Electronic Supplementary Material for "Evolutionary tradeoffs and the structure of allelic polymorphisms"

Table of Contents

Supplementary Section 1: Allelic polymorphisms align with the Pareto front when locales can be occupied by more than one individual

The results presented in the main text used a selection model where each locale could only be inhabited one individual. To examine the sensitivity of the results to this assumption, we ran simulations in which h individuals can occupy each locale. Note that this means that for a population size of N there are N/h locales. We simulated two scenarios. In the first scenario, the optimal phenotypes in the different locales are equally-spaced along the entire Pareto front (in such cases the distance between adjacent optimal phenotypes increases with h, so the average mutation size becomes smaller relative to the distance between the locales). In the second scenario, the spacing between the optimal phenotypes in different locales doesn't depend on h but equals the distance between the archetypes divided by N-1 (identical to what was used in all simulations in the main text), which means that the optimal phenotypes in the available locales occupy only 1/h of the front. In both scenarios, the simulation process was as described in the main text and Section S6, except that we created h duplicates of each locale (Fig S1, S2).

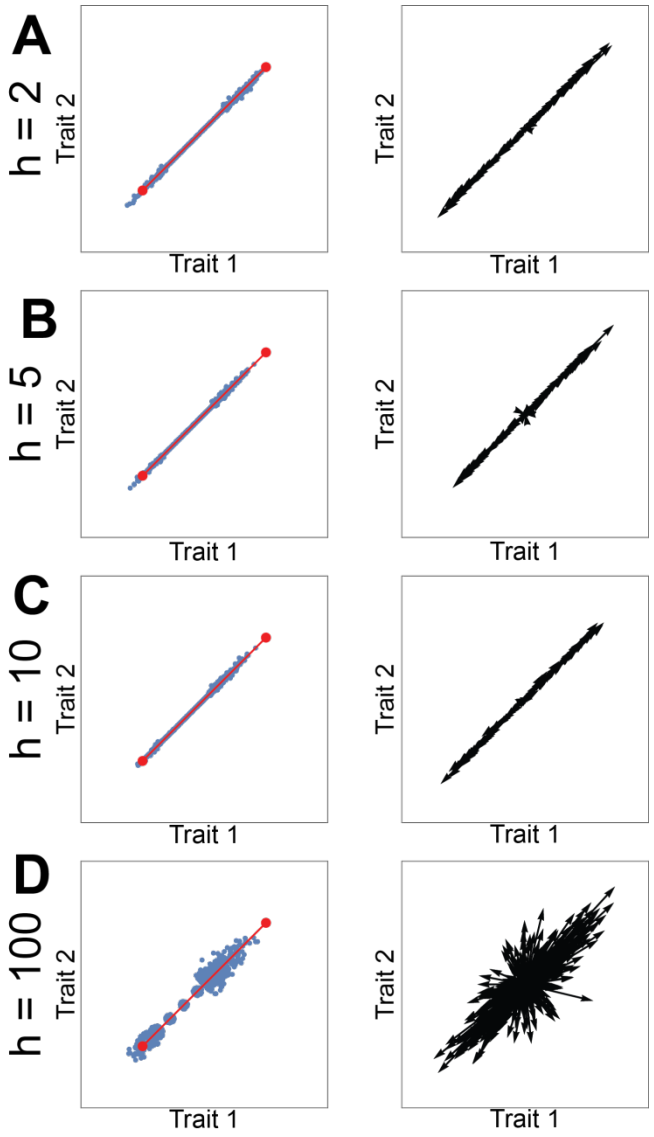

Figure S1. Polymorphisms align with the front when locales can be occupied by more than one individual. N/h locales, occupied each by h individuals, were chosen such that their optimal phenotypes are equally spaced along the Pareto front. The distance between 2 adjacent locales' optimal phenotypes is 100h/N, and mean mutation effect size is 1. Phenotypes (left) and effect of selected allelic polymorphisms (right) after 100,000 simulation-generations for (A) h=2, (B) h=5, (C) h=10 and (D) h=100. Simulation parameters: N=1000, μ=0.05, Q_1=Q_2 =I, infinite recombination.

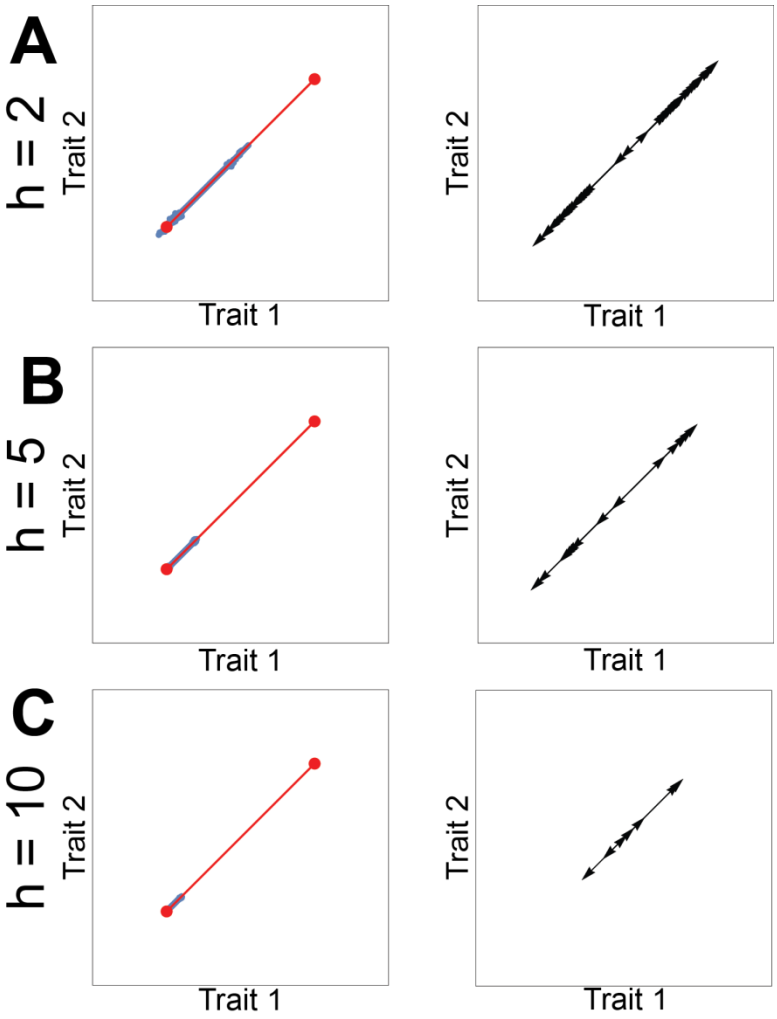

Figure S2. Polymorphisms align with the front when locales can be occupied by more than one individual. N/h locales, occupied each by h individuals, were positioned close to archetype 1. The distance between 2 adjacent locales' optimal phenotypes is 100/N (as all cases studied in the main text), and mean mutation effect size is 1. Phenotypes (left) and effect of selected allelic polymorphisms (right) after 100,000 simulation-generations for (A) h=2, (B) h=5, (C) h=10. Simulation parameters: N=1000, μ=0.05, Q_1=Q_2 =I, infinite recombination.

Supplementary Section 2: Polymorphisms align with the Pareto front in an example in which fitness depends non-linearly on performance functions

The results presented in the main text for evolution under two tasks were obtained by using a fitness function at each locale w_1 that linearly combines the two performance functions, $f_{w_1}(\vec{x}) = w_1 P_1(\vec{x}) + (1 - w_1) P_2(\vec{x})$ $(0 \le w_1 \le 1)$. Shoval et. al [1] show that as long as the fitness function $f_{w_1}(P_1(\vec{x}), P_2(\vec{x}))$ is monotonically increasing with both $P_1(\vec{x})$ and $P_2(\vec{x})$, the maximum of f_{w_1} will reside on the line between the archetypes, even when f is nonlinear. Here, we demonstrate that polymorphisms align with the Pareto front for a class of non-linear fitness-functions that are monotonic in the performance functions: $f_{w_1}(\vec{x}) = P_1(\vec{x})^{w_1} P_2(\vec{x})^{1-w_1}$ $(0 \le w_1 \le 1)$. We use performance functions that decay from the archetype with an Euclidean norm: $P_i(\vec{x}) = m_i - \left(\vec{x} - \overrightarrow{A_1}\right)^2$. For simplicity, to avoid cases where the performances become negative and the power is not well-defined, we look at the equivalent problem of $P_i(\vec{x}) = m_i + \left(\vec{x} - \overrightarrow{A_1}\right)^2$, while trying to minimize $f_{w_1}(\vec{x})$. It is important to note, that unlike the linear case, where an extremum of $f_\alpha(\vec{x})$ is unique and is necessarily a minimum (when looking at $P_i(\vec{x}) = m_i + \left(\vec{x} - \overrightarrow{A_1}\right)^2$), in the multiplicative case the extremum can be either a minimum or a maximum. For example, if we choose $m_1 = m_2 = 0$, each f_α will have its minimum at one of the archetypes. When m_i increases, the range of points along the front that are a minimum increases. We chose to work with m_i = 1000, $A_1 = (0,0)$, $A_2 = 50(\sqrt{2}, \sqrt{2})$. Hence, locale-optima are not spread on the entire Pareto front but rather near its edges (See Fig S3A-B, E). The empty region on the front is due to an effect described in Shoval et al [1], for the case where performance functions have convex regions. Still, the polymorphisms that persist are aligned with the Pareto front (see Fig S3C-D, F).

