Corrected 1 June 2012; see below

www.sciencemag.org/cgi/content/full/science.1217405/DC1

Supplementary Materials for

Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space

O. Shoval, H. Sheftel, G. Shinar, Y. Hart, O. Ramote, A. Mayo, E. Dekel, K. Kavanagh, U. Alon*

*To whom correspondence should be addressed. E-mail: uri.alon@weizmann.ac.il

Published 26 April 2012 on *Science* Express
DOI: 10.1126/science.1217405

This PDF file includes:

Correction: Various typos, grammatical errors, and misplaced references to figures have been corrected, and some points made in the main text have been clarified.

Supplementary Materials

Evolutionary Tradeoffs, Pareto Optimality, and the Geometry of Phenotype Space

O. Shoval[1], H. Sheftel[1], G. Shinar[1], Y. Hart[1], O. Ramote[1], A. Mayo[1], E. Dekel[1], K. Kavanagh[2], and U. Alon[1]

[1]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel.
[2]Biology Department, University of Massachusetts Dartmouth, Dartmouth, MA, USA

correspondence to: uri.alon@weizmann.ac.il

This PDF file includes Methods:

Detecting suites of variation shaped as triangles and lines and evaluating statistical significance

Here we present a statistical test of whether the data resembles a triangle, and whether the data falls on a line.

Detection of triangles

This requires a criterion for triangularity, and a null model in order to obtain a p-value. We use data from *(13)* on Darwin's finch body/beak measurements as an example.

Criterion for triangularity

We sought a criterion that fits the subjective notion that the data occupies a region which is triangular in shape. For this purpose, a reasonable criterion is the ratio of the areas of the convex hull of the data and the minimal triangle that encloses the data. First, we calculated the convex hull of the data (Figure S1A,B) using a standard algorithm (quickhull *(22)*). Then, we calculated the minimal area triangle that encloses the convex hull (Figure S1C), using the algorithm described in *(23)*. The ratio between the area of the minimal enclosing triangle and the convex hull is a measure of the triangularity of the data. We denote this ratio by 't-ratio'. A t-ratio of one occurs when the convex hull is perfectly shaped as a triangle. The larger the t-ratio, the less the data is arranged in a triangle. In the Darwin's finch example, depicted in Figure S1, the t-ratio is 1.07.

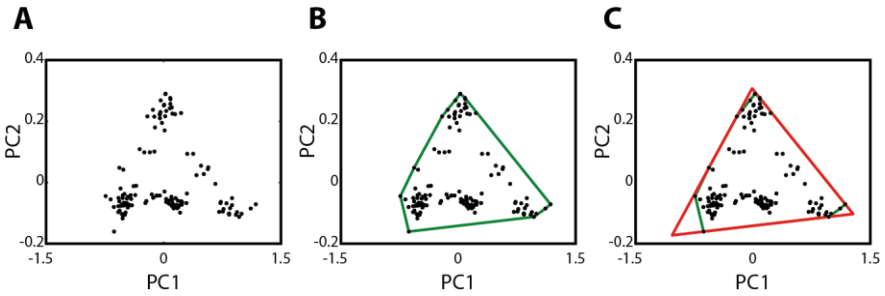

Figure S1. A test for a suite of variation shaped as a triangle on sample data. **A.** The suite of variation. **B.** The convex hull of the data (green). **C.** The minimal area triangle enclosing the convex hull (red).

Generating a randomized dataset

To obtain a significance for the triangularity criterion of a dataset, we compare it to a null model- made of an ensemble of suitably randomize datasets. We chose for this purpose a null model that preserves the statistics of each trait, but that reflects a situation where the traits are independent of each other. The null model thus assumes that the two coordinates of the data (x,y) are independent. We generated a large number (10^4) of randomized datasets as follows: each dataset is comprised of the same number of points N as the original dataset. Each point has an x value drawn from the CDF (cumulative distribution function) of the original data's x values, and a y value drawn from the CDF of the original data's y values. We repeat this process until we have the number of points as in the original data set. In this method the null model's x and y CDFs coincide with the CDFs of the x and y of the original data, but we eliminate the relationship between the x and y values. An example of a randomized dataset produced with this procedure for Darwin's finches dataset is shown in Figure S2.

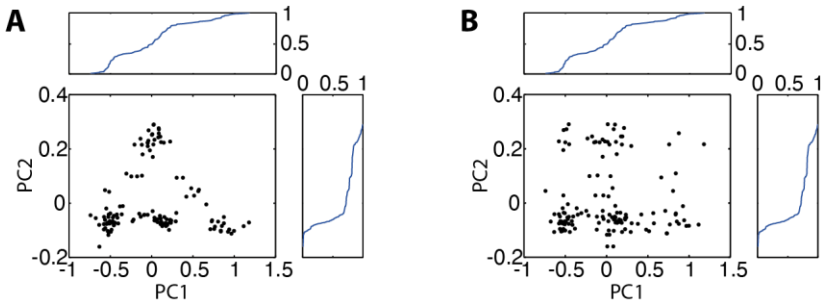

Figure S2. **A.** Cumulative distribution functions (CDF) of x (PC1) and y (PC2) values of Darwin's finch data. **B.** Randomized dataset, with CDFs equal to those of the original dataset.

For each randomized dataset, the t-ratio is computed as described above (Figure S3). In the example depicted in Figure S3, the t-ratio is 1.64, showing that the randomized data is less triangular according to the present criterion than the original dataset (with t-ratio 1.07).

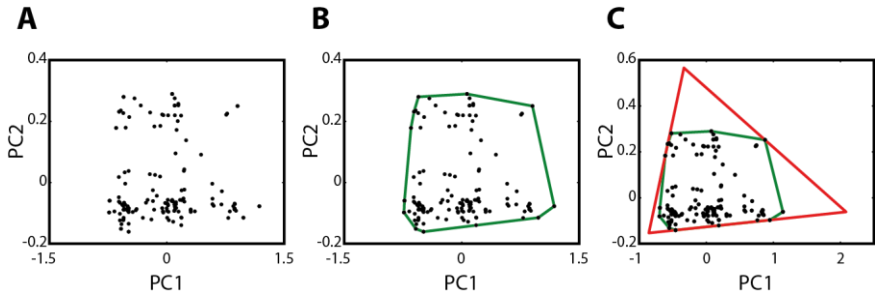

Figure S3. Triangularity test on randomized dataset. **A.** Randomized data based on Darwin's finch data. **B.** Convex hull of the randomized dataset (green). **C.** Minimal area triangle enclosing the dataset (red). Note change of scale in C.

Computing the P-value for the triangularity test

To find the p-value for the triangularity of a dataset, we first compute the t-ratio of the original dataset. We then generate random datasets as described above. For each random dataset we calculate the t-ratio. The resulting p-value is the fraction of datasets for which the t-ratio is smaller than the t-ratio for the original dataset. For the Darwin's finch dataset, the t-ratio is 1.07. The statistics for 10,000 randomized sets is shown in Figure S4. Since for all 10,000 randomized datasets, the t-ratio is higher than the original dataset, the p-value is smaller than 10^{-4}. The z-score (how many standard deviations the original data's t-ratio is from the mean of the 10,000 randomized dataset's t-ratio distribution) is 5.4.

Statistical calculations of this type are dependent on the null hypothesis used: in the present case the null hypothesis is that the traits in the random model are drawn independently from their CDFs. Other null hypotheses and randomization methods can lead to different p-values.

In addition, triangularity can be sensitive to points with extreme values - in the present test, use of the convex hull emphasizes such points. Improvements can be made that are less sensitive to extreme points, e.g. by removing points on the convex hull and repeating the calculation. In judging whether data is triangular for the purpose of linking it to the present theory, one can add to the statistical tests information about the tasks (behaviors) performed by organisms near the predicted vertices. For example, in the bat data below, an extreme data point near the top of the

predicted triangle (archetype 2) increases the significance of the triangle in the present test. Removing that point can increase p-values. However, the fact that all 4 species of bats nearest archetype 2 have the same feeding behavior, which is otherwise very rare in the dataset, adds to the significance of the data in relation to the theory. The other data sets in the study are much less sensitive to removing extreme points. Further work can devise formal statistical tests that employ such behavioral information.

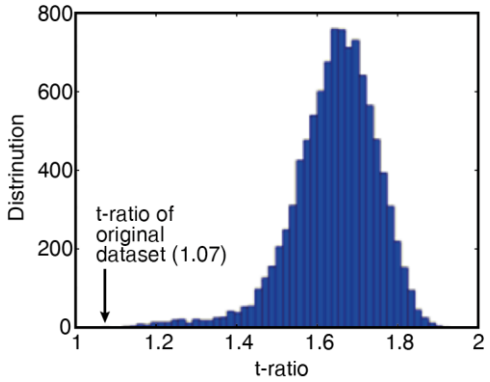

Figure S4. Histogram of t-ratios for 10,000 randomized datasets. Each randomized dataset is based on the cumulative distribution functions of Darwin's finch data, as described in the text. The t-ratio of the original dataset is 1.07, which is lower than the values for all randomized datasets, leading to a p-value $< 10^{-4}$, and a z-score of 5.4.

Software for analyzing data in terms of triangular Pareto fronts and their significance

Software for the analysis described above for triangular suites of variation is available online *(14)*. The software runs as a command line executable on a dataset saved as a comma delimited text file. The software returns the p-value and suggests preliminary archetype trait values, given by the vertices of the minimal area triangle. See user manual on the website *(14)*.

Statistical test for lines

We describe a statistical test of whether a dataset in two dimensions is well described by a line. Principal component analysis (PCA) is used to measure the 'linearity' of the data: PCA returns the variance of the data along the first and second principal components – the ratio between these two variances is a measure of the correlation of the data. The null model used to obtain a p-value statistic, is the same as for triangles above - that each trait is drawn from its distribution independently of the other trait. A randomized dataset is formed using the cumulative distribution functions of trait 1 and 2 (Figure S5). For 10,000 randomized datasets the ratio of major and minor axis is computed, and compared to that of the measured data, resulting in a p-value. Other standard methods for linear regression provide similar results.

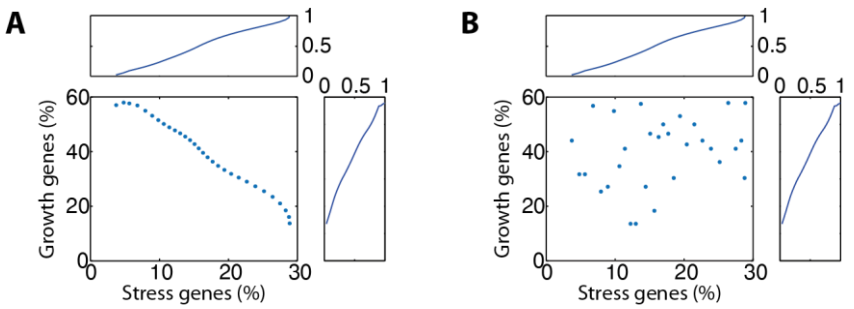

Figure S5. Detection of line shaped Pareto fronts. Sample data of *E. coli* growth vs. stress genes for the no glucose condition (see main text and below). **A.** Cummulative distribution functions (CDF) of trait 1 (stress genes) and trait 2 (growth genes). **B.** Randomized data set, with CDFs equal to those of the original dataset.

Analysis of triangular suites of variation presented in the main text

Darwin's finches (*Geospiza*)

Darwin's finches present one of the hallmarks for the study of evolution. The different species occupy niches that differ mostly in diet. Grant et al. *(13)* used five different measured parameters for multiple individuals in each of the different Darwin's finch species. Parameters measured are: wing length, tarsus length, bill length, bill depth, and bill width. Here we review their results regarding the six species of ground finches (see Table S1 for name abbreviations). Each data point is an average of a certain species in a given island. Data from *(13)*, tables A1-A5.

