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], 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

Correspondence: 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, producing a well-resolved genotype-phenotype map that reaffirms canonical pigmentation genes and unveils novel loci. 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 do not alter trait architecture and preclude 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 [1,2]. 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 [3–9]. 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 [10,11]. Quantifying, and ultimately predicting, how these traits evolve remains a challenge.

Two facets of complex trait variation that may preclude evolutionary predictability are pleiotropy and epistasis. Pleiotropy occurs when a single genomic locus impacts phenotypic variation in multiple traits [12,13]. 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 [14]. 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 actually evolved through time [10,13,15,16].

Epistasis is the interaction of allelic effects at multiple loci and, in the context of predicting trait adaptation, may shift the genotype-phenotype map itself throughout the evolutionary process [17–20]. 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 [17,18,20–28]. Similarly, context-dependent dominance may alter the phenotypic effect of alleles as a population evolves to shifting environmental conditions [29]. 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, an assumption readily applied to approaches using contemporary mapping data to infer past trait evolution [30–34].

In concert, epistasis and pleiotropy may lead to a genetic architecture of adaptation that is not entirely predictable based on the genotype-phenotype map 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 genomic loci throughout evolution itself, represent a powerful method to study pleiotropy, epistasis, and the predictability of complex trait evolution [35–39]. 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 of selection?

Our study comprised a longitudinal analysis of abdominal pigmentation evolution in *Drosophila melanogaster* within an ecologically realistic context. We selected pigmentation as our focal complex trait because of its strong association with phenotypic diversity across the tree of life and implication in adaptive evolution across a range of taxa [40]. For *D. melanogaster* in particular, abdominal pigmentation exhibits latitudinal clines among populations, and rapidly and cyclically evolves in response to seasonally varying selection pressures [41]. 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 [42,43]. 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* [44–46]. Interestingly, clines in pigmentation across space and time do not co-occur with expected signatures of selection at the core pigmentation-associated loci [41]. This pattern is suggestive of underlying pleiotropy and/or epistasis, but impossible to disentangle fully from the confounding effects of demography and linkage among wild populations.

Here, using a single, genetically diverse founding population we quantified pigmentation evolution in mesocosms housed in two environments with distinct selective pressures: a simple, lab-based environment and a dynamic, semi-natural outdoor setting. 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 pigmentation, separately for replicate populations evolving in each environment (Fig. 1). The genotype-phenotype maps we derived from this process 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, outdoor environment. 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 the 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 [24,26]. Furthermore, the pigmentation and fitness-relevant variation subject to selection in our study was evaluated in the same local region where it is maintained naturally and in which we have documented repeatable seasonal evolution of lighter pigmentation from late spring through summer [41]. Finally, the haplotype-informed allele frequency estimates we generated from the reconstituted outbred populations are of the accuracy produced using standard pool-seq methods and sequencing depths of roughly 1000x [47]. Such haplotype-imputed methods of mapping are significantly better powered than more standard, SNP-by-SNP GWA approaches (Long et al. 2025).

The outcrossed population recombined for four generations before being used to seed each of 10 indoor cage replicates maintained in controlled, lab-based conditions [48], as well as 10 replicate outdoor field cages comprising a semi-natural mesocosm exposed to natural environmental fluctuations in Philadelphia, PA (Fig. 1) [48–50]. 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. The replicate populations then expanded and evolved 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. Samples were then used in whole genome sequencing, and the resulting reads subject to a bioinformatic pipeline that leverages the known haplotype structure from our founding set of individuals to compute genome-wide haplotype and allele frequency data in the evolved, pooled samples [47,51]. 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 1.9 M SNPs previously identified in our inbred reference panel.

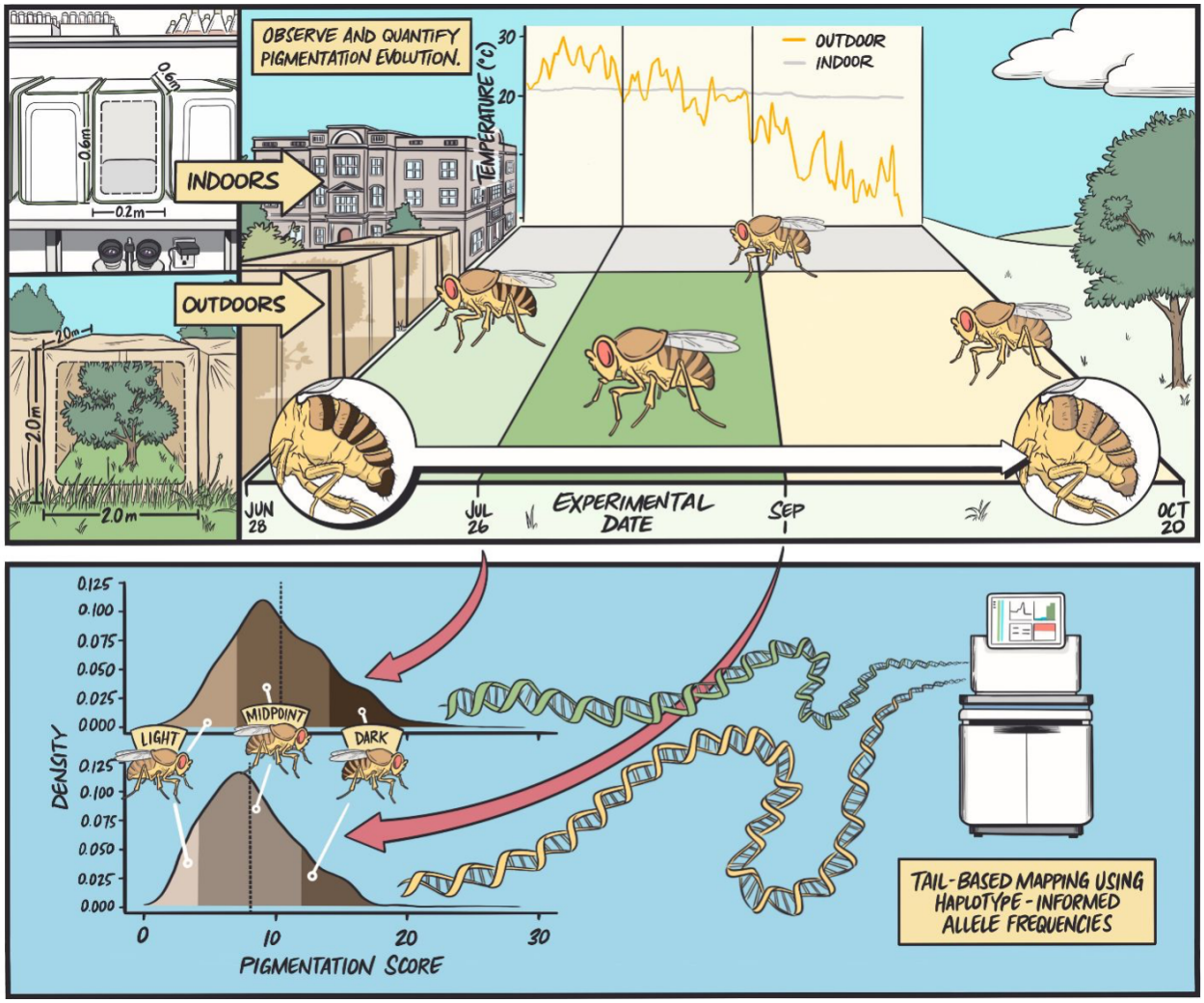


Figure 1 | Experimental design. Ten cages each in an indoor laboratory and in an outdoor, field mesocosm were seeded with an outbred population of *D. melanogaster*. The replicate populations were allowed to evolve from late June to late October, during which time significant pigmentation evolution was observed in each environment (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). Each color fraction from each replicate cage, environment, 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 in the derived outbred population (hereafter, 'baseline'), and among replicate populations at weeks two and eight of evolution in each environment. As expected, our baseline population exhibited variation in pigmentation that was representative of the variation observed across the 76 inbred lines from which it was derived (Fig. 2A). Pigmentation after two generations of evolution resulted in significantly lighter pigmentation across replicates in both the lab and field environments, relative to the baseline (pairwise t-tests, week 2 outdoor $p << 0.001$, week 2 indoor $p << 0.001$). Also as expected, the replicate populations then evolved to become significantly lighter during the six additional weeks of evolution in both the outdoor and indoor environments (pairwise t-tests; week 2 vs. 8 outdoor $p << 0.001$, week 2 vs. 8 indoor $= 0.0245$). Notably, a significant treatment by time interaction on variation in pigmentation throughout the experiment was observed (two-way ANOVA; $F = 17.03$, $p << 0.001$), driven by the accelerated rate and magnitude of lighter pigmentation evolution in the outdoor replicates (Fig. 2A).

