bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title

The evolution of trait correlations constrains phenotypic adaptation to high CO_2 in a eukaryotic alga

Running Title

Biogeochemically important, multi-trait adaptation in a eukaryotic alga

Nathan G. Walworth[1], Jana Hinners[2], Phoebe A. Argyle[3], Suzana G. Leles[1], Martina A. Doblin[3], Sinéad Collins[2], Naomi M. Levine[1]*

[1]Department of Biological Sciences, University of Southern California, Los Angeles, California 90089-0371, USA.

[2] Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom

[3] Climate Change Cluster, University of Technology Sydney, Sydney, NSW 2007, Australia

Corresponding author

* Naomi M. Levine

3616 Trousdale Parkway,

University of Southern California,

Los Angeles, CA 90089

Phone: (213) 821-0745

e-mail: n.levine@usc.edu

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Microbes form the base of food webs and drive biogeochemical cycling. Predicting the effects of microbial evolution on global elemental cycles remains a significant challenge due to the sheer number of interacting environmental and trait combinations. Here we present an approach for modeling the interactive effects of *de novo* biological change and multivariate trait correlation evolution using principal component axes. We investigated the outcome of thousands of possible adaptive walks parameterized using empirical evolution data from the alga *Chlamydomonas* exposed to high CO_2. We found that only a limited number of phenotypes emerged. Applying adaptive trait correlations to the starting population (historical bias) accelerated adaptation while highly convergent, nonrandom phenotypic solutions emerged irrespective of bias. These findings are consistent with a limited set of evolutionary trajectories underlying the vast amount of possible trait combinations (phenotypes). Critically, we demonstrate that these dynamics emerge in an empirically defined multidimensional trait space and show that trait correlations, in addition to trait values, must evolve to explain multi-trait adaptation. Identifying high probability high-fitness outcomes based on trait correlations is necessary in order to connect microbial evolutionary responses to biogeochemical cycling, thereby enabling the incorporation of these dynamics into global ecosystem models.

Keywords: microbial evolution, trait correlations, trait adaptation, phytoplankton, biogeochemistry, PCA

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Microbes play a critical role in regulating biogeochemistry and the global climate. In recent years, there has been a significant increase in global change studies examining the roles of microbial evolution in shaping future biogeochemical cycles. This work has helped to more explicitly integrate the fields of evolution and microbial ecology resulting in both long-term experimental evolution studies with ecologically important microbes and, to a limited extent, the incorporation of adaptation into ecological and ocean circulation models [1-13]. These studies are just the first step in tackling the immensely complex challenge of microbial evolution and its influence on global biogeochemistry. Critically, we still have only a limited understanding of how microbial communities will respond to multi-stressor and fluctuating environmental change, and the sheer number of interacting environmental and trait combinations exceeds our experimental ability to do so [14,15]. Hence, experimental and theoretical methods to reduce dimensionality and extract broad evolutionary patterns across traits and taxa are critical for creating a predictive framework that can both help guide experiments and make more accurate future predictions [5].

Here, we aim to understand how different evolutionary starting points derived from multiple traits and their relationships (historical bias) can constrain overarching evolutionary trajectories of phenotypes (suites of traits) in microbial populations adapting to environmental change. We broadly define bias as standing trait correlations (i.e., relationships) in a population that are heritable and can impact fitness such that, over time, these correlations can constrain the direction of evolution [16]. Since our overall goal is to assess how biogeochemically-important, microbial traits and their relationships will evolve in response to future environmental change, our approach is designed to facilitate future integration into global biogeochemical models. Specifically, our framework can be used to replace an assumption commonly used in

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

biogeochemical models that existing interspecific trait relationships will govern future microbial phenotypes such that fixed tradeoffs determine competitive outcomes across different environmental conditions. In reality, tradeoffs can evolve [17] and microbial populations can display different plastic and evolutionary responses [10,18]. Furthermore, growing evidence demonstrates that intraspecific trait variation can be significant in phytoplankton, and that constraints on trait relationships will bias evolutionary trajectories of biogeochemically important microbial populations in the face of environmental change [19-21].

Seminal research modeling the interaction of complex trait relationships, inheritance, epistasis, and metabolic networks has been conducted on theoretical populations experiencing environmental change [16-20]. These studies have broadly found that an evolving population may be able to access only a subset of phenotypes depending on both its initial trait values and trait correlations. Specifically, these studies (e.g., [16,22,23]) have used quantitative genetics approaches to study adaptive walks accounting for uncertainty inherent in trait variation, genotypic variability, inheritance, and environmental variability. They created theoretical frameworks using multivariate and eigenvector methods to examine evolutionary trade-offs between biological and environmental dimensions over time through primarily accounting for the standing genetic variation. Other theoretical approaches emphasized the role of de novo mutation in a fitness landscape without accounting for standing genetic variation [23,24]. These studies are often entirely theoretical [24,25], or empirically limited by the need to measure the fitness impacts of every possible single mutation. Here, we have created a blended approach at the trait level that models how de novo trait changes map onto standing trait variation, and parameterized our model with empirical trait data from a laboratory evolution experiment.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

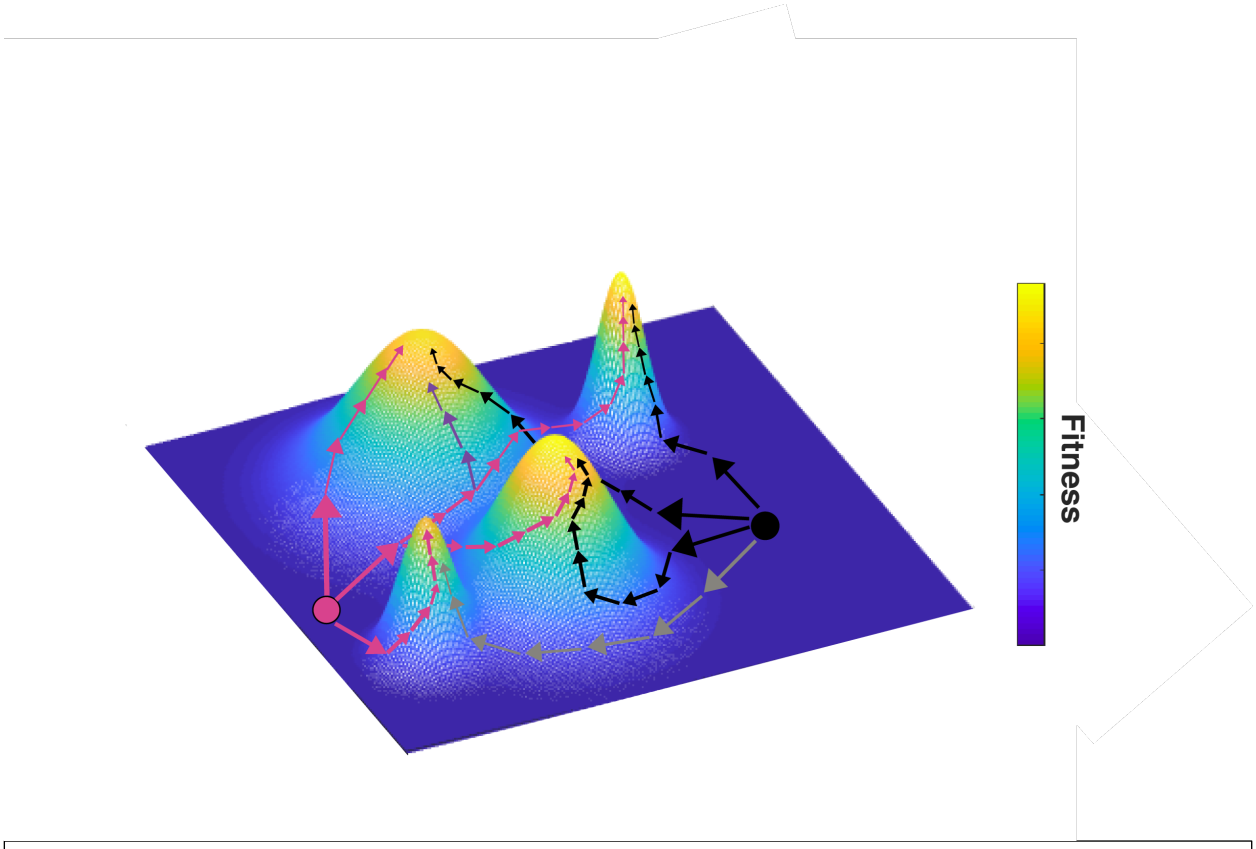


Fig. 1. Comparison of adaptive walks between two different phenotypes in a rugged fitness landscape with four high-fitness peaks Two example starting phenotypes are represented as circles (magenta and black). The x- and y-axis represent dimensions in fitness space (e.g., different traits). The phenotypes start with low fitness (z-axis) and through trait and trait-correlation changes move to higher fitness. The adaptive walk is governed by historical bias, or different initial trait architecture, that impacts the movement of the population within the landscape. As the adaptive walk proceeds, the population moves to the top of one of the fitness peaks. While there are several paths available to each starting phenotype (represented by magenta and black arrows), due to historical bias (trait correlation constraints), some paths can be inaccessible (denoted by the grey and purple arrows). Note that depending on historical bias (i.e., phenotypic starting location), some high fitness peaks are either more difficult to access or completely inaccessible.

Previous studies in developmental bias have used empirical data to demonstrate that biological systems will produce certain phenotypic variants more readily than others in response to a perturbation (mutation or environmental change) due to the inherent structure, composition, and evolutionary history of a population [26,27]. These findings contrast with the long-held assumption of isotropic (i.e. equal) variation [28] and have revealed instead that only a limited part of multivariate phenotypic space (i.e. only certain phenotypes) can be accessed [29]. Critically, this explains why not all viable trait combinations are explored [28]. In summary, a growing body of literature has shown that genetic architecture influences how traits and trait correlations are

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

impacted by environmental shifts and that these shifts produce nonrandom distributions of phenotypes [30-32].

While inaccessible trait combinations (i.e. phenotypes) have been well-documented in the evolutionary literature [23,26,28] [16,22,33], there have been few attempts to investigate the implications of this phenomenon for the evolution of trait and trait correlations of photosynthetic microbes [19,20]. Therefore, we lack a fundamental understanding of how evolutionary dynamics can impact biogeochemical cycling when both trait values and trait correlations evolve. Specifically, constraints on how trait combinations evolve in phytoplankton have the potential to impact rates of carbon cycling and shifts in aquatic ecosystem structure that depend on these microscopic primary producers [34,35].

