Strong Selection, but low repeatability: Temperature-specific effects on genomic predictions of adaptation

Alexandre Rêgo[1,2], Julian Baur[2], Camille Girard-Tercieux[2], Maria de la Paz Celorio-Mancera[3], Rike Stelkens[1], David Berger[2]

[1] Department of Zoology, Stockholm University, Stockholm 114 19, Sweden

[2] Department of Ecology and Genetics, Uppsala University, Uppsala, 752 36, Sweden

[3] Department of Ecology, Environment, and Plant Science (DEEP), Stockholm University, Stockholm 106 91, Sweden

Corresponding author: Alexandre Rêgo
Department of Ecology and Genetics
Uppsala University
Uppsala 752 36, Sweden
Email: denovorego@gmail.com

Abstract

Evolution should be more predictable when natural selection is strong and favors the same outcome. Climate warming is increasing temperatures beyond the optima of many ectotherms, which, due to the inherent non-linear relationship between temperature and the rate of cellular processes, is predicted to impose stronger selection compared to corresponding shifts toward cold temperatures. This suggests that adaptation to climate warming should be relatively predictable. Here, we tested this hypothesis from the level of single-nucleotide polymorphisms to life-history traits, by conducting an evolve-and-resequence experiment on three genetic backgrounds of the seed beetle, *Callosobruchus maculatus*. Indeed, phenotypic evolution was faster and more repeatable at hot, relative to cold, temperature. However, at the genomic level, adaptation to heat was less repeatable than to cold, especially when comparing responses between backgrounds. As a result, genomic predictions of phenotypic (mal)adaptation in populations exposed to hot temperature were highly accurate within, but inaccurate between, genetic backgrounds. These results seem best explained by an increased importance of epistasis during adaptation to heat and imply that the same biophysical mechanisms that increase the repeatability of phenotypic evolution by exerting strong selection at hot temperature, reduce repeatability at the genome level. Thus, predictions of adaptation in key phenotypes from genomic data may become increasingly difficult as climates warm.

Keywords: evolve and resequence; *Callosobruchus maculatus*; thermal adaptation; parallel evolution; selection coefficient; genomic offset; epistasis; repeatability

Introduction

Whether evolution is repeatable, and whether the same genes contribute to recurrent phenotypic adaptations, is a long-standing question with fundamental implications for predicting evolutionary responses to changing environments [1–3]. Recent explorations of the genetic basis of adaptation have shown that, while evolution at the phenotypic level often is repeatable, the genes that contribute to adaptation can be highly contingent on the evolutionary history of populations [2, 4, 5]. Nevertheless, theory [6] and recent evolve and re-sequence (E&R) experiments [2, 7–13] have revealed that broad-scale patterns also emerge at the genomic level; adaptation is often polygenic, and while individual DNA polymorphisms are unlikely to repeatedly contribute to adaptation, the same molecular pathways are often involved, sometimes even over long evolutionary timescales [14]. This type of information is necessary to gain an informed opinion about the adaptive potential of populations under contemporary climate change [1, 15, 16], and is becoming increasingly important given the growing reliance on genomic data to inform conservation practices [17–20]. However, whether genomic estimates of evolutionary potential accurately predict extinction risk and adaptation in key ecological phenotypes remains poorly understood [3, 21, 22].

Several factors can influence the probability of observing repeated evolutionary responses, such as differences in standing genetic variation in natural populations [23–25], or previous differentiation between evolving lineages that may impact the fates of segregating alleles whose fitness effects depend on allele frequencies at other loci (i.e. epistasis) [26–28]. Theory additionally suggests that environmental changes that impose stronger selection are more likely to produce repeatable evolutionary outcomes compared to environments that impose weak selection [2, 6]. However, despite being critical for eco-evolutionary forecasting, the relative importance of these factors in producing (non)repeatable outcomes of evolution remains unknown [3, 29, 30].

As populations experience increased temperature extremes, the evolutionary potential of populations to adapt to both high and low temperatures becomes crucial to their survival. To what extent high and low temper-

atures result in repeatable genomic adaptation, and whether changes at the genomic level are informative of phenotypic evolution, can further inform the management of threatened populations. However, studies which incorporate both phenotypic and genomic information on thermal adaptation are rare [31, 32]. Yet, cases of adaptation to temperature are both amenable to exploring the causes of evolutionary repeatability due to a fairly deep understanding of thermal physiology [33, 34], and crucial for making informed management decisions in future climates. Thermodynamic constraints on protein function often manifest as asymmetries in thermal performance curves of ectotherms (and possibly in endotherms to a lesser-extent; [35]), where performance measures and fitness-correlates more rapidly decline at hotter than colder temperatures at equal distances from a thermal optimum [36]. Populations far-displaced from their thermal optimum towards hotter temperatures may then experience strong increases in selection for beneficial alleles. Increased selection at hot temperatures suggests that evolution may generally be faster and more predictable than at cold temperatures, and that selection may leave a temperature-specific genome-wide signature of adaptation [37–39]. Understanding the evolutionary consequences of this physiological temperature-dependence is of fundamental importance for understanding the prospects of employing genomics data to predict phenotypic adaptation and extinction risk in warming climates.

Here we employed experimental evolution in the seed beetle *Callosobruchus maculatus* to test the hypothesis that adaptation to heat results in greater evolutionary predictability than adaptation to cold. We combine data on phenotypic and genomic change from replicate experimental evolution lines deriving from an evolve-and-resequence experiment explicitly designed to quantify the role of selective determinism versus historical contingency in thermal adaptation. Replicate lines, established from three far-separated geographic populations, were placed at cold (23°C) or hot (35°C) temperature equidistant from their optimal ancestral temperature (29°C). To test for evolutionary repeatability, we estimated evolutionary trajectories at the phenotypic level among seven life-history traits, and at the genomic level across putatively selected SNPs and genes. We then compared evolutionary responses between line replicates (within genetic backgrounds) and geographic populations (between genetic backgrounds) to understand

the role of historical contingency and temperature-specific selection on the repeatability of evolution. Finally, we estimated the correspondence between genomic and phenotypic estimates to evaluate the power of genomic data to predict the level of (mal)adaptation.

Materials and Methods

Study Species

The seed beetle, *Callosobruchus maculatus*, infests human stores of grain legumes. Females attach eggs to the surface of seeds and larvae burrow into and develop within a single seed. Adults then emerge and reproduce, typically within 3-4 weeks, without the need for additional nutritional resources and a full life cycle can be completed within less than a month at benign temperatures. Because *C. maculatus* has been associated with stored legumes for thousands of years, laboratory conditions are a good approximation of its "natural" environment [40]. The experimental lines used for this study are described in detail in [41] and [42]. Briefly, outbred stocks were initially collected from three geographic populations (Brazil, California, and Yemen) and kept in laboratory conditions for >60 generations. The mitochondrial haplotypes of each population were introgressed in an orthogonal fashion into each nuclear background through repeated backcrossings over 16 generations, such that each mitochondrial haplotype existed in equal frequency on each nuclear background [43, 44]. The introgressed lines were kept as separate populations for a subsequent 50 generations, at which point females from each line were backcrossed to males from the source population with the same nuclear background to refresh nuclear genetic diversity. These newly backcrossed lines were used as the three base populations (Brazil, California, and Yemen) for our experimental evolution. At the time of study, these stocks had been reared on black-eyed beans (*Vigna unguiculata*) under standard laboratory rearing conditions of 29°C in 12:12 light/dark cycles and 50% relative humidity, and $N = 300 - 400$, for >200 generations.

Experimental Evolution

Each replicate population was maintained on 250 ml black eyed beans (*Vigna unguiculata*) in a 1 liter glass jar kept in a climate cabinet set at 50% relative humidity at the given evolution regime temperature (23°C or 35°C) (Fig. 1A). Once adults emerged from beans, 600 beetles were transferred to a new jar with fresh beans and allowed to mate and then lay eggs until their death. Beetles were transferred at the peak of hatching, so as not to apply strong direct selection for faster development time. See [41] and [42] for further details.

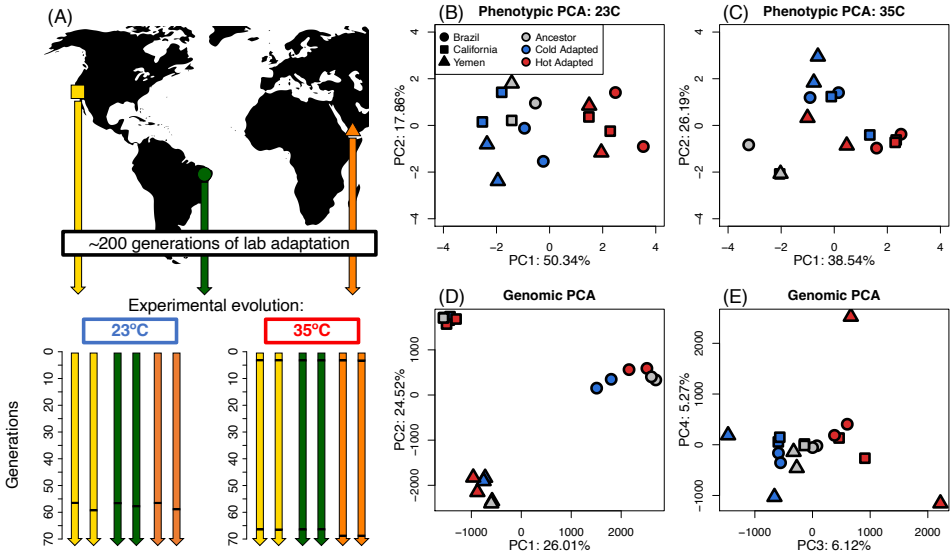

Figure 1: Experimental design. (A) Populations were collected from California, Brazil, and Yemen and adapted to lab conditions for approximately 200 generations. At the beginning of experimental evolution, lines were placed at either 23C or 35C, with two replicates in each treatment per genetic background. Horizontal black bars in the arrows depicting each experimental evolution line indicate the time at which lines were sequenced. The sequencing of heat-adapted lines at generation 3 represent the genomic estimates of ancestral variation. For both the 7 life-history phenotypes (B-C) and for all genomic SNPs (D-E), PCA summaries are presented of evolutionary change between ancestral and evolved replicates. Traits were measured for all lines in a common garden design at both low and high temperature.

Phenotypic data and analysis

We re-analyzed previously collected data on thermal adaptation in seven life-history traits of females [42], including three core traits: lifetime offspring production (LRS), adult weight (AW) and juvenile development time (DT) (Fig. 2A), as well as four rate-dependent traits measured over the first 24 hours of female reproductive lifespan: weight loss (WL), water loss (WVP), early fecundity (eggs laid over the first day of adult female reproduction) and mass-specific metabolic rate (ml CO_2/min/mg). As reported by [42], Hot-adapted lines have evolved greater body mass, weight loss, and reproductive output than cold lines, whereas rate-dependent traits show signs of local adaptation to temperature signified by statistically significant regime-by-assay temperature interactions. The Yemen background is also phenotypically more distinct than the California and Brazil background based on the four rate-dependent traits (Fig. S1).

These data were collected at generation 45 for cold-adapted lines and 60 for hot-adapted lines, but at the same time in a large common garden experiment removing (non-genetic) parental effects and including the two assay temperatures corresponding to the experimental evolution regimes (23°C and 35°C). Phenotypic data from the ancestral populations were not collected at the time of line establishment, but were scored in a separate experiment. Because each line had been lab-adapted for >200 generations, we assume that minimal lab adaptation had occurred after thermal regime establishments, and that these phenotypic data approximate the phenotypes of the founding populations. To estimate any block effects due to differing sampling times, a reference line was used in both the experiment on ancestors and E&R lines, and the relative difference in trait means of this reference line in the two experiments was used to standardize the ancestral data to minimize the possibility that block effects were wrongfully assigned as evolved differences between ancestors and evolved lines. We note that the standardization was made for each trait by averaging over temperatures, and thus, does not contribute to the differences in the repeatability and magnitude of phenotypic adaptation between temperature and genetic background that we study here. Data on LRS was then further gathered in a large second common garden including both ancestors and evolved lines, after 80 (cold),

120 (hot) and 125 (ancestors) generations of evolution. Methods for data collection are thoroughly described in [42].

By including ancestral data, we were able to assess the direction and magnitude of phenotypic evolution. Following [45], we define parallel evolution here in a geometric sense by the angles of phenotypic change vectors between evolving populations. Although it should be emphasized that (non)parallelism is a continuous measure, three general categories can be made of the degree of (non)parallelism. In multi-dimensional trait space, evolution is parallel when the vectors of phenotypic change between evolving populations result in small angles ($\theta < 90°$), and anti-parallel when angles are large ($\theta > 90°$). Evolution can also result in orthogonal (i.e. uncorrelated) changes when $\theta \approx 90°$. Following [46], we calculated the 6×6 inter-population correlation matrix of evolutionary change vectors for each regime, and subsequently arccosine transformed correlations to angles. Because change vectors for each of the two replicate lines of a given origin were calculated using the phenotypic scores of the same ancestral population, shared measurement error in the ancestor will have contributed to more similarity between line replicates of the same origin, relative to similarity in phenotypic change vectors of lines of different origin (with separate ancestors used in calculations of change vectors). The angle between two change vectors can be decomposed into its components; $\theta = cos^{-1}\frac{a_x b_x + a_e b_e}{\sqrt{(a_x^2+a_e^2)(b_x^2+b_e^2)}}$, where a_x and b_x represent the vectors of population mean trait change for evolved lines a and b, and a_e and b_e represent the vectors of standard errors for each trait in the ancestors. This latter error component of the dot product (numerator) sums to zero when measurement errors are uncorrelated, such as the case when using different ancestors in calculations, but equals the squared standard error when using the same ancestor. To produce unbiased comparisons within and between backgrounds we therefore corrected for this contribution of shared measurement error by subtracting this error component from the within estimates of change vectors. For the two cold-adapted Brazil populations, this error component was larger than the observed mean trait change. The angle of phenotypic change between these populations was therefore not estimated due to the statistical uncertainty.

By comparing the phenotypic distance between populations at the start (S_d) of experimental evolution (i.e. between ancestors) to the distance

at the end (E_d) (i.e. between the evolved lines), we estimated whether populations were diverging or converging phenotypically over the course of the experiment. When evolved lines are more distant than the ancestors in trait-space ($E_d - S_d > 0$), populations are diverging, while the opposite ($E_d - S_d < 0$) is indicative of convergence. To estimate evolutionary responses to heat, we compared the Hot lines to their ancestors when reared at 35°C. Similarly, cold-adaptation was estimated by comparing Cold lines to their ancestors when reared at 23°C. To perform the evolutionary change vector analysis, we first mean-scaled all traits into unitless measurements for comparison with each other by dividing ancestral and evolved estimates by the grand-mean across all ancestors measured at the same assay temperature. Hence, rates of evolution are reported as proportional changes relative to ancestral trait means (see [47, 48]). We then calculated the vector of phenotypic change from the ancestral to evolved samples in the 7-dimensional trait-space. Similarly to the error correction conducted for angles, divergence within backgrounds will be overestimated relative to that between backgrounds due to the use of a single shared ancestral population in the former. We therefore corrected within-background distances by subtracting the squared standard errors of distance estimates in the two evolved replicates as $E_d = \sqrt{(a_x - b_x)^2 - (a_e^2 + b_e^2)}$, where a_x and b_x represent vectors of trait means, and a_e and b_e represent the vectors of standard errors, in evolved lines. For instances in which the error component was larger than the observed mean trait change component, the distances were set to zero (in all within-origin comparisons at 23°C and between the two replicate California lines at 35°C).

