator has estimates of fitness for 500 individuals, and estimates of expression for >500 genes in those same 500 individuals, the goal of these approaches is to reduce the problem to estimating selection from far fewer than 500 parameters (so that *N* is greater than *p*).

PCA and Gene Coexpression Modules

The most straightforward approach is likely familiar to many users of selection gradient analysis, principal component analysis (PCA). Because PCA is a widely used technique and familiar to many biologists, we do not consider the mathematical or technical details of its implementation; Jolliffe (2002) provides an extensive coverage. In short, after a PCA, an investigator obtains independent axes capturing variation in the original traits. In many cases, far fewer axes (PCs) are required to describe the data than there were original traits. In the context of gene expression, these PC axes can be used as independent variables to predict relative fitness. An important point is that fewer– ideally far fewer– PC axes must be used than there were original traits, otherwise nothing is gained. Groen et al. (2020; Figure 1, path (a)) used this approach with PC axes, and were able to detect significant selection on several PC axes. They used these findings to detect selection on the expression of genes related to photosynthesis and growth.

One downside of this approach is that a PC axis simultaneously reflects all the individual traits included in the study. A PC score is a weighted average of the expression of all measured genes with the magnitude and direction of the weights differing among principal components, which can make them difficult to interpret. Chong et al. (2018) illustrate a method to 'back-transform' selection estimates for PC scores into selection estimates on the original traits. They argued these are much easier to interpret and suggested the technique would be useful for studies of selection on gene expression, metabolomics, and other high dimensional traits. In brief, one performs some matrix algebra computations involving selection estimates for PC scores and the eigenvectors of the original PCA. This rotation yields an estimate of a selection gradient on individual gene expression traits, accounting for the patterns of correlation among the traits, but only within the portion of multivariate space described by the PC axes included (Chong et al. 2018). Similar calculations can be performed to estimate standard errors for these reconstituted estimates of selection gradients for gene expression.

Henry and Stinchcombe (2025, Figure 1, path (b)) also used PCA to understand selection on gene expression. Like Groen et al. (2020), they regressed relative fitness on PC axes of gene expression. However, rather than using the PC axes as objects of study in themselves, they used the methods described by Chong et al. (2018) to back-transform selection on PC scores into selection gradient estimates for individual genes. In their study of *Ipomoea hederacea* (Ivyleaf morning glory), they had estimates of relative fitness for 96 individuals, and estimates of gene expression for 2,753 genes throughout the genome. The best model used 61 PCs to describe patterns of variation in gene expression, which collectively explained 55% of variation in relative fitness. Turning these back into selection gradients for the expression of individual genes suggest several important, if tentative, findings about selection on gene expression. First, they found a very strong positive relationship between selection differentials and selection gradients for gene expression, suggesting that most of the selection on gene expression was direct, rather than indirect due to the expression of other genes. Second, they found a wide distribution of selection gradients for expression, approximately symmetrical around zero: some genes were under selection for increased expression, and a similar number for decreased expression. Finally, they observed that selection gradients for gene expression were substantially smaller than their past findings of selection on size and life history traits in the same population (Henry and Stinchcombe 2023).

An alternative approach to dimensionality reduction is to first identify gene co-expression modules using programs like WGCNA (Langfelder and Horvath 2008). These modules are constructed by identifying sets of genes whose expression is more

strongly correlated with other genes in the module than with genes in other modules. The expression of the genes within a module can be summarized with PCA– so-called eigen-genes– and the PC1 score of a module can be estimated for each individual in a data set. These PC scores represent a weighted sum of gene expression of the genes within the module. As before, PC scores for a module's expression– which might summarize the expression of dozens to hundreds of genes– can be used as 'traits' in Lande-Arnold style analyses.

