Title: High-resolution mapping of a rapidly evolving complex trait reveals genotype-phenotype stability and an unpredictable genetic architecture of adaptation

Smiley-Rhodes J.A*[1], Bitter M.C.*[1], Berardi S.[2], Oken H.[2], Beltz J.[2], Petrov D.A.[#1,3], and Schmidt P.[#2]

* = equal contribution
# = equal contribution

1 – Department of Biology, Stanford University, Stanford CA 94305
2 – Department of Biology, University of Pennsylvania, Philadelphia, PA 19104
3 – Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

Correspondance: dpetrov@stanford.edu, schmidtp@sas.upenn.edu

Abstract

The extent to which adaptation can be predicted, particularly for traits with complex genetic bases, is unknown. Here, we leveraged a model complex trait, model species, and high-powered longitudinal sampling design to test the efficacy of genomic prediction of complex trait variation and evolution in ecologically relevant settings. We monitored genome-wide allele frequencies and pigmentation variation in genetically diverse populations of *Drosophila melanogaster* across seven generations of evolution in both field mesocosms exposed to natural environmental fluctuations, as well as mesocosms housed in a controlled, lab-based setting. At two time points throughout trait evolution, we conducted a high-powered, tail-based mapping of pigmentation. We produced a well-resolved genotype-phenotype map that reaffirms canonical pigmentation genes and unveils novel loci underpinning a substantial portion of trait variation. While we were able to use this map to correctly infer the direction of pigmentation evolution in both the field and lab mesocosms, the particular loci responding to selection, and thus architecture of adaptation itself, were largely unpredictable. We suggest this unpredictability to be a result of pleiotropic constraint, which was more pronounced in the field, relative to the lab-based, environment. Finally, we quantified a striking stability of the genotype-phenotype map across genetically diverged populations, demonstrating that shifting epistatic landscapes associated with the evolutionary process itself does not alter trait architecture to an extent that precludes phenotypic prediction, provided the mapping is sufficiently powered. In concert our results highlight both the promise and limitations of genomic prediction, and exemplify the challenges of applying lab-based studies of complex traits to their evolutionary dynamics in the wild.

Introduction

Whether or not adaptation can be predictable is an outstanding question in evolutionary biology (Blount et al., 2018; Wortel et al., 2022). This fundamental question permeates numerous issues of societal concern: from the preservation of species in the face of global change to the management of disease outbreaks. Two distinct ways in which adaptation can be predicted are: i) correctly inferring whether or not a focal trait has evolved, given patterns of genomic variation alone; or ii) correctly inferring the presence and direction of selection on genomic loci, given observations of phenotypic change alone. In order to employ either method of prediction, researchers must first quantify the molecular genetic basis of phenotypic variation, and then understand how such phenotypic variation is translated into fitness-relevant differences among genotypes.

Instances in which adaptation has proven to be predictable are almost exclusively in the context of traits exhibiting simple genetic bases, whereby trait variation is predominantly explained by one or a few large effect loci (Colosimo et al., 2005; Enattah et al., 2008; Hoekstra et al., 2006; Hof et al., 2016; Kunte et al., 2014; Nadeau et al., 2016; Reed et al., 2011). While such case studies provide fundamental insight into the evolutionary process, the vast majority of fitness-relevant traits instead exhibit complex genetic bases, whereby phenotypic differences are due to allelic variation at tens to thousands of loci spread throughout the genome (Sella and Barton, 2019; Visscher et al., 2017). Quantifying, and ultimately predicting, how these traits evolve remains a challenge.

Two facets of quantitative trait variation that may preclude evolutionary predictability are pleiotropy and epistasis. Pleiotropy occurs when a single genomic locus impacts phenotypic variation in multiple traits, a phenomenon expected to be pervasive among the large-effect quantitative trait loci that impact fitness in natural populations (Mackay and Anholt, 2024; Walsh and Blows, 2009). As populations adapt, correlated selection across fitness-relevant traits can then impact the evolution of pleiotropic loci, constraining their effect on the evolution of the trait in question (Svensson et al., 2021). In such a scenario, patterns of genetic variation at a mapped locus may behave as effectively neutral during the evolution of the trait itself, or even indicate trait shifts in the opposite direction of how it in fact evolved through time (Albecker et al., 2025; Conover and Schultz, 1995; Sella and Barton, 2019; Walsh and Blows, 2009).

Epistasis is the interaction of allelic effects at multiple loci and, in the context of predicting trait adaptation, may shift the genotype-phenotype map throughout the evolutionary process (Duncan et al., 2019; Huang et al., 2012; Phillips, 2008; Privé et al., 2022). Specifically, epistasis can cause the marginal effect of trait-associated loci to change as a function of genetic background, which varies across populations with distinct demographic histories and patterns of selection (Duncan et al., 2019; Durvasula and Lohmueller, 2019; Guo et al., 2014; Huang et al., 2012; Kosheleva and Desai, 2018; Lohmueller,

2014; Martin et al., 2017; Novembre and Barton, 2018; Patel et al., 2022; Privé et al., 2022; Yair and Coop, 2021). Similarly, context-dependent dominance may alter the phenotypic effect of alleles as a function of shifting environments (Grieshop et al., 2021). It is therefore unclear whether loci mapped within a particular population and at a particular point in time will be those that underpin phenotypic variation when the trait mean shifts, as is often assumed by studies that use contemporary mapping data to infer past trait evolution using genomic data (Abdellaoui et al., 2023; Akbari et al., 2024; Cox et al., 2019; Halldorsson et al., 2022; Mathieson et al., 2015).

In concert, epistasis and pleiotropy lead to a genetic architecture of adaptation that may be much less predictable based on the genotype-phenotype mapping data alone. As these processes and their effects manifest through time, inferring them using static sequencing data can be challenging. Thus, longitudinal studies, which directly examine selection on individual traits and loci throughout the genome over the progression of evolution itself represent a powerful method to study pleiotropy, epistasis, and the predictability of complex trait evolution (Chiu et al., 2016; Kerner et al., 2009; Lenski, 2017; Stroud and Ratcliff, 2025; Wu et al., 2014). Such studies can probe the predictability of adaptation in a systematic manner, answering fundamental questions, such as: are the loci underpinning phenotypic variation in a complex trait the same that respond to selection and underpin trait evolution in nature? How stable is the trait genotype-to-phenotype map throughout evolution and across different environments associated with distinct selection pressures?

Our study comprised a longitudinal analysis of abdominal pigmentation evolution in *Drosophila melanogaster* within an ecologically realistic context. Pigmentation is one of the most conspicuous complex traits associated with phenotypic diversity across the tree of life, demonstrating adaptive evolution across a range of taxa (Kronforst et al., 2012). In *D. melanogaster*, abdominal pigmentation exhibits latitudinal clines across widespread geographic populations, and rapidly and cyclically evolves in response to seasonally varying selection pressures (Berardi et al., 2025). Importantly, abdominal pigmentation has long served as a model complex trait, whereby the biosynthetic pathway of the two main molecules, darkly pigmented melanin and light-colored sclerotin, have been well-characterized (Massey and Wittkopp, 2016; Wright, 1987). More recent genome wide associations studies have pinpointed three core genes in the melanin/sclerotin biosynthetic pathway that explain the majority of intra- and inter-population pigmentation variation: *bric-a-brac* (*bab*), *tan*, and *ebony* (Bastide et al., 2016, 2013; Dembeck et al., 2015). Interestingly, clines in pigmentation across space and time do not appear to be explained by expected signatures of selection at these pigmentation genes of large effect (Berardi et al., 2025), a pattern suggestive of underlying pleiotropy and/or epistasis but challenging to attribute to these factors directly in the face of confounding population demography and linkage.

Here, using a single genetically diverse founding population we quantified pigmentation evolution across seven generations in two environments with distinct selective pressures: a simple, lab-based environment and a dynamic, semi-natural field mesocosm. We quantified evidence of selection on pigmentation, as well as genome-wide single nucleotide polymorphisms (SNPs) as our focal trait itself evolved rapidly over seven generations in each environment. At two points throughout this process, we conducted a tail-based mapping of replicate populations to independently identify trait-associated loci in each of four total mapping populations (Fig. 1). Our genotype-phenotype mapping both validated the role of previously identified candidate pigmentation genes, and also identified novel loci underpinning pigmentation variation in the species. We found that this map could then be used to correctly infer pigmentation evolution through time in each environment given patterns of allele frequency data alone. Interestingly, however, the loci displaying signatures of selection as pigmentation evolved (i.e., the genetic architecture of adaptation) did not always behave in a manner as predicted by mapping, a dynamic that was exacerbated in the complex, field setting. This pattern argues for the strong effect of pleiotropy affecting the genomic predictability, and underlying genetic architecture, of phenotypic evolution in nature. Finally, we observed genotype-phenotype mapping to be remarkably stable and transferable across mapping populations, indicating that marginal effects of mapped loci remain detectable even amidst shifting epistatic landscapes associated with the evolutionary process itself.

Results

Study scheme and power

Our study population was derived via outcrossing a panel of 76 inbred lines originally collected in local Pennsylvania orchards (Behrman et al. 2018). Several key advantages arise from this methodological decision. By allowing evolution to proceed from a single founding population we could evaluate the effect of epistasis on genotype-phenotype map portability directly and in a manner that is not confounded by population-specific demographic histories and patterns of linkage (Lohmueller, 2014; Novembre and Barton, 2018). Furthermore, the pigmentation and fitness-relevant variation subject to selection in our study was evaluated in the same local region from which it originated and is maintained naturally, and in which we have documented repeatable seasonal evolution of lighter pigmentation from late spring through summer (Berardi et al., 2025). Finally, the haplotype-informed allele frequency estimates we are able to generate from the reconstituted outbred populations are of the accuracy produced using standard pool-seq methods and sequencing depths of roughly 1000x (Tilk et al., 2019). Such haplotype-imputed methods of mapping have been recently shown to be significantly greater powered for complex trait mapping than more standard SNP-by-SNP approaches (Long et al. 2025).

The outbred population was allowed to recombine for four generations, before being used to seed each of 10 large indoor cage replicates maintained in controlled, lab-based conditions (Bitter et al., 2024b), as well as 10 replicate outdoor field cages comprising a semi-natural mesocosm exposed to natural environmental fluctuations in Philadelphia, PA (Fig. 1) (Bitter et al., 2024a, 2024b; Rudman et al., 2022). Populations evolving within the indoor environment were subject to increasing population densities, but in the absence of the suite of fluctuating abiotic conditions imposed upon the populations evolved in the outdoor mesocosm. We allowed the replicate populations to expand and evolve for eight weeks (seven generations) in each environment, during which time we monitored patterns of pigmentation variation and genome-wide allele frequencies at 1.9 M single nucleotide polymorphisms (SNPs). At weeks 2 and 8 of the experiment (generations 2 and 7), we conducted tail-based mapping whereby we selectively isolated and sequenced pools of 75 female flies per cage from the 15% lightest, 15% darkest, and representative midpoint of the trait value distribution (hereafter referred to as 'light', 'dark', and 'midpoint' trait value groups, see Methods) from each of the ten replicate populations of the lab and field mesocosms. Whole genome sequencing reads derived from each sample were then subject to a bioinformatic pipeline that leverages knowledge of the haplotype structure from a founding set of individuals (here, our inbred reference panel) to compute genome-wide haplotype and allele frequency data in evolved, pooled samples (Kessner et al., 2013; Tilk et al., 2019). This resulted in, for each sample, estimates of haplotype frequencies within locally unrecombined windows (quantified using the recombination rate and the number of generations since cage founding) and allele frequencies at each of the 1.9 M SNPs we previously identified in our inbred reference panel.

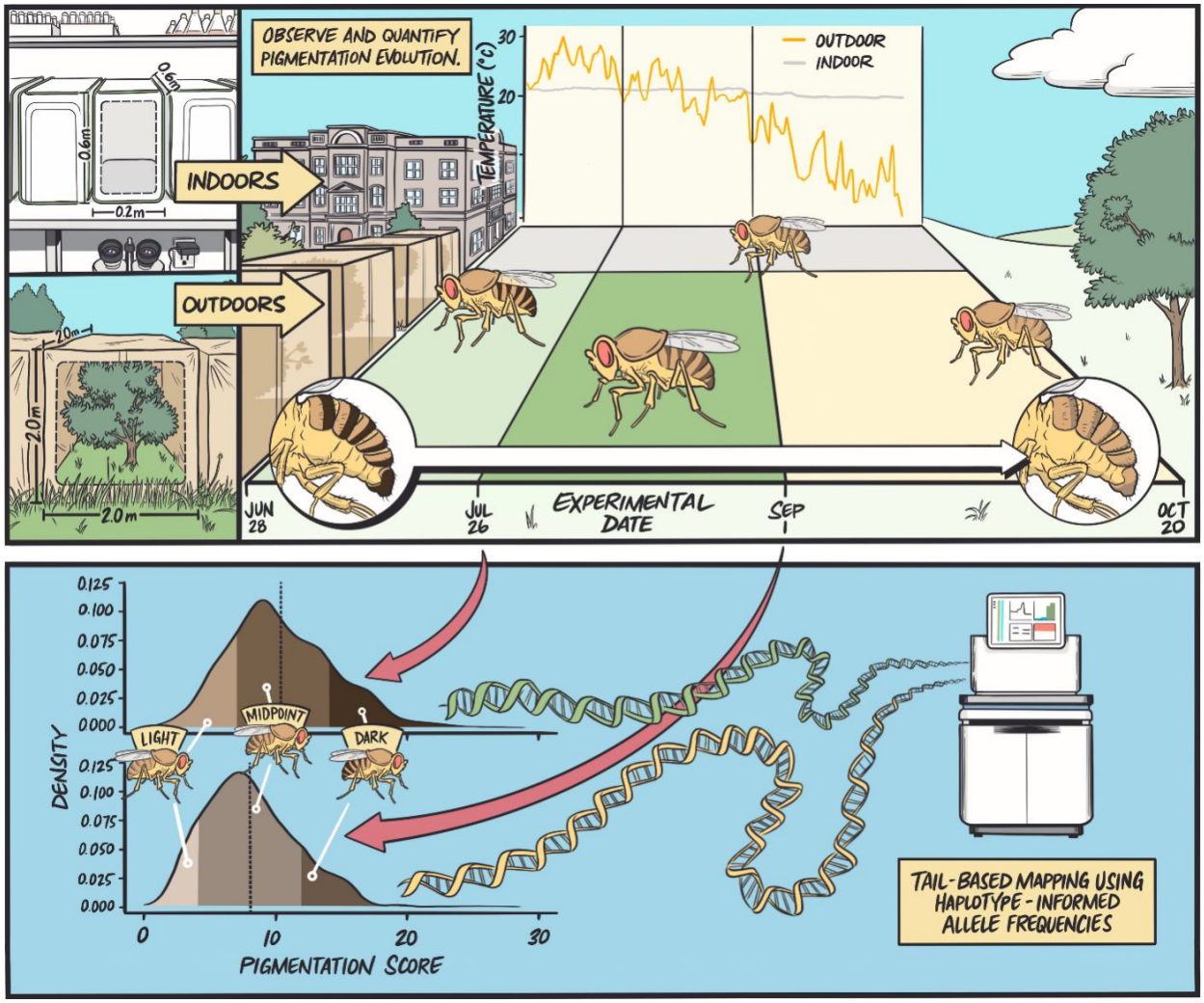


Figure 1 | Experimental design. Ten cages each in an indoor laboratory and in an outdoor mesocosm were seeded with an outbred population of *D. melanogaster*. The replicate populations were allowed to evolve from late June to late October, during which time significant pigmentation evolution was observed (see text). At weeks 2 and 8 of the experiment (generations 2 and 7) we conducted tail-based mapping of pigmentation. Specifically, following two generations of common garden rearing, flies were segregated into color fractions based on the population's distribution of abdominal pigmentation scores (fractions comprised 75 female individuals from the darkest 15%, the lightest 15%, and a representative sample from the trait midpoint) across each of ten replicates per time point and environment. Each color fraction from each replicate and time point was then used in whole-genome sequencing and estimation of high-accuracy, haplotype-informed allele frequencies at 1.9 M SNPs.

Phenotypic and genomic variation across environments and throughout evolution

We first assessed the degree of pigmentation variation in our study population and whether or not it was affected by natural selection (whether direct or indirect) in either the lab or field mesocosm. Specifically, we quantified pigmentation independently for each inbred reference line, across individuals

170 in the derived outbred population (hereafter, 'baseline'), and among replicate populations at weeks two
171 and eight of evolution in each environment. As expected, our baseline population exhibited variation in
172 pigmentation that was representative of the variation observed across the 76 inbred lines from which it
173 was derived (Fig. 2A). Pigmentation after two generations of evolution resulted in significantly lighter
174 pigmentation across replicates in both the lab and field environments, relative to the baseline (pairwise t-
175 tests, week 2 outdoor $p << 0.001$, week 2 indoor $p << 0.001$). Also as expected, the replicate populations
176 then evolved to become significantly lighter during the six additional weeks of evolution in both the
177 outdoor and indoor environments (pairwise t-tests; week 2 vs. 8 outdoor $p << 0.001$, week 2 vs. 8 indoor
178 $= 0.0245$). Notably, a significant treatment by time interaction on variation in pigmentation throughout
179 the experiment was observed (two-way ANOVA; $F = 17.03$, $p << 0.001$), driven by the accelerated rate
180 and magnitude of pigmentation evolution towards lighter pigmentation observed in the outdoor replicates
181 (Fig. 2A).
182 Next, we quantified whether there existed evidence of environment-specific patterns of selection
183 that would, in effect, induce unique epistatic landscapes across lab and field-based mapping populations.
184 Specifically, we ran principal component analysis (PCA; *prcomp* function, R), on the normalized allele
185 frequency data of all 1.9 M SNPs, and then projected samples onto the first two principal components
186 (Fig. 2C-D). Strikingly, the first principal component (6.4% variance explained) segregated samples in
187 accordance with trait group, and along the second principal component (4.8% variance explained) in
188 accordance with environment of selection (Fig. S1). When segregated by chromosome, we observed the
189 same correlation between PC1 and color fraction on all chromosomes except X, and all chromosomes
190 except 2R demonstrated the same relationship between PC2 and treatment (Fig. S2). These patterns were
191 not disrupted when the major cosmopolitan inversions were excluded (Fig. S3). This analysis thus
192 demonstrates the pronounced impact of both phenotypic group and environment of selection on genome
193 wide patterns of variation, while also suggesting that a shared set of loci underpinned variation in
194 pigmentation across mapping populations.

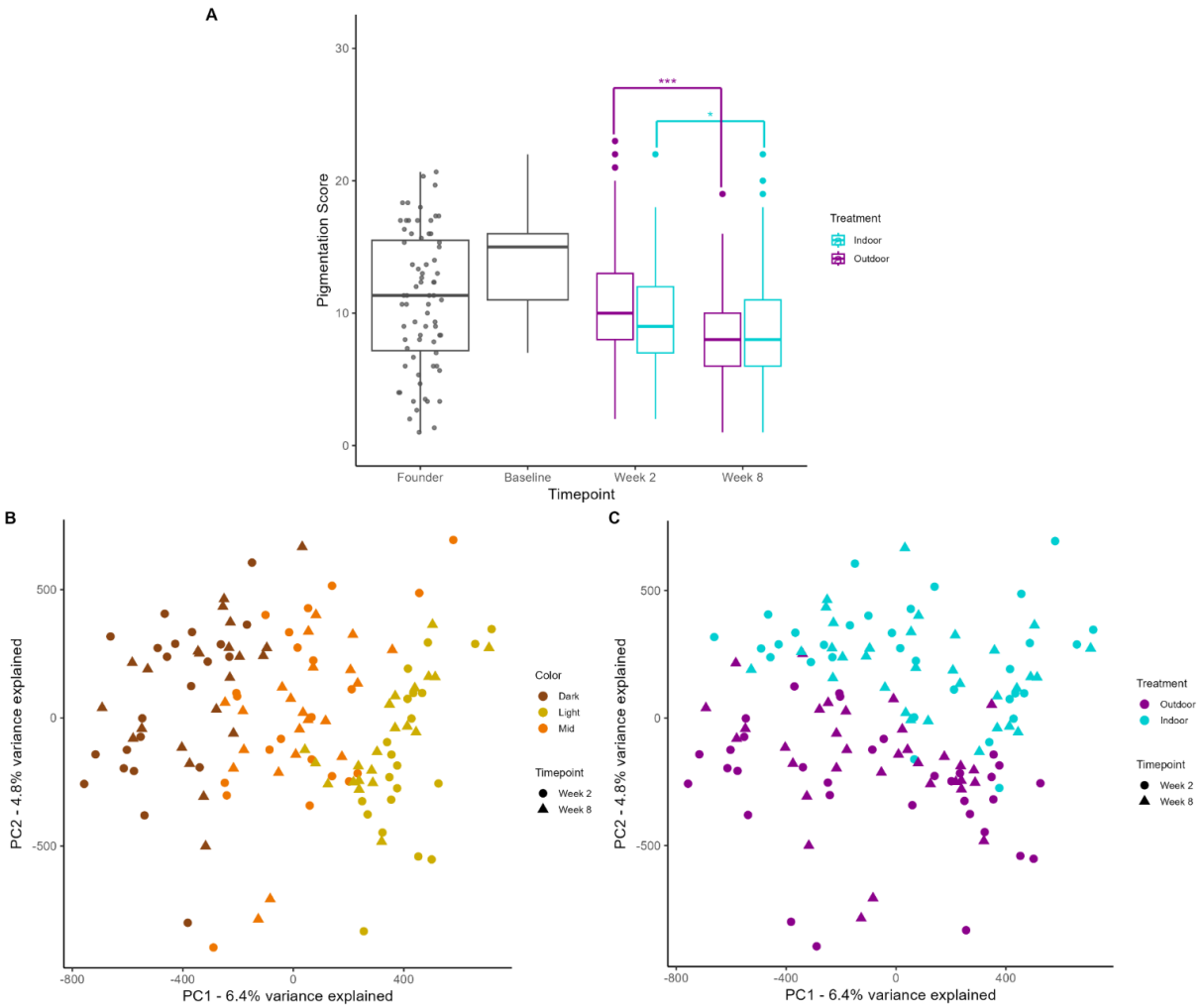


Figure 2 | Phenotypic and genomic variation throughout study period. (A) Distribution of the total pigmentation scores (whereby higher scores indicate increased pigmentation) across samples. "Founder" refers to our 76 inbred lines that we outbred for four generations to generate our baseline (week 0) population. Grey dots each represent the average pigmentation score of N = 3 females from each inbred line. The baseline pigmentation distribution represents the scores of N = 90 females from the initial outbred population which was used to seed our indoor and outdoor cages. From the samples collected during the experiment at weeks 2 and 8 in both environments, we scored 30 females per replicate cage (N = 300 in total for each timepoint/treatment combination) and found a significant shift towards lighter pigmentation in both outdoor and indoor populations (paired t-test; p-values << 0.001). (B-C) Projection of samples used in trait mapping onto the first two principal components generated via PCA of allele frequency data from all 1.9 M SNPs. In (B) samples are colored by the color fraction from which they originate (dark, midpoint, or light) and the shape of the points corresponds to the timepoint when the sample was collected. PC1 value and color fraction identity are significantly associated (linear regression p-value << 0.001). In (C) samples are colored by treatment. PC2 value and treatment are significantly associated (ANOVA p-value << 0.001).

Ancestry-informed mapping provides resolved architecture of abdominal pigmentation

Given the substantial variation in pigmentation in our study population (Fig. 2B), we next aimed to map the underlying genetic architecture of this trait. We initially focused mapping on samples collected at week 8 in the outdoor mesocosm, as these replicates had undergone evolution in a semi-natural environment and also experienced more generations of recombination since initial outbreeding than the week 2 samples, and were thus predicted to generate better resolved genetic maps. Our general approach to mapping was to identify SNPs harboring alleles with systematic frequency changes across the light, midpoint, and dark trait value groups. However, given the finite number of haplotypes from the founding inbred lines used to derive our outbred population, a significant association between a SNP and trait value group could indicate either its linkage to a locally causal allele and/or its broader inbred-line ancestry. Accordingly, to infer the magnitude of the inbred-line ancestry genome-wide, we quantified the average contribution of haplotypes coming from each inbred line to each trait value group and tested for the association of this quantity with the pigmentation score of that inbred line. The results of this are depicted in Figure 3A, where we observe that the light trait value group contains a disproportionate amount of ancestry from the lightest inbred lines and, likewise, the dark trait value group contains a similarly disproportionate amount of ancestry from the darkest inbred lines ($R^2 = 0.032$; $p < 0.01$). To further quantify this ancestry effect on pigmentation, we a ran a simple regression (GLM; logistic link function and quasibinomial error variance) of allele frequencies on trait value group. We took the top 1% of the SNPs (based on nominal p-value from the GLM) and calculated their expected frequency difference between light and dark color fractions based on estimates of average contribution of each inbred line to that sample, as well as the association between that SNP and its presence in the inbred line (Fig 3B). Indeed, we see that the GLM-identified light-associated SNPs tend to be associated with the lighter pigmented lines, making it hard to disentangle the SNP's local association with a causal locus from its association with global line ancestry.