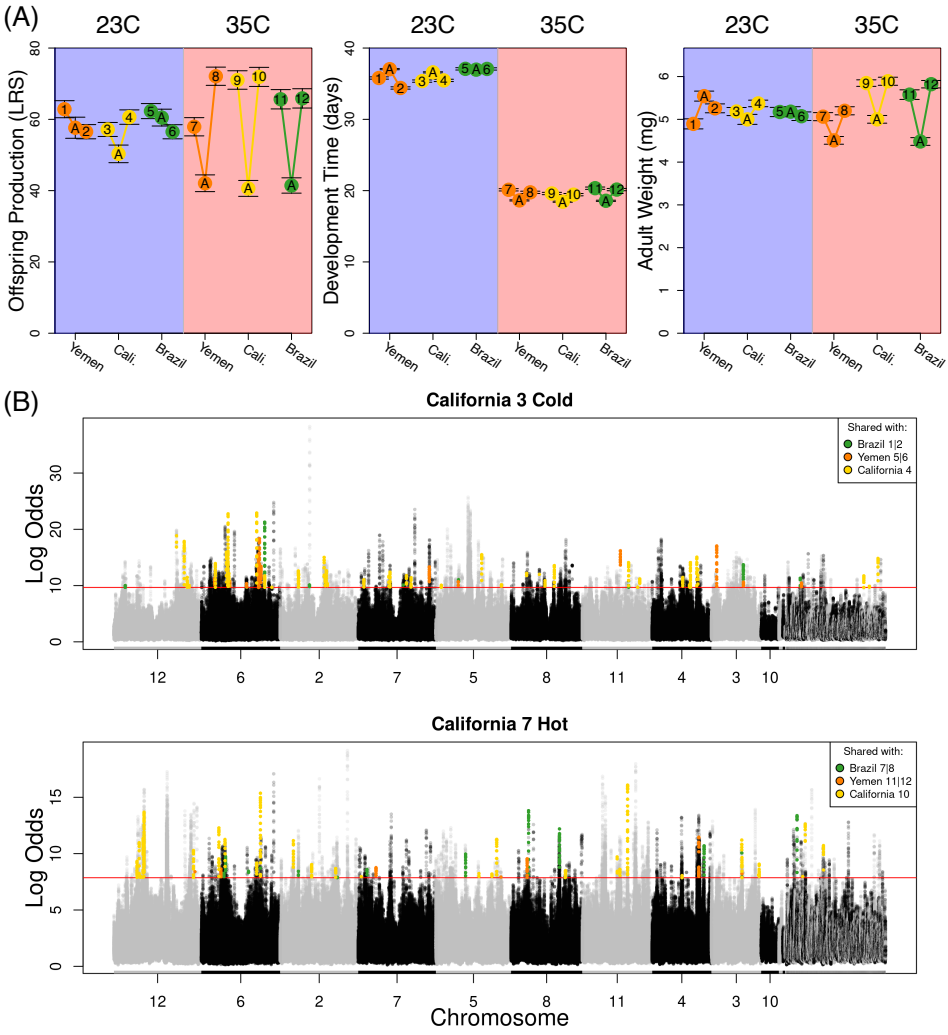

Figure 2: Phenotypic and genomic data. (A) Mean trait measurements for ancestors and evolved lines for 3 of the 7 measured traits (see Fig. S1 for all traits). Phenotypes of evolved lines (1–12) are shown only as measured in their localtemperature regime (cold lines, 1–6, measured at 23°C; hot lines, 7–12, measured at 35°C). Means and one standard error are shown. (B) Manhattan plots of rolling average log odds p-values (window size = 20 SNPs) for example populations (see Fig. S2 for all populations) across all putatively selected sites (n=475,194) using `R` package `RcppRoll` (v.0.3.0). The largest 10 scaffolds (i.e., chromosomes) are labelled. The 0.001th quantile of rolling average p-values is given by the horizontal red line for each population. SNPs whose rolling average p-values are also above their respective 0.001th quantile in other populations are colored. There is a clear polygenic signal of adaptation. Note that the California reference populations shown here exhibit many more SNPs in common with the single other California replicate (yellow points) than those shared with any of the four replicates of different origin (green and orange points).

Genomic data and analysis

For evolved populations, we isolated genomic DNA from pools of 60 male adults between generations 57-59 for cold-adapted lines and generations 67-69 for hot-adapted lines. Within each thermal regime, populations were sampled at the same time but varied in the number of generations of adaptation due to differences in development rate (Fig. 1A). Due to a paucity of beetles remaining after line-establishment, we were not able to sample the direct establishing ancestor. Instead, the third generation of heat-adapting populations were sampled and sequenced as a proxy of the ancestor (hereafter simply referred to as 'ancestor' when considering genomic data); hence comparisons between evolved and ancestral lines represent 60-62 and 64-66 generations of divergence for the cold and hot regime, respectively. Adult beetles were collected and stored at -80°C. We purified DNA from the beetles using the King Fisher Cell and Tissue DNA Kit (Thermo Fisher Scientific). DNA fragment libraries for whole-genome sequencing (WGS) were generated at the Science for Life Laboratory (Stockholm, Sweden) from each pool by following the standard Illumina TrueSeq PCR-free library prep protocol in which 350bp fragments were retained. Purified genomic libraries were then sequenced on the Illumina NovaSeq 6000 platform. In total, two lanes of 150bp paired-end reads were generated.

We used `BWA mem v0.7.9` [49] to align the ∼5.8 billion ∼150-bp DNA sequences to a genome assembly for *C. maculatus* (accession: `PRJEB60338`). After sorting, merging, and de-duplicating (`Picard Tools v2.23.4`, Broad Institute) our alignments, we then identified SNPs using the GATK best practices recommendations [50]. Variants were initially filtered based on GATK hard-filtering recommendations after manual and visual inspection of the data with the following parameters: $QUAL < 20$, $QD < 2$, $FS > 20$, $SOR > 5$, $MQ < 40$, $-5 > MQRankSum > 5$, $-5 > ReadPosRankSum > 5$. A second round of filtering was performed on regions of the genome that do not uniquely map to itself using `GenMap v2.4.1` (with parameters K31,E2) [51], sites with missing data across any sample, or $DP < 20$ across any sample. We retained 10,611,665 variants after filtering (∼10 variants per kb).

To measure the strength of directional selection, we first identified candidate loci consisting of putatively selected SNPs. To determine whether

a SNP is likely under selection, we incorporated temporal information of allele frequency change from the ancestor to evolved hot and cold lines. We estimated the variance effective population size (N_e) during the experiment for every replicate from patterns of allele frequency change from the total set of ∼10 million SNPs using the `R` package `poolSeq` [52]. Given the estimated values of N_e, we then parameterized and tested a null model of evolution by genetic drift to determine the probability of observing the magnitude of allele frequency change for each SNP in each population. We calculated the probability of the observed allele frequency change from the ancestral to evolved populations based on a beta approximation to the basic Wright–Fisher model [53]. Following [54], we assumed $p_t|p_0 \sim \text{beta}(\alpha + 0.001, \beta + 0.001)$, where $\alpha = p_0\frac{1-F}{F}, \beta = (1-p_0)\frac{1-F}{F}$, p_0 and p_t are the allele frequencies at the beginning and end of the interval, $F = 1-(1-\frac{1}{2N_e})^t$, t is the number of generations between samples. We retained SNPs with allele frequency changes more extreme than the 0.01th or 99.99th percentiles of the null distribution per SNP in any population (Fig. 2B), resulting in a total set of 475,194 putatively selected sites among all replicates. Estimation of selection coefficients was performed with `CLEAR` [55]. We generated the input for `CLEAR` from the set of putative SNPs with scripts from the PoPoolation2 pipeline [56]. Estimations of N_e, null model drift estimates, and selection coefficients all require information on the number of generations elapsed between timepoints; for each replicate, we therefore added or subtracted 3 generations for cold- and hot-adapted lines, respectively, to the generation parameter to account for drift that occurred between ancestral lines and our proxies of the ancestors.

Repeatability at the SNP level was measured, similarly to the phenotypic data, by geometric estimates of θ and divergence ($E_d - S_d$) of allele frequency change vectors between populations of the set of putatively selected sites. Note that because two replicate ancestral samples were available per background, error corrections were not implemented. Because we are interested in parallelism due to selection, we focused on the repeatability of evolutionary change occurring in the set of putatively selected SNPs. Repeatability in which genes were targeted by selection between populations was measured by first identifying the set of genes in which selected SNPs fell within protein coding regions, and subsequently calculating pairwise Jaccard

indices ($J(A,B) = \frac{A \cap B}{A \cup B}$) between populations.

Repeatability and historical contingency for different classes of alleles

To explore if certain types of variants are more or less likely to contribute to repeated instances of thermal adaptation, we first identified SNPs which were assigned as being under selection in both replicate lines deriving from a particular origin and regime. Because of this stricter criterion, we relaxed our null drift expectations to include SNPs with allele frequency changes more extreme than the 1st or 99th percentiles of the null distribution, resulting in 271,515–617,395 candidate SNPs being selected. For each genetic background and evolution regime separately, SNPs were then assigned into four categories based on the patterns of evolutionary change observed between regimes: Synergistically pleiotropic (SP) SNPs were those that were selected for in the same direction across regimes, antagonistically pleiotropic (AP) SNPs were those that were selected for in opposite directions across regimes, while private cold (PC) and private hot (PH) SNPs were those selected for in cold- or hot-adapted populations only (Fig. 4A). We then compared the relative abundance of SNPs that showed overlap between the three genetic backgrounds for each category. The same analysis was carried out at the gene-level based on if the SNP fell within the protein coding region of a gene. The biological processes for each gene set were analyzed for enrichment using the `R` package `topGO` (v.2.38.1) and `weight01` algorithm.

We tested whether there was more overlap between origins per SNP classes than expected by chance by performing forward in time drift simulations from all ancestral populations (`R` package `poolSeq` v.0.3.5.9). During each simulation iteration (n=1,000), 10,000 randomly sampled SNPs were sampled to parameterize all lines. Each line was parameterized with the ancestral allele frequencies, the number of generations evolved, and the effective population sizes of respective lines. SNP classes and overlap between lines were calculated as described previously. The observed overlap in SNP classes was tested using a two-sided test to determine significance.

Genomic and phenotypic/fitness offsets

To assess our ability to predict phenotypic evolution from genomic data, we calculated "offsets" for both kinds of data, describing the level of genetic and phenotypic maladaptation relative to a well-adapted reference population [57, 58]. Offsets were calculated for each geographic origin and evolution regime separately. Thus, for each background-regime combination, the reference was chosen as the line replicate with the highest fitness at the temperature to which it had been adapting. We here defined our proxy of fitness as the ratio between lifetime reproductive success and egg-to-adult development time, roughly corresponding to offspring produced per day. Thus, for example, for hot-adapted lines, Yemen 8 had higher fitness (3.65) than Yemen 7 (2.88) at 35°C and was chosen as the reference line for all 6 Yemen populations phenotyped at 35°C. The set of SNPs used to calculate genomic offsets were those putatively selected for in the reference line (more extreme than the 0.01th or 99.99th quantiles of the null drift distributions, see above). Genomic offsets were calculated as the sum of absolute differences between reference and tested lines. The difference in each SNP's allele frequency was scaled by the absolute selection coefficient of that SNP calculated in the reference population such that, for example, SNPs with large differences in allele frequency but relatively small selection coefficients would not contribute greatly to the genomic offset. Phenotypic offsets were calculated as the Euclidean distance of the 7 mean-scaled traits between reference and tested lines. Additionally, we used our fitness metric, $\frac{LRS}{Dev.Time}$, to calculate a relative fitness offsets, again between the reference and tested lines.

Results

Phenotypic Evolution is faster and more parallel at hot temperature

Phenotypic change vector analysis showed that hot-adapted lines exhibit higher evolutionary rates per-generation ($\|\bar{x}_{35}\| = 0.87 \pm 0.14$) compared to cold-adapted lines ($\|\bar{x}_{23}\| = 0.5 \pm 0.07$) ($t_5 = -4.01$, $p = 0.003$, Fig. 3A). Phenotypic changes were also more parallel ($\theta \pm SD$) on average (permutation test, $p < 0.001$) in heat-adapted lines (39.32 ± 19.16) than cold-adapted

lines (67.42 ± 23.3), consistent with the hypothesis that heat should impose greater selection than cold (Fig 3B). Compared to Monte-Carlo simulations of random angles in 7-dimensional space, the mean pairwise angles for both cold- and heat-adapted lines were more parallel than expected by chance (two-sided test, $p < 0.05$, 10^5 iterations). Moreover, parallelism was greater within than between genetic backgrounds for both heat- and cold-adapted lines (Table S3), suggesting a role of historical contingency in dictating the repeatability of phenotypic evolution. The major components contributing to variation in evolutionary trajectories were weight loss in cold-adapted lines and lifetime reproductive success, adult weight, and weight loss in heat-adapted lines (Fig. S5).

Despite increased parallelism during adaptation to heat, cold lines generally exhibited less divergence (or greater convergence) relative to hot lines (Fig 3B) ($Div_{23} = -0.14 \pm 0.29$; $Div_{35} = 0.22 \pm 0.20$, permutation test, $p < 0.001$). While this result, to some extent, can be explained by the overall smaller magnitudes of per-generation phenotypic change in cold-adapted lines relative to hot-adapted lines, it also suggests that non-parallelism in cold populations in part resulted from convergent evolutionary trajectories of differentiated ancestral populations, making cold-evolved lineages become more similar to each other in some phenotypic traits during experimental evolution (Fig. 2A).

The genetic basis of thermal adaptation is polygenic and governed by private alleles

Consistent with stronger selection at warmer temperatures, hot-adapted lines experienced greater rates of allele frequency change at candidate loci and were estimated to have lower N_e than cold lines (Table S1). This pattern was similar for selection coefficients which take into account drift based on estimates of N_e, with heat-adapted lines experiencing stronger average selection than their cold counterparts on all three backgrounds (Table S1).

Thermal adaptation was highly polygenic, with the number of putatively selected SNPs ranging from 21,551–86,561 across lines. The same estimates for genes ranged from 774–2,805 (Table S2). Although it is unlikely that all putatively selected SNPs are causal, several distinct peaks

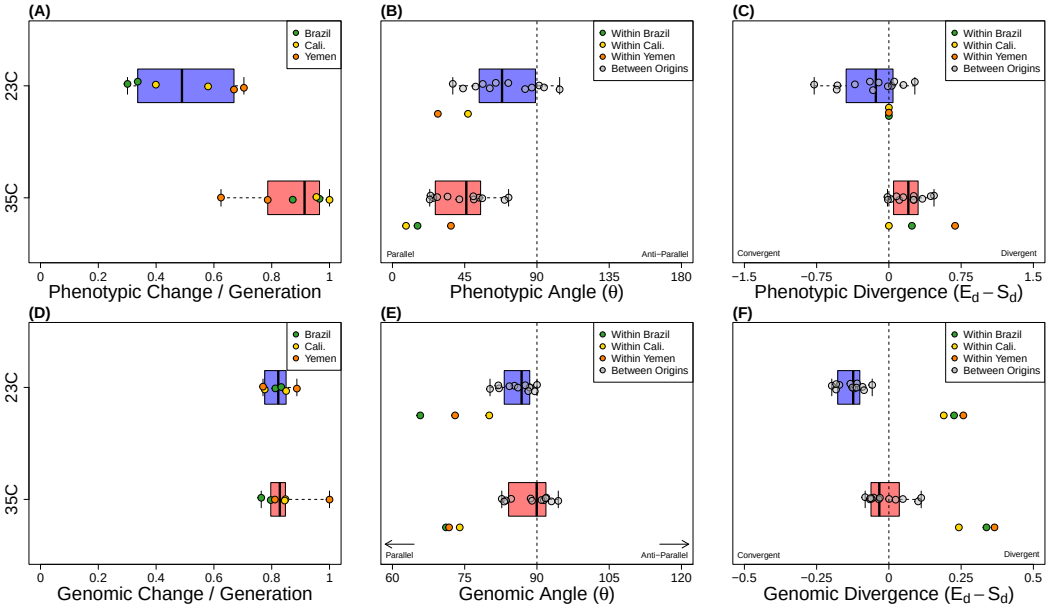

Figure 3: Distribution of all phenotypic and genomic evolutionary rates (A & D), pairwise angles (B & E) and measurements of convergent or divergent evolution (C & F) between lines within evolution regimes. All genomic estimates are based on the set of putatively selected SNPs (n=475,194). Angles and divergences are given for pairwise comparisons between populations of different (grey points) and same (colored points) genetic background. Angles of 0° are considered completely parallel, while angles of 180°are considered completely anti-parallel. Divergence measurements are calculated as the difference in Euclidean distance between population-pairs at the end (E_d) and start (S_d) of experimental evolution, and are divergent when the distance between end points is greater than the distance between start points ($E_d - S_d > 0$), or convergent if the opposite is true ($E_d - S_d < 0$). Evolutionary rates have been scaled by the maximum evolutionary rate in the dataset, while divergences have been scaled by the maximum starting Euclidean distance in the dataset (i.e. relative to the distance between the two most differentiated ancestors). Inestimable phenotypic angles due to sampling error are not shown (Within Brazil, 23°C).

can be observed of SNPs deviating from drift expectations across almost all chromosomes in all populations (Fig. 2B, S2).

Across all six cold-adapted populations there were 296 shared genic targets of selection, which are many more than expected by chance (expected $= 2.33$, $p < 0.001$), while hot-adapted populations shared 51 genic targets (expected $= 0.11$, $p < 0.001$). The 296 cold adaptation genes were enriched for such terms as regulation of Rho protein signal transduction, growth, and homophilic cell adhesion, while the 51 heat adaptation genes were enriched for homophilic cell adhesion, microtubule-based movement, hexose biosynthetic process, and mannose metabolic process (Table S5).

Contrary to typical theoretical assumptions, overall patterns based on the four classified variant categories (i.e. Synergistically Pleiotropic SNPs (SP), Antagonistically Pleiotropic SNPs (AP), SNPs Private to Cold adaptation (PC), and SNPs Private to Heat adaptation (PH), Fig. 4A) indicate that thermal adaptation is not governed by antagonistic pleiotropy. Instead, thermal adaptation was primarily characterized by privately selected alleles (range PC SNPs: 12,215–38,799; range PH SNPs:10,763–26,672). There were also more SNPs evolving in the same direction between temperature regimes (range SP SNPs: 517–3,418) than in opposite direction (range AP SNPs: 617–1,966). Compared to forward in time simulations from ancestral populations (`R` package `poolSeq` v.0.3.5.9), there were more SNPs identified in each variant category for each genetic background (i.e. shared between the two line replicates within each background) than expected solely due to drift (100,000 simulations per population, all $p < 0.001$). However, most candidate SNPs were unique to a particular genetic background (Fig S3). There was only a significant excess of shared SNPs between genetic backgrounds in the PC set (range: 76–656) and PH set (range: 23–70), and no SNPs selected for among all three backgrounds.

