

RESEARCH ARTICLE

A possible genomic footprint of polygenic adaptation on population divergence in seed beetles?

Göran Arnqvist¹  | Ahmed Sayadi^{1,2}

¹Animal Ecology, Department of Ecology and Genetics, EBC, Uppsala University, Uppsala, Sweden

²Rheumatology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

Correspondence

Göran Arnqvist, Animal Ecology, Department of Ecology and Genetics, EBC, Uppsala University, Uppsala, Sweden.

Email: goran.arnqvist@ebc.uu.se

Funding information

H2020 European Research Council, Grant/Award Number: AdG-294333; Svenska Forskningsrådet Formas, Grant/Award Number: 2018-00705 and 2018-05973; Vetenskapsrådet, Grant/Award Number: 621-2014-4523 and 2019-03611

Abstract

Efforts to unravel the genomic basis of incipient speciation are hampered by a mismatch between our toolkit and our understanding of the ecology and genetics of adaptation. While the former is focused on detecting selective sweeps involving few independently acting or linked speciation genes, the latter states that divergence typically occurs in polygenic traits under stabilizing selection. Here, we ask whether a role of stabilizing selection on polygenic traits in population divergence may be unveiled by using a phenotypically informed integrative approach, based on genome-wide variation segregating in divergent populations. We compare three divergent populations of seed beetles (*Callosobruchus maculatus*) where previous work has demonstrated a prominent role for stabilizing selection on, and population divergence in, key life history traits that reflect rate-dependent metabolic processes. We derive and assess predictions regarding the expected pattern of covariation between genetic variation segregating within populations and genetic differentiation between populations. Population differentiation was considerable (mean $F_{ST} = 0.23-0.26$) and was primarily built by genes showing high selective constraints and an imbalance in inferred selection in different populations (positive Tajima's D_{NS} in one and negative in one), and this set of genes was enriched with genes with a metabolic function. Repeatability of relative population differentiation was low at the level of individual genes but higher at the level of broad functional classes, again spotlighting metabolic genes. Absolute differentiation (d_{XY}) showed a very different general pattern at this scale of divergence, more consistent with an important role for genetic drift. Although our exploration is consistent with stabilizing selection on polygenic metabolic phenotypes as an important engine of genome-wide relative population divergence and incipient speciation in our study system, we note that it is exceedingly difficult to firmly exclude other scenarios.

KEYWORDS

accessory gland proteins, *Callosobruchus maculatus*, epistasis, incipient speciation, metabolic rate, seminal fluid proteins

TAXONOMY CLASSIFICATION

Biodiversity ecology, Ecological genetics, Evolutionary ecology, Genomics, Life history ecology

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

What we can measure is by definition uninteresting and what we are interested in is by definition unmeasurable.

Lewontin (1974)

1 | INTRODUCTION

Some 160 years after Darwin (1859) suggested that natural selection plays a key role in the origin of new species, and we now have a wealth of different types of evidence for Darwin's general tenet (Coyne & Orr, 2004; Schluter, 2000). Yet, the relative importance of different types of phenotypic selection (e.g., natural vs. sexual selection) remains to be determined, and we need to improve our general understanding of the nature of the molecular genetic basis of speciation and of the processes within populations that generate genomic divergence (Wolf et al., 2010). In this regard, one strategy is to study divergence between conspecific populations: studying selection in partially reproductively isolated forms can yield insights into selective processes before they become confounded by the accumulation of fixed differences between species (Burri, 2017; Via, 2009).

The large amounts of genome-wide molecular genetic data generated by the last few decades of next-generation sequencing efforts have made it possible to pursue "phenotype-free" bottom-up approaches in an effort to increase our understanding of incipient speciation. A common strategy in speciation genomics is to identify genes or regions of the genome that show unusually high differentiation between, for example, divergent populations. The physical location of detected outlier loci or their functional annotation is then used to make inferences about the processes responsible for divergence (Cruickshank & Hahn, 2014; Nosil & Feder, 2012). The more general supposition of speciation genomics is that incorporating analyses of genetic variation segregating within populations can provide information about the selective processes that generate divergence between populations (Wolf & Ellegren, 2017).

Traditional methods based on outlier detection strategies can sometimes successfully detect the signature of selection if divergence is dominated by a limited set of major effect loci with independent effects and evolution proceeds by selective sweeps, whereby favorable variants are driven to fixation by divergent positive selection. However, issues are considerably more complicated when selection acts on ecologically relevant quantitative traits with a polygenic basis (i.e., complex traits). This is particularly true when net phenotypic selection (integrating over all life history stages, sexes, etc) is stabilizing and populations are adapting to different optima. Here, we expect populations to diverge through relatively minor shifts in allele frequencies across a very large number of loci showing epistasis for fitness (Gavrilets, 2004; Höllinger et al., 2019) and such shifts are generally undetectable using outlier detection methods (Chevin, 2019; Csilléry et al., 2018; Hayward & Sella, 2022). In fact, because of redundancy, where a given optimal phenotype can be produced by an almost infinite number of allelic combinations across a large number of loci, even stabilizing selection which

is uniform between populations is expected to generate suites of compensatory allelic changes across epistatically interacting loci. Such quasi-neutral polygenic drift is expected to result in population divergence, hybrid inferiority, and, eventually, reproductive isolation (Fierst & Hansen, 2010; Gavrilets, 2004; Figure 1).

Our empirical toolkit and theoretical understanding of speciation genomics is still to a large extent based on scenarios of selective sweeps involving one or a few independently acting or linked speciation genes (Burri, 2017; Cruickshank & Hahn, 2014; Nosil & Feder, 2012; Seehausen et al., 2014; Wolf & Ellegren, 2017). This state is unfortunate, given that this scenario is likely quite rare (Korte & Farlow, 2013; Rockman, 2012; Wellenreuther & Hansson, 2016). We often expect (i) net selection on integrative quantitative traits such as life history traits to be stabilizing with intermittent shifts of the optimum and (ii) variation in such traits to have a highly polygenic basis, therefore inevitably showing marked epistasis for

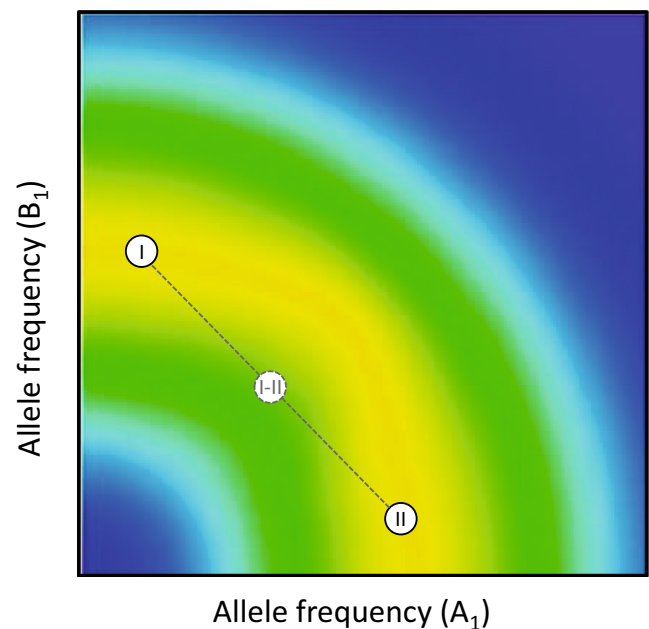


FIGURE 1 For polygenic traits under stabilizing selection, epistatic interactions between loci are inevitable (Barton, 1986, 1989) because the fitness effect of an allele frequency shift in one locus will depend on the allelic state in many other loci (i.e., the genetic background). Because of complex multiplicative allelic effects on fitness across genes, fitness functions across loci tend to be curved (Fierst & Hansen, 2010; Rice, 1998). Here, there is a virtually infinite number of different combinations of allele frequencies across a large number of interacting loci that will all produce the same intermediate optimal phenotype. Populations may thus evolve along multidimensional "ridges" across holey adaptive landscapes (Gavrilets, 2004). While population fitness may remain unaltered as populations move along such non-linear ridges, hybrid populations will tend to fall in "holes" outside the maximum fitness ridges and show depressed fitness. This is illustrated in this simple 3D contour plot of allele frequencies in two loci where two populations (I and II) that occupy different locations along a curved maximum fitness ridge produce lower fitness hybrids (yellow denotes high and blue low fitness) (after Fierst & Hansen, 2010).

fitness (Barghi et al., 2020; Barton, 1986, 1989; Boyle et al., 2017; Brodie, 2000; Fisher, 1930; Nosil et al., 2020; Whitlock et al., 1995). Moreover, the predicted role of selection in divergence critically depends upon genetic architecture (Chevin, 2019). For example, loci under balancing selection are expected to contribute little to population divergence (Via, 2009) and rather tend to show shared polymorphism and low divergence (Fijarczyk & Babik, 2015) when assumed to act independently. In contrast, polygenic scenarios predict that a very large number of epistatically interacting loci encoding traits under stabilizing selection will, collectively, diverge either through selection and adaptation to shifting optima (e.g., Barton, 1989; Höllinger et al., 2019) and/or through quasi-neutral polygenic drift (e.g., Fierst & Hansen, 2010; Gavrilets, 2004; Figure 1).

Detecting the signal of polygenic adaptation in genomic data, however, remains a major challenge (Barghi et al., 2020; Csilléry et al., 2018; Hayward & Sella, 2022). Here, our primary aim is to explore whether a role of stabilizing selection in polygenic population divergence can be unveiled by using an integrative and phenotypically informed approach, focused on analyses of genome-wide genetic variation segregating in divergent populations. We use a well-known insect model system where previous work has demonstrated a prominent role for stabilizing selection on key life history traits (e.g., Berger et al., 2016). We first consult multilocus theory of adaptation and divergence to ask what patterns we might expect if stabilizing selection on polygenic traits indeed plays an important role in population divergence in this system. We then assess whether these patterns are upheld in divergent populations. Importantly, a large body of previous experimental work in this system has generated insights into phenotypic selection and these insights allow us to form phenotypically informed predictions. We focus on broad-scale genome-wide patterns of covariation across genes between standard indices of selection within populations (reflecting allele frequency spectra) and divergence between populations, rather than focusing on identifying single genes or regions by means of outlier detection strategies.

We base our analyses on population resequencing data from three well-studied populations of the seed beetle *Callosobruchus maculatus* (Bruchinae). These populations differ in geographic origin (see Section 2) and are known to have diverged slightly but significantly in a series of life history phenotypes such as metabolic phenotypes (metabolic rate and respiratory quotient) (Immonen et al., 2016), body size (Rankin & Arnqvist, 2008), juvenile growth rate (Arnqvist & Tuda, 2010), egg-adult development time (Dowling et al., 2010), rate of reproductive senescence (Immonen et al., 2016), ejaculate weight (Immonen et al., 2016) and composition (Goenaga et al., 2015), investment in immune function (Dougherty et al., 2017), and environment-specific reproductive fitness (Arnqvist & Tuda, 2010; Hotzy & Arnqvist, 2009). Reproductive incompatibility between populations is also significant but slight: egg-to-adult survival is depressed from some 90% within populations to approximately 80% in between-population crosses (Arnqvist & Tuda, 2010; Fricke & Arnqvist, 2004; Rankin & Arnqvist, 2008). Previous studies show that net stabilizing selection on life history traits is

pronounced in populations of this species (Berger et al., 2014, 2016). This derives in part from fundamental differences in male and female optima: briefly, males show a higher optimal pace-of-life than do females (Arnqvist et al., 2017; Berger et al., 2014, 2016). This is manifested as pronounced sexual dimorphism of rate-dependent processes such as organismal metabolic rate (Arnqvist et al., 2017; Berger et al., 2014), accumulation of lipids during larval development (Lazarević et al., 2012), growth rate and development time (Arnqvist & Tuda, 2010), adult activity (Berger et al., 2014), and adult life span (Fox et al., 2003, 2004) and have resulted in persistent sexually antagonistic selection (Arnqvist & Tuda, 2010; Berg & Maklakov, 2012; Berger et al., 2014, 2016). Other studies have provided support for widespread balancing selection. For example, previous molecular genetic studies suggest an important role for balancing selection in maintaining polymorphism in life history genes in this species (Sayadi et al., 2019), and experimental work has identified sexually antagonistic pleiotropy as an important source of balancing selection (Berger et al., 2016; Grieshop & Arnqvist, 2018). In line with this, Immonen et al. (2017) showed that genes with sex-biased expression are enriched with those involved in metabolic processes and genes showing alternative splicing with sex-biased expression of isoforms also show enrichment of metabolic genes.

