Rapid evolution allows coexistence of highly divergent lineages within the same niche

Ben A. Ward[1,*] and Sinead Collins[2]

[1]School of Ocean and Earth Science, University of Southampton, Southampton, UK
[2]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FL, UK
[*]Corresponding author: b.a.ward@soton.ac.uk

Statement of authorship: BAW designed the study and performed the research, BAW and SC wrote the manuscript.

Running title: Convergent evolution allows coexistence

Article type: Letter

Number of words in abstract: 182

Number of words in main text: 4788

Number of references: 42

Number of figures, tables, and text boxes: 6

Abstract

Marine microbial ecosystems underpin global biogeochemical cycles and play a central role in the regulation of Earth's climate. These communities are extremely diverse, and their taxonomic composition varies considerably across ocean basins. It has however been difficult to establish links between taxonomic diversity and ecosystem function, and the ecological and evolutionary mechanisms underpinning taxonomic variation are not well understood. Here we use an individual-based eco-evolutionary model in which taxonomic diversity emerges as a consequence of evolutionary history. Using this model we are able to show that virtually unlimited genetic divergence can be supported in highly abundant and rapidly evolving assemblages, even in the absence of niche separation. With a steady stream of genetic, epigenetic and plastic heritable changes to phenotype, competitive exclusion may be weakened, allowing sustained coexistence of nearly neutral phenotypes with highly divergent lineages. This response may help to explain observed patterns of taxonomic diversity and functional redundancy - without recourse to hidden dimensions of niche partitioning. In light of these results we suggest that individual-level variability is a key driver of species coexistence and the maintenance of microbial biodiversity.

Keywords: Microbial | Biodiversity | Functional Redundancy | Coexistence | Neutral | Convergent Evolution

Introduction

Marine microbial communities are a fundamental driver of global biogeochemical cycles. Photosynthetic plankton form the energetic foundation of virtually all pelagic ecosystems, while cycling among broader networks of individuals plays a key role in the regulation of Earth's climate (Guidi et al., 2016). While individual metabolic processes and functional traits are often well correlated with environmental conditions (Thomas et al., 2012; Marañón et al., 2012; Ustick et al., 2021; Cohen et al., 2021), our ability to predict when and where individual taxa become important is complicated by an extremely high degree of taxonomic diversity. Indeed, among the approximately 10^{28} microbial cells living in the ocean (Flombaum et al., 2013), recent bioinformatic surveys have identified the existence of up to 150,000 genera of marine eukaryotes in the photic

layer alone (de Vargas et al., 2015).

In addition to this raw taxonomic diversity, globally important metabolisms and functional traits often appear broadly distributed across the tree of life, and in any given environment may be performed equally well by a large number of individual taxa. There is thus a high degree of functional redundancy in marine ecosystems (Louca et al., 2016), with the selection of traits and function occurring irrespective of taxonomic classification. For example, global metagenomic analysis points to high taxonomic dissimilarity among functionally very similar communities (Sunagawa et al., 2015; Louca et al., 2018). Likewise, single-cell genomic analyses have shown extremely high levels of genetic divergence among coexisting cells from the same taxonomic group (Rynearson and Armbrust, 2000; Kashtan et al., 2014).

This pattern of functional redundancy brings a new perspective to a longstanding question in marine microbial ecology, namely "*how it is possible for a number of species to coexist in a relatively isotropic or unstructured environment all competing for the same sorts of materials?*" (Hutchinson, 1961). As initially suggested by Hutchinson himself, many valid solutions to this "paradox" exist (Record et al., 2013). Species compete for (and are limited by) a broad range of chemical and biological factors that enable coexistence (Tilman, 1977). It is also clear that even a well-mixed ocean is neither isotropic nor unstructured (d'Ovidio et al., 2010). Spatial partitioning can thus occur at many different scales and ecological equilibrium is often prevented by external perturbations (Litchman et al., 2009) and internal dynamics (Huisman et al., 1999) such that competitive exclusion can be indefinitely postponed.

The mechanisms above work by partitioning coexisting species into different niches or by separating them in time or physical space, but do not address the potential for more than one species to coexist within a single niche. An alternative perspective, provided by the neutral theory of biodiversity (Hubbell, 2001), suggests that an unlimited degree of diversity can be maintained within the same niche if species have effectively identical fitness in their shared environment.