At the level of genes (Fig 4B), we observed more SP variants than expected by chance in all pairwise comparisons across backgrounds (range: 3–18), including 3 shared genes between all backgrounds ($p < 0.05$). Significant overlap in AP genes was only observed between California and Yemen (n=12, $p < 0.01$). Among PC (range: 133–252) and PH (range: 37–98) gene sets, there were an excess of shared variants among all pairwise origins and between all three origins (all $p < 0.01$). Hence, thermal adaptation seems

best described by polygenic adaptation from "cryptic genetic variation" with effects private to each temperature, and modest repeatability across genetic backgrounds for heat-adaptation, but higher repeatability when evolution proceeded at cold temperature.

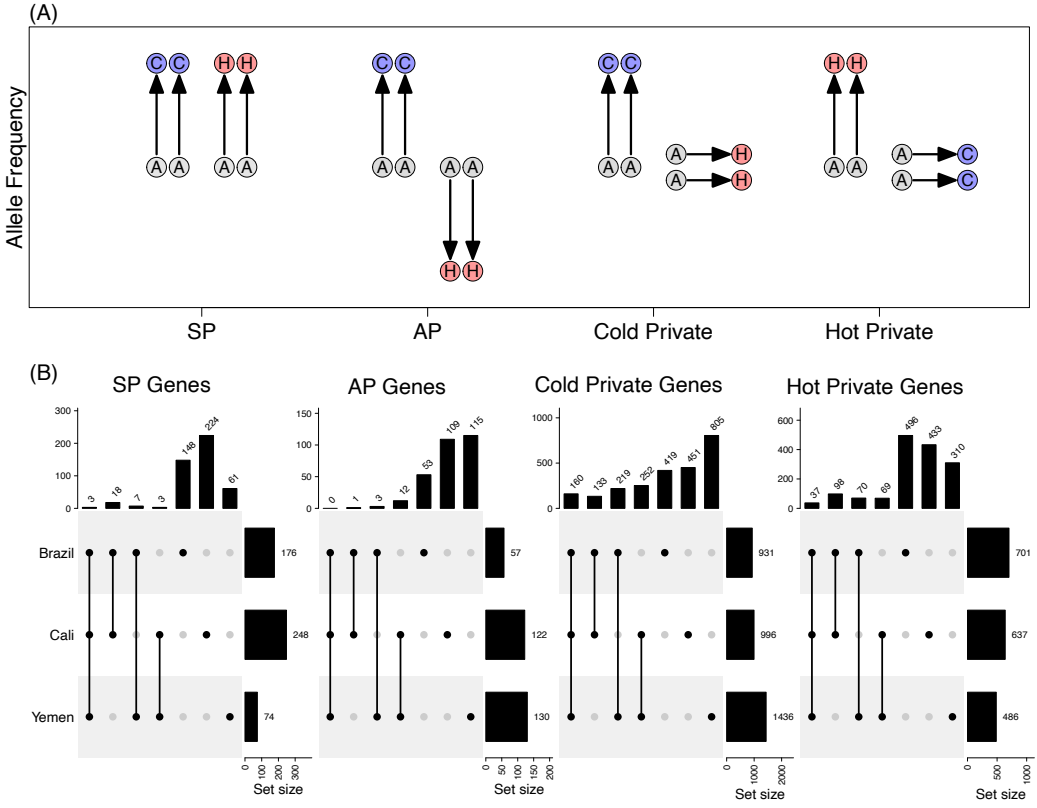

Figure 4: (A) Schematic showing how SNP and gene classes were assigned. For each class, we require that both replicates of a particular background behave similarly (i.e., more extreme than the 1st or 99th percentiles with the same direction of allele frequency change in both replicates). The four classes identified are: 1) Synergistically pleiotropic (SP) SNPs which are selected for in the same direction between regimes, 2) antagonistically pleiotropic (AP) SNPs which are selected in opposite directions between regimes, 3) private cold SNPs which are selected for only in the cold regime, and 4) private hot SNPs which are selected for only in the hot regime. (B) Upset plots of gene sets based on corresponding SNP set classes (see Fig. S3) located in protein-coding regions. There is significant excess of shared genes for private genes, and more gene sharing among genetic backgrounds adapting to cold temperature.

Of the three SP genes which were shared across all origins, we identified a diacylglycerol kinase protein, which has been shown to be involved

in thermosensory behavior in *Caenorhabditis elegans* [59]. Although no AP genes were shared between all origins, between pair-wise comparisons of origins, we identified a cGMP-dependent protein kinase which is associated with thermotolerance in *D. melanogaster* larvae [60], the gene *couch potato* (*cpo*) which is associated with diapause climate adaptation in both *D. melanogaster* [61] and *Culex pipiens* [62], as well as a MAPK-binding protein related to homeostatic stress responses [63].

Gene set enrichment analysis revealed some biological processes of interest among gene set categories (Fig S4). Both AP and SP variants were over-represented for processes such as oxidation-reduction processes, double-strand break repair, and DNA recombination, which may indicate shared general stress responses to suboptimal temperature [33]. Among AP and SP sets, the only biological processes enriched among more than one background was homophilic cell adhesion. The three genetic backgrounds shared many more processes for private (PC and PH) gene sets. PC variants were enriched for such terms as phospholipid metabolic process, chitin metabolic process, transmembrane transport, and growth. PH variants were enriched for such terms as DNA topological change, proteolysis, and negative regulation of cell cycle. Many terms were also shared between the PC and PH set, including regulation of Rho protein signal transduction, Wnt signalling pathway, chitin metabolic process, and ion transport. Notably, with the exception of the Brazil AP gene set, the GO term "Homophilic cell-adhesion via plasma membrane adhesion molecules (GO:0007156)" is found in all gene set categories across all backgrounds. This GO term was also identified in several *Anolis* species in relation to adaptation of thermal niches [64] and cold tolerance in *D. melanogaster* [65]. As the mechanical and functional properties of the plasma membrane are greatly influenced by thermal fluctuations [66], it seems likely that this biological process is a strong candidate for shared targets of selection across thermal gradients.

Genomic evolution is more predictable at cold temperature

Despite greater phenotypic parallelism in heat-adapted lines, hot-adapted lines did not exhibit greater parallelism at the genome level when comparing changes in SNP frequencies ($\bar{\theta}_{23} = 83.42 \pm 6.7$, $\bar{\theta}_{35} = 85.48 \pm 7.8$, permu-

tation test, $p = 0.15$) (Fig 3C). Despite relatively orthogonal measures of parallelism compared to phenotypic estimates, which partly can be ascribed to the difference in dimensionality of the two analyses, the observed pair-wise angles among both cold- and heat-adapted populations were smaller than expected by chance (Monte-Carlo simulations, $p < 0.001$, 10^5 iterations). Just as for the estimates of phenotypic repeatability, within-background comparisons were more parallel ($\bar{\theta}_{23} = 72.96 \pm 7.16$; $\bar{\theta}_{35} = 72.28 \pm 1.5$) than between-origin comparisons, and especially so for heat-adaptation ($\bar{\theta}_{23} = 86.03 \pm 3.25$; $\bar{\theta}_{35} = 88.77 \pm 4.19$), suggesting an important role of historical contingency in dictating genomic repeatability (Table S4)

Cold lines also tended to diverge less (or converge more) in genomic space than Hot lines (permutation test, $p = 0.011$). This pattern is true both within ($Div_{23} = 0.22 \pm 0.03$; $Div_{35} = 0.32 \pm 0.06$) and between backgrounds ($Div_{23} = -0.13 \pm 0.04$; $Div_{35} = -0.01 \pm 0.06$) (Fig 3F). Unlike phenotypic rates of change, genomic rates of change in selected loci were not different between hot- and cold-adapted lines ($t_5 = -0.59$, $p = 0.57$ (Fig. 3D).

Similar to our other estimates of genomic repeatability, Jaccard indices of the overlap of candidate genes under selection between backgrounds (Table S8–S9), indicated that cold lines show greater repeatability (0.32 ± 0.06) compared to hot lines (0.20 ± 0.06) (permutation test, $p < 0.001$) (Fig S6). A simulated model of genetic drift parameterized by initial allele frequencies and estimates of Ne showed that the Jaccard indices calculated were all greater than expected solely under drift (Two-sided test, $p < 0.001$).

The mechanistic basis of temperature-specific evolutionary predictability

That we see greater phenotypic repeatability for hot lines, but higher genomic repeatability across cold lines for all genomic metrics is counter-intuitive and opposite of our expectations. We propose three non-mutually exclusive mechanisms that may contribute to such patterns. First, due to stronger selection at high temperature, hot lines also experience more genetic drift, as evidenced by lower Ne. This could contribute to lower predictability of genomic changes across lines. However, Ne was not strongly

correlated with the repeatability of observed evolutionary trajectories (Fig. S7A–B) nor with divergence at candidate loci within regimes (Fig. S7C–D).

Second, differences in ancestral polymorphism at selected loci may contribute to historical contingencies and low repeatability across geographically separated populations [23]. To test this hypothesis, we re-estimated all measures of genomic repeatability based on only those putatively selected SNPs that segregated in all ancestral populations (n=119,630, or ∼25% of the original set), expecting to see significantly higher repeatability, especially for heat-adapted lines, if differences in ancestral polymorphism have contributed to the observed temperature-specific patterns. Across the original set of SNPs, the three ancestral backgrounds were differentially fixed (i.e., fixed in opposite directions) for between 739–1,625 of the (putatively) selected SNPs, and between 4,225–10,618 (0.04-0.11%) of all SNPs across the genome. However, although genomic repeatability was slightly increased overall for shared polymorphic sites, the greater effect of genetic background on parallelism and divergence observed for heat- relative to cold-adaptation remained (Fig. S9), Table S6). This pattern holds true even under a stricter criterion of shared polymorphic candidate SNPs, in which the top 100 SNPs were required to be at least 10kb away from each other per line to reduce the effects of linkage (Table S7). Additionally, when comparing the number of candidate genes in the PC and PH gene sets (which contain the majority of candidate polymorphisms underlying thermal adaptation), we found no difference in the degree of overlap between the genetic backgrounds with respect to using all SNPs or only those shared among all ancestors (Fig S8, all Chi-square test $p > 0.05$). Hence, differences in shared ancestral polymorphisms seem to have played a minor role in shaping the observed genomic repeatability between temperatures.

Third, it is well recognized that gene interactions (epistasis) can contribute to unpredictable evolutionary outcomes [67], and that such epistasis might depend on the environment [27, 68]. Indeed, epistasis is predicted to be a natural property of metabolic networks, and as all metabolic processes are strongly temperature-dependent, it is possible that epistasis between differentially fixed variants in the three genetic backgrounds and selected variants may have been stronger at hot temperature. In Fig 5 and Supplementary (Fig. S12), we illustrate that our temperature-specific estimates of genomic

adaptation (i.e., excess of PC and PH genes) and repeatability (lower at hot temperature) are consistent with predictions from a simple model of thermal adaptation assuming exponential increases in both enzyme reaction rates (Fig. 5A) and protein unfolding (i.e. loss of enzyme stability; Fig. 5B) with increasing temperature [38, 69]. This heuristic model has been used previously to argue that hot temperature results in increased epistasis by causing proteins to evolve marginal thermostability, such that further mutational or thermal perturbations can be either inconsequential or highly deleterious depending on the wildtype protein [38, 39, 70–72]. Nevertheless, the mechanistic basis underlying the observed temperature-specificity of evolutionary repeatability needs to be investigated further.

Does Genomic Evolution Predict Phenotypic Evolution and (mal)adaptation?

We tested whether genomic offsets between populations are predictive of phenotypic evolution and (mal)adaptation. We focus on offsets for high heat (see Fig. S10 for cold offsets) as we see more clear signs of local adaptation to heat than cold and because phenotypic estimates of cold-adaptation were smaller and therefore estimated with lower precision (Fig. S1).

Genomic offsets were strongly predictive of phenotypic and fitness offsets at hot temperature (Fig. 6) for lines of the same geographic origin as the reference ($\bar{r}_{Pheno} = 0.88$, ANCOVA $F_{1,9} = 13.6$, $p = 0.005$; $\bar{r}_{Fit} = -0.74$, ANCOVA $F_{1,9} = 7.23$, $p = 0.025$). However, when applied to predict adaptation in lines of different geographic origin, the predictive power of genomic data was severely reduced ($\bar{r}_{Pheno} = 0.22$, ANCOVA $F_{1,30} = 0.86$, $p = 0.36$; $\bar{r}_{Fit} = -0.29$, ANCOVA $F_{1,30} = 1.24$, $p = 0.27$). Hence, selected sites identified in a particular origin do not contribute greatly to phenotypic or fitness variation in other origins, which is congruent with estimates of both phenotypic and genomic repeatability being greater within than between origins (Fig. 3).

Interestingly, the genomic offsets were typically larger for within than between population comparisons. This suggests that, while selected sites identified in a given reference line contribute to fitness/phenotypic variation among lines of the same origin, these sites may be mostly fixed (in the di-

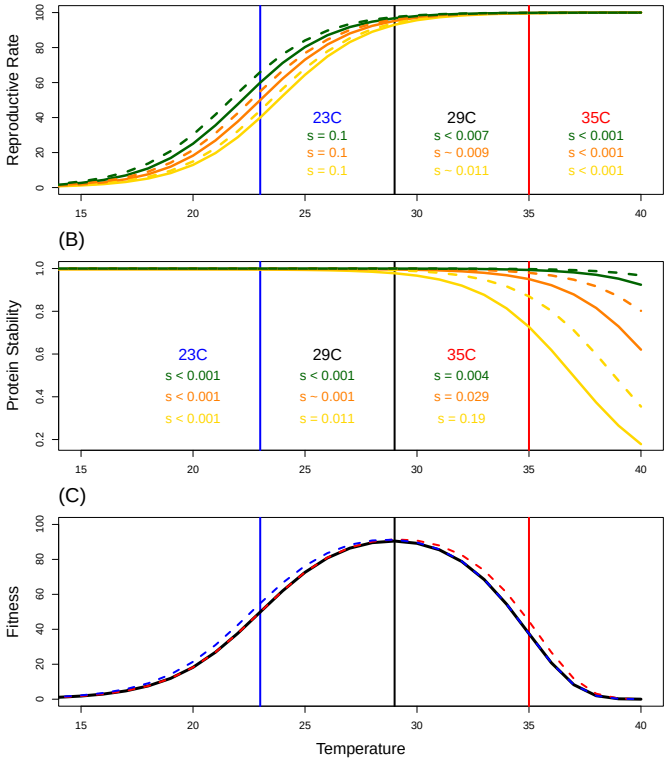

Figure 5: A prediction of temperature-dependent epistasis for fitness based on thermodynamics of enzyme reaction rates and protein folding stability. For each plot, colored vertical lines denote the temperature of our thermal regimes (23°C and 35°C). **A)** Reproductive rate increases exponentially with temperature at the colder range due to relaxation of thermodynamics constraints on enzymatic reactions, but starts to follow a pattern of diminishing returns at warmer temperatures due to ecological constraints placing limits on reproductive output. Illustrated for three genetic backgrounds (green, orange and yellow) with wildtypes (solid lines) and a mutant (broken lines) with a 10% increase in enzymatic reaction rate at 23°C. This results in strong selection on the mutation at cold temperature for all backgrounds as these are far of their optimal reproductive rate, but weak selection at hot temperature at which all backgrounds are close to the maximal achievable reproductive rate determined by ecological factors. Epistasis is weak and selection is similar across backgrounds at all temperatures. **B)** Warm temperatures increase a range of molecular failure rates, here illustrated for the loss of protein stability, which leads to less enzyme ready to catalyse reactions, and more misfolded proteins within cells, leading to protein toxicity and depressed fitness. Illustrated for three genetic backgrounds with stable (green), intermediate (orange) and unstable (yellow) wildtype protein. A mutation increasing stability (broken lines) was introduced on each background. Selection on the mutation is weak at cold temperature for all backgrounds but can become very strong at hot temperatures, depending on the wildtype protein stability, and strong epistasis for fitness results. **C)** Protein fitness, as the product of reproductive rate and molecular failure rate, is depicted for the wildtype (black), and the mutants with increased enzymatic reaction rate (blue) and increased protein stability (red) for the yellow background.

rection of the reference line) in the other origins. It could be, then, that by focusing on SNPs which show putative signs of selection in a limited set of study populations, the importance of already-fixed sites that cannot contribute towards genomic offsets may be greatly underestimated in genomic prediction of (mal)adaptation in geographically distant populations. To explore if this possibility contributed to the poor genomic predictions of maladaptation applied across origins, we also estimated the correlation between the fitness-offsets and the genomic offsets basing the latter on all candidate SNPs per thermal regime (i.e. identified in any of the six replicates), with the reasoning that this would capture and include differentially fixed sites. This resulted in smaller genomic offsets of lines of the same origin as the reference (i.e. lines being more similar to the reference) compared to lines of different origin, demonstrating the more parallel genomic response to selection within origins relative to between origins. However, even when all putatively selected sites were considered, predictive power remained high in the case of within-origin estimates ($r = -0.87$), and low for between-origin estimates ($r = 0.32$) (Fig. S11B), again suggesting that differences in ancestral polymorphisms was not *per se* causing the poor predictability across genetic backgrounds.