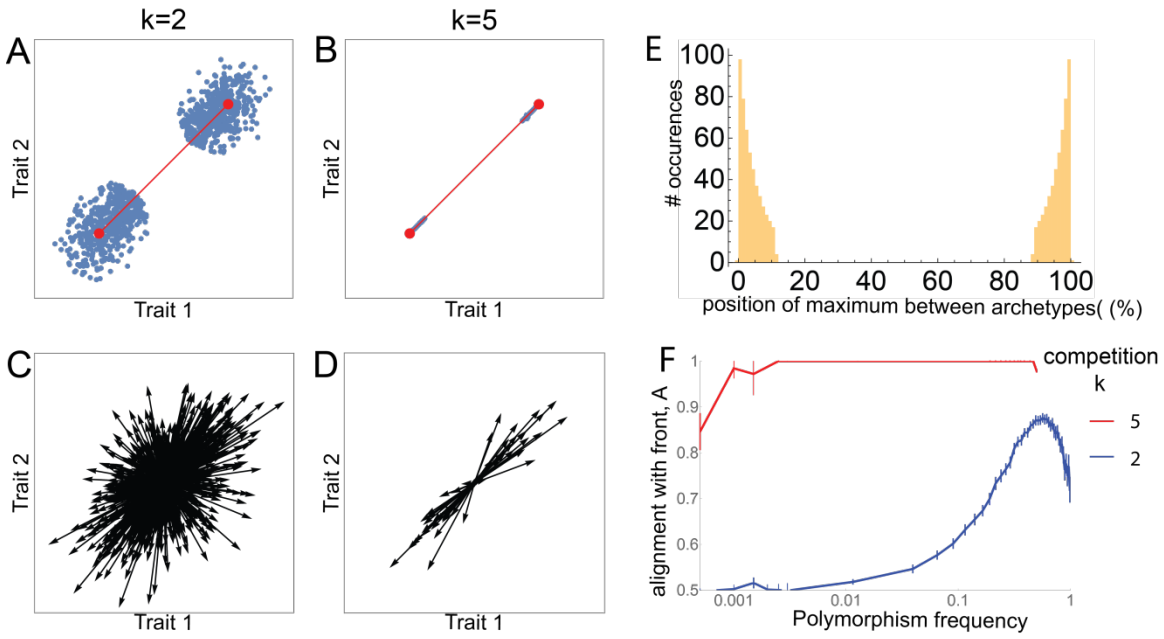

Figure S3. Polymorphisms align with the Pareto front even when fitness does not depend linearly on the performance functions. All results presented here were performed with nonlinear fitness functions of the form $f_{w_1}\hat{}i\ (\vec{x}) = P_1(\vec{x})^{w_1} P_2(\vec{x})^{1-w_1}$, $w_1 = i/(N-1)$, $i \in \{0,1,..,N-1\}$, $P_i(\vec{x}) = m_i + \left(\vec{x} - \overrightarrow{A_i}\right)^2$, $m_i = 1000$, $A_1 = (0,0)$, $A_2 = \sqrt{2}(50,50)$, N=1000, $\mu = 0.05$, infinite recombination. Phenotypes under (A) medium selection (k=2), (B) strong selection (k=5). (C-D) The phenotypic effects of all polymorphisms present in at least 1% of genomes after 100,000 generations for k=2 (C) and k=5 (D). (E) Histogram of the location of the fitness maximum for all locales w_1. Location is given as percent of the distance between the two archetypes. It is seen that no point in the middle of the front maximizes fitness. (F) Alignment of polymorphisms with the front as a function of polymorphism's frequency.

Supplementary Section 3. Simulations where locale maxima are sampled from normal distributions with various standard deviations result in alignment of polymorphisms with the Pareto front

In the simulations described in this manuscript, each locale has its own fitness function (individual selection surface) with locale-specific weights for the different performance

functions: $F_{\{w_i\}} = \sum_i w_i P_i$, where the weights w_i are positive. In the simulations described in the main text, we sampled w_i uniformly with $\Sigma w_i = 1$.

Here, we study the effects of varying the distribution of w_i in simulations of tradeoffs between two tasks in a two-dimensional trait space. In such situations there are two weights, w_1, w_2, such that $w_1 + w_2 = 1$ and $0 \leq w_1, w_2$. To test different distributions of locales, we sampled w_1 from a Gaussian distribution with mean 0.5 and standard deviation S of 0.01, 0.1, 1, 10, (and set $w_2 = 1 - w_1$). Negative w_1 was taken as 0, and $w_1 > 1$ was taken as 1. For small STDs, S = 0.01 and 0.1, genetic variant effects align with the front for population sizes above a few hundred ($\theta < 1°$, Fig S4B). For large standard deviations, S=1,10, the weight distribution results in most locales concentrated near the two archetypes with few in the middle between the archetypes. For this polarized locale distribution, alignment increases with population size, but is weaker (Fig S4B). See Fig S4.

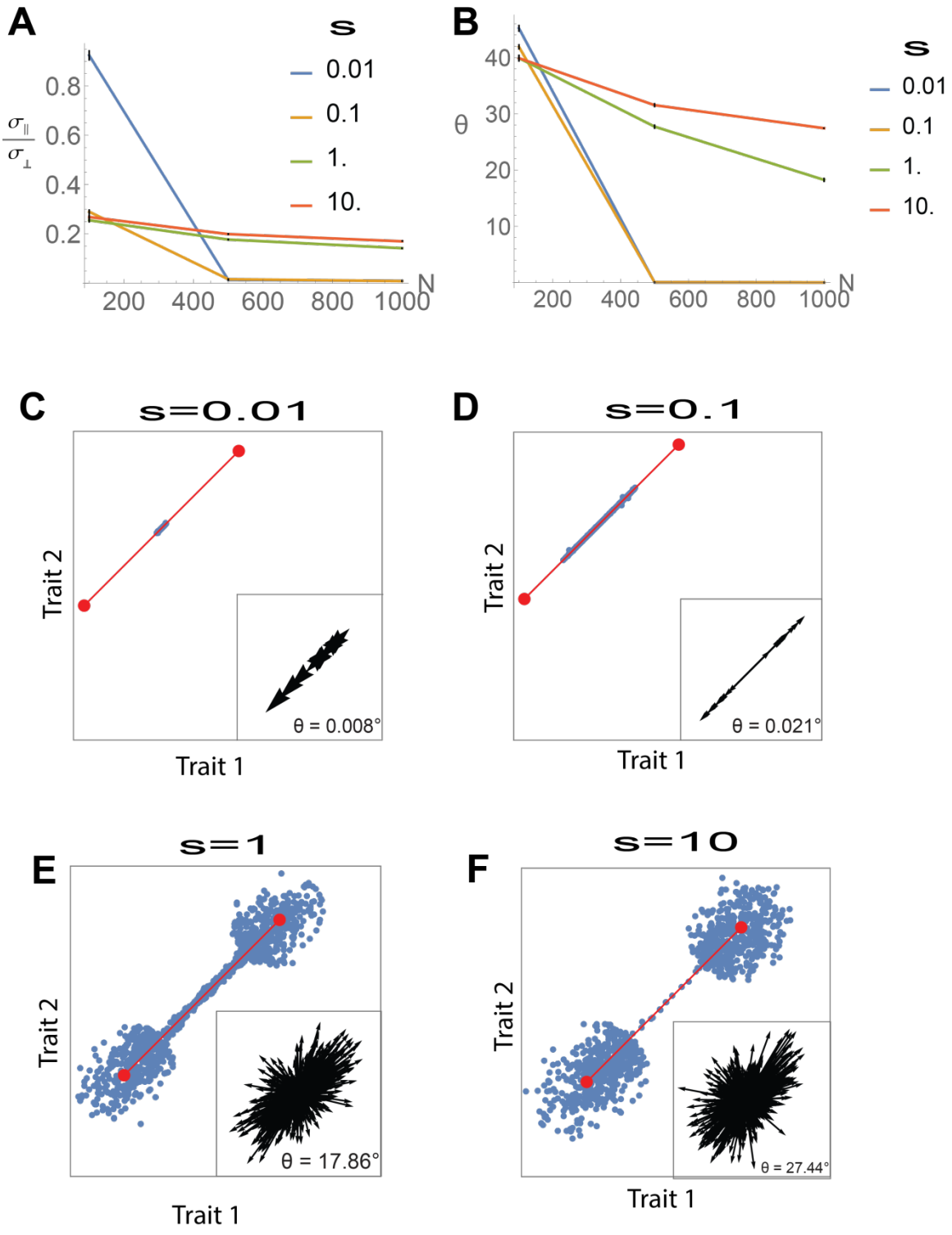

Figure S4. Results stay qualitatively the same when positions of locale-optima are sampled from a normal distribution with various standard deviations. (A). Ratio between standard deviation perpendicular and parallel to the front decreases with population size N, for all standard deviation examined. (B). Median angle of polymorphisms to the front decreases with population size N, for all standard deviations studied. Simulation parameters: N=1000. μ=0.05,

k=2, infinite recombination. (C-G). Phenotypes in the population after 100,000 generations of a simulation performed with $\{w_1\}'s$ sampled from a normal distribution $N(0.5, S^2)$, where S is indicated in the title. Insets: The phenotypic effects of all allelic polymorphisms present in at least 1% of genomes increasingly align with the Pareto front. Each arrow represents the phenotypic effect of one polymorphism (magnified for illustration purposes).

Supplementary Section 4: Polymorphisms align with the front when mating probability depends on fitness

In the simulations presented in the main text, the fitness was used only to determine which individual has the highest fitness in a locale, and hence only the relative order of individuals' fitness in each locale was important. Here, we also study situations in which each locale offers a different maximal fitness, in the sense that the locale affects the mating probability of the individual. When mating probability depends on fitness, the absolute value of fitness is important, and results may depend on the "quality" of different locales (which relates to the maximal fitness in each locale). To account for this, we consider fitness functions of the form: $F_{\{w_i\}} = \mathrm{F}_{0\{\mathrm{w_i}\}} + \sum_{\mathrm{i}} \mathrm{w_i P_i}$, $\mathrm{w_i} \geq 0$, $\sum_{\mathrm{i}} \mathrm{w_i} = 1$, and $F_{0\{w_i\}}$ is a locale-dependent constant, which allows each locale to have a different maximal fitness. $\mathrm{P_i}$ is the performance function at task i. $\mathrm{P_i} = -\mathrm{c}\left(\vec{\mathrm{T}} - \overrightarrow{\mathrm{A_1}}\right)^{\mathrm{T}} \mathrm{Q_i}\left(\vec{\mathrm{T}} - \overrightarrow{\mathrm{A_1}}\right)$, and thus the maximal fitness in a locale is given by $\mathrm{F}_{0\{\mathrm{w_i}\}} - \mathrm{F}_{\{\mathrm{w_i}\}_{\max}}$ where $\mathrm{F}_{\{\mathrm{w_i}\}_{\max}}$= $\mathrm{c}\sum_{\mathrm{i}} \mathrm{w_i}\left(\overrightarrow{\mathrm{T_{max}}} - \overrightarrow{\mathrm{A_1}}\right)^{\mathrm{T}} \mathrm{Q_i}\left(\overrightarrow{\mathrm{T_{max}}} - \overrightarrow{\mathrm{A_1}}\right)$ and $\overrightarrow{\mathrm{T_{max}}}$ is the phenotype that maximizes the fitness at the locale $\{\mathrm{w_i}\}$ (locale-optimum). The coefficient c = 0.001 was chosen to avoid negative fitness values in the region close to the Pareto front.

We examined three settings in which mating probability depends on fitness, for two tasks in a two-dimensional trait space. In the first setting, maximal fitness in all locales ($\mathrm{F}_{0\{\mathrm{w_i}\}} - \mathrm{F}_{\{\mathrm{w_i}\}_{\max}}$) are equal (Fig S5). In the second setting, we assumed maximal fitness is higher in locales whose optimal phenotype (locale-optimum) is close to archetype 1 than in locales with locale-optimum that is close to archetype 2, as illustrated in Fig

S6A. In the third setting, maximal fitness in locales with locale-optimum at the center of the front is higher than the maximal fitness in locales with locale-optimum near the archetypes (Fig. S7A). We find that for all settings, after a sufficient number of simulation generations, phenotypes reside on the front and polymorphisms align with the Pareto front (Fig S5-7).

