Title

Multivariate trait correlational evolution of ecologically important traits underlies constrained 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 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] Plant Functional Biology and Climate Change Cluster, Faculty of Science, 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 both aquatic and terrestrial biogeochemical cycling, thereby significantly influencing the global climate. Predicting how microbes will adapt to global change and the implications for 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 multivariate trait evolution using orthogonal axes to define a trait-scape. We investigate the outcome of thousands of possible adaptive walks within a trait-scape parameterized using empirical evolution data. We find that only a limited number of phenotypes emerge, with some being more probable than others. Populations with historical bias in the direction of selection exhibited accelerated adaptation while highly convergent phenotypes emerged irrespective of the type of bias. Reproducible phenotypes further converged into several high-fitness regions in the collapsed trait-scape, thereby defining probable low-fitness (exclusion) regions. The emergence of nonrandom phenotypic solutions and high-fitness areas in an empirical algal trait-scape confirms that a limited set of evolutionary trajectories underlie the vast amount of possible trait correlation scenarios. Critically, we demonstrate these dynamics in multidimensional trait space and show that trait correlations, in addition to trait values, must evolve to explain the trajectories and outcomes of adaptation in multi-trait space. Investigating microbial evolution through a reduced set of evolvable biogeochemically-important traits and trait relationships lays the groundwork for incorporation into global change-driven ecosystem models where microbial trait dispersal can occur through different inheritance mechanisms. Identifying the probabilities of high-fitness outcomes based on trait correlations will be critical to directly connect microbial evolutionary responses to biogeochemical cycling under dynamic global change scenarios.

Author Summary

Microorganisms drive the base of global food webs and biogeochemical cycling, which influences Earth's climate. Thus, it is critical to understand how their evolutionary responses to global change will affect global environmental processes. Microbial populations are highly diverse and both shape and are shaped by numerous environmental gradients happening simultaneously on different timescales. The sheer number of combinations to experimentally test exceeds our ability to do so, and so theoretical approaches that integrate biological and environmental variability onto a reduced set of representative axes can aid predictions of evolutionary outcomes. Here, we not only show that a finite set of biogeochemically relevant evolutionary outcomes underlie thousands of possible trait correlation combinations, but that the existence of past trait correlations and their subsequent evolvability bias ecologically relevant phenotypic evolution. These phenotypes further converge into high-fitness regions in a collapsed trait landscape derived from these representative axes. The emergence of only a handful of solutions from thousands of possible scenarios is a powerful tool to help constrain predictions of microbial evolutionary trajectories given a vast array of potential trait value and tradeoff combinations. This approach lays the groundwork to embed this framework into larger ecosystem models to examine the effects of these responses on biogeochemical cycling and global climate using trait correlational adaptation vs trait adaptation alone.

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 more 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, and the sheer number of interacting environmental and trait combinations to study 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 robust future predictions [5]. Here, we aim to understand how historical bias (different evolutionary starting points) across biological axes of variation derived from complex traits and their relationships can constrain overarching evolutionary trajectories of phenotypes (suites of traits) in populations adapting to environmental change. We broadly define bias as the standing trait correlations (i.e., relationships) in a population that are heritable and can impact fitness such that, over time, these correlations can generate adaptive constraints that bias 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 a larger biogeochemical framework that can subsequently utilize global

change-driven environmental forcings to drive evolutionary outcomes of microbial populations. Specifically, current models assume that existing interspecific trait relationships will govern phytoplankton phenotypes. However, growing evidence demonstrates that intraspecific variation can be significant in phytoplankton, and that constraints on intraspecific trait relationships will bias evolutionary trajectories of biogeochemically important microbial populations in the face of environmental change [17-19].

Seminal research has been conducted over the years focusing on the interaction of complex trait relationships, inheritance, epistasis, and metabolic networks in organisms 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. Quantitative genetics uses statistical approaches to study adaptive walks accounting for uncertainty inherent in trait variation, genotypic variability, inheritance, and environmental variability [20]. These studies have created theoretical frameworks using multivariate and eigenvector methods to examine evolutionary trade-offs between biological and environmental dimensions over time [21]. Specifically, these frameworks consider how genetic mutations impact the fitness landscape of a population. These studies are often entirely theoretical [22,23], or significantly empirically limited due to the need to measure the fitness impacts of every possible type of mutant. Alternatively, studies in developmental bias and facilitated variation 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) due to the inherent structure, composition, and evolutionary history of a population [24,25]. These findings contrast with the long-held assumption of isotropic (i.e. equal) variation [26] and have revealed that only a limited part of phenotypic space (i.e. only certain phenotypes) can be accessed