Fig. 1 shows an illustrative example of a rugged fitness landscape (i.e., multiple high-fitness peaks) where each peak represents high-fitness trait combinations (phenotypes). In this example landscape, there are 4 equally high-fitness peaks. However, the accessibility of each peak differs depending on the starting location (ancestral trait values) and the initial trajectory, which is dictated by a population's collective trait relationships (Fig. 1, magenta and black circles and paths). Ultimately, to robustly study microbial trait evolution, we need a framework that allows us to estimate probable evolutionary trajectories given both starting trait combinations and trait correlations (historical bias). Below, we introduce such a framework using empirical evolution data from a eukaryotic alga. The TRAit Correlation Evolution (TRACE) model is a first step towards investigating how correlated metabolic traits with clear biogeochemical significance may impact elemental cycling under environmental change (e.g., ocean acidification). Using a trait-based fitness landscape generated using empirical data from an experimental evolution study with the eukaryotic alga *Chlamydomonas reinhardtii*, we found that only a handful of phenotypic

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variants were reproducible both with and without historical bias. Overall adaptive rates (defined as the number of generations to reach maximum fitness) were impacted by the amount and type of bias (trait correlations) in the model. These results indicate that populations harboring trait correlations oriented in (i.e. consistent with) the direction of selection may experience accelerated rates of adaptation. Understanding which trait relationships inform the probability of adaptive microbial phenotypes will be critical for predicting the short- and long-term contribution of biogeochemically-important traits to biogeochemical cycling.

Materials and Methods

PCA

Ancestral and evolved trait values from low-CO_2 and high-CO_2 adapted populations across 5 genotypes of *Chlamydomonas reinhardtii* were obtained from Lindberg et al. (2020) [3] and can be found in Supplementary File 1. We selected four independent ecologically relevant traits: growth rate, respiration, cell size, and daughter cell production. All empirical trait values were standardized for both ancestral and evolved data. Principal component analyses (PCA) were conducted on ancestral traits resulting in 48% and 37% of the variance explained on axes PC1 and PC2, respectively, and 54% and 32% for evolved traits (Fig. 2a). For both ancestral and evolved PCAs, or trait-scapes, there are 6 trait correlations, which can be found in Supplementary File 1. These PCAs served as the trait-based fitness landscape (trait-scape) for the modeled adaptive walk.

To select a start and end point for the adaptive walk, ancestral populations were projected onto the evolved PC axes. A single genotype was selected for the modeling exercise where the observed ancestral trait values defined the start point of the adaptive walk (tan circle in Fig. 2b; row 20 in the ancestral trait value matrix in Supplementary File 1) and the corresponding evolved

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

146 population trait values defined the evolutionary endpoint (red circle in Fig. 2b; row 20 in the
147 evolved trait value matrix in Supplementary File 1).
148
149 *TRACE Model*
150 The TRAit Correlation Evolution (TRACE) model framework simulates the adaptive walk
151 of a microbial population across a trait landscape (trait-scape) towards a high-fitness area. TRACE
152 was adapted from an individual based Fisher model of adaptation [1,36,37]. Each generation, each
153 individual in the population experienced either a change in trait values or changes in trait and trait-
154 correlations. Changes in trait values moved these individuals across the trait-scape while trait-
155 correlations constrained the direction of movement. Selection was imposed based on distance to
156 the evolutionary end-point in the evolved trait-scape (described below), such that the population
157 evolved towards the high fitness region of the trait-scape. In essence, this framework selected for
158 individuals with the smallest overall difference across all trait values from the empirically observed
159 high fitness phenotype. The weighting of the traits was derived from the observed evolved
160 phenotypes evaluated using PCA, such that traits that were not observed to play an important role
161 in fitness in the high-CO_2 environment had low weight. It is important to note that the model did
162 not directly select for trait correlations, but that specific correlations emerged in the population if
163 they provided a fitness advantage in terms of trait dynamics.
164 In the default model simulations (referred to as 90/10), 90% of individuals were randomly
165 chosen to experience a random change in a trait value (while maintaining all existing trait-
166 correlations) while 10% experienced both a trait and trait-correlation change. These changes were
167 drawn from a Gaussian distribution (mean = 0 and standard deviation = 0.05) such that small
168 changes were common and large changes were rare. For each individual, the randomly chosen trait

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

change was added to the existing trait value. Following this first trait change, the remaining 3 trait values were updated using the trait correlations for that individual in that time step. For example, if trait 1 was initially changed, then traits 2, 3, and 4 would subsequently be updated by multiplying the new trait 1 value by the three trait correlations (1v2, 1v3, and 1v4). To test if the sequence of correlational changes influenced adaptive outcomes in our model, we changed the order in which traits were updated and showed that results remained unchanged as expected (Supplementary Fig. 1; Supplementary text).

The remaining 10% of the population experienced both a trait and a trait correlation change. For each individual, one of the six trait correlations was randomly selected to change. Similar to the trait change, a random value was drawn from a Gaussian distribution with a mean of 0 and standard deviation of 0.05 and added to the existing correlation value. Next, one of the two traits associated with that correlation was randomly chosen and a trait change was selected in the same manner as above. Next, we updated the second trait tied to the correlation using the new correlation and trait value (the other 2 trait values were not updated in this generation).

Selection was imposed using distance to the high fitness area (evolutionary end coordinate) as a measure of fitness. Following changes to the respective traits and correlations, all individuals in the population were projected back onto the evolved trait-scape (i.e., evolved PCA) using the evolved factor loadings. The Euclidian distances (z) were calculated for each individual relative to the evolutionary endpoint. Next, fitness was calculated as [1,36]:

$$w(z) = e^{(-z^2)/2} \quad \text{Eq. 1}$$

Finally, individuals were randomly sampled with replacement weighted by fitness to persist to the next generation. This selective approach through probabilistic weighting of fitness was adapted from our previous studies [1,36] inspired by Fisher's model of adaptation [37].

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Model simulations:

The model was initialized with a population of 1000 individuals with the same trait values corresponding to the ancestral trait values. The evolved trait-scape (i.e. evolved PCA), the population starting location (tan circle in Fig. 2b), and the high fitness area (red circle in Fig. 2b) in the evolved trait-scape were defined based on empirical data from the *Chlamydomonas* long-term evolution study [3]. Three different modes of the model were run with varying amounts of starting bias using different starting trait correlations: mixed, ancestral, and evolution (described below). Each model run was conducted for 2000 generations with 100 replicates each. All model parameters are given in Supplementary Table 1. Previous work by us and others have demonstrated that adaptive outcomes using this framework are robust across a wide range of population sizes (Supplementary text) [1,36]. Several sensitivity studies were conducted to test model dynamics. Briefly, we ran the model with a different starting location, varied the order in which traits were updated, removed the influence of trait correlations on evolutionary outcomes, and varied the ratio of trait to trait correlation changes. Detailed explanations and figures can be found in the Supplementary Information.

Mixed mode (no bias) simulations: To first test all possible routes available to travel from the ancestral start point to the evolutionary end point in the evolved trait-scape (Fig. 2b), random correlation values from a standard uniform distribution over the interval (-1,1) were generated and randomly assigned to all individuals within the population. Hence, every individual started with the same 4 trait values but completely random correlation values.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ancestral mode simulations: To test the effects of systematically adding ancestral bias, four ancestral sub-modes were conducted: A1, A2, A3, and A4. For simplicity, we chose to sequentially add back in ancestral correlations based on the empirically calculated significant trait-correlations from most significant to least significant (R^2 = -0.89 to 0.54; see Fig. 2C). For sub-mode A1, random correlation values were generated as above for 5 of the 6 trait correlations, and one empirical ancestral correlation was added back to all individuals. This resulted in a starting population in which each individual contained the same 4 trait values, one ancestral trait correlation value shared across all individuals, and random correlation values for 5 of the 6 trait correlations. The rest of the model steps proceeded as above where all traits and correlations were allowed to change. For A2, all steps were the same except that two empirical ancestral correlations were added. Finally, three and four ancestral correlations were added back for A3 and A4, respectively.

Evolved mode simulations: The same procedure was conducted for the evolved mode but instead empirical evolved correlations were systematically added (modes E1 – E4).

Hierarchical Clustering

Hierarchical clustering with multiscale bootstrap resampling (1,000 replicates) on mean trait correlation values was conducted using R package pvclust [38] using Euclidean distance and the average (UPGMA) method. Principal component analysis using mean correlation values was conducted with package R package vegan [39], and pvclust clusters with approximately unbiased (AU) p-values > 75% were projected onto the PC coordinate plane as convex hulls.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

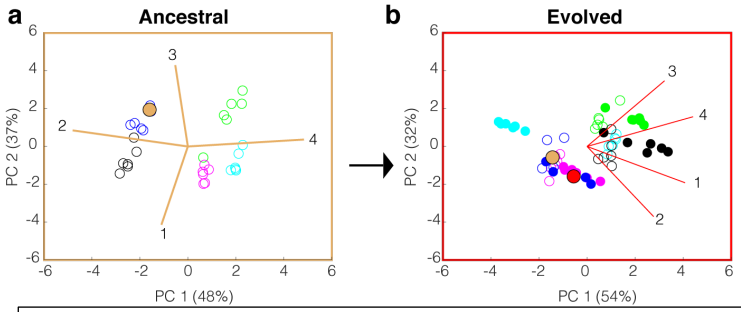
Results

Collapsed multivariate trait-scape

The complexity of multi-dimensional trait evolution requires a tractable framework to understand how trait adaptation might proceed. Previous work has shown that complex trait adaptation and fitness variations can be represented in a reduced dimensional space, specifically using Principle Component axes [20,40]. Building on this work, we created a trait-based landscape or '*trait-scape'* for the green alga *Chlamydomonas reinhardtii* adaptation to high-CO_2 using four ecologically relevant traits (growth rate, respiration, cell size, and daughter cell production). Specifically, using the output from an experimental evolution study [3] with replicate populations of 5 genotypes of *C. reinhardtii*, we demonstrated that both trait values and the correlations between traits changed as the population adapted from a low-CO_2 environment (ancestral environment) to a high-CO_2 environment (evolved environment; Fig. 2). Specifically, all four traits changed to varying degrees depending on the genotype [3], and correlations between traits changed upon high-CO_2 adaptation with some traits becoming correlated (e.g., 1v2) while others becoming uncorrelated (e.g., 2v4; Fig 2c). This resulted in distinct differences between the PCAs (trait-scapes) for the ancestral and evolved populations (Fig. 2). As the specific traits themselves are not relevant for this study, we will hereafter refer to them as traits 1-4. We refer the reader to [3] for an in-depth discussion of the evolution experiment.

To understand how *C. reinhardtii* genotypes adapted to high CO_2, we compared the ancestral genotypes (projection of the ancestral trait values onto the evolved trait-scape) to the evolved genotypes (evolved trait values in evolved trait-scape; Fig. 2b). Fig. 2b shows where replicate populations of ancestral genotypes (empty circles) are located in the evolved trait-scape relative to their corresponding evolved genotypes (filled circles). This analysis demonstrates that

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



c

Correlation

	Trait A	Trait B	Ancestral correlation	Evolved correlation
C1	1	2	X	0.59
C2	1	3	- 0.41	X
C3	1	4	X	0.54
C4	2	3	X	X
C5	2	4	- 0.89	X
C6	3	4	X	0.70

Fig. 2 Principal Component Analysis (PCA) of ancestral and evolved trait values, respectively, and their trait correlations. a) Ancestral PCA calculated from the values of 4 ancestral traits across 5 genotypes where each point represents an independent biological population (i.e. culture) colored by genotype. Percentages along PC1 and PC2 denote the amount of variance explained by each PC axis, respectively. b) Evolved PCA plot calculated from the evolved values of the same 4 traits as in a) across 5 genotypes. Filled circles represent the independent populations of the evolved genotypes. Open circles represent the corresponding populations of the ancestral genotypes in a) projected onto the evolved PC axes. The tan and red filled circles denote the start and end coordinate of the model, respectively. c) Table of all 6 possible trait combinations and their values in their ancestral and evolved genotypes. An 'X' indicates a non-statistically significant trait correlation ($p > 0.05$).