1.1 | What might we expect?

Given the above, we derive eight predictions.

I First, any polygenic trait under stabilizing selection will inevitably be affected by a large number of epistatically interacting loci. If, at one of these loci, a particular allele reaches high frequency in a given population then this will affect selection at other loci within the set (Figure 1). Shifts in allele frequencies across loci are thus predicted to be concerted, such that shifts in some loci will be associated with compensatory shifts in other loci (Hayward & Sella, 2022), and divergence through this process is especially likely under weak epistasis (Fierst & Hansen, 2010). The particular loci showing such shifts, whether caused by selection (Barton, 1999), founder events (Arnqvist et al., 2014) or quasi-neutral polygenic drift (Fierst & Hansen, 2010), are unlikely to be shared across diverging populations as stochastic events will have major effects on precisely which loci that show shifts. Reproductive isolation thus evolves because populations slowly evolve to new and unique combinations of allele frequencies (Barton, 1989), resulting in the accumulation of Bateson–Dobzhansky–Muller (BDM) incompatibilities across many interacting loci (Fierst & Hansen, 2010; Gavrilets, 2004; Figure 1). What pattern might this generate within and across diverging populations? Under a simple additive genetic architecture, long-term stabilizing selection on traits tends to erode underlying polygenic variation (Wright, 1935). However, several plausible factors are capable of promoting polygenic variation, separately or in combination, by effectively generating balancing selection

in loci encoding such traits. These include G×E interactions (Gillespie & Turelli, 1989; Turelli & Barton, 2004), antagonistic pleiotropy (Connallon & Clark, 2014; Patten et al., 2010; Turelli & Barton, 2004), overdominance (Turelli & Barton, 2004), dominance reversal (Connallon & Chenoweth, 2019; Grieshop & Arnqvist, 2018), epistasis (Arnqvist et al., 2014), and frequency-dependent selection (Bürger, 2005; Mani et al., 1990; Turelli & Barton, 2004). This is in accord with genetic studies suggesting that variants segregating at intermediate frequency contribute significantly to polygenic phenotypic variation (e.g., Chakraborty et al., 2018; Charlesworth, 2015; Charlesworth et al., 2007; Charlesworth & Hughes, 2000; Mackay, 2010). Given the above, marked frequency shifts in a given locus, resulting in an excess of rare variants, is then expected to be compensated by relatively minor frequency shifts across many interacting loci. Importantly, because of redundancy, such shifts will very rarely be shared across populations. We suggest that, across the genome, genes that contribute to differentiation through this process should therefore generally tend to show signs of positive selection in one population (i.e., an excess of rare alleles) but belong to a class of genes showing signs of balancing selection in the other population (i.e., more even distributions of allele frequencies). We note that this expected apparent imbalance in selection contrasts with classic predictions for differentiation through divergent positive selection, where genes that build differentiation are expected to show signs of positive selection (for alternative alleles) in both populations.

- II Previous work in seed beetles suggests that balancing selection is widespread among genes that are involved in metabolic pathways and thus dictate metabolic rate and sex-specific pace-of-life phenotypes (see above). Since metabolic pathways involve products of many loci showing direct and indirect epistasis and because there should generally be stabilizing selection for intermediate fluxes, it has been predicted that balancing selection on metabolic genes should be common (Whitlock et al., 1995). We therefore predict that genes that show the predicted footprint of polygenic divergence, that is show an apparent imbalance in selection and simultaneously contribute to population differentiation, should be enriched with genes that are generally involved in metabolic pathways.
- III More specifically, given that organismal metabolic rate is a key phenotype under stabilizing selection in our model system (see above), we predict that genes directly involved in regulating metabolic processes should show clear signs of the pattern described under the first prediction above.
- IV Absolute population differentiation (i.e., d_{XY}) is relatively insensitive to differentiation due to allele frequency shifts, as it ignores diversity within populations by essentially asking how different the most common alleles are in two populations. We therefore expect to see the pattern described under the first prediction above for relative (i.e., F_{ST}) but not necessarily for absolute (i.e., d_{XY}) divergence. Absolute divergence metrics are affected by a multitude of factors and processes and are difficult to interpret in isolation. Many scenarios based on divergent purifying selection predict that loci participating in population divergence should show both high relative and absolute divergence (e.g., Cruickshank & Hahn, 2014). If balancing selection is prevalent, however, this will favor the maintenance of ancestral polymorphism in diverging populations while neutral ancestral polymorphisms are more likely to drift to fixation in diverging populations (Guerrero & Hahn, 2017; Wang et al., 2019). The latter effect would tend to result in loci with less constraint showing relatively high absolute divergence.
- V The considerations above suggest that the genetic basis for variation in any given polygenic life history trait should to some extent be distinct across populations, because of a substantial turnover in the genetic basis of traits during adaptation or quasi-neutral polygenic drift. Some loci should thus contribute more to trait variation in some populations and less in others, a point that reverberates the common observation of limited reproducibility in genotype–phenotype association studies of complex traits across populations of the same species (e.g., Chanock et al., 2007; Schielzeth et al., 2018). It has even been suggested that variation across loci and non-parallelism between replicated populations should be key hallmarks of polygenic adaptation (Barghi et al., 2020). We would thus predict that the degree of repeatability in relative population differentiation should be quite modest (Griffin et al., 2017), because differentiation should be built by partly different sets of loci across different sets of population pairs and the vector describing relative differentiation across genes should thus only partly be shared (Pritchard et al., 2010). Needless to say, however, this prediction is shared with other scenarios. Most importantly, divergence dominated by stochastic processes (e.g., genetic drift) can accommodate much of the same pattern.
- VI While we thus predict repeatability of relative population divergence to be modest at the gene level, due to redundancy, we predict that repeatability of population divergence in terms of functional classes or networks of genes should be considerable (Blount et al., 2018).
- VII As stated above, for polygenic traits under stabilizing selection, drift or adaptation is expected to generate divergence through shifts in quasi-neutral allele frequencies across a large number of segregating loci (Barton, 1999; Fierst & Hansen, 2010). To further help distinguish this scenario from divergence through drift across many independent neutral loci, one might predict that genome-wide relative divergence is better explained by gene-specific metrics that reflect the site frequency spectrum (e.g., Tajima's D) than those that estimate selective neutrality by comparing segregating non-synonymous to synonymous variation (e.g., $\pi N/\pi S$; Kryazhimskiy & Plotkin, 2008).
- VIII The probability that the frequency of a non-synonymous variant in a finite population will ultimately reach unity is a positive function of the strength of selection (Kimura, 1964). For this reason, we often expect variants in large effect loci to show higher fixation probabilities. This dynamic is more complex for

loci encoding polygenic traits under stabilizing selection that experience peak shifts. For example, Hayward and Sella (2022) showed that in many cases alleles with moderate effect sizes are instead predicted to show the greatest contributions to polygenic adaptation and that the fixation probability of such alleles is relatively high. This predicts that the relationship between fixation probability and effect size may be distinct in sets of epistatically interacting genes contributing to traits under stabilizing selection, compared with other sets of genes.

2 | MATERIALS AND METHODS

2.1 | Populations

The populations of the seed beetle *C. maculatus* used here were originally collected in three different geographic regions; California (1980), Yemen (1976), and Brazil (2000). They were all founded by a large number of individuals ($>>100$) and have since been maintained in the laboratory on their natural host *V. unguiculata* at population sizes of >500 individuals and at 29°C, 60%–70% RH and a 12L:12D light cycle. Demographic modeling of genetic variation in natural populations has shown that effective population size is very large in natural populations *C. maculatus* (Kébé et al., 2017), consistent with the fact that it is a cosmopolitan and very abundant crop pest, but inbreeding in laboratory populations is always a potential concern. Four lines of evidence suggest that this is likely not the case in these particular populations. First, quantitative genetic studies suggest that laboratory populations of seed beetles such as these do not suffer from inbreeding depression but harbor levels of genetic variation typical of that of natural populations (Fox et al., 2007). Second, molecular genetic estimates of effective population size in similar laboratory seed beetle populations suggest that effective population size is fairly large ($N_e > 1000$ individuals; Gompert & Messina, 2016). Third, and perhaps most importantly, the mean within-population genome-wide SNP density in coding regions across all genes is approximately 6 SNPs/kb in the studied populations and the level of nucleotide diversity for synonymous sites in all three populations is $\pi_S \approx 0.022$ (Sayadi et al., 2019). This is comparable to, or even larger than, that found in natural population of seed beetles (Tine et al., 2013), as well as in many other insects (e.g., Chen et al., 2017; Gossman et al., 2012), and is inconsistent with significant effects of inbreeding (Charlesworth, 2003).

In studies of divergence between laboratory populations, admixture in the laboratory is a common concern for species like *Drosophila melanogaster* where adults are strong fliers. However, adult *C. maculatus* are poor fliers and very rarely take to the wings in the laboratory and larvae are contained to a single host bean throughout their development. These factors preclude significant admixture, which is consistent with a large number of private SNPs observed in these populations (Sayadi et al., 2019) and the high degree of differentiation seen. Further, effects of recent admixture on the signal of

differentiation would tend to be localized to certain regions of the genome, which we did not observe here (see below).

Callosobruchus maculatus was brought to Asia from West Africa by human farmers of its main host about 2000 years ago and were introduced to the Americas from West Africa by Spanish settlers in the early 1700s (Kébé et al., 2017). This suggests that Yemen originally diverged from ancestral African populations some 20,000 generations ago while California and Yemen diverged some 3000 generations ago. Because the climate differs substantially in the three different regions, the source populations are expected to show local adaptation to differences in temperature and humidity and theory suggests that these divergence times are likely deep enough for the populations to have achieved the peak shifts and have reached an equilibrium phase (Hayward & Sella, 2022). However, the degree of adaptive genetic differentiation has most likely been restricted by gene flow resulting from commercial trade with its host (Kébé et al., 2017). Genetic differentiation between these laboratory populations should thus certainly to some extent reflect differences in life history optima in their parental populations, but also the effects of gene flow, genetic drift, quasi-neutral polygenic drift, and adaptation to common garden conditions in the laboratory. While our data thus does not allow us to estimate the precise extent to which divergence has evolved in the wild and in the laboratory, it is still useful for testing the utility of our approach for detecting a role of peak shifts in polygenic population divergence.