To account for the strong ancestry effects on SNP-pigmentation association (Fig. 3A-B), we ran an ancestry-informed GLM. Specifically, allele frequencies were regressed both on pigmentation trait value and the SNP's association with each inbred line and its genome-wide relative contribution to each trait value group (light, midpoint, or dark). As expected, in Fig. 3B we see that the distribution of expected light-dark frequency differences for SNPs identified by this model is now centered around zero, indicating that the global bias produced by our trait-only GLM is eliminated. We further validated that SNPs identified via this approach were not exclusive to the lightest nor darkest founding inbred lines (colored points, Fig. 2A), but rather present at appreciable frequency across all founding haplotypes (Fig. S4). We then defined a robust set of pigmentation-associated SNPs using permutations of the trait value data and re-running our ancestry-informed GLM (N = 100 permutations). A SNP was considered

245 significant if its nominal p-value from the empirical GLM fell in the top 0.05% of permuted p-values
246 (corresponding to an empirical FDR threshold < 0.0005; Fig. S5). Even at this conservative threshold, we
247 see that trait-associated SNPs are pervasive throughout the genome and distributed across all
248 chromosomal arms (Fig 3C). Hereafter, we refer to pigmentation-associated SNPs identified via this
249 approach as either light or dark-associated given their relative frequency in the light and dark color
250 fractions: the light-associated allele for each mapped SNP is that occurring at higher relative frequency in
251 the light color fraction.

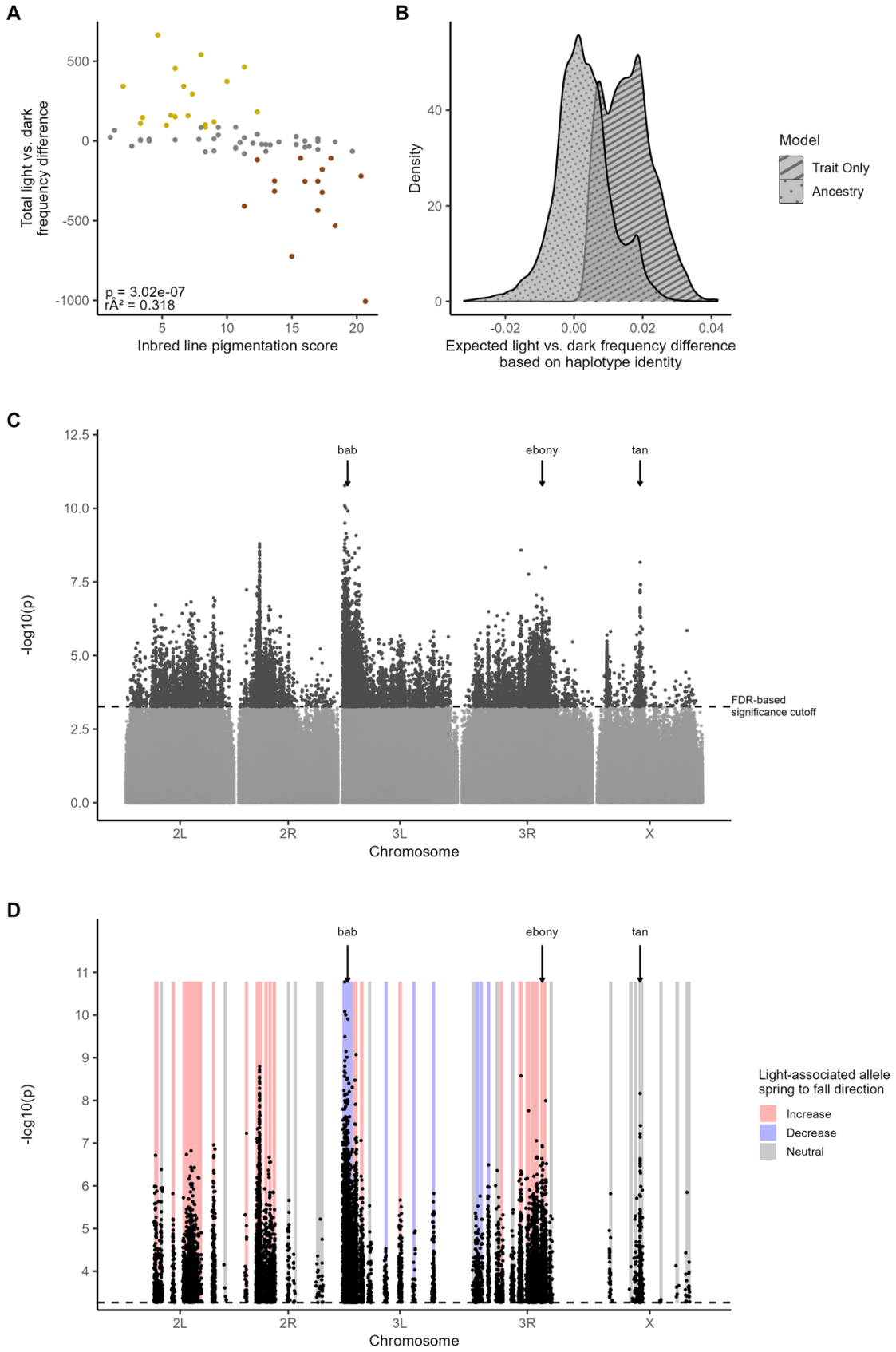
252 We next used our set of pigmentation-associated SNPs throughout the genome to identify the
253 number and location of unlinked loci underpinning variation in the trait. To do so, we first identified the
254 most significant SNP on each chromosomal arm, as well as its surrounding, locally un-recombined region
255 (defined by the recombination rate and number of generations since cage founding, see Methods). We
256 then eliminated all trait-associated SNPs on the chromosome whose relative presence in the light and dark
257 fractions can be explained by haplotype frequencies in the locally un-recombined focal region, in effect
258 eliminating all signal due to linkage at the focal SNP location. From the set of remaining SNPs, we
259 identified the SNP with the most significant nominal p-value and repeated the procedure. This was
260 iterated until we exhausted all trait-associated SNPs on the chromosome and resulted in a total of 76
261 genomic loci, the midpoint of which was the most significant, locus-defining SNP. Merging loci with
262 overlapping start and endpoints, we obtained a final list of 42 loci, ranging in size from 700 Kb to 3 Mb,
263 and distributed across all chromosomal arms (Supplementary Data File 1). This indicates that the genetic
264 architecture of our focal, ecologically relevant complex trait, is not underpinned solely by a single or few
265 large-effect loci, nor exclusively by a very large number of weak-effect loci (e.g., 1000's). Rather, we
266 infer the presence of a moderate number of large to moderate-effect loci underpinning trait variation,
267 while acknowledging there may be further loci of small effect that our approach lacks sufficient power to
268 detect. We once again validated that pigmentation-associated SNPs within the largest-effect loci were not
269 exclusive to those lightest/darkest haplotypes enriched in the trait tails (Fig. 3A), but rather shared across
270 all founding inbred lines (Fig. S6).

271 We quantified the extent to which our identified loci validated previously described, and
272 potentially identified new, candidate genes associated with pigmentation variation in the species.
273 Assuredly, the top loci on chromosomal arms 3L, 3R, and X contained the previously described core
274 genes impacting pigmentation variation within the species: *bab*, *ebony*, and *tan* (Bickel et al., 2011; Jeong
275 et al., 2008; Massey and Wittkopp, 2016; Pool and Aquadro, 2007). A comparison of the association
276 signal between loci identified here and two recent, well-powered association studies of the trait (Bastide
277 et al., 2013; Dembeck et al., 2015) indicated significantly more overlap than expected by chance
278 (permutation $p < 0.001$, see Methods). However, this signal is eliminated when we no longer consider

SNPs within our mapped loci containing *bab*, *ebony*, and *tan* ($p = 0.994$) suggesting little overlap outside of these three core genes. Thus, the extensive enrichment of pigmentation-associated variation outside these candidate loci we observe in our data suggests the presence of novel candidate pigmentation loci identified in our study, which we discuss in the supplement (Supplementary Discussion; Table S1-2).

Pleiotropy and predictions of complex trait evolution

A potential of application of a well-resolved genetic map is to predict whether or not a trait has evolved using genomic data alone. However, this effort could be obscured by underlying pleiotropy at quantitative trait loci, which may constrain their impact on trait evolution and leave the architecture of trait adaptation largely unpredictable. Such pleiotropic interactions may be particularly pronounced in complex environments where selection across multiple fitness-relevant traits may induce correlated selection on pleiotropic loci. Our longitudinal study design afforded the unique opportunity to test this directly by combining our well-resolved genetic mapping data (Fig. 2C-D) with our time-series allele frequency data collected throughout phenotypic evolution of the trait (Fig. 1; 2A). Accordingly, for each of the 42 independent, unlinked loci identified above, we quantified the frequency shifts of the light-associated alleles as the population itself evolved to become significantly lighter in the outdoor mesocosm (Fig. 2A). We then inferred whether observed allele frequency movement within each locus indicated lighter pigmentation evolution by comparing its behavior to a distribution derived from a set of matched control SNPs (see Methods). We found that out of the 42 loci, 18 exhibited an overall increase in light-associated allele frequency and thus a behavior concordant with expectations based on the evolution of lighter pigmentation (i.e., overall increase in frequency of light-associated alleles). For 17 pigmentation-associated loci, however, we did not detect any systematic shift indicative of trait evolution, while 7 loci showed behavior in the direction of darker pigmentation (Fig. 3D). Thus, in sum, while our genome-wide inference correctly predicted evolution of lighter pigmentation in the outdoor mesocosm ($p << 0.001$; Fig. 4A) we observed that at particular loci, and even for entire chromosomal arms (Fig. S7), the genetic basis of pigmentation evolution was unpredictable based on mapping data alone. This likely reflects the role of pleiotropy and constraint during complex trait adaptation in natural settings characterized by a high diversity of selection pressures, such as our outdoor mesocosm (Fig. 1, Fig. 3D. 3D). This unpredictable genetic architecture of adaptation was also evident at the known loci of major effect: allele frequency shifts within the *ebony* locus, as well as the largest effect locus on 2R, predicted evolution of lighter pigmentation; while the locus containing *bab* predicted evolution of darker pigmentation and the *tan* locus predicted that pigmentation would not evolve (arrows in Fig. 3D).


A
Total light vs. dark frequency difference
500
0
-500
-1000
p = 3.02e-07
rÂ² = 0.318
5
10
15
20
Inbred line pigmentation score
B
Density
40
20
0
-0.02
0.00
0.02
0.04
Expected light vs. dark frequency difference based on haplotype identity
Model
Trait Only
Ancestry
C
-log10(p)
12.5
10.0
7.5
5.0
2.5
0.0
bab
ebony
tan
FDR-based significance cutoff
2L
2R
3L
3R
X
Chromosome
D
-log10(p)
11
10
9
8
7
6
5
4
bab
ebony
tan
2L
2R
3L
3R
X
Chromosome
Light-associated allele spring to fall direction
Increase
Decrease
Neutral

313

Figure 3 | Genetic architecture of female abdominal pigmentation. (A) The pigmentation score of the founding inbred lines compared with the light vs. dark frequency differences across all samples for the haplotypes originating from these lines. Inbred line pigmentation is highly correlated to the haplotype's light vs. dark enrichment (linear regression $p << 0.001$). Colored points indicate inbred lines that drive the significance of relationship between pigmentation score and enrichment in color fractions (gold points correspond to inbred lines enriched in light color fraction and dark brown to lines enriched in dark color fractions) (B) Distribution of the difference between average expected light frequencies and average expected dark frequencies for the top 1% of SNPs from a model only considering the impact of trait value on allele frequency variation (Trait-Only), versus a model with an additional term that represents the expected frequency in the sample based on the SNPs' ancestry (Ancestry). (C) Manhattan plot depicting nominal p-values derived from an ancestry-informed mapping of SNP association with pigmentation variation. The dashed horizontal line corresponds to an empirical FDR threshold of 0.0005, derived via permutations. (D) Genomic distribution of SNPs confined within N = 42 unlinked, independent loci underpinning pigmentation variation. SNPs are plotted as a function of their genomic coordinate and regression-derived nominal p-value (as in (C)). Shaded regions denote the stop and end points of each of the identified unlinked, independent loci. Shaded region color then denotes the predominant behavior of light-associated SNPs within each locus through time: red shading indicates loci with light-associated alleles favored throughout evolution of lighter pigmentation (18 total loci; binomial test, FDR < 0.05), blue shading indicates loci in which the light-associated alleles are predominately selected against (7 total loci; binomial test, FDR < 0.05), and grey shading indicates loci with no signature of selection on light-associated alleles (17 total loci; binomial test, FDR > 0.05). The location of the canonical pigmentation genes, *bric-a-brac (bab), ebony,* and *tan*, are indicated on the Manhattan plots. Potential candidate genes underlying the additional loci observed throughout the genome are discussed in the Supplement.

The effect of pleiotropy on genomic prediction differs across complex and simple environments

The pleiotropy we quantified above manifested in a field setting where replicate populations were exposed to a dynamic suite of fluctuating selection pressures. This begs the question of whether the magnitude of pleiotropy would be subdued in a simpler, lab-based environment with fewer selective pressures. Our study design provided a unique opportunity to test this hypothesis via comparing genomic predictions of trait evolution for the replicate populations evolved in the field mesocosm relative to those that evolved within the indoor, lab-based mesocosm (Figs. 1-2). We first quantified whether light-associated alleles were favored during the evolution of lighter pigmentation, finding that their genome-wide frequencies increased significantly throughout trait evolution in each environment, relative to background allele frequency change (paired t-test p-values << 0.001; Fig. 4A; also see Fig. S8-S9). Next, we used the relative effect sizes produced by our mapping data, as well as their magnitude of allele frequency during trait evolution, to provide a genome-wide summary statistic of pigmentation evolution in each environment (Fig. 2A, Methods). This analysis is analogous to the polygenic score-based predictions of trait evolution that are pervasive in studies of human adaptation (Akbari et al., 2024; Cox et

350 al., 2019; Mathieson et al., 2015), and correctly inferred that lighter pigmentation evolved in both the
351 indoor and outdoor environment (Fig. 4B). Interestingly, however, the relative predicted magnitude of
352 phenotypic change was greater in the indoor environment, despite our phenotypic assays showing that
353 replicate populations in this environment evolved to a lesser degree (Fig 2A). Such a disconnect between
354 observed patterns of phenotypic evolution and our genomic inference of the process provides strong
355 evidence of a context-dependent pleiotropy across environments – the more complex environmental
356 conditions in the outdoor mesocosm induces selection on a greater number of traits overall, leading to
357 more pleiotropic constraint at pigmentation-associated loci and poorer genomic prediction.
358 We investigated context-specific pleiotropy further with a focus on the *bab* locus, a canonical
359 pigmentation gene identified by our mapping and for which raw allele frequency trajectories showed
360 evidence of pleiotropic constraint (Fig. 3D). By excluding variation in this locus from our genomic
361 prediction, we observed that the inferred relative magnitude of pigmentation evolution increased in both
362 environments and, strikingly, to a much greater degree in the outdoor environment (Fig 4C). This analysis
363 thus provides a key example of how pleiotropy at a single locus can skew genomic predictions of
364 phenotypic evolution. Furthermore, it elucidates how studying such processes in a natural setting may
365 reveal properties of segregating variation that are masked in simpler, lab-based environments.
366 Interestingly, regenerating these predictions using variation within the canonical pigmentation loci alone
367 led to a spurious prediction of darker evolution in the outdoor environment (Fig. S10), while excluding
368 variation in these loci from our prediction recovered the correct inference of trait evolution (Fig. S11).
369 Thus, the key takeaway of these results is that accurate genomic prediction of trait mapping hinges upon a
370 comprehensive, well-resolved genetic map that includes both the largest-effect loci, as well as those of
371 more moderate or minor impact on trait variation.
372 As our genomic predictions of trait evolution were generated from mapping results derived from
373 samples evolved in the outdoor mesocosm, we next quantified if and how these predictions differed when
374 derived from mapping of the indoor populations sampled at week 8. Using the same mapping procedure,
375 we found that while this independent set of mapped loci correctly predicts trait evolution in the indoor
376 environment, it fails to predict evolution of lighter pigmentation in the outdoor environment (Fig. 4D).
377 This discrepancy in genomic prediction of trait evolution in the outdoor environment (i.e., comparing
378 magenta distributions in Fig. 4B vs. Fig. 4C) potentially indicates epistasis-driven differences in the
379 genotype-phenotype map that arise between environments of selection. Strikingly, but also consistent
380 with those genomic predictions generated using the outdoor mapping data (Fig. 4B-C), we found that the
381 exclusion of *bab* locus variation allowed us to correctly predict trait evolution in the outdoor environment
382 using the indoor mapping data (Fig. 4E). This further exemplifies the pleiotropic constraints of this large
383 effect locus on trait evolution and the consequences therein on genomic prediction.

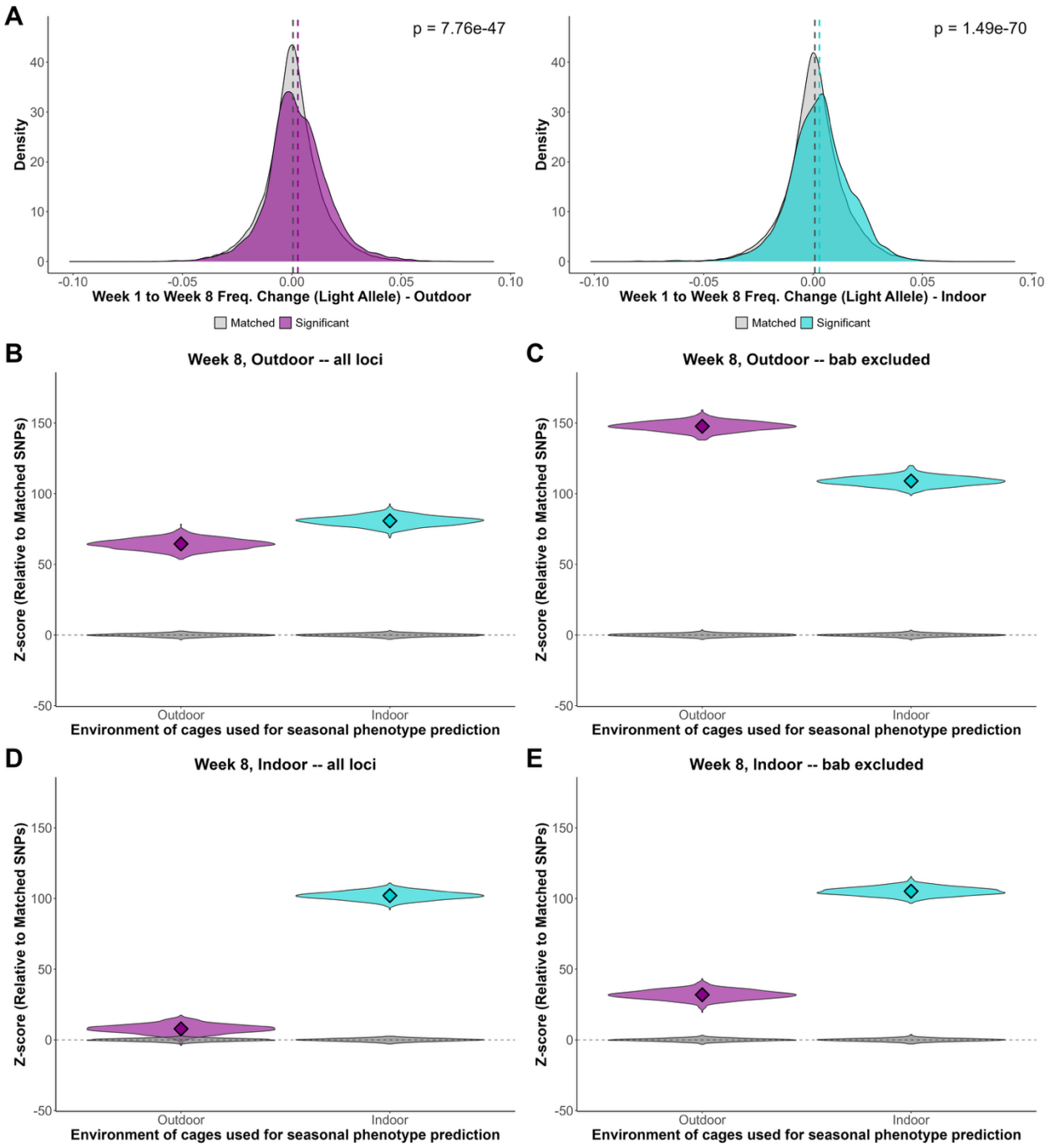


Figure 4 | Prediction of phenotypic evolution in simple and complex environments. (A) Genome-wide allele frequency change at pigmentation-associated SNPs identified in the outdoor, week 8 mapping population and quantified throughout trait evolution in the outdoor (purple) and indoor (blue) mesocosm, relative to matched control shifts (grey distributions). (B) Distribution of Z-scores for bootstrapped estimates of genomic predictions for pigmentation evolution in the outdoor (magenta) and indoor (turquoise) mesocosm using SNPs identified in the outdoor, week 8 mesocosm. Positive Z-scores indicate net lighter pigmentation evolution while negative values would indicate net darker evolution. Z-scores were computed by comparing inferred relative phenotypic change

using mapped SNPs to a that computed using a set of matched control SNPs (grey distributions). (C) Same as in panel (D), but genomic predictions discard pigmentation-associated SNPs within the *bab* locus. (E-F) Same as (D-E) but for pigmentation-associated SNPs identified via mapping the week 8 indoor populations.

Stable genotype-phenotype mapping throughout evolution and across environments of selection

How well do genotype-phenotype maps derived from a single population, at a single point in time, transfer, and allow phenotypic prediction, across genetically diverged populations? This represents another fundamental question in complex trait genetics and evolution that we could address within the context of our longitudinal, multi-environment study. Specifically, we observed a pronounced signature of genome-wide divergence associated between our lab-based, indoor environment and the outdoor, field mesocosm (Fig. 2B), and previously reported substantial evolutionary change within each of these environments throughout the observation period monitored in this study (Bitter et al., 2024a, 2024b). One implication of this environmental and temporal genomic divergence is that it may induce shifts in the epistatic landscape and/or elicit context dependent-dominance of pigmentation-associated loci. In effect, these processes may alter the marginal effect of pigmentation-associated loci and generate distinct genotype-phenotype maps that preclude phenotypic prediction across mapping populations (Novembre and Barton, 2018).

We evaluated the stability of the genotype-phenotype map by re-implementing our mapping approach, as described above, on the replicate populations sampled in outdoor mesocosms at week 2, as well as the indoor populations sampled at weeks 2 and 8. The number of loci underpinning variation across all mapping populations was comparable (ranging from 30-43 independent loci, Supplementary Data File 1; Table S2). Furthermore, as expected, the total genomic distance spanned by these loci declined from weeks 2 to 8 (from ~1 Mb at week 2 to ~700 Kb at week 8) due to additional recombination events that allowed for more localized mapping.