Species name	Abbreviation	English name
Geospiza magnirostris	MAG	Large ground finch
Geospiza fortis	FORT	Medium ground finch
Geospiza fuliginosa	FUL	Small ground finch
Geospiza dificilis	DIFF	Sharp beaked ground finch
Geospiza scandens	SCAN	Cactus finch
Geospiza conirostris	CON	Large cactus finch

Table S1. Abbreviations and English names of Darwin's finches species

Principal component analysis of the log of these five measurements for the ground finches shows that two dimensions explain almost all of the variance *(13)*. The 1st component includes 92.6% of the variance, and the 2nd component includes an additional 6.4%, summing to a total of 99%. The 1st principal component is mostly comprised of total size of the bird and the 2nd component mostly corresponds to the shape of the beak. A high value in component 2 describes a short and wide beak, while a low value is long narrow beak. This is thus an example of a five dimensional measurement dataset that collapses to an excellent approximation onto a two dimensional plane. According to the present theory, since a plane is determined by three points, one expects in the simplest case three archetypes and a triangular suite of variation. Indeed, the suite of variation is triangular with t-ratio 1.07, p-value $< 10^{-4}$, and z-score of 5.4 (Figure S6A). The species near vertices - large ground finch, small ground finch and cactus finch - suggest the nature of the three archetypes, as detailed below. (Figure S6B).

We note that the present theory may also offer an explanation for the linear scaling of finch beak shapes *(24)* as linear blends of archetypal shapes optimal for distinct tasks.

Archetypes

1. Large ground finch (*Geospiza magnirostris*). This finch species mostly feeds on larger and harder seeds than other finches *(25)*.
2. Small ground finch (*Geospiza fuliginosa*). This finch feeds mostly on small seeds *(25)*.

3. Cactus finch (*Geospiza scandens*). The cactus finch's diet depends on location and season. When available these finches feed on the pollen and nectar of Opuntia cacti. In other times *G. scandens* feed mostly on medium sized seeds, berries and worms *(25, 26)*.

The theory also predicts that intermediate species in the triangle have niches (diets) intermediate between the archetype niches. This is supported by observation:

Intermediate species

- Medium ground finch (*Geospiza fortis*). This finch is a generalist with a mixed diet, able to survive on a broad range of seed types. This generalist nature has been suggested to enable *G. fortis* to quickly adapt to changing conditions, such as droughts, affecting the available food sources *(27)*.
- Large cactus finch (*Geospiza conirostris*). The diet of the large cactus finch is comprised mostly of cacti nectar and large seeds *(25, 28)*, which fits a generalist approach with diet between that of the large ground finch and the cactus finch.
- Sharp beaked ground finch (*Geospiza difficilis*). With a long pointed beak, *G.* difficilis displays a generalist diet between *G. fuliginosa* and *G. scandens*, depending on competition from these two other finch species and available diet *(29)*. In addition, in certain cases, *G. difficilis* also feeds on the blood of large seabirds.

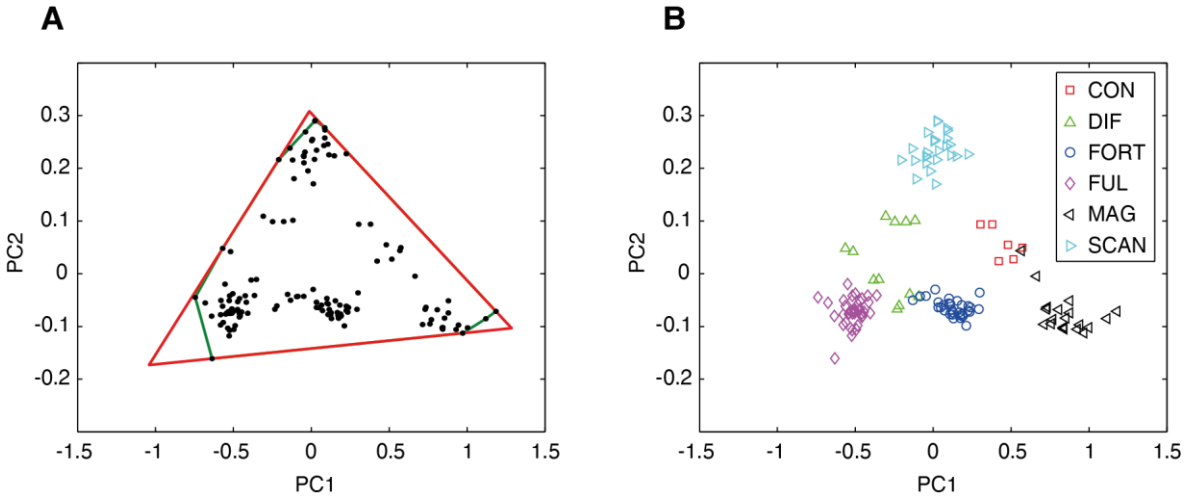

Figure S6. **A.** Darwin's finches suite of variation. Green depicts convex hull, and red is the minimal enclosing triangle (data from *(13)*, tables A1-A5). **B.** Data points colored by species.

Leaf-Cutter ants (*Atta sexdens*)

Atta sexdens present a complex caste system. E.O. Wilson identified a total of 29 different tasks performed by four physical castes, that can be further subdivides to more than seven castes *(15, 30)*. Measurements of poison sac length (normalized to pronotal width) vs. head width (Fig. 3B in the main text) present a suite of variation that resembles a triangle (t-ratio 1.09, p-value<10^{-4} and z-score 3.7, Figure S7). Note that the poison sac is the gland that secretes the recruitment trail pheromone, methyl-4-methylpyrrole-2-carboxylate.

Wilson states that *"disproportionate variation in the size of organs as a function of total body size reflects the function of the organs and can thus be used as a first clue concerning the roles*

of previously unstudied or experimentally intractable organs". We believe that the Pareto front approach may provide a means of providing these clues. The suite of variation suggests three archetypes: a small ant with a relatively small poison sac, a medium sized ant with a relatively large poison sac, and a large ant with a relatively medium sized poison sac.

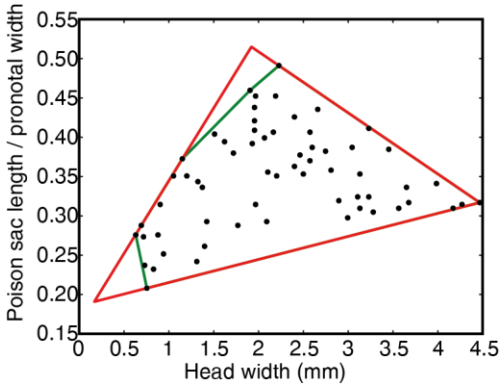

Figure S7. Leaf-Cutter ants suite of variation. Green depicts convex hull, and red is the enclosing triangle.

Archetypes

1. Gardening/nursing. The small ants mostly perform tasks within the nest related to gardening and nursing (Figure S8). It is reasonable that these ants do not require a large poison sac, as they do not leave the nest, and thus do not need to mark their trail.
2. Foragers/excavators. These ants display a relatively large poison sac. Other medium sized ants are within-nest generalists (according to Figure S8).
3. Soldier. Large ants are mostly soldier ants. Since soldier ants also need to leave the nest, a larger poison sac might be needed. Note that the poison sac of the two archetypes, foragers/excavators and soldiers, can be of equal absolute size, because the reported value is normalized to pronotal width.

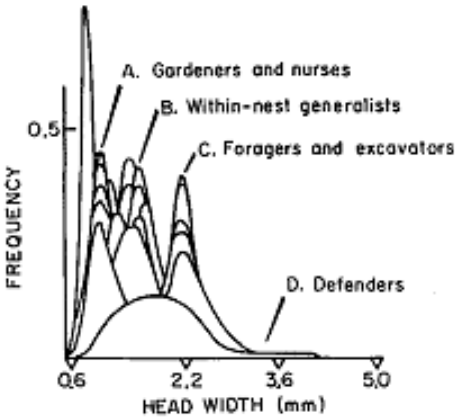

Figure S8. Superimposed polyethism curves of partially discretized caste system at different head width values (reproduced from *(15)*)

Intermediate phenotypes

- Ants with a medium body size and a small poison sac. These are probably the within-nest generalists (Fig S8).

Microbats (Michrochiroptera)

Norberg and Rayner *(16)* gathered data on 215 bat species of 16 families from various sources. They considered each species' mass, food type, and seven parameters describing the wings – width, area, aspect ratio, wing loading, tip length ratio, tip area ratio and tip shape index.

We focus our discussion on the biggest group of bats, the microbats. The rest of the species, megabats (*Pteropodidae*) are a separate sub-order, which are considered to have important differences from microbats (e.g. thought not to use echolocation, and feed mostly on fruit and not insects). Viewing the wing aspect ratio vs. the body mass for a diverse set of microbat species (total of 108 species) suggests a triangular suite of variation (t-ratio 1.13, p-value $< 3 \cdot 10^{-2}$,and z-score 1.34, Figure S9). Analysis suggests three archetypes:

1. Low mass, low wing aspect ratio. These wing proportions are suited for maneuverable flight, and the size suggest feeding on small prey. Species near this corner include *T. discifera, T. tricolor, P. mimus, C. tkonglongyai*. These bat species catch small flying insects at low vegetation level (flight within clutter) *(31–33)*. Some species also glean insects from plant surfaces.
2. Medium mass, high wing aspect ratio. These wing proportions are suited for a more efficient flight with lower energy consumption, but resulting also in lower maneuverability. Four different mollossidae species that are located near this archetype (such as *T. fulminans, T. plicata, M. ater, and T. australis*) hawk for high flying insects *(16)*. Thus we can conclude that the "task" of this archetype is hawking high flying insects (flight without clutter), which requires longer times in flight.
3. High mass, low aspect ratio. These bats feed on large prey, and their lower aspect ratio allows maneuverability for flight in clutter. Species near this archetype (such as *V. spectrum, M. gigas, and C. auritus*) are carnivores, feeding on large prey such as lizards, smaller bats, and frogs *(34, 35)*. In addition these species prey on large insects or fruit. Their hunting strategy seems to be "sit and wait" within vegetation, thus not requiring long flight times.

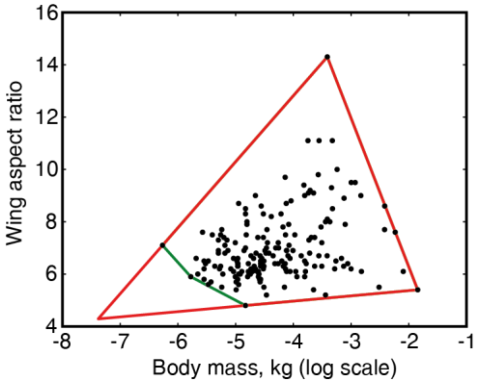

Figure S9. Microbats suite of variation. Green depicts convex hull, and red is the enclosing triangle (most edges of the convex hull coincide with the edges of the triangle).

Analysis of *E. coli* gene expression data of Zaslaver et al. *(18)*

Zaslaver et al. *(18)* measured *E. coli* promoter activity at a high-temporal resolution and accuracy. They used a comprehensive *E. coli* reporter strain library *(36)* in which a full intragenic region controls a green fluorescent protein (rapidly folding gfp), for 1600 different promoters. A robotic assay was used to measure the promoter activity of these strains under different medium conditions. Promoter activity was continuously tracked as cells changed their growth rate from exponential to stationary phase in different media. The promoter activity at each timepoint was calculated by the rate of fluorescence accumulation divided by the number of cells, dGFP/dt/OD, where GFP is gfp fluorescence, and OD is optical cell density (data downloaded from *(18)*).

We analyzed this dataset in order to test the hypothesis that there exist tradeoffs in gene expression data that can result in low-dimensional Pareto fronts that connect archetypes. We considered genes that are expressed in all four conditions, denoted by Zaslaver et al as no glucose, no amino acids (AA), nitrogen limited and ethanol (total of 1199 genes). In order to find the relative instantaneous allocation of transcription resources to each promoter we normalized the promoter activity to the sum of all promoter activities at each time point. We henceforth discuss this normalized promoter activity (normalized to the summed activity of all 1199 genes at each timepoint).

We tested the 200 promoters with most variability across time, which sum up to 96% of the total variability in the dataset. We find that the dynamics of these genes across the four conditions falls into two main clusters (Figure S10), as revealed by ordering the pairwise correlation matrix using a standard hierarchical clustering analysis (Matlab2010b clustergram function). The two clusters correspond to growth genes (ribosomes, transcription and translation, cluster of genes 1-96), and stress/survival genes (oxidative stress, acid stress, iron limitation, etc., cluster of genes 97-178). The correlation matrix reveals that the two clusters are negatively correlated, suggesting a tradeoff, where expression of one of the clusters comes at the expense of the other. In addition genes 179-200 are not correlated with either group.

