Title

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

Running Title

Multi-trait adaptation

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

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 integrating multivariate trait data into a predictive model of trait evolution. 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 the direction of historical bias (existing trait correlations) influenced both the rate of adaptation and the evolved phenotypes (trait combinations). Critically, we use fitness landscapes derived directly from empirical trait values to capture known evolutionary phenomena. This work demonstrates that ecological models need to represent both changes in traits and changes in the correlation between traits in order to accurately capture phytoplankton evolution and predict future shifts in elemental cycling.

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

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 role 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. We still have only a limited understanding of how microbial communities will respond to multi-stressor and fluctuating environmental change. Additionally, 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 framework that can both help guide experiments and make more accurate future predictions [5].

Seminal research in quantitative genetics has investigated the impact of trait variation, genotypic variability, inheritance, epistasis, and environmental variability on adaptive walks using multivariate and eigenvector methods 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. Other theoretical approaches emphasized the role of de novo mutations in a fitness landscape without accounting for standing genetic variation [21,22]. While these studies have generated evolutionary theory [22,23] relevant for marine microbes, the gulf between evolution literature and biological oceanography has widened such that current state-of-the-art evolutionary models are not suited to interface with biogeochemical ODE models [24].

Empirical data has also demonstrated that biological systems 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 [25,26]. These

70 findings contrast with the long-held assumption of isotropic (i.e. equal) variation [27] and have
71 revealed instead that only a limited part of multivariate phenotypic space (i.e. only certain
72 phenotypes) can be accessed [18,20,23,25-28]. Fig. 1 shows an illustrative example of how the
73 accessibility of high-fitness phenotypes (i.e. peaks in a fitness landscape) differs depending on the
74 starting location (ancestral phenotype) and the initial trajectory, which is dictated by a population's
75 collective genetic architecture (Fig. 1, magenta and black circles and paths). In summary, a
76 growing body of literature has shown that genetic architecture influences how traits and trait
77 correlations are impacted by environmental shifts and that these shifts produce nonrandom
78 distributions of phenotypes [29-31]. These studies have substantial consequences for
79 understanding future shifts in marine phytoplankton function and thus carbon cycling and global
80 climate [32,33]. However, there have been few attempts to investigate the implications of this
81 phenomenon for the evolution of trait and trait correlations of photosynthetic microbes [19,20].

82 We introduce a framework for understanding how evolutionary trajectories of phenotypes
83 (suites of traits) in microbial populations are impacted by the evolutionary relationships between
84 traits (historical bias) and the evolution of trait relationships. Here we define bias as standing trait
85 correlations (i.e., relationships) in a population that are heritable and can impact fitness such that,
86 over time, these correlations can constrain the direction of evolution [16]. Critically, this
87 framework is based on measurable trait values and so does not require the quantification of the
88 fitness impacts of mutations. This is important as most marine microbes are not genetically
89 accessible, and mapping the fitness effects of individual genetic variants is resource intensive. Our
90 TRAit Correlation Evolution (TRACE) model is a first step towards investigating how correlated
91 metabolic traits with clear biogeochemical significance may impact elemental cycling under
92 environmental change (e.g., ocean acidification). Specifically, TRACE generates a trait-based

fitness landscape based on empirical trait data thus facilitating future integration into global biogeochemical models, which model marine microbes using trait-based functional groups. Our results indicate that populations harboring trait correlations oriented in (i.e. consistent with) the direction of selection may experience accelerated rates of adaptation. We suggest that incorporating these dynamics into biogeochemical models will be important for accurately predicting the impact of microbial adaptation on rates of biogeochemical cycling in the ocean.

Materials and Methods

Trait-based fitness landscape:

We generated a multi-variate trait-based fitness landscape (trait-scape) using principal component analyses (PCA). Specifically, a trait-scape was created using four independent ecologically relevant traits (growth rate, respiration, cell size, and daughter cell production) from 5 genotypes of high-CO_2 adapted *Chlamydomonas reinhardtii* [3]. PCA was conducted on standardized evolved traits resulting in a total of 86% of variance explained on two axes, 54% and 32% on axes PC1 and PC2, respectively (Fig. 2a). 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; Supplementary File 1) and the corresponding evolved population trait values defined the evolutionary endpoint (red circle in Fig. 2b; Supplementary File 1). Additional simulations were conducted in which the start point was varied. As the specific traits themselves were 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.

116 *TRACE Model Dynamics*

117 The TRAit Correlation Evolution (TRACE) model framework simulates the adaptive walk
118 of a microbial population across a trait landscape (trait-scape) towards a high-fitness area. TRACE
119 was adapted from an individual based Fisher model of adaptation [1,34,35]. Each generation, each
120 individual in the population experienced either a change in traits or changes in traits and trait-
121 correlations. Changes in trait values moved these individuals across the trait-scape while trait-
122 correlations constrained the direction of movement. Selection was imposed based on distance to
123 the evolutionary end-point in the trait-scape (described below), such that the population evolved
124 towards the high fitness region of the trait-scape. In essence, this framework selected for
125 individuals with the smallest overall difference across all trait values from the empirically observed
126 high fitness phenotype. The weighting of the traits was derived from the observed evolved
127 phenotypes evaluated using PCA, such that traits that were not observed to play an important role
128 in fitness in the high-CO_2 environment had low weight. It is important to note that the model did
129 not directly select for trait correlations, but that specific correlations emerged in the population if
130 they provided a fitness advantage in terms of trait dynamics.