PC axes can provide a reduced dimensional space (trait-scape) for understanding how multiple traits simultaneously respond to environmental perturbation, similar to what has been shown in previous studies [20,28,40]. To understand how trait movement within this collapsed multi-dimensional trait-scape can be constrained by historical bias (previous correlations between traits), we developed a statistical model of multi-trait adaptation and investigated probabilities of different emergent evolutionary outcomes.

Simulating TRAit Correlation Evolution

Using the TRACE model, we explored the impact of historical bias (i.e., correlations between traits) on an adaptive walk where the trait-based fitness landscape and start and evolutionary end-points were defined by empirical data. We began with a 'null hypothesis' model in which there was no historical bias (mixed mode simulation) and then systematically added in

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

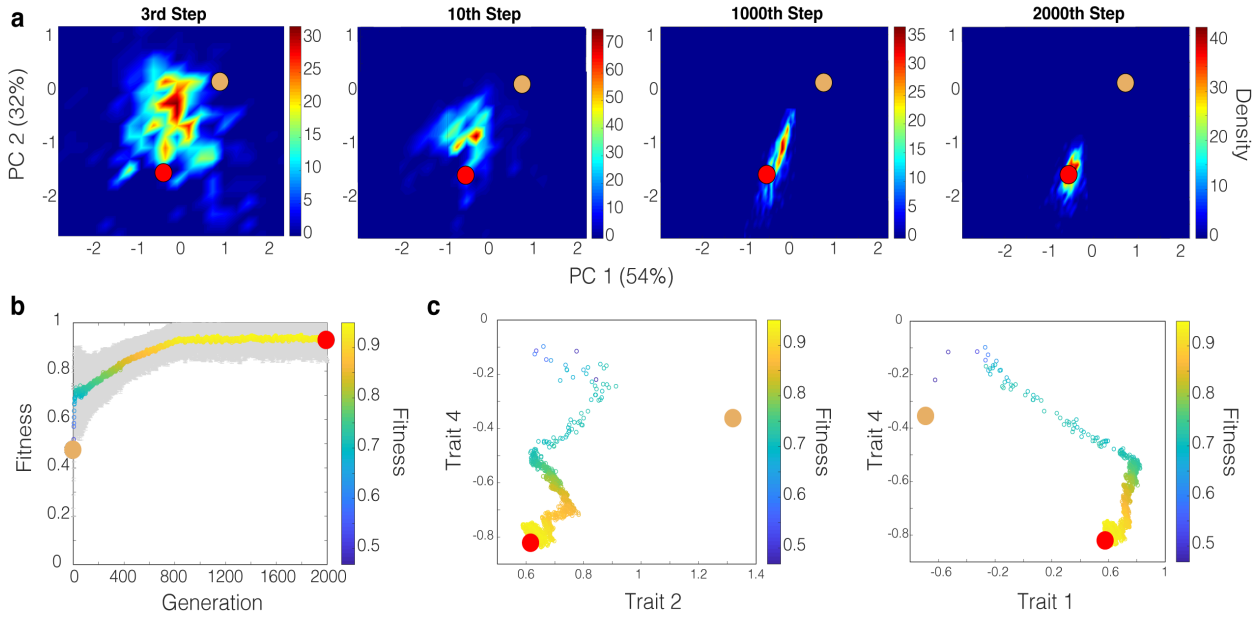


Fig. 3 Representative adaptive walk in evolved PC space of a population of 1000 individuals. a) Density plots of an adaptive walk of a single population for a single run (n = 1000 individuals) starting at the tan dot and ending at the red dot. Each plot represents a different point in time (i.e. generation) in the adaptive walk with the color representing the density of individuals in a given area. b) Fitness plot of the population across the entire adaptive walk with the colored line and grey region representing the mean and standard deviation, respectively. Both the y-axis and color indicate fitness. c) Trait vs. trait plots representing the same adaptive walk where lower fitness denotes the start of the walk and high fitness denotes the end. As in b), each point represents the mean standardized trait value of all individuals at a specific generation, or step.

bias to determine the impact on population level adaptation. An example of model dynamics from a single run in mixed mode is shown in Fig. 3a where a representative population consisting of a thousand individuals moved over time from the ancestral starting phenotype to the evolved high fitness area (Fig. 3a). This resulted in an overall increase in fitness of the population over time (Fig 3b). The underlying dynamics of the model (changes in trait values and trait correlation changes) for 3 representative traits are shown in Fig. 3c.

Consistent with prior studies examining evolution under relatively strong selective pressure [1,36], fitness effects produced from changes at the beginning of the walk were significantly greater than at the end of the walk [41-44] (Fig. 3). As the model ran forward in time, individuals within the population explored the collapsed trait-scape through changes to both traits and their

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

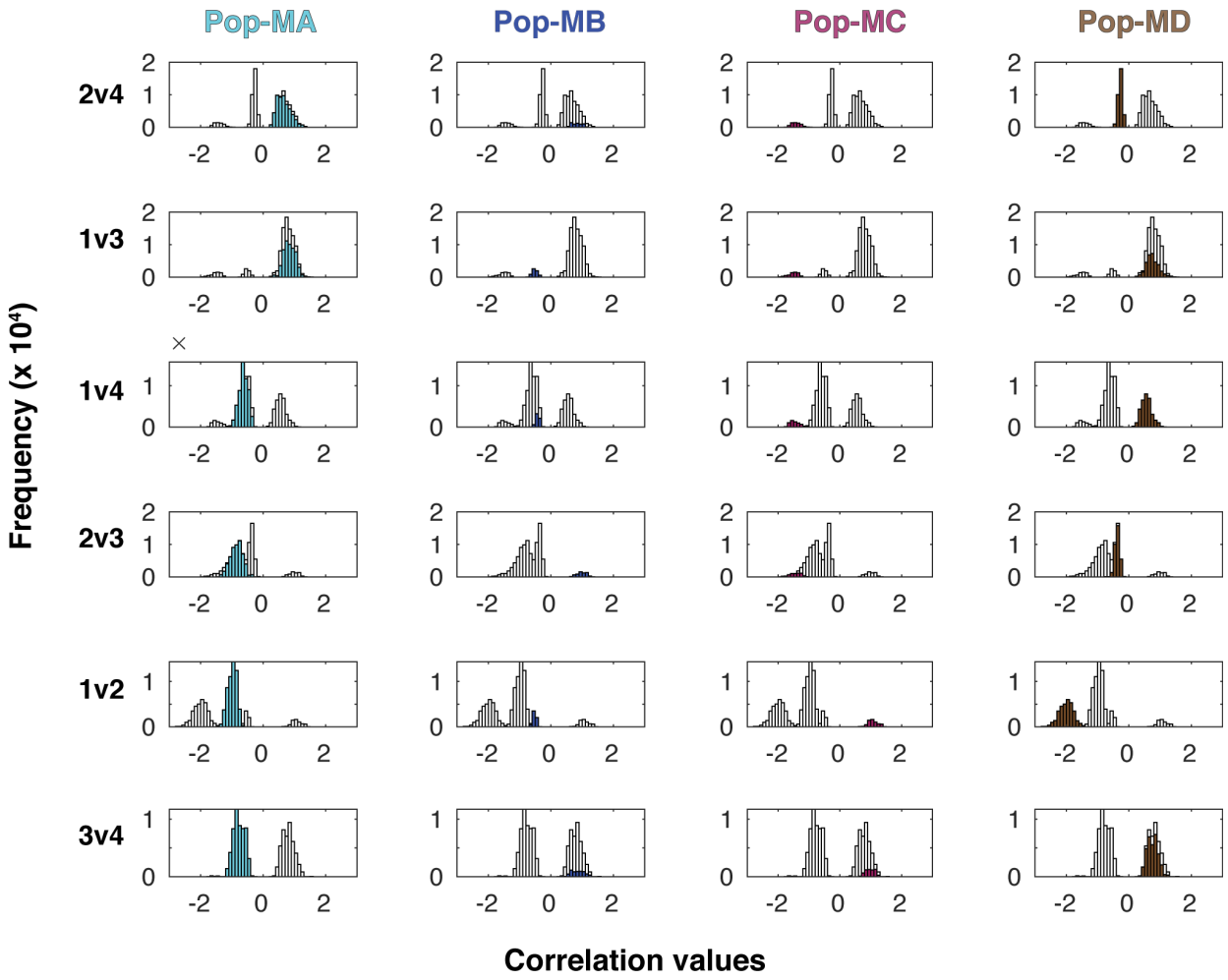


Fig. 4 Four distinct, emergent phenotypes from model runs seeded with no bias. Each row displays one of the six possible trait correlations (2v4, 1v3, 1v4, 2v3, 1v2, and 3v4) with the distribution of the emergent trait correlation values for all individuals in all replicate runs (N=100,000) shown in grey. Highlighted in color in each subplot are the trait correlation values for the individuals belonging to each of the emergent phenotypes, or populations (columns). Each phenotype has a clearly defined set of trait correlation values. For example, the 2v4 mean correlation for Pop-MA was 0.66 +/- 0.22 while the 2v4 mean correlation for pop-MD was -0.28 +/- 0.07. Pop-Ma and Pop-MD were the most accessible phenotypes and so the trait-correlation values associated with these phenotypes had the largest frequency (y-axis, note scale of 10^4).

correlations (Fig. 3). Although some individuals reached a maximum possible fitness of 1 (i.e., the evolutionary end coordinate), the mean population fitness consistently remained below 1 (Fig. 3b). This is due to the fact that the model is simultaneously optimizing multiple traits and their correlations, which inherently introduces small but significant amounts of persistent phenotypic variation. In addition, while average movement of the population was fairly linearly in PC space (Fig. 3a), the trajectory of trait changes was not linear (Fig. 3c).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

289 At the final generation (2000th generation), we examined the distribution of each trait
290 correlation (1v2, 1v3, 1v4, 2v3, 2v4, and 3v4) across all individuals in all replicate runs (1000
291 individuals x 100 replicate runs = 100,000 individuals total). Four distinct phenotypes (i.e., traits
292 + trait correlations for the final population) emerged all with statistically analogous end mean
293 fitness. Figure 4 displays the emergent trait correlations for the four phenotypes (Pop-MA, Pop-
294 MB, Pop-MC, and Pop-MD). As these four phenotypes occurred in the same region of the trait-
295 scape but have distinct trait correlations and, to some extent distinct trait values, we term them
296 'cryptic phenotypes'. In other words, these cryptic phenotypes represent four distinct evolutionary
297 outcomes of different trait correlations + trait values that all converged on the single evolutionary
298 end coordinate in the evolved trait-scape. For some correlations such as trait 1 vs trait 2 (1v2),
299 little to no overlap was observed across each of the 4 phenotypes (Fig. 4, row 5), while for others,
300 several phenotypes shared the same trait correlations. For example, individuals in Pop-MA and
301 Pop-MD shared the same 1v3 correlation (Fig. 4, row 2, columns 1 and 4). In contrast, Pop-MA
302 and Pop-MD have a completely different relationship for 1v4 (Fig. 4, row 3, columns 1 and 4). An
303 example pairwise trait-trait plot is shown in Fig. 5a where the 4 phenotypes can be identified in
304 terms of their trait values. While the four phenotypes distinctly separated in terms of trait 2 and
305 trait 4, there was significant overlap for other traits such as trait 2 vs trait 3 (Supplementary Fig.
306 2). Thus, the cryptic phenotypes shared some trait correlations but diverged in others. These
307 findings are consistent with experimental evolution studies that observed convergent phenotypes
308 derived from a mix of parallel and divergent mutational and transcriptional changes across
309 replicate populations evolving to the same environment [7,45-48]. The emergence of multiple
310 high-fitness phenotypes (e.g., Fig. 5a) occupying a single high fitness area in multivariate space