Next, we quantified whether there existed evidence of environment-specific patterns of selection that would, in effect, induce unique epistatic landscapes across lab and field-based mapping populations. Specifically, we ran principal component analysis (PCA; *prcomp* function, R), on the normalized allele frequency data of all 1.9 M SNPs, and then projected samples onto the first two principal components (Fig. 2B-C). Strikingly, the first principal component (6.4% variance explained) segregated samples in accordance with trait group, and along the second principal component (4.8% variance explained) in accordance with environment of selection (Fig. S1). When segregated by chromosome, we observed the same correlation between PC1 and color fraction on all chromosomes except X, and all chromosomes except 2R demonstrated the same relationship between PC2 and treatment (Fig. S2). These patterns were not disrupted when the major cosmopolitan inversions were excluded (Fig. S3). This analysis thus demonstrates the pronounced impact of both phenotypic group and environment of selection on genome wide patterns of variation, while also suggesting that a shared set of loci underpinned variation in pigmentation across mapping populations.

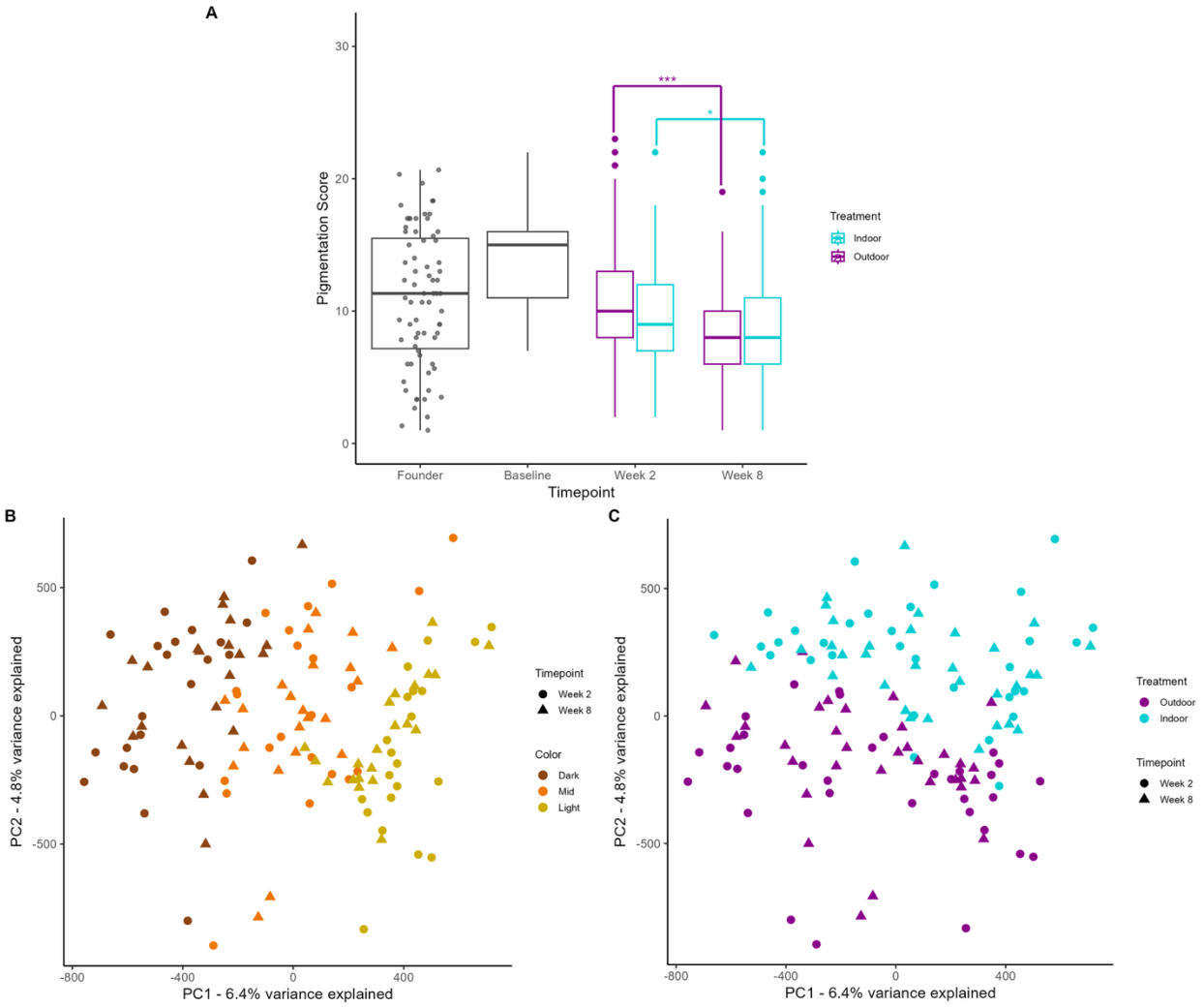


Figure 2 | Phenotypic and genomic variation throughout study period. (A) Distribution of the total pigmentation scores (whereby higher scores indicate darker pigmentation) across samples. "Founder" refers to our 76 inbred lines that we outbred for four generations to generate our baseline (week 0) population. 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).

194

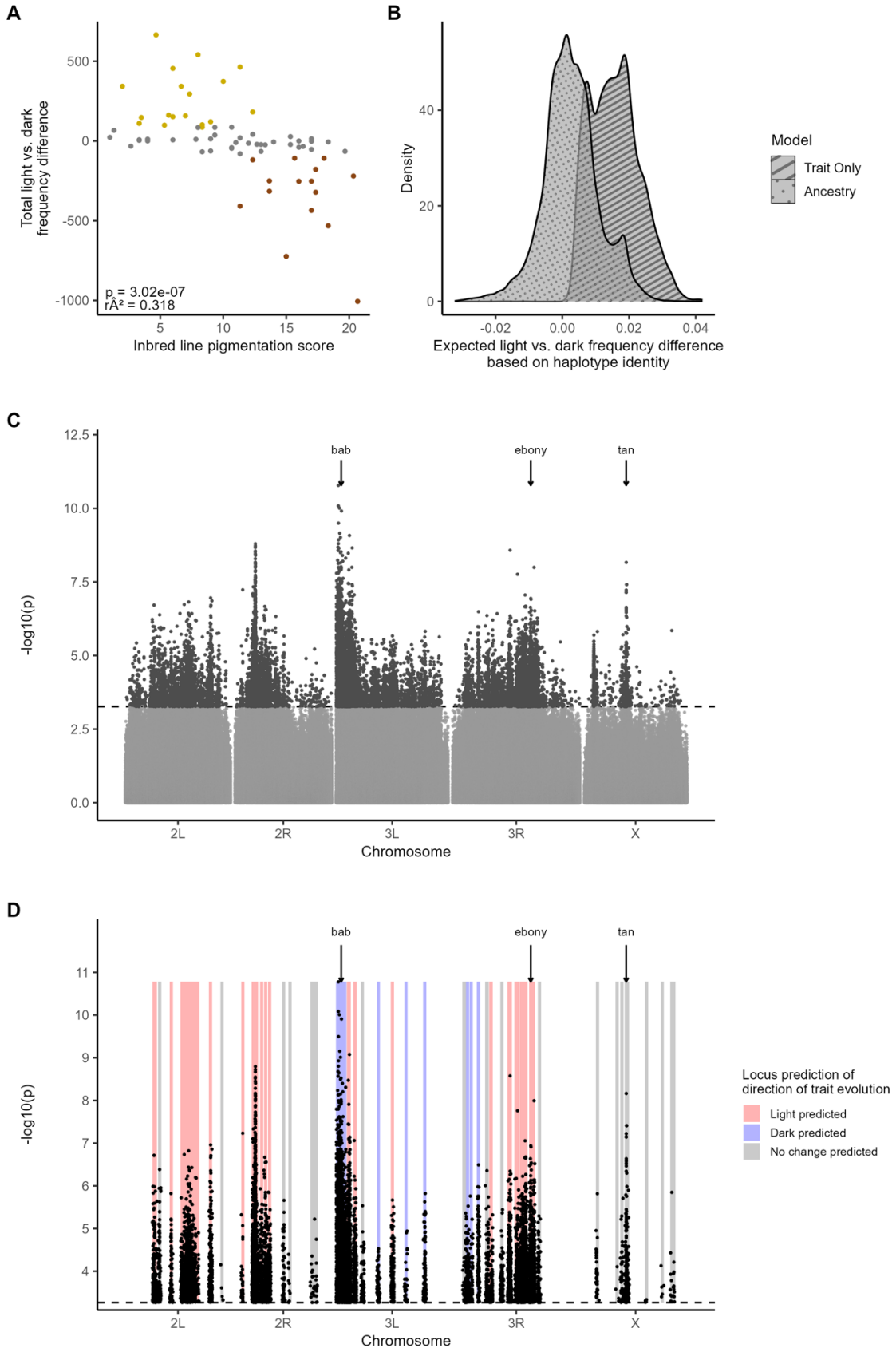
195 *Ancestry-informed mapping provides resolved architecture of abdominal pigmentation*