Discussion

Climate change is disrupting ecological niches and changing species-distribution patterns across the globe [73, 74]. For many taxa, the potential to genetically adapt to these changes is necessary not to face extinction. Approaches to harness genomic data to evaluate extinction risk and adaptive potential of threatened populations are becoming more common, but their limitations are debated [15, 17, 18, 75, 76]. Here, we adapted seed beetles from three different geographic origins for up to 65 generations to cold and hot temperature and assessed both life-history and genomic adaptation to 1) determine the genetic basis of thermal adaptation, 2) quantify the repeatability of evolution across biological levels of organization, and 3) evaluate the prospects of predicting adaptive potential across geographically isolated populations from genomic data. We show that adaptation to both cold and hot temperature is highly polygenic. Evolutionary responses were more re-

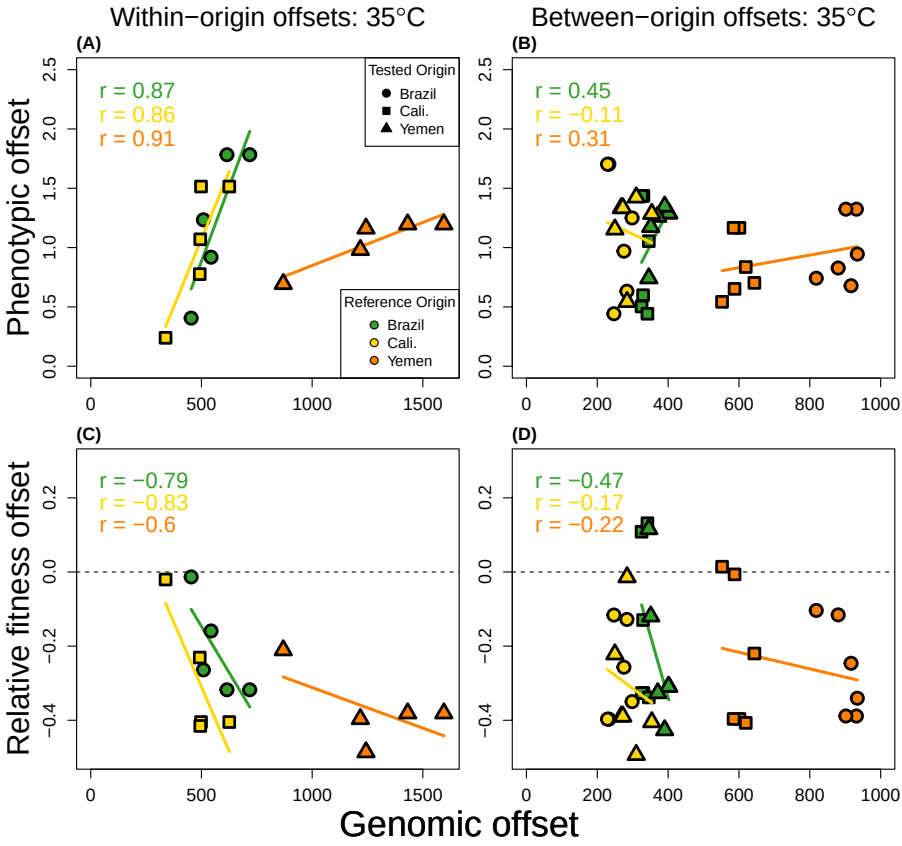

Figure 6: Genomic predictors of phenotypic and relative fitness offsets at hot temperature (35°C). Genomic offsets were calculated per origin based on SNPs whose p-values for allele frequency change fell within the top 0.001th quantile (n=10546–10887). Phenotypic offsets were calculated as the Euclidean distance in scaled trait-space relative to the reference population. Relative fitness offsets were calculated as the relative fitness ($\frac{LRS}{Dev.Time}$) to the reference population. Offsets are organized by predictions within genetic backgrounds (A,C) and between genetic backgrounds (B,D). Correlations and regression lines are colored by reference population. Genomic offsets predict maladaptation within, but not between, backgrounds.

peatable within, than between, genetic backgrounds at both phenotypic and genomic levels, demonstrating an important role of historical contingency in dictating evolutionary predictability. Heat adaptation resulted in both greater phenotypic and genomic rates of change, indicative of high evolutionary potential under future climate warming. However, despite the high repeatability of phenotypic evolution at hot temperature, adaptation at the genomic level was more repeatable at cold temperature, and the effect of genetic background played a relatively larger role for heat-adaptation. As a result, genomic offsets for heat-adaptation, while predicting the level of maladaptation of lines from the same origin, failed to predict maladaptation in lines of different origin. This suggests that the prospects of using data on genomic-phenotype correlations to predict more broad-scale geographic patterns of (mal)adaptation and evolutionary potential under future climate warming may be limited. Complementary hypotheses exist to explain the observed patterns of evolutionary repeatability, which we discuss below.

Genetic basis of thermal adaptation

Thermal adaptation was found to be highly polygenic across all lines and regimes, with several thousand polymorphisms under putative selection, which is predicted to result in high genetic redundancy and low repeatability of adaptation at the genomic level [13, 77]. Although not all selected SNPs are likely causal and are more likely in linkage with true causal sites, several distinct peaks of SNPs deviating from drift expectations were observed across all chromosomes in all lines (Fig. 2B; S2). The patterns of identified SNP classes indicate that thermal adaptation is largely characterized by alleles selected specifically for each thermal regime. Such privately selected alleles comprised the majority of targets, with relatively few alleles changing in the same (SP variants) or opposite direction (AP variants) between thermal regimes. Thus, our findings suggest that evolutionary potential under climate change may be fuelled by alleles with conditional fitness effects that are relatively abundant at ancestral and benign thermal conditions but become exposed to selection at thermal extremes. This notion also corresponds well with predictions of temperature-dependent selection on enzymes based on the thermodynamic basis of protein function [38, 78, 79] (and Fig 5).

A possible exception to this general pattern is the observed divergence along one genomic dimension, explaining roughly 6% of all the variation at selected sites, that occurred in opposing directions for hot and cold lines (Fig. 1E, genomic PC3). This is in contradiction to our analyses of single SNPs and genes, where very few sites showed signs of antagonistic pleiotropy (i.e. the lack of candidate AP genes noted above). It is well recognized that polygenic signals of adaptation can result in stochastic and modest allele frequency changes at single sites, where allelic effects often show epistasis and redundancy for fitness during the process of adaptation, causing a highly unpredictable genomic signal at the level of SNPs [13, 80]. While this points to the general problem of detecting polygenic selection at weakly selected loci, the result also highlights that overall genomic signals can still be picked up and be informative of the basis of adaptation, given sufficient experimental replication.

Our results mirror experimental evolution studies in both *Drosophila melanogaster* [81] and *D. simulans* [25, 82] in which adaptation was polygenic and few targets of selection overlapped between opposing thermal selection regimes. Traditionally it has been assumed that thermal adaptation falls along a single trait optimum axis [83], but our results in conjunction with similar experimental evolution studies [25, 81, 82], suggest that thermal adaptation falls along multiple axes. It is likely that even different magnitudes of cold or heat stress will require differing physiological and mechanistic solutions [84, 85]. Moreover, while immediate plastic responses (physiological and transcriptomic) to chilling temperatures and heat shock have been studied, whether the mechanisms of immediate plastic responses to thermal extremes differ from those targeted by long-term selection is variable among species [86, 87].

(Non)parallelism in adaptation to thermal stress

Evolutionary trajectories were generally more parallel for comparisons within, than between, backgrounds for both phenotypic and genomic adaptation, demonstrating an important role of historical contingencies in shaping evolutionary outcomes. The importance of genetic background for parallelism also varied between thermal regimes (Table S3). For example, Yemen lines ex-

hibited the highest within-origin parallelism at 23°C, but the lowest within-origin parallelism at 35°C. This may indicate that the genetic backgrounds differ in their optima for cold and heat tolerance, as in [82], despite > 200 generations of shared lab adaptation at 29°C. Similar results were found in studies of temperature adaptation in different founders of *Drosophila simulans*, in which the degree of parallelism was determined primarily by the genetic founder of the replicated populations [25, 82]. Likewise, studies of wild populations have also found significant, but modest overlap between climate adaptation loci among independently evolving lineages [88–91].

Biophysical models of protein stability predict that higher temperatures should generate stronger purifying selection as a function of increased constraint on protein evolution [38, 39, 79] and increased positive selection on alleles that buffer thermal stress [33, 92]. Following this, we expected both faster and more repeatable adaptive evolution at hot temperature [6]. We found evidence at the genomic level that high heat indeed imposed stronger directional selection based on estimated selection coefficients (Table S1). However, contrary to our expectations, we observed that heat-adaptation was only more parallel for phenotypes. The efficacy of selection in producing repeatable genomic changes may have been compromised by increased genetic drift at high temperature. However, correlations between estimates of *Ne* and parallelism at both genomic and phenotypic level were weak, suggesting that genetic drift may not be solely responsible for the observed patterns. It is also possible that differences in ancestral standing genetic variation among the three geographically separated backgrounds may have contributed to historical contingency and different evolutionary outcomes in our experiment [23]. However, our estimates of repeatability based only on those SNPs that were shared between all ancestral populations suggest that the observed patterns are unlikely to be solely explained by differences in the access to (adaptive) standing genetic variation in the three ancestors. As a final possibility for the observed temperature-specific genomic repeatability, epistasis between selected SNPs and differentially fixed variants in the three ancestors may have been magnified at hot temperature. Epistatic interactions between different proteins or between different SNPs within the same protein coding sequence are ubiquitous [27, 93, 94], and temperature-dependent increases in such interactions are emergent proper-

ties of the assumptions in simple biophysical models of protein thermodynamics and function [38, 72, 79] (Fig. 5) and have been observed in empirical studies [95]. Although we cannot explicitly test this hypothesis with the current data, the low explanatory power of both genetic drift and differences in ancestral polymorphism, together with the larger difference in repeatability between the hot and cold regime observed between (compared to within) genetic backgrounds, suggest that temperature-dependent epistasis may be causally linked to the lower genomic repeatability observed at hot temperature. This suggests, somewhat paradoxically, that the same thermodynamic mechanism that causes hot thermal extremes to impose strong selection, resulting in repeatability of adaptive outcomes at the phenotype level, may at the same time make genomic evolution more unpredictable by increasing epistatic interactions.

Predicting (mal)adaptation from genomic data

Leveraging genomic datasets to predict the future fates of threatened populations facing climate change is an important aim in ecological and conservation genetics. However, both genetic redundancy [25] and epistasis [96] obfuscate the genotype-to-phenotype map, potentially reducing the power of using genomic data to make inferences about adaptive potential. In accordance with this notion, and the previous observations of a strong influence of genetic background on heat-adaptation, our genomic offsets were predictive of phenotypic maladaptation to heat when applied within a single genetic background, but poor predictors when applied across genetic backgrounds. This suggests that predictions of adaptive potential under future climate change may be problematic when applied on large geographic scales, maybe to the extent that standard estimates of (non-specific) genome-wide genetic diversity may offer similar accuracy in predictions. Indeed, a comparison of several methods for conducting genomic offsets found that sets of randomly selected SNPs better predicted performance than identified candidate SNPs [97].

A possible explanation for low predictability between genomic and phenotypic data is that many changes in allele frequency may not be causal for changes in phenotype. Our seed beetle populations experienced low effec-

tive population sizes ($\bar{N}_e = 215.4$) and strong selection, which likely resulted in both relatively large allele frequency changes due to genetic drift and genetic draft, resulting in skewed associations of genomic change underlying phenotypic change. These results highlight the importance of candidate SNP detection. However, exploring alternative methods of constructing genomic offsets in our lines often gave variable, and worse, predictive power.

The particular metrics used for phenotypic and performance measures should also be given consideration [57, 58]. While whole-genome analyses of predictability can typically incorporate all individual 'units' which compose the response to selection (i.e. changes in allele frequency across the genome), phenotypic or fitness analyses can be sensitive to the set of traits measured [98]. Predictability has been shown to be highest at the level of fitness, relative to underlying traits [99]. The suite of traits we measured is expected to by highly associated with fitness during temperature adaptation, and should thus exhibit high predictability in evolutionary changes. Some traits chosen here have also shown strong parallel responses in *Drosophila* [25]. Nevertheless, we note that our assays did not include juvenile competition traits, that may have been central in driving cold-adaptation [100], potentially explaining poor correlations between genomic- and fitness-offsets at 23°C.

Conclusions

Our work combines evolutionary change at the phenotypic and genomic level, something so far rarely done in thermal adaptation studies [31, 32, 82, 101]. We see that outcomes of both phenotypic change and genomic selection are often repeatable among populations of the same origin, but much less so between populations of differing origins. These patterns are true across all estimates of repeatability and predictability. This aspect of historical contingency places an emphasis on sampling many populations to more confidently describe hypotheses regarding broad patterns of evolutionary repeatability [102].

We found temperature-specific effects on repeatability at both the phenotypic and genomic level. Trajectories of heat-adapted populations tended

to show lower repeatability. Multiple non-mutually exclusive mechanisms may have contributed to this pattern, though evidence points to a role of temperature-specific epistasis. Importantly, theory predicts that many of the same mechanisms that should increase predictability at high temperatures (e.g., strong selection on protein homeostasis under severe heat stress) may contrarily result in reduced genomic predictability due to the biophysical properties of cells resulting in epistasis, and the demographic properties of populations resulting in genetic drift/draft. Our findings suggest that predicting adaptation to future warming climates using genomic data may become increasingly difficult, especially when extrapolating across geographically isolated populations. Although challenging, steps can be taken to enhance predictive power, for instance through methods that detect candidate loci by distinguishing the effects of linkage and drift from selection could provide a pathway to generating more relevant candidate sets for predictive analyses. For genomic predictions, particularly those using offsets, there is a scarcity of experimental evolution studies to validate and refine predictive methods, especially with regards to the role of epistasis in dictating the relevance of generalizing genomic predictions across larger geographic scales. Such studies will be invaluable for advancing predictive accuracy of forecasts of threatened species facing warming climates.

Acknowledgements

This work was funded by grants 2019-05024 from the Swedish Research Council (VR) and 2022-01117 from Formas to DB, and Swedish Research Council 2017-04963 and 2022-03427 to RS. The storage and computations for this work were enabled by resources in projects SNIC 2021/5-125 and SNIC 2020/6-175 provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. Special thanks to Kalle Tunström and Rachel Steward for figure critiques/suggestions, and Kevin Roberts and Loke von Schmalensee for insight into the physiological limits of thermal adaptation.

Literature Cited

1. Lässig, M., Mustonen, V. & Walczak, A. M. Predicting evolution. *Nature Ecology & Evolution* **1,** 0077 (2017).
2. Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science* **362,** eaam5979 (2018).
3. Wortel, M. T. *et al.* Towards evolutionary predictions: Current promises and challenges. *Evolutionary applications* **16,** 3–21 (2023).
4. Storz, J. F. Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics* **17,** 239 (2016).
5. Svensson, E. I. & Berger, D. The role of mutation bias in adaptive evolution. *Trends in ecology & evolution* **34,** 422–434 (2019).
6. Orr, H. A. The probability of parallel evolution. *Evolution* **59,** 216–220 (2005).
7. Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D. & Marx, C. J. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332,** 1190–1192 (2011).
8. Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. & Cooper, T. F. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332,** 1193–1196 (2011).
9. Tenaillon, O. *et al.* The molecular diversity of adaptive convergence. *Science* **335,** 457–461 (2012).
10. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344,** 1519–1522 (2014).
11. Tenaillon, O. *et al.* Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536,** 165–170 (2016).
12. Ono, J., Gerstein, A. C. & Otto, S. P. Widespread genetic incompatibilities between first-step mutations during parallel adaptation of Saccharomyces cerevisiae to a common environment. *PLoS biology* **15,** e1002591 (2017).

13. Barghi, N. *et al.* Genetic redundancy fuels polygenic adaptation in Drosophila. *PLoS biology* **17,** e3000128 (2019).

14. Whiting, J. R. *et al.* The genetic architecture of repeated local adaptation to climate in distantly related plants. *Nature Ecology & Evolution,* 1–15 (2024).

15. Hoffmann, A. A. & Sgro, C. M. Climate change and evolutionary adaptation. *Nature* **470,** 479–485 (2011).

16. Hoffmann, A. A., Sgrò, C. M. & Kristensen, T. N. Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution* **32,** 506–517 (2017).

17. Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M. & Keller, S. R. Genomic prediction of (mal) adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics* **51,** 245–269 (2020).