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

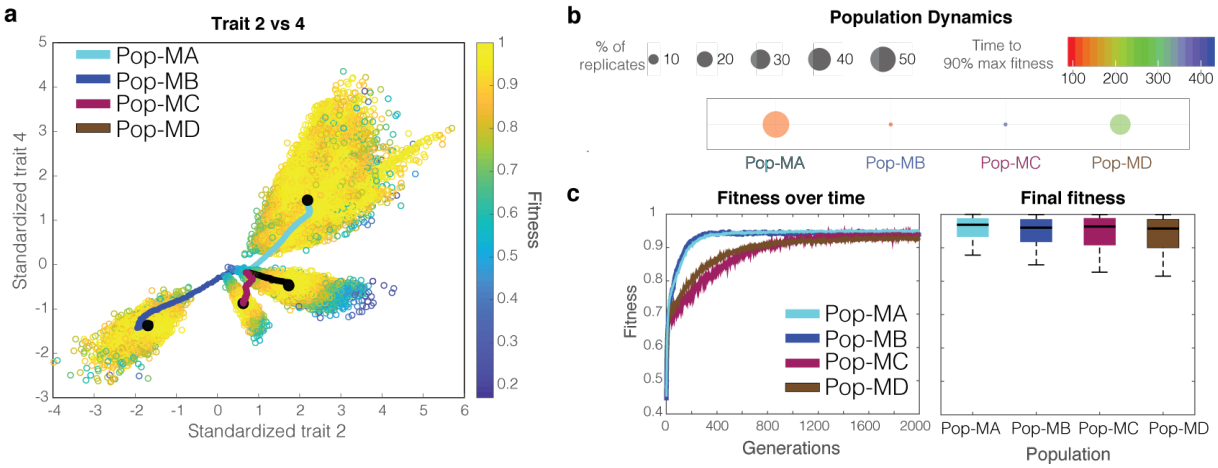


Fig. 5 Representative trait, population, and fitness dynamics for a mixed-mode model run with default model dynamics (90/10) a) Trait vs. trait plot denoting the 4 distinct populations (i.e. phenotypes) that emerged from 100 replicate model runs in mixed mode (i.e., no bias). Each hollow point represents the final trait values of a given individual in the last generation (2000th generation) colored by fitness defined by the trait-scape. Colored lines represent the average trait values at each generation for each population with the black point denoting the final generation. b) Population dynamics of the 4 emergent populations showing the number of replicates (out of 100) that chose specific populations (size of circle) along with each population's rate of adaptation (color of circles). c) The left plot displays the fitness of each population over time while the right displays boxplots representing the distribution of the final fitness values across all individuals of all replicate runs (n = 100,000). Black lines in the boxplots denote the median with the edges denoting the 25th and 75th percentiles, respectively.

demonstrates that our model captures a rugged trait-based fitness landscape with multiple high-fitness peaks (e.g., Fig. 1).

The accessibility of the four phenotypes that emerged from the population without any bias was considerably different. Here we define accessibility as the fraction of replicates that converged on an emergent phenotype. Pop-MA was the most accessible with 55% of replicates converging on this phenotype while Pop-MD was the second most accessible with 33% (Fig. 5b). Pop-MA also exhibited the most variance in trait values within the population (i.e., broadest peak when plotted in more traditional pairwise trait space; e.g., Fig. 1 and Fig. 5a), indicating a relatively larger range of trait values conferring high-fitness with associated Pop-MA's trait correlations. The most accessible phenotype, Pop-MA, also had the fastest rate of fitness gain (Fig. 5b, c), potentially indicating that this phenotype is the most accessible from our experimentally derived

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

starting location. Although Pop-MA and Pop-MB exhibited similar rates of adaptation (Fig. 5c, left plot), Pop-MB was not nearly as accessible with only 6% of the replicates converging on this phenotype (Fig. 5b). Instead, Pop-MD with a slower adaptive rate was the second most accessible phenotype (Fig. 5b, c). Pop-MA and Pop-MD trait correlations were more similar overall than those of Pop-MB.

To examine the impact of ancestral starting point on the emergent phenotypes, we ran the mixed-mode model using a second starting location (i.e., trait values) in the trait-scape that was equidistant to the high fitness area. These model runs converged on 3 of the 4 phenotypes observed with the empirical starting location (Pop-MA, Pop-MB, and Pop-MD). However, shifting the starting location resulted in Pop-MB becoming the most accessible phenotype with the former two most populous phenotypes, Pop-MA and Pop-MD, represented by only 24% and 9% of replicates, respectively (Supplementary Fig. 3). No replicate found Pop-MC and no new phenotypes emerged. These runs indicate that high-fitness areas of the trait-scape were conserved, and that starting an adaptive walk from another location influenced the accessibility of certain phenotypes thereby biasing evolutionary outcomes. The fact that no new populations emerged further supports the ability of this framework to capture the known phenomenon that there are a limited number of accessible phenotypes [28].

We then tested the influence of trait correlational constraints on evolutionary trajectories by randomly changing traits independently of trait correlations (i.e., ignoring trait relationships). Here, every individual experienced a random trait change, but no other traits were updated based on trait correlations resulting in unconstrained movement across the trait-scape. We found that only one phenotype emerged, as expected, but in contrast to the simulations where trait correlations were included (Supplementary Fig. 4; Supplementary text). Here, individuals were unconstrained

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by bias and so were able to quickly move directly to the high-fitness area. This demonstrates that trait correlational constraints can produce different evolutionary strategies (i.e., emergent, cryptic phenotypes), and if constraints are not present, individuals are able to more freely explore phenotypic space and arrive at the high-fitness phenotype more rapidly.

We also tested the sensitivity of the model dynamics to the underlying model assumptions of the frequency of trait and trait correlation changes (default Mode 90/10, 90% of individuals experience a trait change only and 10% experience both a trait + trait correlation change). Specifically, the model was run in two other modes: for every generation, 1) 50% of individuals experienced a trait change and 50% experienced both a trait and correlation change (Mode 50/50) and 2) 10% of individuals experienced a trait change and 90% of individuals experienced both a trait and correlation change (Mode 10/90). All other parameters stayed the same. Mode 50/50 found the same 4 phenotypes as the default Mode 90/10 while Mode 10/90 found the two most accessible phenotypes, Pop-MA and Pop-MD (Supplementary Fig. 5). No new populations emerged. The fact that some combination of the same phenotypes emerged from each of the three independently run modes provides further evidence for a robust, conserved trait-scape with limited high-fitness phenotypes derived from a population with no historical bias.

Adding Historical Bias

The mixed-mode model runs presented above represent a null hypothesis where organisms start with no constraint (e.g., bias) on trait-trait relationships. Next, we assessed the impact of adding trait-correlation bias through the systematic addition of empirical ancestral (sub-modes A1 - A4) and evolved (sub-modes E1 – E4) correlations. For both ancestral and evolved modes, systematically adding more bias (i.e., going from A1 – A4 and E1 – E4, respectively,) changed the accessibility of the high-fitness phenotypes across replicate runs (Fig. 6). In other words, adding

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

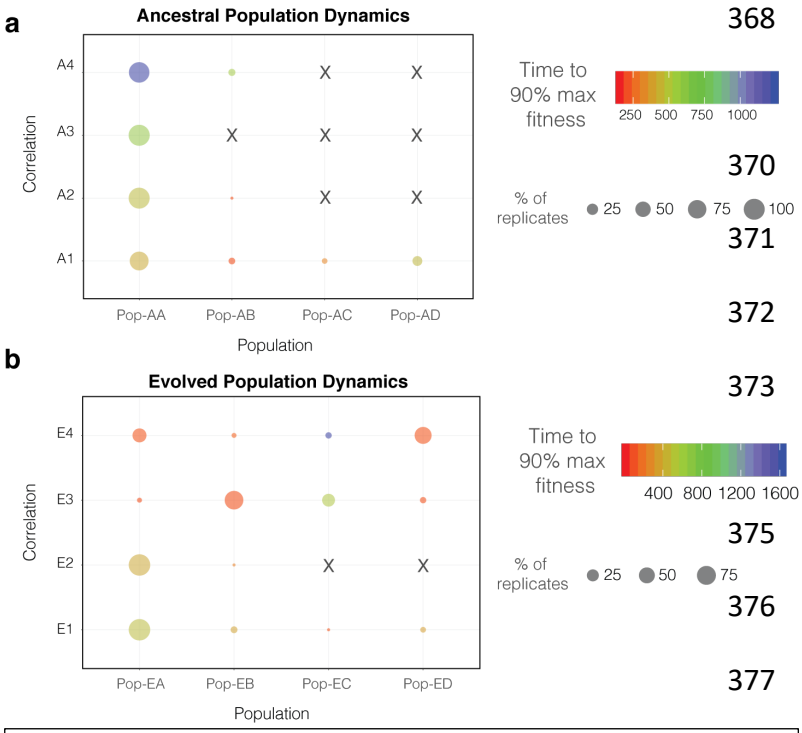


Fig. 6 Dynamics of emergent populations in simulations with different historical bias. a) Bubble plot showing emergent populations as a function of adding empirical ancestral correlations (ancestral bias). Bubble size denotes the number of replicates (out of 100) within a specific population while bubble color represents each population's rate of adaptation (color of circles). b) Same plot as in a) except with adding empirical evolved correlations (evolved bias).

different types of bias influenced adaptive walks across the trait-scape by introducing constraints in the form of trait relationships (e.g., different paths depicted in Fig. 1). However, the type of bias (e.g., ancestral vs evolved correlations) had a different impact on phenotype accessibility. Bias from the ancestral correlations was typically maladaptive and resulted in fewer accessible phenotypes and slower adaptive rates (Fig. 6a). However, bias from the evolved trait relationships (i.e., consistent with the trait-scape) generally resulted in faster adaptive rates and greater overall accessibility to adaptive phenotypes (Fig. 6b). These results are consistent with prior observations where bias (e.g., trait correlations) accelerated adaptive evolution if existing biological orientation aligned with the direction of selection but constrained adaptation if it limited variability in the direction of selection [16,17,28]. Here we show these same modeled dynamics occur when using a collapsed multi-variate trait-space and including trait correlations as a constraint on an adaptive walk. Specifically, depending on a starting population's bias, different phenotypes are more probable than others with some being generally inaccessible as found in other studies [28,29].