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

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

228 and re-running our ancestry-informed GLM (N = 100 permutations). A SNP was considered significant if
229 its nominal p-value from the empirical GLM fell in the top 0.05% of permuted p-values (corresponding to
230 an empirical FDR threshold < 0.0005; Fig. S5). Even at this conservative threshold, we see that trait-
231 associated SNPs are pervasive throughout the genome and distributed across all chromosomal arms (Fig
232 3C). Hereafter, SNPs identified via this approach are referred to as either light or dark-associated in
233 accordance with their relative frequency in the light and dark color fractions – e.g., the light-associated
234 allele for each mapped SNP is that occurring at higher relative frequency in the light color fraction.
235 We next used our set of pigmentation-associated SNPs to identify the number and location of
236 unlinked loci underpinning variation in the trait. Our procedure to do so was as follows: we first identified
237 the most significant SNP on each chromosomal arm, as well as its surrounding, locally un-recombined
238 region (defined by the recombination rate and number of generations since cage founding, see Methods).
239 We then eliminated all trait-associated SNPs on the chromosome whose relative frequency in the light
240 and dark fractions could be explained by haplotype frequencies within the locally un-recombined focal
241 region alone. In effect, this eliminated all GLM-signal due to linkage to the focal SNP. From the set of
242 remaining SNPs, we identified the SNP with the most significant nominal p-value and repeated this
243 procedure. This was iterated until we exhausted all trait-associated SNPs on the chromosome and resulted
244 in a total of 76 genomic loci, the midpoint of which was the most significant, locus-defining SNP.
245 Merging loci with overlapping start and endpoints, we obtained a final list of 42 loci, ranging in size from
246 700 Kb to 3 Mb, and distributed across all chromosomal arms (Supplementary Data File 1). This indicates
247 that the genetic architecture of our focal, ecologically relevant complex trait, is not underpinned solely by
248 a single or few large-effect loci, nor exclusively by a very large number of weak-effect loci (e.g., 1000's).
249 Rather, we infer the presence of a moderate number of large to moderate-effect loci underpinning trait
250 variation, while acknowledging there may be further loci of small effect that our approach lacks sufficient
251 power to detect. We once again validated that pigmentation-associated SNPs within the identified loci
252 were not exclusive to the lightest and darkest haplotypes enriched in the trait tails (Fig. 3A), but rather
253 shared across all founding inbred lines (Fig. S6).
254 We quantified the extent to which our identified loci validated previously described, and
255 potentially identified new, candidate genes associated with pigmentation variation in the species.
256 Assuredly, the top loci on chromosomal arms 3L, 3R, and X contained the previously described core
257 pigmentation genes: *bab*, *ebony*, and *tan* [42,52–54]. A systematic comparison of the association signal
258 between loci identified here and two recent, well-powered GWAS of the trait [45,46] indicated significantly
259 more overlap than expected by chance (permutation $p < 0.001$, see Methods). However, this signal is
260 eliminated when we no longer consider SNPs within the loci containing *bab*, *ebony*, or *tan* ($p = 0.994$)
261 suggesting little overlap outside of these three core genes. Thus, the extensive enrichment of

262 pigmentation-associated variation outside these candidate loci we observe in our data suggests the
263 presence of novel candidate pigmentation loci identified in our study, which we discuss in the supplement
264 (Supplementary Discussion; Table S1-2).
265
266 *Pleiotropy and predictions of complex trait evolution*
267 A potential of application of a well-resolved genetic map is to predict whether or not a trait has
268 evolved using genomic data alone. However, this effort could be obscured by underlying pleiotropy at
269 quantitative trait loci, which may constrain their impact on trait evolution and leave the architecture of
270 trait adaptation largely unpredictable. Such pleiotropic interactions may be particularly pronounced in
271 complex environments where selection across multiple fitness-relevant traits may induces correlated
272 selection [13]. Our longitudinal study design afforded the unique opportunity to test this directly by
273 combining our well-resolved genetic mapping data (Fig. 2C-D) with our time-series allele frequency data
274 collected throughout phenotypic evolution of the trait (Fig. 1; 2A). Accordingly, for each of the 42
275 independent, unlinked loci identified above, we quantified the frequency shifts of the light-associated
276 alleles over the period during which the outdoor replicate populations evolved to become significantly
277 lighter (Fig. 2A). We then inferred whether the observed allele frequency movement within each locus
278 indicated lighter pigmentation evolution by comparing the distribution of shifts to a distribution derived
279 from a set of matched control SNPs (binomial test; see Methods). We found that out of the 42 loci, 18
280 exhibited an overall increase in light-associated allele frequency and thus a behavior concordant with
281 what would be expected given lighter pigmentation evolved. For 17 pigmentation-associated loci,
282 however, we did not detect any systematic shift indicative of trait evolution, while 7 loci showed behavior
283 that indicated the evolution of darker pigmentation (Fig. 3D). Thus, in sum, while our genome-wide
284 inference correctly predicted evolution of lighter pigmentation in the outdoor mesocosm ($p << 0.001$; Fig.
285 4A) we observed that at particular loci, and even for entire chromosomal arms (Fig. S7), the genetic basis
286 of pigmentation evolution was unpredictable based on mapping data alone. The unpredictable genetic
287 architecture of adaptation was also evident at the level of individual loci: allele frequency shifts within the
288 *ebony* locus, as well as the largest effect locus on 2R, predicted evolution of lighter pigmentation; while
289 variation within the locus containing *bab* predicted evolution of darker pigmentation, and variation in the
290 *tan* locus that pigmentation would not evolve (arrows in Fig. 3D). In sum, these observations likely reflect
291 the role of pleiotropy and constraint during complex trait adaptation in natural settings.
292


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
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
Locus prediction of direction of trait evolution
Light predicted
Dark predicted
No change predicted

293

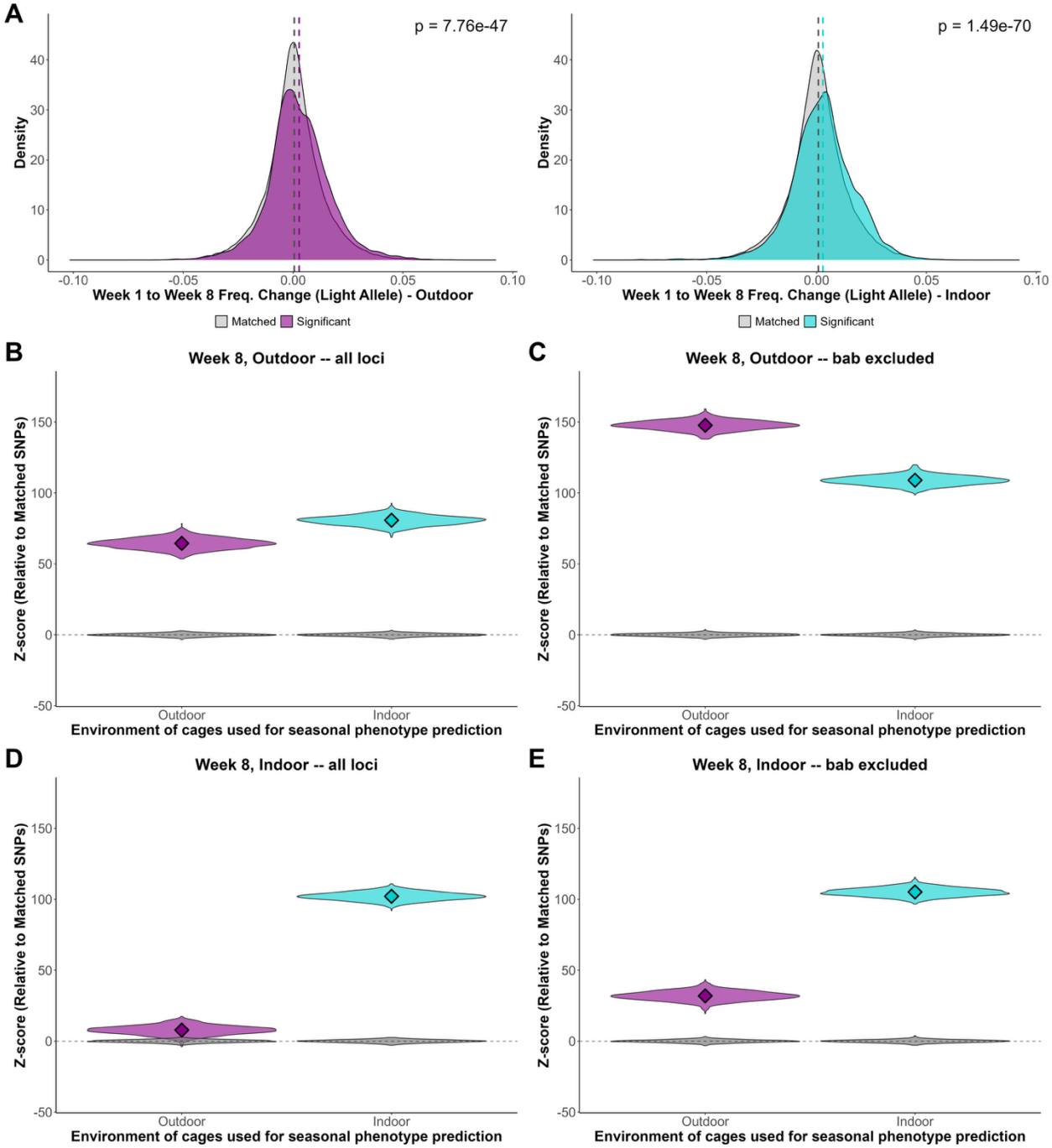
Figure 3 | Genetic architecture of female abdominal pigmentation. (A) The pigmentation score of the founding inbred lines compared with the light vs. dark frequency differences across all samples for the haplotypes originating from these lines. Inbred line pigmentation is highly correlated to the haplotype's 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 mesocosms relative to those that evolved within the indoor, lab-based mesocosms (Figs. 1-2). Accordingly, we first quantified whether light-associated alleles were favored during the evolution of lighter pigmentation in each environment, finding that their genome-wide frequencies increased significantly throughout trait evolution in both the lab and field 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 approach is analogous to polygenic score-based predictions of trait evolution oftentimes implemented in studies of