Several investigators have applied this approach, relating gene coexpression modules to aspects of plant performance, size, or life history traits that are likely to be under strong selection (e.g., Palakurty et al. 2018, Josephs et al. 2020, Brown and Kelly 2022). For example, Brown and Kelly (2022; Figure 1 path (d)) found that PC1 scores from twenty gene coexpression modules could explain 47% of variation in flower size in *Mimulus guttatus.* They used permutation testing to verify that these modules indeed significantly predicted flower size, and that the observed co-expression modules performed significantly better than random groupings of genes of the same size. In other words, the coexpression modules contain biological signal for predicting traits (in this case, flower size). Flower size is not itself a fitness component, but is under strong selection in *Mimulus guttatus* (Mojica and Kelly 2010), suggesting that transcriptomic variation affecting flower size can also potentially affect fitness. Interestingly, while several studies have related eigen-gene expression from coexpression modules to performance and fitness traits, to our knowledge none have used the PC rotations of Chong et al. (2018) to estimate selection gradients for expression of the individual genes within the module.

The use of gene co-expression modules entails both benefits and drawbacks that are worth considering. Co-expression modules have the benefit that individual genes appear in one and only one module. As a consequence, the interpretation of the expression of the entire module is more straightforward than the output of a PCA, where the expression of each gene will load onto all the PC axes. Discrete, non-overlapping modules, in our view, might offer greater biological interpretation of the types of genes (or GO categories) that are associated with any given module. One drawback of coexpression modules, or PC scores summarizing the expression levels of genes within a module (eigen-gene expression), is that the scores summarizing multiple modules are not guaranteed to be uncorrelated across a sample, in contrast to a PCA using all of the data. Consequently, understanding selection on multiple modules simultaneously may require multiple regression and the estimation of selection gradients.

Machine Learning

There is great enthusiasm for machine learning approaches in evolutionary biology (Schrider and Kern 2018). While this field is moving quickly and a full review is beyond our scope here (see Schrider and Kern 2018 for an entry point), there are several features of these algorithms that suggest promise in the context of measuring selection on gene expression. Machine learning approaches often focus on overall prediction rather than individual parameter estimation. In this context, it would be to predict relative fitness from expression of the set of genes for which investigators have expression, rather than hypothesis testing about the individual contribution of any one gene's expression. Several features of the mechanics of how the algorithms work aid this. First, data are often split into "training" and "testing" sets, which can prevent overfitting and noise being fit to the model, and allow an evaluation of the overall performance of the model. Second, many of the approaches identify features (gene expression in this case) in a way that reduces the overall number of parameters that are estimated, which is a start towards addressing the issue of the scale of the transcriptome. Third, in many cases the output of a machine learning algorithm is a measure of importance (e.g., the expression of these genes is important in determining whether an individual survives or dies before reproduction), rather than a parameter with a clear evolutionary interpretation like a selection differential or gradient.

Assuming that as an evolutionary biologist one has managed to implement one of the many machine learning algorithms available, and obtained a list of genes (features) whose expression is related to a fitness component, how does one make that information compatible and conversant with traditional measures of selection like differentials or gradients? One potential way forward is to use this reduced set of genes– that having survived cross-validation, evaluation in the testing data set, and acceptable performance metrics– appear to have expression that predicts relative fitness to estimate selection differentials and gradients the traditional way. In other words, one can use machine learning algorithms to prioritize an important subset of genes for further study, and then traditional selection analysis to estimate selection differentials and gradients.

In Henry and Stinchcombe's (2025) study, they used machine learning classification algorithms to determine which genes' expression were important for determining whether an individual set seed versus failed to set seed. After model fitting, they identified 278 genes whose expression was identified as important for determining whether an individual set seed or failed to set seed; 29 of these genes were also identified with PCA, having strong selection gradients for their expression. Interestingly, the distribution of selection differentials and gradients for the expression of these 29 genes was bimodal, with few instances of weak (near-zero) selection. In other words, the machine learning classifier identified genes whose expression was important for

successfully setting seed and these genes showed the strongest patterns of phenotypic selection.

Regularized regression

Many evolutionary biologists (including ourselves!) find aspects of machine learning to be a bit of a black box: its hard to fully visualize the functions and models being fit by the algorithms. This is especially in the case of neural networks where the output of one function is used as the input for another, in a series of layers. Fortunately, there's a set of statistical techniques closely related to machine learning– and indeed used by some machine learning algorithms– that is similar to the typical statistical toolkits of practicing evolutionary biologists. While to our knowledge regularized regression has not been used to estimate selection on gene expression, several features suggest that it could be useful.