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Meta-analysis of phenotypes across different modeled modes

We analyzed results from the mixed (n=1), ancestral (n=4), and evolutionary (n=4) modes to compare emergent phenotypes with different historical biases across independent runs. We assessed the similarity of the high-fitness phenotypes across all model runs (9 runs with 100 replicates each) using hierarchical clustering with multiscale bootstrap resampling (1,000 replicates) on mean trait correlation values at the 2000th generation (Methods). We also included the empirical data from the ancestral and evolved populations in this analysis. Hierarchical clustering revealed 5 high-confidence clusters (I – V) harboring 93% of the phenotypes (n=26 of 28) with approximately unbiased (AU) p-values > 75 (Fig. 7a). Two phenotypes, Pop-MB and Pop-EB-E1, clustered with II and IV, respectively, albeit with less confidence relative to the high-confidence clusters. The empirical ancestral phenotype did not fall within any of the high-confidence clusters, which is expected as the ancestral phenotype was not well-adapted in the

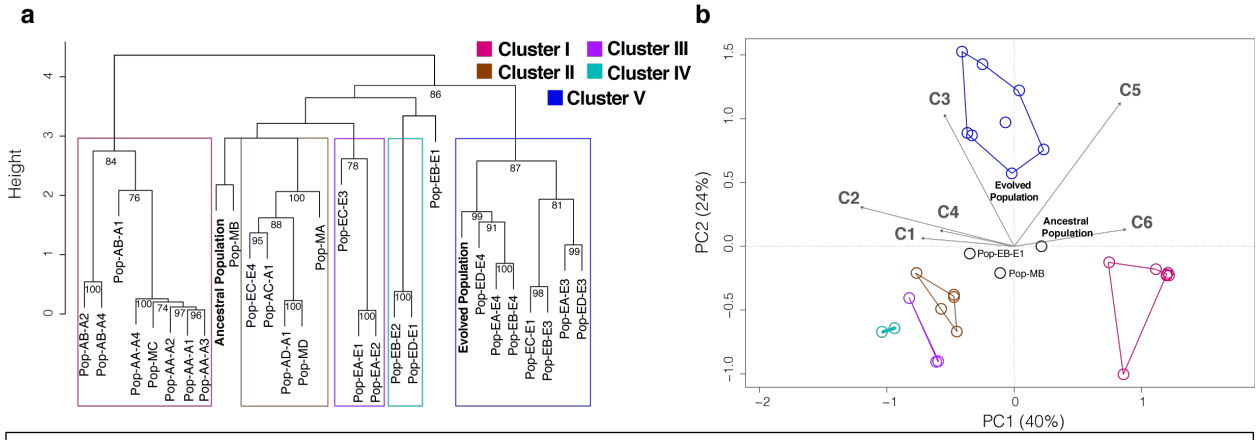


Fig. 7 Hierarchical clustering and Principal Component Analysis of mean trait correlation values calculated across all phenotypes from mixed, ancestral, and evolved mode model simulations a) Hierarchical clustering with multiscale bootstrap resampling (1,000 replicates) on trait correlation values from the emergent phenotypes (Fig. 4) across all model runs (mixed mode, A1-A4, and E1-E4) along with empirical ancestral and evolved correlation values. Approximately Unbiased (AU) p-values > 75 are labeled at the nodes. We identified five overarching clusters with high-confidence AU p-values (colors), which contained even higher confidence sub-clusters. b) Principal component analyses with trait correlation values as in a) with the 5 clusters projected onto the coordinate plane as convex hulls. Percentages on x and y axes denote the percent of explained variance along each axis. Vectors C1-C6 denote correlations 1-6 as defined in Fig. 2C.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

404 evolved trait-scape. In contrast, the empirical evolved population clustered with high-confidence
405 in cluster V, a cluster found by 20% of the replicates including phenotypes from sub-modes E1,
406 E3, and E4 where evolved bias was added. The clustering observed through the hierarchical
407 analysis also emerged through a PC analysis of the population trait correlations. Specifically, we
408 observed 3 general regions of convergence in PC space among the phenotypes, as clusters II, III,
409 and IV collapsed into a small region of the lower left quadrant in the PCA plot (Fig. 7b).
410 Importantly, these convergent regions emerged from thousands of possible trait and correlation
411 values across varying degrees of bias. They provide valuable insight into probable combinations
412 of high-CO_2 adaptive trait correlations along a reduced set of biological axes.
413
414 **Discussion**
415 Here, we combined empirical trait evolution data from a model freshwater alga with a
416 framework that uses eigenvector based methods (principal components) to model multivariate
417 adaptive walks with evolving traits and trait correlations. By leveraging empirical ancestral trait
418 correlations and the observed changes in these correlations as a result of adaptation to high CO_2,
419 we were able to simulate adaptive walks with endpoints anchored in real evolutionary outcomes.
420 The true utility of TRACE lies in its ability to provide insight into multi-trait evolution using trait
421 data from empirical, organismal experiments. Specifically, our model provides a framework for
422 studying the evolution of multiple traits and their potential trade-offs in response to environmental
423 change. This approach generally contrasts with the vast majority of past adaptive walk models that
424 study evolution using hypothetical traits and fitness. Critically, our model captures the same
425 evolutionary phenomena as past models but with a trait-scape characterized by easy-to-quantify
426 and ecologically important traits from globally relevant microbes. From here, we can build on our

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

427 understanding of key sets of multivariate trait relationships under environmental change. This will
428 be critical for determining how evolving microbial processes will influence global
429 biogeochemistry and carbon cycling in the face of global change. For example, TRACE can
430 provide hypotheses as to the degree of evolvability of certain traits and trait correlations under
431 selective gradients (e.g., CO_2) in a multi-trait landscape and suggest potential multivariate trait
432 tradeoffs.

433 We believe that a framework such as TRACE is essential for more robustly predicting
434 biogeochemical shifts as our framework can capture the contrasting responses of microbes in the
435 short- and long-term. Specifically, TRACE allows for the emergence of trade-offs through
436 evolutionary change. For example, experimental evolution study results suggest that microbes
437 might be able to adapt by increasing their carbon use efficiency and/or increasing cell size at
438 warmer temperatures, contrary to their plastic responses [10,18]. However, this adaptive capacity
439 might not be available when considering other stressors, such as nutrient limitation [49]. Therefore,
440 by combining TRACE with experimental evidence on the co-evolution of traits, we can better
441 constrain the adaptive capacity of microbes. These hypotheses can then be tested with targeted
442 laboratory and field experiments and ultimately integrated into larger biogeochemical models to
443 constrain microbial phenotypes and thus trait distributions under different global change scenarios.

444 Applying TRACE to *Chlamydomonas* evolution under high CO_2, we found that a limited
445 set of integrated phenotypes underlie thousands of possible trait correlational scenarios. Upon
446 systematically adding different types (ancestral or evolved) of bias, only certain phenotypes
447 emerged for some trait combinations (e.g., A2 – A4 & E2) while others found all possible
448 phenotypes for a specific mode. These results help elucidate evolutionary trajectories based on
449 trait correlation constraints for ecological and biogeochemical traits of interest. Importantly, they

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

can also help inform future experimental designs aimed at testing the probability of adaptive outcomes across multivariate environments through the analysis of a select set of traits. The combination of both experimental evolution and eigenvector methods like PCA can be a powerful approach to help predict both short- and long-term biological responses to global change. Particularly, this framework can be used to estimate a rugged trait-scape harboring a limited set of phenotypes and identify high-fitness trait-correlation combinations under selective gradients. Due to the seemingly infinite amount of possible interacting biological and environmental variables to test, these evolutionary and mathematical tools that allow us to efficiently combine experiments with modeling will be critical to help predict microbial population responses to future global change scenarios through the lens of evolutionary phenomena.

Data Accessibility

The model code is available at https://github.com/LevineLab

Author contributions

N.G.W., S.C., and N.M.L., and designed research; N.G.W. and N.M.L. performed research; N.G.W., S.C., J.H., P.A.A., M.A.D., and N.M.L. analyzed data; and N.G.W., S.C., J.H., P.A.A., M.A.D., S.G.L., and N.M.L. wrote the paper.

Competing Interests

We declare that we have no competing interests.

Funding This work was supported by the Moore Foundation Grant MMI 7397 (to N.M.L., S.C., M.A.D.) and by The Simons Foundation Grant 509727 (to N.M.L.).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Walworth, N. G., Zakem, E. J., Dunne, J. P., Collins, S. & Levine, N. M. 2020 Microbial evolutionary strategies in a dynamic ocean. *Proc Natl Acad Sci USA*, 1–6. (doi:10.1073/pnas.1919332117)

2. Hinners, J., Hense, I. & Kremp, A. 2019 Modelling phytoplankton adaptation to global warming based on resurrection experiments. *Ecological Modelling* **400**, 27–33. (doi:10.1016/j.ecolmodel.2019.03.006)

3. Lindberg, R. T. & Collins, S. 2020 Quality–quantity trade-offs drive functional trait evolution in a model microalgal 'climate change winner'. *Ecol Lett* **23**, 780–790. (doi:10.1111/ele.13478)

4. Walworth, N. G., Fu, F.-X., Lee, M. D., Cai, X., Saito, M. A., Webb, E. A. & Hutchins, D. A. 2018 Nutrient-Colimited *Trichodesmium* as a Nitrogen Source or Sink in a Future Ocean. *Appl. Environ. Microbiol.* **84**, e02137–17–14. (doi:10.1128/AEM.02137-17)

5. Brennan, G. L., Colegrave, N. & Collins, S. 2017 Evolutionary consequences of multidriver environmental change in an aquatic primary producer. *Proc Natl Acad Sci USA* **114**, 9930–9935. (doi:10.1073/pnas.1703375114)

6. Hutchins, D. A. & Fu, F. 2017 Microorganisms and ocean global change. *Nature Microbiology* **2**, 1–11. (doi:10.1038/nmicrobiol.2017.58)

7. Walworth, N. G., Lee, M. D., Fu, F.-X., Hutchins, D. A. & Webb, E. A. 2016 Molecular and physiological evidence of genetic assimilation to high CO_2 in the marine nitrogen fixer *Trichodesmium*. *Proc. Natl. Acad. Sci. U.S.A.*, 201605202–8. (doi:10.1073/pnas.1605202113)

8. Hutchins, D. A., Walworth, N. G., Webb, E. A., Saito, M. A., Moran, D., McIlvin, M. R., Gale, J. & Fu, F.-X. 2015 Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. *Nat. Commun.* **6**. (doi:10.1038/ncomms9155)

9. Hellweger, F. L., van Sebille, E. & Fredrick, N. D. 2014 Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* **345**, 1346–1349. (doi:10.1126/science.1254421)

10. Schaum, C.-E., Buckling, A., Smirnoff, N., Studholme, D. J. & Yvon-Durocher, G. 2018 Environmental fluctuations accelerate molecular evolution of thermal tolerance in a marine diatom. *Nat. Commun.* **9**, 1–14. (doi:10.1038/s41467-018-03906-5)