The similarities in architecture inferred in the lab and field environments at week 8 are clearly apparent in the mirrored Manhattan plots presented in Fig. 5A, where regions of heightened pigmentation-associated variation are replicable across mapping environments. Notably, loci identified across all four mapping populations contained the canonical pigmentation genes *tan*, *ebony*, and *bab* (Fig. S12), demonstrating consistency of mapping as well as the pronounced effect these genes have on pigmentation variation in this species. We next took two approaches to rigorously quantify, genome-wide, the similarity in loci harboring pigmentation-associated variation across the four independent mapping populations. First, we assessed whether the start and end points of loci identified in each independent mapping population were closer than expected by chance: for each set of loci identified within a particular mapping population, we selected the locus-defining, midpoint SNP (i.e., that with the most significant p-value) and

425 then quantified the number of times it fell within a locus in other three sets of mapped loci (each locus
426 could therefore overlap 0 to 3 times with the remaining mapping populations). The number of overlaps
427 across all loci for a particular mapping population were then summed and compared to an empirical null
428 distribution derived by iterating this procedure N = 2000 times using a set of randomly selected SNPs (see
429 Methods). Across sets of mapped loci, midpoint SNPs were far more likely to fall within a locus of an
430 independent mapping than expected by chance (Fig. 5B), demonstrating that the regions harboring
431 pigmentation-associated variation were conserved across mapping populations. To probe this consistency
432 further, we quantified the distance between locus midpoints among mapping populations. As expected,
433 these distances were significantly closer than expected based on an empirical null (Fig. 5C, see Methods),
434 further indicating the conservation of pigmentation associated loci across mapping populations.
435 Given the pronounced mapping stability at the locus-level, we next interrogated whether
436 pigmentation-associated SNPs were also shared across mapping populations. To do so, we reidentified
437 trait-associated SNPs using our ancestry-informed GLM in groups of 5 (of the 10 total) replicates per
438 mapping population. This then allowed for a balanced comparison of the overlap in trait-associated SNPs
439 within subsets of the same mapping population (e.g., overlap in trait associated SNPs between subsets of
440 five replicates from the outdoor week 8 samples), and between mapping populations (e.g., overlap in trait
441 associated SNPs between subsets of five outdoor week 8 and five indoor week 2 replicates). Strikingly,
442 we found that the amount of overlap in SNPs identified between subsets of the same mapping population
443 did not exceed that between mapping populations, either from the same environment and a different time
444 point or even between environments (Fig. 5D; Fig. S13). In the supplement, we further quantified this
445 SNP-level parallelism using a leave-one-out cross validation scheme (Fig. S14).
446 While our quantification of overlap in pigmentation-associated loci across mapping populations
447 indicated a strong stability of the genotype-phenotype map (Fig. 5B-D), we observed several conspicuous
448 instances in which genomic regions harboring pigmentation-relevant variation for one mapping
449 population appeared to lack pigmentation-associated variation in an independent mapping population.
450 Several of such instances can be observed in the mirrored Manhattan plot in Fig. 5A, where the start and
451 end points of pigmentation loci (indicated by shaded regions) within one environment exhibit non-
452 overlapping start and endpoints with any of the loci identified in the second environment. We suspect
453 these environment-specific loci to potentially indicate epistatic interactions driving differences in the
454 marginal effect of pigmentation-associated SNPs across mapping populations. In the Supplement, we
455 show that such instances of environment-specific loci are not solely due to technical noise, but rather
456 treatment-specific marginal effects of pigmentation-associated loci and variation (Fig. S15 & S16), as
457 first indicated by our genomic prediction results presented above (Fig. 4). Nonetheless, the vast majority
458 of loci are overlapping across populations (e.g., 30 of 43 loci identified in the week 8 outdoor population

overlapped with at least one of the 42 loci in the week 8 indoor population), and thus the dominant signal in the data is one of strong parallelism. Indeed, when we exclude treatment-specific loci from our calculations of predicted phenotypic evolution, our ability to correctly predict trait change across environments is unaffected (Fig. S17).

Finally, given the overall stability of SNP and locus mapping, we queried our ability to predict relative phenotypic values across independent populations using allele frequency data alone. Specifically, we inferred the relative pigmentation of phenotypic groups (i.e., the light, midpoint, and dark color fractions) both within and across mapping populations using our mapping data (see Methods). Strikingly, we found that we predicted relative pigmentation scores with statistically indistinguishable accuracy between, as well as within, mapping populations (Fig 5E; Fig. S18). In concert, these analyses suggest that epistasis or dominance shifts caused by the genetic evolutionary change in our experiment were insufficient to preclude genotype-phenotype map portability and phenotypic prediction. While seemingly at odds with expectations and empirical evidence of limited portability of mapping results across independent populations (de Bakker et al., 2006; Mostafavi et al., 2020; Privé et al., 2022), it bolsters more recent research that have demonstrated increased portability under instances in which mapping is conducted in a high-powered manner and the loci underpinning trait variation are fine mapped (Hu et al., 2025; Wu et al., 2025)

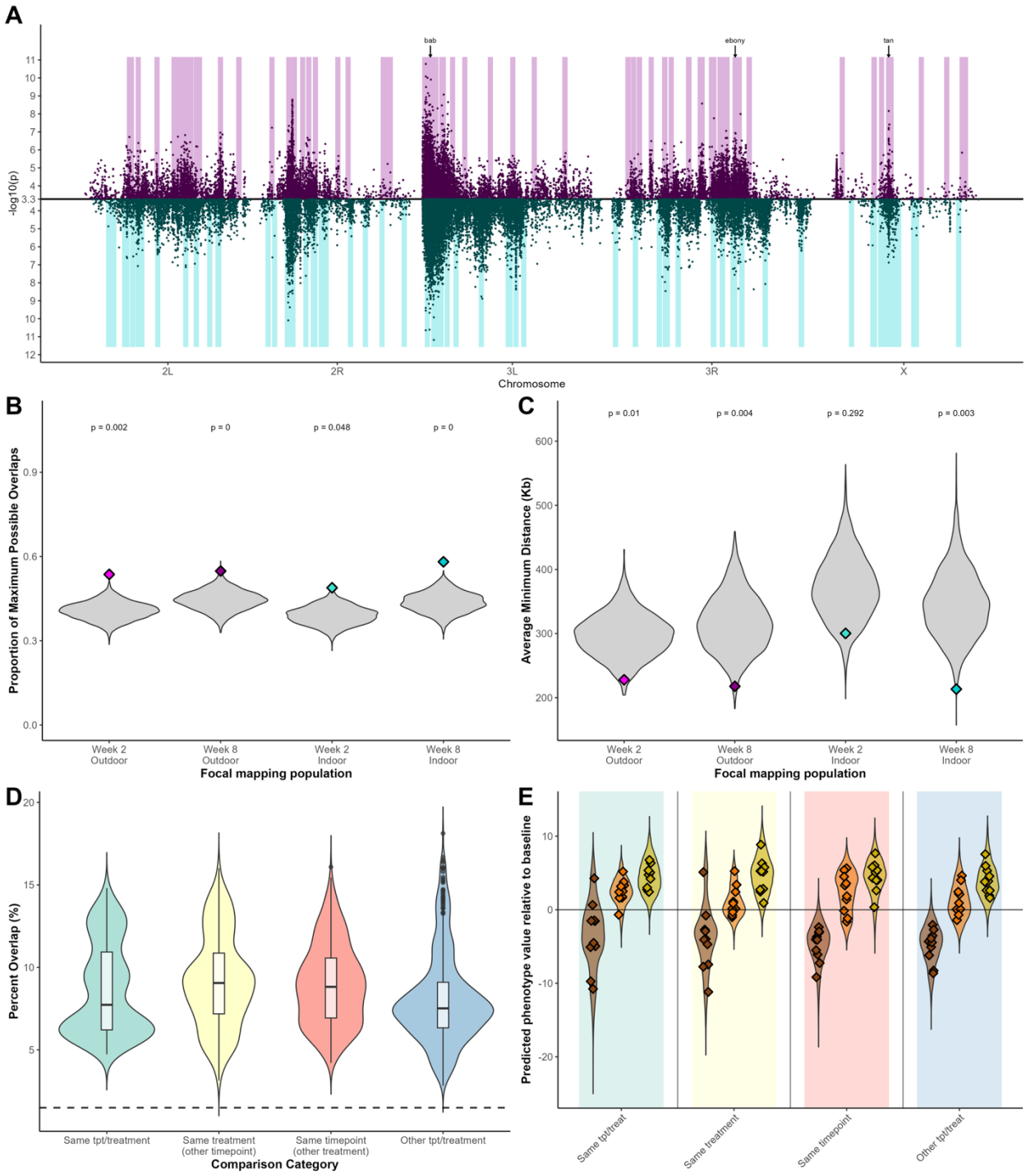


Figure 5 | Epistasis does not preclude portability of trait mapping. (A) Mirrored Manhattan plots of genetic architecture quantified at week 8 in the outdoor (top) and indoor (bottom) environments. Only SNPs with p-value corresponding to empirical FDR < 0 .0005 p-value are depicted. For each mapping, shaded regions correspond to unlinked loci underpinning pigmentation-associated variation (B) Observed (diamonds) vs expected (grey violin distributions) overlap between pigmentation associated loci across mapping populations. (C) Observed (diamonds)

vs expected (grey violin distributions) distances between midpoint SNPs of trait associated loci between mapping populations. (D) Observed percent overlap of trait-associated SNPs identified within and between mapping populations. (E) Predicted relative phenotypic values for dark, midpoint, and light color fractions using genotype-phenotype mapping conducted in the outdoor mesocosm at week 8 where positive and negative values indicate lighter or darker pigmentation relative to the baseline population, respectively. Predictions (diamonds) were conducted for each trait value group and replicate cage. Colored violins represent 1000 bootstrapped predictions of each predicted, relative phenotypic value.

Discussion

The central goal of this study was to rigorously evaluate the predictability of complex trait adaptation. Specifically, we aimed to quantify the extent to which we could correctly infer complex trait evolution given genomic data alone, and also predict which loci came under selection during trait evolution. Leveraging a model complex trait (abdominal pigmentation), a model species (*Drosophila melanogaster)*, and a highly powered, longitudinal study design, we showed that well-resolved genotype-phenotype maps can be used to correctly predict if, and in what direction, a complex trait evolves. However, the specific genetic architecture of adaptation (i.e. which loci show signatures of selection during trait evolution) was largely unpredictable based on mapping data alone. Specifically, despite a genome-wide increase in the frequency of light-associated alleles, pigmentation-associated variation in several loci was observed to be net-neutral, or even selected against in a 'counter-gradient' manner, throughout the evolution of lighter pigmentation (Albecker et al., 2025; Paaby et al., 2014). This dynamic exemplifies underlying pleiotropy in complex trait variation which, under a regime of correlated selection, can ultimately constrain such variation's impact on trait evolution (Walsh and Blows, 2009). It is unsurprising that this dynamic emerged in our outdoor mesocosm, as we have shown that multiple fitness-relevant traits to evolve across seasons in this system (Rudman et al., 2022).

A key additional component of this study was to assess if, and how, pleiotropic constraint differed across distinct environments of selection. Specifically, we paired our outdoor mesocosm with an indoor mesocosm in which the same founding population evolved concurrently in the context of the same general biotic shifts (e.g., increases in population density), but the absence of the multivariate abiotic fluctuations that characterize the outdoor habitat (Fig. 1). In effect, we showed that the pleiotropic constraint of several large-effect pigmentation-associated loci (e.g., *bab*) was alleviated in the simpler, lab-based environment, leading to better overall genomic predictions of trait evolution. This finding thus exemplifies a context-dependence of pleiotropy, whereby more complex environments elicit an increase in pleiotropic constraint. Furthermore, from a practical sense, this implies that generic principles of the evolutionary process itself may be obscured in simplified, lab-based environments and bolster recent calls to study the ecology and evolution of natural populations in realistic settings (Bergelson et al., 2021).

While the impact of pleiotropy was pronounced, we found epistasis to have a less detrimental effect on the prediction of complex trait variation overall. Specifically, we found a striking stability of the genotype-phenotype map across four independent mapping populations sampled at two points in the evolutionary process and evolved in two distinct environments. Furthermore, we showed that relative phenotypic predictions were effectively as accurate across environments of selection, as within. Therefore, at least within the confines of the magnitude and timescale of evolution manipulated here, epistasis (and/or context-dependent dominance shifts) may not inhibit a general transferability of mapping results across genetically diverged populations. While this finding may be at odds with some studies evaluating the portability of mapping data across diverged populations, it bolsters more recent work demonstrating increased portability under instances in which trait architecture has been well-resolved (i.e., fine-mapped) such as in the high-powered mapping conducted here (de Bakker et al., 2006; Hu et al., 2025; Mostafavi et al., 2020; Privé et al., 2022; Wu et al., 2025)

It warrants note that the unprecedented insight into the predictability of evolution was only made possible by several key aspects of our methodological approach: our generation of haplotype-informed allele frequencies at extremely high coverage, in combination with an exceptional amount of biological replicates per mapping treatment and time point, provided the power to produce well-resolved genetic maps (Long et al., 2025; Tilk et al., 2019). Furthermore, our longitudinal analysis of trait variation using a single founding population was the only way to exclude confounding variables such as unique patterns of demographic history and linkage when assessing the role of pleiotropy and epistasis on complex trait prediction (Lohmueller, 2014; Novembre and Barton, 2018). In concert, this study design then allowed us to both reaffirm canonical pigmentation genes, as well as identify new candidates, underpinning variation in this classic complex trait. Crucially, our accurate genomic prediction of trait evolution hinged upon this well-resolved genotype-phenotype map, as prediction using canonical gene variation alone would have led to incorrect inferences of trait evolution. Our study therefore exemplifies the critical roles of study design and power in the rigorous evaluation of fundamental evolutionary dynamics.

Looking forward, our results are pertinent to numerous applications of genomic data: predicting phenotype across disparate populations, inferring past selection using time-series and ancient DNA genomic data, or manipulating contemporary variation to shape future evolutionary trajectories (e.g., augmenting population resilience to future global change conditions). The critical contingency of effectively implementing such approaches, however, will be the procurement of well-resolved genotype-phenotype maps and knowledge of the focal trait variation's pleiotropy and epistasis, a venture that may not be feasible across study systems.

565 **Methods**

566 *Experimental overview*

567 A genetically diverse population of outbred *D. melanogaster* was generated via outcrossing ten
568 male and female individuals from a total of 76 inbred lines, originally collected from Linvilla Orchards,
569 Media, PA (Behrman et al., 2018). The outbred population was allowed to recombine and expand for 4
570 generations, after which 500 males and 500 females were used to seed each of 10 outdoor and indoor
571 replicate cages. The outdoor mesocosm consisted of ten replicate, two m^3 mesh cages, each harboring a
572 single dwarf peach tree, and was exposed to natural environmental conditions, insect, and microbial
573 communities (Bitter et al., 2024a). The indoor mesocosm then consisted of ten replicate, ~0.5 m^3 cages
574 housed in a temperature-controlled laboratory at the University of Pennsylvania. Constant food was
575 supplied to the replicates within each environment, whereby 400 ml of Drosophila media was provided in
576 900 cm^3 aluminum loaf pans within each cage three times per week (Bitter et al., 2024a, 2024b).
577 Monitoring of phenotypic and genomic variation in the replicates began two weeks after cage founding
578 and continued for a period of eight total weeks and six generations of evolution. Generation time
579 estimates were produced using a degree day model (Kamiyama et al., 2020).

580 Population-wide pigmentation variation was quantified in our panel of inbred lines, the
581 recombinant outbred population after 4 generations of expansion and recombination, and the evolving
582 replicate populations in the outdoor and indoor environments at weeks 2 and 8 (20 July and 1 September,
583 separated by a period of 5 generations). To quantify evolved changes in pigmentation, eggs were
584 collected overnight from all replicate cages (indoor/lab and outdoor/field), after which they were
585 transferred to common garden conditions (25˚C, 12L:12D) where they were reared for 2 generations.
586 Following density control of F2 eggs (30 +/-5) in replicate vials, adult female abdominal pigmentation (N
587 = 30 individuals per replicate) was scored as the proportion of pigmentation observed on tergites 5, 6, and
588 7 (David et al., 1990). The sum of these scores across the three tergites provided the total pigmentation
589 score for the assayed individual.

590 We quantified genome-wide allele frequencies throughout evolution weekly in the outdoor
591 environment and at weeks 2 and 8 in the indoor environment. At each collection point, eggs were
592 collected overnight from each cage and then transferred to common garden conditions where larval
593 development took place. 100 random females from the eclosed adult cohort were then collected and
594 stored in ethanol at -20ºC for later DNA extraction and sequencing (see below).

595 At the weeks two and eight sampling points, we conducted a tail-based mapping of pigmentation
596 variation for each mapping population. Specifically, we first quantified pigmentation variation within
597 each treatment using a random sample of sorted, 100 F2 females (3-6 days post-eclosure) and the visual
598 scoring methods described above (David et al., 1990). Then, for each replicate across both environments,

599 we segregated individuals into one of three trait value groups: (1) those exhibiting pigmentation within
600 the upper 15% of the population-wide phenotypic distribution, (2) those exhibiting pigmentation within
601 the lower 15% of the trait distribution and (3) a random sample from those representative of the midpoint
602 (~25th-75th percentile) of the trait distribution. Seventy-five individuals from each trait value group and
603 replicate were then stored in ethanol and at -20ºC for later DNA extraction and sequencing. In total this
604 resulted in 120 samples available for mapping: three trait value groups across ten replicate cages sampled
605 from two environments and at two time points.

606 *Sequencing and allele frequency estimation*

607 Genomic DNA was extracted for both mesocosm time-series and trait mapping samples using
608 NEB Monarch DNA kit. Following sample QC via nanodrop and qubit, whole genome libraries were
609 constructed using Illumina DNA prep kit. Each sample was sequenced to an equal coverage of 8x using
610 Illumina Novaseq 6000. From the resulting sequencing data, we generated haplotype-informed allele
611 frequencies using previously described methods (Kessner et al., 2013; Tilk et al., 2019). This method
612 computes allele frequencies at sites previously identified in the genome sequences of each founding
613 inbred strain using an expectation-maximization algorithm that estimates haplotype frequencies in a
614 pooled sample of mapped sequencing reads. Haplotype inference was conducted in window sizes that
615 varied proportionally to the length of un-recombined haplotype blocks expected as a function of the
616 estimated number of generations since original inbred line outbreeding (Kessner et al., 2013). After
617 filtering sites an average minor allele frequency >= 0.02 in the founder population, and present in at least
618 one evolved sample at a minor allele frequency > =0.01, we retained 1.9 M SNPs for analysis.

619 *Genome-wide patterns of variation across environments and throughout evolution*

620 We first assessed genome-wide patterns of variation across all samples used in trait mapping
621 using principal component analysis (*prcomp* package, R) and normalized allele frequency data across all
622 1.9 M SNPs. We projected samples onto the first two principal components and visualized the data by
623 shaping and coloring points according to different underlying variables (e.g., trait value group, treatment)
624 to infer which variables may dominate variance in allele frequencies across samples. We quantified
625 whether samples segregated along the first two principal components in accordance with trait value group
626 or environment using linear regression and ANOVA, respectively.

627 *Ancestry-informed association analysis of pigmentation variation*

628 Our general approach to mapping pigmentation was to identify SNPs harboring alleles with
629 systematic frequency changes across the light, midpoint, and dark trait value groups. First, however, it
630 was important to consider that due to the finite number of haplotypes from the founding inbred lines used
631 to derive our outbred population, a significant association between a SNP and trait value group could
632 indicate its biological impact on pigmentation variation and/or its inbred-line ancestry. This effect could

633 appear at the level of entire chromosomes due to long range patterns of linkage, as well as more locally
634 around causal loci. We quantified the magnitude of this effect genome-wide by computing the average
635 contribution of haplotypes coming from each inbred line to each trait value group (i.e., the difference in
636 haplotype frequency between the light and dark tails) and quantified the relationship of this value with the
637 pigmentation score of each inbred line using linear regression and Pearson's correlation coefficient.

638 We interrogated the effect of linkage on SNP mapping further by running two generalized linear
639 models (GLM; logistic link function and quasibinomial error variance). The first model only took into
640 consideration the impact of trait value group on variation in allele frequencies (model form: allele
641 frequency ~ trait value; where allele frequencies are normalized by the SNP's effective coverage and the
642 total number of chromosomes sampled). If the effects of ancestry on SNP association were pervasive,
643 SNPs identified via this approach would be enriched in those SNPs expected to have elevated light to
644 dark frequency differences based on their association with particular haplotypes alone. To test this, we
645 selected the top 1% of the SNPs (based on nominal p-value from the GLM) and calculated their expected
646 frequency difference between light and dark color fractions based on estimates of the average contribution
647 of each inbred line to that sample and the association between that SNP and inbred line. We then
648 compared these results to a separate GLM model in which the haplotypic ancestry was accounted for.
649 This model then regressed allele frequencies both on pigmentation trait value, as well as each SNPs
650 association with the line contribution to the sample using average genome-wide haplotype frequencies. In
651 accordance with the previous, ancestry-agnostic, model we used a logistic link function and
652 quasibinomial error variance, and allele frequencies were weighted to account for variation in coverage
653 and the finite number of chromosomes sampled. We then compared the density distributions of expected
654 light-dark frequency differences for SNPs identified via each of these distinct models using a KS test.

655 Following GLM model structure selection, we used permutations to define a final set of trait-
656 associated SNPs. Specifically, we randomly shuffled trait value data and re-ran our regressions N = 100
657 total times to generate an empirical null distribution. We then considered a SNP significant if its nominal
658 p-value from the empirical GLM fell in the top 0.05% of its permuted p-values (corresponding to an
659 empirical FDR threshold < 0.0005).

660 The set of pigmentation-associated SNPs identified via GLM was used to quantify the number
661 and distribution of unlinked loci underpinning trait variation. We developed an iterative procedure for
662 this, whereby we first identified the most significant SNP on each chromosomal arm, as well as its
663 surrounding, locally un-recombined region (defined by the recombination rate and number of generations
664 since cage founding; Tilk *et al*. 2019). We then eliminated all trait-associated SNPs on the chromosomal
665 arm (i.e., those with a significant GLM p-value) whose relative presence in the light and dark fractions
666 can be explained by haplotype frequencies (i.e., ancestry) in the locally un-recombined focal region alone.

667 As the significance of these SNPs could be driven either by linkage to the focal top SNP, or variation in
668 the trait itself, we eliminated these and defined the trait-associated locus as the locally unrecombined
669 region (as defined using number of generations since founder line outbreeding and genome-wide
670 recombination rates) around the lowest p-value SNP. The specific threshold used in SNP elimination was
671 defined via permutations: we iteratively subset the 10 replicates into two equal subsets (N = 5). In one
672 subset, termed 'real-data', light-dark haplotype frequency differences within the focal region were
673 computed as above. In the other subset, the dark-light haplotype frequency difference was calculated (i.e.,
674 termed 'sign-flipped' data). In each subset we sequentially eliminated percentiles of SNPs starting with
675 the top and bottom 2.5% of expected light-dark frequency differences and selected our threshold as the
676 percentile at which the number of true significant SNPs eliminated in the real-date matched that of the
677 sign-flipped data (Fig. S19). From the set of remaining SNPs (those not eliminated in the first iteration),
678 we identified the SNP with the most significant nominal p-value and repeated the procedure. This was
679 iterated until we exhausted all trait-associated SNPs on the chromosome. For each chromosomal arm, we
680 then quantified how the number of trait-associated SNPs decreased as a function of these iterations (Fig.
681 S20). Finally, we merged all loci with overlapping breakpoints to procure a final set of unlinked and
682 independent trait-associated loci.
683 We additionally explored which candidate genes may be underlying the identified loci.
684 Specifically, we identified genes within each locus with a GO term associated with pigmentation
685 (excluding eye pigmentation) or cuticle development/molting. For each of the identified genes, we then
686 quantified whether the number of pigmentation-associated SNPs within the gene (identified via mapping)
687 was greater than expected based on their genome-wide abundance using a binomial test with a Benjamini-
688 Yekutieli false discovery rate correction.
689 *Comparison of mapping to previous association studies*
690 We leveraged results reported by two previous genome wide association studies (Bastide et al.,
691 2013; Dembeck et al., 2015) to quantify the consistency of mapping of pigmentation across disparate
692 populations and studies. In brief, Bastide et al. (2013) conducted a pooled-GWAS approach using wild
693 populations of *D. melanogaster* from Austria and Italy, and focused on female pigmentation on tergite 7.
694 Dembeck et al. (2015) conducted a GWAS using individual *D. melanogaster* lines from the Drosophila
695 Genetics Reference panel (Mackay et al., 2012), focusing on pigmentation variation of tergites A5 and
696 A6. We evaluated whether all SNPs identified as trait-associated in the Dembeck study, and the top 100
697 SNPs from Bastide et al., exhibited more overlap with our unlinked loci than expected by chance.
698 Specifically, we compared the number of SNPs falling within each locus to a random set of SNPs within
699 the chromosome of the same test set size as those obtained from the comparison studies. We iterated this
700 process 1000 times and generated an empirical p-value given the observed overlap, relative to the

701 randomly generated 1000 overlaps. We then re-conducted this analysis, but excluded those within loci
702 containing *tan*, *ebony*, and *bab* to infer the extent to which shared signal across studies was driven by
703 these large effect genes.