Regularized regression is an analytical tool for fitting regressions with many predictors, varying degrees of multicollinearity between the predictors, and limited data (Morrisey 2014; Sztepanacz and Houle 2024). In contrast to ordinary least squares (OLS) univariate or multivariate regressions which estimate parameters by minimizing the sum of squared errors, regularized regressions minimize functions which include a penalty (Morrisey 2014; Sztepanacz and Houle 2024). As a result, individual parameter estimates are shrunk towards zero (i.e., regularized), which also reduces their variance. Parameter estimates obtained from regularized regression are biased compared to least-squares estimates, but the overall model predictive accuracy can be improved in the presence of a bias-variance trade-off. For these reasons, regularized regression approaches are likely to be of use in the case of multicollinearity (Chong et al. 2018; Sztepanacz and Houle 2024).

Sztepanacz and Houle (2024) performed a simulation study that illustrates the potential utility of regularized regression for measuring selection on multiple, potentially highly correlated traits. While their focus was not on gene expression, the lessons apply broadly. They showed that with limited data, and multicollinearity between predictors (as might be expected with the expression of thousands of genes as traits), regularized estimates provided more accurate estimates of the total strength of selection and the overall multivariate direction of selection. The frequentist implementation of regularized regression, however, does not yield traditional measures of uncertainty like standard errors and statistical significance for the individual predictors (Morrissey 2014; Sztepanacz and Houle 2024). While this is a potential limitation for future meta-analyses which require estimates of uncertainty for parameters, it is important to note that the importance of a gene's expression in predicting relative fitness can be judged from the

magnitude of the estimated parameters, especially because regularized regression approaches require the predictor data to be scaled to $\bar{x} = 0$ and $\sigma = 1$. In this manner, genes whose expression leads to large estimated coefficients are worthy of further investigation and follow up. To our knowledge, no one has yet used regularized regression to estimate natural selection on transcriptomes.

Measuring Selection Gradients at the Genotypic scale

In evolutionary quantitative genetics, it is common to distinguish phenotypic selection from response to selection. The former is the relationship between the multivariate phenotype and fitness while the latter is determined by the mapping from genotype to phenotype and requires an additional generation to measure. Separating selection from response enables an operational division of labor. Field studies without a genetic component can characterize selection, usually employing the Lande-Arnold regression framework. Given phenotypic selection estimates, an evolutionary response can be predicted using estimates of additive genetic variances and covariances from genetic experiments. Genetic statistics can be estimated from classical breeding designs or pedigrees or from genomic genotyping of individuals (Lynch and Walsh 1998).

The separation of selection from response is certainly convenient, but it is encumbered with serious assumptions (Morrissey et al. 2010). There are many situations in which it is advantageous to predict fitness from genetic statistics. One downfall of predicting fitness directly from phenotypic traits (of any variety) is the possibility that the relationship may be environmentally induced (Mitchell-Olds and Shaw 1987; Price et al. 1988; Rausher 1992). For plant systems, it is easy to envision that individuals growing in high resource soils (e.g., high N, P, or K) both have higher fitness and also larger values of traits requiring N and P– for example, size, branching, or plant defense traits. In this instance, a naive application of the Lande-Arnold approach would detect selection on these traits even if size, branching, and plant defense have no effect on fitness at all. In this scenario, both fitness and the other phenotypic traits are responding, independently, to soil resource variation, and investigators observe an environmentally induced relationship rather than a causal one. In regression terms, one has omitted a ‘trait’ (in this case, soil NPK concentrations) that is correlated with both the predictors and the response variable, leading to inaccurate parameter estimates. Importantly, such relationships will not lead to responses to selection and evolutionary change (Rausher 1992). It is highly likely that gene expression, as a trait, will be environmentally sensitive to aspects of soils, temperature, weather, abiotic and biotic conditions, and a multitude of other influences. *A priori*, this suggests that the potential for environmental covariances to bias estimates of phenotypic selection on gene expression is high.