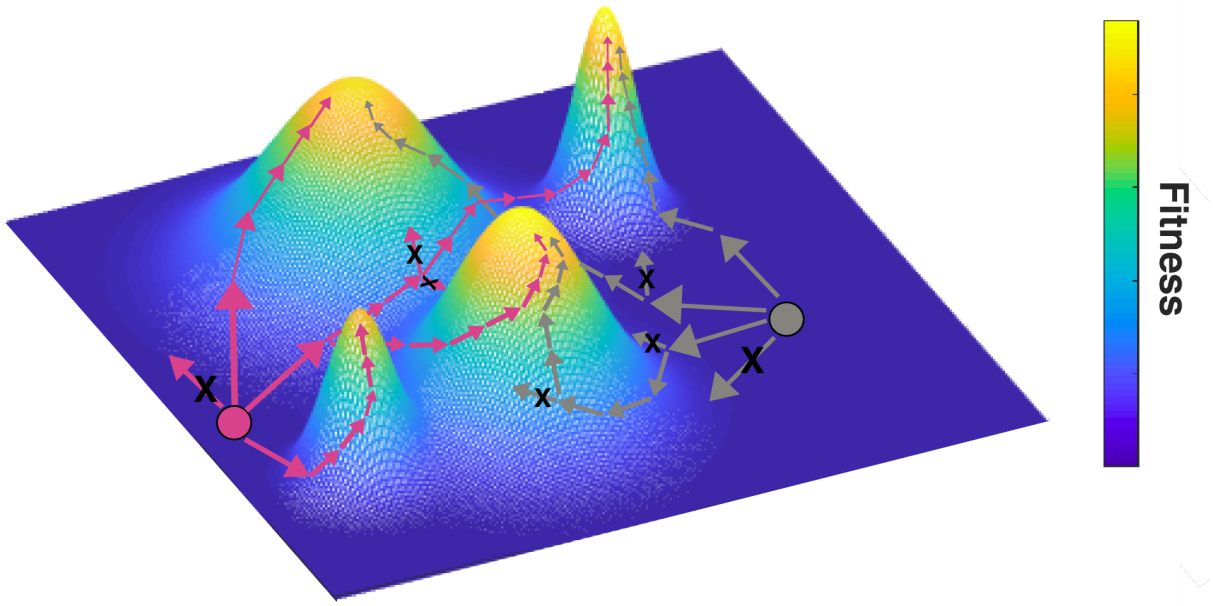


Fig. 1. Comparison of adaptive walks between two different phenotypes in a trait landscape with four high-fitness peaks Two example starting phenotypes are represented as circles (magenta and grey). The phenotypes start with low fitness (z-axis) and through plasticity (trait changes) and adaptation (trait-correlation changes) move to higher fitness. The initial arrows represent 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. Each black 'X' represents an inaccessible path to that phenotype at a specific point in the adaptive walk due to trait correlational constraints. Note that depending on historical bias (i.e., phenotypic starting location), some high fitness peaks are either more difficult to access or completely inaccessible.

contingent upon the aforementioned bias [27]. Critically, this explains why not all trait combinations are achievable [26]. In summary, the rich and growing body of literature from these two fields has shown that genetic architecture significantly influences the impact of environmental shifts and random biological changes (e.g., mutation) on traits and the relationship between traits to produce nonrandom distributions of phenotypes [28-30].

While the phenomenon of restrictions on possible trait combinations is well established, there have been little to no attempts to investigate the implications of this phenomenon for the evolution of ecologically relevant traits and trait relationships in marine microbes and the consequences for biogeochemical cycling. Specifically, the limitation of viable trait combinations

in phytoplankton has the potential to impact both rates of carbon cycling and shifts in marine ecosystem structure.

Fig. 1 shows an illustrative example of a trait landscape where peaks represent high-fitness phenotypes (trait combinations). In this example landscape, there are 4 equally high-fitness peaks. However, the accessibility of each peak differs depending on the starting location and the initial trajectory (Fig. 1, magenta and grey circles). Ultimately, to robustly study phytoplankton trait evolution, we need a framework that allows us to estimate probable evolutionary trajectories given starting trait combinations and trait correlations (historical bias) that set the initial trajectory. Below, we introduce such a framework using empirical evolution data from a eukaryotic alga. Our model is a first step towards investigating how correlated metabolic traits with clear biogeochemical significance may impact chemical cycling under environmental change (e.g., ocean acidification). Specifically, we develop TRACE, a framework that models TRAit Correlation Evolution under selection to constrain evolutionary outcomes for defined traits of interest in microbial adaptation. We validate the model with empirical data from an experimental evolution study with the eukaryotic alga *Chlamydomonas reinhardtii*. We then use the collapsed trait-scape in TRACE to investigate constraints on phenotypic variants and adaptive walks as a function of historical bias (i.e., trait correlations). We find that a handful of phenotypic variants are reproducible even without the introduction of bias. As we increase bias in the model through the systematic addition of empirical ancestral and evolved correlations, we find that overall adaptive rates increased for evolved correlations but not for ancestral ones, consistent with prior studies [26]. These results indicate that populations harboring trait correlations oriented in the direction of selection can accelerate adaptation. Model runs seeded with either ancestral or evolved bias (i.e. correlations) converged on several of the same phenotypes as those seeded with no bias

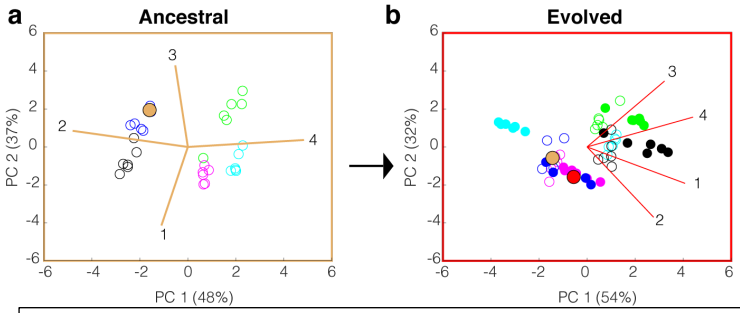
thereby elucidating emergent, highly reproducible evolutionary solutions. Taken together, we show conserved phenotypes that emerge from thousands of possible trait and correlation scenarios across varying levels of historical bias. Understanding which trait relationships inform the probability of adaptive microbial phenotypes will be critical to help predict the short- and long-term contribution biogeochemically-important traits to biogeochemical cycling.

Results

Collapsed multivariate trait space

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 [18,31]. Building on this work, we create a trait landscape or '*trait-scape'* for the green alga *Chlamydomonas reinhardtii* adaptation to high-CO_2 using four independent and ecologically relevant traits (growth rate, respiration, cell size, and daughter cell production). Specifically, using the output from an experimental evolution study [3] with 5 genotypes of *C. reinhardtii*, we demonstrate that both the traits and the correlations between traits evolved between the low-CO_2 environment (ancestral environment) and the 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 (Fig 2c). This results 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 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 [18,26,31]. To understand how trait movement within this collapsed multi-dimensional trait-scape may be constrained by historical bias (previous correlations between traits), we develop a statistical model of multi-trait adaptation and investigate probabilities of different emergent evolutionary outcomes.



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 evolved genotypes. Open circles represent the ancestral genotypes in a) projected into the evolved PC space. 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.