131 In the default model simulations (referred to as 90/10), in each generation 90% of
132 individuals were randomly chosen to experience a random change in a trait value (while
133 maintaining all existing trait-correlations) while 10% experienced both a trait and trait-correlation
134 change. These changes were drawn from a Gaussian distribution (mean = 0 and standard deviation
135 = 0.05) such that small changes were common and large changes were rare. For each individual,
136 the randomly chosen trait change was added to the existing trait value. Following this first trait
137 change, the remaining 3 trait values were updated using the trait correlations for that individual in
138 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).

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. Finally, the second trait tied to the correlation was updated using the new correlation and trait value (the other modeled trait values were not updated in this generation). These trait changes moved the individuals within the trait-scape.

Selection was imposed using the Euclidean distance of each individual to the high fitness area (z_i). Specifically, the fitness of each for each individual (w_i) was calculated as [1,34]:

$$w_i = e^{-z_i^2/2} \qquad \text{Eq. 1}$$

Individuals were then randomly sampled with replacement weighted by w_i such that high fitness individuals were more likely to persist to the next generation, as in [1,34].

Model simulations:

The model trait-scape and the high fitness area (red circle in Fig. 2b) were defined based on empirical data from the *Chlamydomonas* long-term evolution study [3]. The model was initialized with a population of 1000 individuals with either 1) all individuals containing the same trait values corresponding to the empirically observed ancestral trait values (tan circle in Fig. 2b), 2) all individuals containing the same trait values corresponding to an alternative starting location, or 3) a mixed population with multiple different starting trait values based on the empirical data. To

162 explore the impact of historical bias, trait-correlations in the starting populations were initialized
163 in three different ways: mixed mode, ancestral mode, and evolved mode (described below). The
164 modes differed only in the initial conditions not in model dynamics. Each model run was conducted
165 for 2000 generations with 100 replicates each. All model parameters are given in Supplementary
166 Table 1. Previous work by us and others have demonstrated that adaptive outcomes using this
167 framework are robust across a wide range of population sizes (Supplementary information S1)
168 [1,34]. Several sensitivity studies were conducted to test model dynamics (see Supplementary
169 information S2).
170
171 *Mixed mode (no bias) simulations:* To test all possible adaptive walks between the ancestral start
172 point and the evolutionary end point, simulations were conducted with no historical bias.
173 Specifically, each individual was assigned correlation values randomly chosen from a standard
174 uniform distribution over the interval (-1,1). Hence, every individual started with the same 4 trait
175 values but completely random correlation values. See Supplementary information S3 for additional
176 model detail.
177 *Ancestral mode simulations:* To test the effects of systematically adding ancestral bias, four
178 ancestral sub-modes were conducted: A1, A2, A3, and A4. For sub-mode A1, random correlation
179 values were generated as above for 5 of the 6 trait correlations, and one empirical ancestral
180 correlation was used for all individuals. For A2, A3 and A4, all steps were the same except that
181 two, three, and four empirical ancestral correlations were added back, respectively. For simplicity,
182 we chose to sequentially add in ancestral correlations based on the empirically calculated
183 significant trait-correlations from most significant to least significant (R^2 = -0.89 to 0.54; see Fig.
184 2C).

185 *Evolved mode simulations*: The same procedure was conducted for the evolved mode but instead
186 empirical evolved correlations were systematically added (modes E1 – E4).
187 *Hierarchical Clustering*
188 Hierarchical clustering with multiscale bootstrap resampling (1,000 replicates) on mean trait
189 correlation values was conducted using R package pvclust [36] using Euclidean distance and the
190 average (UPGMA) method. Principal component analysis using mean correlation values was
191 conducted with package R package vegan [37], and pvclust clusters with approximately unbiased
192 (AU) p-values > 75% were projected onto the PC coordinate plane as convex hulls.
193
194 **Results**
195 *Multi-dimensional trait evolution*
196 To understand the constraints on phenotype evolution, we must consider how multi-
197 dimensional traits are altered by selection. Previous empirical work has shown that both trait values
198 and the correlations between traits are altered as a population adapts to a new environment. For
199 example, when 5 genotypes of the green alga *Chlamydomonas reinhardtii* were selected under
200 high-CO_2 [3], all quantified traits changed to varying degrees depending on the genotype and the
201 correlations between many traits changed with some traits becoming correlated (e.g., 1v2) while
202 others becoming uncorrelated (e.g., 2v4; Fig 2c).
203 To understand how trait movement within the trait-scape can be constrained by historical
204 bias (previous correlations between traits), we developed a statistical model (TRACE) of multi-
205 trait adaptation and investigated probabilities of different emergent evolutionary outcomes. We
206 began with a 'null hypothesis' model in which there was no historical bias (mixed mode
207 simulation) and then systematically added in bias to determine the impact on population level

208 adaptation. An example of model dynamics from a single run in mixed mode is shown in Fig. 3a
209 where a representative population consisting of a thousand individuals moved over time from the
210 ancestral starting phenotype to the evolved high fitness area (Fig. 3a). This resulted in an overall
211 increase in fitness of the population over time (Fig 3b). The underlying dynamics of the model
212 (changes in trait values and trait correlation changes) for 3 representative traits are shown in Fig.
213 3c.