Fortunately, Rausher (1992; Figure 1, path (g)) provided a solution to this problem: estimating selection using either breeding values or estimates of genotypic values for both phenotypes and fitness. While this approach comes at a cost of sample size and statistical power (Stinchcombe et al. 2002), covariances estimated with breeding values between fitness and phenotypes reflect genetic relationships, rather than environmentally induced ones. The resulting parameter estimates of selection are more accurate, and reflect relationships that have the potential to produce evolutionary change (Stinchcombe et al 2002). While formal studies remain rare (but see Stinchcombe et al. 2002 and Hadfield 2008), existing evidence suggests that many estimates of phenotypic selection on macroscopic traits are highly biased by environmental covariances (Stinchcombe et al. 2002; Kruuk 2002; Morrissey et al. 2012; Hajduk et al. 2020).

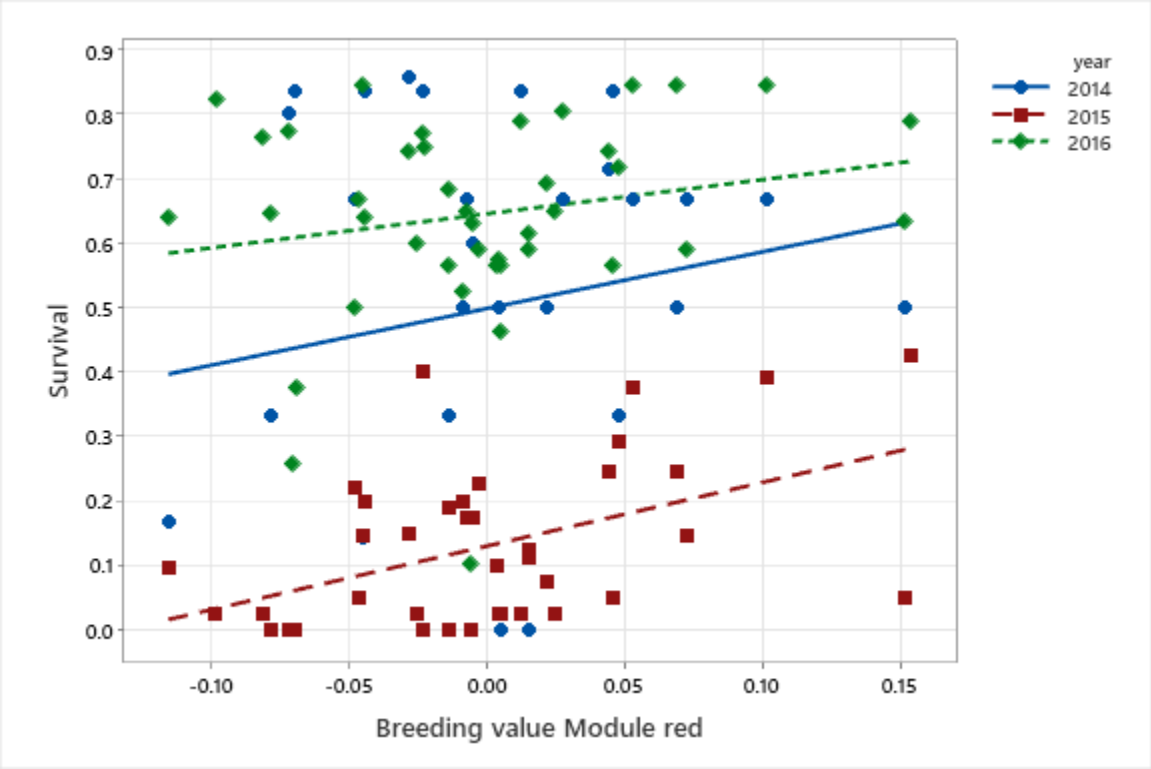

Figure 2. Selection gradients for expression module 'Red' in *M. guttatus* were positive for survival in all three years. The overall effect of Red on survival to flower (all years included) is significantly positive ($F_{1,107} = 6.43$, $p = 0.013$), although only the 2015 regression, where survival was generally low, is significant when considered in isolation ($F_{1, 38} = 11.07$, $p < 0.002$).