Impact of bias

The TRACE framework was adapted from an individual based Fisher model of adaptation [1,32,33]. Here we use a population of 1000 individuals that experience an adaptive walk in a trait-

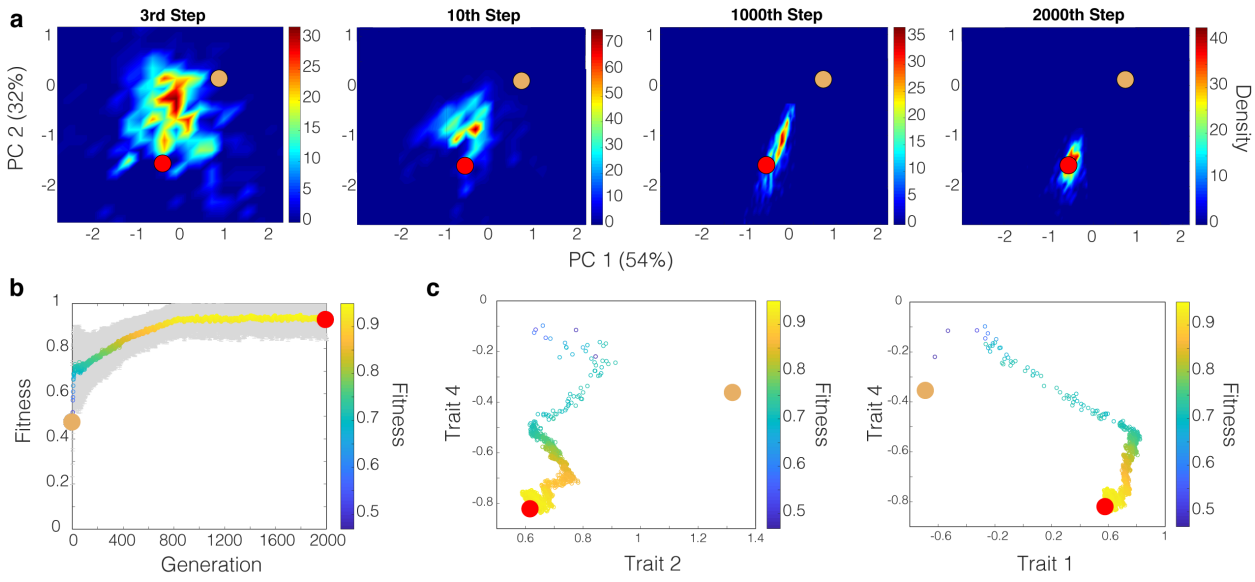


Fig. 3 Representative adaptive walk in evolved PC space of a population of 1000 individuals. a) Density plots of an adaptive walk of a single population starting at the tan circle and ending at the red circle. Each plot represents a different step (i.e. generation) in PC space with the color representing the density of individuals in a given area. Percentage values represent the amount of variance explained of each PC axis. 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-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.

scape towards a high-fitness area (e.g., Fig. 2b, from the tan to red circle) across thousands of generations. 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) [1,32]. The trait-scape, the population starting location, and the high fitness area in the trait-scape were defined based on empirical data from a long-term evolution study. Each generation, each individual in the population experienced either a random change in a trait value but maintained all existing trait-correlations or experienced a trait-correlation change. These changes were drawn from a Gaussian distribution such that small changes were common and large changes were rare. Changes in trait values moved these individuals in the trait-scape. Selection was imposed using distance to the high fitness area (evolutionary end coordinate) as a proxy for fitness (*see Methods for full model description*). In essence, this selects for individuals with the smallest overall difference across all trait values from the empirically observed high fitness phenotype. The weighting of the traits is

derived from the observed evolved phenotypes evaluated using PCA, such that traits that were not observed to play an important role in high-CO_2 fitness had low weight. It is important to note that the model does not directly select for trait correlations, rather specific correlations emerge in the population if they provide a fitness advantage in terms of trait dynamics. We ran the model in 3 different modes to examine all possible trajectories that a population could travel given the same starting trait values but different starting trait correlations, a proxy for different historical biases. Each mode was run for 2000 generations with 100 replicate runs each. All model parameters are given in Supplementary Table 1. An example of model dynamics is shown in Fig. 3a where a representative population consisting of a thousand individuals moves over time from the ancestral start phenotype to the evolved high fitness area (Fig. 3a), which is represented as an overall increase in fitness of the population (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.

Mixed mode – No historical bias

Our null model assumed no historical bias. Specifically, all individuals in the model started with the same trait values and hence a single start point in the PC trait-scape but each with randomly selected correlations between the traits. These random assignments both ensured that the population retained no shared bias and that we explored a vast, unbiased amount of evolutionary choices/orientations (1000 individuals x 100 replicate runs) (Fig. 3a). Consistent with prior studies examining evolution under relatively strong selective pressure [1,32], fitness effects produced from changes at the beginning of the walk were significantly greater than at the end of the walk [34-37]. As the model ran forward in time, individuals within the population explored the collapsed trait-scape through changes to both traits and their 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.