330 human adaptation [31,32,34], and indeed correctly inferred that lighter pigmentation evolved in both the
331 indoor and outdoor environment (Fig. 4B). Interestingly, however, the relative predicted magnitude of
332 phenotypic change was greater in the indoor environment, despite our phenotypic assays showing that
333 replicate populations in this environment evolved to a lesser degree (Fig 2A). We suspect that such a
334 disconnect between observed patterns of phenotypic evolution and the genomic prediction of the process
335 provides strong evidence of a context-dependent pleiotropy across environments – the more complex
336 environmental conditions in the outdoor mesocosm induces selection on a greater number of traits overall,
337 leading to more pleiotropic constraint at pigmentation-associated loci and poorer genomic prediction.
338 We investigated context-specific pleiotropy further with a focus on the *bab* locus, a canonical
339 pigmentation gene identified by our mapping and for which raw allele frequency trajectories showed
340 evidence of pleiotropic constraint (Fig. 3D). By excluding variation in this locus from our genomic
341 prediction, we observed that the inferred relative magnitude of pigmentation evolution increased in both
342 environments and, strikingly, to a much greater degree in the outdoor environment (Fig 4C). This analysis
343 thus provides a key example of how pleiotropy at a single locus can skew genomic predictions of
344 phenotypic evolution. Furthermore, it elucidates how studying such processes in a natural setting may
345 reveal properties of segregating variation that are masked in simpler, lab-based environments.
346 Interestingly, regenerating these predictions using variation within the canonical pigmentation loci alone
347 led to a spurious prediction of darker evolution in the outdoor environment (Fig. S10), while excluding
348 variation in these loci from our prediction recovered the correct inference of trait evolution (Fig. S11).
349 Thus, the key takeaway of these results is that accurate genomic prediction of trait mapping hinges upon a
350 comprehensive, well-resolved genetic map that includes both the largest-effect loci, as well as those of
351 more moderate or minor impact on trait variation.
352 The genomic predictions of trait evolution presented thus far were based on mapping data
353 procured from week 8, outdoor populations. Accordingly, we next quantified if and how these predictions
354 differed when derived from samples evolved to week 8 in the indoor environment. Specifically, using the
355 same mapping procedure, we found that while this independent set of mapped loci correctly predicts trait
356 evolution in the indoor environment, it fails to predict evolution of lighter pigmentation in the outdoor
357 environment (Fig. 4D). This discrepancy in genomic prediction of trait evolution in the outdoor
358 environment (i.e., comparing magenta distributions in Fig. 4B vs. Fig. 4C) potentially indicates epistasis-
359 driven differences in the genotype-phenotype map that arise between environments of selection.
360 Strikingly, but also consistent with those genomic predictions generated using the outdoor mapping data
361 (Fig. 4B-C), we found that the exclusion of *bab* locus variation allowed us to correctly predict trait
362 evolution in the outdoor environment using the indoor mapping data (Fig. 4E). This further exemplifies

363 the pleiotropic constraints of this large effect locus on trait evolution and the consequences therein on
364 genomic prediction.



365 **Figure 4 | Prediction of phenotypic evolution in simple and complex environments**. (A) Genome-wide allele
366 frequency change at pigmentation-associated SNPs identified in the outdoor, week 8 mapping population and
367 quantified throughout trait evolution in the outdoor (purple) and indoor (blue) mesocosm, relative to matched
368 control shifts (grey distributions). (B) Distribution of Z-scores for bootstrapped estimates of genomic predictions for
369 pigmentation evolution in the outdoor (magenta) and indoor (turquoise) mesocosm using SNPs identified in the

370 outdoor, week 8 mesocosm. Positive Z-scores indicate net lighter pigmentation evolution while negative values
371 would indicate net darker evolution. Z-scores were computed by comparing inferred relative phenotypic change
372 using mapped SNPs to a that computed using a set of matched control SNPs (grey distributions). (C) Same as in
373 panel (D), but genomic predictions discard pigmentation-associated SNPs within the *bab* locus. (E-F) Same as (D-E)
374 but for pigmentation-associated SNPs identified via mapping the week 8 indoor populations.
375
376 *Stable genotype-phenotype mapping throughout evolution and across environments of selection*
377 How well do genotype-phenotype maps derived from a single population, at a single point in time,
378 transfer to genetically diverged populations? This represents another fundamental question in complex
379 trait genetics and evolution that we could address within the context of our longitudinal, multi-
380 environment study. Specifically, we observed a pronounced signature of genome-wide divergence
381 associated between our lab-based, indoor environment and the outdoor, field mesocosm (Fig. 2B), and
382 previously reported substantial evolutionary change within each of these environments throughout the
383 observation period monitored in this study [48,49]. One implication of this environmental and temporal
384 genomic divergence is that it may induce shifts in the epistatic landscape and/or elicit context dependent-
385 dominance of pigmentation-associated loci. In effect, these processes may alter the marginal effect of
386 pigmentation-associated loci and generate distinct genotype-phenotype maps that preclude phenotypic
387 prediction across mapping populations [26].
388 We evaluated the stability of the genotype-phenotype map by re-implementing our mapping
389 approach, as described above, on the replicate populations sampled in outdoor mesocosms at week 2, as
390 well as the indoor populations sampled at weeks 2 and 8. The number of loci underpinning variation
391 across all mapping populations was comparable (ranging from 30-43 independent loci, Supplementary
392 Data File 1; Table S2). Furthermore, as expected, the total genomic distance spanned by these loci
393 declined from weeks 2 to 8 (from ~1 Mb at week 2 to ~700 Kb at week 8) due to additional
394 recombination events that allowed for more localized mapping.
395 The similarities in architecture inferred in the lab and field environments at week 8 are clearly
396 apparent in the mirrored Manhattan plots presented in Fig. 5A, where regions of heightened pigmentation-
397 associated variation are replicable across mapping environments. Notably, loci identified across all four
398 mapping populations contained the canonical pigmentation genes *tan*, *ebony*, and *bab* (Fig. S12),
399 demonstrating consistency of mapping as well as the pronounced effect these genes have on pigmentation
400 variation in this species. We next took two approaches to rigorously quantify, genome-wide, the similarity
401 in loci harboring pigmentation-associated variation across the four independent mapping populations.
402 First, we assessed whether the start and end points of loci identified in each independent mapping
403 population were closer than expected by chance: for each set of loci identified within a particular mapping
404 population, we selected the locus-defining, midpoint SNP (i.e., that with the most significant p-value) and

405 then quantified the number of times it fell within a locus in other three sets of mapped loci (each locus
406 could therefore overlap 0 to 3 times with the remaining mapping populations). The number of overlaps
407 across all loci for a particular mapping population were then summed and compared to an empirical null
408 distribution derived by iterating this procedure N = 2000 times using a set of randomly selected SNPs (see
409 Methods). Across sets of mapped loci, midpoint SNPs were far more likely to fall within a locus of an
410 independent mapping than expected by chance (Fig. 5B), demonstrating that the regions harboring
411 pigmentation-associated variation were conserved across mapping populations. To probe this consistency
412 further, we quantified the distance between locus midpoints among mapping populations. As expected,
413 these distances were significantly closer than expected based on an empirical null (Fig. 5C, see Methods),
414 further indicating the conservation of pigmentation associated loci across mapping populations.
415 Given the pronounced mapping stability at the locus-level, we next interrogated whether
416 pigmentation-associated SNPs were also shared across mapping populations. To do so, we reidentified
417 trait-associated SNPs using our ancestry-informed GLM in groups of 5 (of the 10 total) replicates per
418 mapping population. This then allowed for a balanced comparison of the overlap in trait-associated SNPs
419 within subsets of the same mapping population (e.g., overlap in trait associated SNPs between
420 independent subsets of five replicates from the outdoor week 8 samples), and between mapping
421 populations (e.g., overlap in trait associated SNPs between a subsets of five outdoor, week 8 and five
422 indoor, week 2 replicates). Strikingly, we found that the amount of overlap in SNPs identified between
423 subsets of the same mapping population did not exceed that between mapping populations, either from the
424 same environment and a different time point or even between environments (Fig. 5D; Fig. S13). In the
425 supplement, we further quantified this SNP-level parallelism using a leave-one-out cross validation
426 scheme (Fig. S14).
427 While our quantification of overlap in pigmentation-associated loci across mapping populations
428 indicated a strong stability of the genotype-phenotype map (Fig. 5B-D), we observed several conspicuous
429 instances in which genomic regions harboring pigmentation-relevant variation for one mapping
430 population appeared to lack pigmentation-associated variation in an independent mapping population.
431 Several of such instances can be observed in the mirrored Manhattan plot in Fig. 5A, where the start and
432 end points of pigmentation loci (indicated by shaded regions) within one environment exhibit non-
433 overlapping start and endpoints with any of the loci identified in the second environment. We suspect
434 these environment-specific loci to potentially indicate the effects of epistasis altering the marginal effect
435 of pigmentation-associated SNPs across mapping populations. In the Supplement, we show that such
436 instances of environment-specific loci are not solely due to technical noise, but in fact environment-
437 specific marginal effects (Fig. S15 & S16); though the exclusion of these loci from cross-treatment
438 genomic prediction of trait evolution does not alter our inferences (Fig. S17). Over and above this