214 Though TRACE is a novel modeling framework for understanding trait adaptation, the
215 model captures well known dynamics of adaptions. For example, fitness effects produced from
216 changes at the beginning of the walk were significantly greater than at the end of the walk
217 consistent with previous experimental and theoretical work [1,34] [39-42] (Fig. 3). Although some
218 individuals reached a maximum possible fitness of 1 (i.e., the evolutionary end coordinate), the
219 mean population fitness consistently remained below 1 (Fig. 3b). This is due to the fact that the
220 model is simultaneously optimizing multiple traits and their correlations, which inherently
221 introduces small but significant amounts of persistent phenotypic variation. In addition, while
222 average movement of the population was fairly linear in PC space (Fig. 3a), the trajectory of trait
223 changes was not linear (Fig. 3c).

224 *Accessibility of cryptic phenotypes*

225 Four distinct population types (i.e., traits + trait correlations for the final population)
226 emerged across the replicate runs (1000 individuals x 100 replicate runs = 100,000 individuals
227 total). We term these population types (Pop-MA, Pop-MB, Pop-MC, and Pop-MD) 'cryptic
228 phenotypes' as they had statistically similar end mean fitness values and occurred in the same
229 region of the trait-scape but had distinct trait correlations and, to some extent, distinct trait values
230 (Fig. 4 and Supplementary Fig. 1). In other words, these cryptic phenotypes represent four distinct

evolutionary outcomes of different trait correlations + trait values that all converged on the single evolutionary end coordinate in the evolved trait-scape. For some correlations such as trait 1 vs trait 2 (1v2), little to no overlap was observed across each of the 4 population types (Fig. 4, row 5), while for others, several population types shared the same trait correlations. For example, individuals in Pop-MA and Pop-MD shared the same 1v3 correlation (Fig. 4, row 2, columns 1 and 4) but had completely different relationships for 1v4 (Fig. 4, row 3, columns 1 and 4). The observance of emergent, cryptic phenotypes with distinct underlying trait combinations are qualitatively in line with experimental evolution studies that observed convergent phenotypes derived from a mix of parallel and divergent mutational and transcriptional changes across replicate populations evolving to the same environment [7,43-46].

When the model was run without the influence of trait correlations, only one phenotype emerged, as expected, but in contrast to the simulations where trait correlations were included (Supplementary Fig. 2; Supplementary information S2). Here, individuals were unconstrained 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. The emergence of multiple high-fitness phenotypes (e.g., Fig. 5a) occupying a single high fitness area in multivariate space demonstrates that by using PCA for the trait-scape our model captures a rugged trait-based fitness landscape with multiple high-fitness peaks (e.g., Fig. 1).

Not all cryptic phenotypes were equally accessible by the model populations. Here we define accessibility as the fraction of replicates that converged on an emergent population type. When the model was run without bias, Pop-MA was the most accessible with 55% of replicates

converging on this population type 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. 5a), indicating a relatively larger range of trait values conferring high-fitness when associated with Pop-MA's trait correlations. The most accessible population type, Pop-MA, also had the fastest rate of fitness gain (Fig. 5b, c). 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 population type (Fig. 5b). Instead, Pop-MD with a slower adaptive rate was the second most accessible population type (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 trait values on the accessibility of different cryptic phenotypes, we ran the mixed-mode model using 1) a single population at a different starting location (i.e., trait values) in the trait-scape that was equidistant to the high fitness area and 2) 4 different subpopulations (i.e. 4 different starting locations in the trait-scape). These model runs converged on the same cryptic phenotypes observed with the empirical starting location (Pop-MA, Pop-MB, and Pop-MD for run 1 and all four cryptic phenotypes for run 2). However, shifting the starting location did alter the accessibility of the cryptic phenotypes (Supplementary Fig. 3&4). These runs indicate that the high-fitness phenotypes were conserved, and that starting an adaptive walk from another location influenced the accessibility of certain cryptic 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 [27].