11. Ward, B. A., Collins, S., Dutkiewicz, S., Gibbs, S. J., Bown, P., Ridgwell, A., Sauterey, B., Wilson, J. D. & Oschlies, A. 2019 Considering the role of adaptive

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

evolution in models of the ocean and climate system. *Journal of Advances in Modeling Earth Systems*, 1–29. (doi:10.31223/osf.io/srdh3)

12. Schluter, L., Lohbeck, K. T., Gro ger, J. P., Riebesell, U. & Reusch, T. B. H. 2016 Long-term dynamics of adaptive evolution in a globally important phytoplankton species to ocean acidification. *Science Advances* **2**, e1501660–e1501660. (doi:10.1126/sciadv.1501660)

13. Beckmann, A., Schaum, C.-E. & Hense, I. 2019 Phytoplankton adaptation in ecosystem models. *Journal of Theoretical Biology* **468**, 60–71. (doi:10.1016/j.jtbi.2019.01.041)

14. Boyd, P. W., Cornwall, C. E., Davison, A., Doney, S. C., Fourquez, M., Hurd, C. L., Lima, I. D. & McMinn, A. 2016 Biological responses to environmental heterogeneity under future ocean conditions. *Global Change Biol* **22**, 2633–2650. (doi:10.1111/gcb.13287)

15. Boyd, P. W. et al. 2018 Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change-A review. *Global Change Biol* **24**, 2239–2261. (doi:10.1111/gcb.14102)

16. Aguirre, J. D., Hine, E., McGuigan, K. & Blows, M. W. 2014 Comparing G: multivariate analysis of genetic variation in multiple populations. *Heredity* **112**, 21–29. (doi:10.1038/hdy.2013.12)

17. Agrawal, A. A. 2019 A scale-dependent framework for trade-offs, syndromes, and specialization in organismal biology*. *Ecology* (doi:doi.org/10.1002/ecy.2924)

18. Barton, S., Jenkins, J., Buckling, A., Schaum, C.-E., Smirnoff, N., Raven, J. A. & Durocher, G. Y. 2020 Evolutionary temperature compensation of carbon fixation in marine phytoplankton. *Ecol Lett* **23**, 722–733. (doi:10.1111/ele.13469)

19. Brandenburg, K. M., Wohlrab, S., John, U., Kremp, A., Jerney, J., Krock, B. & Van de Waal, D. B. 2018 Intraspecific trait variation and trade-offs within and across populations of a toxic dinoflagellate. *Ecol Lett* **21**, 1561–1571. (doi:10.1111/ele.13138)

20. Malcom, J. W., Hernandez, K. M., Likos, R., Wayne, T., Leibold, M. A. & Juenger, T. E. 2014 Extensive cross-environment fitness variation lies along few axes of genetic variation in the model alga, *Chlamydomonas reinhardtii*. *New Phytol.* **205**, 841–851. (doi:10.1111/nph.13063)

21. Moore, L. R., Post, A. F., Rocap, G. & Chisholm, S. W. 2002 Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* **47**, 989–996. (doi:10.4319/lo.2002.47.4.0989)

22. Walsh, B. 2014 Special issues on advances in quantitative genetics: introduction. *Heredity*, 1–3. (doi:10.1038/hdy.2013.115)

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

23. Anderson, R. W. 1995 Learning and evolution: A quantitative genetics approach. *Journal of Theoretical Biology* **175**, 89–101. (doi:10.1006/jtbi.1995.0123)

24. Tenaillon, O. 2014 The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annu. Rev. Ecol. Evol. Syst.* **45**, 179–201. (doi:10.1146/annurev-ecolsys-120213-091846)

25. Martin, G. & Lenormand, T. 2015 The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. *Evolution* **69**, 1433–1447. (doi:10.1111/evo.12671)

26. Parter, M., Kashtan, N. & Alon, U. 2008 Facilitated Variation: How Evolution Learns from Past Environments To Generalize to New Environments. *PLoS Comput Biol* **4**, e1000206. (doi:10.1371/journal.pcbi.1000206)

27. Kashtan, N., Noor, E. & Alon, U. 2007 Varying environments can speed up evolution. *Proc Natl Acad Sci USA* **104**, 13711–13716. (doi:10.1073/pnas.0611630104)

28. Uller, T., Moczek, A. P., Watson, R. A., Brakefield, P. M. & Laland, K. N. 2018 Developmental Bias and Evolution: A Regulatory Network Perspective. *Genetics* **209**, 949–966. (doi:10.1534/genetics.118.300995)

29. Elhanan Borenstein, D. C. K. 2008 An End to Endless Forms: Epistasis, Phenotype Distribution Bias, and Nonuniform Evolution. *PLoS Comput Biol* **4**, e1000202. (doi:10.1371/journal.pcbi.1000202)

30. Gomulkiewicz, R. & Houle, D. 2009 Demographic and Genetic Constraints on Evolution. *The American Naturalist* **174**, E218–E229. (doi:10.1086/645086)

31. Braendle, C., Baer, C. F. & Félix, M.-A. 2010 Bias and Evolution of the Mutationally Accessible Phenotypic Space in a Developmental System. *PLoS Genet* **6**, e1000877. (doi:10.1371/journal.pgen.1000877)

32. Houle, D., Bolstad, G. H., van der Linde, K. & Hansen, T. F. 2017 Mutation predicts 40 million years of fly wing evolution. *Nature* **548**, 447–450. (doi:10.1038/nature23473)

33. Pavlicev, M., Cheverud, J. M. & Wagner, G. P. 2011 Evolution of adaptive phenotypic variation patterns by direct selection for evolvability. *Proc. R. Soc. B* **278**, 1903–1912. (doi:10.1098/rspb.2010.2113)

34. Monroe, J. G., Markman, D. W., Beck, W. S., Felton, A. J., Vahsen, M. L. & Pressler, Y. 2018 Ecoevolutionary Dynamics of Carbon Cycling in the Anthropocene. *Trends in Ecology & Evolution* **33**, 213–225. (doi:10.1016/j.tree.2017.12.006)

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

582 35. Baltar, F. et al. 2019 Towards Integrating Evolution, Metabolism, and Climate
583 Change Studies of Marine Ecosystems. *Trends in Ecology & Evolution* **34**, 1022–
584 1033. (doi:10.1016/j.tree.2019.07.003)

585 36. Kronholm, I. & Collins, S. 2015 Epigenetic mutations can both help and hinder
586 adaptive evolution. *Mol Ecol* (doi:10.1111/mec.13296)

587 37. Fisher, R. A. 1930 *The Genetical Theory of Natural Selection*. Oxford.

588 38. Suzuki, R. & Shimodaira, H. 2006 Pvclust: an R package for assessing the
589 uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542.
590 (doi:10.1093/bioinformatics/btl117)

591 39. Oksanen, J., Blanchet, F. G. & Kindt, R. 2013 vegan: Community Ecology Package.

592 40. Kirkpatrick, M. 2008 Patterns of quantitative genetic variation in multiple
593 dimensions. *Genetica* **136**, 271–284. (doi:10.1007/s10709-008-9302-6)

594 41. Lenski, R. E. 2017 Convergence and Divergence in a Long-Term Experiment with
595 Bacteria. *The American Naturalist* **190**, S57–S68. (doi:10.1086/691209)

596 42. Avrani, S., Bolotin, E., Katz, S. & Hershberg, R. 2017 Rapid Genetic Adaptation
597 during the First Four Months of Survival under Resource Exhaustion. *Molecular*
598 *Biology and Evolution* **34**, 1758–1769. (doi:10.1093/molbev/msx118)

599 43. Cooper, V. S. & Lenski, R. E. 2000 The population genetics of ecological
600 specialization in evolving *Escherichia coli* populations. *Nature* **407**, 736–739.
601 (doi:10.1038/35037572)

602 44. Elena, S. F. & Lenski, R. E. 2003 Microbial genetics: Evolution experiments with
603 microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* **4**,
604 457–469. (doi:10.1038/nrg1088)

605 45. Jerison, E. R., Ba, A. N. N., Desai, M. M. & Kryazhimskiy, S. 2020 Chance and
606 necessity in the pleiotropic consequences of adaptation for budding yeast. *Nature*
607 *Ecology & Evolution*, 1–20. (doi:10.1038/s41559-020-1128-3)

608 46. Travisano, M., Vasi, F. & Lenski, R. E. 1995 Long-Term Experimental Evolution in
609 *Escherichia coli*. III. Variation Among Replicate Populations in Correlated
610 Responses to Novel Environments. *Evolution* **49**, 189–200.

611 47. Nakatsu, C. H., Korona, R., Lenski, R. E., de Bruijn, F. J., Marsh, T. L. & Forney,
612 L. J. 1998 Parallel and Divergent Genotypic Evolution in Experimental Populations
613 of *Ralstonia* sp. *J. Bacteriol.* **180**, 4325–4331.

614 48. Fong, S. S. 2005 Parallel adaptive evolution cultures of *Escherichia coli* lead to
615 convergent growth phenotypes with different gene expression states. *Genome*
616 *Research* **15**, 1365–1372. (doi:10.1101/gr.3832305)

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

617 49. Gassis, M. A., Kremer, C. T., Klausmeier, C. A. & Litchman, E. 2019 Nitrogen
618 limitation inhibits marine diatom adaptation to high temperatures. *Ecol Lett* **22**,
619 1860–1869. (doi:10.1111/ele.13378)

620

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The evolution of trait correlations constrains phenotypic adaptation to high CO_2 in a eukaryotic alga

Supplementary Information

Nathan G. Walworth[1], Jana Hinners[2], Phoebe Argyle[3], Suzana G. Leles[1], Martina A. Doblin[3], Sinéad Collins[2], Naomi M. Levine[1]*

[1]Department of Biological Sciences, University of Southern California, Los Angeles, California 90089-0371, USA, [2] Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom, [3] Climate Change Cluster, University of Technology Sydney, Sydney, NSW 2007, Australia

Corresponding author

e-mail: n.levine@usc.edu

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Material

Adaptation and population size

Adaptation dynamics in our model framework follow those from our previous work [1,2] where populations initially start far from the optimum and are subjected to predominantly non-neutral changes to traits and/or correlations. We do not explicitly represent drift in the model, which can also impact the adaptive process as each change (e.g., mutation) occurs with an initial frequency of 1/N and can be lost by chance. However, the relative supply of changes (population size x number of changes per generation) used in our model runs is sufficient for selection to overwhelm drift resulting in robust evolutionary results. Accordingly, in prior studies using an analogous adaptive process, we varied population size and/or selection strength over multiple orders of magnitude and demonstrated adaptive outcomes to be robust over the same selection period [1,2].

Sensitivity tests

We conducted several sensitivity tests to examine model dynamics. First, we tested if the the sequence of correlational changes influenced adaptive outcomes in our model. To do this, we conducted model runs where we changed the order in which traits were updated and showed that phenotypic results remained unchanged as expected (Supplementary Fig. 1).