detectable set of environment-specific mapped loci, the vast majority of loci appear shared 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 in mapping.

Finally, given the overall stability of mapping at the SNP and locus-level, we queried our ability to predict relative phenotypic values across independent populations using allele frequency data alone. Specifically, we computed the relative pigmentation of phenotypic groups (i.e., the light, midpoint, and dark color fractions) both within and across mapping populations iteratively using each independent genotype-phenotype map (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 [20,55,56], 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 [57,58]

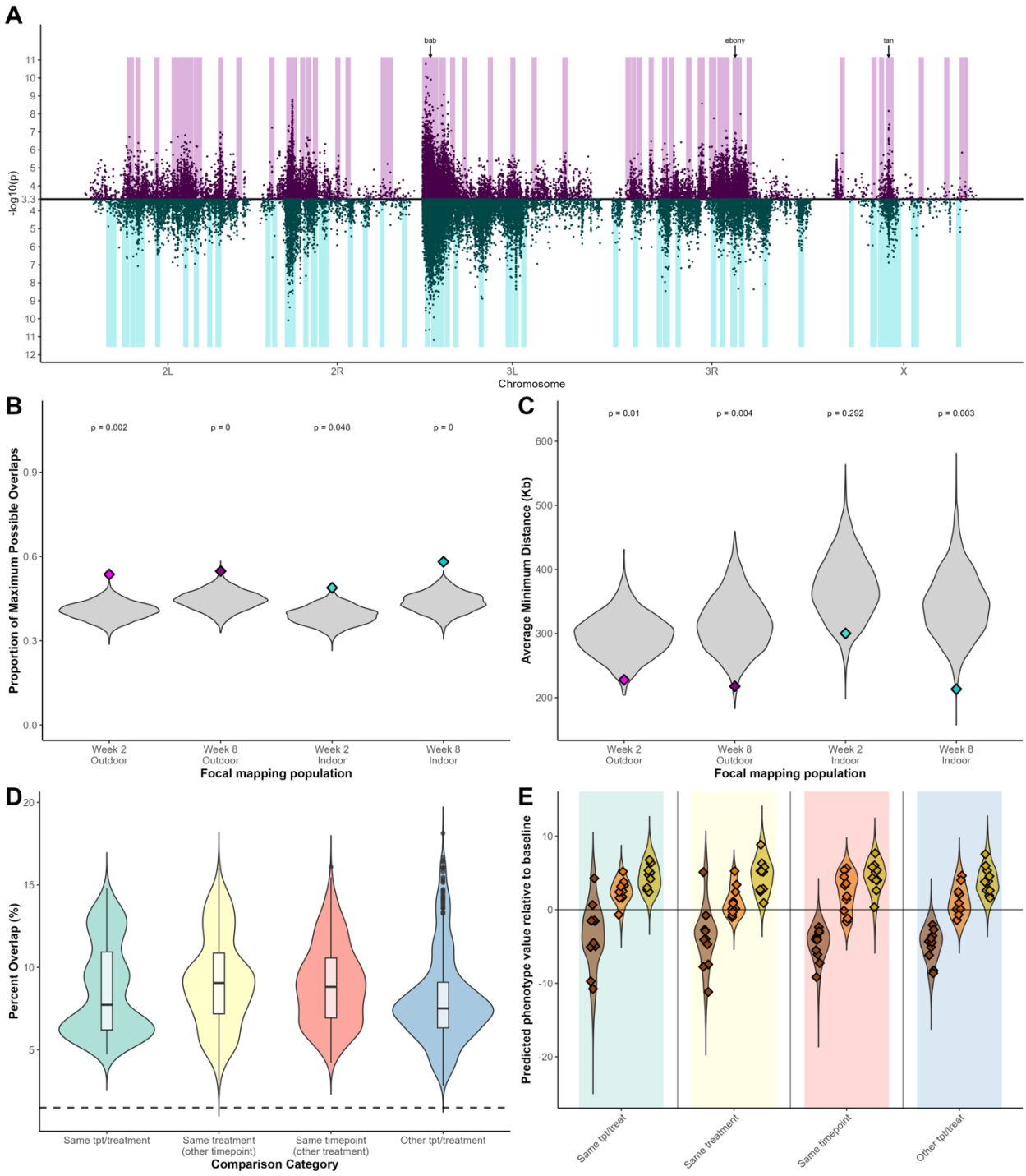


Figure 5 | Epistasis does not preclude portability of trait mapping. (A) Mirrored Manhattan plots of genetic architecture quantified at week 8 in the outdoor (top) and indoor (bottom) environments. Only SNPs with p-value corresponding to empirical FDR < 0 .0005 p-value are depicted. For each mapping, 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 (brown points), midpoint (orange points), and light (yellow points) 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 precise 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 [15,59]. 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 [13]. It is noteworthy that this constraint may not necessarily be a function of a single allele, but representative of a haplotype-level pleiotropy [48] where an allele conferring lighter pigmentation may be in tight linkage to an allele associated with a separate fitness-associated trait that is strongly selected against over the period that lighter pigmentation is favored. Regardless of the underlying mechanism, it is unsurprising that such dynamics emerged in our outdoor mesocosm, as we have shown that multiple fitness-relevant traits evolve across seasons in this system [50].

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

510 in pleiotropic constraint. From a practical standpoint, this implies that generic principles of the
511 evolutionary process itself may be obscured in simplified, lab-based environments and bolster recent calls
512 to study the ecology and evolution of natural populations in realistic settings [60].
513 While the impact of pleiotropy was pronounced, we found epistasis to have a less detrimental
514 effect on the prediction of complex trait variation overall. Specifically, we found a striking stability of the
515 genotype-phenotype map across four independent mapping populations sampled at two points in the
516 evolutionary process and evolved in two distinct environments. Furthermore, we showed that relative
517 phenotypic predictions were effectively as accurate across environments of selection as within. Therefore,
518 at least within the confines of the magnitude and timescale of evolution manipulated here, epistasis
519 (and/or context-dependent dominance shifts) may not inhibit a general transferability of mapping results
520 across genetically diverged populations. While this finding may be at odds with some studies evaluating
521 the portability of mapping data across diverged populations, it bolsters more recent work demonstrating
522 increased portability under instances in which trait architecture has been well-resolved (i.e., fine-mapped)
523 such as in the high-powered mapping conducted here [20,55–58]
524 It warrants note that the unprecedented insight into the predictability of evolution was only made
525 possible by several key aspects of our methodological approach: our generation of haplotype-informed
526 allele frequencies at extremely high coverage, in combination with an exceptional amount of biological
527 replicates per mapping treatment and time point, provided the power to produce well-resolved genetic
528 maps [47,61]. Furthermore, our longitudinal analysis of trait variation using a single founding population was
529 the only way to exclude confounding variables such as unique patterns of demographic history and
530 linkage when assessing the role of pleiotropy and epistasis on complex trait prediction [24,26]. In concert,
531 this study design then allowed us to both reaffirm canonical pigmentation genes, as well as identify new
532 candidates, underpinning variation in this classic complex trait. Crucially, our accurate genomic
533 prediction of trait evolution hinged upon our well-resolved genotype-phenotype map, as prediction using
534 canonical gene variation alone would have led to incorrect inferences of trait evolution. Our study
535 therefore exemplifies the critical roles of study design and power in the rigorous evaluation of
536 fundamental evolutionary dynamics.
537 Looking forward, our results are pertinent to numerous applications of genomic data: predicting
538 phenotype across disparate populations, inferring past selection using time-series and ancient DNA
539 genomic data, or manipulating contemporary variation to shape future evolutionary trajectories (e.g.,
540 augmenting population resilience to future global change conditions). The critical contingency of
541 effectively implementing such approaches, however, will be the procurement of well-resolved genotype-
542 phenotype maps and knowledge of the pleiotropy and epistasis underlying trait variation, a venture that
543 may not be feasible across study systems.