The breeding value regression approach is as applicable to gene expression as it is to macro-phenotypes. This is also true for the compression methods discussed in the previous section (e.g., PCA and WGCNA). They can be applied as readily to breeding value regressions as to phenotypic regressions. To illustrate, we revisit the gene expression modules of Brown and Kelly (2022) which were obtained for homozygous

lines of *Mimulus guttatus*. Many of these same lines were intercrossed to make F_1 plants and then measured for survival and reproductive success in the natural habitat by Troth *et al.* (2018). With additive inheritance, the breeding values (a.k.a. additive genetic values) for the F_1 plants are the average of the values obtained from the parental lines (Lynch and Walsh 1998). Therefore, we can use gene expression estimates obtained in the greenhouse to predict fitness in nature. Over three successive generations, one expression module (Red) was a consistent predictor of survival to flower in each of the three field seasons (**Fig 2;** note that Fig. 2 is path (c) in Fig. 1). In the greenhouse, genotypes with high values for 'Red' are associated with earlier germination (Brown and Kelly 2022).

Unlike phenotypes, which are unique features of individual plants, breeding values are 'portable'; indeed, this feature is at the heart of their success in agricultural applications. Breeding values can be carried across experiments whenever genotypes can be replicated. Portability enables highly powered experiments because gene expression can be studied on large samples of plants grown under controlled conditions (e.g., the greenhouse environment). Also, because breeding values of expression are the predictors of field fitness, we avoid the serious difficulty of environmental factors inducing spurious correlations between phenotype and fitness. The downside is that expression levels in the greenhouse could prove to be the "wrong traits." The same genes could be expressed in different ways under field conditions than under those used to obtain breeding values, or different genes could be expressed in response to different environments. This would be an example of genotype by environment interaction, where the amount of expression, or which genes are expressed, depends on the environmental context (greenhouse or field). Of course, this is always a concern with RNA-seq studies, whether the transcripts sampled from a particular tissue at a particular life stage are the most relevant determinants of phenotype and/or fitness.

To date, there has been a great deal more work on the genetic basis of transcriptional variation than on how this variation affects fitness. Research on the genetics of gene expression has also been confronting the issue of scale. Above, we discussed PCA and WGCNA based on the 'P matrix', the variances and covariances among plants in the expression level of each gene (the phenotype). An alternative approach is to partition the phenotypic variance into genetic and environmental components and then apply the compression to these underlying components. For example, Blows et al (2015) show that the genetic component of variation in expression of 8750 genes of *Drosophila serrata* could be distilled into the contributions of a much smaller number of underlying variables using matrix completion methods.

A distinct but related approach to understanding gene expression evolution is to apply factor analysis or latent factor modelling. These approaches are common in psychology and other disciplines but have received less adoption in evolutionary biology (for exceptions, see McGuigan and Bows 2010; Frichot et al. 2013). In the context of gene expression, the idea is that the expression of each gene is influenced by a limited number of underlying 'factors.' These factors are not directly observed but can be modeled and estimated from data. Variation in factors can be partitioned into genetic and environmental components, and through the mapping from the factors to the expression levels of genes, one can characterize the variances and covariances for the entire transcriptome. The problem thus shifts from analyzing the genetic variances and covariances in the expression of thousands of individual genes to understanding the variances and covariances of a much more limited set of inferred factors. Two implementation methods– Bayesian Sparse Factor Analysis (BSFA; Runcie and Mukherjee 2013; for examples, see Hine et al. 2018, 2022) and MegaLMM (Runcie et al. 2021)– have been developed that are suited to predicting the high dimensional structure of genetic variances and covariances of the transcriptome from a more limited set of variables. These approaches provide a natural means to reduce the dimensionality of the determination of gene expression levels from genetic and environmental influences. Correlations between expression levels emerge when different genes share a common factor.

Factor analysis could be applied to estimate selection on the transcriptome in either of two distinct ways. The first would be to apply BSFA or MegaLMM strictly to the partitioning of transcriptome variation into genetic and environmental components, without including fitness variables in the model. Given estimated breeding values for factors, one could predict field fitness in a way analogous to the *Mimulus* example of Fig 2 (except using factors instead of module PC scores). This approach addresses the scale issue because factors are uncorrelated with each other. Moreover, given the mapping from factors to expression levels, one can extrapolate from selection gradients on factors to gradients on individual genes. The second way would be to apply factor analysis to transcriptomes and fitness measurements simultaneously. This is essentially adding fitness measures to the list of phenotypes (transcript levels). One then estimates genetic and environmental covariances among the expression levels of genes simultaneously with their covariances with fitness. Estimated factors with strong contributions from fitness would be identified as under selection. Genes whose expression loaded heavily on those factors are thus under selection.

