

Meta-analysis: Congruence of genomic and phenotypic differentiation across diverse natural study systems

Zachary T. Wood^{1,2}  | Andrew K. Wiegardt³ | Kayla L. Barton⁴  |
Jonathan D. Clark³ | Jared J. Homola⁵ | Brian J. Olsen^{2,6} | Benjamin L. King⁴ |
Adrienne I. Kovach³ | Michael T. Kinnison^{1,2}

¹School of Biology and Ecology, University of Maine, Orono, ME, USA

²Maine Center for Genetics in the Environment, Orono, ME, USA

³Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH, USA

⁴Department of Molecular & Biomedical Sciences, University of Maine, Orono, ME, USA

⁵Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA

⁶Department of Wildlife, Fisheries, and Conservation Biology, University of Maine, Orono, ME, USA

Correspondence

Zachary T. Wood, School of Biology and Ecology, University of Maine, Orono, ME, USA.

Email: zachary.t.wood@maine.edu

Funding information

NSF OIA 1826777; Maine Agricultural and Forest Experiment Station. This project was supported by the USDA National Institute of Food and Agriculture, Hatch Project number ME022207 through the Maine Agricultural & Forest Experiment Station. This article is Maine Agricultural and Forest Experiment Publication Number 3824.

Abstract

Linking genotype to phenotype is a primary goal for understanding the genomic underpinnings of evolution. However, little work has explored whether patterns of linked genomic and phenotypic differentiation are congruent across natural study systems and traits. Here, we investigate such patterns with a meta-analysis of studies examining population-level differentiation at subsets of loci and traits putatively responding to divergent selection. We show that across the 31 studies (88 natural population-level comparisons) we examined, there was a moderate ($R^2 = 0.39$) relationship between genomic differentiation (F_{ST}) and phenotypic differentiation (P_{ST}) for loci and traits putatively under selection. This quantitative relationship between P_{ST} and F_{ST} for loci under selection in diverse taxa provides broad context and cross-system predictions for genomic and phenotypic adaptation by natural selection in natural populations. This context may eventually allow for more precise ideas of what constitutes “strong” differentiation, predictions about the effect size of loci, comparisons of taxa evolving in nonparallel ways, and more. On the other hand, links between P_{ST} and F_{ST} within studies were very weak, suggesting that much work remains in linking genomic differentiation to phenotypic differentiation at *specific* phenotypes. We suggest that linking genotypes to specific phenotypes can be improved by correlating genomic and phenotypic differentiation across a spectrum of diverging populations within a taxon and including wide coverage of both genomes and phenomes.

KEYWORDS

candidate gene approaches, F_{ST} , GWAS, natural selection, outlier analysis, P_{ST}

1 | INTRODUCTION

Quantifying the relationship between genomic and phenotypic (Box 1) population differentiation is fundamental to characterizing the

genomic basis for phenotypic evolution (Rodríguez-Verdugo et al., 2017). Understanding the association between genes and phenotypes in natural populations also has the potential to reveal generalizable patterns of evolution (Feder & Mitchell-Olds, 2003; Gallant &

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd.

O'Connell, 2020; Rudman et al., 2018). The genomic basis for adaptive evolution also has profound implications for evolutionary conservation, including genetic evolutionary management (Hoffmann et al., 2015; Kinnison & Hairston, 2007) and evolutionary rescue (Carlson et al., 2014). A universal pattern of congruent differentiation in genetic loci and phenotypic traits (i.e., a similar positive relationship between population-level genomic and phenotypic differentiation for traits and loci putatively under selection) in natural populations would have many theoretical and practical benefits, including context for interspecific comparison of genomic and phenotypic differentiation and generalizable patterns of genomic and phenotypic adaptation. Comparing individual results to generalizable patterns would allow us to address questions such as (1) what constitutes "large" differentiation, (2) whether certain loci have relatively strong effects on phenotypes, and (3) whether nonparallel adaptations are similar in their scope of differentiation, if not in trait pathways.

To date, however, the genomic architecture of phenotypic change in most natural populations remains poorly understood, and studies of adaptive population genomics greatly outnumber studies linking genomic change to adaptive phenotypic change (Hendry, 2013, 2016a). Recent technological advances have made sequencing large or whole portions of genomes possible for many nonmodel species (Bolger et al., 2019; Cuperus & Queitsch, 2020; Davey et al., 2011; Goodwin et al., 2016; Russell et al., 2017; Whibley et al., 2021), but are the patterns from these studies generalizable? Specifically, does this growing body of literature support the premise that greater phenotypic differentiation corresponds with greater genomic differentiation in natural organisms (controlling for the number of contributing loci)? Here, we examine this link via standardized measures of genomic differentiation (F_{ST}) and phenotypic differentiation (P_{ST})—while assessing potential interacting effects associated with different study designs (Box 2). While details of species-specific genomic architecture certainly affect this link, we sought generalizable patterns at a broader scale, particularly for when information on these specifics is lacking.

The keystone fact of the Modern Evolutionary Synthesis is the genetic basis for evolution (Fisher, 1930; Huxley, 1942). While phenotypes determine fitness, their heritable, genetic basis controls the response of phenotypes to selection and their persistence in time. In a small but growing number of cases, clear relationships between phenotypes subject to natural selection and their associated genes (e.g., Barrett et al., 2019; Colosimo et al., 2004) have been identified in natural populations. However, the ability to associate genetic variation with phenotypes in natural populations—where environmental conditions are beyond manipulation—remains challenging (Hendry, 2013, 2016a). Nonetheless, substantial progress in linking genetic and phenotypic variation has been made in limited cases (i.e., genome-wide association studies; GWAS: Visscher et al., 2017).

Despite this progress, biologists have struggled to systematically associate genomic data with biologically relevant phenotypes, particularly when pleiotropy, polygenic inheritance, epistasis, and phenotypic plasticity confound their relationship (Pigliucci & Muller, 2010;

BOX 1 What is a phenome?

Numerous papers use the word *phenome*, a phenotypic analogue to *genome* (Bogue et al., 2018; Burnett et al., 2020; Freimer & Sabatti, 2003; Oti et al., 2008). A genome is the combination of all coding material and corresponding noncoding material in an organism, that is, its DNA or RNA. In theory, a genome can be objectively characterized, though sequencing and alignment choices add a layer of subjectivity to the process. A phenome is the combination of all phenotypic traits of an organism (Oti et al., 2008). Phenotypic traits inherently contain a degree of subjectivity, as we must define phenotypes in order to measure them. For example, a phenotype such as bird wing morphology could be defined by wing size, wing mass, wing shape, number of wing feathers, developmental architecture, and/or other attributes. Some phenotypes, such as behavior, become even more difficult to comprehensively describe. The phenome is also theoretically infinitely large, as measures of phenotypes are constrained in number only by our imagination. One means for limiting the size of the phenome is to consider only ecologically relevant functional traits (ERFTs), or functional traits that affect organismal fitness and interactions with the environment (Wood et al., 2021).

Walsh & Lynch, 2018). In a large proportion of studies of adaptation in natural populations, genomic variation is analyzed for signals of selection without any direct quantification of biologically relevant phenotypic trait variation. As such, most literature on the heritable basis of adaptation tends to focus primarily on the characterization of either genomic or phenotypic variation in natural populations, but not explicitly link the two. Genome-wide association studies (GWAS) provide one avenue to explore genotype–phenotype relationships in natural populations, but are often plagued by high false-positive rates and commonly struggle to detect the small genetic effect sizes of many polygenic traits (Chen et al., 2021; Evangelou & Ioannidis, 2013). This growing body of studies has attempted to associate genomic and phenotypic aspects of adaptation in the same diverging populations of organisms. These studies in turn provide a means to assess how genomic and phenotypic variation are distributed among populations experiencing ongoing adaptive differentiation. Genetic and phenotypic differentiation are particularly useful for linking genotypes to phenotypes, as they produce phenotypic and genetic variation, which can then be harnessed statistically for GWAS or outlier studies (Gibson, 2018; Visscher & Goddard, 2019).

While numerous phenotypic traits clearly have a heritable basis, their underlying genomic architecture is rarely fully—or even mostly—explained, leading to what is sometimes called the “missing heritability problem” (Young, 2019; Zuk et al., 2012). This is not entirely unexpected given the great complexity of genomes

BOX 2 F_{ST} and P_{ST}

F_{ST} is the proportion of genetic variation associated with population structure. As populations become more differentiated, the proportions of alleles at various loci will become less similar across populations, leading to increasing between-population variation. In principle, F_{ST} can be calculated as:

$$F_{ST} = \frac{\sum_j (p_j - \bar{p})^2 N_j}{\bar{p}(1 - \bar{p}) \sum_j N_j}$$

where j represents a population; p_j = the frequency of allele p in population j ; \bar{p} = the frequency of allele p across all populations; and N_j = the number of individuals censused in population j .

P_{ST} is a phenotypic analogue for F_{ST} and measures the proportion of phenotypic variation associated with population structure. The more differentiated two populations are for a particular phenotype, the greater the proportion of phenotypic variation will be explained by population structure, and the higher P_{ST} will be. P_{ST} can be calculated as:

$$P_{ST} = 1 - \frac{\sum_j \sum_i (x_{ji} - \bar{x}_j)^2}{\sum_j \sum_i (x_{ji} - \bar{x})^2}$$

where the numerator is the sum of squared deviations of each individual's phenotype (x_{ji}) from the population mean (\bar{x}_j) across all populations; and the denominator is the sum of squared deviations of each individual's phenotype (x_{ji}) from the metapopulation mean (\bar{x}).

P_{ST} is analogous to several other common statistics of variance, including R^2 and F ratios. For a model in which the only independent variable is a categorical variable for population:

$$P_{ST} = R^2$$

For any model including a variable for population:

$$P_{ST} = \frac{f_S \frac{v_{nS}}{v_{dS}}}{1 + \sum_Z f_Z \frac{v_{nZ}}{v_{dZ}}}$$

where f_S = the F ratio for the population variable; v_{nS} and v_{dS} are the numerator and denominator degrees of freedom for the population variable; and Z represents all other covariates in the model.

For mean and standard deviation data:

$$P_{ST} = \frac{\sum_j s_j^2 N_j}{\sum_j s_j^2 N_j + \sum_j (\bar{x}_j - \bar{x})^2 N_j}$$

where j indicates a population; s_j = standard deviation for population j ; N_j = number of individuals censused in population j ; and \bar{x}_j -bar and \bar{x} -bar indicate population-specific and metapopulation means, respectively.

Q_{ST} is calculated the same way as P_{ST} , but only applies to phenotypes for which the heritable component has been isolated, usually via common-rearing experiments.

and phenomes, and the constraints both present for statistical power (López-Cortegano & Caballero, 2019; Uricchio, 2020). This recognized challenge has led to substantial innovation—and thus variability—among investigators and studies in methods used to

associate genomic and phenotypic differentiation (Burt & Munafò, 2021). Despite a growing number of approaches, no clear best practices exist for linking genotype to phenotype across systems. Each method has substantial limitations, and the lack of best practices

adds noise to any attempt to detect underlying trends common across the tree of life (Tam et al., 2019).

For example, while gene-knockout experiments provide an ideal means of studying how variation in a particular candidate gene determines phenotype when the species can be reared in a laboratory setting (Hall et al., 2009), these experiments are prohibitive or unethical for studies of most nonmodel species in natural systems. Further key choices in study system, study design, genomic data collection, and analytical approach all likely influence calculations of genomic and phenotypic differentiation in idiosyncratic ways. Controlling for methodological variation can therefore potentially reveal more generalizable patterns that may help in understanding the relationship between genomic and phenotypic differentiation, allowing for cross-system comparisons and generalizations about responses of natural populations to selection.