We next summed the normalized promoter activities of the genes in each cluster, to get the total allocation of promoter activity to the growth gene cluster and to the stress gene a cluster for each time point in each condition (Figure S10A insets). We find that the promoter activity of these genes varies over time in a trajectory that falls approximately on a straight line (Figure S10B, $p<10^4$ using method described above), and that this line is very similar between the 4 different growth conditions (rms=6%). At one end of the line is the expression program at exponential phase (circles), which occurs when the cells are transferred into fresh medium. Here, most of the expression is devoted to growth genes, and not stress/survival genes. The expression program gradually moves along the line as cells slow growth. It ends at an expression program that devotes promoter activity primarily to stress/survival genes, as cells enter stationary phase (squares). Thus, the instantaneous allocation of promoter activity at each time point is, to a good approximation, a weighted sum of two archetypal expression programs: growth and survival.

The present analysis might also apply to additional studies showing low dimensionality of complex gene expression patterns *(37)*. It may in principle also apply to understanding low dimensional patterns in multi-dimensional data on proteins (e.g. Rubisco which has a linear correlation of parameters across species *(38)*), protein circuits *(4)*, cells and organs, in terms of their multiple tasks.

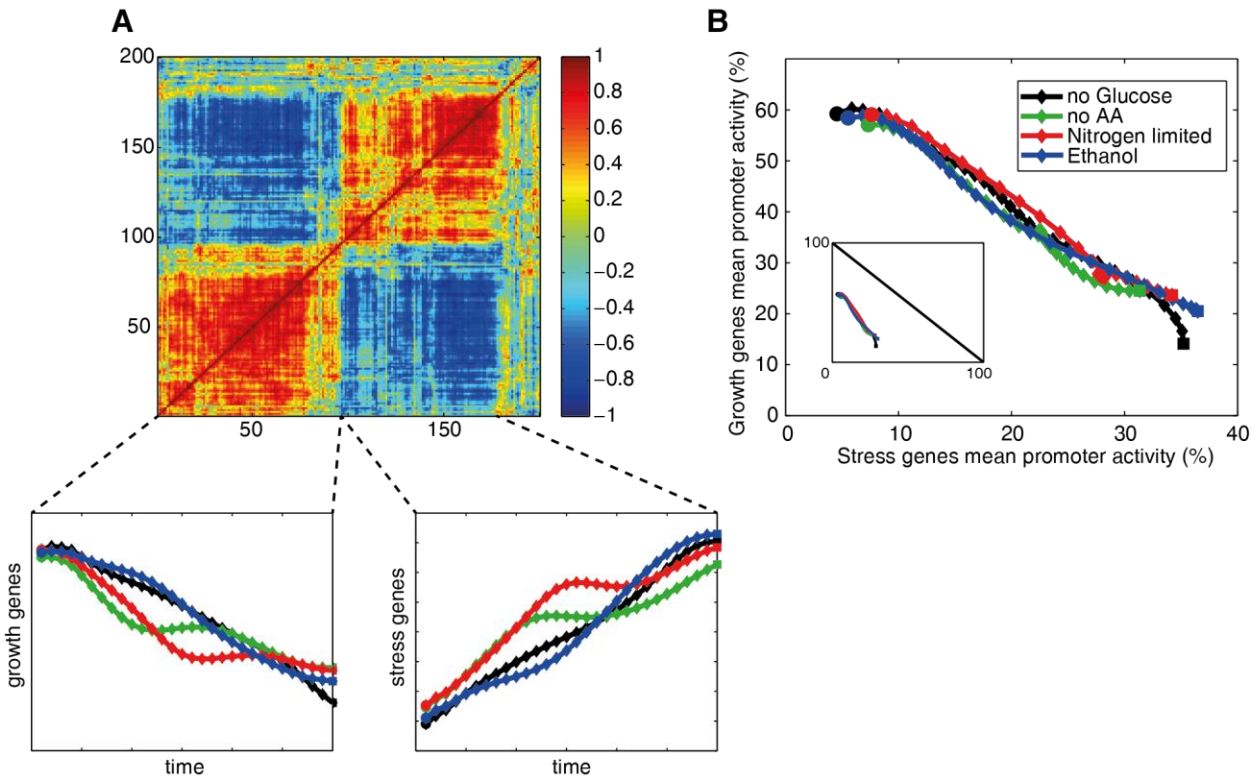

Figure S10. Clustered correlation matrix of 200 most varying genes, across the four growth conditions. **A.** The two clusters of growth and stress genes are anti-correlated. Insets: total normalized (% of total promoter activity) promoter activity of the growth and stress genes as a function of time, for the four conditions (see legend in B.). **B.** Sum of promoter activity of growth gene cluster (% of total promoter activity) vs. stress gene cluster (% of total promoter activity). Circles – time zero, when cells start to grow, squares – end point, when cells reach stationary phase. Inset: Same data with axes reaching 100%. Diagonal line represents 100% total of the 1199 promoter activities.

Relaxing assumption (i) - archetypes that are not isolated points

Up to now we considered cases in which performance functions were maximized by a single isolated phenotype, the archetype. Here we consider cases where performance is maximized by a set of phenotypes (including cases of many-to-one mappings *(39)*, and performance ridges. We consider the case where an archetype includes a continuous set of trait values, i.e. a region in morphospace, and also cases where several non-continuous trait values have archetypal performance.

Performance maximized by a continuous set of phenotypes

Figure S11A depicts a case when the archetype for task 1 is a single point (v_1*), and the archetype for task 2 is a region in phenotype space (v_2*). Here we show that, for performance functions that decrease with Euclidean distance from the archetypes, the Pareto front is the line between v_1* and the closest point on v_2*.

Consider the closest point on v_2* to v_1*, denoted by v_2** (Figure S11B). Consider next a circle centered at v_1* that goes through v_2**. This circle is a contour of equal performance in task 1, with radius C equal to the distance between v_1* and v_2**. We now consider two cases; one comparing a phenotype outside of circle C, and one comparing a phenotype inside circle C.

Phenotype outside C: Consider point s_o which is outside circle C (Figure S11B). It has a larger distance to v_1* than v_2**, since it is outside C. This results in s_o having a lower performance in task 1 than v_2**. It also has lower performance in task 2 than v_2**, because v_2** is in v_2* and thus has maximal performance. Thus v_2** dominates s_o in both tasks. Since this applies for any s_o outside of circle C, all such points are dominated by v_2**.

Phenotype within C: We now consider a point inside circle C, denoted by s_i (Figure S11C). The performance of s_i in task 1 is determined by its distance from v_1*, which we denote by d_1^s: $d_1^s \equiv ||s_i - v_1^*||$. The performance at task 2 is determined by the distance from s_i to v_2***, the closest point on v_2*, which we denote by x: $x \equiv ||s_i - v_2^{***}||$. Note that generally v_2*** does not coincide with v_2**. Due to triangle geometry, the distance from v_1* to v_2***, denoted by y: $y \equiv ||v_1^* - v_2^{***}||$ is shorter than the sum of d_1s and x: $y < d_1^s + x$.

We now examine point A on the line connecting v_1* and v_2**, which has distance x from v_2**. The distance of A from v_1* is d_1^A: $d_1^A \equiv ||v_1^* - A||$. Since v_2*** is outside circle C, its distance to v_1* is longer the distance from v_2** to v_1*: $d_1^A + x < y$. Thus we get that: $d_1^A + x < y < d_1^s + x$, resulting in

$$d_1^A < d_1^s$$

Point A's performance at task 2 is equal to point s_i, as both points are at distance x from v_2*. However A dominates S_i in task 1. Since s_i was an arbitrary point within circle C, we conclude that for every point S_i within circle C, there is a corresponding point on the line connecting v_1* and v_2** that dominates it.

Summary: We find that for any point s_o outside circle C or s_i within C, there is a point on the line connecting v_1* and v_2** that is dominating (i.e. has higher performance in one task, and higher or equal performance in the other). The Pareto front is therefore the shortest line connecting v_1* and v_2* (Figure S11D).

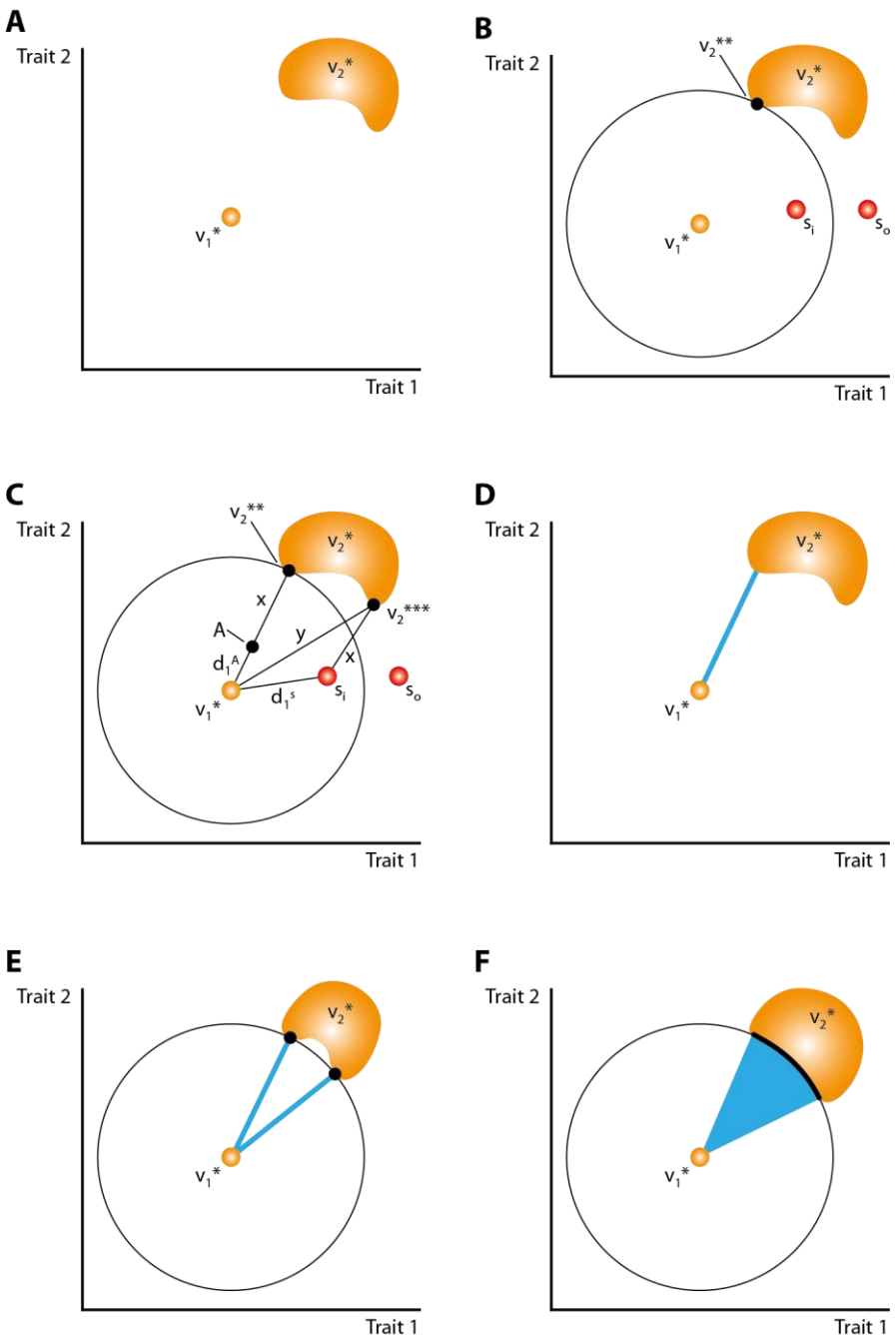

Figure S11. Pareto fronts when an archetype is a region rather than a single point. **A.** The archetype for task 2 is a set of points (region), and the archetype for task 1 is a single phenotype v_1*. **B.** Point s_o is outside the equal performance (in task 1) circle C. Circle C is centered at v_1* with radius equal to the distance between v_1* and the closest point on v_2* (denoted by v_2**). Point s_i is inside circle C. **C.** Geometric constructions used to show that for every point s_i there is a dominating point on the shortest line between v_1* and v_2*. **D.** The Pareto front is the line between v_1* and the closest point on v_2*. **E.** Two points v_2** on v_2* have equal minimal distance to v_1*. In this case the Pareto front includes the two lines from both v_2** points to v_1*. **F.** v_2** might be a continuous set of

points. Note that if v_1* is a point, then v_2** must be arc shaped. The Pareto front in this case is the area enclosed between v_2** and v_1*.