The simultaneous approach has the advantage that the sparse factor model directly estimates the genetic covariance between fitness and gene expression. This is the predicted change in the mean expression level into the next generation (Robertson

1966, Price 1970, 1972). The two-part method is more consistent with the traditional quantitative genetic approach based on regression where we distinguish traits as independent variables (predictors) and fitness components as dependent variables. Oftentimes, the joint distribution for traits can be treated with a multivariate normal distribution. However, fitness components are usually non-normal (e.g. binary for survivorship, negative binomial for fecundity, etc). It may be easier to accommodate the differing distributions for transcript variation and fitness components in a regression framework. A second reason to separate fitness from characterization of transcriptome variation is that we often expect the relationship between trait values and fitness to be non-linear due to stabilizing, disruptive, or correlational selection. Regardless of whether investigators use the simultaneous or two-part approach, we again note that doing so with genetic estimates or breeding values is likely to be superior to purely phenotypic analyses because of the problem of environmentally induced covariances (Rausher 1992).

Conclusions

Several common themes emerged from our overview of techniques for characterizing selection on the transcriptome even though many techniques are still in areas of active development. First, at their heart, most of the approaches we have discussed approach the *N-p* problem through some form of compression and reduction in the number of parameters that have to be estimated. As long as the sample sizes for the number of genes for which expression is measured with sequencing technologies exceed the number of individuals in experiments, some form of data reduction or compression will remain a requirement.

Second, we perceive distinct analysis paths which investigators can take, based on the data in hand and the tractability of the system. For species in which it is possible to perform breeding designs, create known and replicated genotypes, and/or generate inbred lines, analyses based on breeding values should be pursued. In these systems, gene expression can be measured in the greenhouse or growth chamber and fitness estimates obtained from the same genotypes (or relatives with predictable breeding values). In the case of inbred lines, successive estimates of transcriptomes, performance, and fitness could be obtained from immortalized genotypes that are exposed to a variety of growth conditions. In contrast, for species or systems where it is difficult to obtain immortalized genotypes– or where cost constraints preclude characterizing the transcriptomes of many genotypes– estimates of selection on the transcriptome are more akin to the field studies of selection on macro traits that followed the Lande and Arnold (1983) paper. The rich picture of how natural selection acts on morphological, behavioral, and life-history phenotypes is from a set of studies similar in

design to a single-instance measurement of selection on the transcriptome (Henry and Stinchcombe 2025). We have drastically fewer estimates of selection on transcriptomes to characterize its strength, mode, and spatial or temporal consistency, perhaps because the approach and technology are in early development. More than 40 years ago, Arnold (1983) coined the expression “morphology, performance, fitness” in a landmark paper describing how to understand variation in, and selection on, morphology. We suggest that an important area of research in the next 40 years of evolutionary biology will be to explore the mapping from gene expression to phenotype to fitness.

Acknowledgments

We thank our funding sources (NSF grants MCB-1940785 and FAIN 2421689 for JKK; NSERC Canada for JRS) for support. We thank Jacqueline Sztepanacz, Georgia Henry, Emily Josephs, Aneil Agrawal, and Stephen Wright for past and ongoing conversations on gene expression evolution. JRS thanks the Swedish Collegium for Advanced Study, and Goran Arnqvist, Jon Agren, Locke Rowe, Mario Vallejo Marin, and Martin Lascoux for influential discussions during the pre-natal stages of this project. Finally, we thank Dave Des Marais, two anonymous reviewers, Luis Madrigal Roca, and Samson Acoca Pidolle for comments on the manuscript.

References

Arnold SJ. **1983**. Morphology, performance and fitness. *American zoologist* **23**: 347–361.