2.2 | Sequencing and analyses

All sequencing data derives from Sayadi et al. (2019) and we refer in full to that study for details on read data and data availability. Briefly, to assess variation within and divergence between populations, we sequenced pools of individuals from the three populations ($n = 200$ males per population), to an average depth of $\sim 125\times$ per population using the Illumina HiSeq2500 system with a paired-end 125 bp read length. Here, we focus on gene and population-specific estimates of Tajima's D for non-synonymous sites (Figure S4) and for π_N/π_S which we extracted using popoolation and popoolation2 (Kofler, Orozco-Wengel, et al., 2011; Kofler, Pandey, et al., 2011). Tajima's D was standardized and π_N/π_S ratios \log_{10} transformed prior to statistical modeling. Gene-specific relative divergence (F_{ST}) was estimated using the unbiased method proposed by Hivert et al. (2018) as implemented in poolstat, using default options. Absolute divergence (d_{XY}) was estimated as $A_X B_Y + A_Y B_X$, where A and B are the frequencies of two variants at a given site and X and Y represent the two populations being compared (Delmore et al., 2015; Dennenmoser et al., 2017), and gene-specific estimates were derived by summing over the gene (i.e., the window) and dividing by the total number of sites (variable or not) (see Van Doren et al., 2017). To normalize residuals, F_{ST} was logit transformed and d_{XY} \log_{10} transformed prior to statistical modeling.

To identify overrepresentation of Gene Ontology terms, we used a hypergeometric test with a p -value cutoff $<.05$ as implemented

in the GOSTATS package v.2.46.0 (Falcon & Gentleman, 2007). Gene universes are defined in Table S1.

2.3 | Gene classes

Because certain classes of genes are of particular interest in terms of population divergence and incipient speciation, we also explored differentiation in a few focal classes of genes. These were genes encoding (i) male ejaculate proteins (Bayram et al., 2017, 2019), (ii) female reproductive proteins (Bayram et al., 2019), (iii) enzymes involved in digestion and food processing in larvae and (iv) genes located on the sex chromosomes (Sayadi et al., 2019).

3 | RESULTS

3.1 | General patterns of differentiation

The three populations, which were separated some 2000 years ago (see Section 2), showed considerable overall relative differentiation (mean F_{ST} , 95% bias-corrected bootstrap CI; B-C: 0.248, 0.245–0.251; C-Y: 0.234, 0.230–0.236; B-Y: 0.256, 0.252–0.259), comparable in magnitude to previous estimates of differentiation between populations of seed beetles sampled in nature (Duan et al., 2016). To test prediction V, we asked to what extent gene-specific population differentiation is shared between the three possible pair-wise comparisons (i.e., B-C, C-Y, and B-Y). In order to allow explicit assessment of the repeatability of differentiation, we first performed a principal component analysis of the three vectors of pair-wise gene-specific logit F_{ST} across all genes. This yielded a first PC which accounted for 49% of all variance in F_{ST} across genes and all three F_{ST} vectors loaded positively on this first PC (0.68, 0.71, and 0.71). A one-way ANOVA of all three values of gene-specific F_{ST} , using gene as a factor, yielded an intra-class correlation coefficient (repeatability) across the three comparisons of $R = .25$ ($F_{15305,30612} = 1.94$, $p < .001$). Gene identity thus explained some 25% of all the variance in F_{ST} . The same exercise for $\log_{10} d_{XY}$ yielded a first PC that accounted for 89% of all variance in d_{XY} and all three d_{XY} vectors also loaded positively on this first latent variable (0.94, 0.95, and 0.95). The intraclass correlation coefficient here was $R = .85$ ($F_{19589,37752} = 17.81$, $p < .001$) showing that gene identity explained as much as 85% of all the variance in d_{XY} . Moreover, the overall signals of relative and absolute differentiation were quite distinct, as the correlation across genes between the two vectors $PC1_{FST}$ and $PC1_{dXY}$ was as low as $r = .15$. We note that the correlation between F_{ST} and d_{XY} across genes was modest also for specific population pairs (B-C: $r = .20$; C-Y: $r = .19$; B-Y: $r = .27$).

These analyses are consistent with our prediction V, in that population pairs showed very restricted similarity in terms of which genes exhibited high relative differentiation. Some ¼ of all variance in F_{ST} across genes was shared between the three comparisons, implying

that ¼ is due to population specificity. For d_{XY} , in contrast, more than ¾ of all variance was instead shared between the three pair-wise comparisons. In an effort to understand if inherent properties of genes are associated with overall contribution to differentiation, $PC1_{FST}$ was regressed against gene length and GC content. For $PC1_{FST}$, this showed a positive association with both gene length ($\beta' = .15$, $t = 18.7$, $p < .001$) and GC content ($\beta' = .08$, $t = 9.7$, $p < .001$) but the model only accounted for 3% of the variance in $PC1_{FST}$. For $PC1_{dXY}$, neither the association with gene length ($\beta' = .013$, $t = 1.79$, $p = .073$) nor with GC content ($\beta' = -.011$, $t = -1.52$, $p = .127$) was significant and the model only accounted for 0.03% of the variance. The addition of information of gene-specific expression (i.e., FPKM values) did not significantly improve either of these two models ($p > .09$ in both cases). Hence, genes varied substantially and to some extent consistently in their contribution to absolute population differentiation, but gene length, GC content, and gene expression were not strong predictors of their general role in differentiation.

We next asked if genes that generally showed highest overall absolute and relative divergence could be characterized functionally. Gene ontology enrichment analyses of the 400 genes with the highest values of $PC1_{FST}$ showed significant enrichment for genes involved in a variety of biological processes, most of which involved various aspects of molecular transport and/or localization (Table S1). In particular, genes involved in transmembrane transport were overrepresented among the most divergent genes. The 400 genes with the highest values of $PC1_{dXY}$, however, showed no apparent pattern of enrichment (Table S1). Moreover, $PC1_{dXY}$ showed a positive relationship with $\pi N/\pi S$ within the three populations (multiple regression; $F_{3,11459} = 47.11$, $p < .001$; all three β 's positive) while this relationship was strongly negative for $PC1_{FST}$ ($F_{3,11266} = 225.23$, $p < .001$; all three β 's negative). This implies that genes generally showing high absolute differentiation are under less selective constraint while this pattern is reversed for genes showing high relative differentiation.

3.2 | Gene classes

On average, genes encoding male seminal fluid proteins showed a higher absolute differentiation relative to other genes, while this was not true for female reproductive proteins (Figure 2; Table S3). Absolute differentiation was in fact significantly higher for male seminal fluid proteins than for female reproductive proteins (Welch's *t*-test of $\log_{10} d_{XY}$; B vs. C: $t = 3.05$, $p = .002$; B vs. Y: $t = 1.82$, $p = .069$; C vs. Y: $t = 2.62$, $p = .009$). This difference was, however, not seen for relative population differentiation (Welch's *t*-test of logit F_{ST} ; $t < 1.40$ and $p > .16$ in all three cases; Figure 2).

Relative, but not absolute, mean differentiation was higher in genes encoding enzymes involved in digestion in larvae compared with other genes (Table S3; Figure 2). A direct comparison with male seminal fluid proteins showed that these showed significantly higher absolute (Welch's *t*-test of $\log_{10} d_{XY}$; B vs. C: $t = 2.58$, $p = .009$; B vs. Y: $t = 1.32$, $p = .186$; C vs. Y: $t = 2.35$, $p = .019$) but not relative

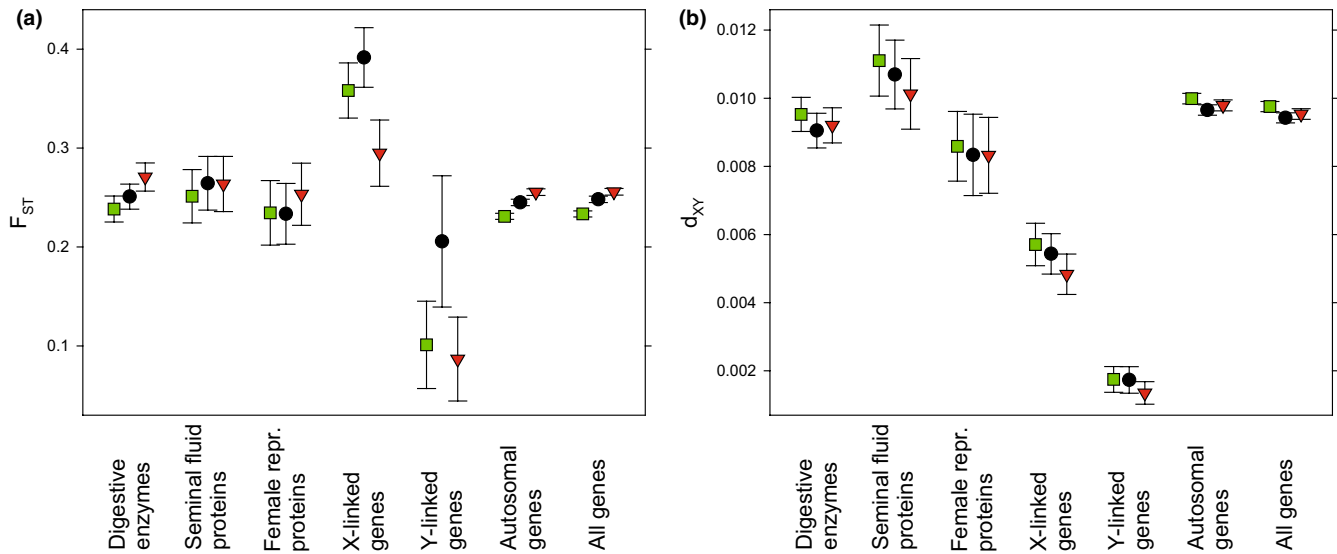


FIGURE 2 Overall population differentiation for functional sets of genes across three *C. maculatus* populations. Shown are mean (\pm bootstrap 95% CI) (a) relative and (b) absolute population differentiation for pairs of populations (green: C vs. Y; black: B vs. C; red: B vs. Y). (a) Relative divergence was highest for X-linked genes but was also somewhat elevated in digestive enzymes and male seminal fluid proteins. (b) Absolute divergence was markedly elevated in male seminal fluid proteins and was reduced in X- and Y-linked genes. See Table S3 for statistical evaluation.

(Welch's t -test of $\logit F_{ST}$; $t < 0.35$ and $p > .72$ in all three cases) differentiation compared with digestive enzymes.

Genes located on the sex chromosomes often, but not always, show higher rates of divergence compared with autosomal loci, which is seen as the combined result of their sex-biased inheritance, lower recombination rates, and lower effective population size (Meisel & Connallon, 2013; Wilson Sayres, 2018). We found that mean relative differentiation was high among X-linked loci (Table S3, Figure 2). In contrast, absolute divergence was very low for X-linked loci, and the elevated relative differentiation seen was primarily a result of the fact that X-linked genes in *C. maculatus* show low nucleotide diversity within populations (Sayadi et al., 2019). This contrast can thus serve to illustrate the fact that relative and absolute measures of population differentiation can sometimes show very different signals (Cruickshank & Hahn, 2014). The low differentiation seen here for Y-linked genes is consistent with previous findings of relatively low nucleotide diversity and efficient purifying selection in most hemizygous genes located on the Y (Sayadi et al., 2019), which is presumably shared among populations to a large extent.