Note that v_2** might not be a single point, i.e. there might be several points on v_2* with the same minimal distance to v_1*. In such cases the Pareto front will be the collection of lines from all points v_2** to v_1* (Figure S11E). Furthermore, v_2** might be a continuous set of points, resulting in a Pareto front including a continuous area (Figure S11F).

It is intriguing that multiple archetypes seem to resolve the otherwise symmetric and neutral relation between points that all maximize a given performance. If that task is the only one under selection, there is no way to choose between the different points, and a population may drift along the maximizing region. The presence of a second task 'chooses' a point in the region that is closest to the other archetype.

Performance maximized by several disjoint regions in morphospace

Peak performance at a task can conceivably be achieved by more than one connected region in morphospace *(39)*. Figure S12A depicts a case where there is a single phenotype for archetype 1, and two sets of phenotypes for archetype 2. In a similar derivation to the previous section, one has that the Pareto front is included in the line from v_1* to the closest point on the closest set of traits for v_2* (Figure S12B).

However, if a phenotype that is included in v_{2b}* slowly evolves and adapts to an intermediate habitat, it generally might need to undergo a significant transformation to reach the Pareto front. Therefore, we anticipate that in such cases, species that are evolving through a changing habitat from v_{2b}* to v_1* will follow the straight line from v_{2b}* to v_1* (Figure S12C). This might be called the 'local' Pareto front, whereas the 'global' Pareto front is the straight line from v_1* to v_{2a}*.

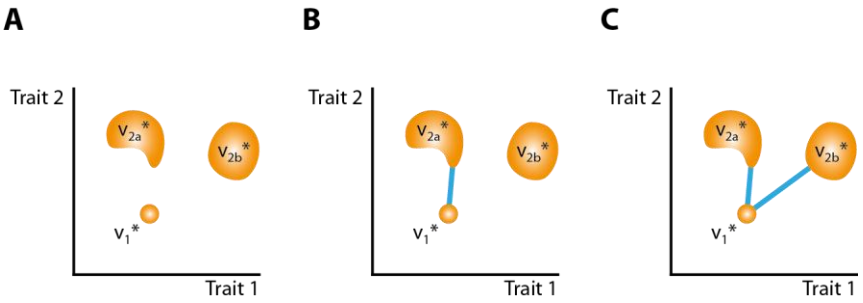

Figure S12. Pareto front in the case of disjoint regions which all optimize performance at a task **A.** Task 2 has two archetype sets – v_{2a}* and v_{2b}*. **B.** The Pareto front is the line from v_1* to the closest archetype (in this example – v_{1a}*). **C.** When evolving from v_{2b}* towards v_1*, a phenotype may follow the Pareto front that does not take into account v_{2a}*.

Relaxing assumption (ii) - non equal performance metrics

For equal distance metrics, the Pareto front is the straight line segment between the archetypes

In this section, we present a graphical method for finding the Pareto front in the case of two traits and two tasks. We use this below to address the effect of different distance metrics on the Pareto front.

For ease of presentation, we begin with Euclidean metrics. Assume that the performance at task 1 is some decreasing function of the Euclidean distance of the phenotype from archetype 1, and similarly for task 2 (Figure S13A,B). In order to find the Pareto front we start from archetype 1 and find the equal-performance contours - all phenotypes with the same performance level p_1. Since performance decreases with Euclidean distance from archetype, the contour is a circle with radius d_1 centered at A_1 (Figure S13C). The point on this circle that is closest to archetype 2 will have highest performance at task 2, and therefore will be part of the Pareto front. This closest point to archetype 2 can be found by the tangent to a circle centered at A2 (Figure S13C). This process can be repeated for every performance value p_1, and thus the Pareto front is the collection of all tangents (Figure S13D). Note that this process is symmetric between A_1 and A_2.

Additionally, if A_2 is within the circle defining performance value p_1, then the point A_2 will have better performance in task 1, as it is closer to A_1, and thus will dominate any point on the circle of performance p_1. Therefore the Pareto front is limited to the line segment between A_1 and A_2.

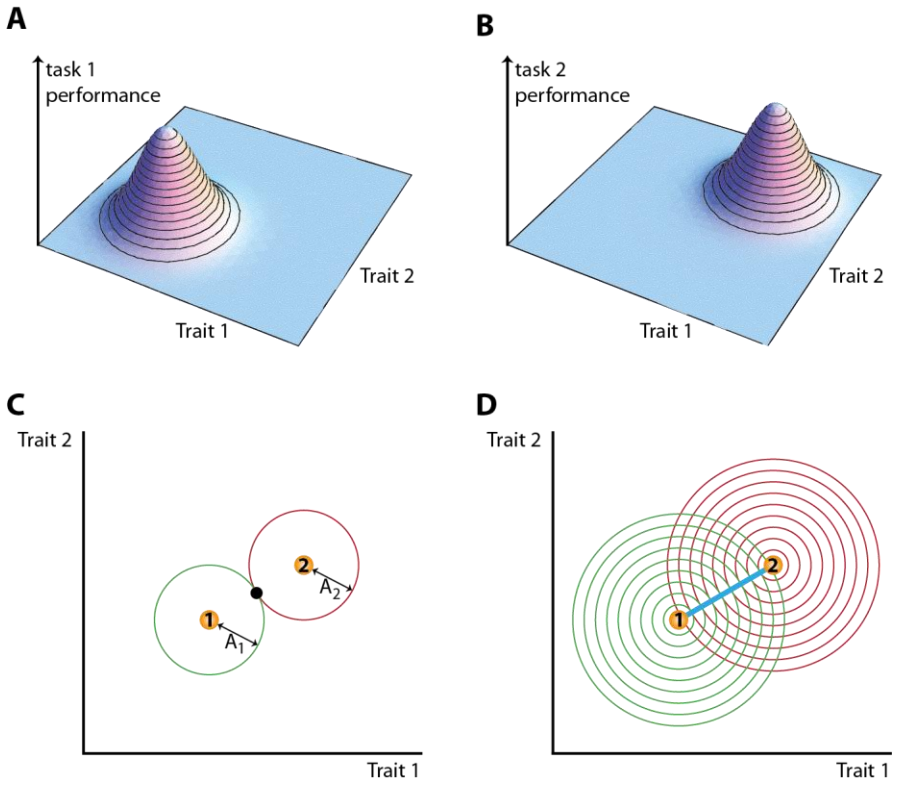

Figure S13. Performance functions decrease with Euclidean distance from the archetypes (phenotypes which maximize performance) for **A.** Task 1, and **B.** Task 2. X and Y axis denote two traits. Performance is z-axis value. **C.** At a distance d_1 from archetype 1 (A_1),

the Pareto optimal point (black) is the tangent point to a circle centered at archetype 2 (A_2). **D.** The Pareto front is the collection of tangents between the performance function contours (blue line).

A similar situation exists when the contours of the performance functions are ellipses rather than circles. This case corresponds to inner-product distance metrics of the form

$$d(v,u)^2 = (v-u)^T M (v-u)$$

Biologically, this means that some traits and trait combinations impact performance more than others. If both performance functions decreases with such generalized distance, with the same matrix M, the Pareto front is also a straight line (Figure S14). In the following section we discuss the case where the distance matrix M is not equal for both performance functions.

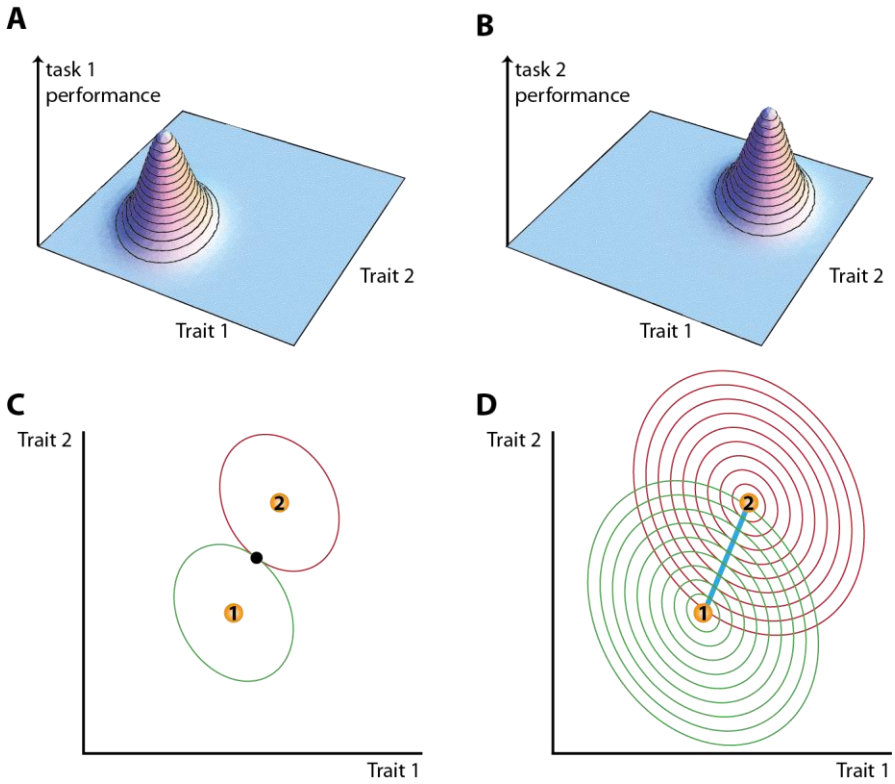

Figure S14. Performance functions with elliptical contours, in which performance decreases with equal distance metrics defined by the same matrix M. **A.** for task 1 and **B.** task 2. **C.** The tangent point between equal-performance ellipses of both tasks is on the Pareto front (black point). **D.** The Pareto front is the collection of tangents between the performance functions (blue line).

<u>Non-equal distance metrics can result in curved Pareto fronts</u>

The performance function might depend in more general ways on the distance from the archetype. For example, some traits may affect performance for one task more strongly than performance for another task. Figure S15 depicts such an example where the distance metric for the two tasks has elliptical contours, but the ellipses are not the same for both tasks (Figure S15A,B). Similarly to the previous derivation, we start from ellipse contour in task 1, and find

the contour of task 2 tangent to it (Figure S15C). The tangent point is on the Pareto front, since any other point has either lower performance in task 1, or lower performance in task 2 or in both.

The collection of tangents for different contours comprises the Pareto front. For two non-equal elliptical distance metrics, the Pareto front is a curved line (Figure S15D).

With higher numbers of tasks, and such non-equal metrics, one expects volumes bounded by curved surfaces. A distinct feature that remains is that the Pareto front has sharp vertices at the archetypes (except in rare, non-generic cases).