Adding Historical Bias

Movement of the populations within the trait-scape was impacted by the historical bias of the ancestral population (i.e., trait correlations). In the mixed-mode model (described above), the population contained a large diversity of trait correlations among individuals with analogous trait values. The robustness of the resulting cryptic phenotypes across model runs indicates that certain trait-correlations confirm a fitness advantage within the empirically defined trait-scape and thus were selected for. To assess how different types of bias (ancestral vs evolved correlations) impacted the accessibility of the cryptic phenotypes and rate of adaptation, we conducted a suite of simulations where bias was systematically added (sub-modes A1 - A4 and sub-modes E1 – E4). For both ancestral and evolved modes, systematically adding more bias (e.g., going from A1 – A4) changed the accessibility of the high-fitness phenotypes across replicate runs (Supplementary Fig. 5). However, the type of bias (e.g., ancestral vs evolved correlations) had a differential impact on cryptic phenotype accessibility. Bias from the ancestral correlations was typically maladaptive and resulted in fewer accessible cryptic phenotypes and slower adaptive rates (Supplementary Fig. 5a). On the other hand, bias from the evolved trait relationships (i.e., consistent with the trait-scape) generally resulted in faster adaptive rates and greater overall accessibility to the cryptic phenotypes (Supplementary Fig. 5b). These results are consistent with prior empirical and theoretical observations in developmental and quantitative genetic studies 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 [16,17,27]. 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 [27,28]. Here we

299 demonstrate these dynamics using a novel framework for modeling multi-variate adaption in
300 phytoplankton based on easy to quantify empirical trait data.
301
302 *Meta-analysis of phenotypes across different modeled modes*
303 We assessed the similarity between emergent cryptic phenotypes across all model
304 simulations (9 modes with 100 replicates each) and showed 5 high-confidence cryptic phenotypes.
305 Specifically, we grouped similar population types using hierarchical clustering with multiscale
306 bootstrap resampling (1,000 replicates) on mean trait correlation values at the 2000th generation
307 (Methods). We also included the empirical data from the ancestral and evolved populations in this
308 analysis. Hierarchical clustering revealed 5 high-confidence clusters (I – V) harboring 93% of the
309 phenotypes (n=26 of 28) with approximately unbiased (AU) p-values > 75 (Fig. 6a). Two
310 population types, Pop-MB and Pop-EB-E1, clustered with II and IV, respectively, albeit with less
311 confidence relative to the high-confidence clusters. The empirical ancestral population did not fall
312 within any of the high-confidence clusters, which is expected as the ancestral population was not
313 well-adapted in the evolved trait-scape. In contrast, the empirical evolved population clustered
314 with high-confidence in cluster V. This empirically observed population type (cluster V) only
315 emerged in the model simulation when evolved bias was present in the starting population and was
316 found by 55% of the replicates from sub-modes E1, E3, and E4.
317 The clustering observed through the hierarchical analysis also emerged through a PC
318 analysis of the population trait correlations. Specifically, we observed 3 general regions of
319 convergence in PC space among the population types, as clusters II, III, and IV collapsed into a
320 small region of the lower left quadrant in the PCA plot (Fig. 6b). Importantly, these convergent
321 regions emerged from thousands of possible trait and correlation values across varying degrees of

322 bias. They provide valuable insight into probable combinations of high-CO_2 adaptive trait
323 correlations along a reduced set of biological axes.
324
325 **Discussion**
326 We need to bridge the gap between evolutionary models and trait-based ecosystems
327 models (ODE models) in order to better predict how marine microbes will adapt to shifts in the
328 environment [11]. This work takes a critical first step in developing a framework (TRACE)
329 which uses empirically derived multi-variate trait-based landscapes to provide insight into the
330 interaction between historical bias (trait correlations) and evolved phenotypes for marine
331 phytoplankton. Critically, TRACE is derived from and provides predictions of easily quantifiable
332 traits – such as those commonly measured by biological oceanographers. Using data from an
333 experimental evolution study with a model green alga, we found that a limited set of integrated
334 phenotypes underlie thousands of possible trait correlational scenarios and that only certain
335 phenotypes were accessible depending on the amount and type of bias. By leveraging empirical
336 ancestral trait correlations and the observed changes in these correlations as a result of adaptation
337 to high CO_2, we were able to simulate adaptive walks with endpoints anchored in real
338 evolutionary outcomes.
339 This study provides a roadmap for future integration of evolutionary theory with
340 biological oceanography. While we focus on a case study using an experimental evolution study
341 to generate the trait-scape, our framework can be used to generate a trait-scape for any given
342 environment A using trait data from an in situ population (assuming that the in situ population is
343 well adapted to environment A). This trait-scape could then be used with the TRACE framework
344 to develop hypotheses as to how a new population from a different environment B might adapt

upon exposure to environment A. This insight can be gained with easily quantifiable trait measurements and without requiring genetic manipulations, which are currently not possible for most marine microbes.

Our work demonstrates that ecological models need to represent both changes in traits (already existing in some ecological models) and changes in the correlation between traits in order to accurately capture phytoplankton evolution. The TRACE framework could be combined with an ODE based ecosystem model to more accurately predict marine microbial adaptation. While the development of a fully integrated TRACE + ODE model is beyond the scope of this paper, we demonstrate that such an integration would result in a substantially different set of phenotypes and impact community dynamics (Supplementary information S4; Supplementary Fig. 6&7). Such an integration provides the added benefit of removing the need to define the evolutionary end point as fitness in an ODE model, which can be dynamically estimated based on prognostically calculated growth and mortality rates.