18. Waldvogel, A.-M. *et al.* Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters* **4,** 4–18 (2020).

19. Ghildiyal, K. *et al.* Genomic insights into the conservation of wild and animal diversity: a review. *Gene,* 147719 (2023).

20. Pearman, P. B. *et al.* Monitoring of species' genetic diversity in Europe varies greatly and overlooks potential climate change impacts. *Nature ecology & evolution* **8,** 267–281 (2024).

21. Hoffmann, A. A., Weeks, A. R. & Sgrò, C. M. Opportunities and challenges in assessing climate change vulnerability through genomics. *Cell* **184,** 1420–1425 (2021).

22. Urban, M. C. *et al.* When and how can we predict adaptive responses to climate change? *Evolution Letters* **8,** 172–187 (2024).

23. Conte, G. L., Arnegard, M. E., Peichel, C. L. & Schluter, D. The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences* **279,** 5039–5047 (2012).

24. Burny, C., Nolte, V., Dolezal, M. & Schlötterer, C. Highly parallel genomic selection response in replicated Drosophila melanogaster populations with reduced genetic variation. *Genome Biology and Evolution* **13,** evab239 (2021).

25. Schlötterer, C. How predictable is adaptation from standing genetic variation? Experimental evolution in Drosophila highlights the central role of redundancy and linkage disequilibrium. *Philosophical Transactions of the Royal Society B* **378,** 20220046 (2023).

26. Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *science* **312,** 111–114 (2006).

27. De Visser, J. A. G., Cooper, T. F. & Elena, S. F. The causes of epistasis. *Proceedings of the Royal Society B: Biological Sciences* **278,** 3617–3624 (2011).

28. Bohutínská, M. *et al.* Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences* **118,** e2022713118 (2021).

29. Agrawal, A. F. & Whitlock, M. C. Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? *Trends in ecology & evolution* **25,** 450–458 (2010).

30. Caruso, C. M. *et al.* What are the environmental determinants of phenotypic selection? A meta-analysis of experimental studies. *The American Naturalist* **190,** 363–376 (2017).

31. Hutter, S., Saminadin-Peter, S. S., Stephan, W. & Parsch, J. Gene expression variation in African and European populations of Drosophila melanogaster. *Genome biology* **9,** 1–15 (2008).

32. Mallard, F., Nolte, V., Tobler, R., Kapun, M. & Schlötterer, C. A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in Drosophila. *Genome biology* **19,** 1–15 (2018).

33. Hochachka, P. W. & Somero, G. N. *Biochemical adaptation: mechanism and process in physiological evolution* (Oxford university press, 2002).

34. Angilletta, M. J. *Thermal adaptation: a theoretical and empirical synthesis* (Oxford University Press, 2009).

35. Levesque, D. L. & Marshall, K. E. Do endotherms have thermal performance curves? *Journal of Experimental Biology* **224,** jeb141309 (2021).

36. Huey, R. B. & Kingsolver, J. G. Evolution of thermal sensitivity of ectotherm performance. *Trends in ecology & evolution* **4,** 131–135 (1989).

37. Drake, J. W. Avoiding dangerous missense: thermophiles display especially low mutation rates. *PLoS genetics* **5,** e1000520 (2009).

38. Agozzino, L. & Dill, K. A. Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proceedings of the National Academy of Sciences* **115,** 9092–9097 (2018).

39. Berger, D., Stångberg, J., Baur, J. & Walters, R. J. Elevated temperature increases genome-wide selection on de novo mutations. *Proceedings of the Royal Society B* **288,** 20203094 (2021).

40. Tuda, M., Kagoshima, K., Toquenaga, Y. & Arnqvist, G. Global genetic differentiation in a cosmopolitan pest of stored beans: effects of geography, host-plant usage and anthropogenic factors. *PLoS One* **9,** e106268 (2014).

41. Immonen, E., Berger, D., Sayadi, A., Liljestrand-Rönn, J. & Arnqvist, G. An experimental test of temperature-dependent selection on mitochondrial haplotypes in Callosobruchus maculatus seed beetles. *Ecology and evolution* **10,** 11387–11398 (2020).

42. Burc, E. *et al.* Life-history adaptation under climate warming magnifies the agricultural footprint of a cosmopolitan insect pest. *bioRxiv,* 2024–03 (2024).

43. Dowling, D. K., Abiega, K. C. & Arnqvist, G. Temperature-specific outcomes of cytoplasmic-nuclear interactions on egg-to-adult development time in seed beetles. *Evolution* **61,** 194–201 (2007).

44. Dowling, D. K., Friberg, U. & Arnqvist, G. A comparison of nuclear and cytoplasmic genetic effects on sperm competitiveness and female remating in a seed beetle. *Journal of Evolutionary Biology* **20,** 2113–2125 (2007).

45. Bolnick, D. I., Barrett, R. D., Oke, K. B., Rennison, D. J. & Stuart, Y. E. (Non) parallel evolution. *Annual Review of Ecology, Evolution, and Systematics* **49,** 303–330 (2018).

46. Watanabe, J. Detecting (non) parallel evolution in multidimensional spaces: angles, correlations and eigenanalysis. *Biology Letters* **18,** 20210638 (2022).

47. Hansen, T. F. & Houle, D. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of evolutionary biology* **21,** 1201–1219 (2008).

48. Berger, D., Postma, E., Blanckenhorn, W. U. & Walters, R. J. Quantitative genetic divergence and standing genetic (co) variance in thermal reaction norms along latitude. *Evolution* **67,** 2385–2399 (2013).

49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25,** 1754–1760 (2009).

50. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20,** 1297–1303 (2010).

51. Pockrandt, C., Alzamel, M., Iliopoulos, C. S. & Reinert, K. GenMap: ultra-fast computation of genome mappability. *Bioinformatics* **36,** 3687–3692 (2020).

52. Jónás, Á., Taus, T., Kosiol, C., Schlötterer, C. & Futschik, A. Estimating the effective population size from temporal allele frequency changes in experimental evolution. *Genetics* **204,** 723–735 (2016).

53. Ewens, W. J. *Mathematical Population Genetics I: Theoretical Introduction* (Springer Science & Business Media, 2012).

54. Gompert, Z. & Messina, F. J. Genomic evidence that resource-based trade-offs limit host-range expansion in a seed beetle. *Evolution* **70,** 1249–1264 (2016).

55. Iranmehr, A., Akbari, A., Schlötterer, C. & Bafna, V. CLEAR: Composition of likelihoods for evolve and resequence experiments. *Genetics* **206,** 1011–1023 (2017).

56. Kofler, R. *et al.* PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PloS one* **6,** e15925 (2011).

57. Gain, C. *et al.* A quantitative theory for genomic offset statistics. *Molecular Biology and Evolution* **40,** msad140 (2023).

58. Lotterhos, K. E. Interpretation issues with "genomic vulnerability" arise from conceptual issues in local adaptation and maladaptation. *Evolution Letters* **8,** 331–339 (2024).

59. Biron, D. *et al.* A diacylglycerol kinase modulates long-term thermotactic behavioral plasticity in C. elegans. *Nature neuroscience* **9,** 1499–1505 (2006).

60. Krill, J. L. & Dawson-Scully, K. cGMP-dependent protein kinase inhibition extends the upper temperature limit of stimulus-evoked calcium responses in motoneuronal boutons of Drosophila melanogaster larvae. *Plos one* **11,** e0164114 (2016).

61. Schmidt, P. S. *et al.* An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* **105,** 16207–16211 (2008).

62. Zhang, Q. & Denlinger, D. L. Elevated couch potato transcripts associated with adult diapause in the mosquito Culex pipiens. *Journal of insect physiology* **57,** 620–627 (2011).

63. Cuevas, B., Abell, A. & Johnson, G. Role of mitogen-activated protein kinase kinase kinases in signal integration. *Oncogene* **26,** 3159–3171 (2007).

64. Sakamoto, F. *et al.* Detection of evolutionary conserved and accelerated genomic regions related to adaptation to thermal niches in Anolis lizards. *Ecology and Evolution* **14,** e11117 (2024).

65. Gerken, A. R., Eller, O. C., Hahn, D. A. & Morgan, T. J. Constraints, independence, and evolution of thermal plasticity: probing genetic architecture of long-and short-term thermal acclimation. *Proceedings of the National Academy of Sciences* **112,** 4399–4404 (2015).

66. Fang, Y., Gong, H., Yang, R., Lai, K. W. & Quan, M. An active biomechanical model of cell adhesion actuated by intracellular tensioning-taxis. *Biophysical journal* **118,** 2656–2669 (2020).

67. Das, S. G. & Krug, J. Unpredictable repeatability in molecular evolution. *Proceedings of the National Academy of Sciences* **119,** e2209373119 (2022).

68. Diaz-Colunga, J., Sanchez, A. & Ogbunugafor, C. B. Environmental modulation of global epistasis in a drug resistance fitness landscape. *Nature Communications* **14,** 8055 (2023).

69. Echave, J. & Wilke, C. O. Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. *Annual review of biophysics* **46,** 85–103 (2017).

70. Chen, P. & Shakhnovich, E. I. Thermal adaptation of viruses and bacteria. *Biophysical journal* **98,** 1109–1118 (2010).

71. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences* **103,** 5869–5874 (2006).

72. Goldstein, R. A. The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins: Structure, Function, and Bioinformatics* **79,** 1396–1407 (2011).

73. Parmesan, C. Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.* **37,** 637–669 (2006).

74. Urban, M. C. Accelerating extinction risk from climate change. *Science* **348,** 571–573 (2015).

75. Radchuk, V. *et al.* Adaptive responses of animals to climate change are most likely insufficient. *Nature communications* **10,** 3109 (2019).

76. Rellstab, C., Dauphin, B. & Exposito-Alonso, M. Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications* **14,** 1202–1212 (2021).

77. Barton, N. H. The "new synthesis". *Proceedings of the National Academy of Sciences* **119,** e2122147119 (2022).

78. Echave, J., Spielman, S. J. & Wilke, C. O. Causes of evolutionary rate variation among protein sites. *Nature Reviews Genetics* **17,** 109–121 (2016).

79. Bloom, J. D. *et al.* Thermodynamic prediction of protein neutrality. *Proceedings of the National Academy of Sciences* **102,** 606–611 (2005).

80. Hayward, L. K. & Sella, G. Polygenic adaptation after a sudden change in environment. *Elife* **11,** e66697 (2022).

81. Tobler, R. *et al.* Massive habitat-specific genomic response in D. melanogaster populations during experimental evolution in hot and cold environments. *Molecular biology and evolution* **31,** 364–375 (2014).

82. Otte, K. A., Nolte, V., Mallard, F. & Schlötterer, C. The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime. *Genome Biology* **22,** 1–25 (2021).

83. Hoffmann, A. A. Physiological climatic limits in Drosophila: patterns and implications. *Journal of Experimental Biology* **213,** 870–880 (2010).

84. Christodoulaki, E., Barghi, N. & Schlötterer, C. Distance to trait optimum is a crucial factor determining the genomic signature of polygenic adaptation. *BioRxiv,* 721340 (2019).

85. Phillips, M. A., Briar, R. K., Scaffo, M., Zhou, S. & Burke, M. K. Strength of selection potentiates distinct adaptive responses in an evolution experiment with outcrossing yeast. *BioRxiv,* 2022–05 (2022).

86. Mäkinen, H., Papakostas, S., Vøllestad, L. A., Leder, E. H. & Primmer, C. R. Plastic and evolutionary gene expression responses are correlated in European grayling (Thymallus thymallus) subpopulations adapted to different thermal environments. *Journal of Heredity* **107,** 82–89 (2016).

87. Sørensen, J. G. *et al.* Pronounced plastic and evolutionary responses to unpredictable thermal fluctuations in Drosophila simulans. *Frontiers in genetics* **11,** 555843 (2020).

88. Gallegos, C., Hodgins, K. A. & Monro, K. Climate adaptation and vulnerability of foundation species in a global change hotspot. *Molecular Ecology* **32,** 1990–2004 (2023).

89. Van Boheemen, L. A. & Hodgins, K. A. Rapid repeatable phenotypic and genomic adaptation following multiple introductions. *Molecular Ecology* **29,** 4102–4117 (2020).

90. Renaut, S., Owens, G. L. & Rieseberg, L. H. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology* **23,** 311–324 (2014).

91. Chaturvedi, S. *et al.* Climatic similarity and genomic background shape the extent of parallel adaptation in Timema stick insects. *Nature ecology & evolution* **6,** 1952–1964 (2022).

92. Feder, M. E., Bennett, A. F. & Huey, R. B. Evolutionary physiology. *Annual review of ecology and systematics* **31,** 315–341 (2000).

93. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: sign epistasis and genetic costraint on evolutionary trajectories. *Evolution* **59,** 1165–1174 (2005).

94. Bank, C. Epistasis and adaptation on fitness landscapes. *Annual Review of Ecology, Evolution, and Systematics* **53,** 457–479 (2022).

95. Li, C. & Zhang, J. Multi-environment fitness landscapes of a tRNA gene. *Nature ecology & evolution* **2,** 1025–1032 (2018).

96. Johnson, M. S., Reddy, G. & Desai, M. M. Epistasis and evolution: recent advances and an outlook for prediction. *BMC biology* **21,** 120 (2023).

97. Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y. & Keller, S. R. Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources* **21,** 2749–2765 (2021).

98. De Lisle, S. P. & Bolnick, D. I. A multivariate view of parallel evolution. *Evolution* **74,** 1466–1481 (2020).

99. Bailey, S. F., Rodrigue, N. & Kassen, R. The effect of selection environment on the probability of parallel evolution. *Molecular biology and evolution* **32,** 1436–1448 (2015).

100. Guedes, R. N., Guedes, N. M. & Smith, R. H. Larval competition within seeds: from the behaviour process to the ecological outcome in the seed beetle Callosobruchus maculatus. *Austral Ecology* **32,** 697–707 (2007).

101. Santos, M. A. *et al.* Experimental Evolution in a Warming World: The Omics Era. *Molecular Biology and Evolution* **41,** msae148 (2024).

102. Czech, L. *et al.* Monitoring rapid evolution of plant populations at scale with Pool-Sequencing. *BioRxiv,* 2022–02 (2022).

Supplemental Material, Figures, and Tables

	Ne	# Putative Sites	# Putative Genes	$\overline{\lvert\Delta AF\rvert}$	$\frac{\overline{\lvert\Delta AF\rvert}}{gen}$	$\overline{\lvert S\rvert}$
Yemen1 (23°C)	244.29	61,458	2,405	0.546	0.009	0.0515
Yemen2 (23°C)	291.21	56,378	2,170	0.511	0.008	0.0490
Cali3 (23°C)	291.33	86,561	2,805	0.484	0.008	0.0440
Cali4 (23°C)	316.67	60,384	2,481	0.501	0.008	0.0374
Brazil5 (23°C)	194.23	34,216	1,490	0.607	0.010	0.0495
Brazil6 (23°C)	215.69	41,629	1,784	0.592	0.010	0.0448
Yemen7 (35°C)	178.38	32,804	1,343	0.663	0.010	0.0540
Yemen8 (35°C)	106.44	21,716	774	0.826	0.014	0.0660
Cali9 (35°C)	189.95	42,695	1,717	0.618	0.010	0.0552
Cali10 (35°C)	193.84	43,445	1,731	0.610	0.010	0.0480
Brazil11 (35°C)	161.86	28,758	1,217	0.658	0.010	0.0602
Brazil12 (35°C)	184.51	21,551	1,146	0.706	0.011	0.0596

Table S1: Estimated Ne, number of putatively selected SNPs, number of putatively selected genes (genes with SNPs falling within protein coding regions), and the mean magnitude of allele frequency change, allele frequency change per generation, and selection coefficient per population. The magnitudes of allele frequency change and selection coefficient estimates are comprised of the putative SNPs identified for each respective line.

	r
Yemen1 (23°c)	0.47
Yemen2 (23°c)	0.46
Cali.3 (23°c)	0.47
Cali.4 (23°c)	0.49
Brazil5 (23°c)	0.54
Brazil6 (23°c)	0.54
Yemen7 (35°c)	0.47
Yemen8 (35°c)	0.59
Cali.9 (35°c)	0.48
Cali.10 (35°c)	0.50
Brazil11 (35°c)	0.53
Brazil12 (35°c)	0.52

Table S2: Estimated Pearson's correlation (r), between allele frequency change and estimated selection coefficient. All p-values are significant ($p < 0.05$).

Phenotypic

Angle 23°C	Brazil	California	Yemen
Within	NA*	47.04	28.27
Between	66.94	67.86	72.28
Angle 35°C	Brazil	California	Yemen
Within	15.63	8.46	36.43
Between	25.07	35.10	53.63
Divergence 23°C	Brazil	California	Yemen
Within	0.00*	0.00*	0.00*
Between	0.02	-0.26	-0.28
Divergence 35°C	Brazil	California	Yemen
Within	0.24	0.00*	0.69
Between	0.16	0.21	0.21

Table S3: Mean phenotypic angles (θ) and divergence ($E_d - S_d$) of within- and between-origin comparisons for each selection regime. Within-origin comparisons are represented by only a single value between the two replicates per origin. Asterisks represent estimates in which estimates could not be properly estimated due to high measurement error relative to vector length.