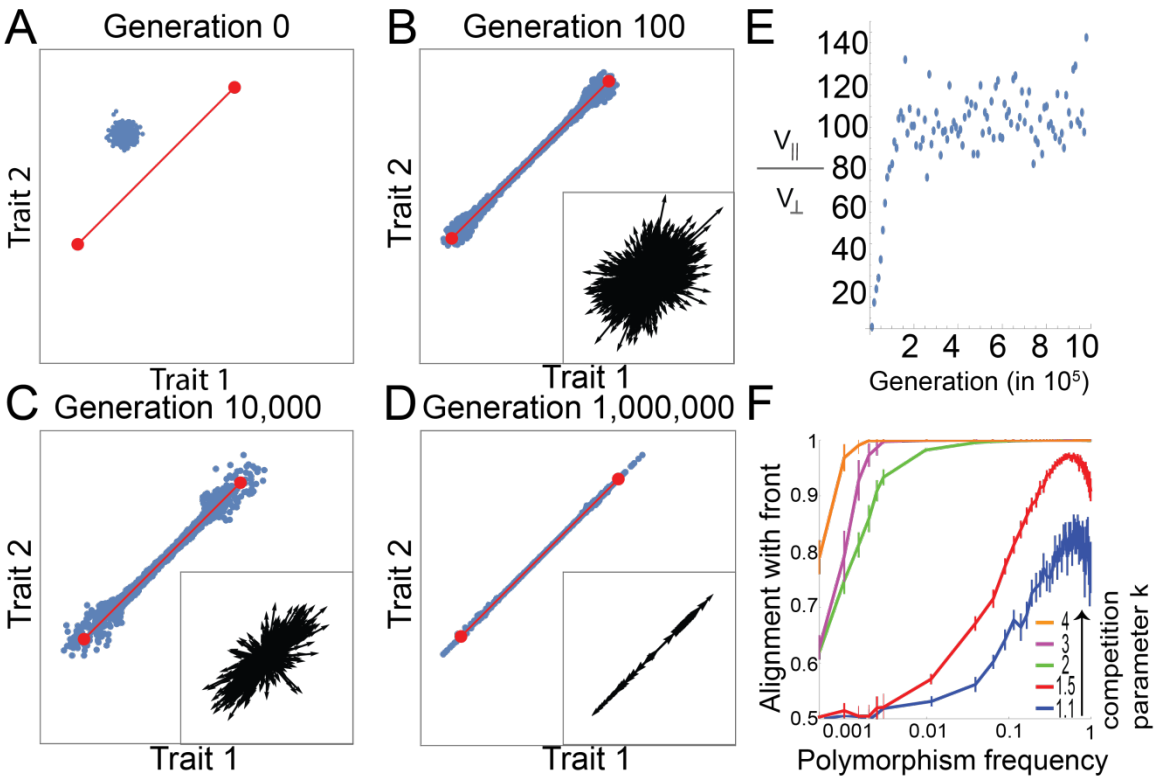

Figure S5. Phenotypes converge to the Pareto front by means of polymorphisms whose effects align with the front, when mating probability depends on fitness and fitness maxima is equal for all locales. (A-D) Snapshot of the phenotypes at different generations in a simulation, for the case of two tasks. The Pareto front is shown in red. Insets: The phenotypic effects of all allelic polymorphisms present in at least 1% of genomes increasingly align with the Pareto front. Each arrow represents the phenotypic effect of one polymorphism (magnified for illustration purposes). At generation zero, no mutation is present at >1% of the genomes. Simulation parameters: N =1000, μ=0.05, k = 2, infinite recombination. Median angle of polymorphisms relative to the front was 33°, 17°, 0.8°, for B-D, respectively. Axes are traits 1 and 2. **(E)** The ratio of the parallel and perpendicular variance of phenotypes with respect to the front increases with generations, and begins to plateau after ~20,000 generations. **(F)** Alignment of allelic polymorphism effects with the front increases with polymorphism frequency (log-linear scale) and selection strength. Simulation parameters are as above except k that varies as indicated. Error bars represent 95% confidence intervals from bootstrapping. Alignment is defined as $A = \frac{\langle |a_{\parallel}| \rangle^2}{\langle |a_{\parallel}| \rangle^2 + \langle |a_{\perp}| \rangle^2}$, where $\langle |a_{\parallel}| \rangle$ and $\langle |a_{\perp}| \rangle$ are the mean parallel and perpendicular component of the mutation effect vectors in each bin of mutation frequency. A = 1 and 0.5 occur when mutations are completely aligned or randomly oriented, respectively.

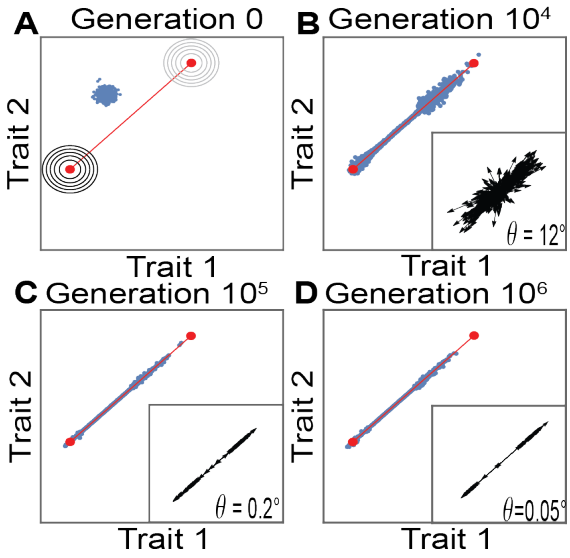

Figure S6. Phenotypes converge to the Pareto front by means of polymorphisms whose effect aligns with the front, when mating probability depends on fitness and fitness maxima is maximal near archetype 1. Simulation results for a setting where maximal fitness decreases linearly by a factor of 2 between locales close to archetype 1 and locales close to archetype 2, for the case of two tasks. **(A-D)** Snapshot of the phenotypes at different generations in a simulation. The Pareto front is shown in red. Insets: The phenotypic effects of all allelic

polymorphisms present in at least 1% of genomes increasingly align with the Pareto front. Each arrow represents the phenotypic effect of one polymorphism. At generation 0, no mutation is present at >1% of the genomes. Simulation parameters: N =1000, μ=0.05, k = 2, Q_1=Q_2=I, infinite recombination. θ is the median angle of polymorphisms relative to the front. Circles in A represent the contours of the fitness functions at locales that require only 1 of the tasks; darker contour color represents higher maximal fitness.

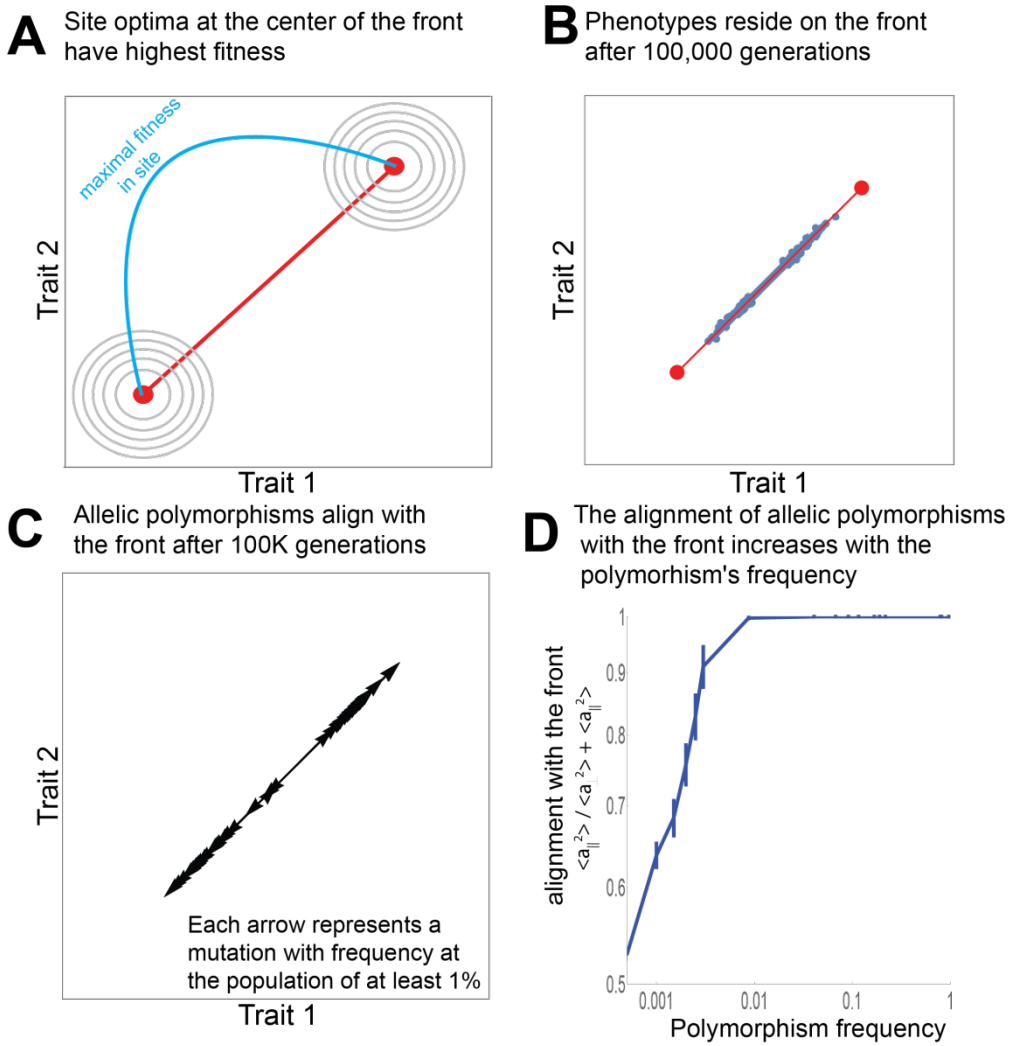

Figure S7. Allelic polymorphisms align with the front where the optimal phenotype in the locale with highest maximal fitness is at the center of the front. (A) We simulated a scenario in which performance functions decay with a Euclidean norm from the archetypes, and maximal fitness is highest for locales with locale-optimum at the center of the front and lowest for locales with locale-optimum near the archetypes (Pareto front is red line, maximal fitness in a locale is represented by the blue curve). (B) Phenotypes at the population after 100,000 generations reside on the Pareto front. (D) Allelic polymorphism at the populations (mutations with frequency > 1%) are aligned with the Pareto front. (D) Alignment of allelic polymorphisms of all frequencies

with the front as a function of the polymorphisms frequency. Errors bars represent 95% confidence interval and are calculated by bootstrapping. Simulation parameters: N=1000,μ=0.05,k=2, Q_1=Q_2=I, infinite recombination, maximal fitness in a locale $0 \leq w_1 \leq 1$ was $2 - 8(w_1 - 0.5)^2$

Supplementary Section 5. Results from simulations where the surviving phenotype at a locale is selected with probability proportional to $exp(\beta F)$

We ran simulation implementing a less strict selection scheme: instead of selecting the phenotype that has highest fitness F_{w_1} in the locale w_1, we select a phenotype with a probability proportional to $exp(\beta F_{w_1})$. Larger β results in higher probability of more-fit phenotypes to be selected, while low β results in a more equal probability for all phenotypes to be selected regardless of their fitness. We find that as β increases, the alignment increases (Fig S8).