Here, we conduct a meta-analysis of 31 studies of natural populations representing 88 unique multi-population comparisons that demonstrate putative genomic and phenotypic differentiation in response to selection. We address two main questions:

How does genomic differentiation at loci under selection explain phenotypic differentiation, both across and within studies? Under ideal conditions, when all the loci underlying a phenotypic trait are identified and the phenotype is accurately quantified, we would expect a strong, positive relationship between P_{ST} and F_{ST} for loci under selection (Brommer, 2011; Kaeuffer et al., 2012; Raeymaekers et al., 2007). However, measuring numerous genotypes and phenotypes inherently leaves much room for error, even beyond methodological nuances, as not all highly differentiated loci will code for highly differentiated phenotypes, and some important loci may exhibit little differentiation, muddying the relationship between P_{ST} and F_{ST} . Fundamental differences in genomic architecture—including the strength of individual loci (many weak vs. few strong), linkage, and gene interactions—across taxa and traits will also obscure the relationship between P_{ST} and F_{ST} (Keane et al., 2011). Finally, some phenotypic differentiation will simply be explained by neutral genomic differentiation (Raeymaekers et al., 2017; Whitlock, 2008; Zhang, 2018).

How do key methodological choices affect the strength of the genome-to-phenome association? Differences in methodological choices for genome-to-phenome studies are likely to affect not only conclusions about the extent of genomic and phenotypic differentiation, but the expected relationship between P_{ST} and F_{ST} as well. As some genomic markers are more likely to fall in or near coding or modifier regions (Box 3), those markers may have stronger relationships with phenotypes. Furthermore, smoothing or adjusting F_{ST} and correcting for false-positive rates may improve statistical error rates, but bias the relationships between P_{ST} and F_{ST} , in part by changing the proportion of the genome characterized as “non-neutral” (Lotterhos & Whitlock, 2014; Luu et al., 2017). Finally, common-rearing experiments may avoid some of these challenges by isolating genetic differences in phenotypes, but they also remove gene-by-environment interactions, which are important genetically based sources of phenotypic variation (Via & Lande, 1985).

2 | METHODS

2.1 | Overview

We used Web of Science (<https://webofknowledge.com>) and the search terms in Table S1 to isolate 88 population-level comparisons that included phenotypic and genotypic data from two or more populations under purported divergent selection (Table 1). We extracted or calculated three metrics from each paper for all possible pairwise comparisons of each phenotype measured between populations: (1) P_{ST} , phenotypic differentiation, (2) neutral F_{ST} , neutral genomic differentiation, and (3) non-neutral F_{ST} (henceforth: nnF_{ST}), genetic differentiation for loci putatively under selection (i.e., candidate genes, outlier loci, or loci associated with a differentiated phenotype in a GWAS) (Wright, 1949). We also included methodological covariates, including the method of determining loci under selection (Box 4), type of genetic marker used (Box 3), the software used to calculate F_{ST} , whether the study included a common-garden design, and the proportion of loci identified as non-neutral.

We used general linear models to examine the relationship between phenotypic differentiation (P_{ST}) and (1) neutral and non-neutral genetic differentiation (F_{ST} , nnF_{ST}), (2) proportion of loci identified as non-neutral, and (3) several methodological choices. The result is several models that assess the degree to which phenotypic and genomic differentiation are congruent (in traits and loci putatively under selection), as well as the role of some potential confounding methodological factors.

2.2 | The database

We used all databases within the online citation database Web of Science and the 25 search terms in Table S1 to find relevant papers that included phenotypic and genomic data from two or more populations undergoing divergent selection. Searches returned anywhere from 0 to 340,088 papers; we retained papers revealed by searches with <700 results (Table S1). We used R version 3.6.1 (R Core Team, 2019) and the packages *metagear* (Lajeunesse, 2016) and *Bibliometrix* (Aria & Cuccurullo, 2017) to screen the abstracts of each paper to determine whether the paper was likely to contain both genomic and phenotypic comparisons for multiple populations. For consistency, the same observer (ZTW) reviewed every abstract. To examine how well our search terms captured the breadth of the relevant literature, we conducted a forward-backward literature search following Koricheva et al. (2013). We examined literature cited by and literature which cited every study included in our meta-analysis for relevance based on the title alone. We used Google Scholar on December 8, 2019, to find literature that cited the papers included in our analysis. We then determined how many of those papers were already captured by our original search terms. Any papers that were not included in our original screening process were then screened based on the

BOX 3 Common genomic markers

Over the past several decades, genetic variation has been assayed using a variety of molecular markers. Rapid advancements in DNA sequencing technologies now allow researchers to cost-effectively sequence whole or large proportions of genomes. The following markers are frequently used to study genomic differentiation.

Microsatellites (msats)

Msats are short (typically 2–4 nucleotide) sequence repeats (typically 5–50 times) of DNA in noncoding regions that vary in length between individuals and thus are generally regarded as neutrally evolving loci (Schlötterer, 2000). Msats are also referred to as short tandem repeats (STRs) and simple sequence repeats (SSRs). Occasionally msats demonstrate signals of selection, likely due to linkage with coding regions experiencing a selective sweep (i.e., genetic hitchhiking). However, in rare cases, msats may be directly involved in phenotype determination by affecting gene expression (Li et al., 2002) or when they occur within a gene (Li et al., 2004).

Amplified fragment length polymorphisms (AFLPs)

AFLP methods assess the fragment profile of DNA that has been amplified after digestion with restriction enzymes. AFLPs provide biallelic genotypes based on presence/absence scoring. Hundreds of loci can be assayed for relatively little cost, making AFLPs a useful tool for studying patterns of selection across the genome without any knowledge of the genome sequence.

Single nucleotide polymorphisms (SNPs)

A SNP is DNA variation that occurs at a single nucleotide position. While most SNPs have no effect on fitness because either they are in neutral regions of the genome or represent a silent mutation (i.e., one that does not affect amino acid coding), some SNPs in coding regions can be highly influential. Sliding windows can be used to measure the collective F_{ST} of numerous nearby SNPs, thus isolating islands of genomic differentiation, rather than single differentiated SNPs.

Quantitative trait loci (QTLs)

QTLs are genomic regions that are significantly associated with a quantitative, or continuous trait with a polygenic basis. These regions are defined using experimental crosses in a process called QTL mapping (Sen & Churchill, 2001). Regions are genotyped using msats or SNP markers.

Haplotypes/Microhaplotypes

Most SNP genotyping methods involve sequencing contiguous sets of nucleotides that may contain multiple polymorphic sites. A microhaplotype genotyping approach considers all the neighboring SNPs present on a single sequencing read to be representative of single allele/haplotype due to the assumption of extremely low recombination rates over short distances (<300 nucleotides). With this approach, a locus that contains multiple SNPs is then analyzed in a multiallelic framework, much like a microsatellite.

abstract as described above; none of the additionally screened papers contained appropriate data for inclusion in our meta-analysis. In total, we screened 4317 papers, retaining 31 papers for analysis (Figure S1). The most common reason for noninclusion of papers (nearly all) was lack of measured phenotypic data. As these data were generally not measured, rather than not reported, we did not request data from authors.

We extracted the following information from each paper: species, phenotypic trait, P_{ST} of the phenotypic trait, number of individuals used to calculate P_{ST} , number of groups the phenotypes were sampled from, number of groups the genotype data were sampled from, F_{ST} of loci under selection ("non-neutral"; nnF_{ST}), F_{ST} of neutral loci, F_{ST} formula, marker type, number of loci under selection, number of neutral loci, and method(s) used to determine which loci

are under selection. For papers which included raw phenotypic measurements (16) we calculated P_{ST} using:

$$P_{ST} = 1 - \frac{\sum_j \sum_i (x_{ij} - \bar{x}_j)^2}{\sum_j \sum_i (x_{ij} - \bar{x})^2} \quad (1)$$

The numerator is the sum of squared deviations of each individual's phenotype from the population mean, and the denominator is the sum of squared deviations of each individual's phenotype from the metapopulation mean. We used F tests to confirm that all reported and calculated P_{ST} values were statistically different than 0, that is, implied phenotypic differentiation.

For papers with more than two study populations, we used all possible pairwise population comparisons for analysis.

TABLE 1 Papers used in this meta-analysis

| Paper ^a | Species | Method ^b | Phens. ^c | Unique comps. ^d | Mean N ^e | Marker type | Analysis methods ^f | Common garden |
|--------------------------------|--------------------------------|---------------------|---------------------|----------------------------|---------------------|-------------|-------------------------------|---------------|
| Defaveri & Merilä (2013) | <i>Gasterosteus aculeatus</i> | C | 1 | 1 | 34.8 | QTL | R | No |
| Morris et al. (2018) | <i>Gasterosteus aculeatus</i> | C | 4 | 1 | 39.2 | QTL | R | No |
| Paccard et al. (2018) | <i>Gasterosteus aculeatus</i> | C | 6 | 1 | 25.4 | MSAT | R | No |
| Le Corre (2005) | <i>Arabidopsis thaliana</i> | C | 12 | 1 | NA | Haplo. | R | Yes |
| Pedersen et al. (2017) | <i>Gasterosteus aculeatus</i> | C, O | 2 | 1 | 116.8 | SNP | N, R | No |
| Ólafsdóttir & Snorrason (2009) | <i>Gasterosteus aculeatus</i> | C, O | 3 | 1 | 46.0 | MSAT | N, R | No |
| Royer et al. (2016) | <i>Yucca</i> spp | G | 3 | 1 | 103.0 | SNP | G | No |
| Johnston et al. (2014) | <i>Salmo salar</i> | G, O | 1 | 1 | 125.8 | SNP | N, B | No |
| Wei et al. (2017) | <i>Brassica napus</i> | G, O | 4 | 2 | 108.6 | SNP, Window | G | Yes |
| Porth et al. (2015) | <i>Populus trichocarpa</i> | G, O | 113 | 1 | 108.3 | SNP | N, B | No |
| Laporte et al. (2015)* | <i>Coregonus clupeaformis</i> | O | 1 | 5 | 30.0 | SNP | N | No |
| Marques et al. (2017) | <i>Gasterosteus aculeatus</i> | O | 1 | 1 | 70.0 | SNP | N | No |
| Raeymaekers et al. (2007)* | <i>Gasterosteus aculeatus</i> | O | 1 | 6 | 30.0 | QTL | N | No |
| Kovi et al. (2015) | <i>Lolium perenne</i> | O | 1 | 1 | 300.0 | SNP | N | Yes |
| Izuno et al. (2017) | <i>Metrosideros polymorpha</i> | O | 2 | 1 | 8.0 | SNP | B | Yes |
| Smith et al. (2008) | <i>Andropodus virens</i> | O | 3 | 1 | 20.9 | AFLP | N | No |
| Qiu et al. (2017) | <i>Phragmites australis</i> | O | 3 | 1 | 9.0 | AFLP | N, B | Yes |
| He et al. (2019) | <i>Banksia attenuata</i> | O | 4 | 1 | 11.0 | SNP | B | No |
| Sra et al. (2019) | <i>Brassica</i> spp | O | 4 | 3 | 270.3 | SNP | N | Yes |
| Hamlin & Arnold (2015)* | <i>Iris hexagona</i> | O | 5 | 22 | 10.1 | SNP | B | No |
| Nakazato et al. (2012) | <i>Solanum peruvianum</i> | O | 5 | 1 | 10.9 | AFLP | N | Yes |
| Hudson et al. (2013)* | <i>Coregonus</i> spp | O | 6 | 9 | 6.8 | AFLP | N, B | No |
| Kaeuffer et al. (2012)* | <i>Gasterosteus aculeatus</i> | O | 9 | 6 | 40.0 | MSAT | R | No |
| Culling et al. (2013)* | <i>Salmo salar</i> | O | 9 | 10 | 36.8 | SNP | N, B | Yes |
| Sedeek et al. (2014) | <i>Ophrys</i> spp | O | 10 | 1 | 28.4 | SNP | B | No |
| Keller et al. (2011) | <i>Populus balsamifera</i> | O | 10 | 3 | 15.2 | SNP | N | Yes |
| N'Diaye et al. (2018) | <i>Triticum turgidum</i> | O | 11 | 1 | 42.7 | SNP | N, B | Yes |
| Eimanifar et al. (2018) | <i>Apis mellifera</i> | O | 13 | 1 | 154.7 | SNP | N | No |
| Porth et al. (2016) | <i>Quercus</i> spp | O | 13 | 1 | 827.5 | SNP | R | No |
| Flanagan et al. (2016) | <i>Syngnathus scovelli</i> | O | 14 | 1 | 21.2 | SNP | R | No |
| Dillon et al. (2013) | <i>Pinus radiata</i> | O | 39 | 1 | 149 | SNP | N | Yes |

*Indicates papers in within-study analysis.