3.3 | Variation within and differentiation between populations

One of our goals was to test our predictions by examining global relationships across genes between the overall properties of site frequency spectra within populations to differentiation between populations. We did so in a series of explicit general linear models of the relationship between F_{ST} or d_{XY} between populations on one hand and metrics summarizing the gene-specific pattern of

segregating genetic variation within populations on the other. We were particularly interested in how common within population molecular estimates indices of selection, here represented by Tajima's D (Tajima, 1989) for non-synonymous sites and by $\pi N/\pi S$, related to estimates of population differentiation. Tajima's D provides a classic and useful summary that relates the frequency spectrum of segregating sites to expectations under the neutral equilibrium model, where positive values correspond to an excess of common polymorphisms consistent with balancing selection and negative values are consistent with a recent selective sweep (Fijarczyk & Babik, 2015). Its absolute value, however, is sensitive to demographic processes and significance testing of single estimates can be difficult. We note that our analyses are based on variation in Tajima's D across genes, which all share a common recent demographic history within populations, and we do not engage in tests of significance of Tajima's D .

The correlation between Tajima's D for non-synonymous sites and $\log \pi N/\pi S$ across all genes was very low (B: $r = -.05$; C: $r = .01$; Y: $r = -.04$), illustrating that these two metrics summarize distinct aspects of the nature of segregating genetic variation within populations.

Our analyses uncovered several patterns. These models are summarized in Tables 1 and 2. Across all genes, Tajima's D predicted both relative and absolute differentiation between populations, but did so in very distinct ways. Variation in F_{ST} was primarily determined by the interaction between Tajima's D in the two populations, where genes showing positive Tajima's D in one population but negative Tajima's D in the other were those genes that overall contributed most to relative population differentiation (Figure 3; see Figure S1 for fitted function). This was true irrespective of whether variation in potentially confounding factors was controlled for or not (Table 1).

TABLE 1 The results of general linear models of the relationship between F_{ST} (logit transformed) between pairs of populations and segregating genetic variation within those populations, across all genes (model A1: $N = 9735$; A2: $N = 8558$; B1: $N = 9888$; B2: $N = 8690$; C1: $N = 10,150$; C2: $N = 8896$)

Model	A: F_{ST} (Brazil - California)					B: F_{ST} (Brazil - Yemen)					C: F_{ST} (California - Yemen)				
	β	SE_{β}	t	p		β	SE_{β}	t	p		β	SE_{β}	t	p	
1	Brazil D_{NS}	.057	0.023	2.52	.012	Brazil D_{NS}	.230	0.021	10.71	<.001	California D_{NS}	.303	0.022	13.76	<.001
	California D_{NS}	.411	0.023	18.14	<.001	Yemen D_{NS}	.355	0.022	16.37	<.001	Yemen D_{NS}	.130	0.022	5.89	<.001
	Brazil D_{NS}^2	.060	0.018	3.38	.001	Brazil D_{NS}^2	.008	0.017	0.47	.640	California D_{NS}^2	-.012	0.017	-0.68	.500
	California D_{NS}^2	-.046	0.018	-2.55	.011	Yemen D_{NS}^2	-.041	0.018	-2.32	.020	Yemen D_{NS}^2	.041	0.018	2.31	.021
	Brazil $D_{NS} \times$ California D_{NS}	-.550	0.021	-26.07	<.001	Brazil $D_{NS} \times$ Yemen D_{NS}	-.566	0.020	-27.93	<.001	California $D_{NS} \times$ Yemen D_{NS}	-.547	0.021	-26.55	<.001
2	Brazil D_{NS}	.003	0.022	0.16	.876	Brazil D_{NS}	.185	0.021	8.82	<.001	California D_{NS}	.273	0.021	12.80	<.001
	California D_{NS}	.368	0.022	16.42	<.001	Yemen D_{NS}	.318	0.021	14.79	<.001	Yemen D_{NS}	.084	0.022	3.89	.022
	Brazil D_{NS}^2	.066	0.017	3.76	<.001	Brazil D_{NS}^2	.012	0.017	0.70	.486	California D_{NS}^2	-.013	0.017	-0.77	.444
	California D_{NS}^2	-.032	0.018	-1.79	.074	Yemen D_{NS}^2	-.041	0.017	-2.34	.019	Yemen D_{NS}^2	.056	0.017	3.20	.001
	Brazil $D_{NS} \times$ California D_{NS}	-.491	0.021	-23.66	<.001	Brazil $D_{NS} \times$ Yemen D_{NS}	-.527	0.020	-26.39	<.001	California $D_{NS} \times$ Yemen D_{NS}	-.480	0.020	-23.98	<.001
GC content	.002	0.004	0.48	.630	GC content	.007	0.003	2.12	.034	GC content	.002	0.003	0.51	.608	
Gene length	4.81E-05	4.60E-06	10.47	<.001	Gene length	3.99E-05	4.31E-06	9.25	<.001	Gene length	4.26E-05	4.38E-06	9.72	<.001	
Brazil pN/pS	-.173	0.044	-3.91	<.001	Brazil pN/pS	-.191	0.037	-5.21	<.001	California pN/pS	-.451	0.047	-9.65	<.001	
California pN/pS	-.416	0.050	-8.26	<.001	Yemen pN/pS	-.238	0.034	-6.91	<.001	Yemen pN/pS	-.084	0.034	-2.45	.014	

Note: The value of (standardized) Tajima's D_{NS} within populations alone explained a sizeable proportion of variance in F_{ST} across genes (model A1: 11%; B1: 13%; C1: 10%) and these effects remained strong when statistically controlling for GC content, gene length and the pN/pS ratios (model 2). Adding information on gene-specific absolute expression (i.e., FPKM values) to model 2 did not significantly improve model fit (A2: $p = .432$; B2: $p = .236$; C2: $p = .830$) and neither did adding information on sex bias in gene expression (i.e., $\log_2 FC$ quartile) (A2: $F_{7,3607} = 1.98$, $p = .053$; B2: $F_{7,3590} = 1.67$, $p = .111$; C2: $F_{7,3740} = 1.41$, $p = .195$).

TABLE 2 The results of general linear models of the relationship between d_{XY} (\log_{10} transformed) between pairs of populations and segregating genetic variation within those populations, across all genes (model A1: $N = 10,096$; A2: $N = 8616$; B1: $N = 10,289$; B2: $N = 8766$; C1: $N = 10,547$; C2: $N = 8968$)

Model	A: d_{XY} (Brazil - California)						B: d_{XY} (Brazil - Yemen)						C: d_{XY} (California - Yemen)					
	β	SE_{β}	t	p	β	SE_{β}	t	p	β	SE_{β}	t	p	β	SE_{β}	t	p		
1	Brazil D_{NS}	.008	0.0032	2.59	.009	.020	0.0033	6.00	<.001	California D_{NS}	.009	0.0031	2.95	.003				
	California D_{NS}	.017	0.0032	5.28	<.001	.032	0.0033	9.76	<.001	Yemen D_{NS}	.023	0.0031	7.55	<.001				
	Brazil D_{NS}^2	-.023	0.0025	-9.35	<.001	-.028	0.0026	-10.59	<.001	California D_{NS}^2	-.038	0.0024	-15.58	<.001				
	California D_{NS}^2	-.037	0.0025	-14.59	<.001	-.035	0.0027	-13.07	<.001	Yemen D_{NS}^2	-.032	0.0025	-12.91	<.001				
	Brazil $D_{NS} \times$ California D_{NS}	.000	0.0030	-0.01	.993	-.007	0.0031	-2.23	.026	California $D_{NS} \times$ Yemen D_{NS}	.000	0.0029	0.06	.953				
	Brazil D_{NS}	.007	0.0031	2.09	.037	.016	0.0033	4.81	<.001	California D_{NS}	.007	0.0031	2.12	.034				
	California D_{NS}	.016	0.0032	5.12	<.001	.033	0.0034	9.71	<.001	Yemen D_{NS}	.024	0.0031	7.78	<.001				
	Brazil D_{NS}^2	-.026	0.0025	-10.51	<.001	-.031	0.0026	-11.82	<.001	California D_{NS}^2	-.038	0.0024	-15.79	<.001				
	California D_{NS}^2	-.037	0.0025	-14.86	<.001	-.037	0.0027	-13.60	<.001	Yemen D_{NS}^2	-.033	0.0025	-13.28	<.001				
	Brazil $D_{NS} \times$ California D_{NS}	-.002	0.0029	-0.56	.573	-.008	0.0031	-2.60	.009	California $D_{NS} \times$ Yemen D_{NS}	.003	0.0029	1.16	.247				
GC content	.001	0.0005	2.01	.045	.000	0.0005	0.92	.356	GC content	.001	0.0005	1.85	.064					
Gene length	-6.00E-06	6.53E-07	-9.19	<.001	-7.09E-06	6.79E-07	-10.44	<.001	Gene length	-6.10E-06	6.35E-07	-9.60	<.001					
Brazil pN/pS	.018	0.0062	2.80	.005	.029	0.0058	5.09	<.001	California pN/pS	.029	0.0066	4.40	<.001					
California pN/pS	.030	0.0070	4.36	<.001	.014	0.0054	2.64	.008	Yemen pN/pS	.011	0.0050	2.21	.027					

Note: The value of (standardized) Tajima's D_{NS} within populations alone explained a significant proportion of variance in d_{XY} across genes (model A1: 4%; B1: 5%; C1: 6%) and these effects remained when statistically controlling for GC content, gene length and the pN/pS ratios (model 2). Adding information on gene-specific absolute expression (i.e., FPKM values) to model 2 did not significantly improve model fit (A2: $p = .413$; B2: $p = .092$; C2: $p = .170$). In contrast, the addition of information on sex bias in gene expression significantly improved model fit in all three cases (i.e., \log_2 FC quartile) (A2: $F_{7,3607} = 10.2, p < .001$; B2: $F_{7,3590} = 9.79, p < .001$; C2: $F_{7,3741} = 8.72, p < .001$).

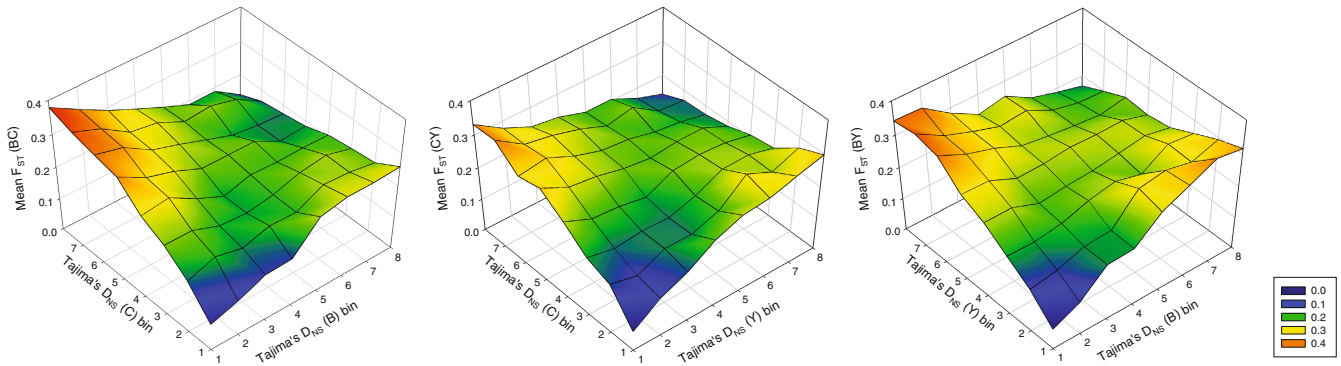


FIGURE 3 3D mesh plots of average F_{ST} across all genes as a function of Tajima's D (based on non-synonymous SNPs) within populations. Here, genes have been binned into eight groups ($N = 1293$ – 1522 genes per bin; see Figure S4) by the value of Tajima's D , to illustrate the fact that genes under positive selection in both populations (i.e., the most negative values of Tajima's D) have the lowest values of F_{ST} overall. Genes showing an apparent signal of balancing selection in one population and positive selection in the other have the highest overall F_{ST} (see Table 1 for statistical evaluation). The comparison B-C is shown to the left, C-Y in the Centre, and B-Y to the right.