Fitness function at locale w_1 was defined as $F_{w_1}(\vec{T}) = 1 + \mathrm{w}_1 P_1(\vec{T}) + (1 - w_1)P_2(\vec{T}) - F_{w_1 max}$. $F_{w_1 max}$ sets the maximal attainable fitness in locale w_1 to 1. $P_i(\vec{T}) = -0.001 (\vec{T} - A_i)^2$, $A_1 = (0,\ 0), A_2 = (100, 0)$.

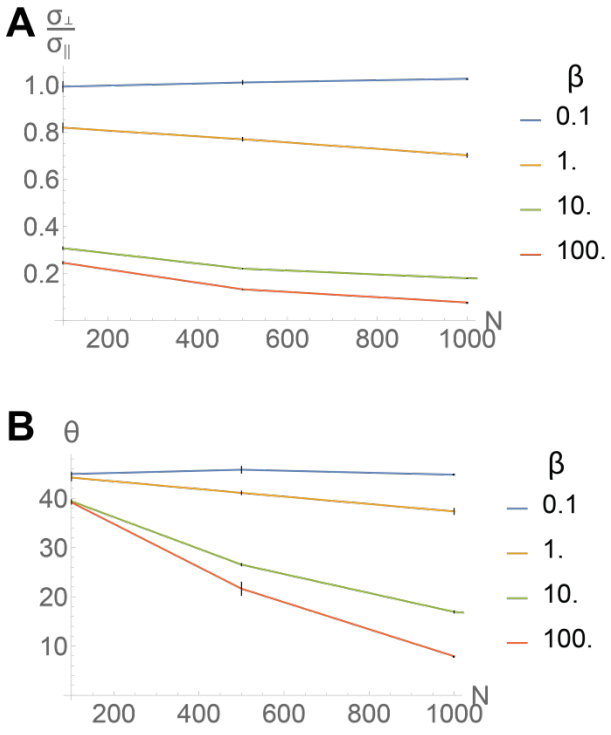

Figure S8. Polymorphisms align with the front and phenotypes reside on the front when the surviving phenotype at a locale is selected according to $exp(\beta F_i)$, for sufficiently large β. (A). Ratio between standard deviation perpendicular and parallel to the front decreases with population size N, for large enough β, and stays close to 1 (which means that phenotypes do not align with the front) for smaller βs. (B). Median angle of polymorphisms to the front decreases with population size N, for large enough β, and stays close to 45° (which means that polymorphisms do not align with the front) for smaller βs. Simulation parameters: Nμ=50, infinite recombination.

Supplementary Section 6: Technical detail on the simulations

Each individual is represented by its genome. The genome is diploid, meaning it is composed of two chromosomes, both containing the same genes. Each gene has a location on the genome, that we represent by a real number $r \in [0,1]$ (infinite site model [2,3]). Each gene can be a wild-type allele, which we model to have no effect on the phenotype, or a mutant allele, which affects the phenotype. A mutant allele is

represented by a vector (with the same dimension as that of the trait-space), which determines the effect that the mutation has on the phenotypes. The effects of mutations are additive, so the effect of the mutation does not depend on the genetic background. The phenotype is the vector sum of all its mutation-effects.

We assume that there are N locales. At each locale, the fitness function is given by $F_{\{w_i\}} = \sum_i w_i P_i \ , w_i \geq 0, \ \sum_i w_i = 1$, where P_i is the performance function at task i.

$P_i = -\left(\vec{T} - \overrightarrow{A_l}\right)^T Q_i \left(\vec{T} - \overrightarrow{A_l}\right)$ (Q_i is the positive-definite matrix that defines the inner-product norm that performance *i* decays from archetype *i* with). Thus,

$F_{\{w_i\}} = -c \sum_i w_i \left(\vec{T} - \overrightarrow{A_l}\right)^T Q_i \left(\vec{T} - \overrightarrow{A_l}\right)$.

The population size is limited by N, the number of available locales. We select the locales such that the weights $\{w_i\}$, representing the relative importance of the different tasks in the locale, are uniformly distributed. Each locale can support only one individual, so that the population size is N. The fitness of an individual is determined by its fitness computed according to the locale it inhabits.

The simulation has 4 steps:

1.Initiation: we start with a population of N individuals, each occupying one locale, where both copies of all genes are wild-types. We add to each individual random mutations (we chose 10, but this step is not necessary). See details on adding mutations in the "mutation" section below.

2.Mating/Recombination: each generation begins with the production of offspring. We model either sexual mating or asexual mating.

In the sexual mating regime, we produce an offspring by selecting 2 parents randomly. To model the mating process, we use either free recombination [4], in the sense that

each allele has a 50% chance of passing to the offspring, with no linkage between alleles, or recombination with linkage, in which recombination was done by swapping chromosomes at a single recombination spot (in both cases each offspring gets one copy of each gene from each parent). We repeat this process to generate a total number of kN offsprings (k>1 is a parameter that determines drift strength).

In the asexual mating regime, we select kN parents randomly, and each offspring is identical to its parent.

3.Mutation: each offspring receives a number of new mutations that is Poisson-distributed with a mean μ. A new mutation appears in a randomly selected locale $r \in [0,1]$. Each mutation has an effect that is a randomly oriented vector in trait space. We used an exponential distribution of mutation effect sizes (length of the effect vector); other distributions [5] did not qualitatively affect the results. We assume an infinite site model [3,6], meaning that each new mutation introduces a new segregating site, and we neglect the possibility of multiple mutations at the same site.

4. Selection: There are kN offsprings, but the N locales support only N individuals. To select these N, we choose a locale at random. The individual that is selected to occupy the locale is the one which has the highest fitness at this locale out of all remaining individuals. We select a locale out of all empty locales, and occupy it with the individual that is most-fit for it out of the individuals that weren't yet selected. We repeat this process until all locales are occupied. The N individuals that occupy the N locales survive for the next generation. Others are removed.

We repeat steps 2-4 for a desired number of generations (typically 10N).

Supplementary Section 7: Allelic polymorphisms align with the front when the individual selection surfaces in locales have ridges that are perpendicular to the front

We ran simulation where the two performance functions, and as a result individual selection surface (ISS) in all locales, have ridges that point perpendicular to the front, rather than circular contours as studies in the main text. We find that also in this case, phenotypes evolve to the front, and selected polymorphisms have phenotypic effects that align with the front (Fig S9).

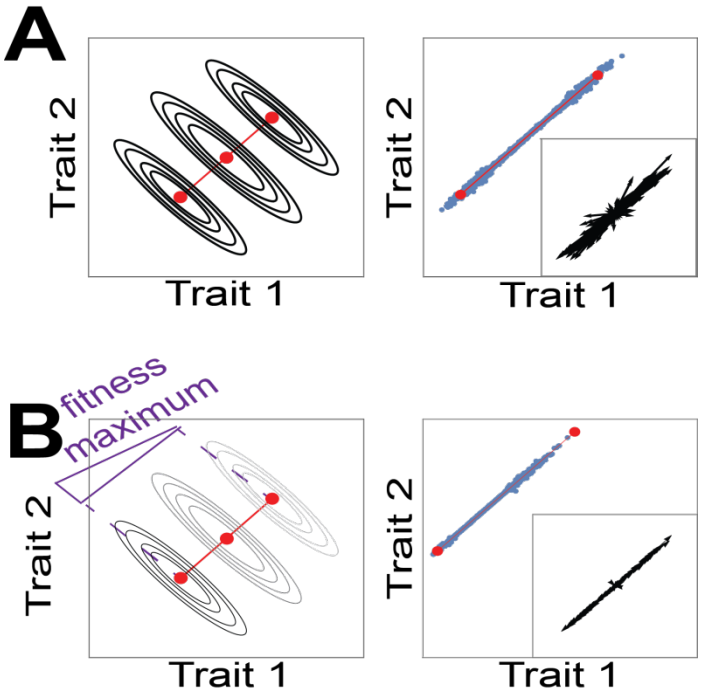

Figure S9: Phenotypic effects of polymorphisms align with the front when performance functions have ridges that are perpendicular to the front. Left panels depict fitness-function contours in 3 different locales: two that require only single tasks, and one that requires both tasks with equal weights. Darker contour color represents higher maximal fitness (as described in supplementary section S4). (A) has equal maximal fitness across all locales, while in (B) maximal fitness is higher in locales with locale-optimum closer to archetype 1, and mating probability depends on the parent's fitness. Right panels show phenotypes after 10^5 generations (blue dots) and the Pareto front (red line), as well as effects of allelic polymorphisms present in at least 1% of the genomes (black arrows).

Supplementary Section 8: Alignment of polymorphisms with the front for various population sizes and mutation rates

To quantify the alignment of polymorphisms effects with the front as a function of their frequency in the population, we define for each mutation its perpendicular and parallel effects with respect to the direction of the Pareto front. We bin polymorphisms according to their frequency, and computed for each bin the alignment with the front defined as A = $\frac{\langle |a_{\parallel}| \rangle^2}{\langle |a_{\parallel}| \rangle^2 + \langle |a_{\perp}| \rangle^2}$, where $\langle |a_{\parallel}| \rangle$ and $\langle |a_{\perp}| \rangle$ are the mean parallel and perpendicular component of the mutation effect vectors. A = 1 occurs when mutations are completely aligned with the Pareto front, and A = 0.5 corresponds to randomly oriented mutations.

Fig. 2F in the main text presents the alignment A for different competition parameters k. We also studied the dependence of the alignment of polymorphisms A (as defined in the main text) with the front on the population size N and mutation rate μ. To study the dependence on N, we ran simulations with population sizes of 100, 500, 1000, and 2000. The mutation rate μ was changed to keep the mutation inflow per generations constant across the simulations. Other parameters were kept constant. See Figure S10A. To study the dependence on μ, we ran simulations with a population sizes of 1000 and μ between 0.01-0.05. See Figure S10B.