Genomic			
Angle 23°C	Brazil	California	Yemen
Within	65.77	80.10	73.00
Between	89.14	87.96	84.5
Angle 35°C	Brazil	California	Yemen
Within	71.10	73.98	71.75
Between	92.70	91.39	86.80
Divergence 23°C	Brazil	California	Yemen
Within	0.226	0.190	0.258
Between	-0.149	-0.122	-0.122
Divergence 35°C	Brazil	California	Yemen
Within	0.338	0.242	0.365
Between	0.014	-0.049	0.008

Table S4: Mean genomic angles (θ) and divergence ($E_d - S_d$) of within- and between-origin comparisons for each selection regime. Within-origin comparisons are represented by only a single value between the two replicates per origin.

GO.ID 23°C	Term	Ann.	Sig.	Exp.	Adj. p-value
GO:0035556	intracellular signal...	195	13	6.73	0.0035
GO:0007156	homophilic cell adhe...	22	4	0.76	0.0062
GO:0035023	regulation of Rho pr...	29	4	1.00	0.0166
GO:0007205	protein kinase C-act...	8	2	0.28	0.0289
GO:0032940	secretion by cell	12	2	0.41	0.0344
GO:0007165	signal transduction	473	32	16.34	0.0410
GO:0007186	G protein-coupled re...	148	11	5.11	0.0487
GO:0040007	growth	20	8	0.69	1.5e-07
GO.ID 35°C	Term	Ann.	Sig.	Exp.	Adj. p-value
GO:0007156	homophilic cell adhe...	22	2	0.12	0.006
GO:0007018	microtubule-based mo...	46	2	0.25	0.025
GO:0019319	hexose biosynthetic ...	5	1	0.03	0.027
GO:0006013	mannose metabolic pr...	5	1	0.03	0.027

Table S5: Gene set enrichment analysis for genic targets of selection for targets shared across all 6 cold-adapted replicates (top) and all 6 hot-adapted replicates (bottom).

Shared Genomic			
Angle 23°C	Yemen	California	Brazil
Within	74.56	75.29	71.67
Between	81.87	83.45	84.03
Angle 35°C	Yemen	California	Brazil
Within	70.52	75.07	69.19
Between	84.69	89.15	90.49
Divergence 23°C	Yemen	California	Brazil
Within	45.53	34.07	36.98
Between	3.9	8.25	7.6
Divergence 35°C	Yemen	California	Brazil
Within	60.36	44.34	57.08
Between	22.39	17.12	18.63

Table S6: Mean genomic angles (θ) and divergence ($E_d - S_d$) of within- and between-origin comparisons for each selection regime using selected sites which were variable in the ancestors of all populations. Within-origin comparisons are represented by only a single value between the two replicates per origin.

23°C	
Within θ	Between θ
71.69	83.66
35°C	
Within θ	Between θ
71.58	86.05

Table S7: Mean genomic angles (θ) of within- and between-origin comparisons for each selection regime using the top 100 selected SNPs per line located at least 10kb away from each other which were variable in the ancestors of all populations (total n=756).

	B1C	B2C	C3C	C4C	Y5C	Y6C
B1C						
B2C	0.37					
C3C	0.25	0.30				
C4C	0.25	0.26	0.46			
Y5C	0.27	0.30	0.37	0.33		
Y6C	0.27	0.29	0.36	0.32	0.39	

Table S8: Pairwise Jaccard indices for gene set similarity between cold-adapted populations. Gene sets are comprised of genes in which significantly selected SNPs (i.e. $p < 0.0001$) are located.

	B7H	B8H	C9H	C10H	Y11H	Y12H
B7H						
B8H	0.26					
C9H	0.20	0.20				
C10H	0.19	0.19	0.34			
Y11H	0.20	0.21	0.23	0.22		
Y12H	0.12	0.15	0.13	0.13	0.19	

Table S9: Pairwise Jaccard indices for gene set similarity between hot-adapted populations. Gene sets are comprised of genes in which significantly selected SNPs (i.e. $p < 0.0001$) are located.

Genic Jaccard Indices

23°C	Yemen	California	Brazil
Within	0.41	0.46	0.35
Between	0.32(±0.04)	0.31(±0.04)	0.28(±0.02)
35°C	**Yemen**	**California**	**Brazil**
Within	0.22	0.35	0.25
Between	0.18(±0.03)	0.20(±0.03)	0.19(±0.03)

Table S10: Mean Jaccard indices (±SD) of within- and between-origin gene set comparisons for each selection regime. Within-origin comparisons are represented by only a single value between the two replicates per origin.

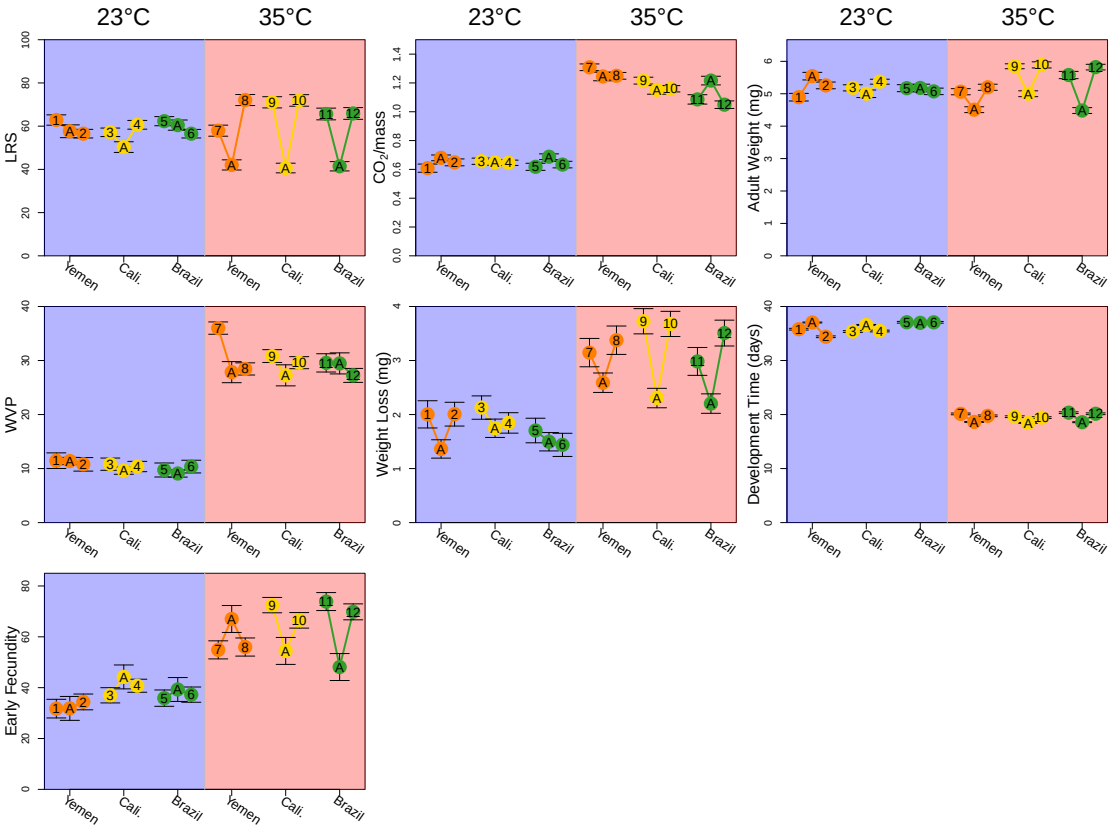

Figure S1: Mean trait measurements for each evolved line. Note that ancestors and evolved replicates (1–12) were measured in their evolved temperature regime, denoted by background color. The ancestral populations ("A") are shared between selection regimes for each respective geographic background. Means are shown as points with bars as standard error.

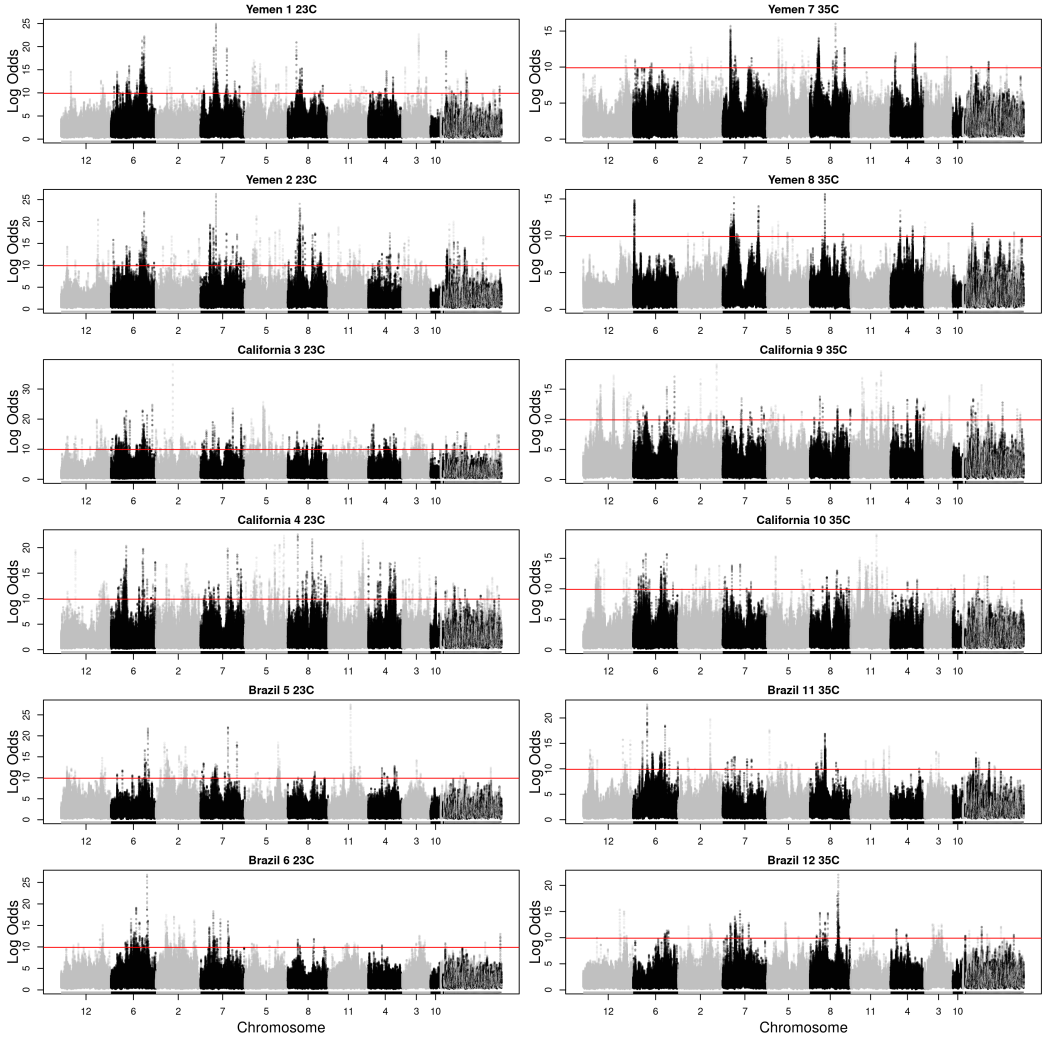

Figure S2: Manhattan plots of rolling average log odds p-values (window size = 20 SNPs) for each population across all selected sites using `R` package `RcppRoll` (v.0.3.0). The largest 10 scaffolds (i.e. chromosomes) are named. The significance threshold of log odds ($log(\frac{0.99995}{0.00005})$) is given by the horizontal red line.

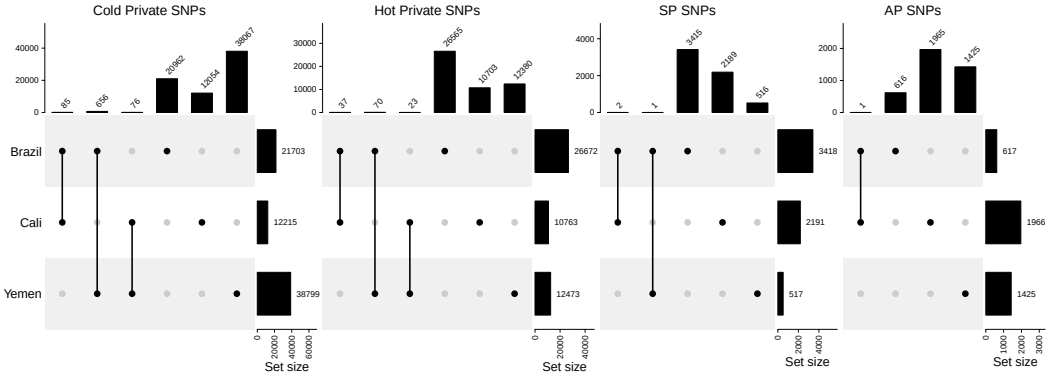

Figure S3: Upset plots of different categories of SNP sets based on allele frequency trajectories of SNPs within and between temperature regimes.

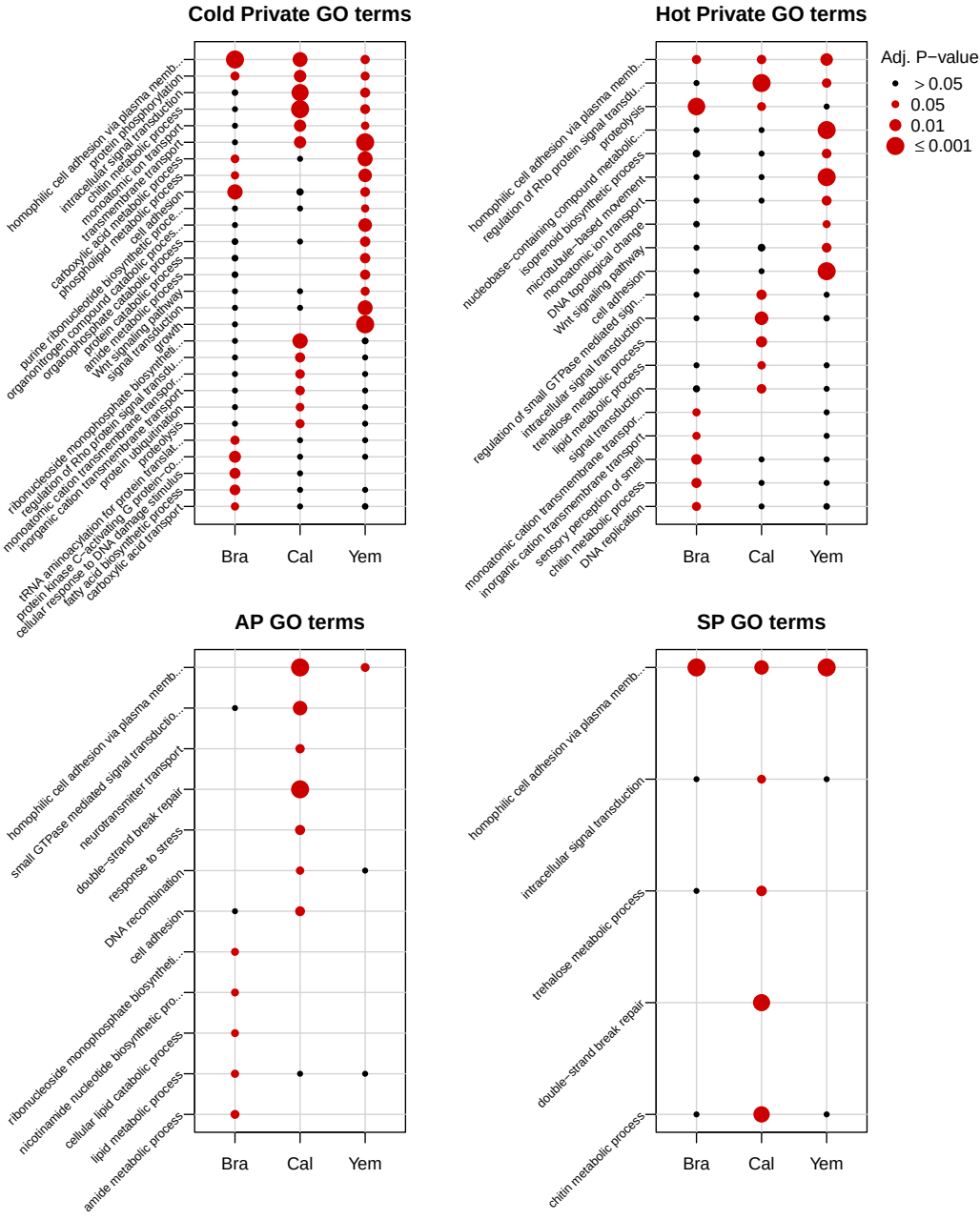

Figure S4: Dot plot indicating significant GO terms per gene set category (based on Fig. 4). Adjusted p-values equal to 1 are not shown.