In contrast, genes showing negative or positive Tajima's D in both populations contributed the least to overall differentiation.

The analyses of variation in Tajima's D across genes are thus consistent with prediction I in showing that genes that exhibit an excess of rare variants in one population and a scarcity of rare variants in the other are those that contribute most to relative population differentiation (F_{ST}), such that an apparent imbalance in inferred gene-specific selection between two populations was associated with differentiation between them. Next, we tested prediction II by asking whether such genes tend to be enriched with genes with particular functions. We selected those genes that showed an absolute value of the difference in Tajima's $D_{NS} > 2.5$ in two populations (imbalance in inferred selection) while simultaneously showing $F_{ST} > 0.3$ between them (high relative differentiation). These three subsets of genes (B vs. C: $N = 238$; B vs. Y: $N = 275$; C vs. Y: $N = 164$) were then subjected to gene ontology enrichment analyses. Analyses of these sets (Table S1) revealed significant enrichment for a series of gene ontology terms, with a striking overrepresentation of genes involved in various metabolic processes (Figure S2). For example, significant terms occurring in at least two out of the three population comparisons (and with enrichment based on at least two genes in the set) included the parental biological process term “single-organism metabolic process” but also the children terms “ncRNA metabolic process” and “organonitrogen compound catabolic process.” Collectively, these analyses are thus consistent with prediction II and suggest that genes that show signs of balancing selection in one population but positive selection in the other and that simultaneously contribute to relative population differentiation tended to be genes involved in a variety of metabolic processes, including catabolism and biosynthesis (Figure S2).

In agreement with prediction VI, the above analyses show that although the repeatability of relative population differentiation was low at the gene level (see above) it was high at the level of broad functional classes of genes: 85% (28/33) of all significantly enriched terms involved metabolic processes (i.e., contained the term “metabolic,” “catabolic,” or “biosynthetic”) (Figure S2).

To assess prediction III, which is founded in the fact that metabolic rate phenotypes have been identified as a key polygenic phenotype in this system, we first selected all genes annotated as being involved in processes that regulates the rate of metabolic organismal pathways (i.e., Biological Process GO:0019222) (in total $N = 269$). Mean logit F_{ST} was higher in metabolic rate genes compared with all other genes in all three population comparisons, and significantly so in two out of three cases (Welch's t -test; B vs. C: $t = 1.13$, $p = .26$; B vs. Y: $t = 2.70$, $p = .007$; C vs. Y: $t = 2.17$, $p = .031$). More importantly, we replicated our three main models (i.e., model 1 in Table 1) using only this subset of genes, predicting that imbalances in Tajima's D between populations would predict relative differentiation between them. These analyses showed that the interaction term between Tajima's D indeed still had the strongest effect on relative differentiation between them (Figure S3; Table S2). Compared with models including all genes (Table 1), models of this subset of genes showed somewhat accentuated effects of the interaction term relative to all other effects in the model in two out of three comparisons (mean effect size [f^2] for the interaction term, B vs. C: from 0.066 to 0.034; B vs. Y: from 0.074 to 0.088; C vs. Y: from 0.067 to 0.082) (Cohen, 2013). Although not entirely conclusive, these analyses are at least broadly consistent with a predicted important role for metabolic rate genes in polygenic differentiation in this system.

The relationship between absolute differentiation between population pairs and indices of selection was very distinct from that of relative differentiation described above. Variance in d_{XY} was primarily determined by the quadratic terms of Tajima's D in both populations (Table 2), such that genes showing intermediate and near-zero values of Tajima's D in both populations were those genes that overall contributed most to absolute population differentiation (Figure 4). Models of variance in d_{XY} were less predictive of differentiation compared with those of F_{ST} but the pattern, again, was robust against the potential impact of confounding factors (Table 2). Moreover, the association between π_N/π_S and d_{XY} described the inverse of that between π_N/π_S and F_{ST} (see signs of coefficients in model 2 in Tables 1 and 2). Genes showing relatively high π_N/π_S

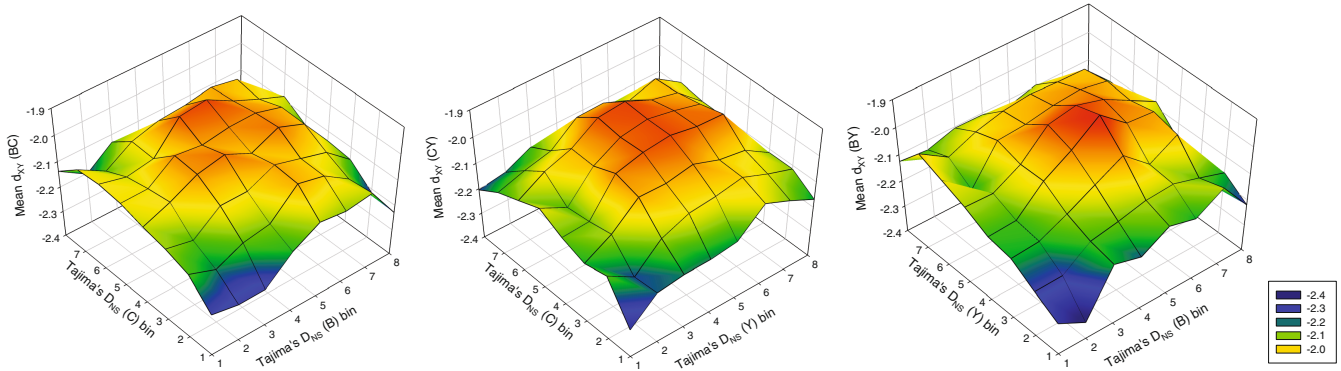


FIGURE 4 3D mesh plots of average d_{XY} (\log_{10} transformed) across all genes as a function of Tajima's D (based on non-synonymous SNPs) within populations. Here, genes have been binned into eight groups ($N = 1293$ – 1522 genes per bin; see Figure S4) by the value of Tajima's D , to illustrate the fact that genes showing the greatest average absolute divergence are those with intermediate values of Tajima's D in both populations. These are generally close to zero (see Table 2 for statistical evaluation). The comparison B-C is shown to the left, C-Y in the Centre, and B-Y to the right.

ratios in both populations, consistent with lower selective constraint (Kryazhimskiy & Plotkin, 2008), were those that contributed most to overall absolute differentiation. These analyses thus suggest that genes that exhibit segregating variation more consistent with the neutral expectation in both populations were those that contributed most to absolute population differentiation. This is consistent with a relatively more important role for genetic drift in generating absolute differentiation.

We note that our models showed that gene-specific relative differentiation is better predicted by Tajima's D than by π_N/π_S (Table 1), consistent with prediction VII. Gene length was associated with both relative and absolute differentiation, but was so in contrasting ways (Tables 1 and 2). While longer genes showed higher relative differentiation on average, they showed lower absolute differentiation.

3.4 | Sex-biased gene expression and population differentiation

Sayadi et al. (2019) showed that the hallmarks of balancing selection, presumably through sexually antagonistic pleiotropy, are particularly pronounced in genes with intermediate female bias in expression in *C. maculatus*. This was true for both within-population indices of selection and the degree of shared polymorphism across populations. They also demonstrated that strong male and female bias in expression are both associated with signs of weak selection, presumably because the efficacy of directional selection is reduced in genes with sex-limited or sex-biased expression (Dapper & Wade, 2020). We therefore also assessed the pattern of population differentiation across sex-biased gene expression, for the subset of genes where data on sex-specific expression is available (Immonen et al., 2017). Although relative differentiation did tend to be higher for genes with female bias in expression, this was not significant for either population comparison (Figure 5).

Genes with sex-biased expression, especially male-biased expression, did show considerably and significantly higher absolute

divergence (Figure 5). This could be interpreted as being consistent with a dominant role for drift in absolute differentiation between the populations studied here (Dapper & Wade, 2020). However, we note that when adding population-specific π_N/π_S ratios as covariates to models of the effect of sex bias in gene expression on d_{XY} (Figure 5), the effects of sex bias in gene expression remained highly significant ($p < .001$ in all three cases). This implies that properties of sex-biased genes other than a lower efficacy of selection contribute to the higher level of absolute differentiation seen in sex-biased genes. One possible interpretation is that female-biased genes to a larger extent show shared polymorphism and lower absolute differentiation due to balancing selection, while male-biased genes to a larger extent contribute to absolute population differentiation through divergent positive selection. In line with this interpretation, Sayadi et al. (2019) found that female-biased genes tend to show overall positive but male-biased genes negative Tajima's D within these populations.

3.5 | Effect size and fixation

Testing prediction VIII is greatly complicated by the fact that we lack explicit measures of the phenotypic effect size of segregating variants, as well as the fact that we lack detailed information of ancestral allele frequencies. To enable at least a provisional test of this prediction, we (i) assume that the number of populations in which a given gene shows non-synonymous polymorphism is inversely related to its fixation probability and (ii) lean on the common but perhaps somewhat dubious assumption that the level of gene expression is overall positively associated with size of allelic effects in that locus across protein-coding genes (e.g., Duret & Mouchiroud, 2000; Gout et al., 2010). We then compared the relationship between gene expression (i.e., overall FPKM values from Immonen et al., 2017) and within-population fixation for non-synonymous sites in genes involved in regulating metabolic processes (Biological Process GO:0019222) with that in all other genes (see Figure 6). This

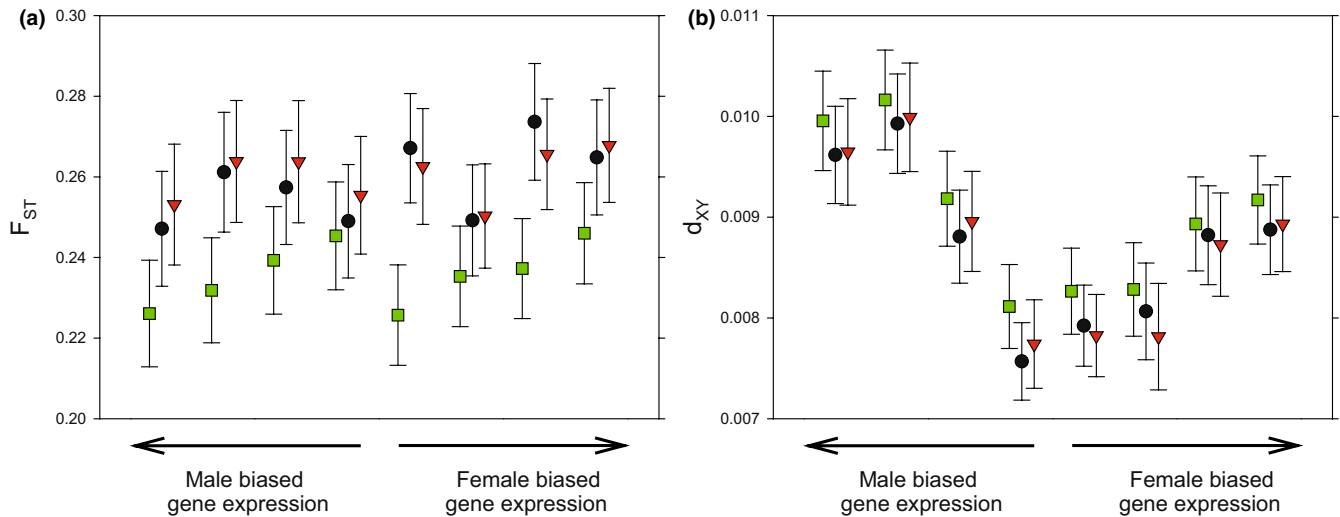


FIGURE 5 The role of sex-biased gene expression for population divergence in three *C. maculatus* populations. Shown are mean ($\pm 95\%$ bootstrap CI) (a) relative and (b) absolute population differentiation for pairs of populations for genes showing different degrees of sex-biased expression (green: C vs. Y; black: B vs. C; red: B vs. Y). Genes were grouped into quartiles based on their \log_2 FC value, separately for FBGs and MBGs, resulting in eight bins in total. Sample size per bin is $n = 591$ – 656 genes. (a) Relative divergence tended to be highest for female-biased genes and intermediately male-biased genes, but the effect of gene expression was not significant in either population comparison (logit F_{ST} ; C vs. Y: $F_{7,4956} = 2.02$, $p = .053$; B vs. C: $F_{7,4956} = 1.47$, $p = .177$; B vs. Y: $F_{7,4956} = 1.57$, $p = .146$). (b) In contrast, both female-biased and especially male-biased genes showed considerably and significantly higher absolute divergence relative to genes with less sex-biased expression ($\log_{10} d_{XY}$; C vs. Y: $F_{7,4985} = 10.67$, $p < .001$; B vs. C: $F_{7,4985} = 10.50$, $p < .001$; B vs. Y: $F_{7,4985} = 9.74$, $p < .001$) (p values from permutation tests).