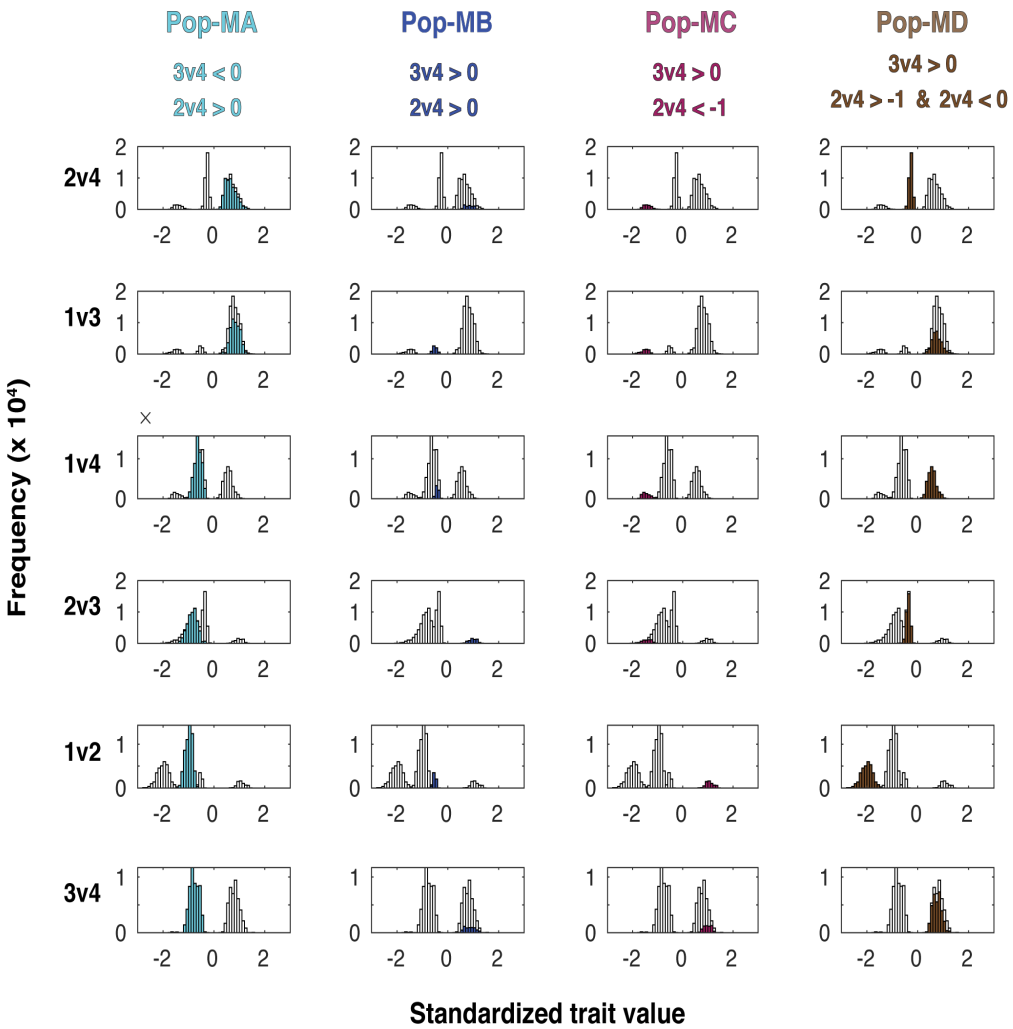


Fig. 4 Four distinct, emergent phenotypes from model runs seeded with no bias Each plot displays the distribution of standardized trait values for a given correlation in a certain phenotype. Phenotypes were binned according to the correlation conditions denoted above each column. Trait values of each population are colored according to their respective correlation conditions. Since a given population may only have to meet 2 correlation conditions (e.g., Pop-MA), colors in other correlations in the population column denote where those individuals fall who meet the 2 correlation conditions. Each column and color represents a phenotype, respectively, with each row denoting one of the six possible correlations. Y-values are 10^4.

At the final generation (2000th generation) of all replicate runs (n = 100), we examined the distribution of each trait correlation (n = 6) across all individuals in all runs (n = 100,000). Four

distinct phenotypes (i.e. traits + trait correlations for the final population) emerged all with statistically analogous end mean fitness (Fig. 4, 1 phenotype per column). As these four phenotypes occurred in the same region of the trait-scape but have distinct trait correlations, and to some extent, trait values, we term them 'cryptic phenotypes'. In other words, these cryptic phenotypes represent four distinct evolutionary solutions of different trait correlations + trait values across all individuals that 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 phenotypes (Fig. 4, row 5), while for others, several phenotypes 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). In contrast, Pop-MA and Pop-MD have a completely different relationship for 1v4 (Fig. 4, row 3, columns 1 and 4). Pairwise trait-trait plots displayed the 4 phenotypes in terms of their trait values (Fig. 5a). In Fig. 5a, the four phenotypes distinctly segregate across the Trait 2 vs Trait 4 values but overlap for other trait values such as Trait 2 vs Trait 3 (Supplementary Fig. 1). Taken together, unique, cryptic phenotypes share some trait correlations but diverge in others. These findings are consistent with other 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 [7,38-41]. The emergence of multiple high-fitness phenotypes (e.g., Fig. 5a) underlying single high fitness area in multivariate space demonstrates that our model captures a multi-peak, rugged fitness landscape.

The accessibility of the four phenotypes that emerged from the starting, non-biased population were 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 in trait space; Fig. 5a), indicating a relatively larger range of trait values conferring high-fitness. The most accessible phenotype, Pop-MA, also had the fastest adaptation rate (Fig. 5b, c), potentially indicating this phenotype to be the most accessible from this starting location. Although Pop-MA and Pop-MB exhibited similar adaptive rates (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 (see below).

To examine the impact of ancestral starting point on the emergent phenotypes, we ran the model with 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 while the former two most populous phenotypes, Pop-MA and Pop-MD, were still accessible, but only 24% and 9% of replicates ended on these phenotypes, respectively (Supplementary Fig. 2). No new phenotypes emerged and no replicate found Pop-MC. 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 confirms the ability of this framework to capture the known phenomenon that there are a limited number of accessible phenotypes.