^bC = candidate gene; G = GWAS; O = outlier.

^cNumber of phenotypes studied.

^dUnique interpopulation comparisons.

^eMean individuals per population.

^fR = raw; N = non-Bayesian; B = Bayesian; G = GWAS-specific.

We organized genomic analysis software into four groups:

1. Raw: raw F_{ST} calculations
2. Bayesian: BayeScan and 2DSFS
3. Non-Bayesian: Arlequin, Lositan, fdist, dfdist, fstat, detsel, InRV
4. GWAS-specific: other software used specifically for GWAS

2.3 | Analyses

We logit-transformed all F_{ST} and P_{ST} data in this study. The logit transformation assumes that small changes at the ends of the range of possibilities (e.g., from $F_{ST} = 0.01$ to 0.02 or 0.98 to 0.99) are more important than small changes in the middle of the range (e.g., from

$F_{ST} = 0.50$ to 0.51). Logit transforming F_{ST} and P_{ST} also standardizes their sensitivity to differentiation—doubling interpopulation variation produces an equal change in $\logit(F_{ST}$ or $P_{ST})$ regardless of the

starting variation (Figure 1). Finally, logit-transforming F_{ST} makes F_{ST} data—which are often right-skewed—roughly normally distributed (Figure S2). All neutral F_{ST} values <0.005 were changed to 0.005 to avoid having zero or negative values, which cannot be handled by a logit transformation.

Due to the large range in number of phenotypes measured in any given paper, we used paper-specific averages for our global P_{ST} - F_{ST} analysis. Specifically, we averaged values for all numeric variables for each unique population-level comparison, allowing multiple points if a comparison was replicated with multiple methods (i.e., different statistical software or a GWAS and outlier approach). This averaging resulted in a final 111 datapoints for 88 unique population-level comparisons across 31 papers.

BOX 4 Genomic methods for detecting divergent selection

Techniques used for detecting selection can broadly be split into complementary approaches that detect either loci of presumably large effect or those that identify patterns of polygenic selection.

Candidate gene approaches

Candidate gene approaches investigate associations between genotypes and phenotypes at specific preordained loci with *a priori* hypothesized functions. Candidate gene approaches are most commonly pursued in species with relatively rich genomic resources. Reproducibility has been challenging for many candidate locus approaches (Tabor et al., 2002).

Outlier approaches

Genetic differentiation (e.g., F_{ST}) outlier tests are most commonly used to identify candidate loci. These tests assume that loci with genetic differentiation values that significantly exceed background (i.e., neutral) differentiation levels are under selection.

Genome-wide association studies (GWAS)

Genome-wide association studies (GWAS) are one suite of methods that identify correlations between genomic data—usually high coverage genomic data—and fitness values, phenotypes, or environmental values (GEAS).

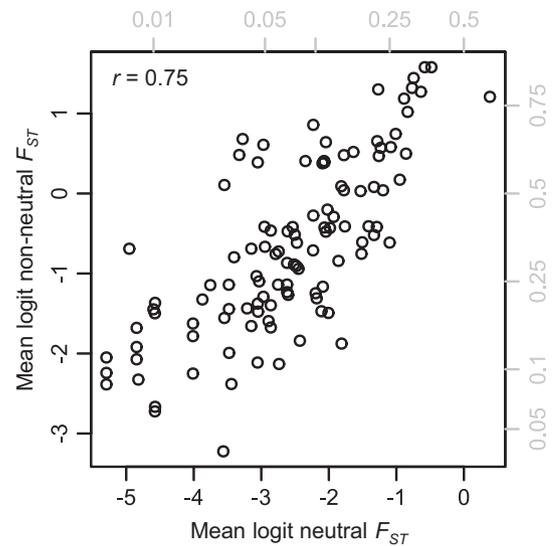


FIGURE 2 Non-neutral and neutral F_{ST} are highly correlated across studies. Gray labels show untransformed F_{ST} values. Each point represents average F_{ST} values for a unique population–population comparison, with multiple points for multiple methods (see text)

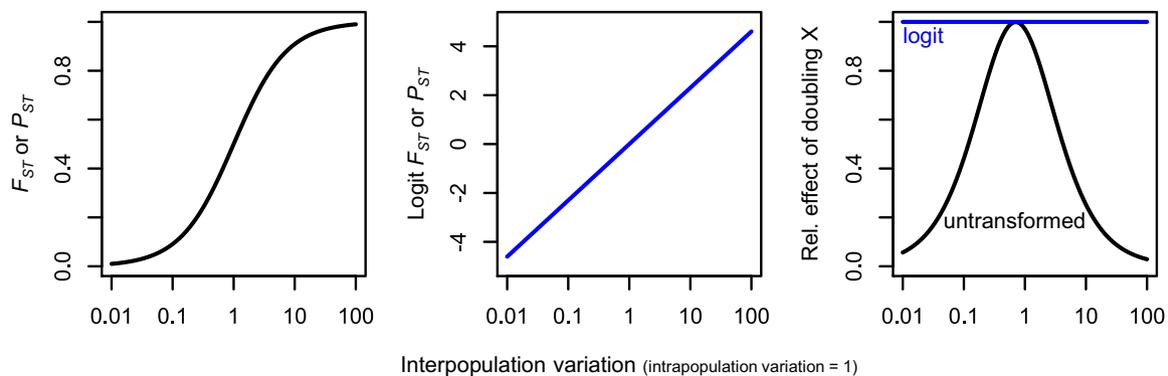


FIGURE 1 Logit transformations are useful for quantifying differentiation. *Left*: F_{ST} and P_{ST} , untransformed, provide limited characterization of differentiation when differentiation is low or high (i.e., interpopulation variation is very small or large relative to intrapopulation variation). *Center*: logit transformations of F_{ST} and P_{ST} , however, provide a log-linear metric of differentiation whose shape is independent of differentiation. *Right*: doubling differentiation (interpopulation variation) has the same effect on $\logit(F_{ST}$ and $P_{ST})$ regardless of starting point, whereas the effect of doubling differentiation on untransformed F_{ST} and P_{ST} depends heavily on starting point

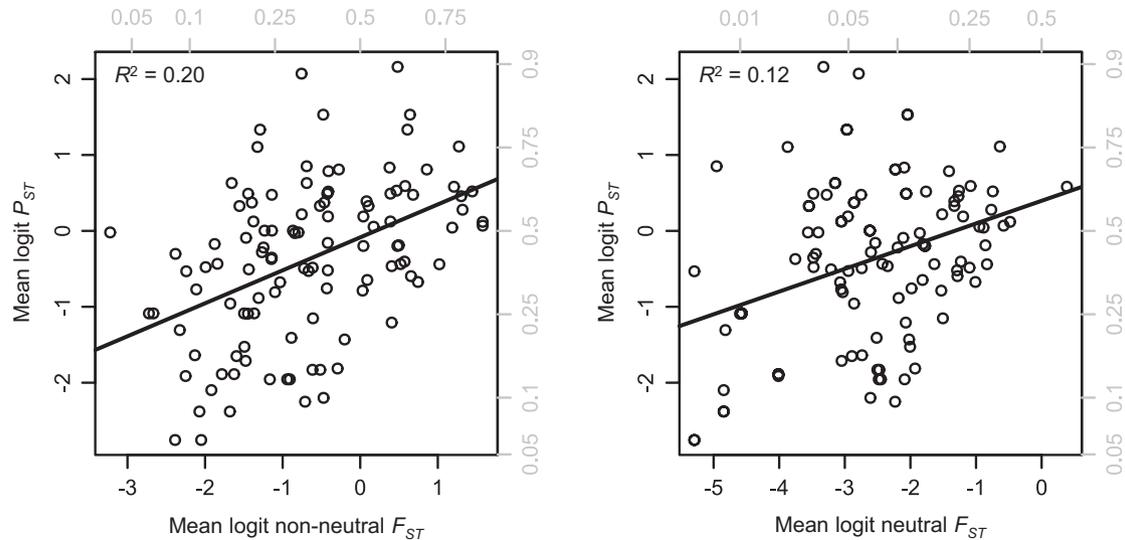


FIGURE 3 Both non-neutral (*left*) and neutral (*right*) F_{ST} predict P_{ST} , but non-neutral F_{ST} is a much stronger predictor of P_{ST} . Gray labels show untransformed F_{ST} and P_{ST} values. Each point represents average P_{ST} and F_{ST} values for a unique population–population comparison, with multiple points for multiple methods (see text)

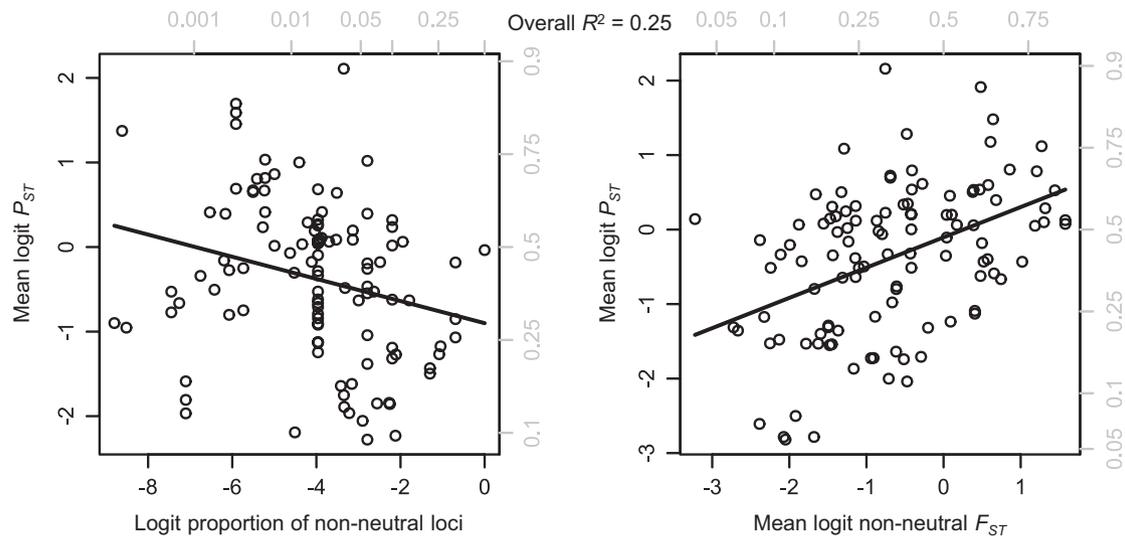


FIGURE 4 *Left*: proportion of non-neutral loci (the ratio of candidate, outlier, or GWAS positive loci to the total number examined) has a negative effect on P_{ST} . Removing this effect allows for a clearer view of the P_{ST} –non-neutral F_{ST} relationship (*right*). There is no significant interaction between proportion of non-neutral loci and non-neutral F_{ST} (Table 2). Gray labels show untransformed F_{ST} , P_{ST} and proportion of non-neutral loci values. Each point represents average P_{ST} , F_{ST} , and proportion of non-neutral loci for a unique population–population comparison, with multiple points for multiple methods (see text)

We then tested these data for relationships between average P_{ST} and average neutral F_{ST} or average non-neutral F_{ST} (nnF_{ST}). As nnF_{ST} and F_{ST} are strongly correlated (Figure 2), including both in the same model is inadvisable; we therefore tested them separately using two general linear models and likelihood ratio tests (Fox & Weisberg, 2011). We compared the models using relative likelihood—as these models were not nested, a likelihood ratio test was not feasible (Burnham & Anderson, 2003).