comparison is relevant given the prediction that metabolic rate should be a polygenic trait under stabilizing selection in our system. We found that both expression level ($LLR_8 = 15.35$, $p < .001$) and gene class ($LLR_1 = 16.45$, $p < .001$) had significant effects on fixation (Generalized Linear Model [GLIM], with Poisson errors, a log link and an empirical dispersion factor). More importantly, however, the interaction between expression level and gene class was significant ($LLR_8 = 3.13$, $p = .002$). Genes involved in the regulation of metabolic processes tended to show the highest probability of fixation for intermediately expressed genes while other genes showed the highest probability of fixation for highly expressed genes (Figure 6), which is at least consistent with prediction VIII. A more complex GLIM, using mean-centered log FPKM and its squared term as continuous predictors along with gene class, confirmed that the curvature of the pattern visualized in Figure 6 differed significantly for the two gene classes (gene class \times FPKM 2 ; $LLR_1 = 3.87$, $p = .049$) and was convex for metabolic rate genes but concave for other genes.

3.6 | Linked selection

Although the overall pattern described by the above analyses is robust, more detailed inferences about the significance of various terms potentially suffer from the common implicit, but unrealistic, assumption that genes segregate independently during divergence. However, the facts that (i) *C. maculatus* has a fairly large (1.2 gigabases) and repeat-rich (70%) genome with 10 chromosome pairs ($2n = 18 + XX/XY$), (ii) the population divergence is fairly deep

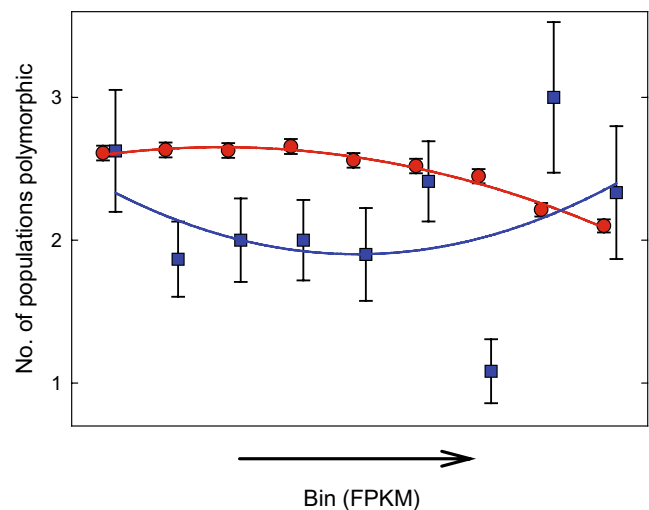


FIGURE 6 The number of populations in which genes show non-synonymous within-population SNPs as a function of increased expression level. Shown are predicted mean (\pm SE) polymorphism for genes involved in the regulation of metabolic processes (blue squares) and all other genes (red circles). Genes were grouped into nine equally sized bins ($N = 552$ – 556 genes per bin), based on their FPKM value. The probability of fixation is here assumed to be inversely related to the ordinate, such that more highly expressed non-metabolic genes were more likely to be fixed for non-synonymous sites within populations. Genes involved in the regulation of metabolic processes showed a significantly different pattern, with highest degree of fixation for intermediately expressed genes. Curves represent fitted squared functions (see text).

(>3000 generations) and differentiation substantial (F_{ST} 's 0.25) and (iii) GC content, often used as a proxy for recombination rate (e.g., Rousselle et al., 2019), was not a strong predictor of gene-specific differentiation; all suggest that linked selection is likely of relatively minor importance in our genome-wide analyses. However, we lack a detailed recombination map of the *C. maculatus* genome, and it is therefore not possible to explicitly model if and how variation in recombination rate across the genome affects divergence. The fact that the genome assembly, although showing a high fraction of well-assembled genes, is highly fragmented (Sayadi et al., 2019) also precludes scans of genomic islands of differentiation (Burri, 2017; Cruickshank & Hahn, 2014).

4 | DISCUSSION

Our main aim was to apply an integrative and phenotypically informed “magnifying glass” approach (sensu Via, 2009) to population genomic data in an attempt to shed some light on the role of stabilizing selection on polygenic life history traits in the evolution of incipient speciation in an insect. Although this is clearly a very challenging task, the patterns documented were well aligned with the predictions made. Overall, thus, we suggest that our findings are consistent with a scenario whereby populations have diverged in polygenic life history traits under stabilizing selection, which we suspect is the likely cause of the emerging reproductive isolation seen between the populations studied here (Figure 1). Yet, we note that a large number of interacting factors and processes should affect population divergence and it currently not possible to firmly assess how unique our predictions are for the scenario considered. Our approach and analyses did generate a series of insight. To us, two main findings stood out.

First, imbalances in inferred selection (i.e., positive Tajima's D in one population and negative in the other) were positively related to F_{ST} across genes. As detailed in the introduction, we suggest that this pattern is predicted for divergence in polygenic traits under stabilizing selection. It is, however, clearly very difficult to entirely exclude the possibility that other processes and effects might also be capable of generating, or at the very least contributing to, this empirical pattern and it is currently not possible to fully evaluate this prediction. For example, there may be intrinsic mathematical dependencies between F_{ST} and properties of the underlying allele-frequency distributions (Jakobsson et al., 2013) that could contribute to complex and non-intuitive relationships between the two. Dedicated modeling efforts would be required to fully evaluate this possibility. Such simulation would, however, be complex and would need to explore variation in a large number of interacting factors and processes (see below) and are beyond the scope of our current contribution. Similarly, genetic drift is predicted to generate higher F_{ST} for loci with higher initial minor allele frequencies, as the Wright-Fisher diffusion process is contingent upon the binomial variance. Loci with more balanced polymorphism at the time of divergence may thus show higher F_{ST} after drift and may simultaneously tend to

show an excess of rare variants in one population but retain a more even distributions of alleles in the other. These sorts of effects are, again, very difficult to account for, but we note that (i) entering π_N/π_S to our models to account for the strength of positive selection did not reduce the effects of the interaction between Tajima's D in our models, (ii) π_N/π_S was negatively, rather than positively, related to F_{ST} and (iii) there was a striking and significant functional enrichment of genes showing high F_{ST} and an imbalance in Tajima's D . These observations are, collectively, difficult to reconcile with a dominating role for random genetic drift in generating the pattern seen.

In addition, we expect the relationship between within-population site frequency spectra and between-population polygenic differentiation to depend on a large number of factors, such as, for example, time since divergence, mutation rate, recombination rate, effective population size, demography, number of genes involved, distribution of allelic effects, strength of phenotypic selection, and differences in optima (Barghi et al., 2020; Barton, 1986, 1989; Fierst & Hansen, 2010; Höllinger et al., 2019). There might thus be conditions where the predicted pattern is most apparent, but it is unclear exactly what those are and explicit modeling efforts aimed at clarifying this would be challenging but helpful. It is also not clear whether the genome-wide pattern that we document is commonly observed among other potential empirical examples of polygenic population divergence. Although a number of studies of genome-wide divergence have estimated both F_{ST} and Tajima's D , few have explicitly related within-population Tajima's D to between-population F_{ST} and among those that have (e.g., Franks et al., 2016) none seem to have modeled the interactive effect predicted and observed here.

Second, phenotype-free bottom-up approaches in speciation genomics suffer from well-known limitations, in particular when adaptive divergence involves polygenic traits under stabilizing selection (e.g., Hayward & Sella, 2022; Korte & Farlow, 2013; Rockman, 2012; Wellenreuther & Hansson, 2016). Several authors have suggested that vertical integration of information and data across different levels of biological organization can help alleviate some of these limitations (e.g., Lee et al., 2014; Pardo-Diaz et al., 2015; Seehausen et al., 2014). For example, phenotypically informed genomic approaches, where an understanding of phenotypic selection is used to generate a priori predictions against which genomic data is analyzed and interpreted, can enrich and improve our inferential basis (Pavlidis et al., 2012; Savolainen et al., 2013). For polygenic traits, Hayward and Sella (2022) suggested that we should perhaps aim to identify traits, rather than specific loci, showing adaptive divergence. Based on a large body of work on phenotypic selection and trait divergence within and between *C. maculatus* populations, we predicted that polygenic life history traits that are determined by metabolic processes should be at the center of population divergence and that such traits should be under net stabilizing selection. Our results match these predictions, as (i) genes showing high F_{ST} and an imbalance in Tajima's D s were enriched with genes involved in metabolic processes, (ii) this enrichment was repeatable over population

comparisons, and (iii) genes known to be involved in the regulation of metabolic rate showed both a higher F_{ST} , a strong association between imbalance in Tajima's D_s and F_{ST} and a distinct pattern of fixation as a function of gene expression. In this sense, our findings can be seen as illustrating that a thorough understanding of phenotypic selection can be used to inform inferences based on genomic data.

Needless to say, using a phenotypically informed approach might be difficult in many systems, where detailed information on phenotypic selection and divergence in life history traits is lacking. Yet, even in systems where little a priori information on phenotypic selection and divergence is available, some general predictions could be made. For example, one would expect that metabolic genes should often stand out. This is because metabolic networks are inherently polygenic and because we generally expect net selection on metabolic fluxes to be stabilizing (Whitlock et al., 1995). Similar predictions could perhaps be made about genes involved in other complex processing, developmental or signaling pathways (Csilléry et al., 2018).

Gene-specific absolute and relative differentiation quantify different aspects of differentiation (Cruickshank & Hahn, 2014) and several studies have found that they are at most only weakly related (e.g., Irwin et al., 2016; Talla et al., 2019; Van Doren et al., 2017; Vijay et al., 2017). Our findings parallel these studies. Moreover, we found that d_{XY} showed a distinct signal in terms of how it related to inferred selection within populations. We found that d_{XY} was consistently highest for genes showing signs of less constraint (higher $\pi N/\pi S$) and allele frequency spectra more consistent with neutrality (near zero Tajima's D_{NS}). In addition, genes with a generally high d_{XY} across all population comparisons (i.e., high $PC1_{d_{XY}}$) were those with high $\pi N/\pi S$ and this set showed no apparent enrichment in terms of gene function, which also suggests that stochastic processes may be major contributors to absolute differentiation. Interestingly, the relatively high repeatability of d_{XY} across the three population comparisons would then imply a lower, rather than higher, redundancy for genes showing less constraint. On the one hand, thus, our collective findings confirm an important role for drift in absolute population differentiation. On the other hand, they also suggest that those genes that show high absolute differentiation may not generally be the most important contributors to incipient genetic incompatibility and reproductive isolation, by the mere fact that they tend to show signs of neutrality.