Blows MW, Allen SL, Collet JM, Chenoweth SF, McGuigan K. 2015. The phenome-wide distribution of genetic variance. *The American naturalist* **186:** 15–30.

Brown KE, Kelly JK. **2022**. Genome-wide association mapping of transcriptome variation in *Mimulus guttatus* indicates differing patterns of selection on *cis*- versus *trans*-acting mutations. *Genetics* **220**: :iyab189.

Campbell-Staton SC, Cheviron ZA, Rochette N, Catchen J, Losos JB, Edwards SV. 2017. Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. *Science* **357**(6350): 495-498.

Chong VK, Fung HF, Stinchcombe JR. **2018**. A note on measuring natural selection on principal component scores. *Evolution letters* **2**: 272–280.

Frichot E, Schoville SD, Bouchard G, François O. **2013**. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular biology and evolution* **30**: 1687–1699.

Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **525**(7569): 372-375.

Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. **2018**. Erratum: Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* **555**: 688.

Gilad Y, Oshlack A, Rifkin SA. **2006**. Natural selection on gene expression. *Trends in Genetics* **22**: 456–461.

Groen SC, Ćalić I, Joly-Lopez Z, Platts AE, Choi JY, Natividad M, Dorph K, Mauck WM 3rd, Bracken B, Cabral CLU, *et al.* 2020. The strength and pattern of natural selection on gene expression in rice. *Nature* **578**: 572–576.

Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating. Proceedings of the Royal Society B: Biological Sciences 275:723–734.

Hajduk GK, Walling CA, Cockburn A, Kruuk LEB. **2020**. The 'algebra of evolution': the Robertson-Price identity and viability selection for body mass in a wild bird population. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **375**: 20190359.

Hamann E, Pauli CS, Joly-Lopez Z, Groen SC, Rest JS, Kane NC, Purugganan MD, Franks SJ. 2021. Rapid evolutionary changes in gene expression in response to climate fluctuations. *Molecular Ecology* **30**(1): 193-206.

Henry GA, Stinchcombe JR. **2023**. Strong selection is poorly aligned with genetic variation in *Ipomoea hederacea*: implications for divergence and constraint. *Evolution* **77**: 1712–1719.

Henry GA, Stinchcombe JR. **2025**. Predicting fitness-related traits using gene expression and machine learning. *Genome biology and evolution* **17**: evae275.

Hill WG. **2010**. Understanding and using quantitative genetic variation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **365**: 73–85.

Hine, E., D. E. Runcie, K. McGuigan, and M. W. Blows. 2018. Uneven distribution of mutational variance across the transcriptome of *Drosophila serrata* revealed by high-dimensional analysis of gene expression. Genetics 209:1319–1328.

Hine, E., D. E. Runcie, S. L. Allen, Y. Wang, S. F. Chenoweth, M. W. Blows, and K. McGuigan. 2022. Maintenance of quantitative genetic variance in complex, multitrait phenotypes: the contribution of rare, large effect variants in 2 *Drosophila* species. Genetics 222:iyac122**Jolliffe IT**. **2002**. *Principal Component Analysis*. NY, NY, USA: Springer.

Josephs EB, Lee YW, Stinchcombe JR, Wright SI. **2015**. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 15390–15395.

Josephs EB, Lee YW, Wood CW, Schoen DJ, Wright SI, Stinchcombe JR. **2020**. The evolutionary forces shaping *cis*- and *trans*-regulation of gene expression within a population of outcrossing plants. *Molecular biology and evolution* **37**: 2386–2393.

King M-C, Wilson AC. **1975**. Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P. **2001**. The strength of phenotypic selection in natural populations. *The American naturalist* **157**: 245–261.

Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM. **2012**. Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary ecology* **26**: 1101–1118.

Kruuk LEB, Slate J, Pemberton JM, Brotherstone S, Guinness F, Clutton-Brock T. **2002**. Antler size in red deer: Heritability and selection but no evolution. *Evolution* **56**: 1683–1695.

Lande R, Arnold S. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210-1226.

Langfelder P, Horvath S. **2008**. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**: 559.