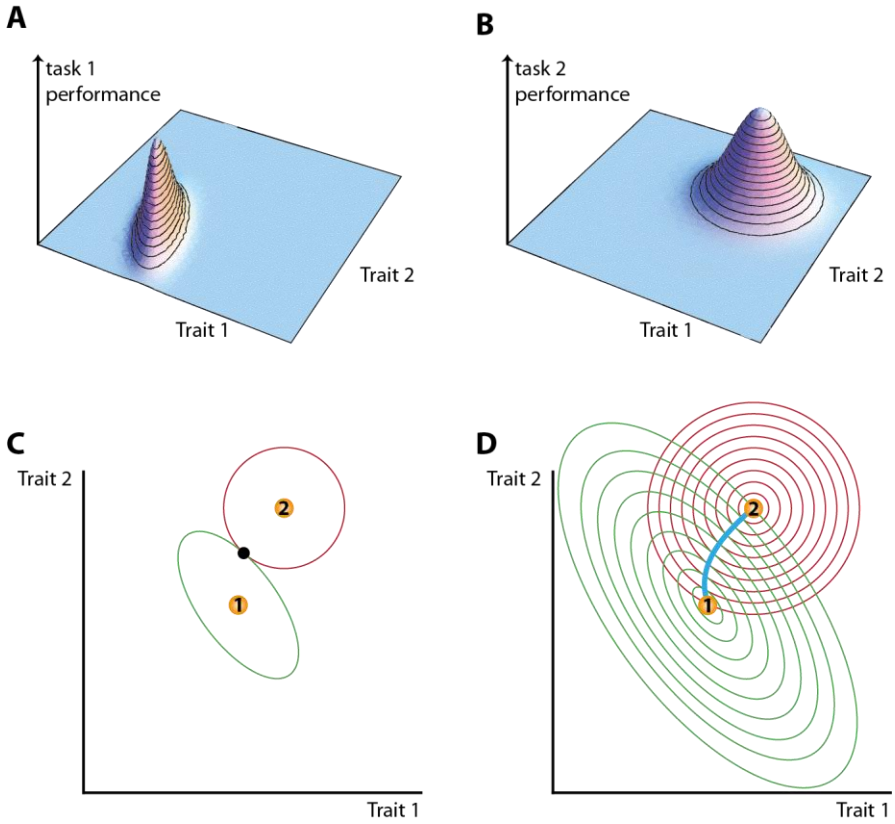

Figure S15. The Pareto front with non-equal distance metrics. **A.** Performance at task 1 depends on elliptical metrics. **B.** Performance at task 2 depends on Euclidean metric. **C.** The tangent point between two equal-performance contours is on the Pareto front (black point). **D.** The Pareto front (blue line) is the collection of tangents of the performance contours of the two tasks.

We now calculate the deviation of the curved Pareto front from a straight line. Each ellipse-shaped contour is characterized by two parameters: a tilt angle θ, and eccentricity (axis ratio) λ (Figure S16A). For a given pair of norms (each norm is defined by a pair (θ,λ)) we calculated the relative rms deviation of the curved Pareto front from the straight line segment between the archetypes (Figure S16B). Deviations from a straight line are sizable only when the ellipses incline towards each other at an angle of 45^{o} relative to the line, and both have large eccentricity (Figure S16C). Large eccentricity results when some trait combinations affect performance much more strongly than other trait combinations.

We calculated the Pareto front for 100,000 randomly chosen pairs of norms (each with a random angle theta between 0^o and 360^o and λ (ratio of major to minor axis) uniformly sampled in the range [0.1,1] - ranging from circles to very eccentric ellipses). The mean relative rms deviation from a straight line is 0.066±0.059(std) (Figure S16D). For a uniform distribution of λ in the range [0.01,1] the rms deviation is 0.084±0.083. A log-uniform distribution in the range [0.1,1] gives an rms of 0.1±0.082 and a log-uniform distribution in the range [0.01,1] results in an rms of 0.213±0.206. In summary, deviations from straight lines in the case of different inner-product norms are expected to be usually mild.

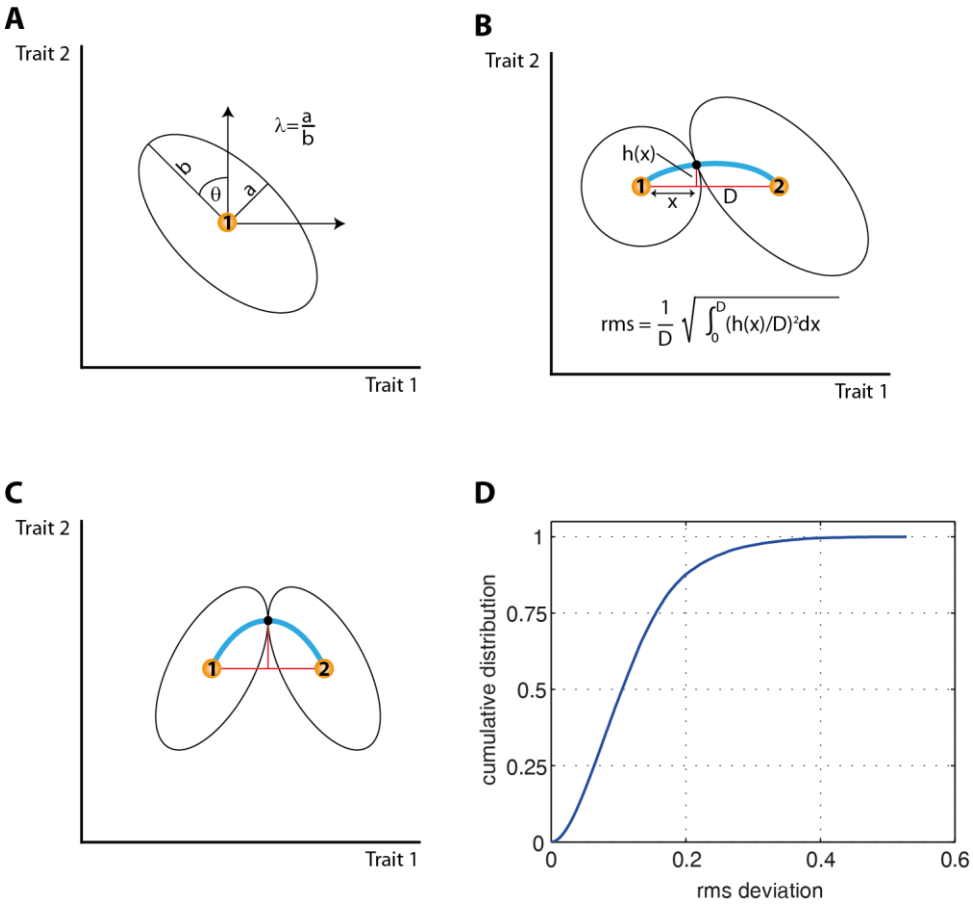

Figure S16. **A.** Each ellipse-shaped contour is characterized by two parameters: a tilt angle θ, and eccentricity (axis ratio) λ. **B.** The rms deviation of the curved Pareto front from the straight line between archetypes. **C.** Example of a case when the Pareto front is highly curved – eccentric metrics tilted towards each-other **D.** Cumulative distribution of rms deviation from the straight line connecting the archetypes (λ uniformly distributed in the range [0.1,1]).

A note on norms and allometry in log space

In the present theory, straight lines occur when performance decreases with a metric- that is, with an inner-product norm distance - from the optimum. To be a norm, a function *n(v)* must satisfy three mathematical conditions: $n(0)=0$, $n(av)=|a|n(v)$ for $a \in \Re$, and the triangle inequality $n(v+v') \leq = n(v)+n(v')$.

In some cases, a norm may be found only in log space. For example, if there are two traits X and Y, and performance decays with

$$(\log X - \log X_0)^2 + (\log Y - \log Y_0)^2$$

a straight line is expected in log space. Note that the decay in this case responds ratiometrically to changes in traits

$$(\log(X/X_0))^2 + (\log(Y/Y_0))^2$$

In contrast, if the performance decays with Euclidean distance such as

$$(X - X_0)^2 + (Y - Y_0)^2$$

A line is expected in linear (not log) space.

An important case in which decrease of performance functions can be approximately described by an inner-product distance metric occurs when the performance functions are differentiable and decrease relatively slowly from their maxima. In such a case, one can use series expansion to describe the performance function. A series expansion to second order results in

$$P(v) \approx P(v^*) + \frac{1}{2}(v - v^*)^T M (v - v^*)$$

where M is the matrix of second derivatives (Hessian matrix), which is negative definite since v^* is the maximum of *P*. The first derivative terms are zero because v^* is a maximum. The second order term is precisely an inner-product norm decrease of performance.

When performance decreases gradually over several orders of magnitude of trait values, expanding in the log of the traits often leads to a greater range over which this series approximation is reasonably accurate, thus resulting in

$$P(\log(v)) \approx P(\log(v^*)) + \frac{1}{2}(\log(v) - \log(v^*))^T M (\log(v) - \log(v^*))$$

To the extent that this expansion applies, straight lines in log coordinates are expected.

Finally, detailed biophysical analyses, such as the work of West, Brown and Enquist *(12)*, provide examples in which logarithmic transformation of trait values naturally emerge. One may be able to use such approaches to directly calculate performance functions.

A note on normalizing the data

Many morphological studies normalize traits by size (or a proxy to size). Normalizing ensures that the data corresponds to differences in ratios (or shapes), rather than to organism size. This highlights the tasks for which the different relative sizes are important. In the example of rodent molars, Kavanagh et al. *(11)* normalized molars by M1 (Figure 1C, main text). Normalizing by a different proxy to size, total molars' size (M1+M2+M3), also yields a straight line (Figure S17A), and conclusions hold. In both normalizations, herbivores are at one end of the line, faunivores at the other, and omnivores in the middle.

Interestingly, when one does not normalize by size, the data still falls approximately on a line (Figure S17B), but it is a very different line. This line is in the three dimensional space whose axes are M1, M2 and M3. The line corresponds primarily to the size of the rodent, one of the most variable features between species. Not normalizing by size highlights tasks for which total body size is important for the tasks: indeed the points do not seem to fall on the line according to diet.

In the paper, we chose in all cases the normalization of the original references *(11, 13, 15, 16)*. Another example for normalizing by size in our study is the measurements of leaf-cutter ants (Fig 3B, main text), in which gland size is normalized to pronotal width, a measure of body size. Also for bat wings, the aspect ratio is dimensionless – ratio of wing area and the square of its length. This normalizes out the size of the bat. Similarly, for bacteria we normalized promoter activity at each timepoint by the sum of all promoter activities.

In two of the examples we consider, one of the axes is explicitly body size (mass) - as in the original data for finches and bats. Indeed, in these cases, (log) body mass seems to be important for the tasks inferred in the study: eating soft (usually smaller) vs. hard (usually larger) seeds, for example, is known to correlate with body mass differences *(25)*.

In summary, the choice of normalizing or not, will affect the tasks that can be inferred from the observed patterns in the data.

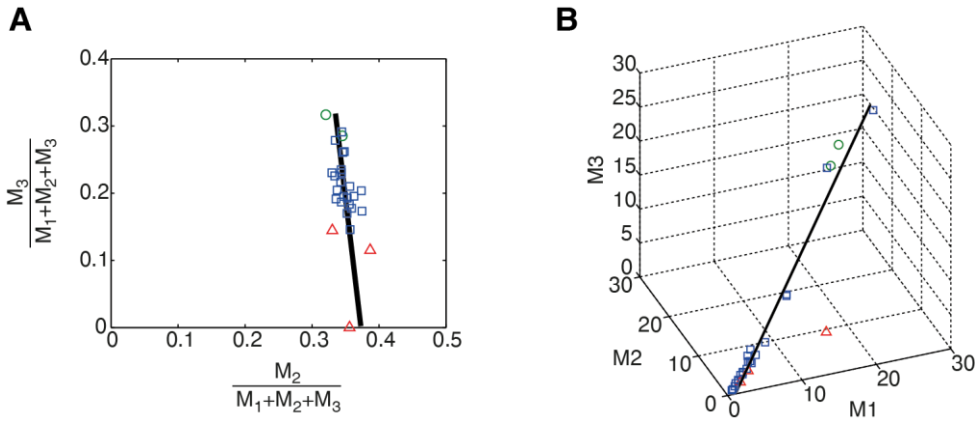

Figure S17. **A.** Rodent molars normalized by sum of molar sizes. Data is spread on a straight line according to diet (circles – herbivores, squares – omnivores, triangles – faunivores). **B.** Unnormalized rodent molar data.

Selected phenotypes in the case of tasks with no tradeoff

Two tasks with no tradeoff result in a single phenotype

Consider a case where task 1 is achieved by a set of traits, and task 2 by a distinct set of traits, such that there is no tradeoff. Each set of traits can achieve its optimal performance, and the resulting phenotype is optimal for both tasks.

For example, the performance at task 1 might be maximized by some value of trait 1, but not depend on the value of trait 2 (Figure S18A), and similarly for task 2 (Figure S18C). For example, bird beak shape might depend on diet, whereas bird plumage might depend on camouflage needs, and the two traits may not face a tradeoff. In this case, a single phenotype can maximize performance in both tasks (Figure S18B). Thus in a habitat which requires performance in both tasks, the unique intersection phenotype will be most adaptive.