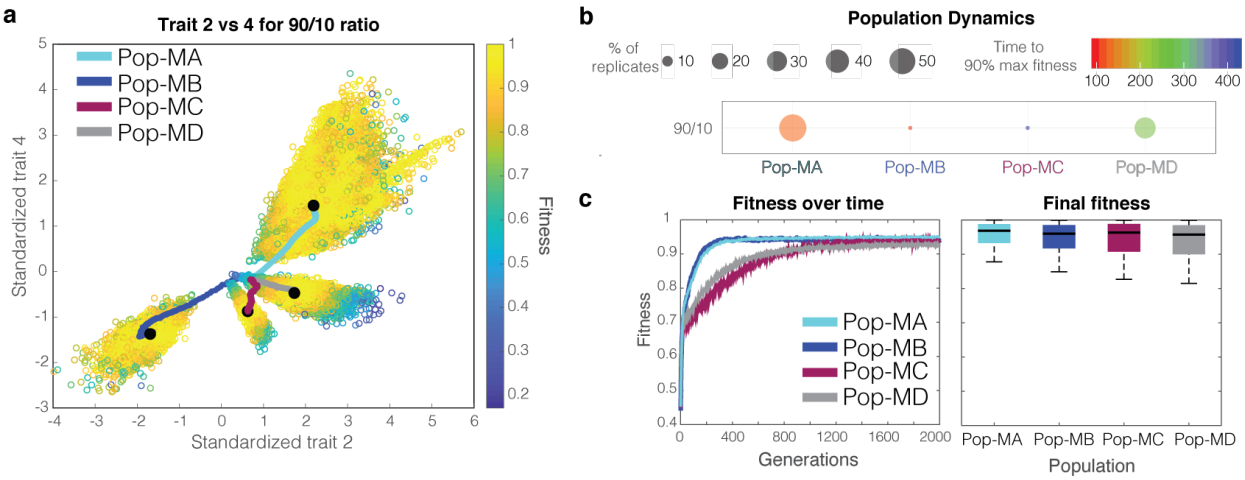


Fig. 5 Representative trait, population, and fitness dynamics for a model run in mixed mode a) Trait-trait plot denoting the 4 distinct populations (i.e. phenotypes) that emerged from 100 replicate model runs in mixed mode (i.e. no bias). Each hollow point represents the final trait values of a given individual in the last generation (2000) colored by fitness. 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.

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) and found that only one phenotypic scenario emerged where individuals were able to explore a broad distribution of trait correlation ranges unconstrained while reaching the high fitness area faster than all other model runs that were comprised of correlation-dependent phenotypes (Supplementary Fig. 3). This demonstrates that trait correlational constraints produce different evolutionary strategies (i.e. emergent, cryptic phenotypes), and if not present, individuals have the ability to explore the range of phenotypic space unconstrained. This in turn accelerates "adaptation" through one "phenotype" (e.g., population) comprised of individuals spanning the range of correlational distributions.

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). Specifically the model was run in two other sub-modes: 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). 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. 4). 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 Bias

The mixed-mode model runs above represent a null hypothesis where organisms have 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 correlations. We first created four sub-modes (A1, A2, A3, and A4) in which correlations derived from the observed ancestral population (Fig. 2) were added sequentially. To test if the sequence of correlational changes influenced adaptive outcomes in our model, we ran the model with the reversed order of correlation changes (Supplementary Fig. 5). For sub-mode A1, random correlation values were generated for 5 of the 6 trait correlations, and the strongest 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

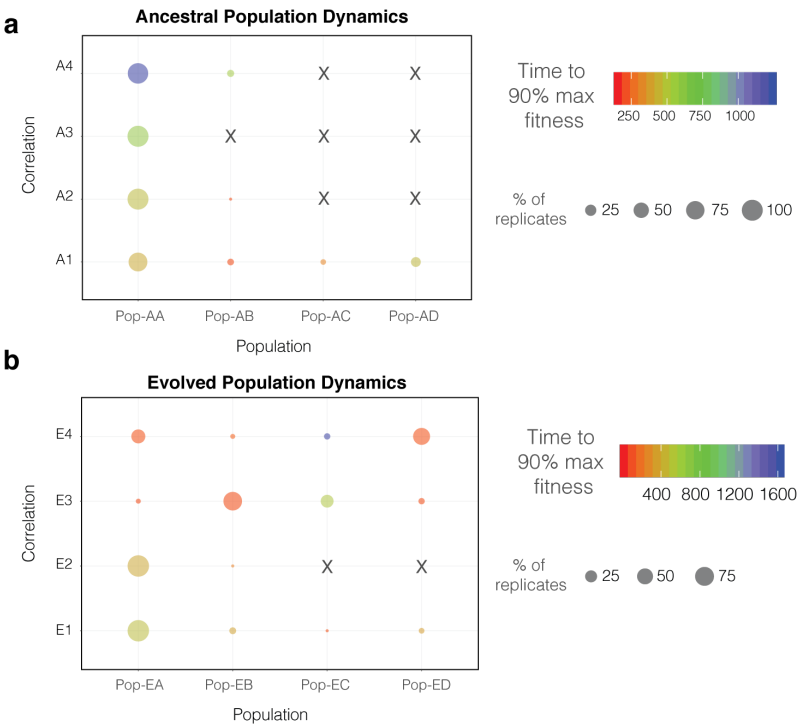


Fig. 6 Population dynamics of emergent populations across different historical bias a) Bubble plot showing emergent populations as a function of adding empirical ancestral correlations (ancestral bias). Bubble size denotes the number of replicates (out of 100) within a specific population while bubble 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).

the same except that two empirical ancestral correlations were shared by all individuals in the starting population. The same was done for A3 and A4. Finally, we conducted the same procedure but using the trait correlations derived from the evolved populations (Fig. 2) to generate sub-modes E1 – E4. The results from both sets of simulations yielded the same phenotypes as described above.

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 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 had a different impact on phenotype accessibility. Bias (trait correlations) 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) resulted in faster adaptive rates and greater overall accessibility to adaptive phenotypes (Fig. 6b). These results are consistent with previous

work, which found that bias (e.g., trait correlations) will accelerate adaptive evolution if existing biological orientation aligns with the direction of selection but may constrain adaptation if it limits variability in the direction of selection [16,17,28]. Hence, depending on a starting population's bias, different phenotypes are more probable than others with some being generally inaccessible as found in other studies [26,27].

Meta-analysis of phenotypes across model runs

We assess the similarity of the high-fitness phenotypes across all 90/10 model runs (9 runs with 100 replicates each) using hierarchical clustering with multiscale bootstrap resampling (1,000 replicates) on mean trait correlation values (Methods). We also included the empirical ancestral and evolved populations in this analysis. Hierarchical clustering revealed 5 high-fitness clusters (I – V) harboring multiple phenotypes 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-fitness clusters. The empirical ancestral phenotype did not fall within one of the high-confidence clusters, which is expected as the ancestral phenotype is not well-adapted in the evolved trait-scape. In contrast, the empirical evolved population clustered with high-confidence in cluster V, a cluster found by 20% of the replicates including phenotypes from sub-modes E1, E3, and E4 where evolved bias was added. The clustering observed through the hierarchical analysis also emerged through a PC analysis of the population trait correlations. Specifically, we observed 3 general areas of high-fitness, as clusters II, III, and IV collapsed into