Liu, Xuanyao, Li, Yang I. and Pritchard, Jonathan K. 2019. Trans effects on gene expression can drive omnigenic inheritance. *Cell* 177 (4): 022-1034.e6 https://doi.org/10.1016/j.cell.2019.04.014

Lynch M, Walsh B. **1998**. *Genetics and the analysis of quantitative traits*. Sunderland, MA, USA: Sinauer Associates.

McGuigan K, Blows MW. **2010**. Evolvability of individual traits in a multivariate context: partitioning the additive genetic variance into common and specific components. *Evolution* **64**: 1899–1911.

Mitchell-Olds T, Shaw RG. **1987**. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* **41**: 1149–1161.

Mojica JP, Kelly JK. **2010**. Viability selection prior to trait expression is an essential component of natural selection. *Proceedings of the Royal Society B: Biological Sciences* **277**: 2945–2950.

Morrissey MB, Kruuk LEB, Wilson AJ. **2010**. The danger of applying the breeder's equation in observational studies of natural populations. *Journal of evolutionary biology* **23**: 2277–2288.

Morrissey MB, Parker DJ, Korsten P, Pemberton JM, Kruuk LEB, Wilson AJ. **2012**. The prediction of adaptive evolution: Empirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution* **66**: 2399–2410.

Morrissey MB. **2014**. In search of the best methods for multivariate selection analysis. *Methods in Ecology and Evolution* **5**: 1095–1109.

Oleksiak MF, Churchill GA, Crawford DL. **2002**. Variation in gene expression within and among natural populations. *Nature genetics* **32**: 261–266.

Palakurty SX, Stinchcombe JR, Afkhami ME. **2018**. Cooperation and coexpression: How coexpression networks shift in response to multiple mutualists. *Molecular ecology* **27**: 1860–1873.

Price GR. **1970**. Selection and covariance. *Nature* **227**: 520–521.

Price GR. **1972**. Extension of covariance selection mathematics. *Annals of human genetics* **35**: 485–490.

Price T, Kirkpatrick M, Arnold SJ. **1988**. Directional selection and the evolution of breeding date in birds. *Science* **240**: 798–799.

Rausher MD. 1992. The measurement of selection on quantitative traits: biases due to the environmental covariances between traits and fitness. *Evolution* **46**: 616-626.

Rifkin SA, Houle D, Kim J, White KP. 2005. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* 438: 220–223.

Robertson A. 1966. A mathematical model of the culling process in dairy cattle. *Animal production* **8:** 95–108**.**

Runcie DE, Mukherjee S. **2013**. Dissecting high-dimensional phenotypes with Bayesian Sparse Factor Analysis of genetic covariance matrices. *Genetics* **194**: 753–767.

Runcie DE, Qu J, Cheng H, Crawford L. **2021**. MegaLMM: Mega-scale linear mixed models for genomic predictions with thousands of traits. *Genome biology* **22**: 213.

Schrider DR, Kern AD. 2018. Supervised machine learning for population genetics: A new paradigm. *Trends in genetics* 34: 301–312.

Siepielski AM, Gotanda KM, Morrissey MB, Diamond SE, DiBattista JD, Carlson SM. 2013. The spatial patterns of directional phenotypic selection. *Ecology letters* 16: 1382–1392.

Slotte T, Bataillon T, Hansen TT, St Onge K, Wright SI, Schierup MH. **2011**. Genomic determinants of protein evolution and polymorphism in *Arabidopsis*. *Genome biology and evolution* **3**: 1210–1219.

Stinchcombe JR, Rutter MT, Burdick DS, Tiffin P, Rausher MD, Mauricio R. **2002**. Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *The American naturalist* **160**: 511–523.

Svensson EI. **2023**. Phenotypic selection in natural populations: what have we learned in 40 years? *Evolution* **77**: 1493–1504.

Sztepanacz JL, Houle D. **2024**. Regularized regression can improve estimates of multivariate selection in the face of multicollinearity and limited data. *Evolution letters* **8**: 361–373.

Troth A, Puzey JR, Kim RS, Willis JH, Kelly JK. 2018. Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science* **361**(6401): 475-478.

Wright SI, Yau CBK, Looseley M, Meyers BC. **2004**. Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Molecular biology and evolution* **21**: 1719–1726.