As nnF_{ST} was superior to neutral F_{ST} in predicting P_{ST} (Figure 3), we investigated the proportion of non-neutral loci as a covariate in the P_{ST} – nnF_{ST} relationship. A study that found one locus with high F_{ST} has different implications for P_{ST} than a similar study finding

hundreds of loci with high F_{ST} . We therefore fit the following general linear model:

$$\text{logit}(P_{ST}) = \beta_0 + \beta_F \text{logit}(nnF_{ST}) + \beta_L \text{logit}\left(\frac{N_{nn}}{N_{total}}\right) + \beta_{FL} \text{logit}(nnF_{ST}) \text{logit}\left(\frac{N_{nn}}{N_{total}}\right) \quad (2)$$

P_{ST} and nnF_{ST} are described above; N_{nn} = number of non-neutral loci; N_{total} = total number of loci examined; and β -terms are coefficients determined during model fitting.

We tested all effects in the above model using type II likelihood ratio tests. As only the first-order effect of proportion non-neutral

loci was significant (see Results), we removed the interaction and refit the model. We also included a first-order effect of proportion of non-neutral loci in all subsequent analyses.

We tested for the effects of four methodological variables—common-garden rearing, genetic marker type, broad methodological approach (candidate gene, outlier, GWAS), and analytical method (i.e., software choice)—on P_{ST} and the P_{ST} - nnF_{ST} slope. We did not test for an effect of nnF_{ST} p -value threshold, as p -value threshold did not have a strong effect on nnF_{ST} (Figure S3). We tested each of the methodological variables separately to avoid overfitting, as there were only 31 papers in our dataset. We fit the following general linear model for each methodological variable:

$$\text{logit}(P_{ST}) = \beta_M + \beta_{FM} \text{logit}(nnF_{ST}) + \beta_L \text{logit}\left(\frac{N_{nn}}{N_{total}}\right) \quad (3)$$

P_{ST} , nnF_{ST} , N_{nn} , and N_{total} are described above; β -terms are coefficients determined during model fitting: β_M indicates a method-specific intercept and β_{FM} indicates a method-specific P_{ST} - nnF_{ST} slope.

We tested all effects in each model using type II likelihood ratio tests.

We also examined P_{ST} - nnF_{ST} trends within studies, with the goal of elucidating a P_{ST} - nnF_{ST} relationship for individual phenotypes across numerous populations. We winnowed our master database down to all paper-phenotype-method combinations that had at least five P_{ST} - nnF_{ST} datapoints (18 paper-phenotype-method combinations total). We then fit the following general linear model across all datapoints from the winnowed database:

$$\text{logit}(P_{ST}) = \beta_Z + \beta_{FZ} \text{logit}(nnF_{ST}) + \beta_L \text{logit}\left(\frac{N_{nn}}{N_{total}}\right) \quad (4)$$

P_{ST} , nnF_{ST} , N_{nn} , and N_{total} are described above; β -terms are coefficients determined during model fitting: β_Z indicates a phenotype-specific intercept and β_{FZ} indicates a phenotype-specific P_{ST} - nnF_{ST} slope—that is, β_Z and β_{FZ} took a unique value for each paper-phenotype-method combination.

We fit one β_L (proportion of non-neutral loci) slope across all paper-phenotype-method combinations, rather than fitting a unique β_L term for each (as we did for β_Z) due to the overall small sample size and the lack of variation in proportion of non-neutral loci for three papers. We tested the slopes of each phenotype-specific P_{ST} - nnF_{ST} relationship (β_{FZ}) using t tests. As the relatively small number of points within each study certainly lowered the power to detect significant P_{ST} - nnF_{ST} relationships, we also examined the distribution of P_{ST} - nnF_{ST} slopes for the within-study analysis.

3 | RESULTS

Non-neutral F_{ST} (nnF_{ST}) was superior to neutral F_{ST} in predicting P_{ST} with higher slope (0.43 vs. 0.30) and R^2 (0.20 vs. 0.12) (Figure 3). Both had statistically significant relationships with P_{ST} (likelihood

TABLE 2 Type II likelihood ratio tests for model predicting P_{ST}

| Variable | χ^2 | df | p |
|---|----------|----|--------|
| logit(non-neutral proportion of loci) | 7.2 | 1 | 0.007 |
| logit(non-neutral F_{ST}) | 24.8 | 1 | <0.001 |
| logit(proportion of loci) \times logit(non-neutral F_{ST}) | 0.9 | 1 | 0.338 |

ratio test—non-neutral F_{ST} : $\chi^2 = 27.3$; df = 1; $p < 0.001$. Likelihood ratio test—neutral F_{ST} : $\chi^2 = 15.0$; df = 1; $p < 0.001$). However, the nnF_{ST} model outperformed the neutral F_{ST} model, with the neutral model having a likelihood of 0.005 with respect to the nnF_{ST} model.

Proportion of non-neutral loci also had a significant negative effect on P_{ST} , but had no significant interaction with nnF_{ST} (Figure 4; Table 2). Including proportion of non-neutral loci in a model with nnF_{ST} (without the nonsignificant interaction) increased the model R^2 from 0.20 to 0.25.

Of our four methodological variables (common-garden rearing, genetic marker type, broad method, and software analysis method), only marker type had a significant effect on P_{ST} (Figure 5; Table 3). P_{ST} was highest (for any given value of non-neutral F_{ST}) for AFLPs, then QTLs, which were closely followed by SNPs and msats. This model result does not, however, imply that some marker types caused higher P_{ST} , rather it indicates that P_{ST} was higher for a given estimate of nnF_{ST} (or more intuitively, nnF_{ST} was estimated as lower for a given value of P_{ST}) for some markers. Marker type did not have a significant interaction with nnF_{ST} .

Within studies, we found no significant relationships between P_{ST} and nnF_{ST} for any phenotypes, even with proportion of non-neutral loci included in the model (Figure 6; Table S2). The mean and standard error for the P_{ST} - nnF_{ST} slope within studies were 0.05 and 0.44, respectively, indicating an average P_{ST} - nnF_{ST} slope close to zero for individual phenotypes within studies (Figure S4).

4 | DISCUSSION

4.1 | Linking genotypic and phenotypic differentiation across and within studies

Here, we show that there is a discernible positive relationship between metrics that measure genomic and phenotypic differentiation when applied to loci and traits putatively under selection (Figure 3). In spite of vast confounding variation within our data set (Table 1)—including species biology, study design, marker type used, statistical approach, and software used—we were able to demonstrate a significant relationship across a diverse range of taxa (i.e., plants, vertebrates, arthropods). This relationship suggests that natural selection acting on the phenome drives evolutionary differentiation on the level of the genome in predictable, somewhat universal ways across clades. The reverse is therefore likely true for the evolutionary processes of drift, mutation, and gene flow,

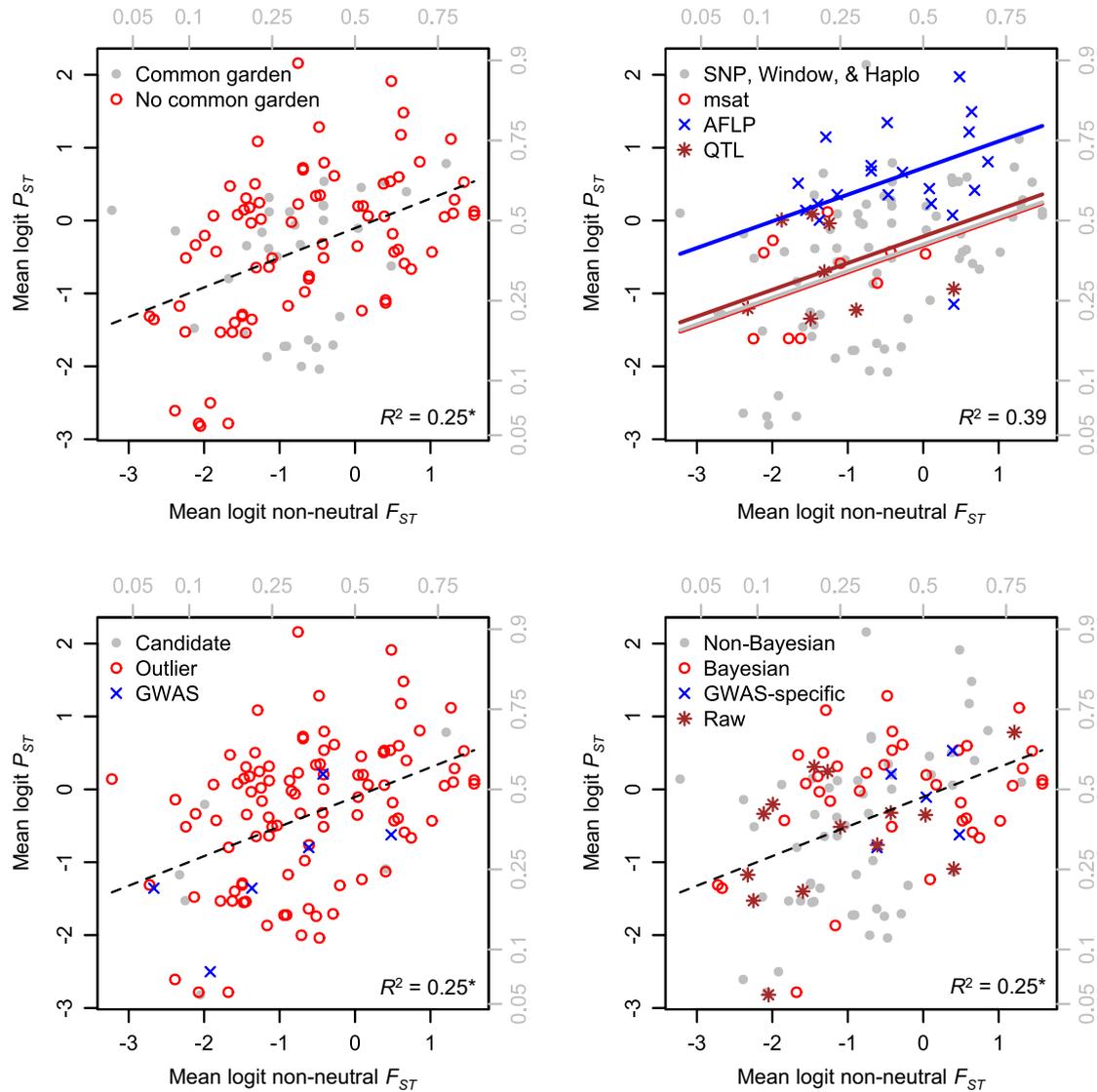


FIGURE 5 Effects of common-garden experimentation (*top left*), marker type (*top right*), broad method (*bottom left*), and analysis method (*bottom right*) on P_{ST} and the P_{ST} -non-neutral F_{ST} slope. Only marker type had a significant effect on P_{ST} , and none of the four variables had a significant effect on the P_{ST} -non-neutral F_{ST} slope (Table 2). Gray labels show untransformed F_{ST} and P_{ST} values. Each point represents average P_{ST} and F_{ST} values for a unique population–population comparison, with multiple points for multiple methods (see text). Variation due to proportion of non-neutral loci is removed in each panel. *Only marker type (*top right*) had a significant effect on P_{ST} ; R^2 values and trendlines for the other three models are from the base (Figure 4) model

in which genomic change may drive phenotypic change. This work expands on similar findings of congruent genomic and phenotypic differentiation within taxa (Brommer, 2011; Kaeuffer et al., 2012; Raeymaekers et al., 2007).