However, while the neutral theory provides a useful null hypothesis for observed patterns of diversity, it is often criticised on the grounds that even tiny differences in fitness must eventually lead to competitive exclusion (in the absence of other mechanisms; Hardin, 1960; Loreau, 2004). This is argued to be particularly true in microbial populations, for which huge population sizes tend to diminish the importance of stochastic effects that might delay exclusion (Louca et al., 2018).

While these ecological considerations suggest that neutrality is an unlikely outcome

in microbial communities, the degree to which species can coexist is also known to be affected by evolution (Kremer and Klausmeier, 2017). Laboratory cultures have been shown to display a high level of phenotypic convergence among traits that are strongly correlated with fitness (Blount et al., 2018), suggesting differences in many trait values and their associated fitness may be minimised through time. On one hand, convergent evolution can maintain diversity by eliminating the fitness differences that lead to exclusion (Scheffer and Nes, 2006; Hubbell, 2006). On the other, the same processes can eliminate complementary differences in phenotype that support coexistence, thus driving a steady decline in biodiversity (Shoresh et al., 2008; Sauterey et al., 2014). Among these modelling studies, a common feature is that the evolving community is represented as discrete populations differentiated by ecophysiological traits. This precludes the examination of potentially important processes of birth, death and mutation occurring at the individual level, or of the substantial variation known to underlie a given set of trait values. These individual level processes require consideration. For example, individual-based models (IBMs) have shown that phenotypic noise among individuals in large populations may be sufficient to add variation to the outcomes of local competitions, allowing extended coexistence of highly similar populations (or even populations of equal average fitness) within the same niche (Menden-Deuer et al., 2021). This suggests that competitive exclusion may proceed much more slowly given realistic levels of noise between genotype and phenotype when populations have the same or very similar average fitnesses (although this does not explain why small differences in average fitness would not eventually lead to exclusion).

In this article we address questions of functional and taxonomic diversity using an ecological and evolutionary (eco-evo) model that makes no prior assumptions regarding the differentiation of populations, species or ecotypes. Instead, the community is resolved at the individual level, with species and populations treated as emergent properties based on genetic rather than phenotypic distance. To achieve this the model includes a neutral genomic component that allows us to track descent and diversity under a range of scenarios. With simulations based on realistic ecophysiological parameters, we show that virtually unlimited diversity is a natural consequence of highly abundant evolving populations, with rapid trait evolution allowing lineages to avoid population bottlenecks despite sharp changes in environmental conditions.

An individual-based model of microbial evolution

The eco-evolutionary model provides a very simple representation of a closed marine microbial ecosystem, with state variables for nutrients, individual phytoplankton cells and organic detritus (Beckmann et al., 2019). The phytoplankton community is represented as a collection of individual cells that take up nutrients and increase in size as a function of their environmental conditions and ecophysiological traits. Cells divide into two daughter cells once they have doubled in biomass relative to a predefined threshold. Cells die through a stochastic process, producing organic detritus that is remineralised to inorganic nutrient at a fixed linear rate. Individual cells differ only in terms of their optimal temperature for growth, which is passed from generation to generation with some error, allowing for evolution by selection (Figure 1). Here, heritable variation is modelled as a random walk in a one-dimensional trait space, which represents the organisms' thermal optima. Heritable changes in trait values may be attributable to any combination of genetic and epigenetic mutations, as well as transgenerational plasticity that can affect the trait in question. These changes need not correspond directly to genetic point mutation rates, but rather to the per-generation rate of trait value change, which can be affected by all or some of these processes. Hereafter we refer to heritable trait changes generically as "mutation", regardless of the molecular cause of the change. A more detailed description of the individual-based model can be found in Appendix A and Beckmann et al. (2019).

In addition to the model components laid out by Beckmann et al. (2019), each simulated individual is assigned two heritable but ecologically-neutral characteristics: a binary string that undergoes a single random bit flip at each generation, and a 'colour trait' encoded as a three-element vector (red, green and blue) that also varies randomly from generation to generation (see Methods). The binary genome can be thought of as representing a two-base equivalent to a non-coding RNA or DNA sequence. Given that (a) genomes are identical at the point of division, (b) changes in the genomes are not under selection, and (c) genomes acquire mutations at a fixed rate, the binary genome can be used as a molecular clock. Changes through time accrue according to a 2-base version of the Jukes and Cantor (1969) model of base substitutions (Appendix A). The colour trait is included primarily for visualisation, with closely-related individuals appearing with similar colours (Figure 1).

Phenotypic and genotypic diversity within a single niche.