We also compared differentiation across certain focal functional classes of genes, using annotations based on experimental data in *C. maculatus* (Bayram et al., 2017, 2019; Sayadi et al., 2019). We were particularly interested in estimating the relative potential roles of sexual selection and natural selection in driving population divergence. Sexual selection in *C. maculatus* is to a large extent generated by postmating sexual selection (Fritzsche & Arnqvist, 2013) and a key male phenotype in mediating competitive fertilization success among males is the ejaculate proteome (Yamane et al., 2015). Males transfer more than 300 proteins to females during mating (Bayram et al., 2019), many of which are accessory gland proteins, and

compositional variation in this multidimensional phenotype is striking: the relative abundance of seminal fluid proteins varies markedly across populations and is also related to male fertilization success (Goenaga et al., 2015). Seminal fluid protein genes are thus candidate genes for sexual selection in males. Yet, Sayadi et al. (2019) found little evidence for selection on seminal fluid genes within populations, apart from a signal of stronger purifying selection, and our study shows that these genes generally only show a slightly elevated rate of relative differentiation between populations. They did, however, consistently show a high absolute differentiation (Table S3). This may at least partly reflect functional differences rather than drift and thus contribute to reproductive isolation, given that seminal fluid proteins are known to be functionally important for male reproductive success in this species (Goenaga et al., 2015; Yamane et al., 2015). In agreement with this, and unlike in the entire gene set, $PC1_{d_{XY}}$ did not show a positive relationship with $\pi N/\pi S$ in the three populations across all seminal fluid protein genes (multiple regression; $F_{3,100} = 1.29$, $p = .282$; two out of three β 's even negative) suggesting that drift does not generally promote divergence in male seminal fluid genes. Taken together, these findings are somewhat inconclusive. Male reproductive proteins are known to often evolve rapidly (e.g., Avila et al., 2011) and while it seems very likely that some seminal fluid proteins indeed diverge rapidly as a result of divergent positive sexual selection in *C. maculatus*, this is seemingly manifested as absolute rather than relative differentiation at the scale investigated here.

Genes encoding enzymes that larvae employ to detoxify and digest legume host seeds are prime candidate genes for host adaptation and natural selection in these herbivorous insects (Zhu-Salzman & Zeng, 2015). We found that genes encoding enzymes with digestive function in the larval gut showed markedly elevated relative, but not absolute, differentiation compared to other genes (Table S3). This is in line with a key role for natural selection in population divergence. We conclude that our findings are consistent with important roles for both sexual and natural selection in divergence, but our results also suggest that the mode of selection and differentiation may differ for the two classes of candidate genes inspected here. It is currently unclear, however, whether and how such differences translate into differences in their importance in terms of generating reproductive isolation.

This study represents an attempt to evaluate the role of selection for polygenic population divergence that extends conventional assessments of the role of selective sweeps involving independently acting genes. Although some of the patterns documented here can clearly have alternative or complimentary explanations, we feel that our collective findings are encouraging and suggest that polygenic adaptation may leave a detectable footprint in the covariance between allele frequency spectra within populations and divergence between populations. We found that relative population differentiation was to a large extent built by genes showing excess of rare variants in one population and simultaneously a more even distribution of allele frequencies in the other. Importantly, these genes were enriched with genes with a metabolic function and this

enrichment was repeatable across population comparisons. Across many interacting loci, differences in allele frequencies are likely to result in BDM incompatibilities in crosses between populations, manifested as suboptimal metabolic phenotypes. We suggest that more genome-wide studies that explicitly relate the properties of within-population allele frequency spectra to between-population divergence against a phenotypically informed backdrop are worthwhile, in that they may provide insights into the genetics of incipient speciation. Yet, it is also clear that more efforts to explicitly model these scenarios are needed to more firmly identify the empirical patterns expected.

AUTHOR CONTRIBUTIONS

Göran Arnqvist: Conceptualization (lead); data curation (supporting); formal analysis (equal); funding acquisition (lead); investigation (lead); methodology (lead); project administration (lead); writing – original draft (lead); writing – review and editing (lead). **Ahmed Sayadi:** Data curation (lead); formal analysis (equal); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting).

ACKNOWLEDGMENTS

This work was supported by grants from the European Research Council ERC (GENCON AdG-294333 to GA), the Swedish Research Council VR (621-2014-4523 to GA) and FORMAS (2018-00705 to GA). We are grateful to N. Barton, T. Hansen, and the GENCON lab group for constructive discussions and to E. Arnqvist for support. The computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council, through grant agreement no. 2018-05973.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The annotated genome assembly of *C. maculatus* used here (Sayadi et al., 2019) is available from the European Nucleotide Archive under accession PRJEB30475. Pool-seq raw sequencing data have been deposited at the NCBI sequence read archive, under the accession number PRJNA503561.

ORCID

Göran Arnqvist  <https://orcid.org/0000-0002-3501-3376>

REFERENCES

Arnqvist, G., Stojković, B., Rönn, J. L., & Immonen, E. (2017). The pace-of-life: A sex-specific link between metabolic rate and life history in bean beetles. *Functional Ecology*, *31*, 2299–2309.

Arnqvist, G., Vellnow, N., & Rowe, L. (2014). The effect of epistasis on sexually antagonistic genetic variation. *Proceedings of the Royal Society of London B*, *281*, 20140489.

Arnqvist, G., & Tuda, M. (2010). Sexual conflict and the gender load: Correlated evolution between population fitness and sexual

dimorphism in seed beetles. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 1345–1352.

Avila, F. W., Sirot, L. K., LaFlamme, B. A., Rubinstein, C. D., & Wolfner, M. F. (2011). Insect seminal fluid proteins: Identification and function. *Annual Review of Entomology*, *56*, 21–40.

Barghi, N., Hermisson, J., & Schlötterer, C. (2020). Polygenic adaptation: A unifying framework to understand positive selection. *Nature Reviews Genetics*, *21*, 769–781. <https://doi.org/10.1038/s41576-020-0250-z>

Barton, N. H. (1986). The maintenance of polygenic variation through a balance between mutation and stabilizing selection. *Genetical Research*, *47*, 209–216.

Barton, N. H. (1989). The divergence of a polygenic system under stabilizing selection, mutation and drift. *Genetical Research*, *54*, 59–77.

Barton, N. H. (1999). Clines in polygenic traits. *Genetics Research*, *74*, 223–236.

Bayram, H., Sayadi, A., Goenaga, J., Immonen, E., & Arnqvist, G. (2017). Novel seminal fluid proteins in the seed beetle *Callosobruchus maculatus* identified by a proteomic and transcriptomic approach. *Insect Molecular Biology*, *26*, 58–73.

Bayram, H., Sayadi, A., Immonen, E., & Arnqvist, G. (2019). Identification of novel ejaculate proteins in a seed beetle and division of labour across male accessory reproductive glands. *Insect Biochemistry and Molecular Biology*, *104*, 50–57.

Berg, E. C., & Maklakov, A. A. (2012). Sexes suffer from suboptimal lifespan because of genetic conflict in a seed beetle. *Proceedings of the Royal Society of London B*, *279*, 4296–4302.

Berger, D., Berg, E. C., Widegren, W., Arnqvist, G., & Maklakov, A. A. (2014). Multivariate intralocus sexual conflict in seed beetles. *Evolution*, *68*, 3457–3469.

Berger, D., Martinossi-Allibert, I., Grieshop, K., Lind, M. I., Maklakov, A. A., & Arnqvist, G. (2016). Intralocus sexual conflict and the tragedy of the commons in seed beetles. *The American Naturalist*, *188*(4), E98–E112.

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science*, *362*, eaam5979.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*, 1177–1186.

Brodie, E. D. (2000). Why evolutionary genetics does not always add up. In J. B. Wolf, E. D. Brodie, & M. J. Wade (Eds.), *Epistasis and the evolutionary process* (pp. 3–19). Oxford University Press.

Bürger, R. (2005). A multilocus analysis of intraspecific competition and stabilizing selection on a quantitative trait. *Journal of Mathematical Biology*, *50*, 355–396.

Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, *1*, 118–131.

Chakraborty, M., VanKuren, N. W., Zhao, R., Zhang, X., Kalsow, S., & Emerson, J. J. (2018). Hidden genetic variation shapes the structure of functional elements in *Drosophila*. *Nature Genetics*, *50*, 20–25.

Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., Hirschhorn, J. N., Abecasis, G., Altshuler, D., Bailey-Wilson, J. E., Brooks, L. D., Cardon, L. R., Daly, M., Donnelly, P., Fraumeni, J. F., Jr., Freimer, N. B., Gerhard, D. S., Gunter, C., Guttmacher, A. E., ... Collins, F. S. (2007). Replicating genotype-phenotype associations. *Nature*, *447*, 655–660.

Charlesworth, B. (2015). Causes of natural variation in fitness: Evidence from studies of *Drosophila* populations. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 1662–1669.

Charlesworth, B., & Hughes, K. A. (2000). The maintenance of genetic variation in life-history traits. In R. S. Singh & C. B. Krimbas (Eds.), *Evolutionary genetics* (pp. 369–392). Cambridge University Press.

Charlesworth, B., Miyo, T., & Borthwick, H. (2007). Selection responses of means and inbreeding depression for female fecundity in *Drosophila melanogaster* suggest contributions from intermediate-frequency alleles to quantitative trait variation. *Genetics Research*, *89*, 85–91.