Our major finding—of a positive relationship between P_{ST} and non-neutral F_{ST} (nnF_{ST}) despite the noise of diverse study systems and study designs—has encouraging implications for evolutionary biology. First and foremost, this relationship unsurprisingly supports the genomic basis for phenotypic evolution, suggesting that phenotypic differentiation often has some underlying genomic basis. Of course, this point does not rule out additional environmental contributions like phenotypic plasticity and transgenerational epigenetics. It also suggests that the genomic patterns behind contemporary phenotypic evolution are at least somewhat comparable

among taxa on average, even if their characterization is incomplete. Specifically, despite the litany of confounding factors described below, we still found a significant relationship between P_{ST} and nnF_{ST} , and nnF_{ST} explained a meaningful proportion of the variation in P_{ST} . With further refinement of genomic methods, standardization and reporting of phenotypic data, and clearer details of how study systems differ in terms of genetic architecture (including the strength of individual loci (many weak vs. few strong), linkage, and gene interactions), this relationship should get clearer.

Practically, our results also suggest that reasonably comparing genomic and phenotypic differentiation across taxa should be possible given standardization of methods and reporting. Such comparisons could prove useful in several situations: first, our results give context for what can be considered large or small differentiation by comparing

TABLE 3 Type II likelihood ratio tests for effects of methodological choices on P_{ST} and the P_{ST} - F_{ST} slope

| Variable | χ^2 | df | p |
|---|----------|----|--------|
| Common-garden model | | | |
| Common garden | 1.3 | 1 | 0.260 |
| logit(non-neutral F_{ST}) | 24.6 | 1 | <0.001 |
| Common garden \times logit(non-neutral F_{ST}) | 1.3 | 1 | 0.251 |
| logit(non-neutral proportion of loci) | 6.1 | 1 | 0.014 |
| Marker model | | | |
| Marker | 24.0 | 3 | <0.001 |
| logit(non-neutral F_{ST}) | 22.6 | 1 | <0.001 |
| Marker \times logit(non-neutral F_{ST}) | 2.9 | 3 | 0.406 |
| logit(non-neutral proportion of loci) | 3.0 | 1 | 0.083 |
| Method model | | | |
| Method | 4.1 | 2 | 0.129 |
| logit(non-neutral F_{ST}) | 20.6 | 1 | <0.001 |
| Method \times logit(non-neutral F_{ST}) | 0.3 | 2 | 0.849 |
| logit(non-neutral proportion of loci) | 8.9 | 1 | 0.003 |
| Analytical method model | | | |
| Analytical method | 2.3 | 3 | 0.521 |
| logit(non-neutral F_{ST}) | 19.6 | 1 | <0.001 |
| Analytical method \times logit(non-neutral F_{ST}) | 4.4 | 3 | 0.222 |
| logit(non-neutral proportion of loci) | 1.5 | 1 | 0.224 |

any given study to the distribution of other studies along the shared axis of genomic and phenomic differentiation presented here (i.e., Figure 4, right panel). Second, the relative phenotypic effect size of a particular genetic locus or set of loci can be captured—at least in part—by looking at the relative size of P_{ST} and F_{ST} compared to the values predicted by our model. Finally, as mounting evidence suggests that adaptation in response to selection is generally nonparallel (Bolnick et al., 2018), similar patterns of linked genomic and phenotypic divergence could allow nonparallel adaptation to be compared in terms of degree of differentiation, rather than differentiation in specific traits.

Furthermore, this study contributes to the mounting evidence that contemporary differentiation due to natural selection can provide sufficient, perhaps ideal, phenotypic and genomic variation for linking genomes to phenomes (Evangelou & Ioannidis, 2013). Indeed, the congruence of phenotypic and genomic differentiation across diverse study systems suggests such techniques can be fairly reasonable across taxa. Divergent selection not only produces genomic variation but also generates targeted variation at loci at or near regions that code for responding phenotypes. Thus, differentiation due to natural selection—especially among closely related populations—may generate ideal patterns of genomic variation for genome–phenome association studies by reducing genomic variation at unimportant loci.

4.2 | Confounding factors

While we have demonstrated a relationship between genomic and phenotypic differentiation, much variation remains, implying the

presence of numerous or influential confounding factors. We also acknowledge limitations in our study for identifying these confounding factors; our study only contained data from 31 papers, nor do we have a balanced design covering equal number of papers for each combination of organism, methods, markers, phenotypes, and analysis tool. What follows are our hypotheses for the major sources of variation aside from the genome–phenome mechanism of interest. In general, we believe that variation in the shape of the genomic to phenotypic differentiation relationship comes from three distinct sources: underlying biological, genomic methodological, and phenotypic methodological differences among studies and study systems.

4.2.1 | Confounding biological factors

Inherent to the biology of phenotypes are factors that make a universal genome–phenome relationship challenging to elucidate. These factors likely contribute to the problem of missing heritability—that numerous phenotypic traits with quantifiable heritability have genetic underpinnings that remain elusive (Zuk et al., 2012):

1. *Few strong vs. many weak loci.* Variation in the strength, number, and interaction of loci underlying phenotypes will affect the nature of the P_{ST} - F_{ST} relationship. For example, hundreds of loci likely underlie body size in animals (Kenney-Hunt et al., 2006), with small changes in many loci (which are challenging to detect) cumulatively leading to large changes in body size. On the other hand, some traits—like stickleback lateral

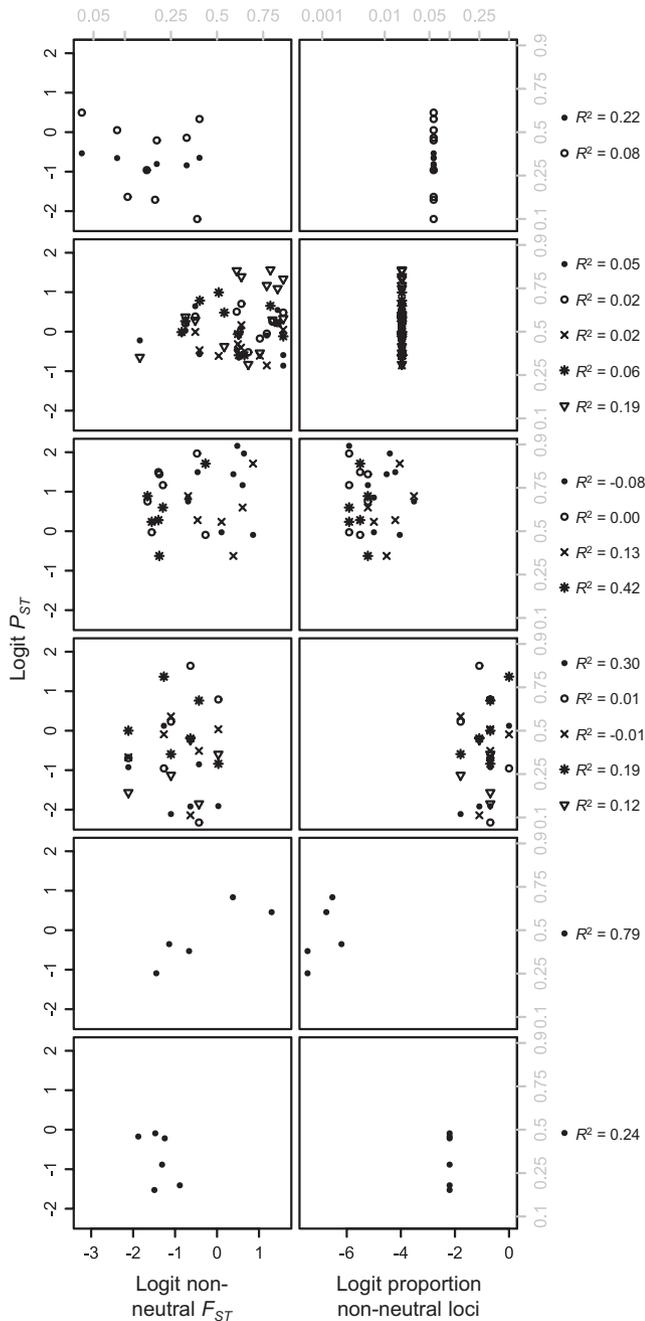


FIGURE 6 Neither non-neutral F_{ST} (left) nor proportion of non-neutral loci (right) predicts P_{ST} consistently well within studies. Each row represents a study; each symbol type represents a phenotype. Data are taken from (top to bottom): Culling et al. (2013), Hamlin & Arnold (2015), Hudson et al. (2013), Kaueffer et al. (2012), Laporte et al. (2015), Raeymaekers et al. (2007). Gray labels show untransformed F_{ST} , P_{ST} , and proportion of non-neutral loci values. R^2 values were calculated based on Equation 4, with a few negative values when study-specific trends in non-neutral F_{ST} vs. proportion of non-neutral loci were the opposite of trends across studies

plate number—can be influenced by a few loci of major effect, which will be much easier to detect (Cresko et al., 2007). We would expect the slope of the P_{ST} - F_{ST} relationship to be much shallower in the first example compared to the second,

even if both phenotypes had an equal additive genetic basis. Furthermore, dominance and epistasis will allow differentiation at some genes to amplify or override differentiation at other genes (Holland, 2007). Developing a P_{ST} - F_{ST} model that is robust to these variations will require high genomic coverage for F_{ST} data (to ensure all differentiated loci are found) and methods elucidating gene interactions (Pecanka et al., 2017; Ritchie & Van Steen, 2018).