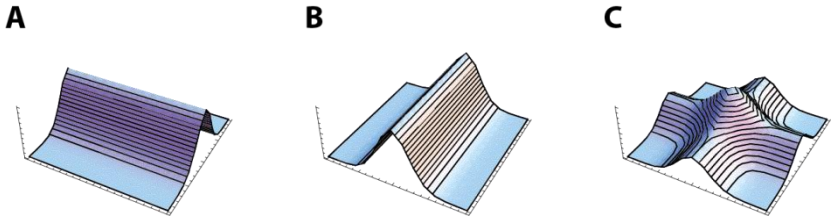

Figure S18. **A.** Performance at task 1 depends only on trait 1. Note no change in performance with changing trait 2. **B.** Same as A, but here performance at task 2 depends only on trait 2. **C.** The two performance functions added.

Tradeoffs within a set of tasks, but not with other tasks

Consider the case in which a given set of traits performs k_1 different tasks, and a separate distinct set of traits performs k_2 tasks. For example, beaks perform tasks such as eating seeds and eating insects, whereas toes in the bird foot perform tasks such as walking and grasping branches. In this case, each set of traits lies on its own Pareto front. The entire phenotype space is an external product of these two fronts.

For example, consider traits x_1 and x_2, which perform two different tasks T_1 and T_2. The Pareto front is a line between the archetypes A_1 and A_2 (Figure S19A). Two other traits, x_3 and x_4, perform distinct tasks S_1 and S_2 and lie on a line between archetypes B_1 and B_2 (Figure S19B). In the four dimensional space (x_1,x_2,x_3,x_4), each phenotype can be characterized by two coordinates: the distance of its (x_1,x_2) traits along the A_1-A_2 line (marked as 'a'), and the distance of the two other traits (x_3,x_4), along the second task line between B_1 and B_2 (marked as 'b'). These two coordinates are independent of each other, since there is no tradeoff between the two sets of tasks (feet tasks and beak tasks).

Consider data on species within the 4-dimensional trait space (x_1,x_2,x_3,x_4). Principle component analysis on this four dimensional dataset, would show that the data lies on a plane, due to the two available degrees of freedom (Figure S19C) – one for feet and one for the beak. Within this plane, we expect to find a quadrangle shaped suite of variation, with vertices that correspond to the four compound archetypes: A_1B_1, A_2B_1, A_1B_2, and A_2B_2. That is, specialists who perform task A_1 or A_2 with the first two traits, and B_1 or B_2 with the second traits, with all combinations

possible. Thus, birds can be found that use beaks for eating seeds and toes for grasping, beaks for seeds and toes for walking, etc.

In general, if a set of traits performs k_1 tasks, and a distinct set of traits performs k_2 tasks, the total degrees of freedom would be the sum of degrees of freedom for each system: $d=d_1+d_2$. The degrees of freedom of each system corresponds to the number of tasks by $d_1=k_1-1$, and $d_2=k_2-1$. Projecting on this space $d=d_1+d_2$, using principal component analysis, one expects the data to fall within a polytope with $k_1 \cdot k_2$ vertices (archetypes), since all combinations of archetypes are possible.

Generally, if traits can be clustered into sets such that within each set traits show tradeoffs with each other, but not with traits in other sets, the resulting Pareto front will be an outer product of Pareto fronts for each set.

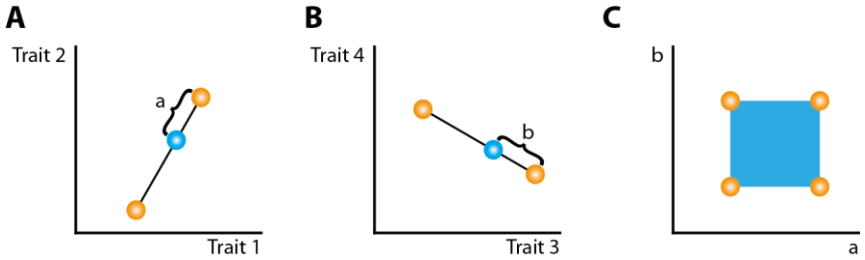

Figure S19. **A.** Trait space of (x1,x2). The two tasks T1 and T2 show a tradeoff, resulting in two distinct archetypes A1 and A2. The resulting Pareto front is the line between the two archetypes, with one degree of freedom, a, the distance of the phenotype on the Pareto front from archetype A_1. **B.** Same as A, for traits (x_3,x_4), tasks S_1 and S_2, and archetypes B_1 and B_2. The phenotype has one degree of freedom, b. **C.** Phenotypes on the Pareto have two degrees of freedom, a and b, and thus fall on a two dimensional plane. The vertices of the suite of variation are all the combinations of archetypes.

Method for detecting which traits may show a tradeoff

In practice, given a high dimensional data set, one can detect which traits may show a tradeoff, and which may not, without additional information. To do this, one can proceed as follows:

1. Test all pairs of traits for correlation, to detect possible lines
2. Test all triplet of traits for 3-way correlations, to detect planes,
3. Repeat for 4-way correlations etc., up to available computational power

next,

1. Cluster traits which correlate with each other. These traits are expected to show tradeoffs, and to fall within Pareto fronts
2. Describe the entire morphospace as an external product of these Pareto fronts.

Normally, however, experience and knowledge of the system can suggest which subsystems (and their traits) are likely not to have tradeoffs (e.g. beaks and toes), and thus may suggest that a separate analysis of the traits for each subsystem is warranted. This seems to be the practice followed in the ecological papers used in the main text.

The Pareto front and weighted averages of the archetypes

Calculating the weights for phenotypes on the Pareto front

We show here that the Pareto front is the weighted average of archetypes, and calculate the weights. The performance functions $P_i(v)$ are assumed to monotonically decrease with an inner-product norm away from their maxima, the archetype v_i^*. Thus $P_i(v) = P_i(d_i)$, where the squared-distance is $d_i = (v - v_i^*)^T M (v - v_i^*)$, and M is a positive-definite matrix (Euclidean distance $d_i = (v - v_i^*)^2$ is when $M=I$). Fitness in habitat h is an increasing function of all of the performances $F_h(v) = F_h(P_1(v),...,P_k(v))$. Examples of such increasing functions include $F = \alpha P_1 + \beta P_2$ and $F = P_1^{\alpha} P_2^{\beta}$ with $\alpha, \beta > 0$.

Phenotypes that maximize fitness obey $\frac{dF_h}{dv} = 0$, which yields, using the chain rule, $\sum_{i=1}^{k} 2 \frac{\partial F_h}{\partial P_i} \frac{\partial P_i}{\partial d_i} M (v - v_i^*) = 0$. The solution is a weighted average of the archetypes $v = \sum_{i=1}^{k} \theta_i v_i^*$ with weights $\theta_i = \frac{\partial F_h}{\partial P_i} \frac{\partial P_i}{\partial d_i} \Big/ \sum_{j=1}^{k} \frac{\partial F_h}{\partial P_j} \frac{\partial P_j}{\partial d_j}$. Note that $\sum_{i=1}^{k} \theta_i = 1$. Furthermore, $\theta_i \geq 0$ because $\frac{\partial F_h}{\partial P_i} \geq 0$ (fitness increases with performance) and $\frac{\partial P_i}{\partial d_i} \leq 0$ (performance decreases with distance from its archetype).

Equivalently, the Pareto front is the convex hull of the archetypes (see below, *(40)*). Mathematically similar problems have been studied in economics and optimization theory related to the Fermat-Weber problem *(40)*.

Out of many tasks a system performs, the observed front is determined, to a good approximation, only by those tasks with highest impact on fitness $\frac{\partial F_h}{\partial P_i} \frac{\partial P_i}{\partial d}$. Having numerous tasks with separated archetypes and large and comparable impact seems unlikely, which may explain the prevalence of lines and triangles over more complex shapes. In other words, for more than three tasks to visibly impact the shape of the Pareto front, they must all have large and comparable fitness and performance gradients, and have archetypes distant enough to be noticeable in the data. The higher the resolution and size of the dataset, the more feasible it is to notice archetypes of minor tasks with small weights.

Developmental constraints and other evolutionary forces

The well-known limitations of selectionist (also called adaptationist) theories apply in the present case. These limitations refer to different effects that can keep actual phenotypes away from the predictions. These effects include genetic drift, important for example when populations are small and/or selection is weak *(41, 42)*. Migration, mutation, lack of time and genetic variation (including phylogenetic inertia), or local maxima in fitness can all keep points away from their predicted position. Physical constraints can limit the feasible morphospace *(7, 43)*.

Developmental constraints can limit the region of morphospace which is reachable within a given timeframe *(10, 44)*.

Developmental constraints are often invoked to explain patterns in shapes across organisms *(10-12)*. One may consider the possibility that developmental pathways have evolved to produce the range of phenotypes that were adaptive in the different environments sampled by the ancestral line of the taxon. In an extreme case, the developmental pathway would 'encode' the Pareto front, in the sense that mutations and perturbation would, with high probability, be canalized along the front and not away from it. This may explain why in some systems, such as rodent molars, perturbing development leads to phenotypes arranged on the same curve defined by different species *(11)*. In other systems, such as *Bicyclus* butterfly wing spots and wing shape, selective breeding can readily produce phenotypes off the allometric line, suggesting that short-term evolution of linear relations between traits is not limited by developmental constraints *(45, 46)*. Instead, natural selection is constantly at play to maintain the line *(45, 46)*. This conclusion seems to hold in almost all studies of developmental constraints reviewed in *(10)*. Developmental constraints do seem to bias the phenotypic effects of mutations in certain directions, but this constraint is relative and not absolute- and thus breeding experiments can reach points off the suite of variation.

Theorem: The Pareto front is identical with the convex hull defined by the archetypes

Introduction

Here we provide a theorem that identifies the "Pareto front" with the convex hull of the archetypes, in a defined setting: The conclusion of the theorem holds for any finite-dimensional vector space possessing an inner product, provided that this space is also endowed with the corresponding inner-product norm and norm topology.

We denote the real numbers by $\Re$, and we designate by $\Re^k$ the *k*-dimensional Euclidean space.

Suppose that the points $x_1,\ldots,x_M$ residing in a finite dimensional inner-product (and inner-product normed) real vector space V represent archetype phenotypes of a biological system. Suppose also that each point in V represents some phenotype (not necessarily an archetype). We assume that the fitness function $f : V \to \Re$ is given, for each $x \in V$, by

$$f(x) := g\left(\|x - x_1\|,\ldots,\|x - x_M\|\right), \qquad 1$$

where $g : \Re^M \to \Re$ is a monotonically decreasing function in each of its arguments:

$$b > a \Rightarrow g(a_1,\ldots,a_{i-1},b,a_{i+1},\ldots,a_M) < g(a_1,\ldots,a_{i-1},a,a_{i+1},\ldots,a_M) \qquad (i = 1,\ldots,M).$$

We want to study the Pareto front of the function *f*. Loosely speaking, the Pareto front of *f* is the set of all points *x* in *V* for which every point different from *x* is "farther away" than *x* from at least one of the archetypal phenotypes.

In the following we show that the Pareto front of the archetypal phenotypes is identical with their convex hull.

Some definitions

We let *V* denote a finite-dimensional vector space over the real numbers. We assume that *V* carries an inner product "·". We also assume that *V* has the norm $\| \ \|$ derived from the inner product in the usual way: $\|x\| := \sqrt{x \cdot x}, \forall x \in V$.

We shall assume that *V* carries the norm topology. Therefore, *V* is a Hausdorff locally convex topological vector space. (A topological space is *Hausdorff* if any two distinct points reside in disjoint neighborhoods; a topological vector space is *locally convex* if every neighborhood of the zero vector contains a convex neighborhood of the zero vector.)

Throughout this section, we will use the symbol *X* to denote a finite subset of *V* containing *M* distinguished vectors (corresponding to the archetypal phenotypes): $X = \{x_1, ..., x_M\}$.

Definition 1. A point $x \in V$ is *Pareto with respect to X* if for each $y \in V$, $y \neq x$, there exists $x_i \in X$ such that $\|y - x_i\| > \|x - x_i\|$.

Definition 2. The *Pareto front of X*, denoted *P(X)*, is the set of all points in *V* which are Pareto with respect to *X*.