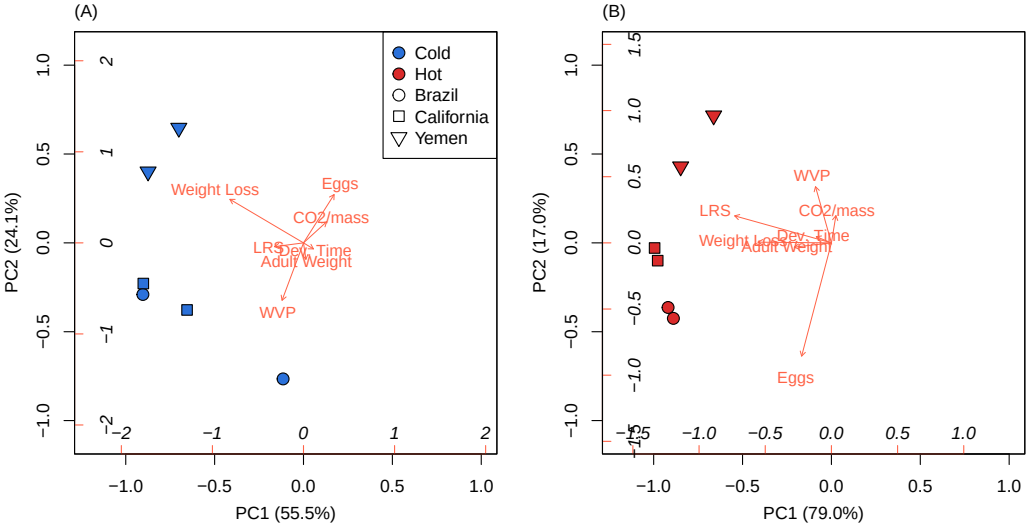

Figure S5: PCA biplot of evolutionary change trajectories showing scores (points) and coefficients (arrows) for (A) cold-adapted populations and (B) hot-adapted populations. The inner-axis represents coefficients, while outer-axis represents scores.

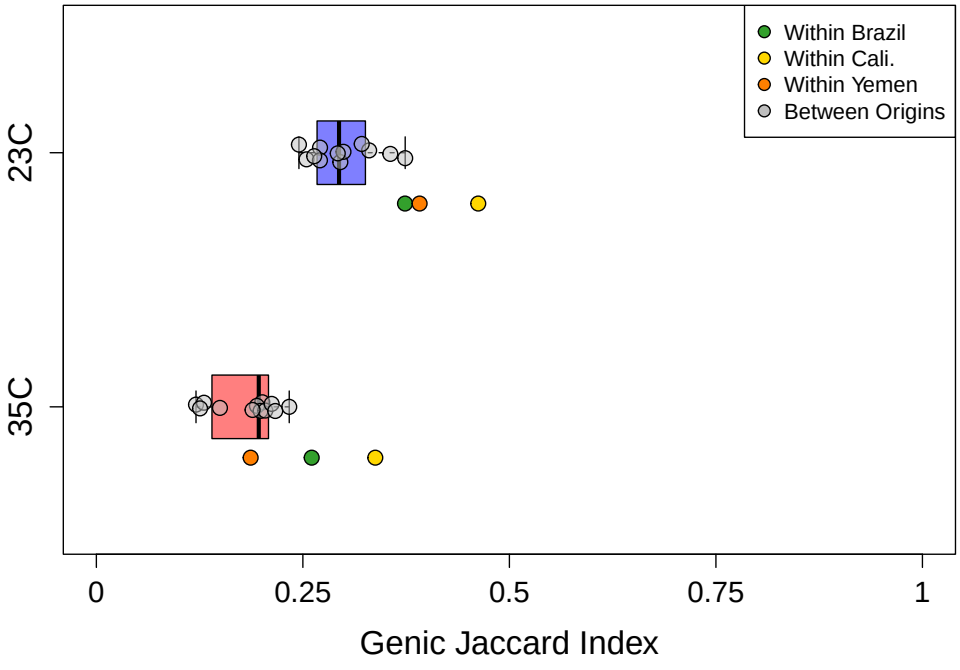

Figure S6: Distribution of Jaccard indices for genic targets of selections between lines within temperature regimes. Distributions of angles and divergences are separated by between-origin (grey points) within-origin (colored points) pairwise comparisons. A Jaccard index of 0 indicates no overlap in genic targets, while an index of 1 indicates complete overlap in genic targets.

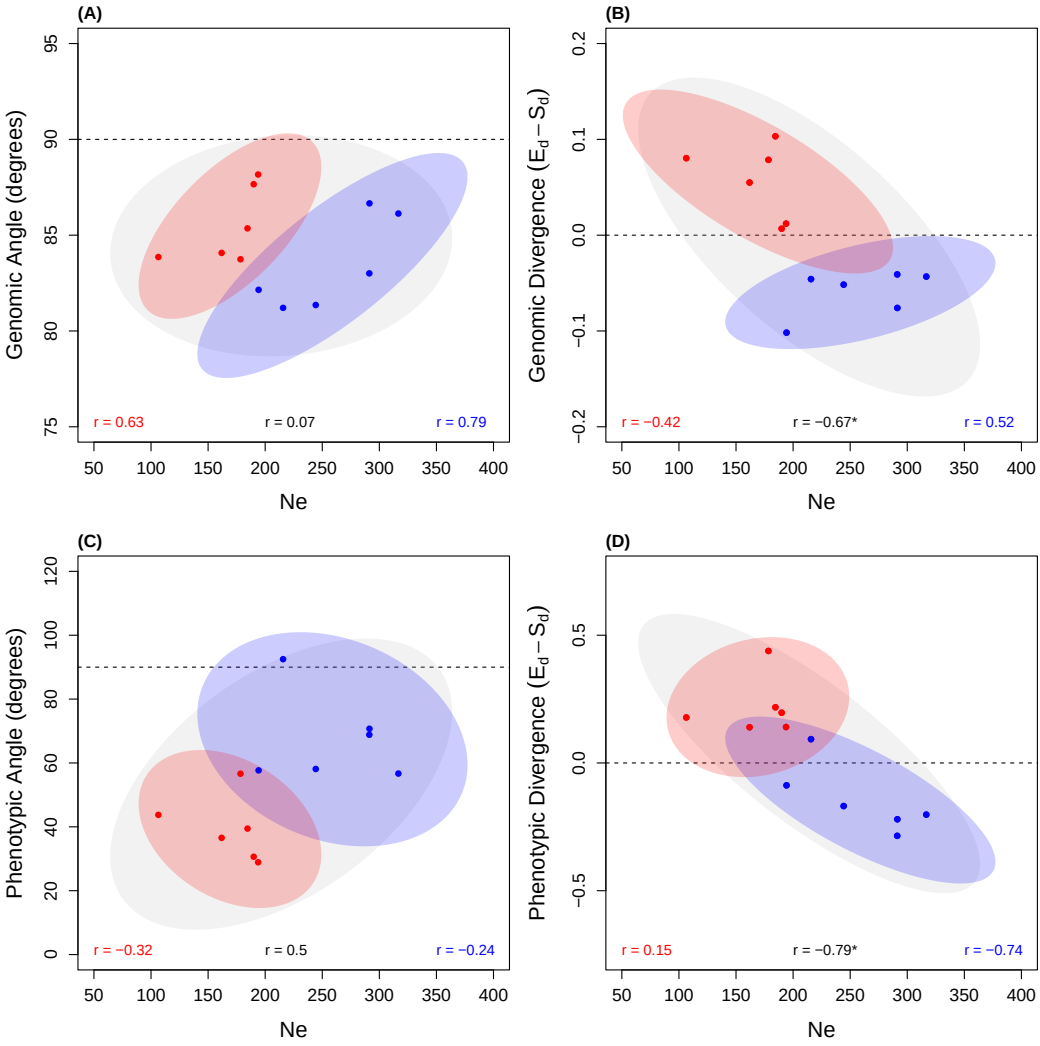

Figure S7: Correspondance between mean measures of phenotypic (A–B) and genomic (C–D) repeatability and estimates of effective population size (Ne) per line. In each plot, points are colored by regime. Confidence ellipses (95%) and correlations are given both per regime (red, blue) and for all lines (grey). Significant correlations are designated with an asterisk.

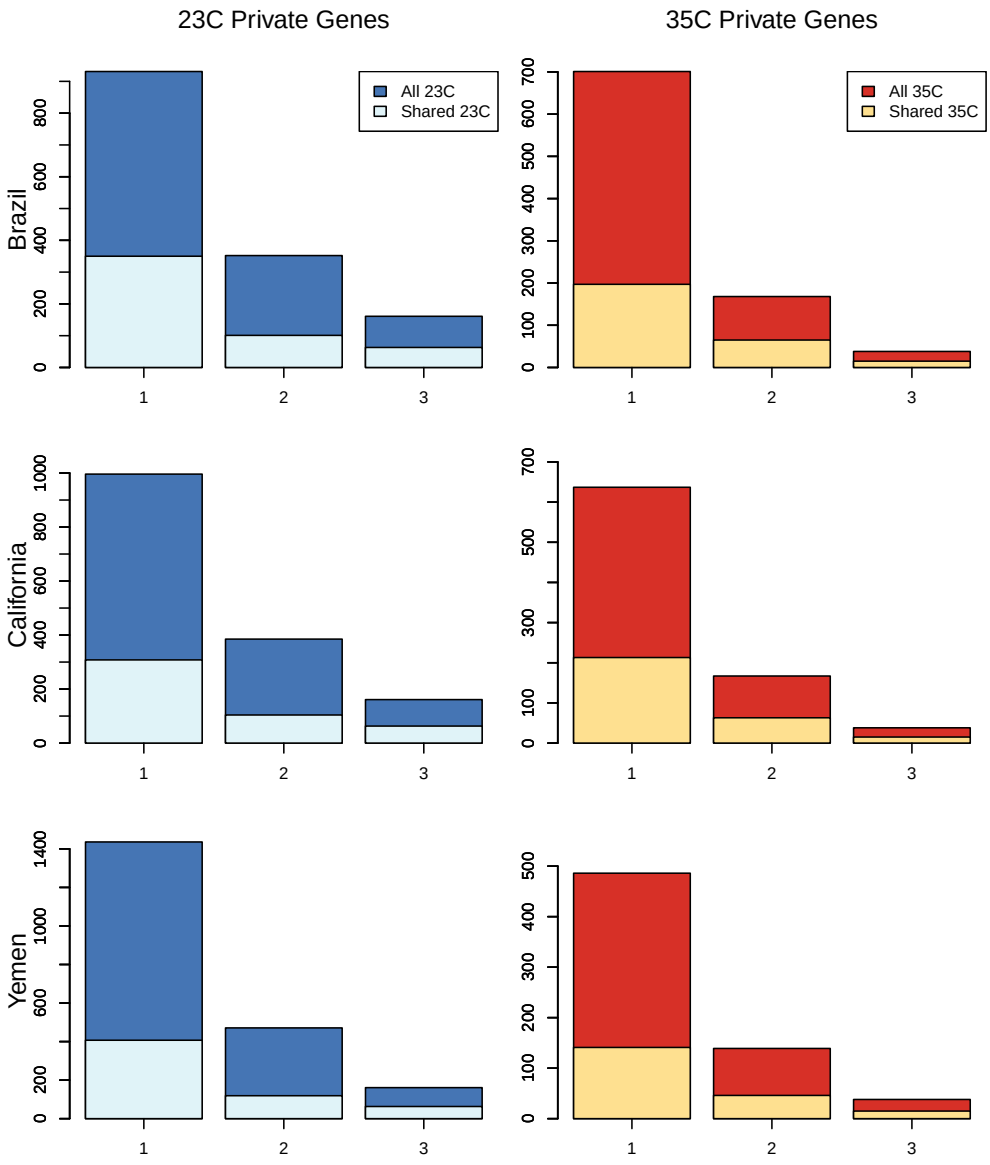

Figure S8: Distributions of the number of privately selected genes for each thermal regime shared between populations, ranging from gene mutations which only appear in a single population (1) to those that appear across all three backgrounds (3). The distributions for both the entire set of selected sites (n=475,194) and sites which are polymorphic across all ancestral backgrounds (n=119,630) are shown in different shades. Chi-square tests show no significant differences between the distributions of all sites and shared sites for any comparison ($p > 0.05$).

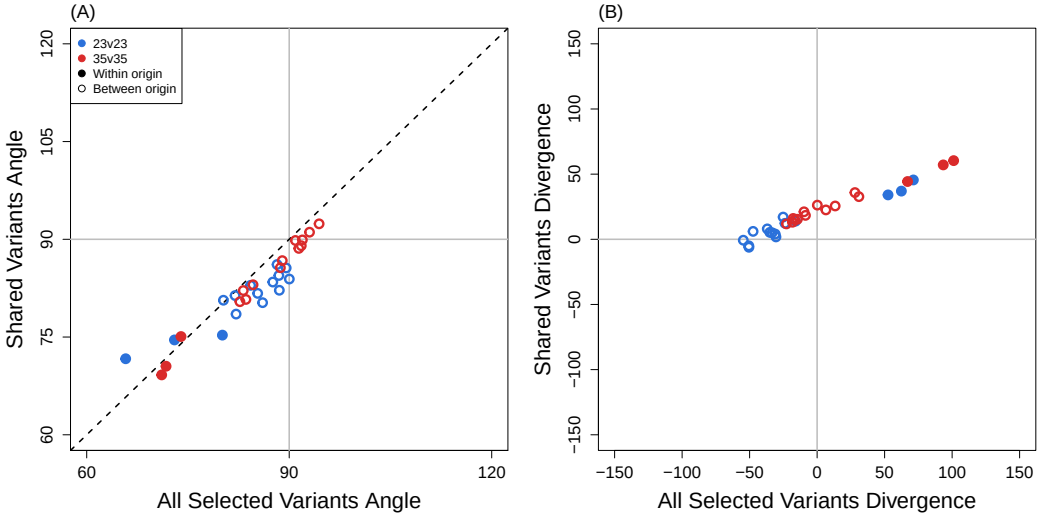

Figure S9: Comparison of repeatability measures between the set of SNPs selected in any one population versus the set of SNPs selected in any one population, but also segregating in all populations. The black dashed line represents the identity line.

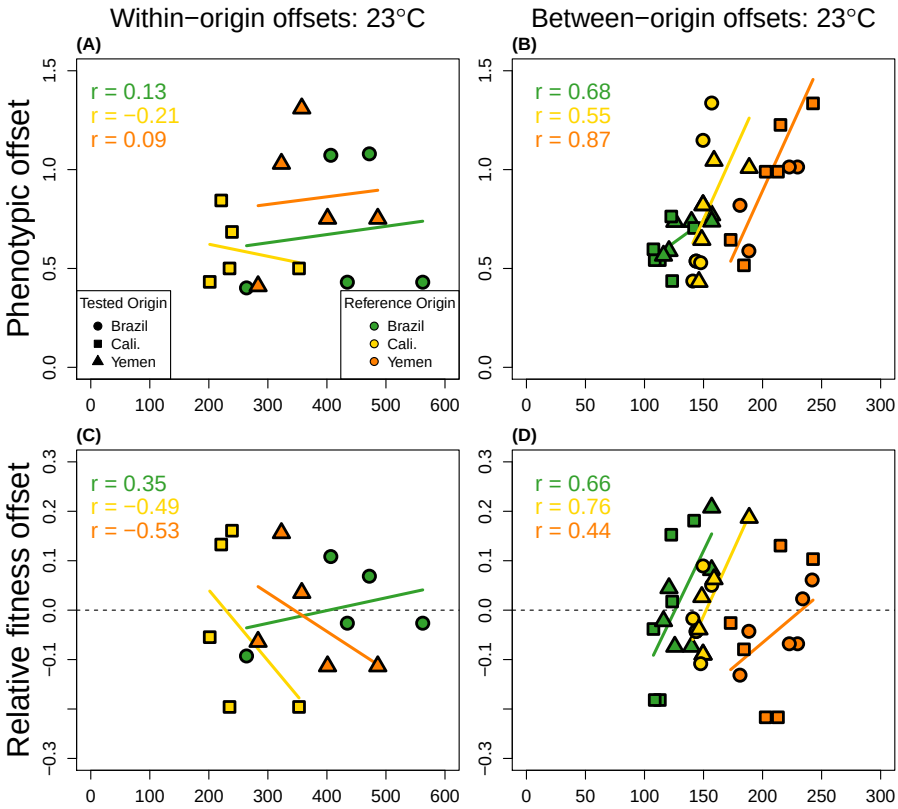

Figure S10: Genomic predictors of phenotypic and relative fitness offsets at high heat. Genomic offsets were calculated per reference origin based on SNPs whose p-values for allele frequency change fell within the top 0.001th quantile (n=10546–10887). Phenotypic offsets were calculated as the Euclidean distance in scaled trait-space to the reference population. Relative fitness offsets were calculated as the relative fitness ($\frac{LRS}{Dev.Time}$) to the reference population. The reference populations consist of a single replicate of the respective line evolved at 35°C. As all lines have been measured in common garden experiments at both 23°C and 35°C, all lines have been used to calculate offsets at each assay temperature. Offsets are organized by within-origin comparisons (A,C) and between-origin comparisons (B,D). Correlations and regression lines are colored by reference population.