544
545
546 **Methods**
547 *Experimental overview*
548 A genetically diverse population of outbred *D. melanogaster* was generated via outcrossing ten
549 male and female individuals from a total of 76 inbred lines, originally collected from Linvilla Orchards,
550 Media, PA [62]. The outbred population was allowed to recombine and expand for 4 generations, after
551 which 500 males and 500 females were used to seed each of 10 outdoor and indoor replicate cages. The
552 outdoor mesocosm consisted of ten replicate, 2x2x2 m mesh cages, each harboring a single dwarf peach
553 tree, and was exposed to natural environmental conditions, insect, and microbial communities [49]. The
554 indoor mesocosm then consisted of ten replicate, ~0.5 m^3 cages housed in a temperature-controlled
555 laboratory at the University of Pennsylvania. Constant food was supplied to the replicates within each
556 environment, whereby 400 ml of Drosophila media was provided in 900 cm^3 aluminum loaf pans within
557 each cage three times per week [48,49]. Monitoring of phenotypic and genomic variation in the replicates
558 began two weeks after cage founding and continued for a period of eight total weeks and six generations
559 of evolution. Generation time estimates were produced using a degree day model [63].
560 Population-wide pigmentation variation was quantified in our panel of inbred lines, the
561 recombinant outbred population after 4 generations of expansion and recombination, and the evolving
562 replicate populations in the outdoor and indoor environments at weeks 2 and 8 (20 July and 1 September,
563 separated by a period of 5 generations). To quantify evolved changes in pigmentation, eggs were
564 collected overnight from all replicate cages (indoor/lab and outdoor/field), after which they were
565 transferred to common garden conditions (25˚C, 12L:12D) where they were reared for 2 generations.
566 Following density control of F2 eggs (30 +/-5) in replicate vials, adult female abdominal pigmentation (N
567 = 30 individuals per replicate) was scored as the proportion of pigmentation observed on tergites 5, 6, and
568 7[64]. The sum of these scores across the three tergites provided the total pigmentation score for the assayed
569 individual.
570 We quantified genome-wide allele frequencies throughout evolution weekly in the outdoor
571 environment and at weeks 2 and 8 in the indoor environment. At each collection point, eggs were
572 collected overnight from each cage and then transferred to common garden conditions where larval
573 development took place. 100 random females from the eclosed adult cohort were then collected and
574 stored in ethanol at -20°C for later DNA extraction and sequencing (see below).
575 At the weeks two and eight sampling points, we conducted a tail-based mapping of pigmentation
576 variation for each mapping population. Specifically, we first quantified pigmentation variation within
577 each treatment using a random sample of sorted, 100 F2 females (3-6 days post-eclosure) and the visual

578 scoring methods described above [64]. Then, for each replicate across both environments, we segregated
579 individuals into one of three trait value groups: (1) those exhibiting pigmentation within the upper 15% of
580 the population-wide phenotypic distribution, (2) those exhibiting pigmentation within the lower 15% of
581 the trait distribution and (3) a random sample from those representative of the midpoint (~25th-75th
582 percentile) of the trait distribution. Seventy-five individuals from each trait value group and replicate
583 were then stored in ethanol and at -20ºC for later DNA extraction and sequencing. In total this resulted in
584 120 samples available for mapping: three trait value groups across ten replicate cages sampled from two
585 environments and at two time points.

586 *Sequencing and allele frequency estimation*

587 Genomic DNA was extracted for both mesocosm time-series and trait mapping samples using
588 NEB Monarch DNA kit. Following sample QC via nanodrop and qubit, whole genome libraries were
589 constructed using Illumina DNA prep kit. Each sample was sequenced to an equal coverage of 8x using
590 Illumina Novaseq 6000. From the resulting sequencing data, we generated haplotype-informed allele
591 frequencies using previously described methods [47,51]. This method computes allele frequencies at sites
592 previously identified in the genome sequences of each founding inbred strain using an expectation-
593 maximization algorithm that estimates haplotype frequencies in a pooled sample of mapped sequencing
594 reads. Haplotype inference was conducted in window sizes that varied proportionally to the length of un-
595 recombined haplotype blocks expected as a function of the estimated number of generations since original
596 inbred line outbreeding [51]. After filtering sites an average minor allele frequency >= 0.02 in the founder
597 population, and present in at least one evolved sample at a minor allele frequency > =0.01, we retained
598 1.9 M SNPs for analysis.

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

600 We first assessed genome-wide patterns of variation across all samples used in trait mapping
601 using principal component analysis (*prcomp* package, R) and normalized allele frequency data across all
602 1.9 M SNPs. We projected samples onto the first two principal components and visualized the data by
603 shaping and coloring points according to different underlying variables (e.g., trait value group, treatment)
604 to infer which variables may dominate variance in allele frequencies across samples. We quantified
605 whether samples segregated along the first two principal components in accordance with trait value group
606 or environment using linear regression and ANOVA, respectively.

607 *Ancestry-informed association analysis of pigmentation variation*

608 Our general approach to mapping pigmentation was to identify SNPs harboring alleles with
609 systematic frequency changes across the light, midpoint, and dark trait value groups. First, however, it
610 was important to consider that due to the finite number of haplotypes from the founding inbred lines used
611 to derive our outbred population, a significant association between a SNP and trait value group could

612 indicate its biological impact on pigmentation variation and/or its inbred-line ancestry. This effect could
613 appear at the level of entire chromosomes due to long range patterns of linkage, as well as more locally
614 around causal loci. We quantified the magnitude of this effect genome-wide by computing the average
615 contribution of haplotypes coming from each inbred line to each trait value group (i.e., the difference in
616 haplotype frequency between the light and dark tails) and quantified the relationship of this value with the
617 pigmentation score of each inbred line using linear regression and Pearson's correlation coefficient.
618 We interrogated the effect of linkage on SNP mapping further by running two generalized linear
619 models (GLM; logistic link function and quasibinomial error variance). The first model only took into
620 consideration the impact of trait value group on variation in allele frequencies (model form: allele
621 frequency ~ trait value; where allele frequencies are normalized by the SNP's effective coverage and the
622 total number of chromosomes sampled). If the effects of ancestry on SNP association were pervasive,
623 SNPs identified via this approach would be enriched in those SNPs expected to have elevated light to
624 dark frequency differences based on their association with particular haplotypes alone. To test this, we
625 selected the top 1% of the SNPs (based on nominal p-value from the GLM) and calculated their expected
626 frequency difference between light and dark color fractions based on estimates of the average contribution
627 of each inbred line to that sample and the association between that SNP and inbred line. We then
628 compared these results to a separate GLM model in which the haplotypic ancestry was accounted for.
629 This model then regressed allele frequencies both on pigmentation trait value, as well as each SNPs
630 association with the line contribution to the sample using average genome-wide haplotype frequencies. In
631 accordance with the previous, ancestry-agnostic, model we used a logistic link function and
632 quasibinomial error variance, and allele frequencies were weighted to account for variation in coverage
633 and the finite number of chromosomes sampled. We then compared the density distributions of expected
634 light-dark frequency differences for SNPs identified via each of these distinct models using a KS test.
635 Following GLM model structure selection, we used permutations to define a final set of trait-
636 associated SNPs. Specifically, we randomly shuffled trait value data and re-ran our regressions N = 100
637 total times to generate an empirical null distribution. We then considered a SNP significant if its nominal
638 p-value from the empirical GLM fell in the top 0.05% of its permuted p-values (corresponding to an
639 empirical FDR threshold < 0.0005).
640 The set of pigmentation-associated SNPs identified via GLM was used to quantify the number
641 and distribution of unlinked loci underpinning trait variation. We developed an iterative procedure for
642 this, whereby we first identified the most significant SNP on each chromosomal arm, as well as its
643 surrounding, locally un-recombined region (defined by the recombination rate and number of generations
644 since cage founding; Tilk *et al*. 2019). We then eliminated all trait-associated SNPs on the chromosomal
645 arm (i.e., those with a significant GLM p-value) whose relative presence in the light and dark fractions