704 *Predicting evolution using mapping results*

705 We quantified whether mapped SNPs displayed a behavior during lighter pigmentation evolution
706 that would be expected based on this mapping data alone. Specifically, we computed the frequency shifts
707 for all light-associated alleles at pigmentation-associated SNPs between weeks 1 and 8 of evolution in the
708 outdoor mesocosm, separately for of the unliked, trait-associated loci. We compared the resulting
709 distribution to one derived via a set of matched control SNPs (matched on chromosome and starting
710 frequency (+-/ 2.5%), and excluding those SNPs within 500 Kb of the respective target SNP) using a
711 binomial test. In effect, we quantified whether the proportion of light-associated SNPs within each locus
712 shifting in a manner concordant with that predicted by mapping (i.e., increasing in frequency as lighter
713 evolution evolved) was significantly greater than expected based on our matched control distribution.

714 We then compared how our genomic predictions of trait evolution differed between the semi-
715 natural, outdoor mesocosms and relatively simple, indoor environment. First, we quantified the trajectory
716 of trait-associated SNPs (identified in our outdoor, week 8 mapping population) separately in the outdoor
717 and indoor mesocosm throughout the period of lighter pigmentation evolution (weeks 1 through 8). We
718 compared each resulting distribution to a set of matched control SNPs using a binomial test to infer
719 whether the direction and magnitude of light-favored allele frequency shifts exceeded background allele
720 frequency movement. This analysis was then re-conducted separately for SNPs within each of the top loci
721 on chromosomal arms 2L and 2R, as well as the loci containing the genes *bab*, *ebony*, and *tan* on
722 chromosomal arms 3L, 3R, and X, respectively. Furthermore, the distribution of trait-associated SNP
723 trajectories throughout trait evolution within each environment were compared directly to each other
724 using a paired t-test.

725 Finally, we generated genomic predictions of trait evolution using an approach that explicitly
726 accounted for the observed effect-size of each SNP on trait variation observed via mapping. Specifically,
727 we summed the frequency shifts of light-associated alleles, this time weighting each allele's shift by the
728 associated SNP's GLM regression coefficient. This summation then represents an effective shift in the
729 polygenic score of the population's pigmentation through time, whereby positive values indicate the
730 evolution of lighter pigmentation and negative values, darker pigmentation. We bootstrapped these
731 estimates N = 1000 times and their significance was then assessed using Z-scores computed via a
732 comparison of the estimate to a distribution generated via N = 1000 matched control SNP sets (matched
733 on chromosome and starting frequency). We considered the empirical, bootstrapped value significant if its
734 confidence intervals did not overlap with the matched control distribution.

735

736 *Stability of mapping throughout evolution and across environments of selection*

737 We next sought to determine the stability of the genotype-phenotype map across independent
738 mapping populations (i.e., each mapping environment and timepoint). Accordingly, we re-conducted our
739 mapping procedure, as described above, for the indoor weeks 2 and 8, as well as the outdoor week 2
740 mapping population. First, we asked if the proximity of loci identified across mapping populations were
741 more similar than expected by chance. To address this, for each of the four genotype-phenotype maps we
742 identified the top SNP of each locus (the SNP with the lowest p-value, and also the locus midpoint) and
743 summed the number of times it fell within a locus identified in one of the independent mapping
744 populations. This then produced a score, ranging from 0-3, for each locus and these scores were summed
745 across all loci to produce an empirical overlap score for each mapping population (note that to increase
746 power of this analysis, we selected focal SNPs using the set of pre-merged loci for each mapping
747 population). To determine if the observed overlaps were greater than expected by chance, we generated
748 an empirical null distribution via random sampling of sets of SNPs for each genotype-phenotype map and
749 re-computing overlap scores with each independent mapping (a procedure iterated N = 2000 times for
750 each of the four genotype-phenotype maps). We subsequently interrogated whether the distance between
751 locus-defining midpoint SNPs across sets of loci were closer than expected by chance. Specifically, as
752 before, we first identified the top SNPs for each set of loci from every genotype-phenotype map, and then
753 computed its distance to the closest top SNP of each of the remaining three mappings. This distance
754 metric was averaged to produce an average proximity score which, as above, was compared to a
755 distribution of scores derived via a random set of SNPs (N = 2000 times).

756 We next aimed to quantify the extent to which the identity of specific pigmentation-associated
757 SNPs was shared across independent mapping populations. To do so, we compared the overlap in
758 identified SNPs within and between independent mapping populations. To conduct this analysis in a
759 balanced manner (i.e., compare sets of SNPs identified using the same number of replicate
760 cages/populations), we iteratively subset our set of 10 replicates from each mapping population into
761 groups of 5 replicates and identified pigmentation associated SNPs in each subset using the ancestry-
762 informed regression introduced above. This was iterated with each possible combination of subset
763 replicates, yielding 126 total sets of significant SNPs per mapping population. The percentage of shared
764 SNPs was then calculated for subsets of the same mapping population (e.g. overlap between
765 pigmentation-associated SNPs identified using different subsets of 5 replicate populations from the
766 outdoor, week 8 replicates), and compared to those observed between independent mapping populations
767 (e.g. overlap between sets of pigmentation associated SNPs identified using 5 replicate populations from
768 outdoor week 8 replicates and sets of SNPs from independent treatment and/or time point). Whether the

769 amount of overlap in identified SNPs within vs. between mapping populations differed was evaluated
770 with a planned contrast using a one-way ANOVA model.
771 We next sought to quantify whethe there existed mapped loci with environment-specific marginal
772 effects on pigmentation variation. Accordingly, we first isolated those loci from the outdoor and indoor
773 week 8 mapping populations that had non-overlapping breakpoints with any of the loci in the opposing
774 environments (N = 13 total unique outdoor loci, and N = 13 total unique indoor loci). We then identified
775 all pigmentation-associated SNPs within each unique locus and quantified the light dark-frequency
776 difference for the light-associated allele for the environment in which it was identified, as well as the
777 environment that did not contain the locus. Evidence for environment-specific marginal effects of each
778 locus was then determined by comparing the distribution of light-dark frequency shifts for each locus
779 between environments (using a Wilcoxon Signed Rank Test), as well as to a distribution derived from an
780 equally sized set of matched control SNPs (using a Mann Whitney U test).
781 Finally, we sought to determine if genotype-phenotype maps produced at a particular point in
782 evolution, and within a particular environment of selection, could correctly predict phenotypes of
783 individuals sampled at an independent time and environment. To do so, we used those SNPs identified in
784 our outdoor week 8 mapping and calculated the relative expected phenotypic score for trait value groups
785 (e.g., light, midpoint, and dark color fractions) within this focal mapping, as well as all remaining,
786 independent mapping populations. Specifically, for each light-associated allele at all pigmentation-
787 associated SNPs, we calculated the difference in frequency between the baseline population and each
788 pooled sample. This allele frequency change was weighted in accordance with the SNPs effect size on
789 trait variation as computed via mapping (i.e., by multiplying that by the regression coefficient of our
790 ancestry-informed GLM by the light-favored allele frequency change of that sample). These values were
791 then summed across all SNPs for each sample, providing a metric of the relative lightness of the sample,
792 as predicted by mapping (values greater than 0 indicating lighter pigmentation and less than 0 darker). We
793 then bootstrapped this value 1000 times for each sample. The accuracy of relative predicted phenotypic
794 scores was then evaluated using three criteria: (i) whether the dark samples had, on average, predicted
795 phenotypic values below 0, (ii) whether the light samples had, on average, predicted phenotypic values
796 above 0, and (iii) if there existed a significant linear association between trait value group and relative
797 predicted phenotypic value. The latter was tested using a simple linear regression with the model trait
798 value group ~ phenotype prediction, where trait values groups were assigned a numeric value (dark = 0,
799 midpoint = 1, and light = 2).
800
801 **Code and data availability:** All code necessary to replicate the analyses presented in this manuscript are
802 available at the following: https://github.com/jrhodes613/PigMap2021 .

Founder line sequences are available at the NCBI with the accession ID PRJNA722305. Upon manuscript acceptance, sequences generated via pooled sequencing for this study will be available at the NCBI with the accession ID PRJNA1031645 and phenotypic and allele frequency data will be uploaded to a dryad repository.

Author contributions:

Conceptualization: MCB, DAP, PS

Experimental methodology: MCB, SB, HO, JB, DAP, PS

Formal analysis: MCB, JASR

Writing of manuscript: MCB, JASR, DAP, PS

Manuscript editing: all authors

Acknowledgements: We thank Lauren McIntyre, Jonathan Pritchard, and all members of the Petrov and Schmidt lab for helpful feedback of experimental design and analysis and interpretation. We thank Caitlynn T. Tran and Andy V. Huynh for assistance in pigmentation phenotyping. We thank Heloise Bastide for an extended list of pigmentation-associated loci and results from Bastide et al 2013 (doi: 10.1371/journal.pgen.1003534).

Funding: We are grateful to our funding organizations: the National Science Foundation (NSF PRFB 2109407 to M.C.B.), the National Institutes of Health (NIH R35GM118165-07 to D.A.P. and NIH R01GM137430 to P.S.), and the Chan Zuckerburg Biohub.

References

Abdellaoui, A., Yengo, L., Verweij, K.J.H., Visscher, P.M., 2023. 15 years of GWAS discovery: Realizing the promise. The American Journal of Human Genetics 110, 179–194. https://doi.org/10.1016/j.ajhg.2022.12.011

Akbari, A., Barton, A.R., Gazal, S., Li, Z., Kariminejad, M., Perry, A., Zeng, Y., Mittnik, A., Patterson, N., Mah, M., Zhou, X., Price, A.L., Lander, E.S., Pinhasi, R., Rohland, N., Mallick, S., Reich, D., 2024. Pervasive findings of directional selection realize the promise of ancient DNA to elucidate human adaptation. https://doi.org/10.1101/2024.09.14.613021

Albecker, M.A., Bittar, T.B., Trussell, G.C., Lotterhos, K.E., 2025. A Quantitative Survey of Cogradient and Countergradient Variation in Nature. The American Naturalist 000–000. https://doi.org/10.1086/737682

Bastide, H., Betancourt, A., Nolte, V., Tobler, R., Stöbe, P., Futschik, A., Schlötterer, C., 2013. A Genome-Wide, Fine-Scale Map of Natural Pigmentation Variation in Drosophila melanogaster. PLOS Genetics 9, e1003534. https://doi.org/10.1371/journal.pgen.1003534

Bastide, H., Lange, J.D., Lack, J.B., Yassin, A., Pool, J.E., 2016. A Variable Genetic Architecture of Melanic Evolution in Drosophila melanogaster. Genetics 204, 1307–1319. https://doi.org/10.1534/genetics.116.192492

Behrman, E.L., Howick, V.M., Kapun, M., Staubach, F., Bergland, A.O., Petrov, D.A., Lazzaro, B.P., Schmidt, P.S., 2018. Rapid seasonal evolution in innate immunity of wild Drosophila melanogaster. Proceedings of the Royal Society B: Biological Sciences 285, 20172599. https://doi.org/10.1098/rspb.2017.2599

Berardi, S., Rhodes, J.A., Berner, M.C., Greenblum, S.I., Bitter, M.C., Behrman, E.L., Betancourt, N.J., Bergland, A.O., Petrov, D.A., Rajpurohit, S., Schmidt, P., 2025. Drosophila melanogaster pigmentation demonstrates adaptive phenotypic parallelism over multiple spatiotemporal scales. Evolution Letters qraf008. https://doi.org/10.1093/evlett/qraf008

Bergelson, J., Kreitman, M., Petrov, D.A., Sanchez, A., Tikhonov, M., 2021. Functional biology in its natural context: A search for emergent simplicity. eLife 10, e67646. https://doi.org/10.7554/eLife.67646

Bickel, R.D., Kopp, A., Nuzhdin, S.V., 2011. Composite Effects of Polymorphisms near Multiple Regulatory Elements Create a Major-Effect QTL. PLOS Genetics 7, e1001275. https://doi.org/10.1371/journal.pgen.1001275

Bitter, M.C., Berardi, S., Oken, H., Huynh, A., Lappo, E., Schmidt, P., Petrov, D.A., 2024a. Continuously fluctuating selection reveals fine granularity of adaptation. Nature 1–8. https://doi.org/10.1038/s41586-024-07834-x

Bitter, M.C., Greenblum, S., Rajpurohit, S., Bergland, A.O., Hemker, J.A., Betancourt, N.J., Tilk, S., Berardi, S., Oken, H., Schmidt, P., Petrov, D.A., 2024b. Pervasive fitness trade-offs revealed by rapid adaptation in large experimental populations of Drosophila melanogaster. https://doi.org/10.1101/2024.10.28.620721

Blount, Z.D., Lenski, R.E., Losos, J.B., 2018. Contingency and determinism in evolution: Replaying life's tape. Science 362, eaam5979. https://doi.org/10.1126/science.aam5979

Chiu, Y.-F., Justice, A.E., Melton, P.E., 2016. Longitudinal analytical approaches to genetic data. BMC Genetics 17, S4. https://doi.org/10.1186/s12863-015-0312-y

Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., Kingsley, D.M., 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. Science 307, 1928–1933. https://doi.org/10.1126/science.1107239

Conover, D.O., Schultz, E.T., 1995. Phenotypic similarity and the evolutionary significance of countergradient variation. Trends in Ecology & Evolution 10, 248–252. https://doi.org/10.1016/S0169-5347(00)89081-3

Cox, S.L., Ruff, C.B., Maier, R.M., Mathieson, I., 2019. Genetic contributions to variation in human stature in prehistoric Europe
. PNAS 116, 21484–21492. https://doi.org/10.1073/pnas.1910606116

David, J.R., Capy, P., Gauthier, J.-P., 1990. Abdominal pigmentation and growth temperature in Drosophila melanogaster: Similarities and differences in the norms of reaction of successive segments. Journal of Evolutionary Biology 3, 429–445. https://doi.org/10.1046/j.1420-9101.1990.3050429.x

de Bakker, P.I.W., Burtt, N.P., Graham, R.R., Guiducci, C., Yelensky, R., Drake, J.A., Bersaglieri, T., Penney, K.L., Butler, J., Young, S., Onofrio, R.C., Lyon, H.N., Stram, D.O., Haiman, C.A., Freedman, M.L., Zhu, X., Cooper, R., Groop, L., Kolonel, L.N., Henderson, B.E., Daly, M.J., Hirschhorn, J.N., Altshuler, D., 2006. Transferability of tag SNPs in genetic association studies in multiple populations. Nat Genet 38, 1298–1303. https://doi.org/10.1038/ng1899

Dembeck, L.M., Huang, W., Magwire, M.M., Lawrence, F., Lyman, R.F., Mackay, T.F.C., 2015. Genetic Architecture of Abdominal Pigmentation in Drosophila melanogaster. PLOS Genetics 11, e1005163. https://doi.org/10.1371/journal.pgen.1005163

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., Domingue, B., 2019. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun 10, 3328. https://doi.org/10.1038/s41467-019-11112-0

Durvasula, A., Lohmueller, K.E., 2019. Negative selection on complex traits limits genetic risk prediction accuracy between populations. bioRxiv 721936. https://doi.org/10.1101/721936

Enattah, N.S., Jensen, T.G.K., Nielsen, M., Lewinski, R., Kuokkanen, M., Rasinpera, H., El-Shanti, H., Seo, J.K., Alifrangis, M., Khalil, I.F., Natah, A., Ali, A., Natah, S., Comas, D., Mehdi, S.Q., Groop, L., Vestergaard, E.M., Imtiaz, F., Rashed, M.S., Meyer, B., Troelsen, J., Peltonen, L., 2008. Independent Introduction of Two Lactase-Persistence Alleles into Human Populations Reflects Different History of Adaptation to Milk Culture. The American Journal of Human Genetics 82, 57–72. https://doi.org/10.1016/j.ajhg.2007.09.012

Grieshop, K., Ho, E.K.H., Kasimatis, K.R., 2021. Dominance reversals and the maintenance of genetic variation. https://doi.org/10.48550/arXiv.2109.01571

Guo, Z., Tucker, D.M., Basten, C.J., Gandhi, H., Ersoz, E., Guo, B., Xu, Z., Wang, D., Gay, G., 2014. The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet 127, 749–762. https://doi.org/10.1007/s00122-013-2255-x

Halldorsson, B.V., Eggertsson, H.P., Moore, K.H.S., Hauswedell, H., Eiriksson, O., Ulfarsson, M.O., Palsson, G., Hardarson, M.T., Oddsson, A., Jensson, B.O., Kristmundsdottir, S., Sigurpalsdottir, B.D., Stefansson, O.A., Beyter, D., Holley, G., Tragante, V., Gylfason, A., Olason, P.I., Zink, F., Asgeirsdottir, M., Sverrisson, S.T., Sigurdsson, B., Gudjonsson, S.A., Sigurdsson, G.T., Halldorsson, G.H., Sveinbjornsson, G., Norland, K., Styrkarsdottir, U., Magnusdottir, D.N., Snorradottir, S., Kristinsson, K., Sobech, E., Jonsson, Helgi, Geirsson, A.J., Olafsson, I., Jonsson, P., Pedersen, O.B., Erikstrup, C., Brunak, S., Ostrowski, S.R., Thorleifsson, G., Jonsson, F., Melsted, P., Jonsdottir, I., Rafnar, T., Holm, H., Stefansson, H., Saemundsdottir, J., Gudbjartsson, D.F., Magnusson, O.T., Masson, G., Thorsteinsdottir, U., Helgason, A., Jonsson, Hakon, Sulem, P., Stefansson, K., 2022. The sequences of 150,119 genomes in the UK Biobank. Nature 607, 732–740. https://doi.org/10.1038/s41586-022-04965-x

Hoekstra, H.E., Hirschmann, R.J., Bundey, R.A., Insel, P.A., Crossland, J.P., 2006. A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. Science 313, 101–104. https://doi.org/10.1126/science.1126121

Hof, A.E. van't, Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., Saccheri, I.J., 2016. The industrial melanism mutation in British peppered moths is a transposable element. Nature 534, 102–105. https://doi.org/10.1038/nature17951

Hu, S., Ferreira, L.A.F., Shi, S., Hellenthal, G., Marchini, J., Lawson, D.J., Myers, S.R., 2025. Fine-scale population structure and widespread conservation of genetic effect sizes between human groups across traits. Nat Genet 57, 379–389. https://doi.org/10.1038/s41588-024-02035-8

Huang, W., Richards, S., Carbone, M.A., Zhu, D., Anholt, R.R.H., Ayroles, J.F., Duncan, L., Jordan, K.W., Lawrence, F., Magwire, M.M., Warner, C.B., Blankenburg, K., Han, Y., Javaid, M., Jayaseelan, J., Jhangiani, S.N., Muzny, D., Ongeri, F., Perales, L., Wu, Y.-Q., Zhang, Y., Zou, X., Stone, E.A., Gibbs, R.A., Mackay, T.F.C., 2012. Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proceedings of the National Academy of Sciences 109, 15553–15559. https://doi.org/10.1073/pnas.1213423109

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., Carroll, S.B., 2008. The Evolution of Gene Regulation Underlies a Morphological Difference between Two Drosophila Sister Species. Cell 132, 783–793. https://doi.org/10.1016/j.cell.2008.01.014

Kamiyama, M.T., Bradford, B.Z., Groves, R.L., Guédot, C., 2020. Degree day models to forecast the seasonal phenology of Drosophila suzukii in tart cherry orchards in the Midwest U.S. PLOS ONE 15, e0227726. https://doi.org/10.1371/journal.pone.0227726

Kerner, B., North, K.E., Fallin, M.D., 2009. Use of Longitudinal Data in Genetic Studies in the Genome-wide Association Studies Era: Summary of Group 14. Genet Epidemiol 33, S93–S98. https://doi.org/10.1002/gepi.20479

Kessner, D., Turner, T.L., Novembre, J., 2013. Maximum Likelihood Estimation of Frequencies of Known Haplotypes from Pooled Sequence Data. Mol Biol Evol 30, 1145–1158. https://doi.org/10.1093/molbev/mst016

Kosheleva, K., Desai, M.M., 2018. Recombination Alters the Dynamics of Adaptation on Standing Variation in Laboratory Yeast Populations. Mol Biol Evol 35, 180–201. https://doi.org/10.1093/molbev/msx278

Kronforst, M.R., Barsh, G.S., Kopp, A., Mallet, J., Monteiro, A., Mullen, S.P., Protas, M., Rosenblum, E.B., Schneider, C.J., Hoekstra, H.E., 2012. Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. Pigment Cell & Melanoma Research 25, 411–433. https://doi.org/10.1111/j.1755-148X.2012.01014.x

Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D.H., Martin, A., Reed, R.D., Mullen, S.P., Kronforst, M.R., 2014. doublesex is a mimicry supergene. Nature 507, 229–232. https://doi.org/10.1038/nature13112

Lenski, R.E., 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J 11, 2181–2194. https://doi.org/10.1038/ismej.2017.69

Lohmueller, K.E., 2014. The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits. PLOS Genetics 10, e1004379. https://doi.org/10.1371/journal.pgen.1004379

Long, A.D., Hanson, K.M., Macdonald, S.J., 2025. The Illusion of Polygenicity in Poolseq studies: Insufficient Power Can Mask Simple Genetic Architectures. https://doi.org/10.1101/2025.07.23.666414

Mackay, T.F.C., Anholt, R.R.H., 2024. Pleiotropy, epistasis and the genetic architecture of quantitative traits. Nat Rev Genet 25, 639–657. https://doi.org/10.1038/s41576-024-00711-3

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., Richardson, M.F., Anholt, R.R.H., Barrón, M., Bess, C., Blankenburg, K.P., Carbone, M.A., Castellano, D., Chaboub, L., Duncan, L., Harris, Z., Javaid, M., Jayaseelan, J.C., Jhangiani, S.N., Jordan, K.W., Lara, F., Lawrence, F., Lee, S.L., Librado, P., Linheiro, R.S., Lyman, R.F., Mackey, A.J., Munidasa, M., Muzny, D.M., Nazareth, L., Newsham, I., Perales, L., Pu, L.-L., Qu, C., Ràmia, M., Reid, J.G., Rollmann, S.M., Rozas, J., Saada, N., Turlapati, L., Worley, K.C., Wu, Y.-Q., Yamamoto, A., Zhu, Y., Bergman, C.M., Thornton, K.R., Mittelman, D., Gibbs, R.A., 2012. The Drosophila melanogaster Genetic Reference Panel. Nature 482, 173–178. https://doi.org/10.1038/nature10811

Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., Kenny, E.E., 2017. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. The American Journal of Human Genetics 100, 635–649. https://doi.org/10.1016/j.ajhg.2017.03.004

Massey, J., Wittkopp, P.J., 2016. The genetic basis of pigmentation differences within and between Drosophila species. Curr Top Dev Biol 119, 27–61. https://doi.org/10.1016/bs.ctdb.2016.03.004

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E.R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J.L., de Castro, J.M.B., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M.A.R., Roodenberg, J., Vergès, J.M., Krause, J., Cooper, A., Alt, K.W., Brown, D., Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R., Reich, D., 2015. Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528, 499–503. https://doi.org/10.1038/nature16152

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J.K., Przeworski, M., 2020. Variable prediction accuracy of polygenic scores within an ancestry group. eLife 9, e48376. https://doi.org/10.7554/eLife.48376

Nadeau, N.J., Pardo-Diaz, C., Whibley, A., Supple, M.A., Saenko, S.V., Wallbank, R.W.R., Wu, G.C., Maroja, L., Ferguson, L., Hanly, J.J., Hines, H., Salazar, C., Merrill, R.M., Dowling, A.J., ffrench-Constant, R.H., Llaurens, V., Joron, M., McMillan, W.O., Jiggins, C.D., 2016. The gene

cortex controls mimicry and crypsis in butterflies and moths. Nature 534, 106–110. https://doi.org/10.1038/nature17961