This study demonstrates that shifts in trait-correlations are fundamental for understanding the evolved phenotype and provides a novel framework for linking easily quantifiable trait measurements to predictions of evolved phenotypes for marine phytoplankton. This is particularly exciting because we can create trait-scapes from field data and use TRACE to understand how invading microbial populations may be able to adapt (i.e. create new phenotypes). Importantly, this approach 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. We are at a critical juncture where we need ecosystem and biogeochemical models to incorporate evolutionary dynamics in order to robustly predict future shifts in ecosystem dynamics [33]. Due to the seemingly infinite amount of possible interacting biological and environmental variables to

368 test, evolutionary and mathematical tools that allow us to efficiently combine experiments with
369 modeling will be critical to help predict microbial population responses to future global change
370 scenarios through the lens of evolutionary phenomena.
371
372
373
374
375

Figure Legends

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.

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 associated with 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 (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 ($p > 0.05$).

Fig. 3 Representative adaptive walk in the evolved trait-scape 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 higher fitness denotes the end. As in b), each point represents the mean standardized trait value of all individuals at a specific generation, or step.

Fig. 4 Four distinct, emergent population types (cryptic 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 population types (columns). Each population type 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 population types and so the trait-correlation values associated with these population types (cryptic phenotypes) had the largest frequency (y-axis, note scale of 10^4).

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 population types (i.e. cryptic 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.

Fig. 6 Hierarchical clustering and Principal Component Analysis of mean trait correlation values calculated across all emergent population types 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 population types (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.

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.).

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

22. 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)

23. 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)

24. Ward, B. A., Collins, S., Dutkiewicz, S., Gibbs, S., Bown, P., Ridgwell, A., Sauterey, B., Wilson, J. D. & Oschlies, A. 2019 Considering the Role of Adaptive Evolution in Models of the Ocean and Climate System. *Journal of Advances in Modeling Earth Systems* **11**, 3343–3361. (doi:10.1029/2018MS001452)

25. 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)

26. 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)

27. 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)

28. 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)

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

30. 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)

31. 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)

32. 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)

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

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

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

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

37. Oksanen, J., Blanchet, F. G. & Kindt, R. 2013 vegan: Community Ecology Package. *Community ecology …*

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

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

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

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

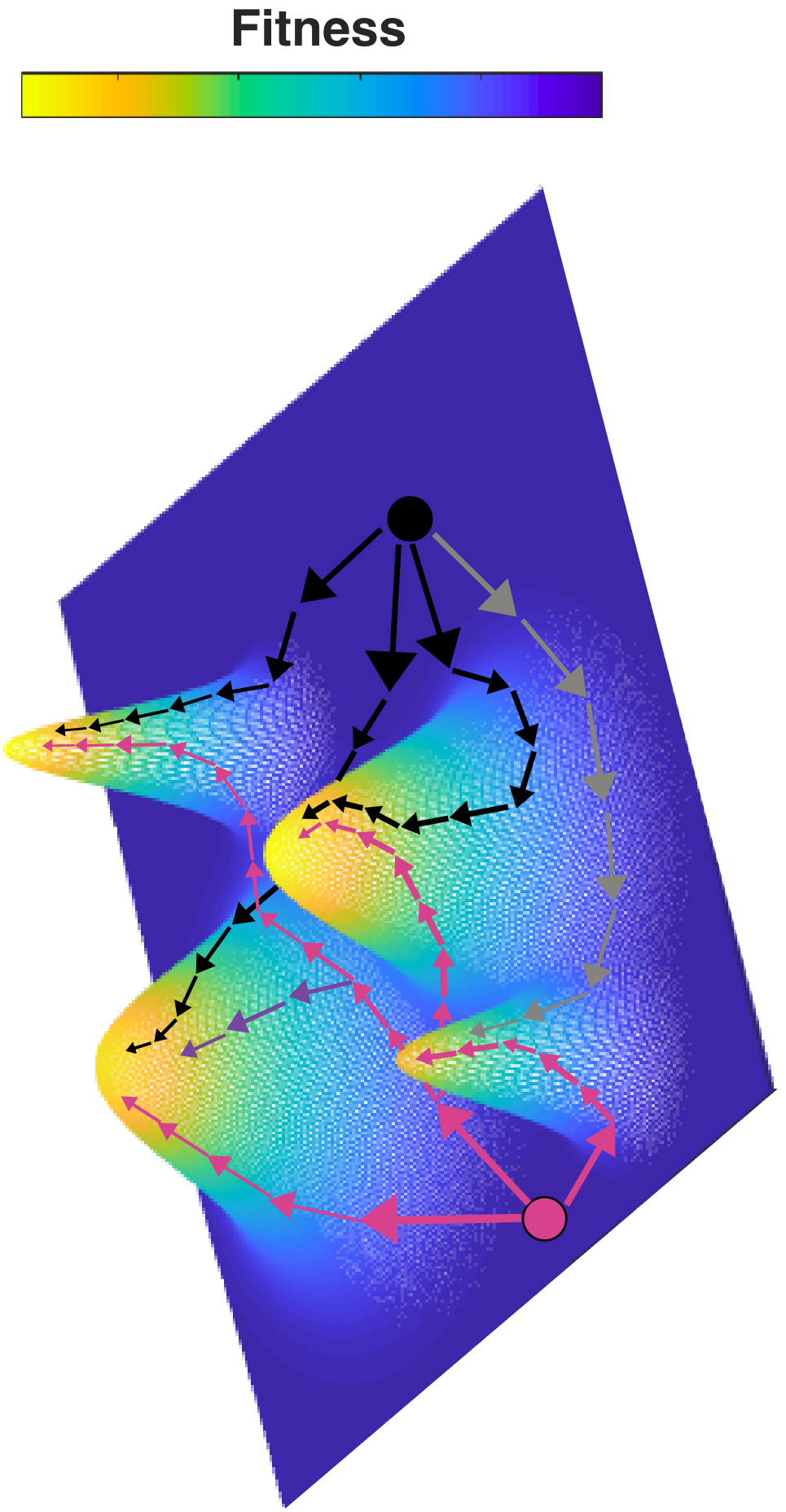
42. Elena, S. F. & Lenski, R. E. 2003 Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* **4**, 457–469. (doi:10.1038/nrg1088)