- Charlesworth, D. (2003). Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358, 1051–1070.
- Chen, J., Glémin, S., & Lascoux, M. (2017). Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution*, 34, 1417–1428.
- Chevin, L. M. (2019). Selective sweep at a QTL in a randomly fluctuating environment. *Genetics*, 213, 987–1005.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press.
- Connallon, T., & Chenoweth, S. F. (2019). Dominance reversals and the maintenance of genetic variation for fitness. *PLoS Biology*, 17, e3000118.
- Connallon, T., & Clark, A. G. (2014). Balancing selection in species with separate sexes: Insights from Fisher's geometric model. *Genetics*, 197, 991–1006.
- Coyne, J. A., & Orr, A. H. (2004). *Speciation*. Sinauser Associates, Inc.
- Cruikshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23, 3133–3157.
- Csilléry, K., Rodríguez-Verdugo, A., Rellstab, C., & Guillaume, F. (2018). Detecting the genomic signal of polygenic adaptation and the role of epistasis in evolution. *Molecular Ecology*, 27, 606–612.
- Dapper, A. L., & Wade, M. J. (2020). Relaxed selection and the rapid evolution of reproductive genes. *Trends in Genetics*, 36, 640–649.
- Darwin, C. (1859). *On the origin of species by means of natural selection or the preservation of favored races in the struggle for life*. J. Murray.
- Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., Guigó, R., & Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, 24, 1873–1888.
- Dennenmoser, S., Vamosi, S. M., Nolte, A. W., & Rogers, S. M. (2017). Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (*Cottus asper*) revealed by Pool-seq. *Molecular Ecology*, 26, 25–42.
- Dougherty, L. R., van Lieshout, E., McNamara, K. B., Moschilla, J. A., Arnqvist, G., & Simmons, L. W. (2017). Sexual conflict and correlated evolution between male persistence and female resistance traits in the seed beetle *Callosobruchus maculatus*. *Proceedings of the Royal Society B: Biological Sciences*, 284(1855), 20170132.
- Dowling, D. K., Meerupati, T., & Arnqvist, G. (2010). Cytonuclear interactions and the economics of mating in seed beetles. *The American Naturalist*, 176, 131–140.
- Duan, C. X., Li, W. C., Zhu, Z. D., Li, D. D., Sun, S. L., & Wang, X. M. (2016). Genetic differentiation and diversity of *Callosobruchus chinensis* collections from China. *Bulletin of Entomological Research*, 106, 124–134.
- Duret, L., & Mouchiroud, D. (2000). Determinants of substitution rates in mammalian genes: Expression pattern affects selection intensity but not mutation rate. *Molecular Biology and Evolution*, 17, 68–74.
- Falcon, S., & Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics*, 23, 257–258.
- Fierst, J. L., & Hansen, T. F. (2010). Genetic architecture and postzygotic reproductive isolation: Evolution of Bateson–Dobzhansky–muller incompatibilities in a polygenic model. *Evolution*, 64, 675–693.
- Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, 24, 3529–3545.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press.
- Fox, C. W., Czesak, M. E., & Wallin, W. G. (2004). Complex genetic architecture of population differences in adult lifespan of a beetle: Nonadditive inheritance, gender differences, body size and a large maternal effect. *Journal of Evolutionary Biology*, 17, 1007–1017.
- Fox, C. W., Dublin, L., & Pollitt, S. J. (2003). Gender differences in lifespan and mortality rates in two seed beetle species. *Functional Ecology*, 17, 619–626.
- Fox, C. W., Scheibly, K. L., Smith, B. P., & Wallin, W. G. (2007). Inbreeding depression in two seed-feeding beetles, *Callosobruchus maculatus* and *Stator limbatus* (Coleoptera: Chrysomelidae). *Bulletin of Entomological Research*, 97, 49–54.
- Franks, S. J., Kane, N. C., O'Hara, N. B., Tittes, S., & Rest, J. S. (2016). Rapid genome-wide evolution in *Brassica rapa* populations following drought revealed by sequencing of ancestral and descendant gene pools. *Molecular Ecology*, 25, 3622–3631.
- Fritzsche, K., & Arnqvist, G. (2013). Homage to Bateman: Sex roles predict sex differences in sexual selection. *Evolution*, 67, 1926–1936.
- Fricke, C., & Arnqvist, G. (2004). Divergence in replicated phylogenies: The evolution of partial post-mating prezygotic isolation in bean weevils. *Journal of Evolutionary Biology*, 17, 1345–1354.
- Gavrilets, S. (2004). *Fitness landscapes and the origin of species*. Princeton University Press.
- Gillespie, J. H., & Turelli, M. (1989). Genotype-environment interactions and the maintenance of polygenic variation. *Genetics*, 121, 129–138.
- Goenaga, J., Yamane, T., Rönn, J., & Arnqvist, G. (2015). Within-species divergence in the seminal fluid proteome and its effect on male and female reproduction in a beetle. *BMC Evolutionary Biology*, 15, 266.
- Gompert, Z., & Messina, F. J. (2016). Genomic evidence that resource-based trade-offs limit host-range expansion in a seed beetle. *Evolution*, 70, 1249–1264.
- Gossmann, T. I., Keightley, P. D., & Eyre-Walker, A. (2012). The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biology and Evolution*, 4, 658–667.
- Gout, J. F., Kahn, D., Duret, L., & Paramecium Post-Genomics Consortium. (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genetics*, 6, e1000944.
- Grieshop, K., & Arnqvist, G. (2018). Sex-specific dominance reversal of genetic variation for fitness. *PLoS Biology*, 16, e2006810.
- Griffin, P. C., Hangartner, S. B., Fournier-Level, A., & Hoffmann, A. A. (2017). Genomic trajectories to desiccation resistance: Convergence and divergence among replicate selected *Drosophila* lines. *Genetics*, 205, 871–890.
- Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, 26, 5362–5368.
- Hayward, L. K., & Sella, G. (2022). Polygenic adaptation after a sudden change in environment. *eLife*, 11, e66697.
- Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from pool-seq data. *Genetics*, 210, 315–330.
- Höllinger, I., Pennings, P. S., & Hermisson, J. (2019). Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genetics*, 15, e1008035.
- Hotzy, C., & Arnqvist, G. (2009). Sperm competition favors harmful males in seed beetles. *Current Biology*, 19, 404–407.
- Immonen, E., Liljestrand-Rönn, J., Watson, C., Berger, D., & Arnqvist, G. (2016). Complex mitonuclear interactions and metabolic costs of mating in male seed beetles. *Journal of Evolutionary Biology*, 29, 360–370.
- Immonen, E., Sayadi, A., Bayram, H., & Arnqvist, G. (2017). Mating changes sexually dimorphic gene expression in the seed beetle *Callosobruchus maculatus*. *Genome Biology and Evolution*, 9, 677–699.
- Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H., & Owens, G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology*, 25, 4488–4507.

- Jakobsson, M., Edge, M. D., & Rosenberg, N. A. (2013). The relationship between F_{ST} and the frequency of the most frequent allele. *Genetics*, *193*, 515–528.
- Kébé, K., Alvarez, N., Tuda, M., Arnqvist, G., Fox, C. W., Sembène, M., & Espindola, A. (2017). Global phylogeography of the insect pest *Callosobruchus maculatus* (Coleoptera: Bruchinae) relates to the history of its main host, *Vigna unguiculata*. *Journal of Biogeography*, *44*, 2515–2526.
- Kimura, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability*, *1*, 177–232.
- Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlötterer, C. (2011). PoPoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*, *6*, e15925.
- Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-seq). *Bioinformatics*, *27*, 3435–3436.
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, *9*, 29.
- Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genetics*, *4*, e1000304.
- Lazarević, J., Tucić, N., Šešlija Jovanović, D., Večeřa, J., & Kodrík, D. (2012). The effects of selection for early and late reproduction on metabolite pools in *Acanthoscelides obtectus* Say. *Insect Science*, *19*, 303–314.
- Lee, Y. W., Gould, B. A., & Stinchcombe, J. R. (2014). Identifying the genes underlying quantitative traits: A rationale for the QTN programme. *AoB Plants*, *6*, plu004.
- Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. Columbia University Press.
- Mackay, T. F. (2010). Mutations and quantitative genetic variation: Lessons from *Drosophila*. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*, 1229–1239.
- Mani, G. S., Clarke, B. C., & Shelton, P. R. (1990). A model of quantitative traits under frequency-dependent balancing selection. *Proceedings of the Royal Society of London B*, *240*, 15–28.
- Meisel, R. P., & Connallon, T. (2013). The faster-X effect: Integrating theory and data. *Trends in Genetics*, *29*, 537–544.
- Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 332–342.
- Nosil, P., Villoutreix, R., de Carvalho, C. F., Feder, J. L., Parchman, T. L., & Gompert, Z. (2020). Ecology shapes epistasis in a genotype–phenotype–fitness map for stick insect colour. *Nature Ecology and Evolution*, *4*, 1673–1684.
- Pardo-Díaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, *6*, 445–464.
- Patten, M. M., Haig, D., & Ubeda, F. (2010). Fitness variation due to sexual antagonism and linkage disequilibrium. *Evolution*, *64*, 3638–3642.
- Pavlidis, P., Jensen, J. D., Stephan, W., & Stamatakis, A. (2012). A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution*, *29*, 3237–3248.
- Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*, R208–R215.
- Rankin, D. J., & Arnqvist, G. (2008). Sexual dimorphism is associated with population fitness in the seed beetle *Callosobruchus maculatus*. *Evolution*, *62*, 622–630.
- Rice, S. H. (1998). The evolution of canalization and the breaking of von Baer's laws: Modeling the evolution of development with epistasis. *Evolution*, *52*, 647–656.
- Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*, *66*, 1–17.
- Rousselle, M., Laverré, A., Figueat, E., Nabholz, B., & Galtier, N. (2019). Influence of recombination and GC-biased gene conversion on the adaptive and nonadaptive substitution rate in mammals versus birds. *Molecular Biology and Evolution*, *36*, 458–471.
- Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*, 807–820.
- Sayadi, A., Martinez Barrio, A., Immonen, E., Dainat, J., Berger, D., Tellgren-Roth, C., Nystedt, B., & Arnqvist, G. (2019). The genomic footprint of sexual conflict. *Nature Ecology and Evolution*, *3*, 1725–1730.
- Schielzeth, H., Rios Villamil, A., & Burri, R. (2018). Success and failure in replication of genotype–phenotype associations: How does replication help in understanding the genetic basis of phenotypic variation in outbred populations? *Molecular Ecology Resources*, *18*, 739–754.
- Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*, 176–192.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.
- Talla, V., Johansson, A., Dincă, V., Vila, R., Friberg, M., Wiklund, C., & Backström, N. (2019). Lack of gene flow: Narrow and dispersed differentiation islands in a triplet of *Leptidea* butterfly species. *Molecular Ecology*, *28*, 3756–3770.
- Tine, E. M., Diome, T., Khadim, K., Ndong, A., Douma, A., Ketoh, G., Sanon, A., & Sembene, M. (2013). Genetic diversity of *Callosobruchus maculatus* Fabricius (cowpea weevil) populations in various agro-ecological areas of five countries in west African sub-region. *South Asian Journal of Experimental Biology*, *3*, 71–83.
- Turelli, M., & Barton, N. H. (2004). Polygenic variation maintained by balancing selection: Pleiotropy, sex-dependent allelic effects and G × E interactions. *Genetics*, *166*, 1053–1079.
- Van Doren, B. M., Campagna, L., Helm, B., Illera, J. C., Lovette, I. J., & Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology*, *26*, 3982–3997.
- Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9939–9946.
- Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H., & Wolf, J. B. (2017). Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Molecular Ecology*, *26*, 4284–4295.
- Wang, B., Mojica, J. P., Perera, N., Lee, C. R., Lovell, J. T., Sharma, A., Adam, C., Lipzen, A., Barry, K., Rokhsar, D. S., Schmutz, J., & Mitchell-Olds, T. (2019). Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in *Boechera stricta*. *Genome Biology*, *20*, 1–15.
- Wellenreuther, M., & Hansson, B. (2016). Detecting polygenic evolution: Problems, pitfalls, and promises. *Trends in Genetics*, *32*, 155–164.
- Whitlock, M. C., Phillips, P. C., Moore, F. B. G., & Tonsor, S. J. (1995). Multiple fitness peaks and epistasis. *Annual Review of Ecology and Systematics*, *26*, 601–629.
- Wilson Sayres, M. A. (2018). Genetic diversity on the sex chromosomes. *Genome Biology and Evolution*, *10*, 1064–1078.
- Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*, 87–100.
- Wolf, J. B., Lindell, J., & Backström, N. (2010). Speciation genetics: Current status and evolving approaches. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *365*, 1717–1733.
- Wright, S. (1935). Evolution of populations in approximate equilibrium. *Journal of Genetics*, *30*, 257–266.

- Yamane, T., Goenaga, J., Rönn, J., & Arnqvist, G. (2015). Male seminal fluid substances affect sperm competition success and female reproductive behavior in a seed beetle. *PLoS One*, *10*, e0123770.
- Zhu-Salzman, K., & Zeng, R. (2015). Insect response to plant defensive protease inhibitors. *Annual Review of Entomology*, *60*, 233–252.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Arnqvist, G., & Sayadi, A. (2022). A possible genomic footprint of polygenic adaptation on population divergence in seed beetles? *Ecology and Evolution*, *12*, e9440. <https://doi.org/10.1002/ece3.9440>