2. *Genotype-environment interactions.* Environmental effects (intra- and transgenerational), including phenotypic plasticity, can also muddle the P_{ST} - F_{ST} relationship. Indeed, genotype-environment effects account for a large portion of variation in many phenotypes (Forsman, 2015; Hendry, 2016b). Cogradient plasticity can increase P_{ST} , resulting in an apparently stronger P_{ST} - F_{ST} relationship, while countergradient plasticity can decrease P_{ST} , resulting in an apparently weaker P_{ST} - F_{ST} relationship (Ghalambor et al., 2007). Plasticity that is unrelated to the gradient in question can still weaken the P_{ST} - F_{ST} relationship simply by adding noise to P_{ST} (Brommer, 2011). Common-rearing approaches can not only remove plastic effects, but also muddle genetic differentiation in plastic capacity (also known as gene-by-environment interactions), thus underestimating P_{ST} . These opposing potential consequences of common-rearing approaches may explain why common rearing had no significant effect on P_{ST} or the P_{ST} - nnF_{ST} slope in our study (Figure 5). As with many issues in biology, a solution here is to consider results within the context of the specific study organism and examine genetic differentiation, plasticity, and genetic differentiation in plastic capacity (i.e., gene-by-environment interactions) through reciprocal-transplant or multi-environment common-rearing studies.

4.2.2 | Biases in identifying loci under selection

Different methodological approaches to determining genetic differentiation associated with selection introduce noise to genotype-phenotype relationships between studies. However, as the genome is a relatively concrete feature of an organism, the bias introduced by the choice of genetic and analytical methods can be reduced by systematically identifying appropriate methodology.

1. *Methods used for identifying differentiated loci.* While we found no evidence for favoring any particular method of identifying differentiated loci, we did find a significant *negative* relationship between the proportion of non-neutral loci identified and P_{ST} (Figure 4). If most loci had similar phenotypic effect sizes, we would expect a positive relationship between the proportion of non-neutral loci and P_{ST} , as differentiation that involves more loci of the same phenotypic effect size should generate stronger phenotypic differentiation. We offer two hypotheses as to the observed negative relationship between P_{ST} and the proportion of non-neutral loci. The negative relationship may be linked to the effect size of loci (López-Cortegano & Caballero, 2019).

In this case, studies that detected few loci of large effect would have high P_{ST} values and low proportions of non-neutral loci, while studies detecting many loci of small effects would have higher proportions of non-neutral loci, but likely lower P_{ST} , thus generating a negative relationship between the two. Alternatively, the observed negative relationship may instead be linked to methodology, as studies with more liberal classification of loci as non-neutral (i.e., high false-positive rates) would report a higher proportion of non-neutral loci despite relatively low levels of P_{ST} . If this hypothesis is confirmed, more liberal classification of loci as non-neutral may require down-weighting of non-neutral F_{ST} .

2. **Marker choice.** Our results confirm that marker choice induces significant variation into the nnF_{ST} - P_{ST} relationship. While marker choice did not significantly affect the P_{ST} - nnF_{ST} slope, including an effect of marker choice on P_{ST} raised the model R^2 from 0.25 to 0.39. This apparent effect may be due to the correlation of marker type and genomic coverage, as high genomic coverage (e.g., SNPs) resulted in a lower value of P_{ST} for a given value of nnF_{ST} than low genomic coverage (e.g., AFLPs). This result suggests that historical low-coverage approaches associated with certain marker types may have underestimated nnF_{ST} , leading to higher P_{ST} values for a given value of nnF_{ST} (or more intuitively, lower nnF_{ST} estimates for a given value of P_{ST}).
3. **Estimating F_{ST} .** While our results indicated no particular best F_{ST} estimation method in terms of linking P_{ST} to nnF_{ST} , we note that our observed P_{ST} - nnF_{ST} relationship is almost certainly muddled by noise generated by varying software and software settings used to estimate F_{ST} .

4.2.3 | Bias in measuring phenotype

Unlike the genome, which has an objective, finite definition as a nucleotide sequence, the “phenome” is inherently subjective (Box 1). Though the phenotype is the object of selection and a physical property determined in part by the genome, different phenotypes must be recognized and defined on a case-by-case basis, and it is unlikely that the simple metrics used by researchers to define traits fully capture the more complex integrated phenotypes that are truly under selection. Moreover, investigators may be inconsistent in how they capture traits from study to study. Therefore, we recommend reducing the subjectivity of phenotypic data by standardizing the measurement of traits within taxonomic groups and by capturing a wider array of phenotypes within studies:

1. **Data standardization.** Variable methods for measuring complex phenotypes can make comparisons across studies challenging. For example, features like “body shape” may be quantified several different ways even within a particular clade. As relevant phenotypes are study-dependent, some variability in measurements is to be expected. However, it is crucial that authors report methods detailing their phenotyping protocols

so that phenotypic data that are comparable across studies can be more easily extracted. Even methods for calculating P_{ST} vary, seemingly arbitrarily, from paper to paper, often without reporting of assumed values of some variables included in calculations (such as heritability). Some disciplines, like macroecology and studies of vertebrate museum specimens, utilize simplified, standardized phenotypic measurements for a particular taxon (Schneider et al., 2019). Adopting similar protocols across and within study systems will increase the power of meta-analyses and allow for broader comparisons of phenotypic differentiation. For example, standardized descriptions and databases of mutant phenotypes have been developed for a few model taxa (Bogue et al., 2018; Davis et al., 2012; Smith et al., 2004). We also encourage researchers within certain study systems to explore correlative statistical approaches—like structural equation modeling—to describe the relationship among particular phenotypic measurements and how they might relate to the latent phenotypic trait of interest (e.g., body shape). Large strides have been made in behavioral ecology with these methods to understand what phenotypic traits are being measured by different quantifications of behavior (Dingemanse et al., 2010), and similar approaches should be possible for any complex phenotypic trait.

2. **More extensive phenotypic data collection.** Many phenotypes are highly plastic (Forsman, 2015), and detecting causative loci even for those phenotypes with a strong genetic basis may be difficult with reduced-representation genome-sequencing methods. Therefore, much as increasing the completeness of genome sequencing increases the chance of finding differentiated loci when they are present, expanding phenomic coverage by measuring more biologically relevant phenotypes increases the chance of finding phenotypes with a strong genetic basis (but, also like genomic methods, requires appropriate multiple comparison statistical corrections). Furthermore, expanded coverage of the phenome will provide additional useful information like an estimate of the background neutral phenotypic differentiation and correlations in degree of differentiation among suites of differentiated traits, both of which could increase the power of association studies. Finally, the more traits measured, the more likely that multivariate phenotypes can be identified that come closer to the latent, integrated phenotypes under selection.

4.3 | Data archiving

Standardization of data collection and publication methods is necessary to ensure reproducibility and to allow more broad-scale analyses of the genome-to-phenome association like we have done here. Thanks to increasingly common data sharing requirements by journals and broadly standardized data archiving efforts such as GenBank or Dryad, genotypic data are widely available; however, the relatively modest number of studies analyzed here reflects that only a small proportion of papers adequately publish associated

phenotypic data. Perhaps the best resource is the Database of Genotypes and Phenotypes (dbGaP) that is used by the human genetics community (Tryka et al., 2014). Going forward, it is imperative that authors ensure relevant phenotypic data and metadata are collected and archived with genotypic data at the time of publication.

Some metadata accessibility issues are common to both genetic and phenotypic data. In particular, thorough metadata and scripts on bioinformatic and analytical pipelines—particularly those including phenotypes (i.e., GWAS)—are often not published in sufficient detail. The inclusion of metadata and workflows for bioinformatic and statistical analyses will improve reproducibility and ensure data are accessible for future analyses as technologies evolve (Broman et al., 2017; Sandve et al., 2013).

4.4 | Increasing power to link genomes to phenomes

Jarringly, we found no significant relationships between P_{ST} and nnF_{ST} for individual phenotypes within studies (Figure 6; Table S2). This result, we hypothesize, suggests that loci identified as differentiated are largely not responsible for the differentiation in phenotypes documented in these studies. This lack of trend is unlikely to be a statistical artifact due to small sample size, as the average intrastudy P_{ST} - nnF_{ST} slope was strikingly close to zero (Figure S4). We speculate that a scarcity of population-level replication may constrain our ability to link genomic differentiation to phenotypic differentiation. Using relatively few populations to identify diverging loci may mask important loci (through lack of variation) and lead us to focus on spurious loci (through random variation).

Based on our ability to infer a general P_{ST} - nnF_{ST} relationship across study systems, we suggest several approaches to establishing genome-to-phenome relationships within study systems with greater power. These approaches focus on correlating genomic and phenotypic differentiation across metapopulations and involve:

1. Replicate correlations of P_{ST} and nnF_{ST} across a multitude of populations spanning a differentiation spectrum.
2. Broad coverage of phenotype (to correspond with broad genomic coverage), including sampling of as many biologically relevant and evolutionarily independent phenotypes as feasible.
3. A standardization of traits documented within taxonomic groups.

Correlating P_{ST} and nnF_{ST} across a spectrum of differentiation ensures that genotypes and phenotypes are not only associated, but clearly *differentiate congruently* across landscapes, providing more thorough evidence for the genome-phenome functional link. Having a gradient of differentiation (i.e., numerous nnF_{ST} and P_{ST} values) avoids potentially spurious genome-phenome relationships generated by cherry-picking highly differentiated populations, which may be responsible for the observed weak P_{ST} - nnF_{ST} relationship within studies. Measuring numerous phenotypes increases the likelihood of finding a phenotype that is strongly determined by diverging loci, as long as appropriate statistical corrections are

used to avoid false positives, and also helps us better understand how a given trait diverges in reference to the rest of the phenome. With enough careful measurement, we should be able to describe the relationship between nnF_{ST} and nnP_{ST} , the non-neutral components of phenotype. Currently, our only approach to this is to use our limited understanding of the system to select what we judge as the most differentiated traits. Through careful methodological choices, broader measurement of phenotypes, and a metapopulation approach, genome-to-phenome associations in natural populations can become a more powerful and accessible tool for understanding contemporary evolution.

ACKNOWLEDGEMENTS

The authors thank Robb Brumfield, William Cresko, and Wesley Warren for critical feedback on this project. This project was supported by the USDA National Institute of Food and Agriculture, Hatch Project number ME022207 through the Maine Agricultural & Forest Experiment Station. This article is Maine Agricultural and Forest Experiment Publication Number 3824.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTION

Project design: All authors. Project funding: BJO, BLK, AIK, MTK. Literature search: AKW, ZTW. Data collection: ZTW, AKW, KLB, JDC. Analysis: ZTW. Figure making: ZTW, AKW, KLB, JDC. Lead writing: ZTW, AKW, KLB, JDC, JJH. Editing and revision: All authors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article.