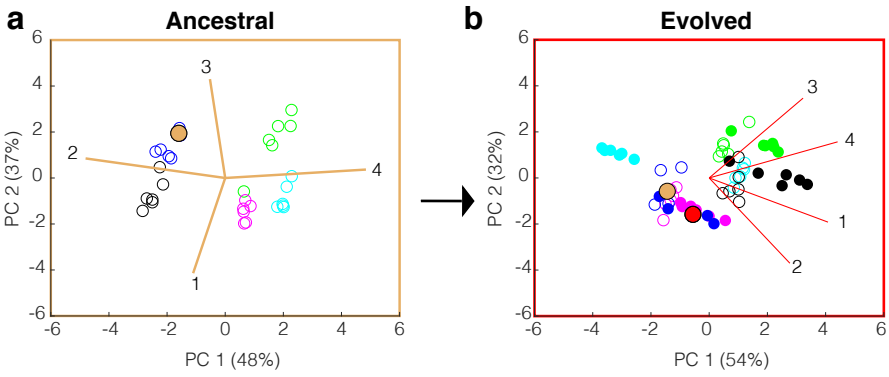
43. Jerison, E. R., Ba, A. N. N., Desai, M. M. & Kryazhimskiy, S. 2020 Chance and necessity in the pleiotropic consequences of adaptation for budding yeast. *Nature Ecology & Evolution*, 1–20. (doi:10.1038/s41559-020-1128-3)