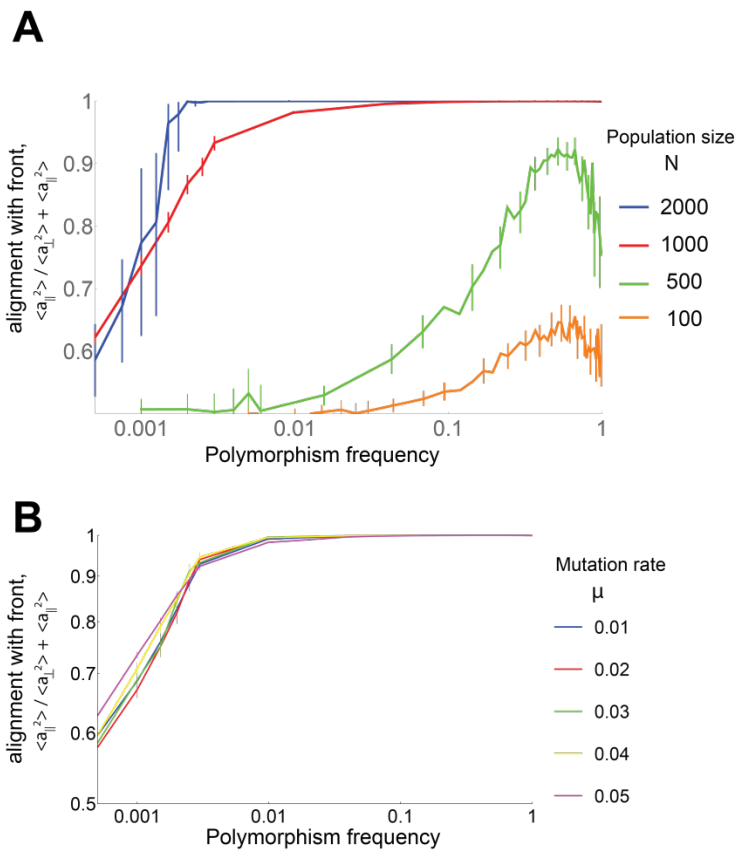

Figure S10. Segregating mutations have effects that are aligned with the Pareto front. Graphs present the alignment of the allelic polymorphism's phenotypic effects with the front (A = $\frac{\langle |a_{\parallel}| \rangle^2}{\langle |a_{\parallel}| \rangle^2 + \langle |a_{\perp}| \rangle^2}$) as a function of the polymorphisms frequency among genomes (log-linear scale). Different curves correspond to simulations run with different population sizes (A), or different mutation sizes (B), as indicated in the legend. All simulations were done with k = 2, Q_1=Q_2 =I, μN = 50 (A), or N=1000 (B), and infinite recombination. **(A)** Alignment increases with polymorphism frequency for N=1000 and N=2000, while for lower N alignment increases until polymorphisms penetrate around half the population and then slightly decreases (see Supplementary Section 9 for more details). Alignment also increases with population size N. **(B)** Alignment does not depend significantly of mutation rate. Error bars are estimated using bootstrapping and represent 95% confidence intervals.

Supplementary Section 9: Alignment with the front is greatest at mid-range frequency when selection is not strong

We see in Fig 2F and Fig. S10 that for weaker selection (population size N=1000 and k=1.1 or k=1.5, or N=100 or N=500 with k=2), the maximal alignment with the front is attained for polymorphisms of frequency of about 0.5 rather than at frequency of 1. We hypothesize that the cause for this effect is that polymorphisms with a larger perpendicular component spend a shorter amount of time at mid-frequencies, and hence they are less likely to be sampled. This is because negative selection on them is stronger, and hence those that don't fixate quickly disappear. When simulations are run with parameters that enhance selection (larger k or larger N), this phenomena is not observed (alignment is maximal at frequency of 1), because non-parallel polymorphisms don't manage to reach mid-range frequencies. This hypothesis agrees with detailed observation of the fixation dynamics of polymorphisms in the simulations (not shown).

Supplementary Section 10: When the Pareto front is curved, polymorphisms align with a local region on the front

We studied the distribution of allelic polymorphism under a tradeoff situation that results in a curved front. A curved front occurs when the performance functions have eccentric contours that point at an angle with respect to each other. To explain this point, note a straight line segment between the two archetypes is the resulting Pareto front when the performance functions decay with the same distance metric away from the archetypes. The distance metric of a task i is defined by an inner product norm: $\left(\vec{T}-\overrightarrow{A_i}\right)^T Q_i \left(\vec{T}-\overrightarrow{A_i}\right)$,

where $\overrightarrow{A_i}$ is the archetype at task i, $\vec{T}$ is the phenotype and $Q_i$is a positive definite matrix. Generally, Q_i can be represented as $\begin{pmatrix} Cos(\theta_i) & -Sin(\theta_i) \\ Sin(\theta_i) & Cos(\theta_i) \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \lambda_i^2 \end{pmatrix}\begin{pmatrix} Cos(\theta_i) & Sin(\theta_i) \\ -Sin(\theta_i) & Cos(\theta_i) \end{pmatrix}$. Here the angle θ_i determines the orientation of the elliptical contours of the performance function, and the parameter $\lambda_i$determines their eccentricity.

We proved in Sheftel et al. [7] that when the two performance functions depend on different matrices Q_1 and Q_2 (and thus have elliptic contours that point at an angle with respect to each other), the Pareto front is a section of a hyperbola that connects the archetypes instead of a straight line segment. This hyperbola doesn't deviate much from a straight line as long as contours are not too eccentric (both λs are not very large, see quantitative criteria in Sheftel et. al (2)).

We simulated evolution under tradeoff between two tasks that generate a curved front. We set the same eccentricity for both performance functions (same λ), and had their contours point at an angle of 0.5 relative to the Pareto front ($\theta_1 = \frac{\pi}{4} + 0.5$ and $\theta_2 = \frac{\pi}{4} - 0.5$). We start with medium selection (k=2). When λ =1.1 or 1.5, the Pareto front is mildly curved, phenotypes occupy the entire front and polymorphisms align relatively well with the Pareto front (Fig S11A-D). When λ is larger, λ =4, the Pareto front is highly curved, and the phenotypes occupy only the flat area at the top of the Pareto front (Fig S11E-F). Polymorphisms are no longer aligned with the line between archetypes for higher values of λ such as λ =10 (Fig S11G-H). If phenotypes occupied the entire curved front, many offspring would fall far from the front and would be sub-optimal. This is why selected phenotypes remain in a region where offspring are likely to remain near the front. A different situation may result if instead of random mating we used mating between neighboring locales.

The phenotypes seem to track the shape of the curved front more precisely the stronger the selection (higher k). This happens since k determines the number of offsprings, and larger number of offsprings increases the chance of performing well in a wider range of locales. Polymorphisms still align with the front, but alignment decreases (Fig S12).

We also studied curved fronts when mating probability depend on fitness, and the maximal fitness in different locales is not equal, but rather higher for locales with locale-optima near one of the archetypes (as descries in section S4, see Fig S13). In this case, phenotypes and polymorphism can align with a direction dictated by the local region on the front that is closer to the archetypes with the higher fitness (Figure S13).

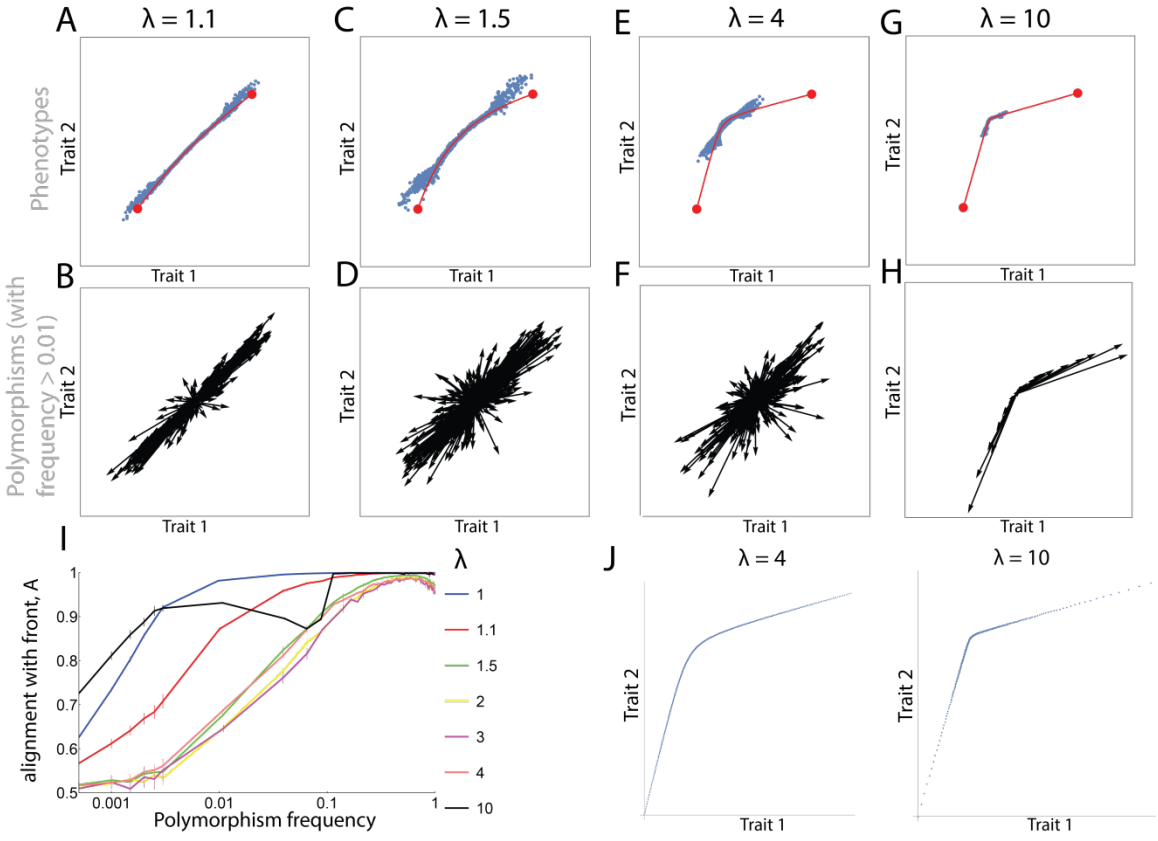

Figure S11. Polymorphisms align with the direction between the archetypes when the Pareto front is slightly curved, and selection is medium. (A,C,E,G) Snapshot of the phenotypes after 100,000 simulation-generations for different curvatures of the Pareto front. The curvature increases as the parameter λ increases. All simulation used the performance functions:

$P_i(\vec{x}) = (\vec{x} - \overrightarrow{A_i}) \begin{pmatrix} Cos[\theta] & -Sin[\theta] \\ Sin[\theta] & Cos[\theta] \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \lambda^2 \end{pmatrix} \begin{pmatrix} Cos[\theta] & Sin[\theta] \\ -Sin[\theta] & Cos[\theta] \end{pmatrix} (\vec{x} - \overrightarrow{A_i})$, where A_i is the archetype number i. θ_1 was set to $\frac{\pi}{4} + 0.5$ and θ_2 to $\frac{\pi}{4} - 0.5$. The Pareto front is shown in red. Simulation parameters: N =1000, $\mu = 0.05$, k = 2. **(B,D,F,H)** The phenotypic effects of all polymorphisms present in at least 1% of genomes at the stated generation. **(I)** Alignment of polymorphisms with the line between the archetypes for different λs. **(J-K)** Each point represents a locale-optimum, obtained for a different w_1 ($F_{w_1} = F_{0w_1} + (1 - w_1)P_1 + w_1P_2$) for $0 \leq w_1 \leq 1$). Each panel depicts all locales used for a certain λ. **(J)** $\lambda = 4$ **(K)** $\lambda = 10$. Locale-optima are not equally spaced along the Pareto front, even though the w_1′s, the weighting of the different tasks, are equally spaced between zero and one. As λ increases, the spacing becomes less uniform.