Definition 3. The *convex hull of X*, denoted *CH(X)*, is the set

$$CH(X) := \left\{ x \in V : x = \sum_{n=1}^{M} \alpha_n x_n, \alpha_n \geq 0 \ (n = 1, ..., M), \sum_{n=1}^{M} \alpha_n = 1 \right\}$$

To prove the result of this section, we shall need the Hahn-Banach Theorem of analysis:

Theorem 1. (Hahn-Banach)**.** Let *V* be a Hausdorff locally convex topological vector space, and let *A* and *B* be non-empty disjoint closed convex subsets of *V* with *B* compact. Then there exists a continuous linear function $h : V \to \Re$ and a number $\gamma \in \Re$ such that $h(a) < \gamma, \forall a \in A$ and $h(b) > \gamma, \forall b \in B$.

Intuitively, the Hahn-Banach Theorem says is that it is always possible to separate the sets described in the statement of the theorem by a hyperplane.

Remark 1. We recall from linear algebra that in the case of a finite dimensional inner product space *W*, for each linear function $T : W \to \Re$ there exists a unique vector $w \in W$ such that $Tu = w \cdot u, \forall u \in W$.

Result

Theorem 2. *P(X) = CH(X).*

Proof. Suppose that $x \in CH(X)$. Suppose also, contrary to what is to be proved, that $x \notin P(X)$. Then there exists $y \in V$, $y \neq x$, such that

$$(y - x_n) \cdot (y - x_n) \leq (x - x_n) \cdot (x - x_n) \ (n = 1, ..., M). \qquad 2$$

From equation (2) we have that

$$y \cdot y - 2(y - x) \cdot x_n - x \cdot x \leq 0 \ (n = 1, ..., M). \qquad 3$$

By supposition, there are numbers $\alpha_1,...,\alpha_M, \alpha_n \geq 0\ (n=1,...,M), \sum_{n=1}^{M}\alpha_n = 1,$ such that $x = \sum_{n=1}^{M}\alpha_n x_n$. Multiplying both sides of each inequality in (3) by the corresponding α_n and summing the resulting inequalities we obtain

$$y\cdot y - 2y\cdot x + x\cdot x \leq 0. \qquad 4$$

From (4) we have that

$$(y-x)\cdot(y-x) \leq 0. \qquad 5$$

Recall that the inner product of a vector with itself is always nonnegative. Therefore, equality obtains in (5). Recall also that the inner product of a vector with itself is zero only if the vector is zero. We therefore have that $y - x = 0$, in contradiction with the inequality $y \neq x$. Hence, $x \in P(X)$, and because the choice of x was arbitrary, we have that $CH(X) \subset P(X)$.

To prove the opposite inclusion, suppose that $x \in P(X)$ but, contrary to what is to be proved, $x \notin CH(X)$. Then $CH(X)$ and $\{x\}$ are nonempty, disjoint, closed and convex subsets of V, with $\{x\}$ compact. By the Hahn-Banach Theorem and Remark 1, there exists a non-zero vector $v \in V$ and a number $\gamma \in \mathfrak{R}$ such that

$$v\cdot x > \gamma, \qquad 6$$
$$v\cdot x_n < \gamma\ (n=1,...,M).$$

Inequality (6) implies that

$$v\cdot(x-x_n) > 0\ (n=1,...,M). \qquad 7$$

We define the number t by

$$t = \min_{x_n \in X} \frac{2v\cdot(x-x_n)}{v\cdot v}. \qquad 8$$

From (7) we have that $t > 0$.

Recall that by supposition $x \in P(X)$. Then there exists $x_i \in X$ such that

$$(x - tv - x_i)\cdot(x - tv - x_i) > (x-x_i)\cdot(x-x_i) \qquad 9$$

By rearranging inequality (9) we obtain the contradiction

$$t > \frac{2v\cdot(x-x_i)}{v\cdot v} \geq t, \qquad 10$$

implying that our assumption $x \notin CH(X)$ is false. Because our choice of $x \in P(X)$ was arbitrary, we have that $P(X) \subset CH(X)$.

□

Remark 2. Note that Theorem 2 obtains for any finite dimensional vector space V possessing some inner product and having a norm (and norm topology) induced by the inner product.

Remark 3. Note that the "*CH(X)⊂P(X)"* direction of proof of Theorem 2 does not require that *V* be finite-dimensional. The proof of the opposite inclusion does require that *V* be finite dimensional, because use is made of Remark 1.

Calculation of the weights θ_i for the case of an additive fitness function

Here we provide a concrete example for the relation between the weights that describe the position of a phenotype on the front, and the relative impact of each task on fitness. As a simple example consider two performance functions, $P_1(v)$ and $P_2(v)$, each maximized by a different archetype (Figure S20A,E). Consider the following fitness function that linearly combines the two performance functions according to a parameter φ, that corresponds to the fraction of time the organism performs task 1 in its habitat and $1-\varphi$ the fraction of time it performs task 2:

$$F_\varphi = \varphi P_1(v) + (1-\varphi) P_2(v)$$

As shown above, the optimal phenotype maximizing $f_\varphi(v)$ is a convex combination of the two archetypes

$$v_\varphi = \theta_\varphi v_1^* + (1-\theta_\varphi) v_2^*$$

We next calculate the phenotype weights θ_φ as a function of the habitat parameter φ. From the derivation in the beginning of this section, we find that

$$\theta_\varphi = \frac{\varphi \frac{\partial P_1}{\partial d_1}}{\varphi \frac{\partial P_1}{\partial d_1} + (1-\varphi)\frac{\partial P_2}{\partial d_2}}$$

When $\varphi=0$ *or* $\varphi=1$, the fitness function equals the corresponding performance function (Figure S20A,E), which results in the optimal phenotypes being the archetypes (i.e. $\theta=0,1$). This analysis shows that the values of θ can be different from φ. $\theta=\varphi$ only when $\frac{\partial P_1}{\partial d_1} = \frac{\partial P_2}{\partial d_2}$ at the optimal phenotype v_h. For equal performance functions, each centered at the corresponding archetype, this occurs at $\varphi=0.5$, Figure S20C. No matter what the performance functions, however, the weight θ is an increasing function of the parameter φ, so that the more often a task is performed relative to others (or the more heavily it contributes to the fitness function), the closer the phenotype to the corresponding archetype.

Figure S21 depicts the dependence of θ on the habitat parameter φ, for two identical performance functions shaped as Gaussians with width σ, $P_i = \exp\left(-d_i/\sigma^2\right)$ (recall that d_i is the distance squared from the archetype). For low σ values, the phenotypes will be concentrated near the archetypes (red line), high σ values result in $\theta_\varphi=\varphi$ (black line), and intermediate values of σ results in weights that are biased towards the archetypes (blue line). Note that at low σ values the optimal phenotypes might only include the archetypes, as discussed in the following section.

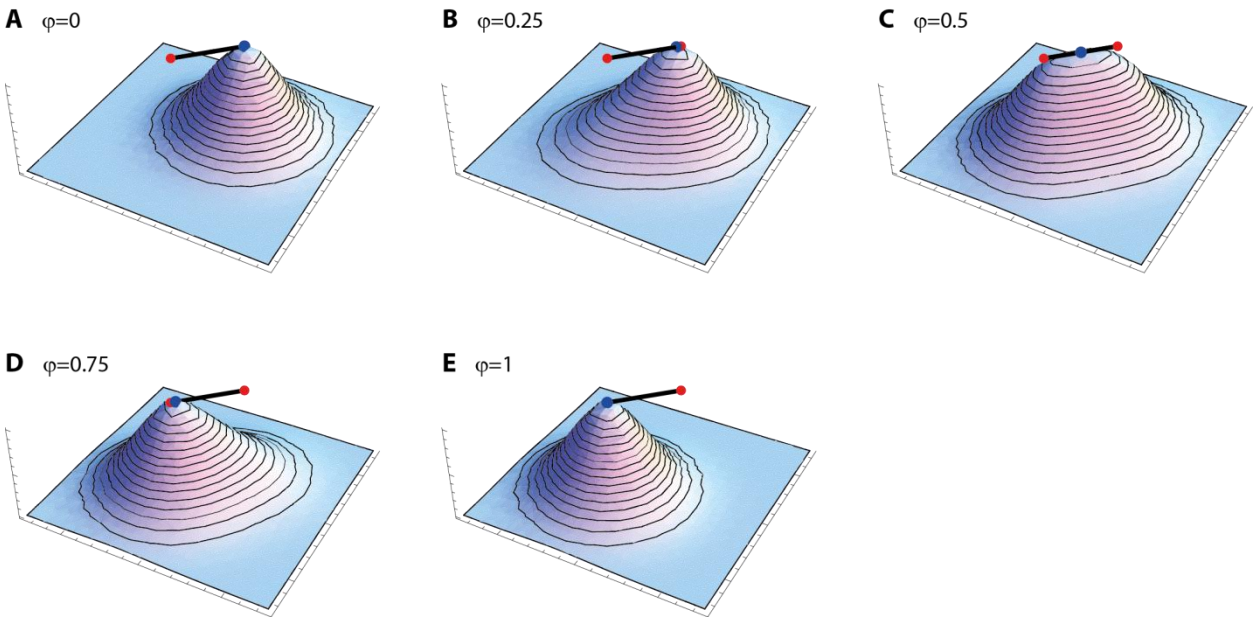

Figure S20. Fitness functions at different habitats described by the parameter φ that weighs the two performances in an additive fitness function. **A.** $\varphi=0$, **B.** $\varphi=.25$, **C.** $\varphi=0.5$, **D.** $\varphi=0.75$ **E.** $\varphi=1$. Red dots denote locations of the archetypes. Blue dot is the location of the phenotype maximizing the fitness function.

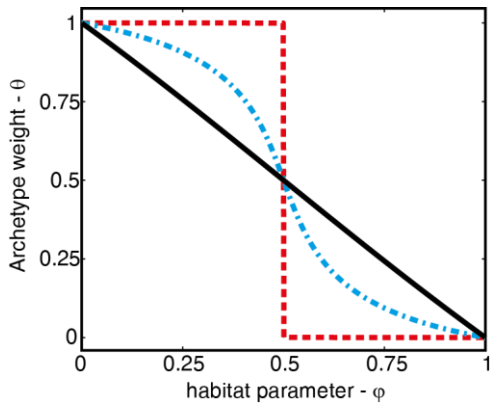

Figure S21. Phenotype weights θ vs. habitat parameter φ. Performance functions are equal and Gaussian shaped with width σ. Different lines denote different values of σ - high (σ^2=1000, solid black), low (σ^2=10, dashed red), and intermediate (σ^2=50, dash-dot blue). Distance between archetypes equals $6\sqrt{2}$.

<u>The set of optimal phenotypes can, in some cases, include only a subset of the convex hull of the archetypes</u>

The derivation of the optimal phenotype required that it maximize the fitness function, dF/dv=0. However, in certain cases, the solution can be a minimizing phenotype, due for example to a positive second derivative. To illustrate this, consider again a fitness function that linearly combines the two performance functions according to a parameter φ:

$$F_\varphi(v) = \varphi P_1(v) + (1-\varphi) P_2(v)$$

In order to find the optimal phenotype we differentiate with respect to *v*:

$$\nabla_v F_\varphi(v) = \varphi \nabla_v P_1(v) + (1-\varphi) \nabla_v P_2(v)$$

and at the optimal phenotype, v_φ, the derivative is zero $\nabla_v F_\varphi(v)\big|_{v=v_\varphi} = 0$. The sign of the second derivative (at the optimal phenotype v_φ) describes the phenotype as maximizing or minimizing the fitness

$$\nabla_v^2 F_\varphi(v) = \varphi \nabla_v^2 P_1(v) + (1-\varphi) \nabla_v^2 P_2(v)$$

Thus the curvature of the performance function can result in minimal fitness. As an example, consider two equal Gaussian shaped performance functions with width σ. The second derivative of a Gaussian becomes positive at a distance larger than $\sigma/\sqrt{2}$ from the archetype. Thus, if the typical width σ of the performance function is much smaller than the distance between the archetypes, then we find that the phenotype minimizes fitness (Figure S22, Figure S21 red dashed line). For every value of φ, one of the archetypes has the highest fitness. This results in the selected phenotypes being one of the archetypes.

On the other hand, if the two archetypes are closer than the width $\sigma/\sqrt{2}$ of the performance functions, the Pareto front can be fully populated, because the second derivative is negative at every point on the line, so that every point on the line is a maximum in a certain habitat (Figure S20).