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

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

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

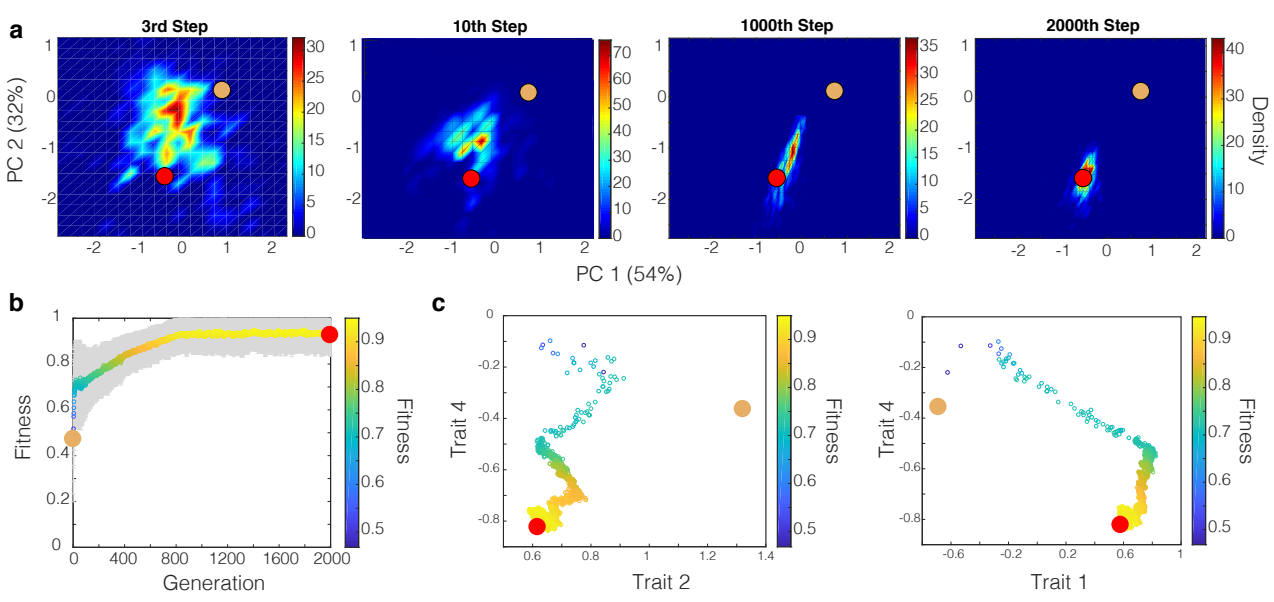

Fitness

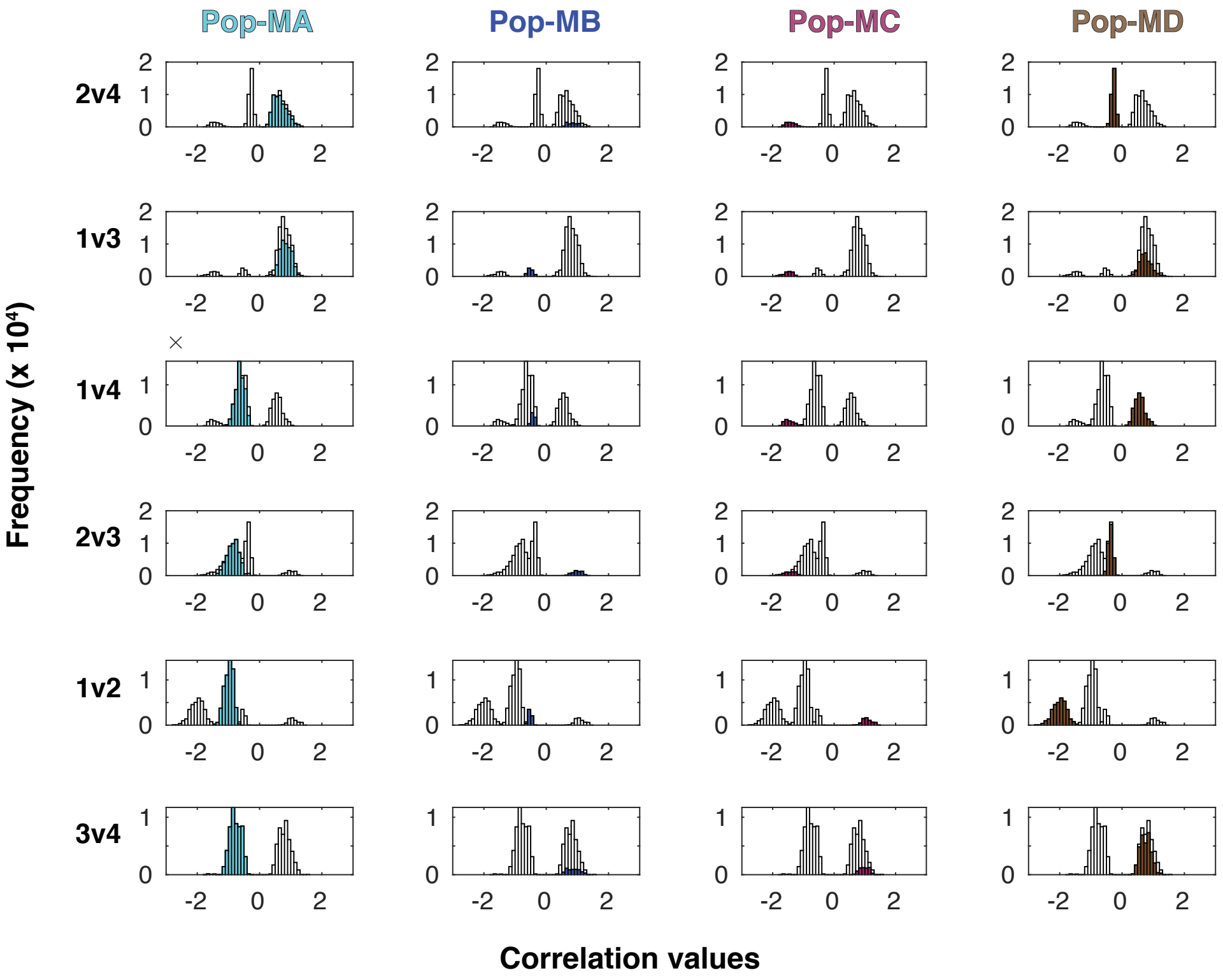


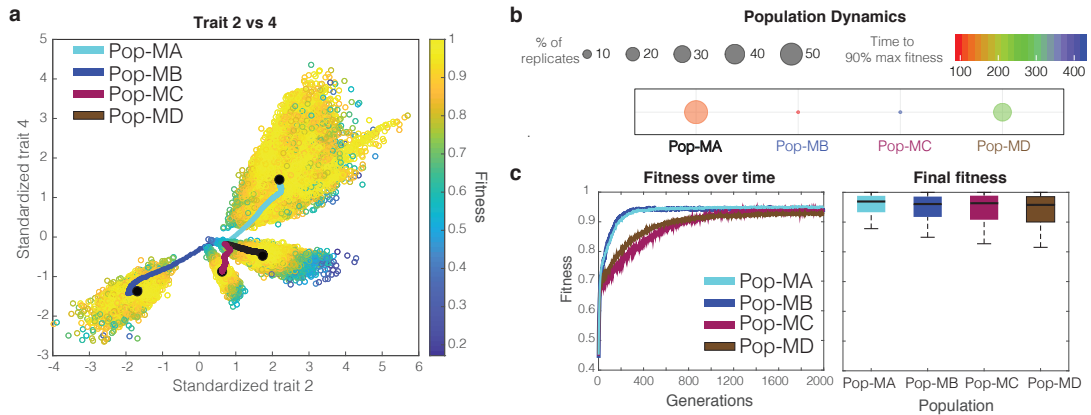
c

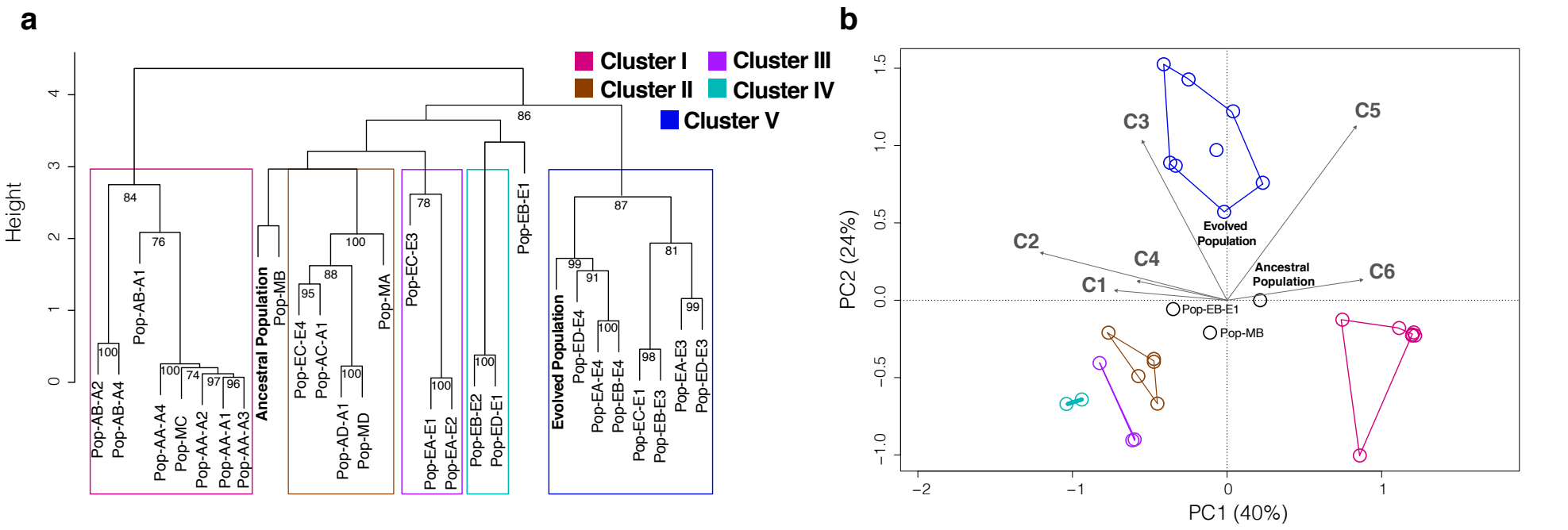
Correlation

	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


a
3rd Step
10th Step
1000th Step
2000th Step
PC 2 (32%)
PC 1 (54%)
Density
b
Fitness
Generation
c
Trait 4
Trait 2
Trait 1


Pop-MA
Pop-MB
Pop-MC
Pop-MD
2v4
1v3
1v4
2v3
1v2
3v4
Frequency (x 10⁴)
Correlation values


a
Trait 2 vs 4
Pop-MA
Pop-MB
Pop-MC
Pop-MD
Standardized trait 4
Standardized trait 2
Fitness
b
Population Dynamics