We also examined the impact of trait correlational constraints on evolutionary trajectories (Supplementary Fig. 4) by randomly changing traits independently of trait correlations (i.e. ignoring trait relationships). For these runs, each individual experienced a random trait change, but no other traits were updated. Hence, individuals were unconstrained by bias and so were able to quickly move directly to the high-fitness area. We found that only one phenotype emerged, as

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expected (Supplementary Fig. 4). This demonstrates that trait correlational constraints are critical for producing different evolutionary strategies (i.e. emergent, cryptic phenotypes), and if constraints are not present, individuals are able to more freely explore phenotypic space and arrive at the high-fitness phenotype more rapidly.

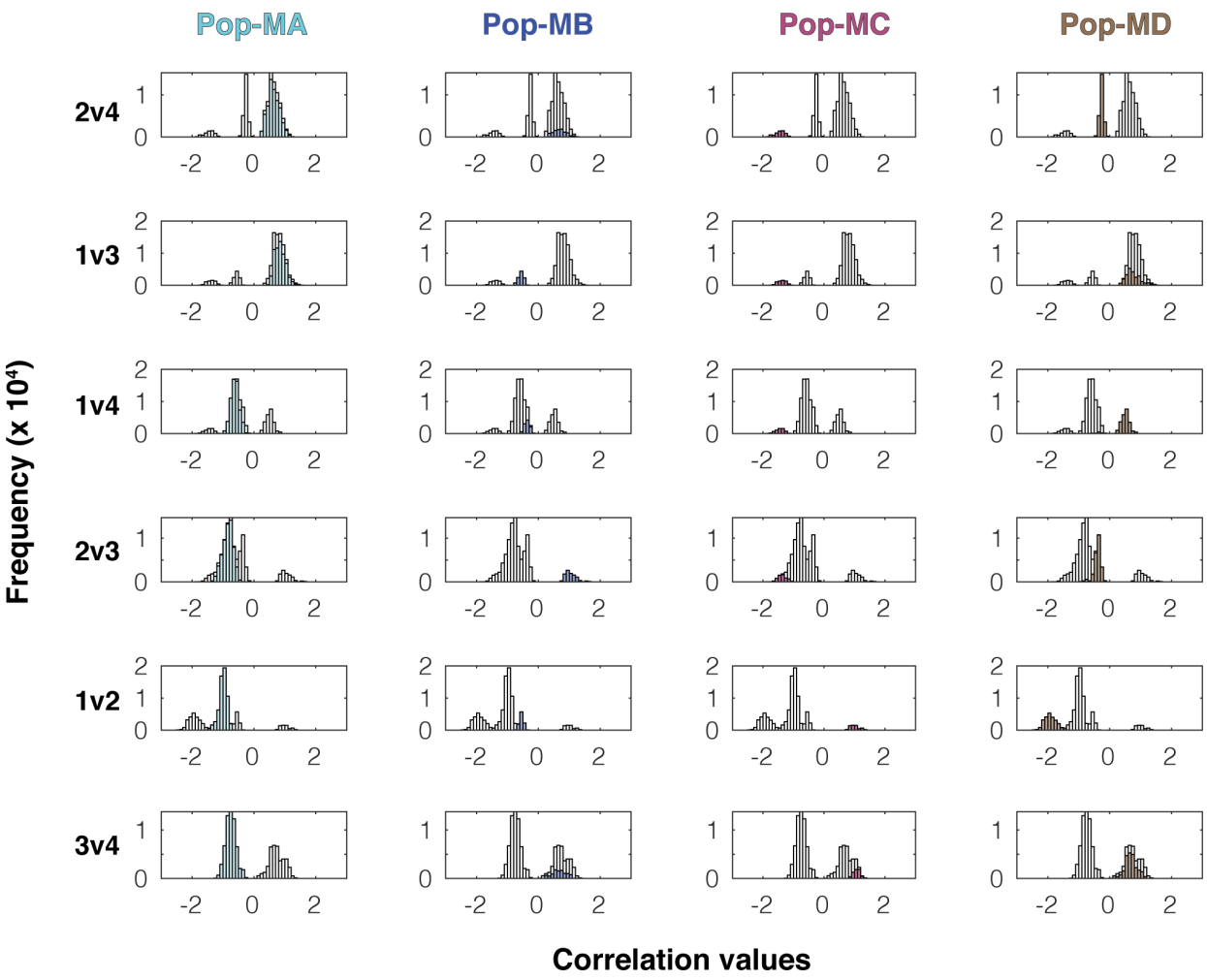
Finally, we tested the sensitivity of the model dynamics by varying the ratios of trait and trait correlation changes (Supplementary Fig 5). In addition to the default Mode 90/10 (90% of individuals experience a trait change only and 10% experience both a trait + trait correlation change) two other modes were run: Mode 50/50 and Mode 10/90. For each generation in Mode 50/50, 50% of individuals experienced a trait change and 50% experienced both a trait and correlation change. For Mode 10/90, 10% of individuals experienced a trait change and 90% of individuals experienced both a trait and correlation change. All other parameters stayed the same. Mode 50/50 found the same 4 phenotypes as the default Mode 90/10 while Mode 10/90 found the two most accessible phenotypes, Pop-MA and Pop-MD (Supplementary Fig. 5). No new populations emerged. The fact that some combination of the same phenotypes emerged from each of the three independently run modes provides further evidence for a robust, conserved trait-scape with limited high-fitness phenotypes derived from a population with no historical bias.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 1 Parameter values used for main text model simulations.

Parameter	Description	Value
N	Size of population	1000
t	Number of generations	2000
tgrad	Standard deviation of trait change	0.05
cgrad	Standard deviation of correlation change	0.05
$\mathbf{N_{trait}}$	Number of trait changes	[900, 500, 100]
$\mathbf{N_{corr}}$	Number of correlation changes	[900, 500, 100]
$\mathbf{N_{runs}}$	Number of replicate runs	100

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

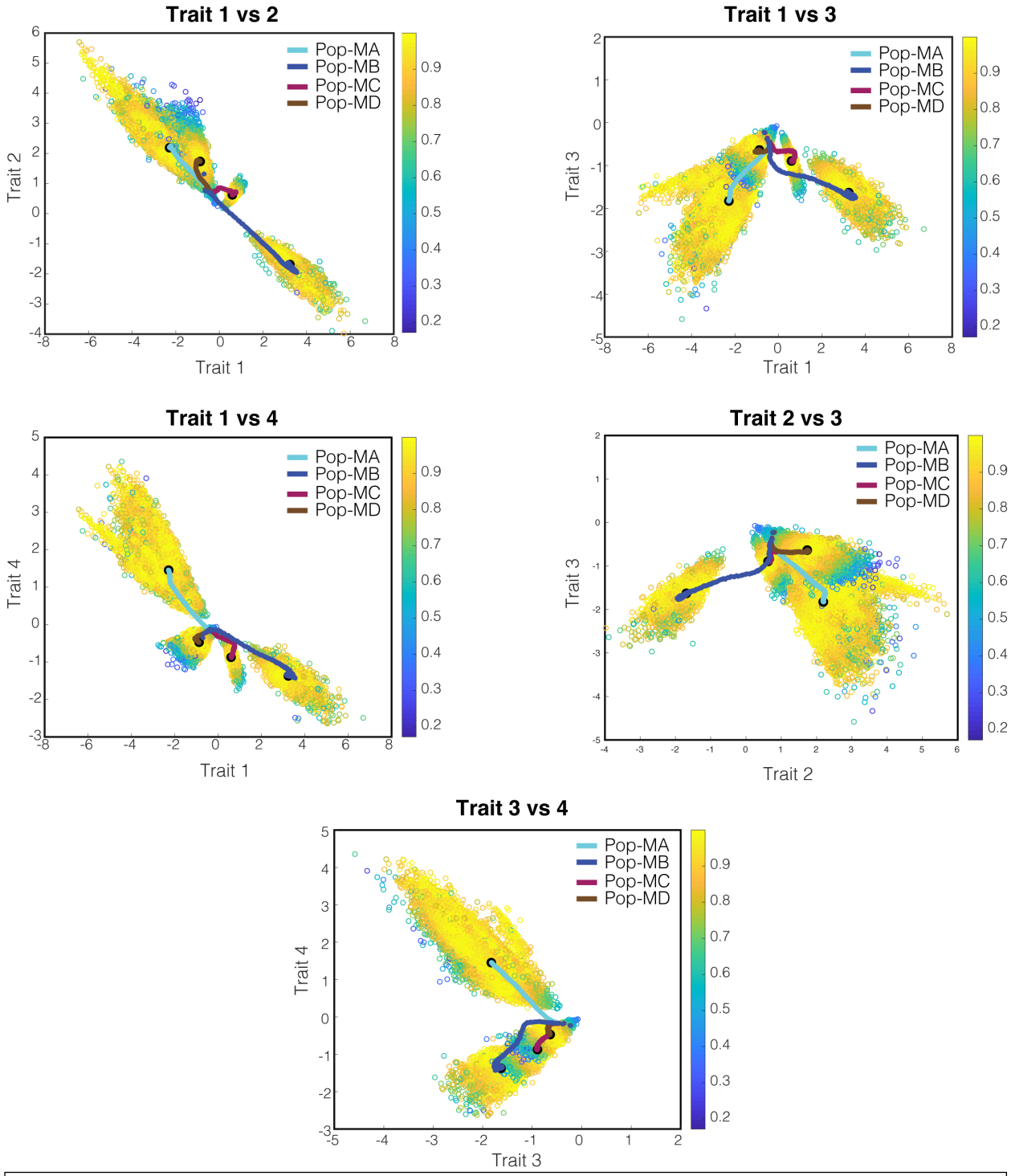


Supplementary Fig. 1 Four distinct, emergent phenotypes (i.e. populations) from model runs seeded with no bias These model runs were identical to "Mixed mode" in the main text (Fig. 4) except the order of updating traits and correlations was reversed. Each row displays one of the six possible trait correlations (2v4, 1v3, 1v4, 2v3, 1v2, and 3v4) with the distribution of the emergent trait correlation values for all individuals in all replicate runs (N=100,000) shown in grey (y-axis, note scale of 10^4). Highlighted in color in each subplot are the trait correlation values for the individuals belonging to each of the emergent phenotypes, or populations (columns). Each phenotype has a clearly defined set of trait correlation values. Here, the 4 same populations emerged as in Fig. 4 of the main text.