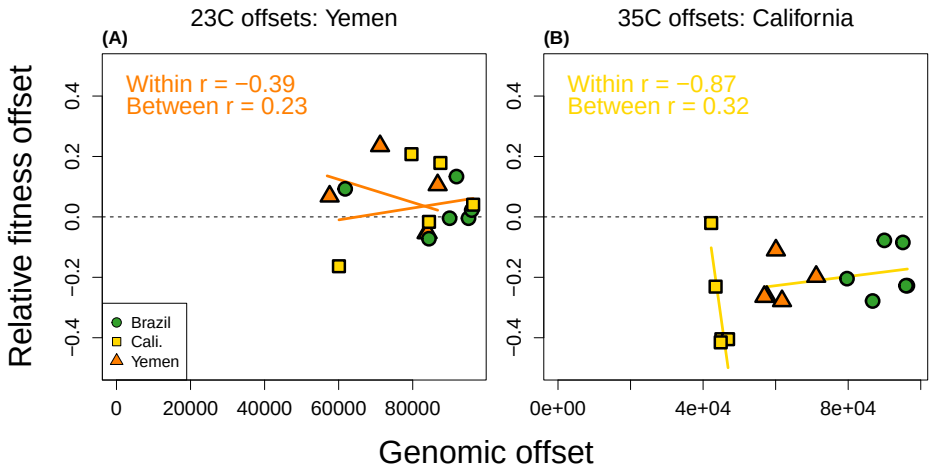

Figure S11: Genomic predictors of relative fitness offsets for populations showing the greatest performance per thermal regime. Genomic offsets were calculated per reference origin based on all putatively selected SNPs identified for each respective thermal regime (Yemen at 23°C, California at 35°C), but not scaled per-SNP by selection coefficients as in other genomic offsets. Relative fitness offsets were calculated as the relative fitness ($\frac{LRS}{Dev.Time}$) to the reference population. Correlations and regression lines are separated by comparisons of the same and other populations.

A hypothesis for how temperature extremes affect epistasis for fitness based on the thermodynamics of enzyme performance.

Biological rates of ectotherms show an empirically well-described, roughly exponential, increase with temperature that closely mirrors the thermodynamic properties of enzyme reactions [1, 2]. This pattern occurs because biological rates are governed at the molecular level by the catalytic reaction rate, k_{cat}. Based on transition state theory [3, 4]:

$$k_{cat} = k_0 e^{-\Delta G^a / RT} \tag{1}$$

where ΔG^a is the Gibbs free energy of activation required for the enzymatic reaction to occur (kcal mol^{-1}), R is the universal gas constant (0.002 kcal mol^{-1}), T is temperature measured in Kelvin, and $k_0 = \kappa k_B T/h$ where κ is a rate- and species-specific constant, k_B is the Boltzmann constant, and h is Planck's constant. ΔG^a is comprised by an enthalpy term (ΔH^a) and a temperature-dependent entropy term (ΔS^a):

$$\Delta G^a(T) = \Delta H^a - T\Delta S^a \tag{2}$$

Warm temperature increases the entropy term ($\Delta S^a T$), reducing ΔG^a. Equations S1 and S2 thus describe an exponential increase in reaction rate with temperature.

Strength of selection on catalytic rates and epistasis for fitness at cold temperatures

To form a hypothesis for how natural selection acts on catalytic rates at different temperatures, we incorporate ecological realism into predictions by assuming that the temperature-driven exponential increase in reaction rates will not have a 1:1 mapping with reproductive output, due to constraints on reproductive output set by other factors than temperature. At some point, suboptimal co-expression of correlated traits (evolutionary constraints), and/or limited nutrients or reproductive opportunities (ecological constraints) should make fitness pay-offs follow a pattern of diminishing returns with increases in catalytic rates [5–7] until temperature-dependent

physiological rates are no longer the limiting factor on reproductive output (Fig. S1A). We describe these dynamics by assuming that reproductive rates are a logistic function of temperature, where the initial exponential increase in reproductive rate is explained by temperature-driven increases in catalytic rates and a release from thermodynamic constraints as temperatures warm, but where diminishing returns result from ecological constraints at higher temperatures:

$$R_B = \frac{B_{max}}{1 + \frac{(B_{max} - B_{T_{ref}})}{B_{T_{ref}}} \times e^{\frac{-\Delta S(T - Tref)}{RT}}} \tag{3}$$

where B_{max} is the maximum reproductive rate of the organism, and $B_{T_{ref}}$ is reproductive output as a function of ΔG^a at the reference temperature, T_{ref}. We introduce a beneficial mutation that increases reproductive rate at the reference temperature (i.e., $B_{T_{ref}}$) by 10% through effects on ΔH^a ($\Delta\Delta H^a = -0.055$ kcal/mol). We then estimate selection on this mutation across temperatures by comparing wildtype reproductive rate to that of the mutant (Fig. S1A, B). To estimate epistasis and the dependency of natural selection on genetic background, we performed these calculations on three genetic backgrounds with wildtype $B_{T_{ref}}$ set to 40, 50 or 60 at $T_{ref} = 23°C$, with $B_{max} = 100$ in all cases. We further set ΔS to 0.50 kcal/mol K^{-1} based on values from the literature [8, 9].

This scenario generates strong positive selection on the introduced mutation across all three backgrounds at cold temperature ($s = 0.1$), while selection is weaker at intermediate benign temperature ($s \approx 0.005$) and at hot temperature the mutation is effectively neutral ($s \approx 0$) (Fig. S1B). This effect arises because cold temperatures act rate-limiting on reproductive output on all three backgrounds to an extent that far exceeds any rate-limiting ecological process. Once ecological constraints come into play at warmer temperatures, selection on the mutation weakens, but to similar extent across backgrounds, and little epistasis results.

Strength of selection on protein stability and epistasis for fitness at hot temperatures

Organisms that experience temperatures that exceed their thermal optimum suffer from increased cellular stress and many physiological processes fail. This results in an exponential increase in molecular failure rates and cellular damage [1, 10]. One major component of cellular damage and decreased fitness at hot temperature is attributed to a reduction in the proportion of functional enzyme due to reversable inactivation via protein unfolding [11–15]. While this loss of protein stability is reversible, the reduction of properly folded protein leaves fewer protein molecules ready for work [14], and can cause excessive cellular toxicity as misfolded proteins clog up the cellular environment [16]. While several processes contribute to the observed exponential increase in death rates with temperature [1], here we use the example of protein folding stability to illustrate how thermodynamics predict temperature-dependent selection due to the rich empirical data allowing parameterization of the calculations. The proportion of folded enzyme ready to catalyse reactions follows a Boltzmann probability as a function of the Gibbs free energy of folding, ΔG^f, which is thus a measure of protein stability [12]

$$P_{fold} = \prod_{1}^{i} = \frac{1}{1 + e^{-\frac{\Delta G^f_{T_{ref}} - \Delta Sf(T - T_{ref})}{RT}}} \tag{4}$$

where stability (ΔG^f) is given for each of a set of i proteins that act in sequence and contribute multiplicatively to a biological rate [8, 14]. At a benign temperature of 25°C, most natural proteins occur in functional (properly folded) state, and mean $\Delta G^f \approx 7$ kcal mol^{-1} [12, 17]. Warm temperatures increase entropy ($\Delta S^f(T - T_{ref})$) , reducing ΔG^f, which leads to a rapid non-linear reduction in the fraction of functional enzyme (Fig. S1C). We introduce a beneficial mutation that increases folding stability at the reference temperature (T_{ref}) and estimate selection on this mutation across temperatures by comparing the fraction of properly folded enzyme of the wildtype to that of the mutant (Fig. S1C). For the effect size of the mutation, we chose to increase folding stability by decreasing ΔH^f by 1 kcal mol^{-1} ($\Delta\Delta H^f = -1.0$ kcal/mol) based on empirical estimates [9, 17, 18]. We did this on three genetic backgrounds with the wildtype ΔH^f in the

enzyme receiving a mutation equal to 7, 9, or 11 kcal mol^{-1} at $T_{ref} = 23°C$, which are typical stabilities observed across ectotherms (Dill et al. 2011). For other parameters we assumed $\Delta S = 0.50$ kcal mol^{-1} K^{-1} and $i = 100$ proteins (e.g. [8, 17]).

In this scenario, the mutation acts as effectively neutral on all backgrounds at cold temperature ($s \approx 0$). At intermediate temperature, there is weak selection on the background with the most unstable wildtype configuration ($s \approx 0.01$) whereas the mutation is still effectively neutral on the other two backgrounds. At hot temperature the mutation is under strong selection on the backgrounds with low ($s > 0.1$) and intermediate ($s \approx 0.03$) wildtype stability, whereas the mutation is still effectively neutral ($s \approx 0$) on the background with most stable wildtype configuration (Fig. S1D), signifying substantial epistasis for fitness. This effect arises because proteins evolve marginal stability up until a point where selection is too weak to increase stability further [14, 19]. Around this zone of marginal stability, new mutations with conditional effects that are weak an effectively neutral at colder temperatures can accumulate, which affects subsequent selection on new secondary genetic variants.

Protein folding can also be compromised by acute cold stress [1, 20]. However, such cold temperatures are typically not experienced during the active seasons of tropical and sub-tropical insects [21, 22], as the case for *C. maculatus* [23], and were therefore not modelled. Moreover, stressful temperatures affect organisms in a multitude of ways [1, 2, 5]. Thus, our particular example using protein folding is merely meant to illustrate a general principle and was chosen based on the simple fact that the modelled relationships are empirically well-supported in qualitative sense (see above). Other physiological phenomena that scale with temperature, such as metabolic expenditure [24] and the production of harmful reactive oxygen species [25], are likely to simultaneously contribute to increasing molecular failure rates at stressful temperature. If such processes follow similar patterns with temperature (i.e. marginally robust at the thermal optimum and governed by the same thermodynamics principles), similar predictions for the strength of selection and epistasis for fitness would result.

Yet, we have only illustrated predictions based on general, and very simplified, principles founded in thermodynamics. Our predictions assume

multiplicative action between genes (or pathways) in effects on catalytic rates (Fig. S1A) and protein stability (Fig. S1C) that result in epistasis for fitness via non-linear (and temperature-dependent) mapping to fitness. We argue that these effects are general and may explain broad patterns in the average extent of epistasis for fitness across temperatures. However, at the more fine-grained scale, molecular epistasis is the result of intricate and interdependent molecular processes which are likely to cause epistatic interactions among genes governing interacting pathways, which are likely omnipresent at any temperature.

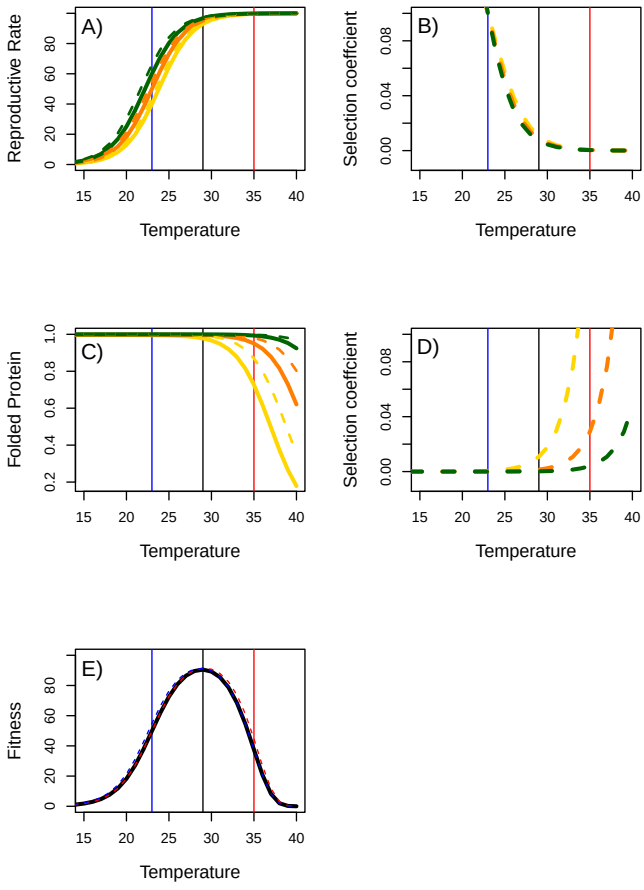

Figure S12: A hypothesis for temperature-dependent epistasis for fitness. A) Reproductive rate increases exponentially with temperature at the colder range due to relaxation of thermodynamics constraints on reaction rates, but starts to follow a pattern of diminishing returns at warmer temperatures due to ecological constraints. Illustrated for three genetic backgrounds (green, orange and yellow) with wildtypes (solid lines) and a mutant (broken lines) with a 10% increase in reproductive rate at 23°C. B) This results in strong selection on the mutation at cold temperature, but weak selection at hot temperature, and no epistasis. C) Warm temperatures increase molecular failure rates, resulting in decreased protein stability. Illustrated for three genetic backgrounds with stable (green), intermediate (orange) and unstable (yellow) wildtype protein. A mutation increasing stability (broken lines) was introduced on each background. D) Selection on the mutation is weak at cold temperature for all backgrounds but can become under very strong selection at hot temperatures, depending on the wildtype protein stability, and strong epistasis for fitness results. E) Protein fitness as the product of reproductive rate and protein stability is depicted for the wildtype (black), and the mutants with increased reaction rate (blue) and increased protein stability (red) for the yellow background.

Supplementary Literature Cited

1. Hochachka, P. W. & Somero, G. N. *Biochemical adaptation: mechanism and process in physiological evolution* (Oxford university press, 2002).
2. Angilletta, M. J. *Thermal adaptation: a theoretical and empirical synthesis* (Oxford University Press, 2009).
3. Evans, M. G. & Polanyi, M. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Transactions of the Faraday Society* **31,** 875–894 (1935).
4. Eyring, H. The activated complex in chemical reactions. *The Journal of Chemical Physics* **3,** 107–115 (1935).
5. Clarke, A. Is there a universal temperature dependence of metabolism? *Functional Ecology* **18,** 252–256 (2004).
6. Berger, D., Walters, R. & Gotthard, K. What limits insect fecundity? Body size-and temperature-dependent egg maturation and oviposition in a butterfly. *Functional Ecology* **22,** 523–529 (2008).
7. Dell, A. I., Pawar, S. & Savage, V. M. Systematic variation in the temperature dependence of physiological and ecological traits. *Proceedings of the National Academy of Sciences* **108,** 10591–10596 (2011).
8. Chen, P. & Shakhnovich, E. I. Thermal adaptation of viruses and bacteria. *Biophysical journal* **98,** 1109–1118 (2010).
9. Dill, K. A., Ghosh, K. & Schmit, J. D. Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences* **108,** 17876–17882 (2011).
10. Jørgensen, L. B., Ørsted, M., Malte, H., Wang, T. & Overgaard, J. Extreme escalation of heat failure rates in ectotherms with global warming. *Nature* **611,** 93–98 (2022).
11. Bloom, J. D. *et al.* Thermodynamic prediction of protein neutrality. *Proceedings of the National Academy of Sciences* **102,** 606–611 (2005).
12. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nature Reviews Genetics* **6,** 678–687 (2005).

13. Bershtein, S., Serohijos, A. W. & Shakhnovich, E. I. Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations. *Current opinion in structural biology* **42,** 31–40 (2017).

14. Echave, J. & Wilke, C. O. Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. *Annual review of biophysics* **46,** 85–103 (2017).

15. Agozzino, L. & Dill, K. A. Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proceedings of the National Academy of Sciences* **115,** 9092–9097 (2018).

16. Allan Drummond, D. & Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nature Reviews Genetics* **10,** 715–724 (2009).

17. Chen, P. & Shakhnovich, E. I. Lethal mutagenesis in viruses and bacteria. *Genetics* **183,** 639–650 (2009).

18. Zeldovich, K. B., Chen, P. & Shakhnovich, E. I. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proceedings of the National Academy of Sciences* **104,** 16152–16157 (2007).

19. Goldstein, R. A. The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins: Structure, Function, and Bioinformatics* **79,** 1396–1407 (2011).

20. Feder, M. E., Bennett, A. F. & Huey, R. B. Evolutionary physiology. *Annual review of ecology and systematics* **31,** 315–341 (2000).

21. Deutsch, C. A. *et al.* Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences* **105,** 6668–6672 (2008).

22. Johansson, F., Orizaola, G. & Nilsson-Örtman, V. Temperate insects with narrow seasonal activity periods can be as vulnerable to climate change as tropical insect species. *Scientific Reports* **10,** 8822 (2020).

23. Baur, J., Zwoinska, M., Koppik, M., Snook, R. R. & Berger, D. Heat stress reveals a fertility debt owing to postcopulatory sexual selection. *Evolution Letters* **8,** 101–113 (2024).

24. Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. Effects of size and temperature on metabolic rate. *science* **293,** 2248–2251 (2001).

25. Dowling, D. K. & Simmons, L. W. Reactive oxygen species as universal constraints in life-history evolution. *Proceedings of the Royal Society B: Biological Sciences* **276,** 1737–1745 (2009).