646 can be explained by haplotype frequencies (i.e., ancestry) in the locally un-recombined focal region alone.
647 As the significance of these SNPs could be driven either by linkage to the focal top SNP, or variation in
648 the trait itself, we eliminated these and defined the trait-associated locus as the locally unrecombined
649 region (as defined using number of generations since founder line outbreeding and genome-wide
650 recombination rates) around the lowest p-value SNP. The specific threshold used in SNP elimination was
651 defined via permutations: we iteratively subset the 10 replicates into two equal subsets (N = 5). In one
652 subset, termed 'real-data', light-dark haplotype frequency differences within the focal region were
653 computed as above. In the other subset, the dark-light haplotype frequency difference was calculated (i.e.,
654 termed 'sign-flipped' data). In each subset we sequentially eliminated percentiles of SNPs starting with
655 the top and bottom 2.5% of expected light-dark frequency differences and selected our threshold as the
656 percentile at which the number of true significant SNPs eliminated in the real-date matched that of the
657 sign-flipped data (Fig. S19). From the set of remaining SNPs (those not eliminated in the first iteration),
658 we identified the SNP with the most significant nominal p-value and repeated the procedure. This was
659 iterated until we exhausted all trait-associated SNPs on the chromosome. For each chromosomal arm, we
660 then quantified how the number of trait-associated SNPs decreased as a function of these iterations (Fig.
661 S20). Finally, we merged all loci with overlapping breakpoints to procure a final set of unlinked and
662 independent trait-associated loci.
663 We additionally explored which candidate genes may be underlying the identified loci.
664 Specifically, we identified genes within each locus with a GO term associated with pigmentation
665 (excluding eye pigmentation) or cuticle development/molting. For each of the identified genes, we then
666 quantified whether the number of pigmentation-associated SNPs within the gene (identified via mapping)
667 was greater than expected based on their genome-wide abundance using a binomial test with a Benjamini-
668 Yekutieli false discovery rate correction.
669 *Comparison of mapping to previous association studies*
670 We leveraged results reported by two previous genome wide association studies [45,46] to quantify
671 the consistency of mapping of pigmentation across disparate populations and studies. In brief, Bastide et
672 al. (2013) conducted a pooled-GWAS approach using wild populations of *D. melanogaster* from Austria
673 and Italy, and focused on female pigmentation on tergite 7. Dembeck et al. (2015) conducted a GWAS
674 using individual *D. melanogaster* lines from the Drosophila Genetics Reference panel [65], focusing on
675 pigmentation variation of tergites A5 and A6. We evaluated whether all SNPs identified as trait-
676 associated in the Dembeck study, and the top 100 SNPs from Bastide et al., exhibited more overlap with
677 our unlinked loci than expected by chance. Specifically, we compared the number of SNPs falling within
678 each locus to a random set of SNPs within the chromosome of the same test set size as those obtained
679 from the comparison studies. We iterated this process 1000 times and generated an empirical p-value

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

Predicting evolution using mapping results

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

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

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

714

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

716 We next sought to determine the stability of the genotype-phenotype map across independent
717 mapping populations (i.e., each mapping environment and timepoint). Accordingly, we re-conducted our
718 mapping procedure, as described above, for the indoor weeks 2 and 8, as well as the outdoor week 2
719 mapping population. First, we asked if the proximity of loci identified across mapping populations were
720 more similar than expected by chance. To address this, for each of the four genotype-phenotype maps we
721 identified the top SNP of each locus (the SNP with the lowest p-value, and also the locus midpoint) and
722 summed the number of times it fell within a locus identified in one of the independent mapping
723 populations. This then produced a score, ranging from 0-3, for each locus and these scores were summed
724 across all loci to produce an empirical overlap score for each mapping population (note that to increase
725 power of this analysis, we selected focal SNPs using the set of pre-merged loci for each mapping
726 population). To determine if the observed overlaps were greater than expected by chance, we generated
727 an empirical null distribution via random sampling of sets of SNPs for each genotype-phenotype map and
728 re-computing overlap scores with each independent mapping (a procedure iterated N = 2000 times for
729 each of the four genotype-phenotype maps). We subsequently interrogated whether the distance between
730 locus-defining midpoint SNPs across sets of loci were closer than expected by chance. Specifically, as
731 before, we first identified the top SNPs for each set of loci from every genotype-phenotype map, and then
732 computed its distance to the closest top SNP of each of the remaining three mappings. This distance
733 metric was averaged to produce an average proximity score which, as above, was compared to a
734 distribution of scores derived via a random set of SNPs (N = 2000 times).

735 We next aimed to quantify the extent to which the identity of specific pigmentation-associated
736 SNPs was shared across independent mapping populations. To do so, we compared the overlap in
737 identified SNPs within and between independent mapping populations. To conduct this analysis in a
738 balanced manner (i.e., compare sets of SNPs identified using the same number of replicate
739 cages/populations), we iteratively subset our set of 10 replicates from each mapping population into
740 groups of 5 replicates and identified pigmentation associated SNPs in each subset using the ancestry-
741 informed regression introduced above. This was iterated with each possible combination of subset
742 replicates, yielding 126 total sets of significant SNPs per mapping population. The percentage of shared
743 SNPs was then calculated for subsets of the same mapping population (e.g. overlap between
744 pigmentation-associated SNPs identified using different subsets of 5 replicate populations from the
745 outdoor, week 8 replicates), and compared to those observed between independent mapping populations
746 (e.g. overlap between sets of pigmentation associated SNPs identified using 5 replicate populations from
747 outdoor week 8 replicates and sets of SNPs from independent treatment and/or time point). Whether the

amount of overlap in identified SNPs within vs. between mapping populations differed was evaluated with a planned contrast using a one-way ANOVA model.

We next sought to quantify whether there existed mapped loci with environment-specific marginal effects on pigmentation variation. Accordingly, we first isolated those loci from the outdoor and indoor week 8 mapping populations that had non-overlapping breakpoints with any of the loci in the opposing environments (N = 13 total unique outdoor loci, and N = 13 total unique indoor loci). We then identified all pigmentation-associated SNPs within each unique locus and quantified the light dark-frequency difference for the light-associated allele for the environment in which it was identified, as well as the environment that did not contain the locus. Evidence for environment-specific marginal effects of each locus was then determined by comparing the distribution of light-dark frequency shifts for each locus between environments (using a Wilcoxon Signed Rank Test), as well as to a distribution derived from an equally sized set of matched control SNPs (using a Mann Whitney U test).

Finally, we sought to determine if genotype-phenotype maps produced at a particular point in evolution, and within a particular environment of selection, could correctly predict phenotypes of individuals sampled at an independent time and environment. To do so, we used those SNPs identified in our outdoor week 8 mapping and calculated the relative expected phenotypic score for trait value groups (e.g., light, midpoint, and dark color fractions) within this focal mapping, as well as all remaining, independent mapping populations. Specifically, for each light-associated allele at all pigmentation-associated SNPs, we calculated the difference in frequency between the baseline population and each pooled sample. This allele frequency change was weighted in accordance with the SNPs effect size on trait variation as computed via mapping (i.e., by multiplying that by the regression coefficient of our ancestry-informed GLM by the light-favored allele frequency change of that sample). These values were then summed across all SNPs for each sample, providing a metric of the relative lightness of the sample, as predicted by mapping (values greater than 0 indicating lighter pigmentation and less than 0 darker). We then bootstrapped this value 1000 times for each sample. The accuracy of relative predicted phenotypic scores was then evaluated using three criteria: (i) whether the dark samples had, on average, predicted phenotypic values below 0, (ii) whether the light samples had, on average, predicted phenotypic values above 0, and (iii) if there existed a significant linear association between trait value group and relative predicted phenotypic value. The latter was tested using a simple linear regression with the model trait value group ~ phenotype prediction, where trait values groups were assigned a numeric value (dark = 0, midpoint = 1, and light = 2).

Code and data availability: All code necessary to replicate the analyses presented in this manuscript are 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, 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 Hayes Oken for experimental support, Marianna Karageorgi and Emily Behrman for assistance with the degree day model, as well as 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

1. Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).
2. Wortel, M. T. *et al.* Towards evolutionary predictions: Current promises and challenges. *Evol Appl* **16**, 3–21 (2022).
3. Colosimo, P. F. *et al.* Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science* **307**, 1928–1933 (2005).

4. Enattah, N. S. *et al.* 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 (2008).
5. Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. & Crossland, J. P. A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science* **313**, 101–104 (2006).
6. Hof, A. E. van't *et al.* The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**, 102–105 (2016).
7. Kunte, K. *et al.* doublesex is a mimicry supergene. *Nature* **507**, 229–232 (2014).
8. Nadeau, N. J. *et al.* The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature* **534**, 106–110 (2016).
9. Reed, R. D. *et al.* optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**, 1137–1141 (2011).
10. Sella, G. & Barton, N. H. 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 (2019).
11. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**, 5–22 (2017).
12. Mackay, T. F. C. & Anholt, R. R. H. Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat Rev Genet* **25**, 639–657 (2024).
13. Walsh, B. & Blows, M. W. Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annu. Rev. Ecol. Evol. Syst.* **40**, 41–59 (2009).