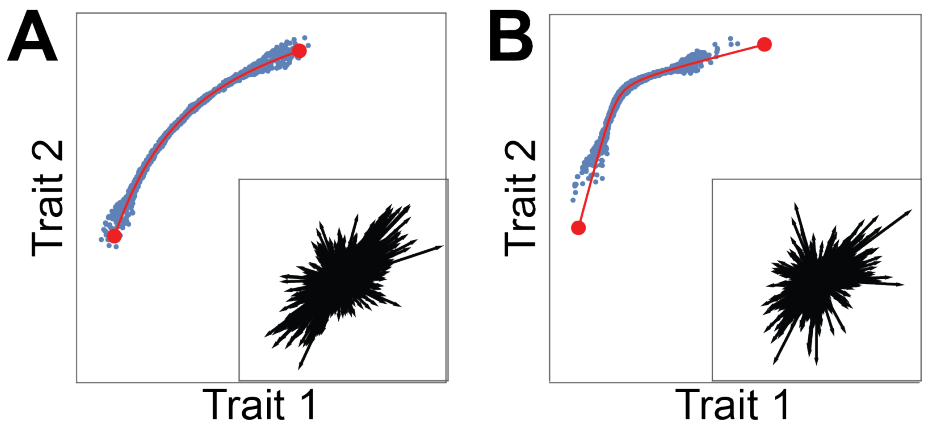

Figure S12. Polymorphisms align with the direction between the archetypes when the Pareto front is slightly curved, and selection is strong. (A-B) Phenotypes (blue points) in the population after 100,000 simulation generations. Pareto fronts are red curved and archetypes are red points. Insets: polymorphisms found in more than 1% of the genomes. Each arrow represents the phenotypic effect of 1 polymorphism. Simulation parameters: N=1000, k=5, μ=0.05. All simulation used the performance functions: $P_i(\vec{x}) = (\vec{x} - \overrightarrow{A_i}) \begin{pmatrix} Cos[\theta] & -Sin[\theta] \\ Sin[\theta] & Cos[\theta] \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \lambda^2 \end{pmatrix} \begin{pmatrix} Cos[\theta] & Sin[\theta] \\ -Sin[\theta] & Cos[\theta] \end{pmatrix} (\vec{x} - \overrightarrow{A_i})$, where A_i is the archetype number i. θ_1 was set to $\frac{\pi}{4} + 0.5$ and θ_2 to $\frac{\pi}{4} - 0.5$. **(A)** $\lambda = 1.5$ **(B)** $\lambda = 4$.

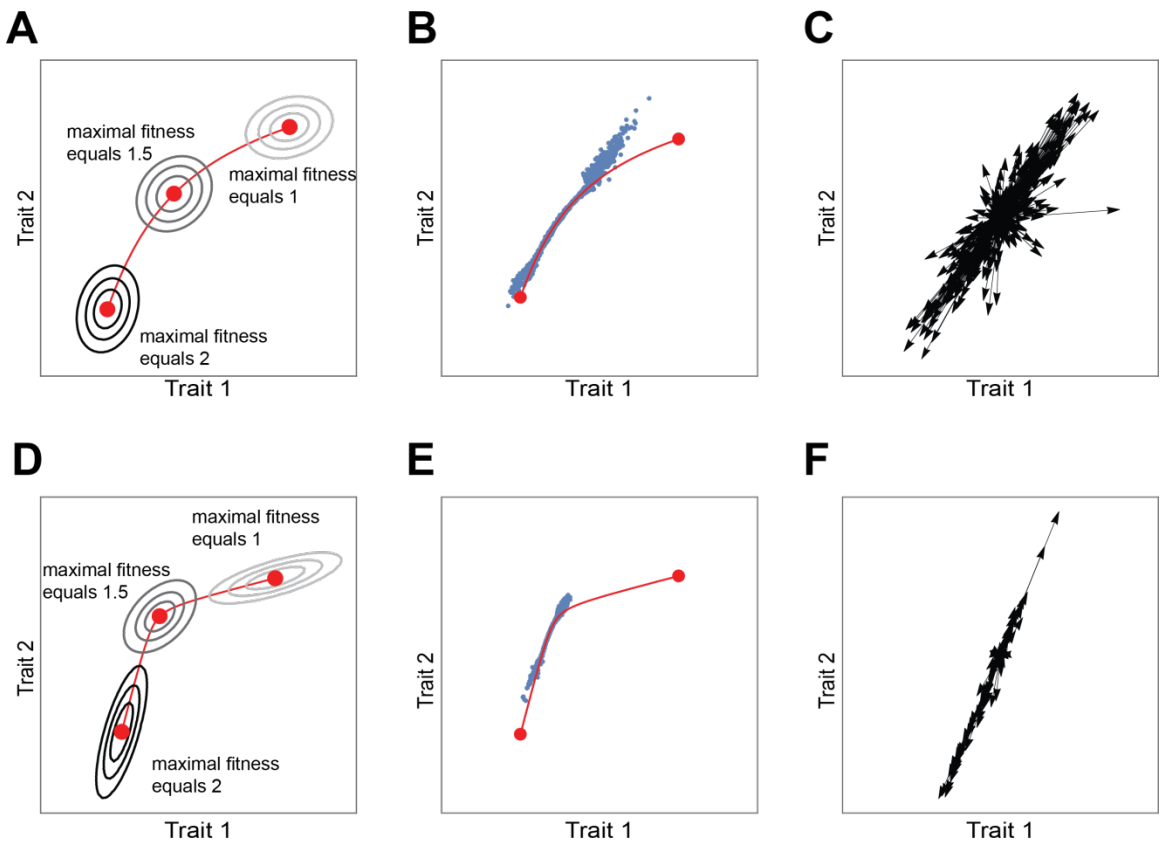

Figure S13. Polymorphisms align with a local direction on the front. Left column presents the contours of the fitness function at three different locales. Darker contour color represents higher maximal fitness. Middle column shows the phenotypes in the population after 100,000 simulation generations. Right columns are the polymorphisms found in more than 1% of the genomes. Each arrow represents the phenotypic effect of 1 polymorphism. Simulation parameters: N=1000, k=2, μ=0.05, maximal fitness in the locale with archetype 1 as a locale-optimum is twice the maximal fitness in the locale with archetype 2 as a locale-optimum. Parameters of the matrices Q_1 and Q_2 were: $\theta_1 = \frac{\pi}{4} + 0.5$, $\theta_2 = \frac{\pi}{4} - 0.5$. In the upper row $\lambda = 1.5$ for both performance functions, in the lower row $\lambda = 4$ for both Q_1 and Q_2.

Supplementary Section 11: Polymorphisms align with the triangular Pareto front when maximal fitness in different locales is not equal.

We studied a scenario with three tasks, where the mating probability depends on fitness (as described in Supplementary Section 4), and maximal fitness of locale optimum

located at archetype 3 is 2, 5, or 10-fold higher than the maximal fitness of locale-optimum located at archetypes 1 and 2. Fitness in each locale $\{w_1 w\}$ is $F_{\{w_1,w_2\}} = F_{0\{w_1,w_2\}} + w_1 P_1 + w_2 P_2 + (1 - w_1 - w_2) P_3$, $(w_1 \geq 0, w_2 \geq 0, w_1 + w_2 \leq 1)$. We find that in all cases polymorphism effects align with the front (Fig S14).

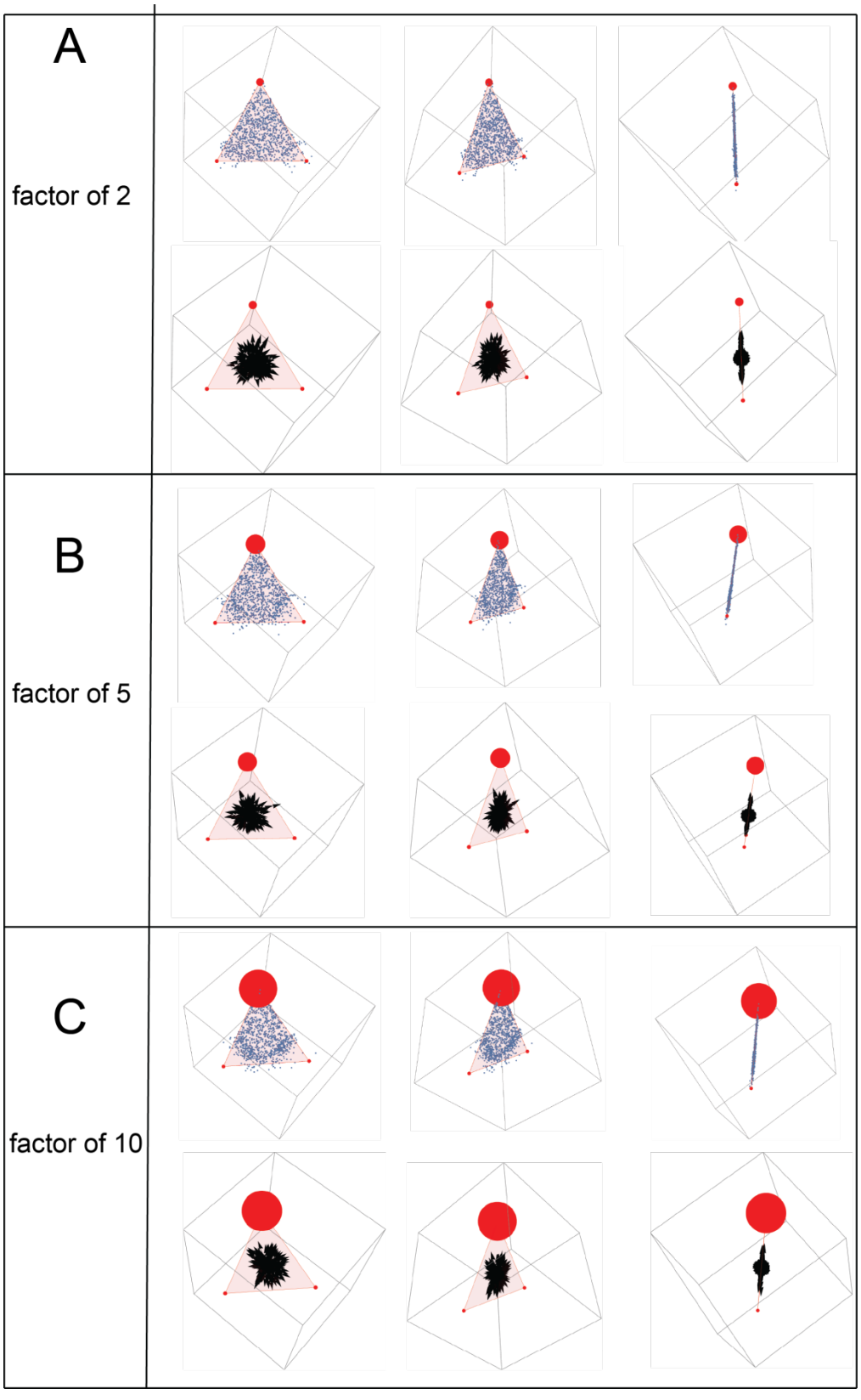
A
factor of 2
B
factor of 5
C
factor of 10

Figure S14. Polymorphisms align with the triangular Pareto front when maximal fitness in different locales is not equal. We studied a scenario where the maximal fitness at a locale with locale -optimum at archetype 3 is 2, 5, or 10-fold higher than the maximal fitness at a locale whose locale -optimum is at archetypes 1 and 2, and the decay in fitness between them is linear. In all figures, Pareto front is a red triangle, archetypes are red points whose size corresponds to the value of maximal fitness at the locale whose locale -optimum is at the archetype. (A-C). Phenotypes and polymorphisms after 100,000 generations from three different angles, for a 2-fold (A), 5-fold (B) and 10-fold (C) difference between the fitness at the archetypes.