Novembre, J., Barton, N.H., 2018. Tread Lightly Interpreting Polygenic Tests of Selection. Genetics 208, 1351–1355. https://doi.org/10.1534/genetics.118.300786

Paaby, A.B., Bergland, A.O., Behrman, E.L., Schmidt, P.S., 2014. A highly pleiotropic amino acid polymorphism in the Drosophila insulin receptor contributes to life-history adaptation. Evol 68, 3395–3409. https://doi.org/10.1111/evo.12546

Patel, R.A., Musharoff, S.A., Spence, J.P., Pimentel, H., Tcheandjieu, C., Mostafavi, H., Sinnott-Armstrong, N., Clarke, S.L., Smith, C.J., Durda, P.P., Taylor, K.D., Tracy, R., Liu, Y., Johnson, W.C., Aguet, F., Ardlie, K.G., Gabriel, S., Smith, J., Nickerson, D.A., Rich, S.S., Rotter, J.I., Tsao, P.S., Assimes, T.L., Pritchard, J.K., 2022. Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits. The American Journal of Human Genetics 109, 1286–1297. https://doi.org/10.1016/j.ajhg.2022.05.014

Phillips, P.C., 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9, 855–867. https://doi.org/10.1038/nrg2452

Pool, J.E., Aquadro, C.F., 2007. The genetic basis of adaptive pigmentation variation in Drosophila melanogaster. Molecular Ecology 16, 2844–2851. https://doi.org/10.1111/j.1365-294X.2007.03324.x

Privé, F., Aschard, H., Carmi, S., Folkersen, L., Hoggart, C., O'Reilly, P.F., Vilhjálmsson, B.J., 2022. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. The American Journal of Human Genetics 109, 12–23. https://doi.org/10.1016/j.ajhg.2021.11.008

Reed, R.D., Papa, R., Martin, A., Hines, H.M., Counterman, B.A., Pardo-Diaz, C., Jiggins, C.D., Chamberlain, N.L., Kronforst, M.R., Chen, R., Halder, G., Nijhout, H.F., McMillan, W.O., 2011. optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333, 1137–1141. https://doi.org/10.1126/science.1208227

Rudman, S.M., Greenblum, S.I., Rajpurohit, S., Betancourt, N.J., Hanna, J., Tilk, S., Yokoyama, T., Petrov, D.A., Schmidt, P., 2022. Direct observation of adaptive tracking on ecological time scales in Drosophila. Science 375, eabj7484. https://doi.org/10.1126/science.abj7484

Sella, G., Barton, N.H., 2019. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. Annual Review of Genomics and Human Genetics 20, 461–493. https://doi.org/10.1146/annurev-genom-083115-022316

Stroud, J.T., Ratcliff, W.C., 2025. Long-term studies provide unique insights into evolution. Nature 639, 589–601. https://doi.org/10.1038/s41586-025-08597-9

Svensson, E.I., Arnold, S.J., Bürger, R., Csilléry, K., Draghi, J., Henshaw, J.M., Jones, A.G., De Lisle, S., Marques, D.A., McGuigan, K., Simon, M.N., Runemark, A., 2021. Correlational selection in the age of genomics. Nat Ecol Evol 5, 562–573. https://doi.org/10.1038/s41559-021-01413-3

Tilk, S., Bergland, A., Goodman, A., Schmidt, P., Petrov, D., Greenblum, S., 2019. Accurate Allele Frequencies from Ultra-low Coverage Pool-Seq Samples in Evolve-and-Resequence Experiments. G3 (Bethesda) 9, 4159–4168. https://doi.org/10.1534/g3.119.400755

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 101, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Walsh, B., Blows, M.W., 2009. Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. Annu. Rev. Ecol. Evol. Syst. 40, 41–59. https://doi.org/10.1146/annurev.ecolsys.110308.120232

Wortel, M.T., Agashe, D., Bailey, S.F., Bank, C., Bisschop, K., Blankers, T., Cairns, J., Colizzi, E.S., Cusseddu, D., Desai, M.M., van Dijk, B., Egas, M., Ellers, J., Groot, A.T., Heckel, D.G., Johnson, M.L., Kraaijeveld, K., Krug, J., Laan, L., Lässig, M., Lind, P.A., Meijer, J., Noble, L.M., Okasha, S., Rainey, P.B., Rozen, D.E., Shitut, S., Tans, S.J., Tenaillon, O., Teotónio, H., de Visser, J.A.G.M., Visser, M.E., Vroomans, R.M.A., Werner, G.D.A., Wertheim, B., Pennings,

P.S., 2022. Towards evolutionary predictions: Current promises and challenges. Evol Appl 16, 3–21. https://doi.org/10.1111/eva.13513

Wright, T.R.F., 1987. The Genetics Of Biogenic Amine Metabolism, Sclerotization, And Melanization In *Drosophila Melanogaster**, in: Scandalios, J.G., Caspari, E.W. (Eds.), Advances in Genetics, Molecular Genetics of Development. Academic Press, pp. 127–222. https://doi.org/10.1016/S0065-2660(08)60008-5

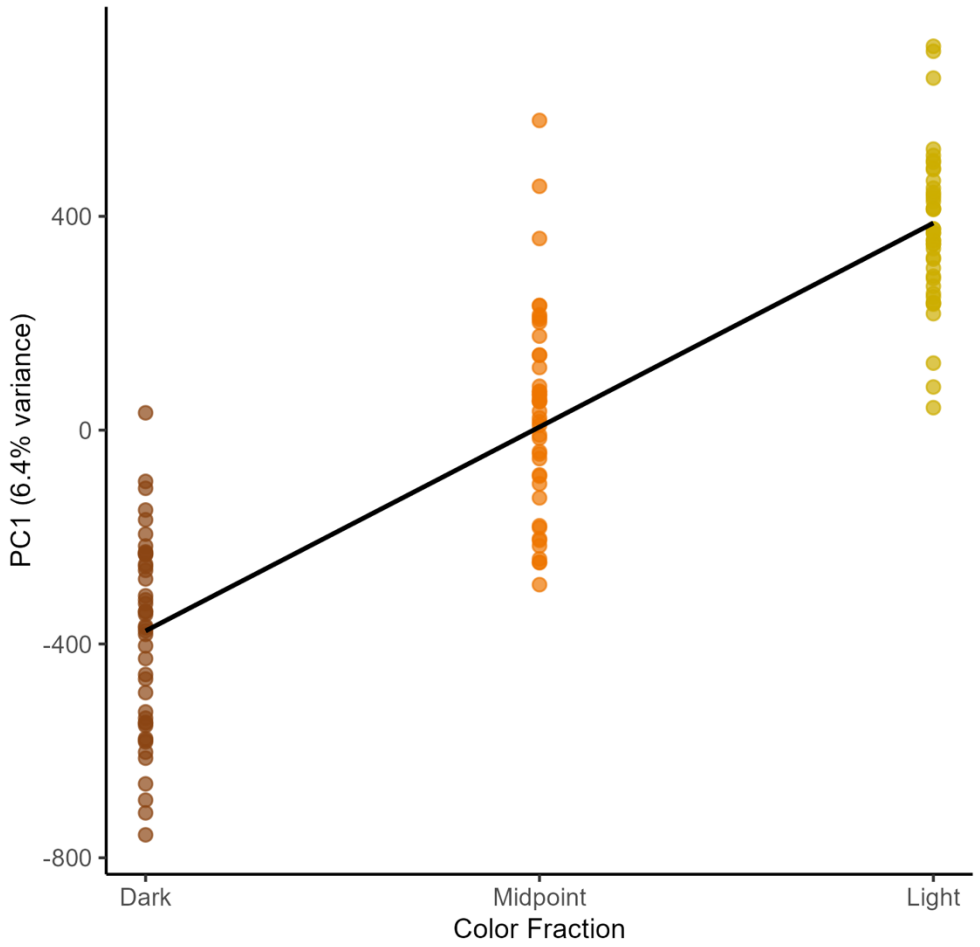
Wu, Y., Zheng, Z., Thibaut, L., Lin, T., Feng, Q., Cheng, H., Yengo, L., Goddard, M.E., Wray, N.R., Visscher, P.M., Zeng, J., 2025. Genome-wide fine-mapping improves identification of causal variants. medRxiv 2024.07.18.24310667. https://doi.org/10.1101/2024.07.18.24310667

Wu, Z., Hu, Y., Melton, P.E., 2014. Longitudinal Data Analysis for Genetic Studies in the Whole-Genome Sequencing Era. Genetic Epidemiology 38, S74–S80. https://doi.org/10.1002/gepi.21829

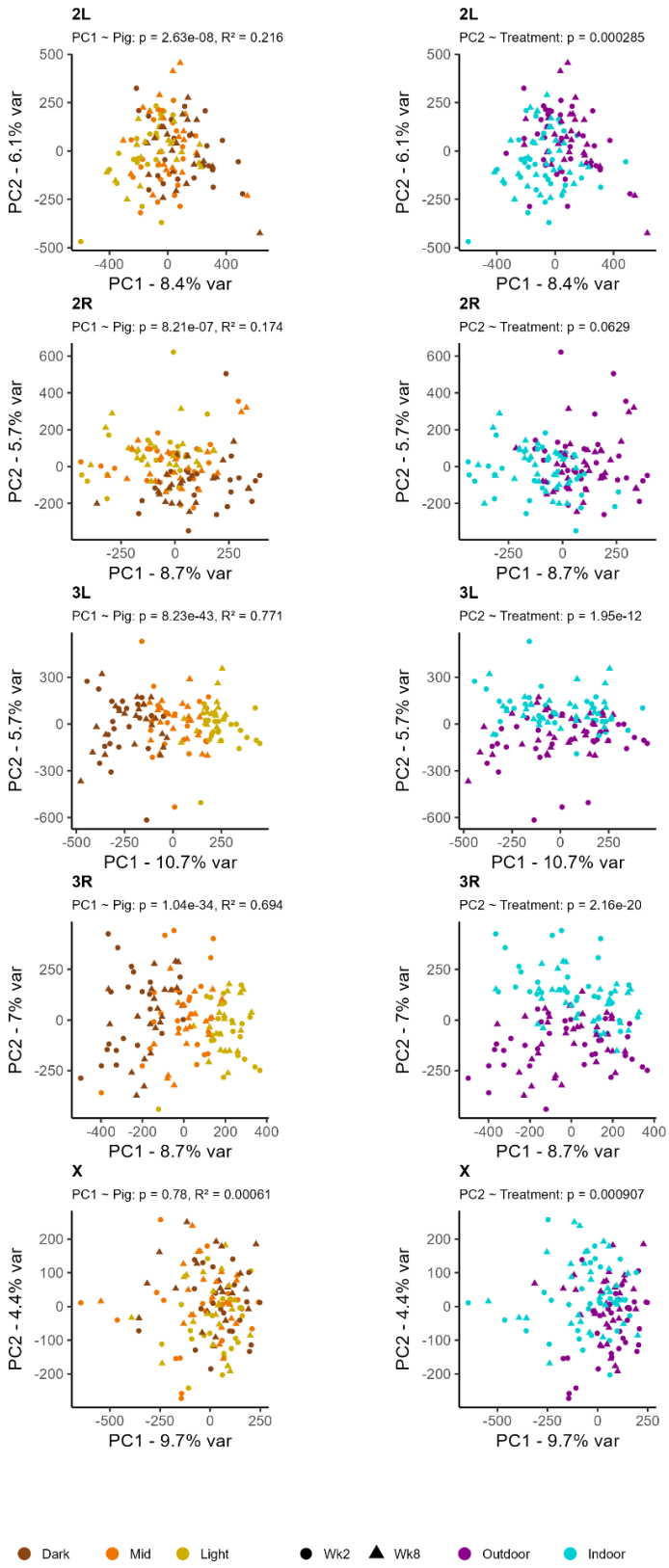
Yair, S., Coop, G., 2021. Population differentiation of polygenic score predictions under stabilizing selection. https://doi.org/10.1101/2021.09.10.459833

Supplementary Information for Smiley-Rhodes* and Bitter*, *et al.*

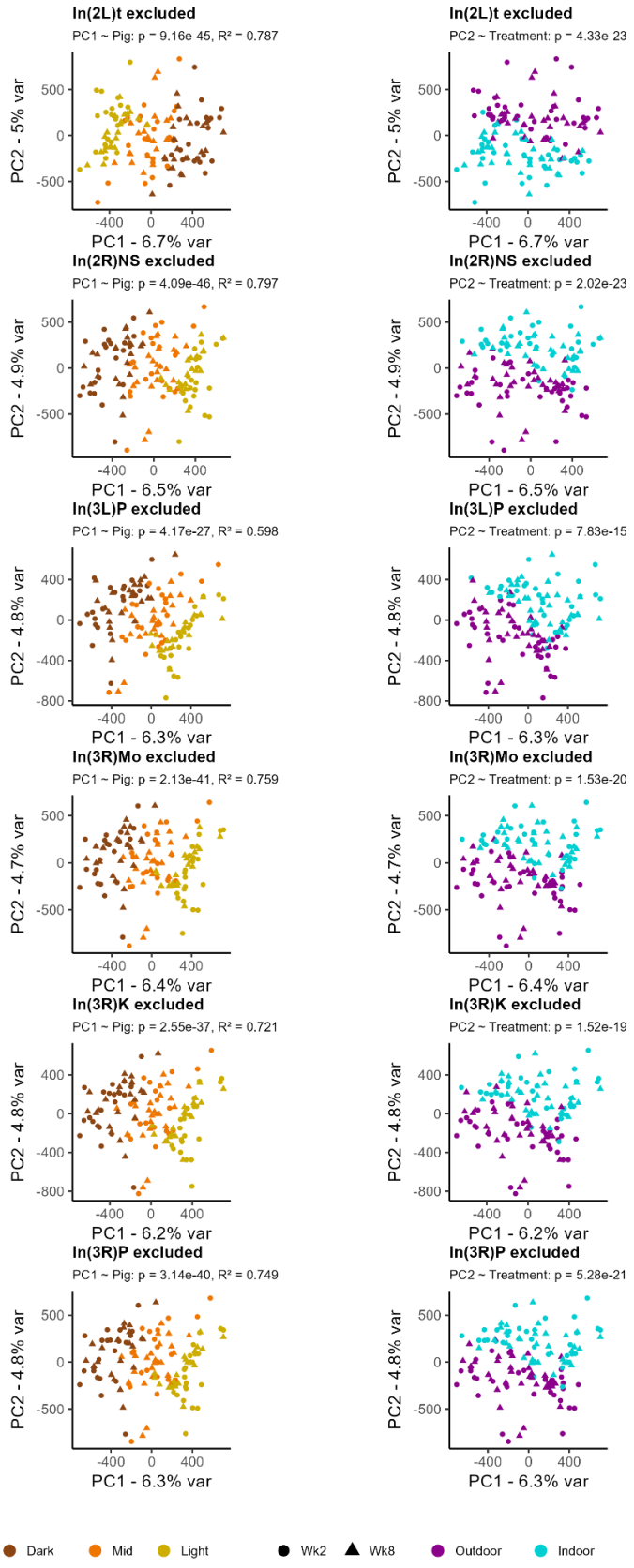
Supplementary Figures



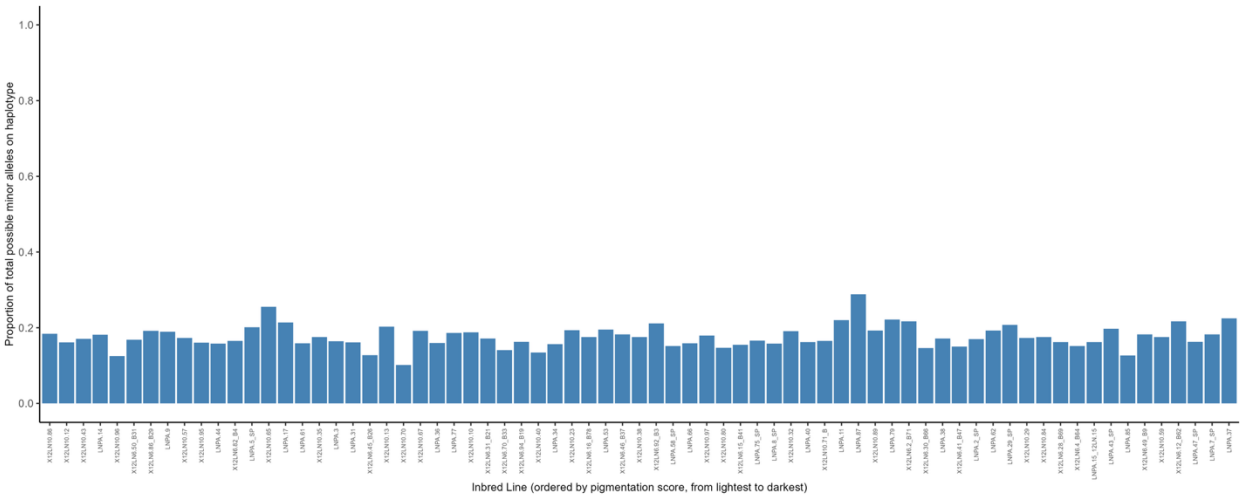
Supplementary Figure 1 | Relationship between color fraction and PC1. Samples plotted by color fraction (dark, midpoint, or light) and the coordinate along the first principal component generated via PCA of allele frequency data across all SNPs and samples used in mapping. A linear regression line (line in figure) between trait value and PC1 is significant ($R^2 = 0.762$; $p << 0.001$).


2L
PC1 ~ Pig: p = 2.63e-08, R² = 0.216
PC2 - 6.1% var
PC1 - 8.4% var
2L
PC2 ~ Treatment: p = 0.000285
PC2 - 6.1% var
PC1 - 8.4% var
2R
PC1 ~ Pig: p = 8.21e-07, R² = 0.174
PC2 - 5.7% var
PC1 - 8.7% var
2R
PC2 ~ Treatment: p = 0.0629
PC2 - 5.7% var
PC1 - 8.7% var
3L
PC1 ~ Pig: p = 8.23e-43, R² = 0.771
PC2 - 5.7% var
PC1 - 10.7% var
3L
PC2 ~ Treatment: p = 1.95e-12
PC2 - 5.7% var
PC1 - 10.7% var
3R
PC1 ~ Pig: p = 1.04e-34, R² = 0.694
PC2 - 7% var
PC1 - 8.7% var
3R
PC2 ~ Treatment: p = 2.16e-20
PC2 - 7% var
PC1 - 8.7% var
X
PC1 ~ Pig: p = 0.78, R² = 0.00061
PC2 - 4.4% var
PC1 - 9.7% var
X
PC2 ~ Treatment: p = 0.000907
PC2 - 4.4% var
PC1 - 9.7% var
Dark
Mid
Light
Wk2
Wk8
Outdoor
Indoor

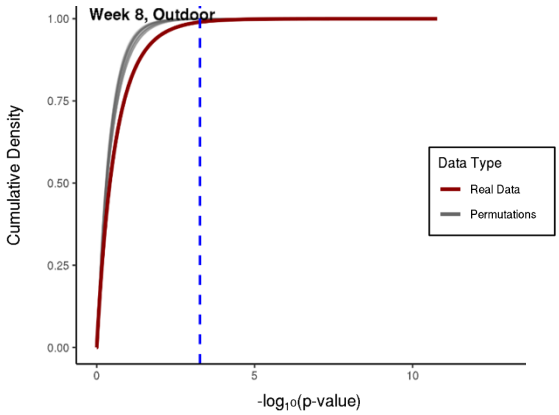
Supplementary Figure 2 | PCA by chromosome. Projection of samples used in trait mapping onto the first two principal components generated via PCA of allele frequency data from all SNPs located on each chromosomal arm. In the column on the left, samples are colored by the color fraction from which they originate (dark, midpoint, or light) and the shape of the points corresponds to the timepoint when the sample was collected. In the right column, samples are colored by treatment.


In(2L)t excluded
PC1 ~ Pig: p = 9.16e-45, R² = 0.787
In(2L)t excluded
PC2 ~ Treatment: p = 4.33e-23
In(2R)NS excluded
PC1 ~ Pig: p = 4.09e-46, R² = 0.797
In(2R)NS excluded
PC2 ~ Treatment: p = 2.02e-23
In(3L)P excluded
PC1 ~ Pig: p = 4.17e-27, R² = 0.598
In(3L)P excluded
PC2 ~ Treatment: p = 7.83e-15
In(3R)Mo excluded
PC1 ~ Pig: p = 2.13e-41, R² = 0.759
In(3R)Mo excluded
PC2 ~ Treatment: p = 1.53e-20
In(3R)K excluded
PC1 ~ Pig: p = 2.55e-37, R² = 0.721
In(3R)K excluded
PC2 ~ Treatment: p = 1.52e-19
In(3R)P excluded
PC1 ~ Pig: p = 3.14e-40, R² = 0.749
In(3R)P excluded
PC2 ~ Treatment: p = 5.28e-21
PC2 - 5% var
PC2 - 4.9% var
PC2 - 4.8% var
PC2 - 4.7% var
PC1 - 6.7% var
PC1 - 6.5% var
PC1 - 6.3% var
PC1 - 6.4% var
PC1 - 6.2% var
Dark
Mid
Light
Wk2
Wk8
Outdoor
Indoor

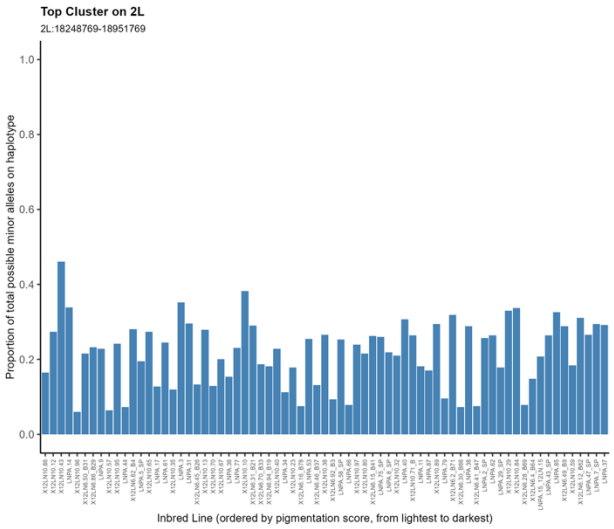
Supplementary Figure 3 | PCA without variation within inversions. Projection of samples used in trait mapping onto the first two principal components generated via PCA of allele frequency data from all SNPs except those located in each of six common inversions found in wild populations. In the column on the left, samples are colored by the color fraction from which they originate (dark, midpoint, or light) and the shape of the points corresponds to the timepoint when the sample was collected. In the right column, samples are colored by treatment.

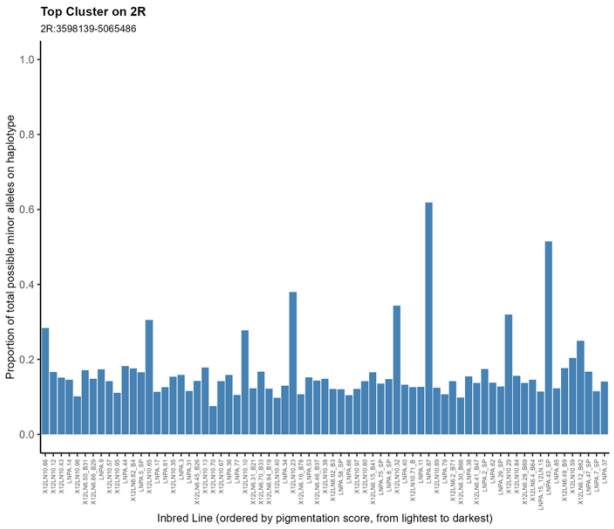


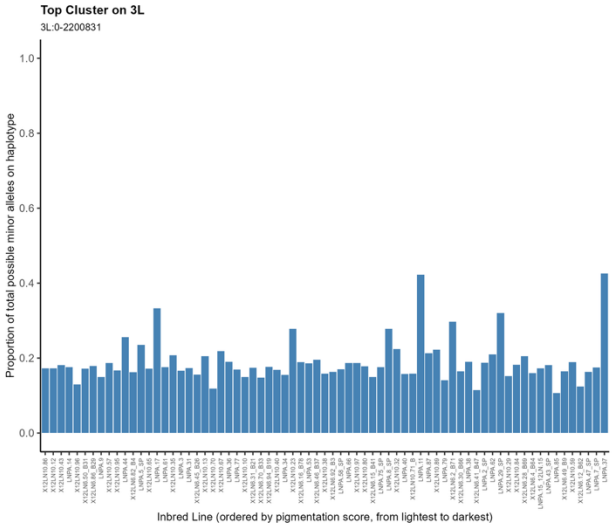
Supplementary Figure 4 | Distribution of pigmentation-associated SNPs across inbred reference panel. The proportion of all pigmentation-associated, significant SNPs (minor allele, identified via our ancestry-informed GLM conducted on the week 8, outdoor mapping population) present on each of the 76 founding inbred lines. Inbred lines are ordered from lightest to darkest pigmentation, as quantified via manual scoring. As can be observed, those SNPs identified by our ancestry-informed GLM are not exclusive to either the lightest or darkest lines, but rather present at appreciable frequency across all haplotypes.