835 14. Svensson, E. I. *et al.* Correlational selection in the age of genomics. *Nat Ecol Evol* **5**, 562–
836 573 (2021).
837 15. Albecker, M. A., Bittar, T. B., Trussell, G. C. & Lotterhos, K. E. A Quantitative Survey of
838 Cogradient and Countergradient Variation in Nature. *The American Naturalist* 000–000
839 (2025) doi:10.1086/737682.
840 16. Conover, D. O. & Schultz, E. T. Phenotypic similarity and the evolutionary significance of
841 countergradient variation. *Trends in Ecology & Evolution* **10**, 248–252 (1995).
842 17. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human
843 populations. *Nat Commun* **10**, 3328 (2019).
844 18. Huang, W. *et al.* Epistasis dominates the genetic architecture of Drosophila quantitative
845 traits. *Proceedings of the National Academy of Sciences* **109**, 15553–15559 (2012).
846 19. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and
847 evolution of genetic systems. *Nat Rev Genet* **9**, 855–867 (2008).
848 20. Privé, F. *et al.* Portability of 245 polygenic scores when derived from the UK Biobank and
849 applied to 9 ancestry groups from the same cohort. *The American Journal of Human*
850 *Genetics* **109**, 12–23 (2022).
851 21. Durvasula, A. & Lohmueller, K. E. Negative selection on complex traits limits genetic risk
852 prediction accuracy between populations. *bioRxiv* 721936 (2019) doi:10.1101/721936.
853 22. Guo, Z. *et al.* The impact of population structure on genomic prediction in stratified
854 populations. *Theor Appl Genet* **127**, 749–762 (2014).
855 23. Kosheleva, K. & Desai, M. M. Recombination Alters the Dynamics of Adaptation on
856 Standing Variation in Laboratory Yeast Populations. *Mol Biol Evol* **35**, 180–201 (2018).

24. Lohmueller, K. E. The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits. *PLOS Genetics* **10**, e1004379 (2014).

25. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics* **100**, 635–649 (2017).

26. Novembre, J. & Barton, N. H. Tread Lightly Interpreting Polygenic Tests of Selection. *Genetics* **208**, 1351–1355 (2018).

27. Patel, R. A. *et al.* Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits. *The American Journal of Human Genetics* **109**, 1286–1297 (2022).

28. Yair, S. & Coop, G. *Population Differentiation of Polygenic Score Predictions under Stabilizing Selection*. 2021.09.10.459833 https://www.biorxiv.org/content/10.1101/2021.09.10.459833v1 (2021) doi:10.1101/2021.09.10.459833.

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

30. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. *The American Journal of Human Genetics* **110**, 179–194 (2023).

31. Akbari, A. *et al.* Pervasive findings of directional selection realize the promise of ancient DNA to elucidate human adaptation. 2024.09.14.613021 Preprint at https://doi.org/10.1101/2024.09.14.613021 (2024).

879 32. Cox, S. L., Ruff, C. B., Maier, R. M. & Mathieson, I. Genetic contributions to variation in
880 human stature in prehistoric Europe. *PNAS* **116**, 21484–21492 (2019).
881 33. Halldorsson, B. V. *et al.* The sequences of 150,119 genomes in the UK Biobank. *Nature* **607**,
882 732–740 (2022).
883 34. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**,
884 499–503 (2015).
885 35. Chiu, Y.-F., Justice, A. E. & Melton, P. E. Longitudinal analytical approaches to genetic
886 data. *BMC Genetics* **17**, S4 (2016).
887 36. Kerner, B., North, K. E. & Fallin, M. D. Use of Longitudinal Data in Genetic Studies in the
888 Genome-wide Association Studies Era: Summary of Group 14. *Genet Epidemiol* **33**, S93–
889 S98 (2009).
890 37. Lenski, R. E. Experimental evolution and the dynamics of adaptation and genome evolution
891 in microbial populations. *ISME J* **11**, 2181–2194 (2017).
892 38. Stroud, J. T. & Ratcliff, W. C. Long-term studies provide unique insights into evolution.
893 *Nature* **639**, 589–601 (2025).
894 39. Wu, Z., Hu, Y. & Melton, P. E. Longitudinal Data Analysis for Genetic Studies in the
895 Whole-Genome Sequencing Era. *Genetic Epidemiology* **38**, S74–S80 (2014).
896 40. Kronforst, M. R. *et al.* Unraveling the thread of nature's tapestry: the genetics of diversity
897 and convergence in animal pigmentation. *Pigment Cell & Melanoma Research* **25**, 411–433
898 (2012).
899 41. Berardi, S. *et al.* Drosophila melanogaster pigmentation demonstrates adaptive phenotypic
900 parallelism over multiple spatiotemporal scales. *Evolution Letters* qraf008 (2025)
901 doi:10.1093/evlett/qraf008.

902 42. Massey, J. & Wittkopp, P. J. The genetic basis of pigmentation differences within and
903 between Drosophila species. *Curr Top Dev Biol* **119**, 27–61 (2016).
904 43. Wright, T. R. F. The Genetics Of Biogenic Amine Metabolism, Sclerotization, And
905 Melanization In *Drosophila Melanogaster**. in *Advances in Genetics* (eds Scandalios, J. G.
906 & Caspari, E. W.) vol. 24 127–222 (Academic Press, 1987).
907 44. Bastide, H., Lange, J. D., Lack, J. B., Yassin, A. & Pool, J. E. A Variable Genetic
908 Architecture of Melanic Evolution in Drosophila melanogaster. *Genetics* **204**, 1307–1319
909 (2016).
910 45. Bastide, H. *et al.* A Genome-Wide, Fine-Scale Map of Natural Pigmentation Variation in
911 Drosophila melanogaster. *PLOS Genetics* **9**, e1003534 (2013).
912 46. Dembeck, L. M. *et al.* Genetic Architecture of Abdominal Pigmentation in Drosophila
913 melanogaster. *PLOS Genetics* **11**, e1005163 (2015).
914 47. Tilk, S. *et al.* Accurate Allele Frequencies from Ultra-low Coverage Pool-Seq Samples in
915 Evolve-and-Resequence Experiments. *G3 (Bethesda)* **9**, 4159–4168 (2019).
916 48. Bitter, M. C. *et al.* Pervasive fitness trade-offs revealed by rapid adaptation in large
917 experimental populations of Drosophila melanogaster. 2024.10.28.620721 Preprint at
918 https://doi.org/10.1101/2024.10.28.620721 (2024).
919 49. Bitter, M. C. *et al.* Continuously fluctuating selection reveals fine granularity of adaptation.
920 *Nature* 1–8 (2024) doi:10.1038/s41586-024-07834-x.
921 50. Rudman, S. M. *et al.* Direct observation of adaptive tracking on ecological time scales in
922 Drosophila. *Science* **375**, eabj7484 (2022).
923 51. Kessner, D., Turner, T. L. & Novembre, J. Maximum Likelihood Estimation of Frequencies
924 of Known Haplotypes from Pooled Sequence Data. *Mol Biol Evol* **30**, 1145–1158 (2013).

925 52. Bickel, R. D., Kopp, A. & Nuzhdin, S. V. Composite Effects of Polymorphisms near
926 Multiple Regulatory Elements Create a Major-Effect QTL. *PLOS Genetics* **7**, e1001275
927 (2011).
928 53. Jeong, S. *et al.* The Evolution of Gene Regulation Underlies a Morphological Difference
929 between Two Drosophila Sister Species. *Cell* **132**, 783–793 (2008).
930 54. Pool, J. E. & Aquadro, C. F. The genetic basis of adaptive pigmentation variation in
931 Drosophila melanogaster. *Molecular Ecology* **16**, 2844–2851 (2007).
932 55. de Bakker, P. I. W. *et al.* Transferability of tag SNPs in genetic association studies in
933 multiple populations. *Nat Genet* **38**, 1298–1303 (2006).
934 56. Mostafavi, H. *et al.* Variable prediction accuracy of polygenic scores within an ancestry
935 group. *eLife* **9**, e48376 (2020).
936 57. Hu, S. *et al.* Fine-scale population structure and widespread conservation of genetic effect
937 sizes between human groups across traits. *Nat Genet* **57**, 379–389 (2025).
938 58. Wu, Y. *et al.* Genome-wide fine-mapping improves identification of causal variants.
939 *medRxiv* 2024.07.18.24310667 (2025) doi:10.1101/2024.07.18.24310667.
940 59. Paaby, A. B., Bergland, A. O., Behrman, E. L. & Schmidt, P. S. A highly pleiotropic amino
941 acid polymorphism in the Drosophila insulin receptor contributes to life-history adaptation.
942 *Evol* **68**, 3395–3409 (2014).
943 60. Bergelson, J., Kreitman, M., Petrov, D. A., Sanchez, A. & Tikhonov, M. Functional biology
944 in its natural context: A search for emergent simplicity. *eLife* **10**, e67646 (2021).
945 61. Long, A. D., Hanson, K. M. & Macdonald, S. J. The Illusion of Polygenicity in Poolseq
946 studies: Insufficient Power Can Mask Simple Genetic Architectures. 2025.07.23.666414
947 Preprint at https://doi.org/10.1101/2025.07.23.666414 (2025).

62. Behrman, E. L. *et al.* Rapid seasonal evolution in innate immunity of wild Drosophila melanogaster. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20172599 (2018).

63. Kamiyama, M. T., Bradford, B. Z., Groves, R. L. & Guédot, C. Degree day models to forecast the seasonal phenology of Drosophila suzukii in tart cherry orchards in the Midwest U.S. *PLOS ONE* **15**, e0227726 (2020).

64. David, J. R., Capy, P. & Gauthier, J.-P. 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 (1990).

65. Mackay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178 (2012).