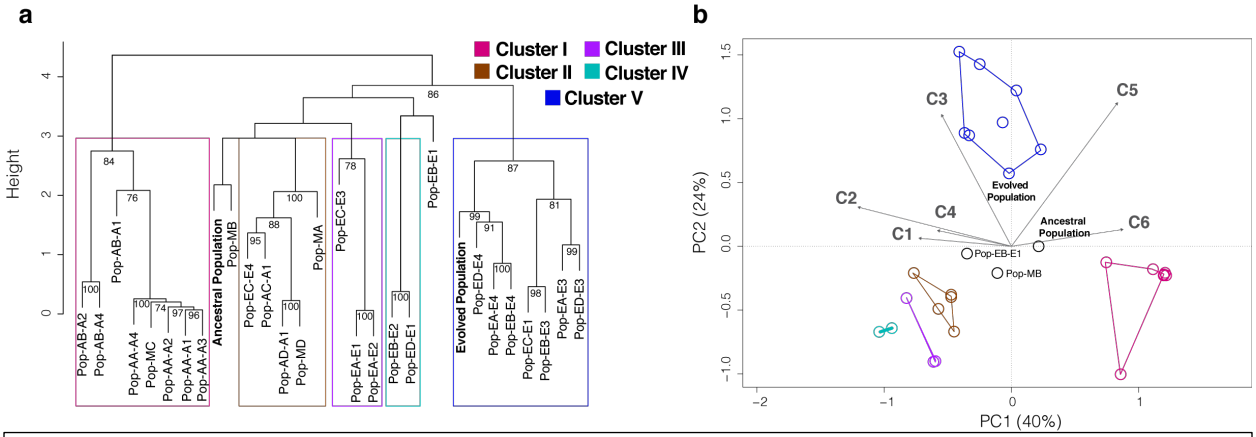


Fig. 7 Hierarchical clustering and Principal Component Analysis of mean trait correlation values calculated across all phenotypes from independent model runs a) Hierarchical clustering with multiscale bootstrap resampling (1,000 replicates) on mean trait correlation values across all individuals from independent model runs (mixed mode, A1-A4, and E1-E4) along with empirical ancestral and evolved correlation values. Approximately Unbiased (AU) p-values (Methods) > 75 are labeled at the nodes. We identified five overarching clusters with AU p-values > 75, which contained even higher confidence sub-clusters. b) Principal component analyses with mean 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 identified in Fig. 2C.

a small region of the lower left quadrant in a PCA including all 90/10 model runs (Fig. 7b). Here we show convergent regions of high-fitness derived from thousands of possible trait and correlation values across varying degrees of bias. These areas representing integrated phenotypes along a reduced set of biological axes offer valuable insight into emergent, constrained evolutionary trajectories given a starting set of traits and their relationships.

Discussion

Here, we combined empirical evolution trait data of a freshwater alga with a framework that uses principal components to model multivariate adaptive walks with evolving traits and trait correlations. By leveraging empirical ancestral and evolved trait correlations to high CO_2, we were able to recreate adaptive walks anchored in real evolutionary outcomes. We found that although different individuals across all model runs reached a maximum fitness of ~1 at different times near the evolutionary end coordinate, many of them were knocked off the maximum through simultaneous random changes to both traits and their correlations (e.g., Fig. 3b, Fig. 5c). These

persistent changes represent a combination of small selective changes refining fitness as well as random mutation-like changes (e.g., drift). These concurrent changes to traits and their correlations inherently produce phenotypic variance that natural selection acts on and may be part of a process that produces and/or maintains genetic heterogeneity. This variance can aid in producing cryptic phenotypes within a high-fitness region in a trait-scape, as we observed above. This could help buffer populations against biological phenomena such as antagonistic pleiotropy from changing environmental conditions. Moreover, positive selection in certain environments may preferentially act on trait relationships relative to individual traits similar to selection on epistatic interactions rather than individual genes in heterogeneous environments [42]. Future studies incorporating more traits and their correlations may help determine if high-fitness, connective ridges exist between these high-fitness areas or if these areas are more akin to high-fitness peaks separated by deeper valleys, which are improbable to cross. With a greater understanding of evolutionary relationships among biogeochemically relevant traits under global change stressors, we can work towards identifying more probable phenotypic outcomes, or at least, those that are improbable due to evolutionary and environmental constraints.

The true utility of TRACE lies in its ability to be applied to specific sets of real-world traits of interest by leveraging data from empirical, organismal experiments. Specifically, it provides a framework for studying the evolution of multiple traits in response to environmental change. This approach generally contrasts the vast majority of past adaptive walk models that only included hypothetical traits and fitness to study evolution. Critically, our model captured the same evolutionary phenomena as past models but with a trait-scape characterized by easy-to-quantify and ecologically important traits from globally important microbes. From here, we can build on our understanding of key sets of multivariate trait relationships and their co-evolution in

ubiquitously distributed microbial populations as they interact with ecological, biological, and physicochemical processes. This will be critical for determining how evolving microbial processes will influence global biogeochemistry and carbon cycling in the face of global change. Since it is impossible to model all possible combinations of interactions, using our approach for constraining probable evolutionary outcomes will be important for generating and testing hypotheses related to global change-driven effects on microbial communities. For example, TRACE can provide hypotheses as to the degree of evolvability of certain traits and trait correlations under selective gradients (e.g., CO_2) in a multi-trait landscape and suggest potential multivariate trait tradeoffs. These hypotheses can then be tested with targeted laboratory and field experiments. This will provide critical new knowledge connecting multi-trait evolutionary outcomes to larger ecological and biogeochemical processes across changing environments. This knowledge can ultimately be integrated into larger biogeochemical models to constrain microbial phenotypes and thus trait distributions under different global change scenarios.