ORCID

Zachary T. Wood  <https://orcid.org/0000-0001-7369-9199>

Kayla L. Barton  <https://orcid.org/0000-0001-6446-8997>

REFERENCES

- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11, 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Barrett, R. D. H., Laurent, S., Mallarino, R., Pfeifer, S. P., Xu, C. C. Y., Foll, M., Wakamatsu, K., Duke-Cohan, J. S., Jensen, J. D., & Hoekstra, H. E. (2019). Linking a mutation to survival in wild mice. *Science*, 363, 499–504. <https://doi.org/10.1126/science.aav3824>
- Bogue, M. A., Grubb, S. C., Walton, D. O., Philip, V. M., Kolishovski, G., Stearns, T., Dunn, M. H., Skelly, D. A., Kadakkuzha, B., TeHennepe, G., Kunde-Ramamoorthy, G., & Chesler, E. J. (2018). Mouse Phenome Database: An integrative database and analysis suite for curated empirical phenotype data from laboratory mice. *Nucleic Acids Research*, 46, D843–D850. <https://doi.org/10.1093/nar/gkx1082>
- Bolger, A. M., Poorter, H., Dumschott, K., Bolger, M. E., Arend, D., Osorio, S., Gundlach, H., Mayer, K. F. X., Lange, M., Scholz, U., & Usadel, B. (2019). Computational aspects underlying genome to phenome analysis in plants. *The Plant Journal*, 97, 182–198. <https://doi.org/10.1111/tbj.14179>

- Bolnick, D. I., Barret, R. D. H., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non)parallel evolution. *Annual Review of Ecology, Evolution, and Systematics*, 49, 303–330. <https://doi.org/10.1146/annurev-ecolsys-110617-062240>
- Broman, K., Cetinkaya-Rundel, M., Nussbaum, A., Paciorek, C., Peng, R., Turek, D., & Wickham, H. (2017). Recommendations to funding agencies for supporting reproducible research. *American Statistical Association*, 2, 1–4.
- Brommer, J. E. (2011). Whither Pst? The approximation of Qst by Pst in evolutionary and conservation biology. *Journal of Evolutionary Biology*, 24, 1160–1168. <https://doi.org/10.1111/j.1420-9101.2011.02268.x>
- Burnett, K. G., Durica, D. S., Mykles, D. L., Stillman, J. H., & Schmidt, C. (2020). Recommendations for advancing genome to phenome research in non-model organisms. *Integrative and Comparative Biology*, 60, 397–401. <https://doi.org/10.1093/icb/icaa059>
- Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer Science & Business Media.
- Burt, C., & Munafò, M. (2021). Has GWAS lost its status as a paragon of open science? *PLoS Biology*, 19, e3001242. <https://doi.org/10.1371/journal.pbio.3001242>
- Carlson, S. M., Cunningham, C. J., & Westley, P. A. H. (2014). Evolutionary rescue in a changing world. *Trends in Ecology & Evolution*, 29, 521–530. <https://doi.org/10.1016/j.tree.2014.06.005>
- Chen, Z., Boehnke, M., Wen, X., & Mukherjee, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. *G3: Genes, Genomes, Genetics*, 11(2), jkaa056. <https://doi.org/10.1093/g3journal/jkaa056>
- Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biology*, 2, e109. <https://doi.org/10.1371/journal.pbio.0020109>
- Cresko, W. A., McGuigan, K. L., Phillips, P. C., & Postlethwait, J. H. (2007). Studies of threespine stickleback developmental evolution: Progress and promise. *Genetica*, 129, 105–126. <https://doi.org/10.1007/s10709-006-0036-z>
- Culling, M. A., Freamo, H., Patterson, K. P., Berg, P. R., Lien, S., & Boulding, E. G. (2013). Signatures of selection on growth, shape, parr marks, and SNPs among seven Canadian Atlantic salmon (*Salmo salar*) populations. *The Open Evolution Journal*, 7, 1–16. <https://doi.org/10.2174/1874404401307010001>
- Cuperus, J. T., & Queitsch, C. (2020). Editorial overview: Technology development as a driver of biological discovery. *Current Opinion in Plant Biology*, 54, A1–A4. <https://doi.org/10.1016/j.pbi.2020.06.001>
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510. <https://doi.org/10.1038/nrg3012>
- Davis, A. P., Wiegers, T. C., Rosenstein, M. C., & Mattingly, C. J. (2012). MEDIC: A practical disease vocabulary used at the comparative toxicogenomics database. *Database*, 2012, bar065.
- Defaveri, J., & Merilä, J. (2013). Evidence for adaptive phenotypic differentiation in Baltic Sea sticklebacks. *Journal of Evolutionary Biology*, 26, 1700–1715. <https://doi.org/10.1111/jeb.12168>
- Dillon, S. K., Nolan, M. F., Matter, P., Gapare, W. J., Bragg, J. G., & Southerton, S. G. (2013). Signatures of adaptation and genetic structure among the mainland populations of *Pinus radiata* (D. Don) inferred from SNP loci. *Tree Genetics and Genomes*, 9, 1447–1463. <https://doi.org/10.1007/s11295-013-0650-8>
- Dingemans, N. J., Dochtermann, N., & Wright, J. (2010). A method for exploring the structure of behavioural syndromes to allow formal comparison within and between data sets. *Animal Behavior*, 79, 439–450. <https://doi.org/10.1016/j.anbehav.2009.11.024>
- Eimanifar, A., Brooks, S. A., Bustamante, T., & Ellis, J. D. (2018). Population genomics and morphometric assignment of western honey bees (*Apis mellifera* L.) in the Republic of South Africa. *BMC Genomics*, 19, 615.
- Evangelou, E., & Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, 14, 379–389. <https://doi.org/10.1038/nrg3472>
- Feder, M. E., & Mitchell-Olds, T. (2003). Evolutionary and ecological functional genomics. *Nature Reviews Genetics*, 4, 649–655. <https://doi.org/10.1038/nrg1128>
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press.
- Flanagan, S. P., Rose, E., & Jones, A. G. (2016). Population genomics reveals multiple drivers of population differentiation in a sex-role-reversed pipefish. *Molecular Ecology*, 25, 5043–5072. <https://doi.org/10.1111/mec.13794>
- Forsman, A. (2015). Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity*, 115, 276–284. <https://doi.org/10.1038/hdy.2014.92>
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. SAGE Publications.
- Freimer, N., & Sabatti, C. (2003). The human phenome project. *Nature Genetics*, 34, 15–21. <https://doi.org/10.1038/ng0503-15>
- Gallant, J. R., & O'Connell, L. A. (2020). Studying convergent evolution to relate genotype to behavioral phenotype. *Journal of Experimental Biology*, 223, 1–11. <https://doi.org/10.1242/jeb.213447>
- Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21, 394–407. <https://doi.org/10.1111/j.1365-2435.2007.01283.x>
- Gibson, G. (2018). Population genetics and GWAS: A primer. *PLoS Biology*, 16, e2005485. <https://doi.org/10.1371/journal.pbio.2005485>
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17, 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Hall, B., Limaye, A., & Kulkarni, A. B. (2009). Overview: Generation of gene knockout mice. *Current Protocols in Cell Biology*, 44(1), 1–17. <https://doi.org/10.1002/0471143030.cb1912s44>
- Hamlin, J. A. P., & Arnold, M. L. (2015). Neutral and selective processes drive population differentiation for *Iris hexagona*. *Journal of Heredity*, 106, 628–636.
- He, T., Lamont, B. B., Enright, N. J., D'Agui, H. M., & Stock, W. (2019). Environmental drivers and genomic architecture of trait differentiation in fire-adapted *Banksia attenuata* ecotypes. *Journal of Integrative Plant Biology*, 61, 417–432.
- Hendry, A. P. (2013). Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity*, 111, 456–466. <https://doi.org/10.1038/hdy.2013.75>
- Hendry, A. P. (2016a). *Eco-evolutionary dynamics*. Princeton University Press.
- Hendry, A. P. (2016b). Key questions on the role of phenotypic plasticity in eco-evolutionary dynamics. *Journal of Heredity*, 107, 25–41. <https://doi.org/10.1093/jhered/esv060>
- Hoffmann, A., Griffin, P., Dillon, S., Catullo, R., Rane, R., Byrne, M., Jordan, R., Oakeshott, J., Weeks, A., Joseph, L., Lockhart, P., Borevitz, J., & Sgrò, C. (2015). A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, 2, 1. <https://doi.org/10.1186/s40665-014-0009-x>
- Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology*, 10, 156–161. <https://doi.org/10.1016/j.pbi.2007.01.003>
- Hudson, A. G., Vonlanthen, P., Bezault, E., & Seehausen, O. (2013). Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish. *BMC Evolutionary Biology*, 13, 108. <https://doi.org/10.1186/1471-2148-13-108>
- Huxley, J. (1942). *Evolution, the modern synthesis*. G. Allen & Unwin Limited.
- Izuno, A., Kitayama, K., Onoda, Y., Tsujii, Y., Hatakeyama, M., Nagano, A. J., Honjo, M. N., Shimizu-Inatsugi, R., Kudoh, H., Shimizu, K. K.,