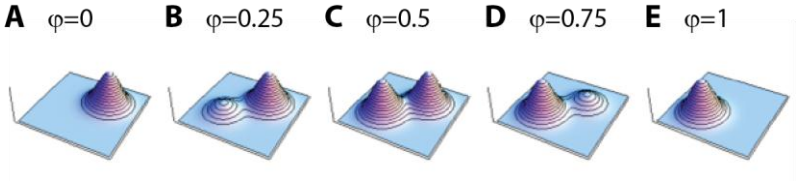

Figure S22. For 'narrow' performance functions (σ^2=1, distance between archetypes equals $2\sqrt{2}$), the extremal phenotype minimizes fitness. This results in the selected phenotypes being the archetypes. **A.** $\varphi=0$, **B.** $\varphi=.25$, **C.** $\varphi=0.5$, **D.** $\varphi=0.75$ **E.** $\varphi=1$.

Measurement of a subset of traits

A given biological system may be characterized by hundreds of quantitative traits. Often, one can only measure a subset of these traits. A virtue of the convex-hull-shaped fronts in the present study is that they can be seen in many cases even if only some traits are measured and others not.

For example, if the front is a triangle (three tasks) in high dimensional morphospace, then measuring any two traits will be generally be enough to see the triangle. This is because the projection of the triangle onto any plane, defined by the two measured traits, will still be a triangle (Figure S23). Only in the singular case where the triangle is perpendicular to the plane, will its projection be a line segment (not a triangle), and one would 'miss' one of the archetypes. The same is true of two tasks: a projection of a line segment on any plane is also a line segment (or, in a singular case, a point).

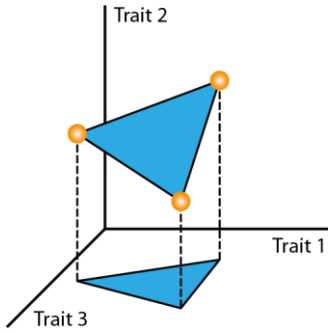

Figure S23. The projection of a convex hull remains a convex hull. If one would measure only traits 1 and 3, a triangle remains.

In the cases of higher numbers of tasks, a similar situation is found. Consider the case of four tasks ($k=4$) and three traits ($n=3$). According to the present theory we expect the suite of variation to be included in a tetrahedron with vertices at the four archetypes (Fig 2C). If we measure only two traits of this system we will get the projection of the suite of variation on these two traits (Figure S24A). Instead of a tetrahedron, we now measure a rectangle, where each vertex is the projection of the archetype on the two remaining traits. Note that the projection of a convex hull is also a convex hull. Thus we are still left with significant information – the values of the measured traits are still convex combinations of the archetypes.

In some cases the projection of the suite of variation might result in an archetype within the interior of the lower dimension convex hull. Consider the case depicted in Figure S24B. The center archetype (at the top of the tetrahedron) is obscured in the lower dimension convex hull. Thus in such cases where we measure not all relevant traits, we might 'miss' an archetype.

Note that there might also be cases where the degrees of freedom of the biological system are limited for physical reasons. For example, a case where there are four tasks, but only two traits can be changed in the biological system. These cases will show similar results to the above discussion.

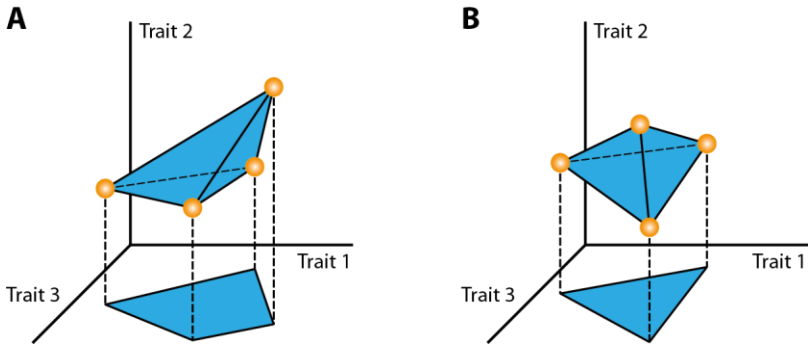

Figure S24. **A.** In the case of four tasks and three traits we expect the suite of variation to be shaped as a tetrahedron. When only two traits are measured, the projection of the tetrahedron becomes a rectangle. **B.** In certain archetype configurations, the projection might result in a hidden archetype.

References and Notes

1. S. J. Arnold, Morphology, Performance and Fitness. *Am. Zool.* **23**, 347 (1983).

2. G. F. Oster, E. O. Wilson, *Caste and Ecology in the Social Insects* (Princeton University Press, NJ, 1979).

3. K. D. Farnsworth, K. J. Niklas, Theories of optimization, form and function in branching architecture in plants. *Funct. Ecol.* **9**, 355 (1995). doi:10.2307/2389997

4. H. El Samad, M. Khammash, C. Homescu, L. Petzold, Optimal performance of the heat-shock gene regulatory network, *Proc. 16th IFAC World Congress* (2005).

5. M. C. Kennedy, Functional–structural models optimize the placement of foliage units for multiple whole-canopy functions. *Ecol. Res.* **25**, 723 (2010). doi:10.1007/s11284-009-0658-6

6. R. E. Steuer, *Multiple criteria optimization: theory, computation, and application* (Wiley, New York, 1986).

7. G. R. McGhee, *The geometry of evolution: adaptive landscapes and theoretical morphospaces* (Cambridge University Press, Cambridge, 2007).

8. Supplementary materials are available on *Science* Online.

9. S. J. Gould, Allometry and size in ontogeny and phylogeny. *Biol. Rev. Camb. Philos. Soc.* **41**, 587 (1966). doi:10.1111/j.1469-185X.1966.tb01624.x Medline

10. C. P. Klingenberg, Evolution and development of shape: integrating quantitative approaches. *Nat. Rev. Genet.* **11**, 623 (2010). Medline

11. K. D. Kavanagh, A. R. Evans, J. Jernvall, Predicting evolutionary patterns of mammalian teeth from development. *Nature* **449**, 427 (2007). doi:10.1038/nature06153 Medline

12. G. B. West, J. H. Brown, B. J. Enquist, A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122 (1997). doi:10.1126/science.276.5309.122 Medline

13. P. R. Grant, I. Abbott, D. Schluter, R. L. Curry, L. K. Abbott, Variation in the size and shape of Darwin's finches. *Biol. J. Linn. Soc. Lond.* **25**, 1 (1985). doi:10.1111/j.1095-8312.1985.tb00384.x

14. Software that analyzes data in terms of triangles and their significance, and suggests archetype trait values, is available for download at www.weizmann.ac.il/mcb/UriAlon.

15. E. O. Wilson, Caste and division of labor in leaf-cutter ants (Hymenoptera: Formicidae: Atta). *Behav. Ecol. Sociobiol.* **7**, 143 (1980). doi:10.1007/BF00299520

16. U. M. Norberg, J. M. V. Rayner, Ecological Morphology and Flight in Bats (Mammalia; Chiroptera): Wing Adaptations, Flight Performance, Foraging Strategy and Echolocation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **316**, 335 (1987). doi:10.1098/rstb.1987.0030

17. T. Nyström, Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? *Mol. Microbiol.* **54**, 855 (2004). doi:10.1111/j.1365-2958.2004.04342.x Medline

18. A. Zaslaver *et al*., Invariant distribution of promoter activities in Escherichia coli. *PLOS Comput. Biol.* **5**, e1000545 (2009). doi:10.1371/journal.pcbi.1000545 Medline

19. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099 (2010). doi:10.1126/science.1192588 Medline

20. N. Geva-Zatorsky *et al.*, Protein dynamics in drug combinations: a linear superposition of individual-drug responses. *Cell* **140**, 643 (2010). doi:10.1016/j.cell.2010.02.011 Medline

21. D. Schluter, Adaptive Radiation Along Genetic Lines of Least Resistance. *Evolution* **50**, 1766 (1996). doi:10.2307/2410734

22. C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469 (1996). doi:10.1145/235815.235821

23. V. Klee, M. C. Laskowski, Finding the smallest triangles containing a given convex polygon. *J. Algorithms* **6**, 359 (1985). doi:10.1016/0196-6774(85)90005-7

24. O. Campàs, R. Mallarino, A. Herrel, A. Abzhanov, M. P. Brenner, Scaling and shear transformations capture beak shape variation in Darwin's finches. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3356 (2010). doi:10.1073/pnas.0911575107 Medline

25. P. R. Grant, *Ecology and evolution of Darwin's finches* (Princeton University Press, Princeton, 1999).

26. S. J. Millington, P. R. Grant, Feeding ecology and territoriality of the Cactus Finch Geospiza scandens on Isla Daphne Major, Galapagos. *Oecologia* **58**, 76 (1983). doi:10.1007/BF00384545

27. P. R. Grant, B. R. Grant, Evolution of character displacement in Darwin's finches. *Science* **313**, 224 (2006). doi:10.1126/science.1128374 Medline

28. B. R. Grant, Selection on Bill Characters in a Population of Darwin's Finches: Geospiza conirostris on Isla Genovesa, Galapagos. *Evolution* **39**, 523 (1985). doi:10.2307/2408650

29. D. Schluter, P. R. Grant, Ecological Correlates of Morphological Evolution in a Darwin's Finch, Geospiza difficilis. *Evolution* **38**, 856 (1984). doi:10.2307/2408396

30. B. Holldobler, E. O. Wilson, *The Ants* (Belknap Press of Harvard University Press, ed. 1st, Cambridge, 1990).

31. J. S. Findley, D. E. Wilson, Observations on the neotropical disk-winged bat, Thyroptera tricolor spix. *J. Mammal.* **55**, 562 (1974). doi:10.2307/1379546 Medline

32. G. Neuweiler, Auditory adaptations for prey capture in echolocating bats. *Physiol. Rev.* **70**, 615 (1990). Medline

33. A. M. Hutson, S. P. Mickleburgh, P. A. Racey, Eds., *Microchiropteran bats: global status survey and conservation action plan* [International Union for Conservation of Nature Species Survival Commission (IUCN/SSC) Chiroptera Specialist Group, Gland, Switzerland, 2001].

34. D. Navarro, D. E. Wilson, Vampyrum Spectrum. *Mammal. Species* **184**, 1 (1982).

35. W. S. Hudson, D. E. Wilson, Macroderma gigas. *Mamm. Species* **260**, 1 (1986). doi:10.2307/3503920

36. A. Zaslaver *et al*., A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat. Methods* **3**, 623 (2006). doi:10.1038/nmeth895 Medline

37. T. Bollenbach, R. Kishony, Resolution of gene regulatory conflicts caused by combinations of antibiotics. *Mol. Cell* **42**, 413 (2011). doi:10.1016/j.molcel.2011.04.016 Medline

38. Y. Savir, E. Noor, R. Milo, T. Tlusty, Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3475 (2010). doi:10.1073/pnas.0911663107 Medline

39. P. C. Wainwright, M. E. Alfaro, D. I. Bolnick, C. D. Hulsey, Many-to-One Mapping of Form to Function: A General Principle in Organismal Design? *Integr. Comp. Biol.* **45**, 256 (2005). doi:10.1093/icb/45.2.256 Medline

40. H.W. Kuhn, On a pair of dual nonlinear programs, in: Methods of nonlinear programming, Ed. J. Abadie (North-Holland, Amsterdam, 1967) 38-54.

41. Lande, Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314 (1976). doi:10.2307/2407703

42. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, 1985).

43. D. M. Raup, Geometric Analysis of Shell Coiling: General Problems. *J. Paleontol.* **40**, 1178 (1966).

44. J. Maynard Smith *et al*., Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution. *Q. Rev. Biol.* **60**, 265 (1985). doi:10.1086/414425

45. W. A. Frankino, B. J. Zwaan, D. L. Stern, P. M. Brakefield, Natural selection and developmental constraints in the evolution of allometries. *Science* **307**, 718 (2005). doi:10.1126/science.1105409 Medline

46. C. E. Allen, P. Beldade, B. J. Zwaan, P. M. Brakefield, Differences in the selection response of serially repeated color pattern characters: standing variation, development, and evolution. *BMC Evol. Biol.* **8**, 94 (2008). doi:10.1186/1471-2148-8-94 Medline