Applying TRACE to *Chlamydomonas* evolution to high CO_2, we found that a limited set of integrated phenotypes underlie thousands of possible trait correlational scenarios. Upon systematically adding different types (ancestral or evolved) of bias, only certain phenotypes emerged for some trait combinations (e.g., A2 – A4 & E2) while others found all possible phenotypes for a specific mode. These results help elucidate evolutionary trajectories based on trait correlation constraints for ecological and biogeochemical traits of interest. Importantly, they also help to inform future experimental designs aiming to test 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 to identify high-fitness trait-correlation combinations under selective gradients. As random biological changes are mapped onto existing phenotypic dimensions with different proportional variation, certain phenotypes or adaptive paths are more readily available than others, thereby producing nonrandom outcomes. Because predicting evolutionary outcomes is extremely difficult, understanding which trait combinations are improbable or impossible helps us to focus on investigating more probable combinations. Future studies can begin to build robust knowledge of select microbial traits and their evolutionary relationships, which can be linked to their dispersal and impact to elemental cycling. 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.

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 downloaded 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 resulting in 48% and 37% of the variance explained on axes PC1 and PC2, respectively, for ancestral correlations and 54% and 32% for evolved correlations. For both ancestral and

evolved PCAs, or trait-scapes, there are 6 trait correlations, which can be found in Supplementary File 1.

To select a start and end point for the adaptive walk, ancestral populations were projected into the evolved PC space. We selected a single genotype to use for the start and end point of the walk (tan circle in Fig. 2b; row 20 in the ancestral trait value matrix in Supplementary File 1) as the start point, and the corresponding evolved population as the evolutionary endpoint (red circle in Fig. 2b; row 20 in the evolved trait value matrix in Supplementary File 1).

TRACE

The model was initialized with a population of 1000 individuals with the same trait values corresponding to the ancestral trait values. Three different modes were run with varying amounts of starting bias using different starting trait correlations: mixed, ancestral, and evolution. The mixed mode corresponds to no bias; the ancestral and evolution modes each had 4 sub-modes where ancestral and evolutionary bias (i.e. trait correlations) were systematically added (explained in detail below). Each model run was conducted for 2000 generations with 100 replicates each. All model parameters are given in Supplementary Table 1.

Mixed mode – No bias

To first test all possible routes available to travel from the ancestral start point to the evolutionary end point in evolved the trait-scape (Fig. 2b), we generated random correlation values from a standard uniform distribution within the open interval (-1,1) and randomly assigned them to all individuals. Hence, every individual started with the same 4 trait values but completely random correlation values (n = 6 correlations). For each generation, 90% of the individuals were randomly chosen, and for each individual, 1 trait was randomly selected to change by drawing a random value from a normal distribution with mean of 0 and standard deviation of 0.05. This trait change

was added to the existing trait value. Following this first trait change, we updated the remaining 3 traits using the trait correlations for that individual in that time step. For example, if trait 1 was initially changed, then we would update traits 2, 3, and 4 using each of their respective correlations with trait 1. We conducted additional runs where we reversed the order of updating the traits and/or correlations but results remained the same (Supplementary Fig. 5).

For the remaining 10% of the population, we changed both a trait and a trait correlation. For each individual, we randomly sampled a correlation to change. Similar to the trait change, we drew a random value from a normal distribution with mean of 0 and standard deviation of 0.05 and added it to the existing correlation value. Next, we randomly chose one of the two traits associated with that correlation and changed it in the same manner as above. Next, we updated the second trait tied to the correlation using the new correlation and trait value.

Following these changes in traits, all individuals in the population were projected back onto the evolved trait-scape using the evolved factor loadings. The Euclidian distances (z) were calculated for each individual relative to the evolutionary endpoint. Next fitness was calculated after as [1,32].

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

Different individuals either moved further or closer to the endpoint depending on the nature of the trait and/or correlation changes (Fig. 3a). The closer the individual moved, the higher its fitness (Fig. 3). 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 work [1,32] inspired by Fisher's model of adaptation [33]. Previous work by us and others have demonstrated that adaptive outcomes using this framework are robust relative to large changes in population sizes (Supplementary text) [1,32]. Several sensitivity

studies were conducted. We ran the mixed mode with a different starting point that was equidistant to the evolutionary end point, as the original starting point. To test different adaptive rates applied to traits and trait correlations, we altered the ratio of changes where instead of changing a trait in 90% individuals and a correlation + trait in 10% of them (90/10), we changed a trait for 50% and a trait + correlation for the other 50% (50/50), and finally changed a trait for 10% and a trait + correlation for 90%.

Ancestral mode

To test the effects of systematically adding ancestral bias, we ran the ancestral mode in 4 sub-modes with 100 replicate runs each: A1, A2, A3, and A4. For simplicity, we chose to sequentially add back in ancestral correlations from most significant to least significant. For 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. So, now each individual contained random correlation values for 5 of the 6 trait correlations, and one ancestral correlation value shared across all of them. The rest of the model steps proceeded as above. For A2, all steps were the same except that we added two empirical ancestral correlations. Finally, the same was done for A3 and A4.

Evolved mode

We conducted the same procedure for the evolved mode but instead systematically added empirical evolved correlations for 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 [43] using Euclidean distance and the average (UPGMA) method. Principal component analysis using mean correlation values was

conducted with package R package vegan [44], and pvclust clusters with AU p-values > 75 were projected onto the PC coordinate plane as convex hulls.

ACKNOWLEDGMENTS. This work was supported by the Moore Foundation Grant MMI 7397 (to N.M.L., S.C., M.A.D.).