- & Isagi, Y. (2017). The population genomic signature of environmental association and gene flow in an ecologically divergent tree species *Metrosideros polymorpha* (Myrtaceae). *Molecular Ecology*, *26*, 1515–1532.
- Johnston, S. E., Orell, P., Pritchard, V. L., Kent, M. P., Lien, S., Niemelä, E., Erkinaro, J., & Primmer, C. R. (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, *23*, 3452–3468.
- Kaeuffer, R., Peichel, C., Bolnick, D., & Hendry, A. (2012). Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution*, *66*, 402–418. <https://doi.org/10.1111/j.1558-5646.2011.01440.x>
- Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellåker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., ... Adams, D. J. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, *477*, 289–294. <https://doi.org/10.1038/nature10413>
- Keller, S. R., Soolanayakanahally, R. Y., Guy, R. D., Silim, S. N., Olson, M. S., & Tiffin, P. (2011). Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, *Populus balsamifera* L. (Salicaceae). *American Journal of Botany*, *98*, 99–108.
- Kennedy-Hunt, J. P., Vaughn, T. T., Pletscher, L. S., Peripato, A., Routman, E., Cothran, K., Durand, D., Norgard, E., Perel, C., & Cheverud, J. M. (2006). Quantitative trait loci for body size components in mice. *Mammalian Genome*, *17*, 526–537. <https://doi.org/10.1007/s00335-005-0160-6>
- Kinnison, M. T., & Hairston, N. G. (2007). Eco-evolutionary conservation biology: Contemporary evolution and the dynamics of persistence. *Functional Ecology*, *21*, 444–454. <https://doi.org/10.1111/j.1365-2435.2007.01278.x>
- Koricheva, J., Gurevitch, J., & Mengersen, K. (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton University Press.
- Kovi, M. R., Fjellheim, S., Sandve, S. R., Larsen, A., Rudi, H., Asp, T., Kent, M. P., & Rognli, O. A. (2015). Population structure, genetic variation, and linkage disequilibrium in perennial ryegrass populations divergently selected for freezing tolerance. *Frontiers in Plant Science*, *6*, 929. <https://doi.org/10.3389/fpls.2015.00929>
- Lajeunesse, M. J. (2016). Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for R. *Methods in Ecology and Evolution*, *7*, 323–330.
- Laporte, M., Rogers, S. M., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P.-A., Dalziel, A. C., Chebib, J., & Bernatchez, L. (2015). RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in lake whitefish (*Coregonus clupeaformis*) species pairs. *G3: Genes, Genomes, Genetics*, *5*, 1481–1491.
- Le Corre, V. (2005). Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: Comparison with markers and traits. *Molecular Ecology*, *14*, 4181–4192. <https://doi.org/10.1111/j.1365-294X.2005.02722.x>
- Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology*, *11*, 2453–2465. <https://doi.org/10.1046/j.1365-294X.2002.01643.x>
- Li, Y.-C., Korol, A. B., Fahima, T., & Nevo, E. (2004). Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution*, *21*, 991–1007. <https://doi.org/10.1093/molbev/msh073>
- López-Cortegano, E., & Caballero, A. (2019). Inferring the nature of missing heritability in human traits using data from the GWAS catalog. *Genetics*, *212*, 891–904. <https://doi.org/10.1534/genetics.119.302077>
- Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, *23*, 2178–2192.
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*, 67–77.
- Marques, D. A., Lucek, K., Haesler, M. P., Feller, A. F., Meier, J. I., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Genomic landscape of early ecological speciation initiated by selection on nuptial colour. *Molecular Ecology*, *26*, 7–24.
- Morris, M. R. J., Bowles, E., Allen, B. E., Jamniczky, H. A., & Rogers, S. M. (2018). Contemporary ancestor? Adaptive divergence from standing genetic variation in Pacific marine threespine stickleback. *BMC Evolutionary Biology*, *18*, 113. <https://doi.org/10.1186/s12862-018-1228-8>
- N'Diaye, A., Haile, J. K., Nilsen, K. T., Walkowiak, S., Ruan, Y., Singh, A. K., Clarke, F. R., Clarke, J. M., & Pozniak, C. J. (2018). Haplotype loci under selection in Canadian durum wheat germplasm over 60 years of breeding: Association with grain yield, quality traits, protein loss, and plant height. *Frontiers in Plant Science*, *9*, 1589. <https://doi.org/10.3389/fpls.2018.01589>
- Nakazato, T., Franklin, R. A., Kirk, B. C., & Housworth, E. A. (2012). Population structure, demographic history, and evolutionary patterns of a green-fruited tomato, *Solanum peruvianum* (Solanaceae), revealed by spatial genetics analyses. *American Journal of Botany*, *99*, 1207–1216. <https://doi.org/10.3732/ajb.1100210>
- Ólafsdóttir, G. Á., & Snorrason, S. S. (2009). Parallels, nonparallels, and plasticity in population differentiation of threespine stickleback within a lake. *Biological Journal of the Linnean Society*, *98*, 803–813. <https://doi.org/10.1111/j.1095-8312.2009.01318.x>
- Oti, M., Huynen, M. A., & Brunner, H. G. (2008). Phenome connections. *Trends in Genetics*, *24*, 103–106. <https://doi.org/10.1016/j.tig.2007.12.005>
- Paccard, A., Wasserman, B. A., Hanson, D., Astorg, L., Durston, D., Kurland, S., Apgar, T. M., El-Sabaawi, R. W., Palkovacs, E. P., Hendry, A. P., & Barrett, R. D. H. (2018). Adaptation in temporally variable environments: stickleback armor in periodically breaching bar-built estuaries. *Journal of Evolutionary Biology*, *31*, 735–752. <https://doi.org/10.1111/jeb.13264>
- Pecanka, J., Jonker, M. A., Bochdanovits, Z., & Van Der Vaart, A. W. (2017). A powerful and efficient two-stage method for detecting gene-to-gene interactions in GWAS. *Biostatistics*, *18*, 477–494. <https://doi.org/10.1093/biostatistics/kxw060>
- Pedersen, S. H., Ferchaud, A.-L., Bertelsen, M. S., Bekkevold, D., & Hansen, M. M. (2017). Low genetic and phenotypic divergence in a contact zone between freshwater and marine sticklebacks: Gene flow constrains adaptation. *BMC Evolutionary Biology*, *17*, 130. <https://doi.org/10.1186/s12862-017-0982-3>
- Pigliucci, M., & Muller, G. (2010). *Evolution - The extended synthesis*. MIT Press.
- Porth, I., Garnier-Géré, P., Klápště, J., Scotti-Saintagne, C., El-Kassaby, Y. A., Burg, K., & Kremer, A. (2016). Species-specific alleles at a β -tubulin gene show significant associations with leaf morphological variation within *Quercus petraea* and *Q. robur* populations. *Tree Genetics and Genomes*, *12*, 81. <https://doi.org/10.1007/s11295-016-1041-8>
- Porth, I., Klápště, J., McKown, A. D., La Mantia, J., Guy, R. D., Ingvarsson, P. K., Hamelin, R., Mansfield, S. D., Ehlting, J., Douglas, C. J., & El-Kassaby, Y. A. (2015). Evolutionary quantitative genomics of *Populus trichocarpa*. *PLoS One*, *10*, e0142864. <https://doi.org/10.1371/journal.pone.0142864>
- Qiu, T., Jiang, L., Li, S., & Yang, Y. (2017). Small-scale habitat-specific variation and adaptive divergence of photosynthetic pigments in different alkali soils in reed identified by common garden and genetic tests. *Frontiers in Plant Science*, *7*, 2016. <https://doi.org/10.3389/fpls.2016.02016>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Raeymaekers, J. A. M., Chaturvedi, A., Hablützel, P. I., Verdonck, I., Hellemans, B., Maes, G. E., De Meester, L., & Volckaert, F. A. M.

- (2017). Adaptive and non-adaptive divergence in a common landscape. *Nature Communications*, 8, 267. <https://doi.org/10.1038/s41467-017-00256-6>
- Raeymaekers, J. A. M., Van Houdt, J. K. J., Larmuseau, M. H. D., Geldof, S., & Volckaert, F. A. M. (2007). Divergent selection as revealed by P(ST) and QTL-based F(ST) in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. *Molecular Ecology*, 16, 891–905. <https://doi.org/10.1111/j.1365-294X.2006.03190.x>
- Ritchie, M. D., & Van Steen, K. (2018). The search for gene-gene interactions in genome-wide association studies: Challenges in abundance of methods, practical considerations, and biological interpretation. *Annals of Translational Medicine*, 6, 157. <https://doi.org/10.21037/atm.2018.04.05>
- Rodríguez-Verdugo, A., Buckley, J., & Stapley, J. (2017). The genomic basis of eco-evolutionary dynamics. *Molecular Ecology*, 26, 1456–1464. <https://doi.org/10.1111/mec.14045>
- Royer, A. M., Streisfeld, M. A., & Smith, C. I. (2016). Population genomics of divergence within an obligate pollination mutualism: Selection maintains differences between Joshua tree species. *American Journal of Botany*, 103, 1730–1741. <https://doi.org/10.3732/ajb.1600069>
- Rudman, S. M., Barbour, M. A., Csilléry, K., Gienapp, P., Guillaume, F., Hairston Jr, N. G., Hendry, A. P., Lasky, J. R., Rafajlović, M., Räsänen, K., Schmidt, P. S., Seehausen, O., Therkildsen, N. O., Turcotte, M. M., & Levine, J. M. (2018). What genomic data can reveal about eco-evolutionary dynamics. *Nature Ecology & Evolution*, 2, 9–15. <https://doi.org/10.1038/s41559-017-0385-2>
- Russell, J. J., Theriot, J. A., Sood, P., Marshall, W. F., Landweber, L. F., Fritz-Laylin, L., Polka, J. K., Oliferenko, S., Gerbich, T., Gladfelter, A., Umen, J., Bezanilla, M., Lancaster, M. A., He, S., Gibson, M. C., Goldstein, B., Tanaka, E. M., Hu, C.-K., & Brunet, A. (2017). Non-model model organisms. *BMC Biology*, 15, 55. <https://doi.org/10.1186/s12915-017-0391-5>
- Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Computational Biology*, 9, e1003285. <https://doi.org/10.1371/journal.pcbi.1003285>
- Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109, 365–371. <https://doi.org/10.1007/s004120000089>
- Schneider, F. D., Fichtmueller, D., Gossner, M. M., Güntsch, A., Jochum, M., König-Ries, B., Le Provost, G., Manning, P., Ostrowski, A., Penone, C., & Simons, N. K. (2019). Towards an ecological trait-data standard. *Methods in Ecology and Evolution*, 10, 2006–2019. <https://doi.org/10.1111/2041-210X.13288>
- Sedeek, K. E. M., Scopece, G., Staedler, Y. M., Schönenberger, J., Cozzolino, S., Schiestl, F. P., & Schlüter, P. M. (2014). Genic rather than genome-wide differences between sexually deceptive Ophrys orchids with different pollinators. *Molecular Ecology*, 23, 6192–6205.
- Sen, S., & Churchill, G. A. (2001). A statistical framework for quantitative trait mapping. *Genetics*, 159, 371–387. <https://doi.org/10.1093/genetics/159.1.371>
- Smith, C. L., Goldsmith, C.-A.-W., & Eppig, J. T. (2004). The mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biology*, 6, R7.
- Smith, T. B., Milá, B., Grether, G. F., Slabbekoorn, H., Sepil, I., Buermann, W., Saatchi, S., & Pollinger, J. P. (2008). Evolutionary consequences of human disturbance in a rainforest bird species from Central Africa. *Molecular Ecology*, 17, 58–71. <https://doi.org/10.1111/j.1365-294X.2007.03478.x>
- Sra, S. K., Sharma, M., Kaur, G., Sharma, S., Akhtar, J., Sharma, A., & Banga, S. S. (2019). Evolutionary aspects of direct or indirect selection for seed size and seed metabolites in *Brassica juncea* and diploid progenitor species. *Molecular Biology Reports*, 46, 1227–1238. <https://doi.org/10.1007/s11033-019-04591-3>
- Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nature Reviews Genetics*, 3, 391–397. <https://doi.org/10.1038/nrg796>
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 20, 467–484. <https://doi.org/10.1038/s41576-019-0127-1>
- Tryka, K. A., Hao, L., Sturcke, A., Jin, Y., Wang, Z. Y., Ziyabari, L., Lee, M., Popova, N., Sharopova, N., Kimura, M., & Feolo, M. (2014). NCBI's database of genotypes and phenotypes: dbGaP. *Nucleic Acids Research*, 42, D975–D979. <https://doi.org/10.1093/nar/gkt1211>
- Uricchio, L. H. (2020). Evolutionary perspectives on polygenic selection, missing heritability, and GWAS. *Human Genetics*, 139, 5–21. <https://doi.org/10.1007/s00439-019-02040-6>
- Via, S., & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, 39, 505–522. <https://doi.org/10.1111/j.1558-5646.1985.tb00391.x>
- Visscher, P. M., & Goddard, M. E. (2019). From R.A. Fisher's 1918 paper to GWAS a century later. *Genetics*, 211, 1125–1130. <https://doi.org/10.1534/genetics.118.301594>
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *American Journal of Human Genetics*, 101, 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*. Oxford University Press.
- Wei, D., Cui, Y., He, Y., Xiong, Q., Qian, L., Tong, C., Lu, G., Ding, Y., Li, J., Jung, C., & Qian, W. (2017). A genome-wide survey with different rapeseed ecotypes uncovers footprints of domestication and breeding. *Journal of Experimental Botany*, 68, 4791–4801. <https://doi.org/10.1093/jxb/erx311>
- Whibley, A., Kelley, J. L., & Narum, S. R. (2021). The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. *Molecular Ecology Resources*, 21, 641–652. <https://doi.org/10.1111/1755-0998.13312>
- Whitlock, M. C. (2008). Evolutionary inference from QST. *Molecular Ecology*, 17, 1885–1896.
- Wood, Z. T., Palkovacs, E. P., Olsen, B. J., & Kinnison, M. T. (2021). The importance of eco-evolutionary potential in the Anthropocene. *BioScience*, biab010. <https://doi.org/10.1093/biosci/biab010>
- Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, 15, 323–354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>
- Young, A. I. (2019). Solving the missing heritability problem. *PLoS Genetics*, 15, e1008222. <https://doi.org/10.1371/journal.pgen.1008222>
- Zhang, J. (2018). Neutral theory and phenotypic evolution. *Molecular Biology and Evolution*, 35, 1327–1331. <https://doi.org/10.1093/molbev/msy065>
- Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, 109, 1193–1198. <https://doi.org/10.1073/pnas.1119675109>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wood, Z. T., Wiegardt, A. K., Barton, K. L., Clark, J. D., Homola, J. J., Olsen, B. J., King, B. L., Kovach, A. I., & Kinnison, M. T. (2021). Meta-analysis: Congruence of genomic and phenotypic differentiation across diverse natural study systems. *Evolutionary Applications*, 14, 2189–2205. <https://doi.org/10.1111/eva.13264>