% of replicates
10
20
30
40
50
Time to 90% max fitness
100
200
300
400
Pop-MA
Pop-MB
Pop-MC
Pop-MD
c
Fitness over time
Fitness
Generations
Final fitness
Population


a
Cluster I
Cluster II
Cluster III
Cluster IV
Cluster V
Height
0
1
2
3
4
86
84
76
100
100
74
97
96
Pop-AB-A2
Pop-AB-A4
Pop-AB-A1
Pop-AA-A4
Pop-MC
Pop-AA-A2
Pop-AA-A1
Pop-AA-A3
Ancestral Population
Pop-MB
100
88
95
100
Pop-EC-E4
Pop-AC-A1
Pop-AD-A1
Pop-MD
Pop-MA
78
100
Pop-EC-E3
Pop-EA-E1
Pop-EA-E2
100
Pop-EB-E2
Pop-ED-E1
Pop-EB-E1
87
99
91
100
81
98
99
Evolved Population
Pop-ED-E4
Pop-EA-E4
Pop-EB-E4
Pop-EC-E1
Pop-EB-E3
Pop-EA-E3
Pop-ED-E3
b
C1
C2
C3
C4
C5
C6
Evolved Population
Ancestral Population
Pop-EB-E1
Pop-MB
PC2 (24%)
PC1 (40%)
1.5
1.0
0.5
0.0
−0.5
−1.0
−2
−1
0
1