References

1. Walworth NG, Zakem EJ, Dunne JP, Collins S, Levine NM. Microbial evolutionary strategies in a dynamic ocean. Proc Natl Acad Sci USA. 2020; 1–6. doi:10.1073/pnas.1919332117

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

3. Lindberg RT, Collins S. Quality–quantity trade-offs drive functional trait evolution in a model microalgal "climate change winner." Ecol Lett. 2020;23: 780–790. doi:10.1111/ele.13478

4. Walworth NG, Fu F-X, Lee MD, Cai X, Saito MA, Webb EA, et al. Nutrient-Colimited *Trichodesmium* as a Nitrogen Source or Sink in a Future Ocean. Appl Environ Microbiol. 2018;84: e02137–17–14. doi:10.1128/AEM.02137-17

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

6. Hutchins DA, Fu F. Microorganisms and ocean global change. Nature Microbiology. 2017;2: 1–11. doi:10.1038/nmicrobiol.2017.58

7. Walworth NG, Lee MD, Fu F-X, Hutchins DA, Webb EA. Molecular and physiological evidence of genetic assimilation to high CO_2 in the marine nitrogen fixer *Trichodesmium*. Proc Natl Acad Sci USA 2016 doi:10.1073/pnas.1605202113

8. Hutchins DA, Walworth NG, Webb EA, Saito MA, Moran D, McIlvin MR, et al. Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. Nat Commun. 2015 doi:10.1038/ncomms9155

9. Hellweger FL, van Sebille E, Fredrick ND. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. Science. 2014;345: 1346–1349. doi:10.1126/science.1254421

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

11. Ward BA, Collins S, Dutkiewicz S, Gibbs SJ, Bown P, Ridgwell A, et al. Considering the role of adaptive evolution in models of the ocean and climate system. Journal of Advances in Modeling Earth Systems. 2019;: 1–29. doi:10.31223/osf.io/srdh3

12. Schluter L, Lohbeck KT, Gro ger JP, Riebesell U, Reusch TBH. Long-term dynamics of adaptive evolution in a globally important phytoplankton species to ocean acidification. Science Advances. 2016;2: e1501660–e1501660. doi:10.1126/sciadv.1501660

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

14. Boyd PW, Cornwall CE, Davison A, Doney SC, Fourquez M, Hurd CL, et al. Biological responses to environmental heterogeneity under future ocean conditions. Global Change Biol. 2016;22: 2633–2650. doi:10.1111/gcb.13287

15. Boyd PW, Collins S, Dupont S, Fabricius K, Gattuso J-P, Havenhand J, et al. Experimental strategies to assess the biological ramifications of multiple drivers of global ocean change-A review. Global Change Biol. 2018;24: 2239–2261. doi:10.1111/gcb.14102

16. Aguirre JD, Hine E, McGuigan K, Blows MW. Comparing G: multivariate analysis of genetic variation in multiple populations. Heredity. 2014;112: 21–29. doi:10.1038/hdy.2013.12

17. Brandenburg KM, Wohlrab S, John U, Kremp A, Jerney J, Krock B, et al. Intraspecific trait variation and trade-offs within and across populations of a toxic dinoflagellate. Ecol Lett. 2018;21: 1561–1571. doi:10.1111/ele.13138

18. Malcom JW, Hernandez KM, Likos R, Wayne T, Leibold MA, Juenger TE. Extensive cross-environment fitness variation lies along few axes of genetic variation in the model alga, *Chlamydomonas reinhardtii*. New Phytol. 2014;205: 841–851. doi:10.1111/nph.13063

19. Moore LR, Post AF, Rocap G, Chisholm SW. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Limnol Oceanogr. 2002;47: 989–996. doi:10.4319/lo.2002.47.4.0989

20. Walsh B. Special issues on advances in quantitative genetics: introduction. Heredity. 2014;: 1–3. doi:10.1038/hdy.2013.115

21. Anderson RW. Learning and evolution: A quantitative genetics approach. Journal of Theoretical Biology. 1995;175: 89–101. doi:10.1006/jtbi.1995.0123

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

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

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

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

26. Uller T, Moczek AP, Watson RA, Brakefield PM, Laland KN. Developmental Bias and Evolution: A Regulatory Network Perspective. Genetics. 2018;209: 949–966. doi:10.1534/genetics.118.300995

27. Elhanan Borenstein DCK. An End to Endless Forms: Epistasis, Phenotype Distribution Bias, and Nonuniform Evolution. PLoS Comput Biol. 2008;4: e1000202. doi:10.1371/journal.pcbi.1000202

28. Gomulkiewicz R, Houle D. Demographic and Genetic Constraints on Evolution. The American Naturalist. 2009;174: E218–E229. doi:10.1086/645086

29. Braendle C, Baer CF, Félix M-A. Bias and Evolution of the Mutationally Accessible Phenotypic Space in a Developmental System. PLoS Genet. 2010;6: e1000877. doi:10.1371/journal.pgen.1000877

30. Houle D, Bolstad GH, van der Linde K, Hansen TF. Mutation predicts 40 million years of fly wing evolution. Nature. 2017;548: 447–450. doi:10.1038/nature23473

31. Kirkpatrick M. Patterns of quantitative genetic variation in multiple dimensions. Genetica. 2008;136: 271–284. doi:10.1007/s10709-008-9302-6

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

33. Fisher RA. The Genetical Theory of Natural Selection. Oxford; 1930.

34. Lenski RE. Convergence and Divergence in a Long-Term Experiment with Bacteria. The American Naturalist. 2017;190: S57–S68. doi:10.1086/691209

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

36. Cooper VS, Lenski RE. The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature. 2000;407: 736–739. doi:10.1038/35037572

37. Elena SF, Lenski RE. Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet. 2003;4: 457–469. doi:10.1038/nrg1088

38. Jerison ER, Ba ANN, Desai MM, Kryazhimskiy S. Chance and necessity in the pleiotropic consequences of adaptation for budding yeast. Nature Ecology & Evolution. 2020;: 1–20. doi:10.1038/s41559-020-1128-3

39. Travisano M, Vasi F, Lenski RE. Long-Term Experimental Evolution in *Escherichia coli*. III. Variation Among Replicate Populations in Correlated Responses to Novel Environments. Evolution. 1995;49: 189–200.

40. Nakatsu CH, Korona R, Lenski RE, de Bruijn FJ, Marsh TL, Forney LJ. Parallel and Divergent Genotypic Evolution in Experimental Populations of *Ralstonia* sp. J Bacteriol. 1998;180: 4325–4331.

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

42. Westphal LL, Lau J, Negro Z, Moreno IJ, Mohammed WI, Lee H, et al. Adaptation of *Escherichia coli* to long-term batch culture in various rich media. Research in Microbiology. 2018;169: 145–156. doi:10.1016/j.resmic.2018.01.003

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

44. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB et al. vegan: Community Ecology Package. 2013 (https://github.com/vegandevs/vegan).