Supplementary Section 12: Polymorphisms with epistatic effects between the two copies of the same allele, that have a large component perpendicular to the front, are selected against.

We consider the effects of tradeoffs on the non-additive interaction (epistasis) between two alleles of the same gene. We modelled this type of epistasis by assigning to each mutation *i* a randomly oriented vector $\overrightarrow{d_i}$, such that the heterozygote mutant effect is $\overrightarrow{m_i}$ and the homozygote is $2\overrightarrow{m_i} + \overrightarrow{d_i}$ (Fig S15A-C). We computed the alignment of the epistasis vectors $\overrightarrow{d_i}$ with the front, $D = \frac{\langle |d_{\parallel}| \rangle^2}{\langle |d_{\parallel}| \rangle^2 + \langle |d_{\perp}| \rangle^2}$, where $\langle |d_{\parallel}| \rangle$ and $\langle |d_{\perp}| \rangle$ are the mean parallel and perpendicular component of the epistasis effect vectors**.** For common polymorphisms, for which a sizable fraction of the population is homozygous [8], *D* approaches 1. Thus, epistatic effects of prevalent allelic polymorphisms are also oriented along the Pareto front (Fig S15D-G).

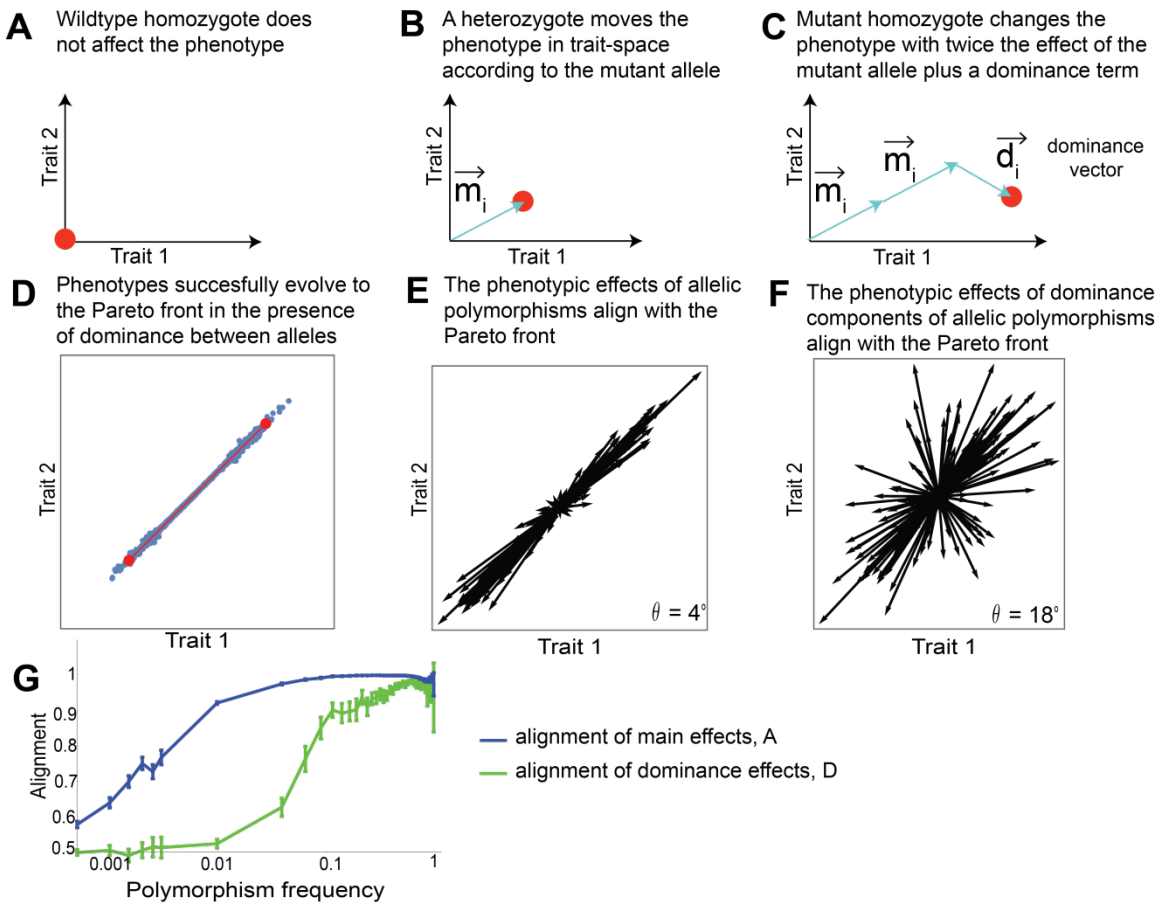

Figure S15. Epistatic effects between two copies of the same allele, which have a large component perpendicular to the front, are selected against. (A-C) Definition of epistatic effects between two alleles $\vec{d_i}$, showing the phenotypic effect of wildtype homozygote, heterozygote and mutant homozygote. **(B)** Phenotypes after 100,000 simulation generations (parameters: $N = 1000, \mu = 0.05, k = 2$, infinite recombination). **(E)** Phenotypic effect of the additive effect $\vec{m_i}$ of polymorphisms (frequency > 1%) in the population, θ is the median angle (in degrees) of polymorphisms relative to the front. **(F)** Phenotypic effect of the dominance $\vec{d_i}$ of polymorphisms (frequency > 1% in the population), θ is the median angle (in degrees) of polymorphisms relative to the front. **(G)** The alignment of dominance with the front, ($\frac{\langle |d_{\parallel}| \rangle^2}{\langle |d_{\parallel}| \rangle^2 + \langle |d_{\perp}| \rangle^2}$, green plot) rises with polymorphism frequency. Error bars are estimated using bootstrapping and represent 95% confidence intervals. Blue graph represents the alignment A of the main effects for comparison.

Supplementary Section 13: Molecular mechanism that can generate a triangular front

We study a regulatory mechanism where three pathways are regulated by three regulators X_i that compete over a limiting factor R. This mechanism is inspired by bacterial gene regulation, in which the regulators X_i are sigma factors that compete over binding to RNA polymerase R and provide it with specificity to bind to gene promoters.

The binding reactions can be described by a mass action model:

$$[\mathrm{R_f}] + [\mathrm{X_i}] \xrightarrow{k^i_{ON}} [\mathrm{RX_i}]$$

$$[\mathrm{RX_i}] \xrightarrow{k^i_{OFF}} [\mathrm{R_f}] + [\mathrm{X_i}]$$

For i=1,2,3, where $[\mathrm{R_f}]$ is the concentration of free R (unbound to any X_i). At equilibrium

$[\mathrm{RX_i}] = \frac{k^i_{ON}}{k^i_{OFF}}[\mathrm{R_f}][\mathrm{X_i}]$.

We redefine $\mathrm{X_i}$ as $\frac{k^i_{ON}}{k^i_{OFF}}[\mathrm{X_i}]$, to obtain $[\mathrm{RX_i}] = [\mathrm{R_f}][\mathrm{X_i}]$.

The total concentration of R in the cell, $\mathrm{R_T}$, is the sum of the unbound and bound forms:

$$[\mathrm{R_T}] = [\mathrm{R_f}] + \sum_i [\mathrm{RX_i}]$$

We consider the case in which binding is tight (regulator concentrations are much higher than their $K_d=\frac{k^i_{ON}}{k^i_{OFF}}$ for binding to R) and hence free R is limiting. This case is defined by very little free R: $\mathrm{R_f} \ll \sum_i [\mathrm{RX_i}]$. In this case,

$$[R_T] \cong \sum_i [RX_i] = \sum_i [R_f][X_i]$$

and hence

$$[R_f] = \frac{[R_T]}{\sum_i [X_i]}$$

The output traits are the expression of gene j, T_j. The expression depends on the concentration of the bound complex X_i R through the parameters ω_{ij}. In bacterial regulation, ω_{ij} represent the rate of transcription initiation from the gene promoter by RNA polymerase bound to the sigma factor Xi. The total transcription initiation rate is

$$T_j = \sum_i \omega_{ij}[RX_i] = \sum_i \omega_{ij}\,[R_f][X_i] = \sum_i \omega_{ij} \frac{[R_T][X_i]}{\sum_m [X_m]}$$

For convenience we define R=$[R_T]$, remove the brackets to obtain the equation in the main text:

$$(1)\ T_j = R \sum_i \omega_{ij} \frac{X_i}{\sum_m X_m}$$

The simulations that model the molecular mechanism presented here are identical to the simulation presented in the other sections in their selection and recombination schemes. They differ from the previously-described simulations in the mutation scheme, mutation effect, and genotype->phenotype mapping. Each mutation affects one of the parameters R, X_i, or ω_{ij}. We use product-rule mutations: each mutation multiplies the value of the parameter by a number drawn from a log-normal distribution with a mean of 1 and a standard deviation of 1. Product rule mutations are thought to be a better approximation for the impact of mutations on biochemical parameters than mutations in which a random number is added (sum-rule) [9]. Using standard sum-rule mutations yields the

same conclusions. The phenotype $\vec{T}$ depends on the parameters as described in Eq 1. Simulations were performed with the three archetypes: $A_1 = \{2,1,1\}$, $A_2 = \{1,2,1\}$, $A_3 = \{1,1,2\}$, providing a triangle that is at an angle to all trait axes, and performance functions that decay with Euclidean distance from the archetypes.

This model can be generalized to any number of regulators X_i, and the resulting front is a polytope with a number of vertices equal to the number of regulators (a line is provided by two regulators, a tetrahedron by four and so on).

Supplementary Section 14: Parameters ω_{ij} of the molecular mechanism we model become fixed during the simulation

In the main text we described a molecular mechanism that we simulated to study epistasis. We mention that the weights ω_{ij} become fixed at some point and the variation in phenotypes is due to mutations in the X factors. To show that ω_{ij} become fixed, we plot for each ω_{ij} its standard deviation divided by the mean in the population. After about 1000 generations, this value becomes very small for all the weights (Fig S16).