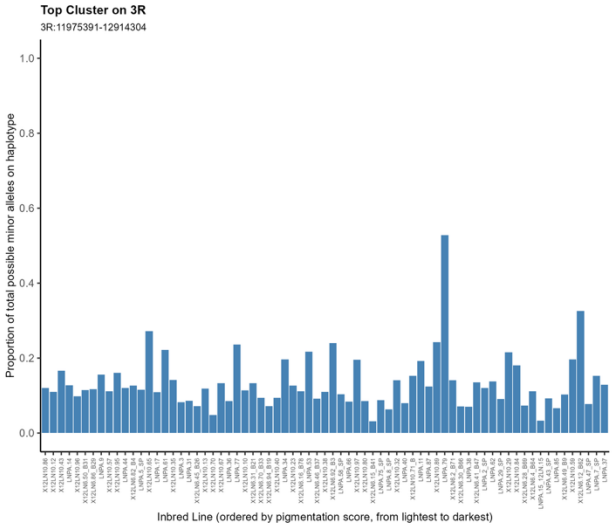


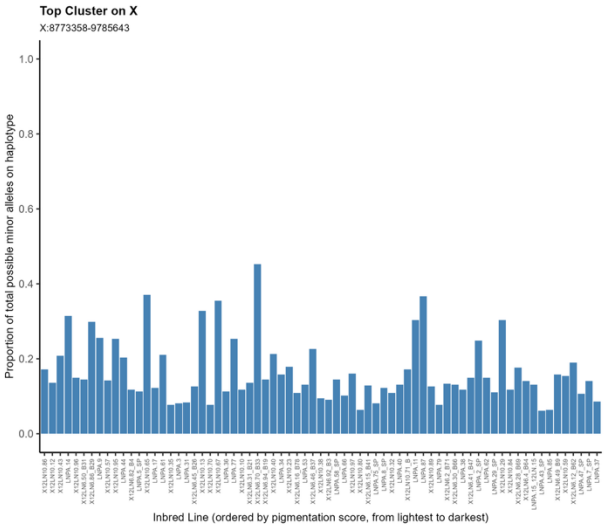
Supplementary Figure 5 | GLM empirical and permuted p-value cumulative distribution functions. CDFs of log10 p-values of SNPs resulting from a GLM performed with real data (red) and 100 GLMs performed with shuffled, permuted data (grey). The cutoff, determined as the top 0.05% of permuted p-values, is drawn as a blue dotted line for each timepoint and treatment.


Top Cluster on 2L
2L:18248769-18951769
Proportion of total possible minor alleles on haplotype
Inbred Line (ordered by pigmentation score, from lightest to darkest)

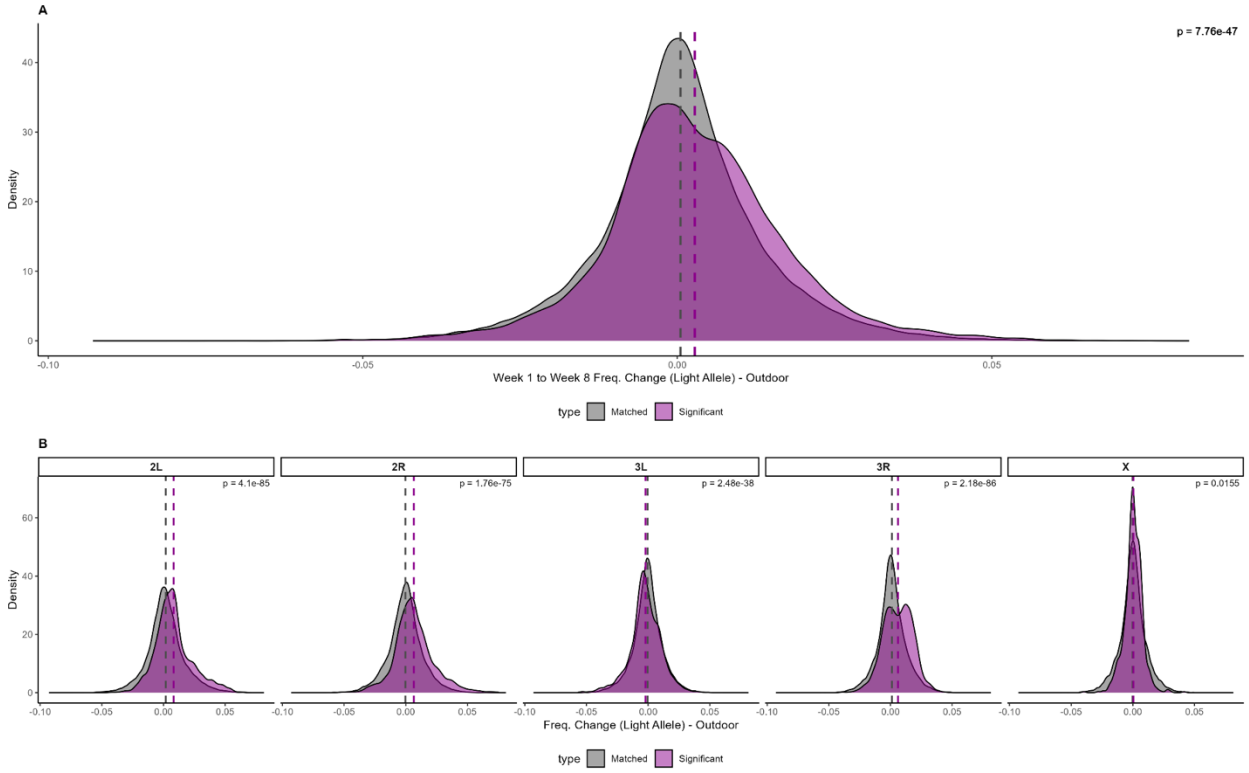

Top Cluster on 2R
2R:3598139-5065486
Proportion of total possible minor alleles on haplotype
Inbred Line (ordered by pigmentation score, from lightest to darkest)


Top Cluster on 3L
3L:0-2200831
Proportion of total possible minor alleles on haplotype
Inbred Line (ordered by pigmentation score, from lightest to darkest)

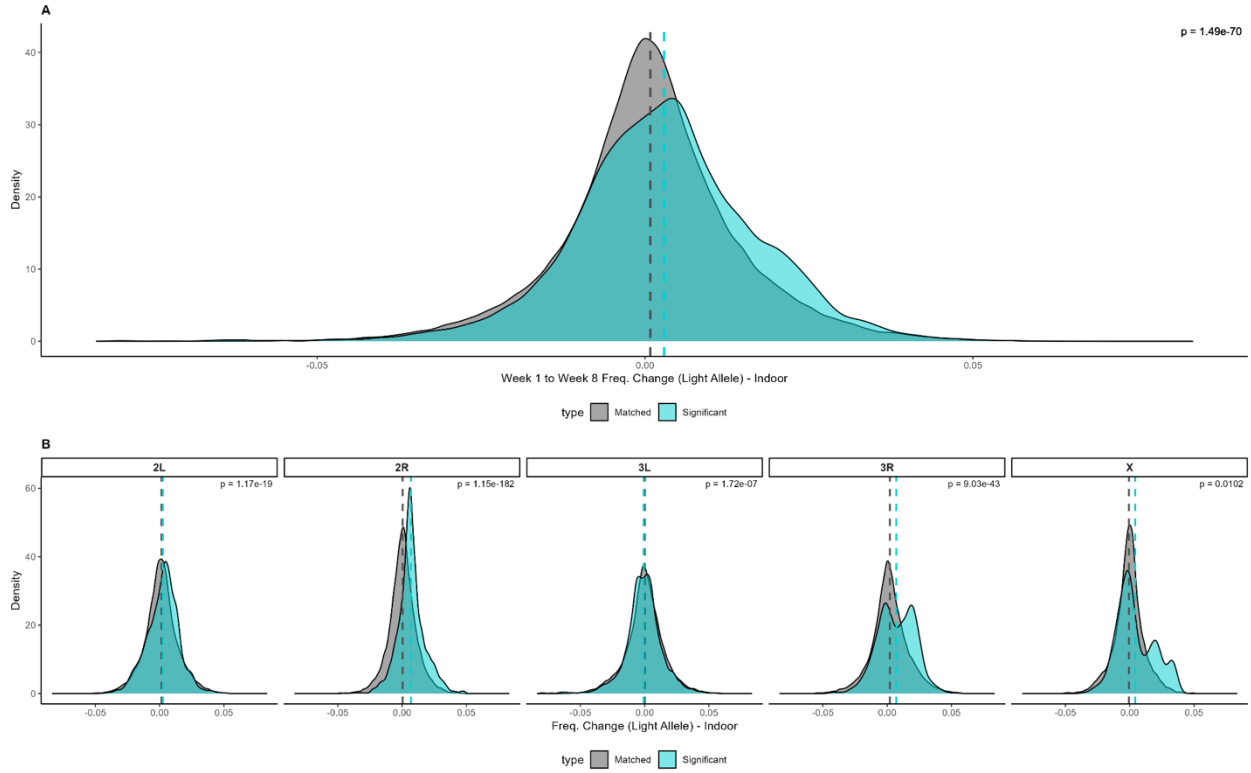

Top Cluster on 3R
3R:11975391-12914304
Proportion of total possible minor alleles on haplotype
Inbred Line (ordered by pigmentation score, from lightest to darkest)


Top Cluster on X
X:8773358-9785643
Proportion of total possible minor alleles on haplotype
Inbred Line (ordered by pigmentation score, from lightest to darkest)

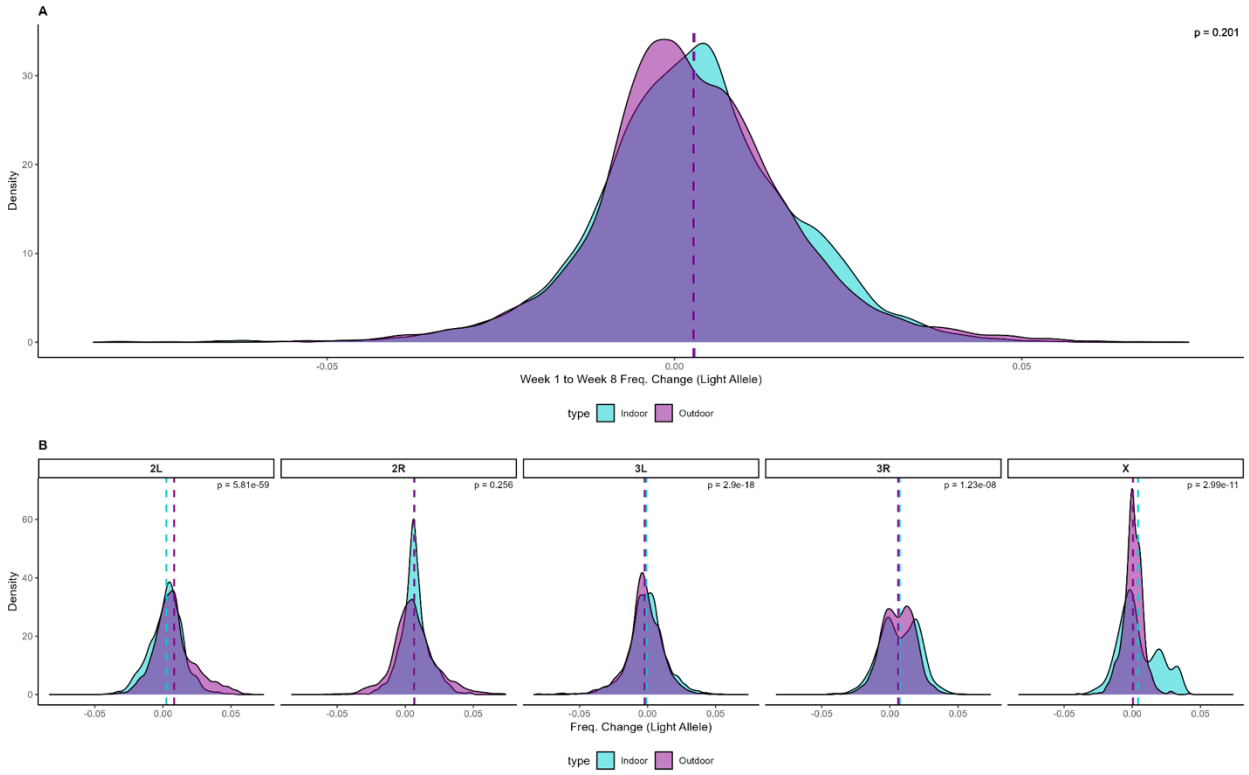
Supplementary Figure 6 | Distribution of pigmentation-associated SNPs across inbred reference panel for the largest-effect, independent loci identified on each chromosomal arm. The proportion of all pigmentation-associated SNPs within each of the top independent loci (clusters) on each chromosomal arm (minor allele, identified via our ancestry-informed GLM conducted on the week 8, outdoor mapping population) present on each of the 76 founding inbred lines. Inbred lines are ordered from lightest to darkest pigmentation score as quantified via manual scoring.



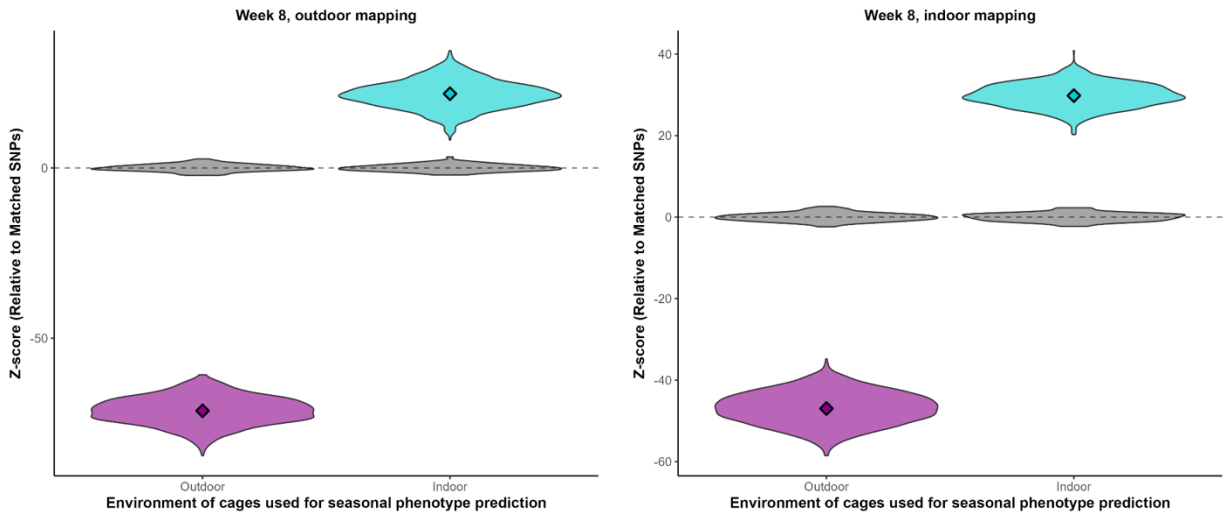
Supplementary Figure 7 | Seasonal shifts of trait-associated SNPs in outdoor cages. A) Distribution of light-associated allele frequency shifts at all trait-associated SNPs (identified at Week 8 in outdoor cages) between Weeks 1 and 8 of evolution in the outdoor field environment. Distributions of frequency shifts at trait-associated SNPs are dark magenta, while distributions derived from matched controls are grey. Reported p-values are derived from a comparison of trait-associated and matched control SNPs using a binomial test. B) Same as (A), but segregated by chromosomal arm.



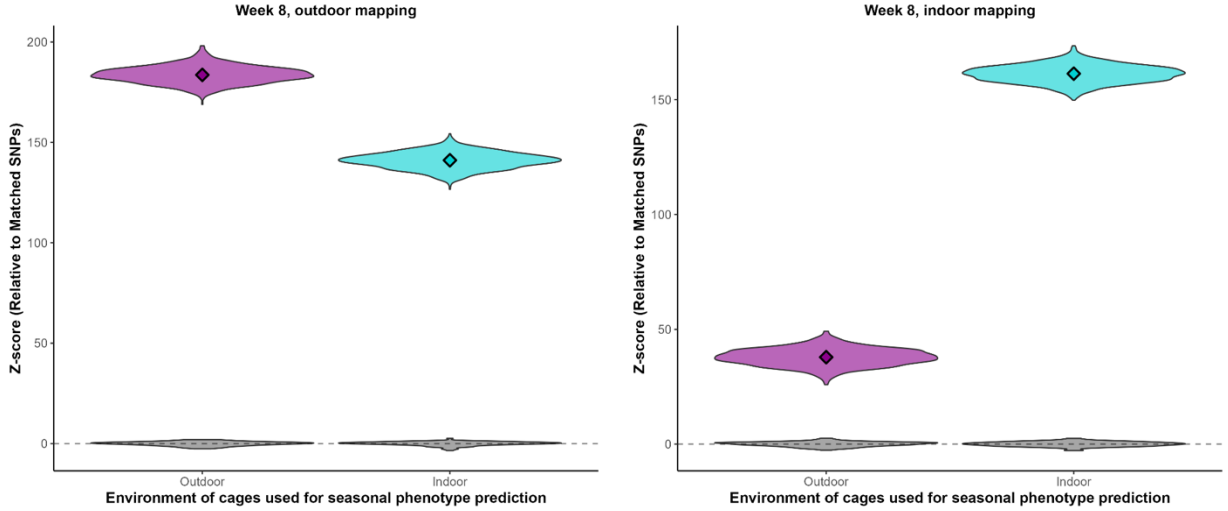
Supplementary Figure 8 | Seasonal shifts of trait-associated SNPs in indoor cages. A) Distribution of light-associated allele frequency shifts at all trait-associated SNPs (identified at Week 8 in outdoor cages) between Weeks 1 and 8 of evolution in the indoor lab environment. Distributions of frequency shifts at trait-associated SNPs are turquoise, while distributions derived from matched controls are grey. Reported p-values are derived from a comparison of trait-associated and matched control SNPs using a binomial test. B) Same as (A), but segregated by chromosomal arm.



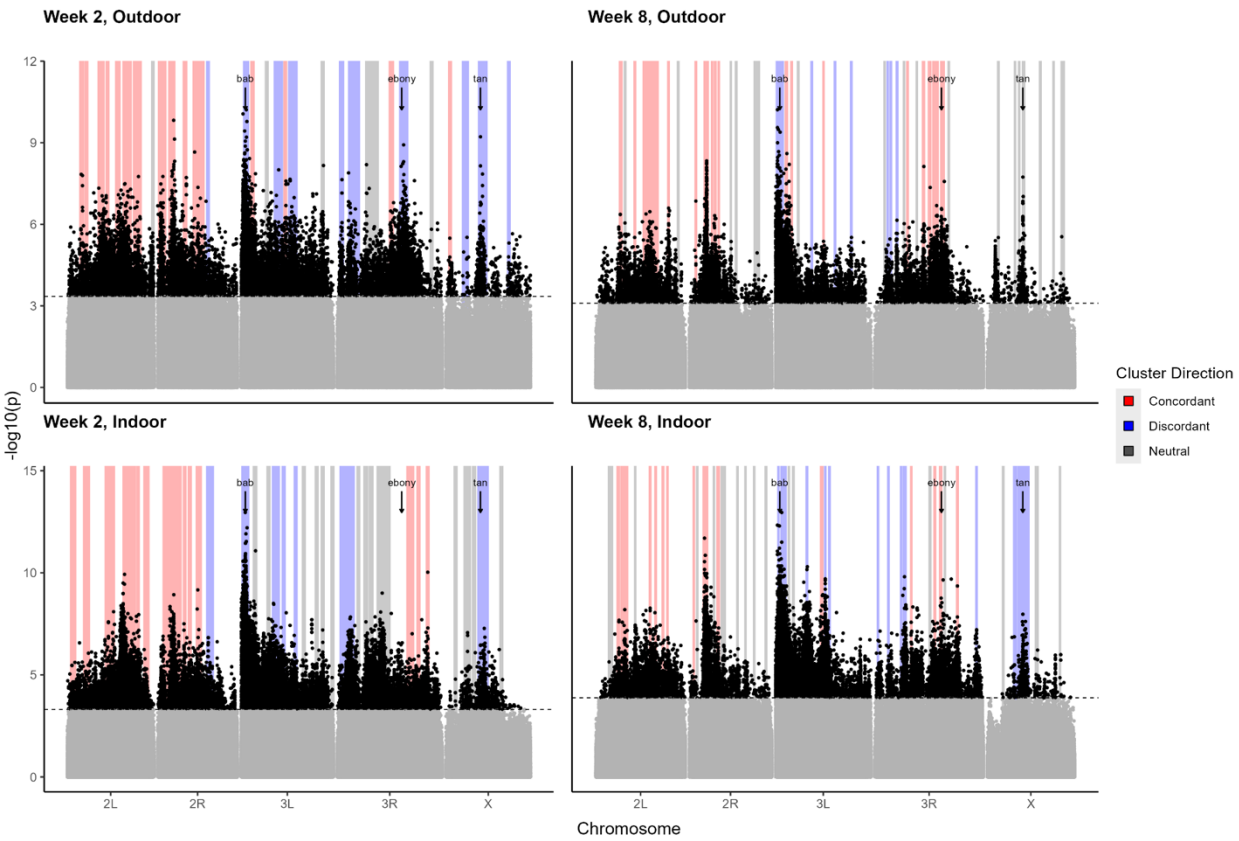
Supplementary Figure 9 | Seasonal shifts of trait-associated SNPs in outdoor versus indoor cages. A) Distribution of light-associated allele frequency shifts at all trait-associated SNPs (identified at Week 8 in outdoor cages) between Weeks 1 and 8 of evolution in the outdoor field vs. indoor lab environment across the whole genome. Distributions of frequency shifts at trait-associated SNPs in the outdoor cages are magenta, while distributions in indoor cages are turquoise. The indoor and outdoor distributions were compared using a two-sided t-test. B) Same as (A), but segregated by chromosomal arm.