85

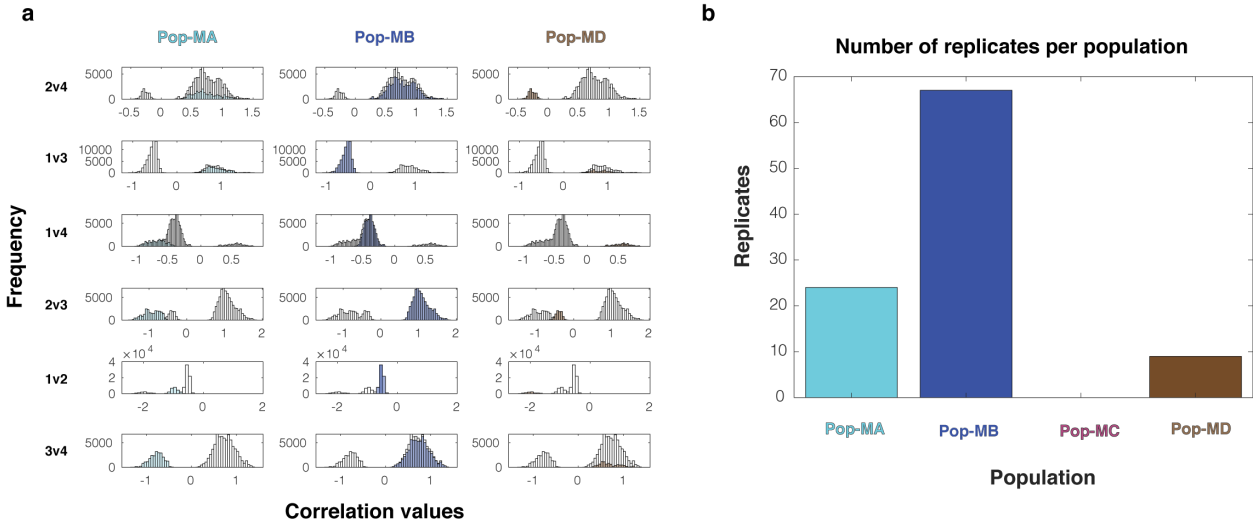
bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

86



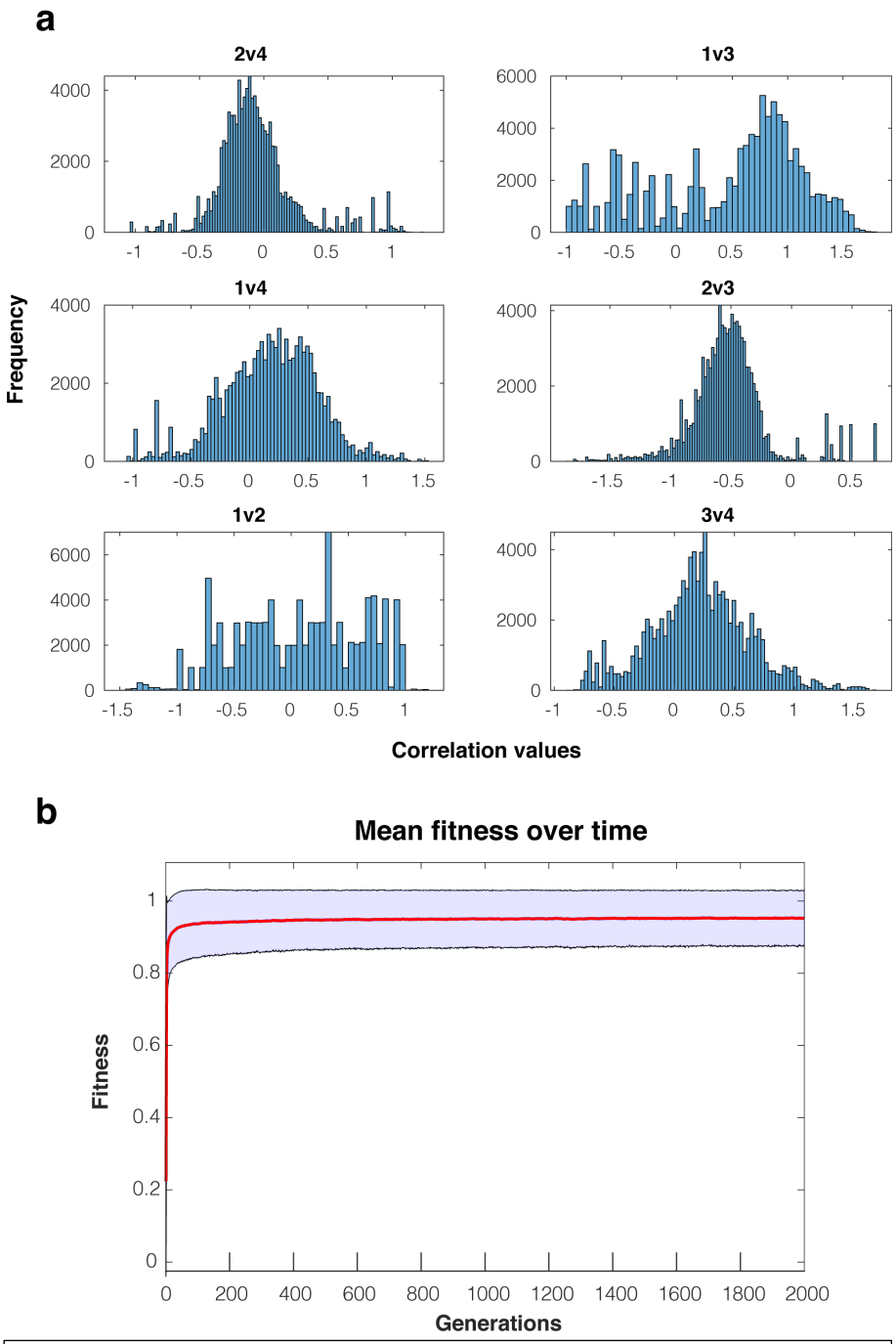
Supplementary Fig. 2 Trait vs. trait plots for mixed mode simulations a) Trait vs. trait plots denoting the 4 distinct populations (i.e. phenotypes) that emerged from 100 replicate model runs in mixed mode (i.e., no bias). Each hollow point represents the final trait values of a given individual in the last generation (2000th) colored by fitness. Colored lines represent the average trait values at each generation for each population with the black point denoting the final generation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



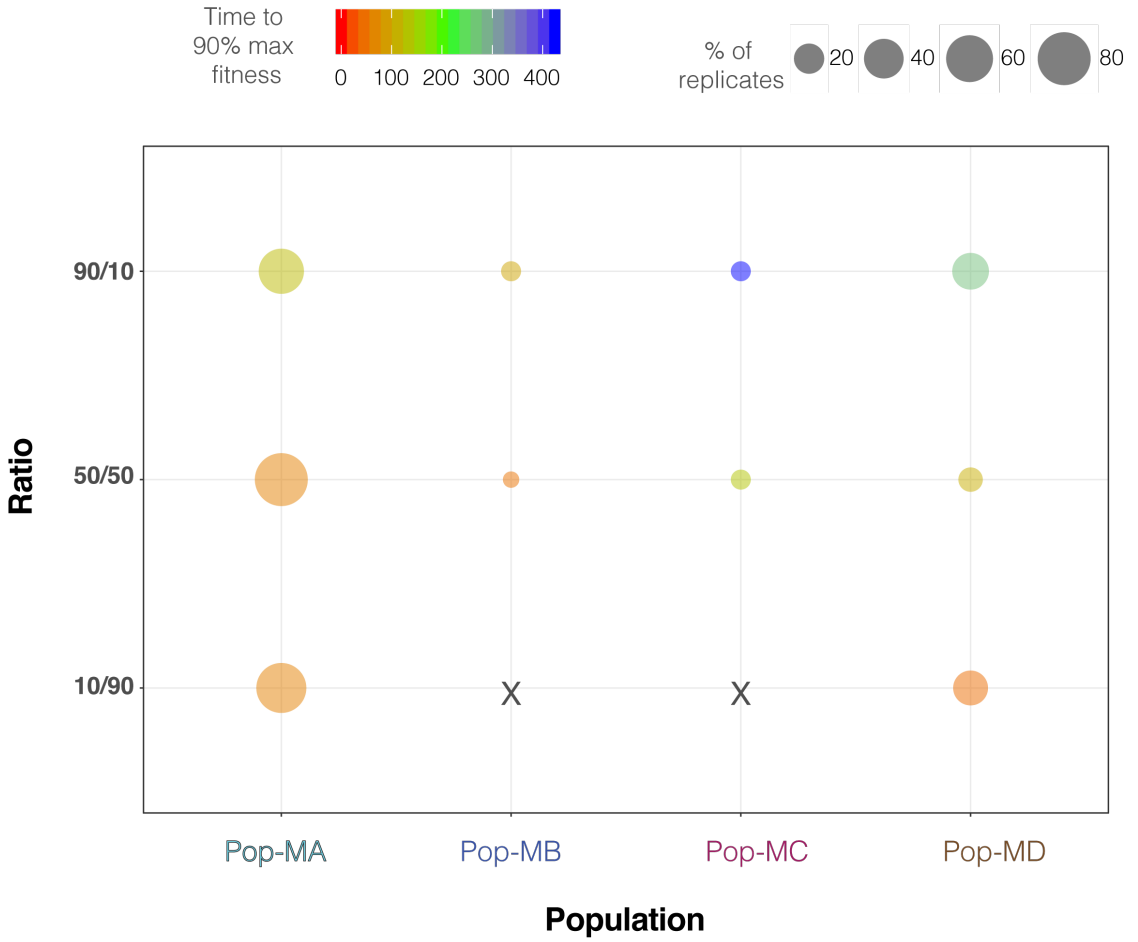
Supplementary Fig. 3 Impact of starting location on emergent phenotypes (i.e. populations) in mixed-mode model simulations. a) Shown are emergent populations resulting from mixed-mode model simulations as in the main text except starting from a different, equidistant ancestral phenotype in PCA space. Each row displays one of the six possible trait correlations (2v4, 1v3, 1v4, 2v3, 1v2, and 3v4) with the distribution of the emergent trait correlation values for all individuals in all replicate runs (N=100,000) shown in grey. Highlighted in color in each subplot are the trait correlation values for the individuals belonging to each of the emergent phenotypes, or populations (columns). Each phenotype has a clearly defined set of trait correlation values. b) Shown is a bar plot denoting the number of replicates within each population (out of 100 model runs). When starting from a different ancestral start point, Pop-MC did not emerge in any replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Fig. 4 Trait correlational distributions from model runs that changed traits independently of trait correlations. a) Shown are trait correlational distributions per correlation. At each generation, 1 trait per individual was randomly changed but no other traits were updated using the trait correlations resulting in one phenotypic solution. b) Shown is the mean fitness of all individuals over time across all replicate runs (1000 individuals x 100 replicate runs). The red line indicates the mean while the light blue area denotes the smoothed standard deviation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Fig. 5 Population dynamics of emergent populations across different ratios of trait and trait correlation changes. The bubble plot shows emergent populations as a function of varying the ratio of trait and trait correlation changes for mixed-mode simulations. 90/10 denotes 90% of individuals in a population experiencing a trait change and 10% both a trait and correlation change per generation. Similarly, 50/50 denotes 50% experiencing a trait change and 50% both a trait and correlation change. Finally, 10/90 denotes 10% experiencing a trait change and 90% both a trait and correlation change. Bubble size denotes the number of replicates (out of 100) within a specific population while bubble color represents each population's rate of adaptation (color of circles), or time to reach 90% of its maximum fitness.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.04.237230; this version posted December 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

145 **References**

146 1. Kronholm, I. & Collins, S. 2015 Epigenetic mutations can both help and hinder adaptive
147 evolution. *Mol Ecol* (doi:10.1111/mec.13296)

148 2. Walworth, N. G., Zakem, E. J., Dunne, J. P., Collins, S. & Levine, N. M. 2020 Microbial
149 evolutionary strategies in a dynamic ocean. *Proc Natl Acad Sci USA*, 1–6.
150 (doi:10.1073/pnas.1919332117)

151