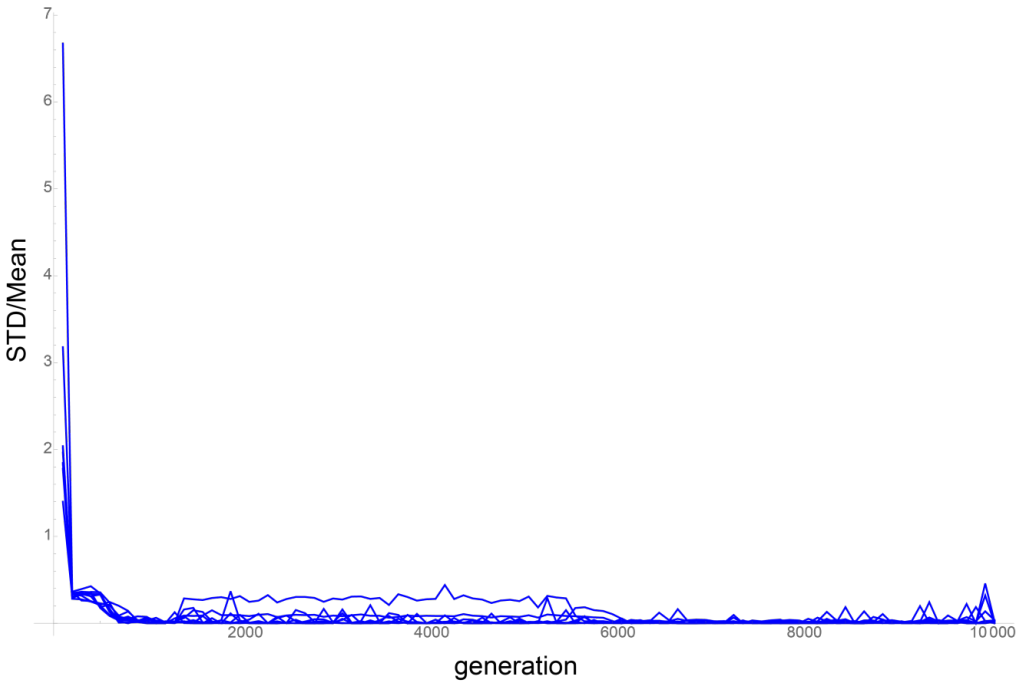

Figure S16. The weights ω_{ij} fixate after around 1000 generation. Each plot represents the standard deviation divided by the mean of a variable ω_{ij} in the population as a function of generation. After 1000 generations this value becomes close to zero, which signifies that ω_{ij}s are relatively constant in the population at this stage.

Supplementary Section 15: Artificial selection simulations

In figures 5ABD, and Fig. S17, we present results of artificial selection simulations. In the artificial selection simulations, we used as an initial population a population that was adapted to a 2-task Pareto front in a 2D trait-space (Fig 5A,B) or a 3-task Pareto front in a 3D trait-space (Fig 5A, S17), with parameters of N=1000, μ=0.05, and k=2 (2 tasks, Fig S17) or k=5 (3 tasks, Fig S17, Fig 5A), or N=100, μ=0.5, k=2 (2 tasks, Fig 5B). The phenotypes and genotypes of such a population are presented at Fig 2A-D for 2 tasks and Fig 3A-F for 3 tasks. For the artificial selection simulation results presented in Fig. 5A and S17B, we took as an initial population the 100 offsprings that were closest to archetype 1. For the results presented at Fig. 5B,D and S17D, we took the entire

population. For the results presented at Fig 5A, we took a target that is distanced from the starting population equally along the front and perpendicular to the front. For the results presented at 5D and S17B, we took 2 targets: one distanced parallel to the front and one distanced perpendicular to the front. At each artificial selection generation, we selected the fraction p of individuals closest to the target, and then recombined the population (without mutations) to generate a number of offsprings that equals the initial population size. For simulations using triangular front we used p=0.1, while for 2-dimensional fronts we used p=0.5. We repeated this process for a number of generations as specified in the relevant captions.

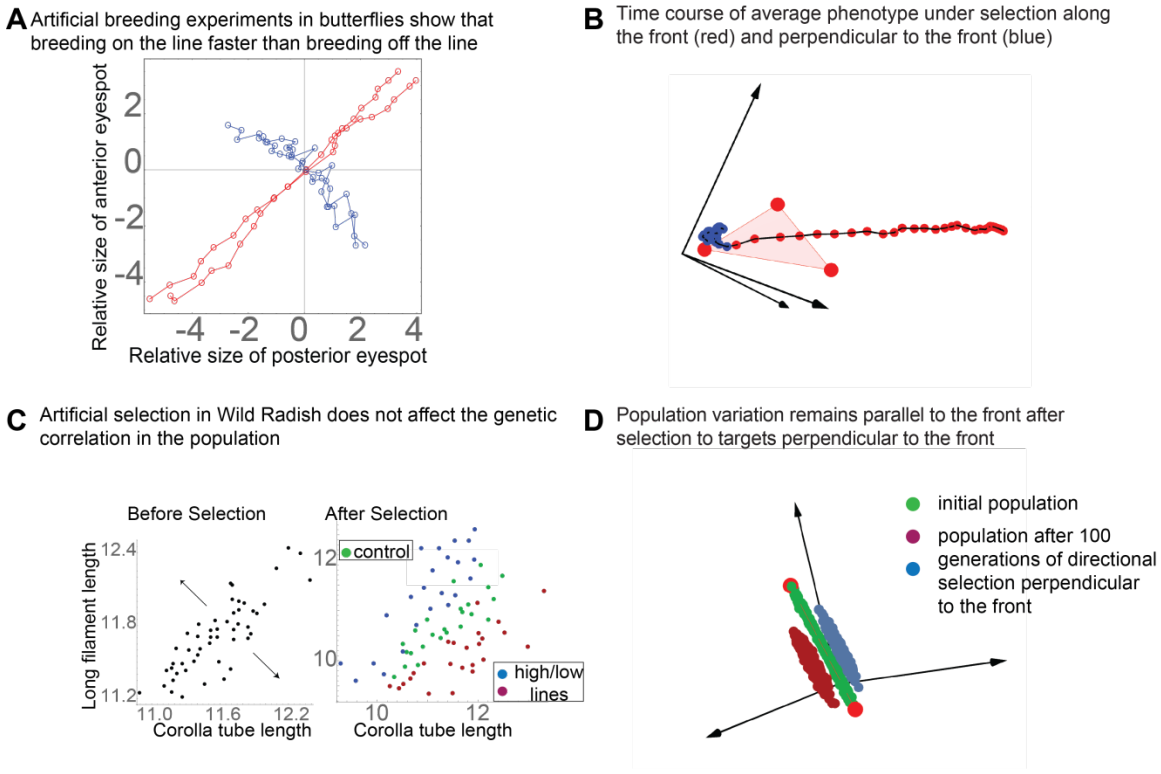

Figure S17. The `response of a simulated population adapted to the Pareto front to artificial selection. (A) Artificial selection experiments performed by [10] on butterfly eyespot-size show that response to selection along the main axis of phenotypic variation of the natural population is faster and larger than response to selection perpendicular to this axis. The x and y axes represent the normalized sizes of the two dorsal eye-spots, and the mean phenotype of the initial population is at the origin. Adapted from Allen et al. 2008 [10]. **(B)** Time course of mean phenotype in simulations under selection towards a far target along front (red) and perpendicular

to the front (blue). Each point represents the phenotype mean at subsequent generations of selection. Initial population was taken as the 10% closest phenotypes to archetype 1, out of a population adapted to the front for 100,000 generations in a simulation run with parameters $N = 1000, \mu = 0.05, k = 5$. Each generation, 1000 offspring were produced and the 10% closest to the target were selected. **(C)** left panel: variation of corolla tube length and long filament length in Wild Radish. Each point is the mean of all offsprings of one sire. Right panel: results of artificial selection for 5-6 generations in the directions represented by the arrows in the left panel. Adapted by Agrawal et. al [11] from Conner 2003 [12] **(D)** A population (green) adapted to the triangular Pareto front as in (B) was selected to far targets perpendicular to the front. The final populations (blue, red) after 100 generations of selection have variation that is on a plane parallel to the original Pareto front.

Supplementary Section 16. Details on the construction of Figure 5F

We used the data presented in Fig. 5B-C in Mallarino et al. [13]. From Fig. 5C we obtained beak length, width and depth relative to *G.difficilis*. Fig 5B shows for each of BMP, calmodulin, TGFβIIr, β-catenin and Dkk3 if it had a positive (+), negative(-), or no effect (0) on one of the beak's dimensions. For polymorphism representation, a "+" was assigned the value "1", a "-" was assigned a value of -1, and 0 was assigned a value of 0.

References

1. Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, Dekel E, Kavanagh K, Alon U. 2012 Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space. *Science* **336**, 1157–1160. (doi:10.1126/science.1217405)
2. Kimura M. 1969 The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations. *Genetics* **61**, 893–903.
3. Watterson GA. 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276. (doi:10.1016/0040-5809(75)90020-9)
4. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006 Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. *Mol. Biol. Evol.* **23**, 1891–1901. (doi:10.1093/molbev/msl051)
5. Eyre-Walker A, Keightley PD. 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618. (doi:10.1038/nrg2146)

6. Kimura M. 1969 The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations. *Genetics* **61**, 893–903.

7. Sheftel H, Shoval O, Mayo A, Alon U. 2013 The geometry of the Pareto front in biological phenotype space. *Ecol. Evol.* **3**, 1471–1483. (doi:10.1002/ece3.528)

8. Gillespie JH. 2004 *Population Genetics: A Concise Guide*. 2nd edition. Baltimore, Md: Johns Hopkins University Press.

9. Friedlander T, Mayo AE, Tlusty T, Alon U. 2013 Mutation rules and the evolution of sparseness and modularity in biological systems. *PloS One* **8**, e70444.

10. Allen CE, Beldade P, Zwaan BJ, Brakefield PM. 2008 Differences in the selection response of serially repeated color pattern characters: Standing variation, development, and evolution. *BMC Evol. Biol.* **8**, 94. (doi:10.1186/1471-2148-8-94)

11. Agrawal AA, Conner JA, Rasmann S. 2010 Tradeoffs and negative correlations in evolutionary ecology. In *Evolution since Darwin: the first 150*, pp. 243–268.

12. Conner JK. 2003 Artificial Selection: A Powerful Tool for Ecologists. *Ecology* **84**, 1650–1660. (doi:10.1890/0012-9658(2003)084[1650:ASAPTF]2.0.CO;2)

13. Mallarino R, Grant PR, Grant BR, Herrel A, Kuo WP, Abzhanov A. 2011 Two developmental modules establish 3D beak-shape variation in Darwin's finches. *Proc. Natl. Acad. Sci.* **108**, 4057–4062. (doi:10.1073/pnas.1011480108)