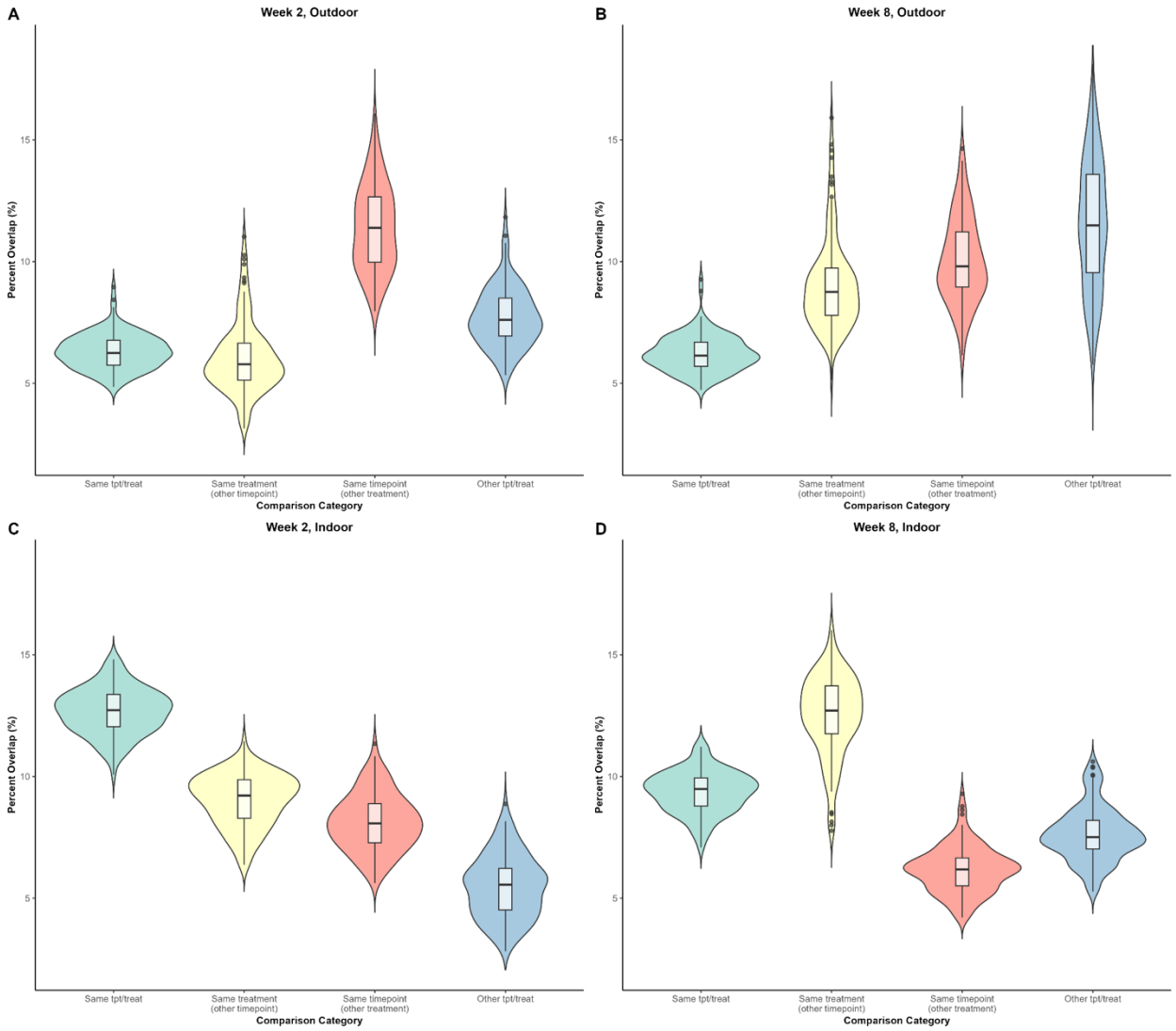
Supplementary Figure 10 | Phenotypic prediction with only *tan*, *ebony*, and *bab*, using mapping from each timepoint/treatment subset. Relative, predicted phenotypic change between Weeks 1 and 8 in both indoor and outdoor environments, based on only the three clusters that contain *tan, ebony*, or *bab*. Each quadrant uses a different mapping (for example, from Week 8 in the outdoor cages) and within each quadrant, we use the mapping to predict the phenotypic change in the outdoor environment (magenta) and the indoor environment (turquoise). The predicted change is denoted with diamonds, while the violins indicate the distribution of 1000 bootstrapped values (colored for trait-associated SNPs and grey to matched control SNPs).



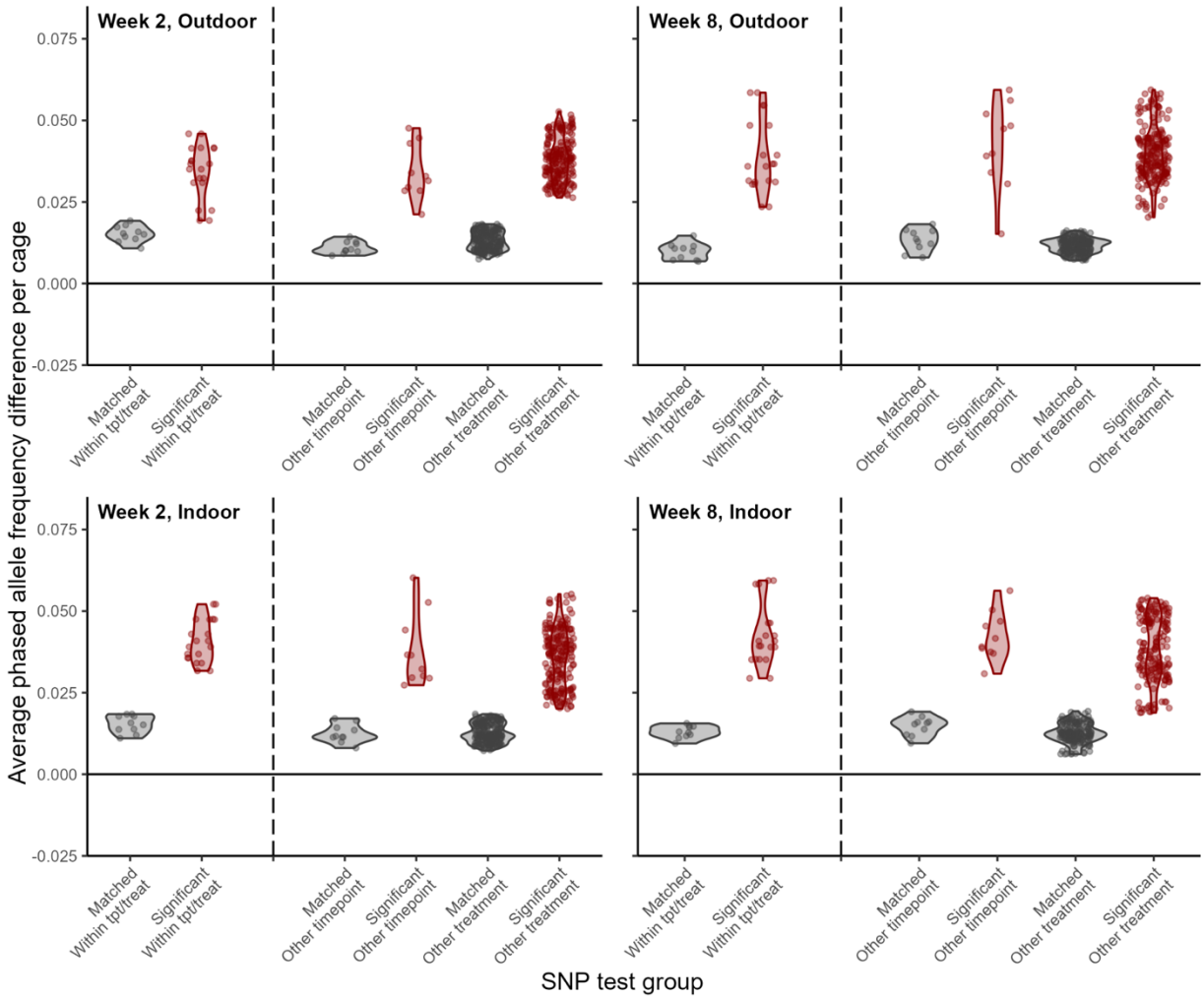
Supplementary Figure 11 | Phenotypic prediction excluding *tan*, *ebony*, and *bab*. Relative, predicted phenotypic change between Weeks 1 and 8 in both outdoor (magenta) and indoor (turquoise) environments, generated using our trait mapping results derived at weeks 8 in the outdoor (left) and indoor environment, and using variation from every mapped locus excluding those containing *tan, ebony*, or *bab*. Predicted phenotypic change are quantified as Z-scores (computed relative to matched control shifts) whereby positive Z-scores indicate net lighter pigmentation evolution while negative values would indicate net darker evolution. Z-scores were computed by comparing inferred relative phenotypic change using mapped SNPs to a that computed using a set of matched control SNPs (grey distributions).



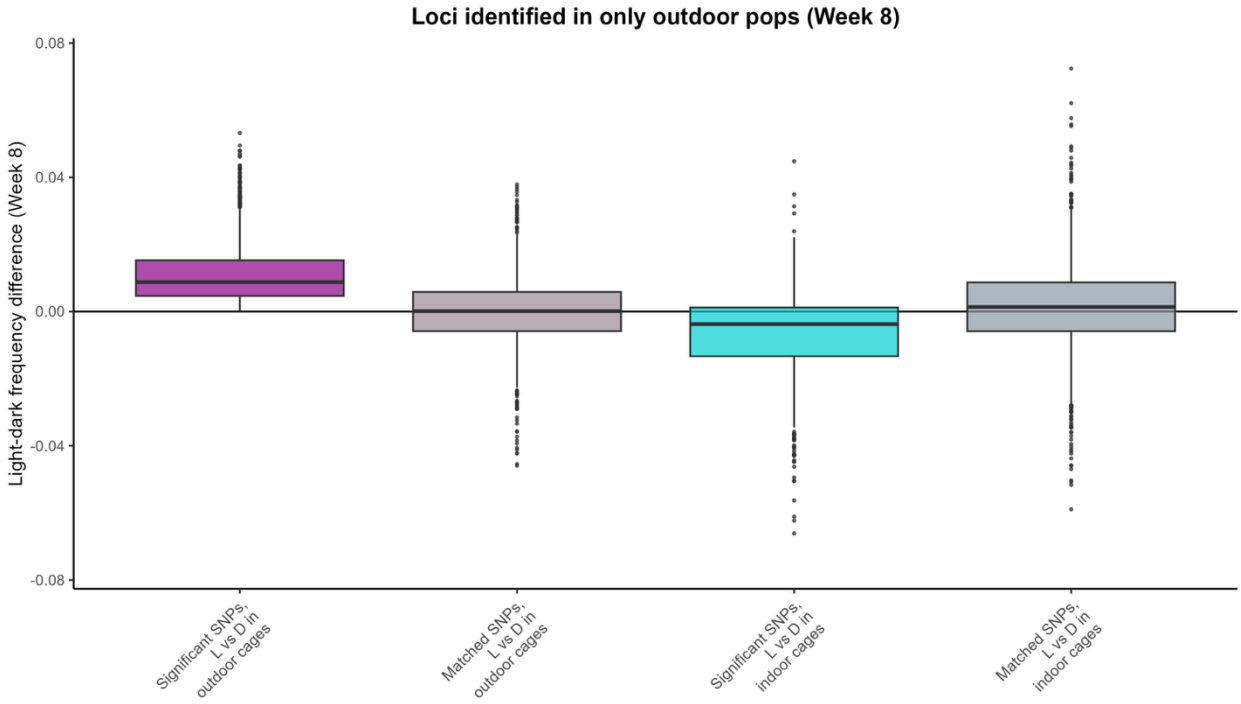
Supplementary Figure 12 | Manhattan plots produced across all mapping populations. Manhattan plot depicting nominal p-value's derived from an ancestry-informed mapping of SNP association with pigmentation variation, separately for each mapping population (week 2 indoor, week 2 outdoor, week 8 indoor, week 8 outdoor). The dashed horizontal line corresponds to an empirical FDR threshold of 0.0005, derived via permutations. Shaded regions denote the stop and end points of each of the independent loci, while bar color denotes behavior of SNPs within each locus through time as the populations evolved lighter pigmentation within the outdoor mesocosm: red shading indicates loci with light-associated alleles favored throughout evolution of lighter pigmentation, blue shading indicates loci in which the light-associated alleles are predominately selected against and grey shading indicates loci with no signature of selection on light associated alleles. The location of the canonical pigmentation genes, *bric-a-brac (bab)*, *ebony*, and *tan*, are indicated on the Manhattan plots. Potential candidate genes underlying the remaining large effect loci observed throughout the genome are discussed in the Supplement.



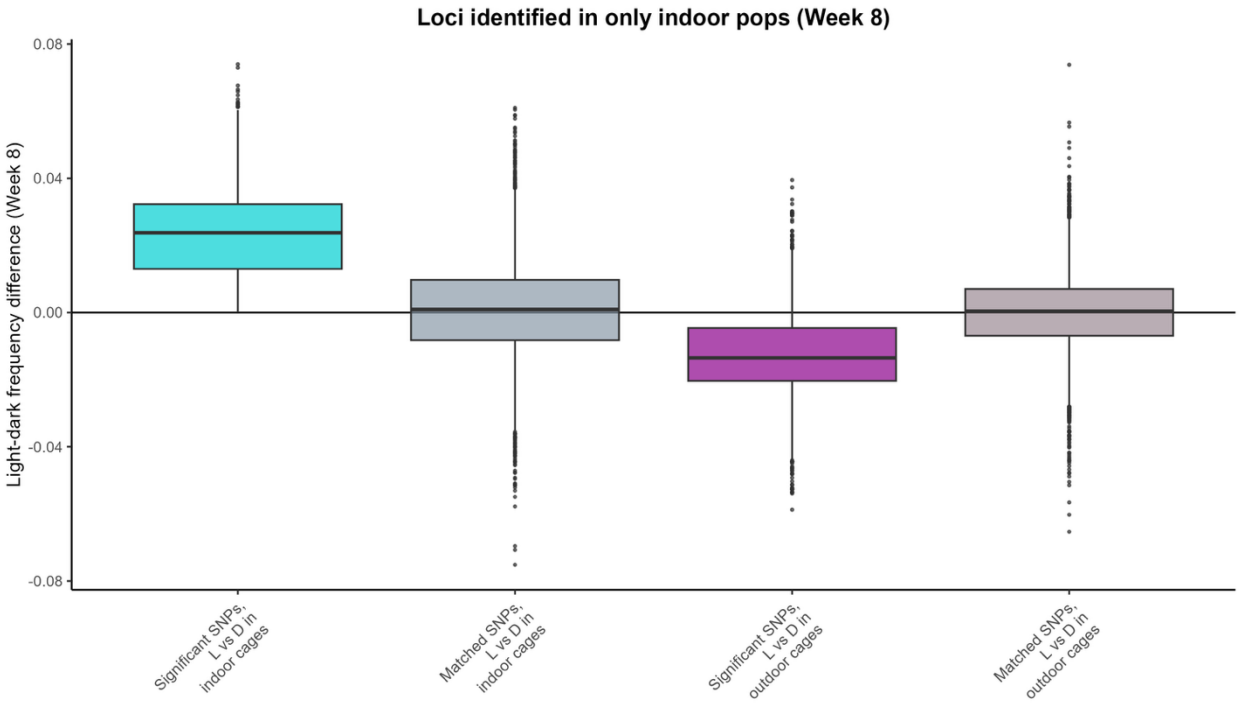
Supplementary Figure 13 | Shared significant SNPs across mapping populations. Overlap percentage from each pair of GLMs that each use five of the cages from a specific timepoint and treatment combination. Significant SNP overlap is either considered between GLMs using cages within the same timepoint/treatment (green), the same treatment but the other timepoint (yellow), the same timepoint but the other treatment (red), or cages from both the other timepoint and treatment (blue). The box plots represent the real percentage from each paired comparison, while the colored violins represent our bootstrapping results.



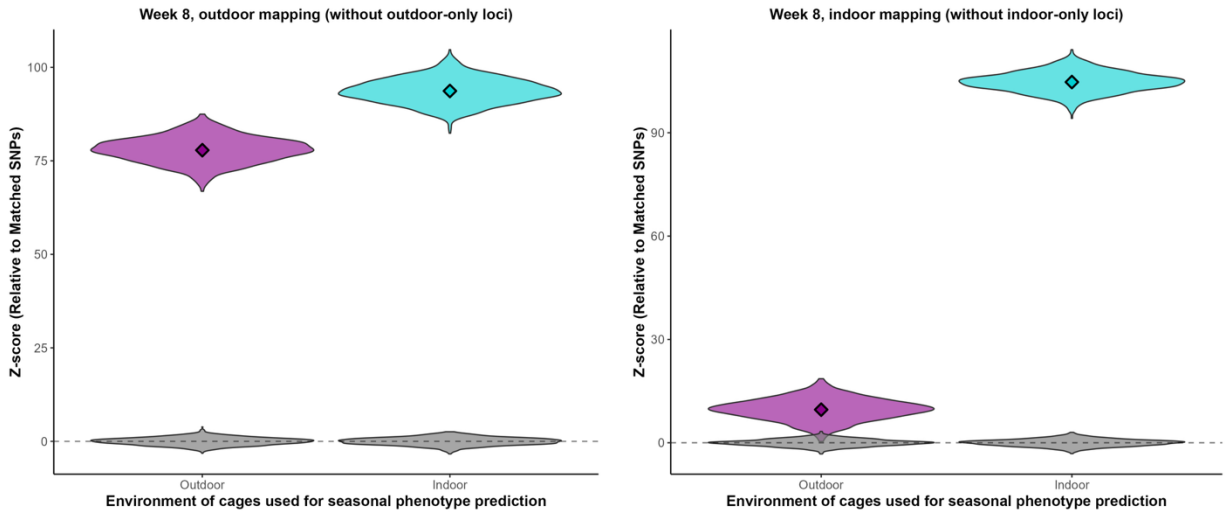
Supplementary Figure 14 | Leave-one-out across the full genome, for all timepoint/treatment combinations. Results from leave-one-out validation of mapping within and across treatments and time points for all timepoint/treatment combinations. Depicted are the median allele frequency difference for each test cage between light and dark samples of trait-associated SNPs (red), as well as a set of matched control SNPs (grey).



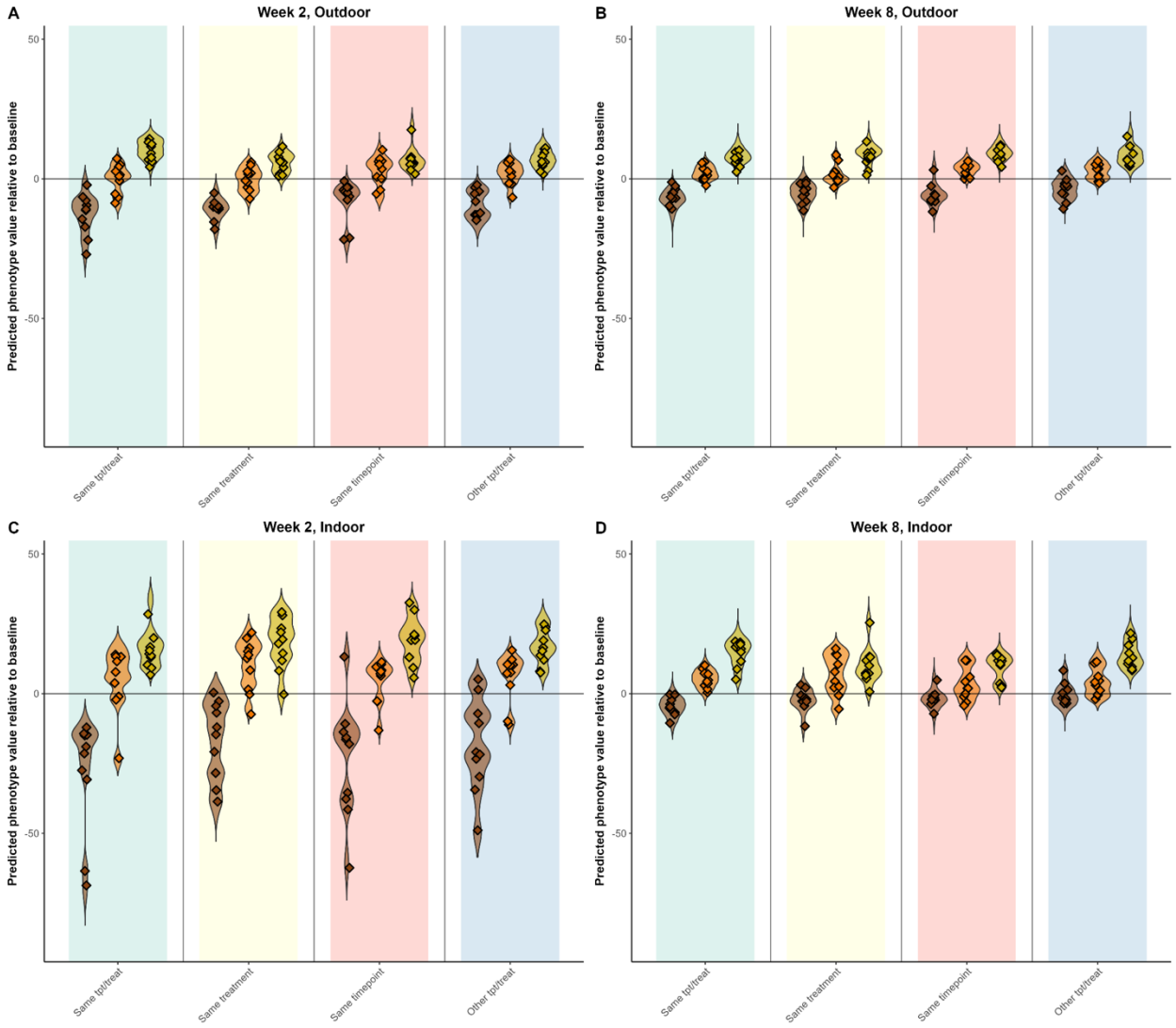
Supplementary Figure 15 | Evidence of field-specific pigmentation-associated loci. Distribution of light-dark allele frequencies (boxes denote the 50% interquartile range, dark horizontal lines correspond to the median, and points indicate outliers) for pigmentation associated SNPs located within mapped loci occurring the outdoor, week 8 mapping population that share no overlapping breakpoints with any loci in the indoor, week 8 mapping population. The magenta distribution corresponds to allele frequency differences within the outdoor cages, while the turquoise to the indoor environment. Grey distributions then represent the light-dark allele frequency differences at matched control SNPs for each environment. The light-dark frequency differences are significantly greater than the matched control shifts for the outdoor environment (Mann-Whitney U test p-value = 0.003), and significantly less than matched control SNPs for the indoor environment (Mann-Whitney U test p-value < 0.001). Furthermore, the outdoor and indoor frequency differences are significantly different than each other (Wilcoxon Signed Rank Test p-value < 0.001), indicating that light-associated alleles in outdoor environment are putatively enriched in the dark color fraction of the indoor enviornment for this set of loci.



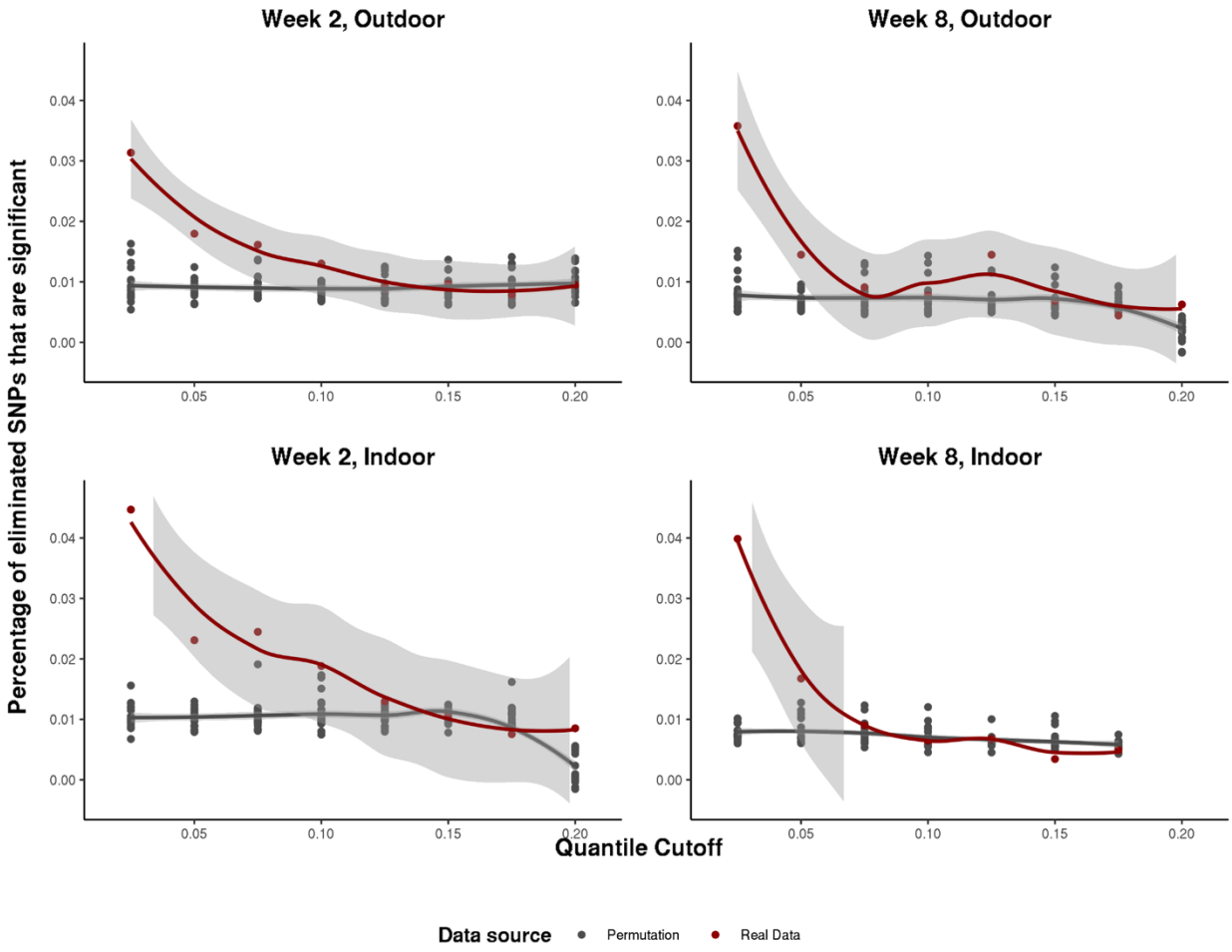
Supplementary Figure 16 | Evidence of lab-specific pigmentation-associated loci. Distribution of light-dark allele frequencies (boxes denote the 50% interquartile range, dark horizontal lines correspond to the median, and points indicate outliers) for pigmentation associated SNPs located within mapped loci occurring the indoor, week 8 mapping population that share no overlapping breakpoints with any loci in the outdoor, week 8 mapping population. The magenta distribution corresponds to allele frequency differences within the outdoor cages, while the turquoise to the indoor environment. Grey distributions then represent the light-dark allele frequency differences at matched control SNPs for each environment. The light-dark frequency differences are significantly greater than the matched control shifts for the indoor environment (Mann-Whitney U test p-value < 0.001), and significantly less than matched control SNPs for the outdoor environment (Mann-Whitney U test p-value < 0.001). Furthermore, the indoor and outdoor frequency differences are significantly different than each other (Wilcoxon Signed Rank Test p-value < 0.001), indicating that light-associated alleles in indoor environment are putatively enriched in the dark color fraction of the outdoor enviornment for this set of loci.



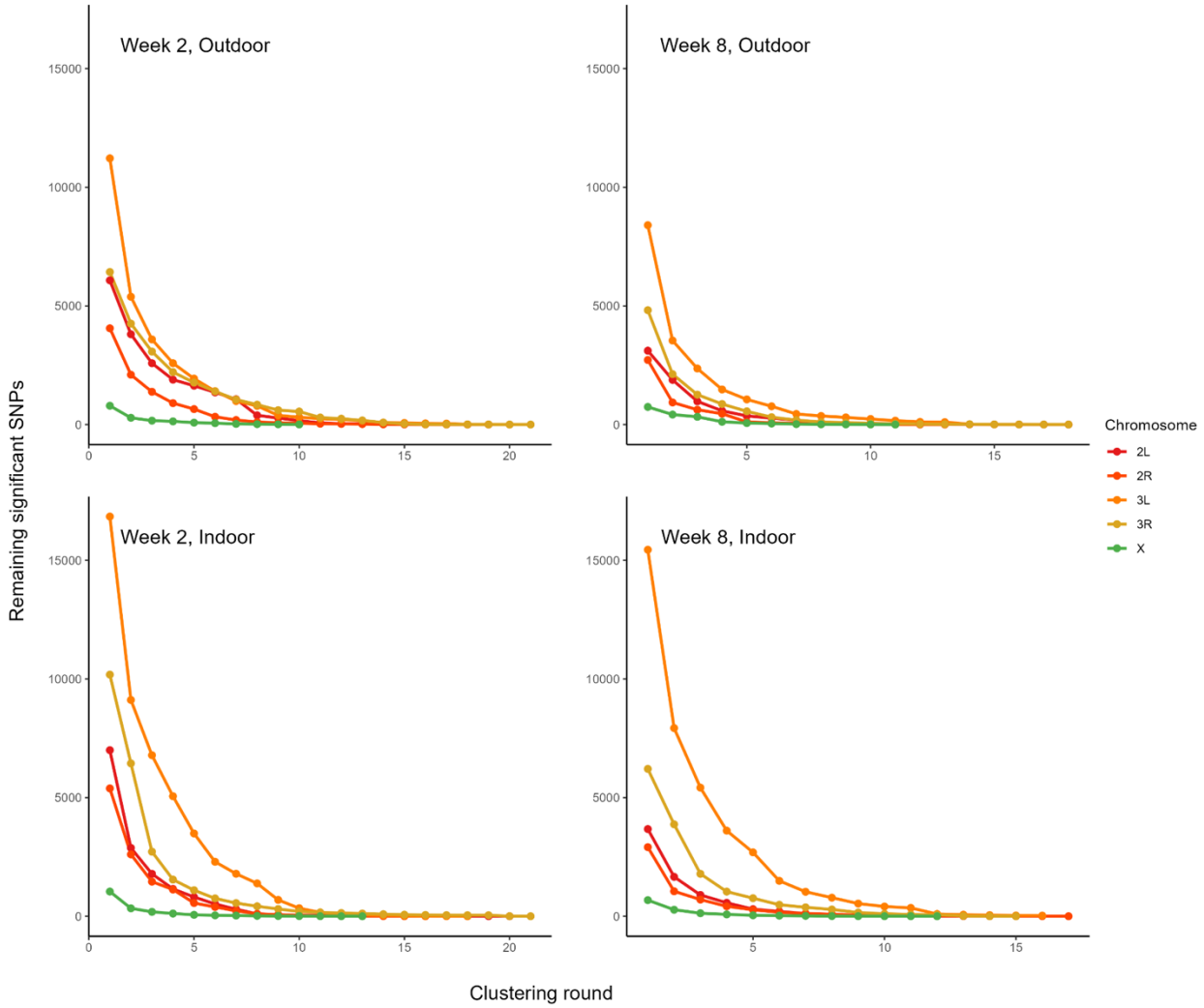
Supplementary Figure 17 | Phenotypic prediction excluding treatment-specific loci. Relative, predicted phenotypic change between Weeks 1 and 8 in both outdoor (magenta) and indoor (turquoise) environments, generated using our trait mapping results derived at weeks 8 in the outdoor (left) and indoor environment, and using variation from every mapped locus excluding those that are identified only in the outdoor cages at week 8 or identified only in the indoor cages at week 8. Predicted phenotypic change are quantified as Z-scores (computed relative to matched control shifts) whereby positive Z-scores indicate net lighter pigmentation evolution while negative values would indicate net darker evolution. Z-scores were computed by comparing inferred relative phenotypic change using mapped SNPs to a that computed using a set of matched control SNPs (grey distributions).



Supplementary Figure 18 | Phenotypic predictions of color samples across timepoint and treatment. Using the mapping from each timepoint/treatment subset and considering all significant SNPs, we generate a phenotypic prediction score for each color fraction sample. A negative value indicates an expected phenotypic shift towards dark pigmentation and a positive value a shift towards lighter pigmentation. We consider our prediction "correct" when the majority of dark samples (dark brown) have a negative predicted value, the majority of light samples (gold) have a positive one, and a linear regression (where we encode dark = 0, midpoint = 1, and light = 2) is significant and has a positive slope. We predict the phenotype of all samples from either within the same timepoint/treatment (green), the same treatment but the other timepoint (yellow), the same timepoint but the other treatment (red), or from both the other timepoint and treatment (blue).



Supplementary Figure 19 | Percentage of top SNPs eliminated in first round of unlinked locus identification with different quantile cutoffs. In our clustering method, the SNPs with the highest light-dark difference as predicted by haplotype frequencies in the highest signal region are eliminated sequentially. Here, we look at the percentage of the eliminated SNPs that are significant SNPs in the first round of clustering with cutoffs at different quantiles using real data versus permutations.



Supplementary Figure 20 | Signal elimination throughout clustering. Plot of the number of remaining significant SNPs before each round of clustering for each chromosome, showing the elimination of linked significant SNPs throughout the clustering process.

Supplementary Tables

Table S1 | List of pigmentation (Pig) and cuticle-associated genes located within mapped loci identified in the week 8, outdoor mapping population that are enriched in pigmentation-associated, mapped SNPs (binomial test, FDR < 0.05). Also reported is the locus ID (corresponding to the locus information provided in Supplementary Data File 1) within which each gene resides, the distance of each gene to the locus midpoint SNP, the number of SNPs within or within 10 Kb of the gene, the proportion of SNPs within the gene that are pigmentation associated (Sig. SNP proportion), and the Benjamini-Yekutieli false discovery rate corrected p-value for observed enrichment of pigmentation-associated SNPs.

Locus ID	FBID	Gene Name	Pig GO?	Cuticle GO?	Distance to midpoint SNP (Mb)	Total SNPs in gene ± 10Kb	Sig. SNP Proportion	FDR
TP8.E.2L.3	FBgn0003255	rickets		✓	0.50	956	0.0251	0.0019
TP8.E.2R.1	FBgn0002533	Larval cuticle protein 2		✓	0.12	330	0.0636	3.13×10^{-9}
	FBgn0002534	Larval cuticle protein 3		✓	0.12	322	0.0621	1.27×10^{-8}
	FBgn0002535	Larval cuticle protein 4		✓	0.12	315	0.0603	5.15×10^{-8}
	FBgn0002531	Larval cuticle protein 1		✓	0.13	344	0.0436	1.10×10^{-4}
	FBgn0002567	lightoid	✓		0.62	422	0.0284	0.0383
TP8.E.3L.1	FBgn0004870	bric a brac 1	✓		0.55	1951	0.1630	2.80×10^{-265}
	FBgn0035281	Cuticular protein 62Bc		✓	1.36	601	0.0499	1.44×10^{-10}
	FBgn0035280	Cuticular protein 62Bb		✓	1.35	577	0.0433	1.20×10^{-7}
	FBgn0035279	Cuticular protein 62Ba		✓	1.34	634	0.0426	4.36×10^{-8}
TP8.E.3L.2	FBgn0035398	Cht7		✓	0.43	772	0.0311	6.90×10^{-5}
TP8.E.3L.5	FBgn0035512	Cuticular protein 64Ac		✓	0.10	511	0.1683	1.66×10^{-72}
	FBgn0035511	Cuticular protein 64Ab		✓	0.10	508	0.1673	1.64×10^{-71}
	FBgn0035513	Cuticular protein 64Ad		✓	0.10	478	0.1611	1.62×10^{-63}
	FBgn0035510	Cuticular protein 64Aa		✓	0.10	551	0.1506	5.41×10^{-66}
TP8.E.3R.2	FBgn0000527	ebony	✓	✓	0.00	625	0.1296	3.55×10^{-59}
TP8.E.3R.3	FBgn0038714	Cuticular protein 92A		✓	0.53	578	0.0329	3.36×10^{-4}
TP8.E.X.1	FBgn0086367	tan	✓		0.01	382	0.1545	1.74×10^{-47}

Table S2 | List of pigmentation (Pig) and cuticle-associated genes located within mapped loci identified in the week 8, indoor mapping population that are enriched in pigmentation-associated, mapped SNPs (binomial test, FDR < 0.05). Also reported is the locus ID (corresponding to the locus information provided in Supplementary Data File 1) within which each gene resides, the distance of each gene to the locus midpoint SNP, the number of SNPs within or within 10 Kb of the gene, the proportion of SNPs within the gene that are pigmentation associated (Sig. SNP proportion), and the Benjamini-Yekutieli false discovery rate corrected p-value for observed enrichment of pigmentation-associated SNPs.

Locus ID	FBID	Gene Name	Pig GO?	Cuticle GO?	Distance to midpoint SNP (Mb)	Total SNPs in gene ± 10Kb	Sig. SNP Proportion	FDR
TP8.I.2L.3	FBgn0053302	Cuticular protein 31A		✓	0.03	642	0.0685	1.05×10^{-14}
TP8.I.2R.1	FBgn0002534	Larval cuticle protein 3		✓	0.43	322	0.1770	1.58×10^{-40}
	FBgn0002535	Larval cuticle protein 4		✓	0.44	315	0.1619	3.09×10^{-34}
	FBgn0002533	Larval cuticle protein 2		✓	0.43	330	0.1485	4.15×10^{-31}
	FBgn0002531	Larval cuticle protein 1		✓	0.43	344	0.1279	3.87×10^{-25}
TP8.I.3L.1	FBgn0000543	ecdysoneless		✓	0.69	535	0.0897	1.08×10^{-20}
	FBgn0035279	Cuticular protein 62Ba		✓	0.25	634	0.0757	1.08×10^{-17}
	FBgn0035280	Cuticular protein 62Bb		✓	0.26	577	0.0676	7.67×10^{-13}
	FBgn0022702	Chitinase 2		✓	0.18	350	0.0657	2.17×10^{-7}
	FBgn0035281	Cuticular protein 62Bc		✓	0.27	601	0.0416	1.74×10^{-4}
TP8.I.3L.2	FBgn0035875	Cuticular protein 66Cb		✓	0.30	429	0.0559	1.82×10^{-6}
TP8.I.3L.5	FBgn0004870	bric a brac 1	✓		0.33	1951	0.1978	1.44×10^{-295}
TP8.I.3L.6	FBgn0261797	Dynein heavy chain 64C		✓	0.23	782	0.0665	7.88×10^{-17}
TP8.I.3R.2	FBgn0027508	tankyrase		✓	0.34	537	0.0503	2.38×10^{-6}
TP8.I.3R.3	FBgn0000527	ebony	✓	✓	0.07	625	0.0432	3.97×10^{-5}
	FBgn0038819	Cuticular protein 92F		✓	0.49	464	0.0409	0.0024
TP8.I.3R.6	FBgn0038901	Bursicon	✓	✓	0.36	325	0.0554	8.00×10^{-5}
TP8.I.X.1	FBgn0086367	tan	✓		0.01	382	0.1440	3.10×10^{-34}

Supplementary Discussion

Novel pigmentation-associated loci identified by high-powered, tail-based mapping

Here we describe in further detail genes occurring within our mapped loci that likely play a functional role in pigmentation variation. As nearly 80% of total genes in the genome were located within the start and endpoints of the loci identified across all mapping populations, we concentrate this discussion on those genes with a gene ontology (GO) term associated with pigmentation (excluding eye-specific pigmentation GO terms) and/or with cuticle development and molting, as well as confine our assessment to mapping conducted at week 8. Additionally, for each putative candidate gene, we ensured that the proportion of mapping significant SNPs within or nearby (within 10 Kb) gene breakpoints were significantly greater than expected based on the genome-wide distribution of significant SNPs (binomial test).

As expected, *bric-a-brac* (*bab*), *ebony*, and *tan* each fell within mapped loci and contained a significant enrichment of pigmentation-associated SNPs. *Ebony* and *tan* are opposing enzymes that catalyze dopamine to a precursor of sclerotin (N-β-alanyl-dopamine, NBAD) and NBAD back to dopamine, respectively (Wright, 1987). Thus, together they control the amount of NBAD available to create light pigmented sclerotin instead of darkly pigmented melanin. In contrast, *bab* is a set of two duplicated transcription factors (*bab1* and *bab2*) that repress pigmentation (Couderc et al., 2002). For example, *bab* is expressed in tergites A2 through A6 in female *D. melanogaster,* but only A2 through A4 in males, giving rise to the characteristic, male-only, completely pigmented terminal tergites for which *D. melanogaster* is named (Kopp et al., 2000). Binding sites for *Abdominal-B* (*Abd-B*) and *doublesex* (dsx) are located in the long intron of bab1, and the binding of *doublesex* to *bab1* mediates this sex-specific pigmentation (Williams et al., 2008). Additionally, previous studies have shown that the regulation of both *bab* and *tan* is influenced by temperature, providing a likely explanation for the plasticity of pigmentation across developmental temperatures observed in *D. melanogaster (Castro et al., 2018; Gibert et al., 2016)*. For *bab*, the *hox* gene and transcription factor Abd-B has been shown to mediate this temperature-dependent expression (Gibert et al., 2007). We also identify Abd-B in our study which is located within the top locus on chromosomal arm 3R.

On 2R, our largest peak (which shows high consistency across timepoints and treatments), overlies larval cuticle proteins 1, 2, and 3 (*Lcp1*, *Lcp2*, *Lcp3*) – three out of four larval cuticle protein genes that form a set of two pairs of genes (*Lcp1* and *Lcp2*, and *Lcp3* and *Lcp4*) that each act as a separate transcriptional unit and are transcribed at different periods during the third instar stage (Kimbrell et al., 1988). *Lcp3* and *Lcp4* accumulate first, followed by *Lcp1* and *Lcp2*, and this is likely part of a set of changes that shift the third instar soft cuticle to a harder, less flexible one for the wandering stage (Kimbrell et al., 1988). Even though the larval cuticle is not preserved into the adult stage, certain pigmentation genes (namely yellow) that are important for adult pigmentation are also expressed and influence pigmentation in the larval stages (Kornezos and Chia, 1992).

We find further evidence of pigmentation-associated SNPs in loci concentrated in and around *Bursicon* (*Burs*), *rickets* (*rk*), and *Partner of bursicon* (*Pburs*). These three genes are the core genes vital to the cuticular hardening after eclosion. *Burs* and *Pburs* dimerize and bind to rickets to initialize the final tanning or hardening of the cuticle (Baker and Truman, 2002; Dewey et al., 2004). Prior to this, the eclosed fly is entirely white, with very little visible pigmentation.

Tanning then occurs when deposited cuticular proteins are linked together and the pigmentation becomes stark and visible. Despite this direct connection to pigmentation of the cuticle, this mapping study represents the first to directly identify pigmentation-relevant variation in these genes.

In only the outdoor week 8 mapping population, we find more significant SNPs than expected around and within a gene on chromosomal arm 3L that encodes for a transmembrane chitinase, *Chitinase 7* (*Cht7*) (Zhu et al., 2008). It appears that *Cht7* is primarily involved in chitin organization in the developing cuticle, as RNAi knockdown of *Cht7* has shown the old cuticle to be disassembled successfully, but the density and organization of the new cuticle is compromised (Noh et al., 2018). Another chitinase gene, *Chitinase 2* (*Cht2*), is nearby *Cht7* on 3L, but only exhibited a significant enrichment of significant SNPs in the indoor mapping population. *Cht2*, like *Cht7*, seems to play a role in organizing the developing cuticle and not in disassembling the old cuticle (Pesch et al., 2017).

In the indoor week 8 mapping population, we identified loci with high signal in and around *tankyrase*, *Dynein heavy chain 64c* (*Dhc64c*), and *ecdysoneless*. Both *tankyrase* and *Dhc64c* are involved in the proper expression and localization of *Wg* (Feng et al., 2014; Kim et al., 2021). *Wg* and other components in, or affected by, the *Wnt* pathway are known to be associated with pigmentation differences within and between species (Massey and Wittkopp, 2016). *Ecdysoneless*, while it does affect the production of ecdysone as its name implies, actually has a broader role in correctly localizing pre-mRNA processing factor 8 (*Prp8*), a highly conserved splicing factor (Claudius et al., 2014; Erkelenz et al., 2021).

Supplementary References:

Baker, J.D., Truman, J.W., 2002. Mutations in the Drosophila glycoprotein hormone receptor, rickets, eliminate neuropeptide-induced tanning and selectively block a stereotyped behavioral program. J Exp Biol 205, 2555–2565. https://doi.org/10.1242/jeb.205.17.2555

Castro, S.D., Peronnet, F., Gilles, J.-F., Mouchel-Vielh, E., Gibert, J.-M., 2018. bric à brac (bab), a central player in the gene regulatory network that mediates thermal plasticity of pigmentation in Drosophila melanogaster. PLOS Genetics 14, e1007573. https://doi.org/10.1371/journal.pgen.1007573

Claudius, A.-K., Romani, P., Lamkemeyer, T., Jindra, M., Uhlirova, M., 2014. Unexpected Role of the Steroid-Deficiency Protein Ecdysoneless in Pre-mRNA Splicing. PLOS Genetics 10, e1004287. https://doi.org/10.1371/journal.pgen.1004287

Couderc, J.-L., Godt, D., Zollman, S., Chen, J., Li, M., Tiong, S., Cramton, S.E., Sahut-Barnola, I., Laski, F.A., 2002. The bric à brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. Development 129, 2419–2433. https://doi.org/10.1242/dev.129.10.2419

Dewey, E.M., McNabb, S.L., Ewer, J., Kuo, G.R., Takanishi, C.L., Truman, J.W., Honegger, H.-W., 2004. Identification of the Gene Encoding Bursicon, an Insect Neuropeptide Responsible for

Cuticle Sclerotization and Wing Spreading. Current Biology 14, 1208–1213. https://doi.org/10.1016/j.cub.2004.06.051

Erkelenz, S., Stanković, D., Mundorf, J., Bresser, T., Claudius, A.-K., Boehm, V., Gehring, N.H., Uhlirova, M., 2021. Ecd promotes U5 snRNP maturation and Prp8 stability. Nucleic Acids Res 49, 1688–1707. https://doi.org/10.1093/nar/gkaa1274

Feng, Y., Li, X., Ray, L., Song, H., Qu, J., Lin, S., Lin, X., 2014. The Drosophila tankyrase regulates Wg signaling depending on the concentration of Daxin. Cell Signal 26, 1717–1724. https://doi.org/10.1016/j.cellsig.2014.04.014

Gibert, J.-M., Mouchel-Vielh, E., Castro, S.D., Peronnet, F., 2016. Phenotypic Plasticity through Transcriptional Regulation of the Evolutionary Hotspot Gene tan in Drosophila melanogaster. PLOS Genetics 12, e1006218. https://doi.org/10.1371/journal.pgen.1006218

Gibert, J.-M., Peronnet, F., Schlötterer, C., 2007. Phenotypic Plasticity in Drosophila Pigmentation Caused by Temperature Sensitivity of a Chromatin Regulator Network. PLOS Genetics 3, e30. https://doi.org/10.1371/journal.pgen.0030030

Kim, J.-Y., Tsogtbaatar, O., Cho, K.-O., 2021. Dynein Heavy Chain 64C Differentially Regulates Cell Survival and Proliferation of Wingless-Producing Cells in Drosophila melanogaster. Journal of Developmental Biology 9, 43. https://doi.org/10.3390/jdb9040043

Kimbrell, D.A., Berger, E., King, D.S., Wolfgang, W.J., Fristrom, J.W., 1988. Cuticle protein gene expression during the third instar of *Drosophila melanogaster*. Insect Biochemistry 18, 229–235. https://doi.org/10.1016/0020-1790(88)90087-X

Kopp, A., Duncan, I., Carroll, S.B., 2000. Genetic control and evolution of sexually dimorphic characters in Drosophila. Nature 408, 553–559. https://doi.org/10.1038/35046017

Kornezos, A., Chia, W., 1992. Apical secretion and association of the Drosophila yellow gene product with developing larval cuticle structures during embryogenesis. Molec. Gen. Genet. 235, 397–405. https://doi.org/10.1007/BF00279386

Massey, J., Wittkopp, P.J., 2016. The genetic basis of pigmentation differences within and between Drosophila species. Curr Top Dev Biol 119, 27–61. https://doi.org/10.1016/bs.ctdb.2016.03.004

Noh, M.Y., Muthukrishnan, S., Kramer, K.J., Arakane, Y., 2018. A chitinase with two catalytic domains is required for organization of the cuticular extracellular matrix of a beetle. PLOS Genetics 14, e1007307. https://doi.org/10.1371/journal.pgen.1007307

Pesch, Y.-Y., Riedel, D., Behr, M., 2017. *Drosophila* Chitinase 2 is expressed in chitin producing organs for cuticle formation. Arthropod Structure & Development, Arthropod Cuticle Function: from Molecules to Structures to Ecology 46, 4–12. https://doi.org/10.1016/j.asd.2016.11.002

Williams, T.M., Selegue, J.E., Werner, T., Gompel, N., Kopp, A., Carroll, S.B., 2008. The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in Drosophila. Cell 134, 610–623. https://doi.org/10.1016/j.cell.2008.06.052

Wright, T.R.F., 1987. The Genetics Of Biogenic Amine Metabolism, Sclerotization, And Melanization In *Drosophila Melanogaster**, in: Scandalios, J.G., Caspari, E.W. (Eds.), Advances in Genetics, Molecular Genetics of Development. Academic Press, pp. 127–222. https://doi.org/10.1016/S0065-2660(08)60008-5

Zhu, Q., Arakane, Y., Banerjee, D., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., 2008. Domain organization and phylogenetic analysis of the chitinase-like family of proteins in three species of insects. Insect Biochemistry and Molecular Biology, Special Issue on the Tribolium Genome 38, 452–466. https://doi.org/10.1016/j.ibmb.2007.06.010