Beckmann et al. (2019) initially ran their model with a total nutrient load of 5 μM N and a constant environmental temperature of 15°C. The model converged to a steady state with individuals occupying a Gaussian distribution of thermal optima (15 ± 0.855°C) centred on the environmental temperature. We repeated this experiment, running the model for 1000 years and obtaining an identical trait distribution.

Using the neutral binary genome to estimate the genealogy of this population, Figure 2 shows the estimated pairwise distance matrix for 1000 individuals sampled at the end of the 1000 year simulation. Although the simulation only includes a single thermal niche, we see multiple distinct genotypic clusters coexisting within that niche, each with many tens of thousands of generations worth of genetic divergence from the others.

In order to explain this prolonged coexistence within a single niche, we will examine mechanisms of phenotypic and genetic diversity within the simulation.

Within niche phenotypic diversity

Figure 3a shows the simulated distribution of traits at the end of the 1000 year simulation. In a system without mutation, selection would drive the system towards dominance by a single optimally-adapted phenotype. This can be seen Figure 3b, in which the dashed line shows the expected net growth rates of different phenotypes at an ecological equilibrium (when nutrients are depleted to the minimum level required to support the best-adapted phenotype; Tilman, 1980). This fitness landscape shows that only the optimal phenotype can achieve a non-negative net growth rate, and thus all other phenotypes should eventually go extinct. While the associated timescales of extinction (calculated as the inverse of the net growth rate and shown by the solid line in Figure 3b) indicate that some phenotypes close to the optimum may take an extremely long time to go extinct, this is not sufficient to explain the trait distribution seen in Figure 3a – in a simulation of 1000 years duration the timescales of exclusion suggest a much narrower distribution of surviving phenotypes.

Figure 3c shows that the equilibrium trait distribution is maintained by a "mutation-selection balance" (Zhang and Hill, 2005), with imperfect heritability of traits serving to level out differences in net growth rate across the trait axis. A net excess of births over deaths around the optimal phenotype is exactly balanced by a mutational flux of individuals towards less favourable parts of the trait axis. This flux likewise supports a net excess of deaths relative to births further away from the optimal trait value.

Overall, the opposing forces of mutation and selection serve to flatten the fitness landscape (the solid line showing zero net growth in Figure 3c), which in theory allows unlimited coexistence across the trait space. In practice, the breadth of the trait distribution is limited by the increasing likelihood of extinction for less well-adapted (and hence less abundant) phenotypes. Nonetheless, the constant divergent flux of individuals provides a degree of standing trait variability.

Within niche genotypic diversity

Is this mutational flattening of the fitness landscape sufficient to support the sustained divergence of genotypes seen in Figure 2? To explore this question we modified the IBM to track the evolutionary trajectories of all simulated lineages, recording the time and phenotype (i.e. thermal optimum) associated with every cell division throughout the simulation.

This is shown in Figure 4a, which shows both the emergent abundance distribution during the first fifteen years of the 'constant temperature' simulation described above and the evolutionary trajectories of 20 individuals that were sampled during the fifteenth year of that simulation. Each of these sampled cells can be tracked back through the generations to the initial seed, providing an exact genealogy with complete information regarding phenotypic changes at each generation.

The plotted trajectories in Figure 4a indicate that individual lineages, while centred around the optimal temperature, show considerable phenotypic variability throughout the simulation. This pattern again occurs through a balance of mutation and selection, as lineages move around the optimal trait value in a constrained random walk. Here the introduction of trait variability is tempered at each generation as individuals with thermal optima further from the environmental temperature are less likely to successfully reproduce.

The simulated pattern of descent suggests two related consequences. First, individual lineages are not associated with a single constant fitness on which selection can consistently act over long periods (even though the trait itself may be strongly and consistently correlated with fitness). Second, different lineages tend to exhibit similar average fitness over reasonably long time scales (decades or more). As a consequence, our simulations show extended coexistence of divergent lineages (Figure 2). While such a high degree of lineage divergence should be expected within a homogeneous population (Kingman, 1982), it occurs here for a group of competing and evolving lineages with clear differences in phenotype and associated fitness. In the following, we will test

7

whether this mechanism also applies in a temporally-varying environment, under which changing conditions might serve to accelerate competitive exclusion.

Dynamic environmental forcing

Beckmann et al. (2019) explored the behaviour of the model in response to a number of alternative environmental forcing scenarios. We repeat those experiments here with identical model parameters (Table A.1), but over a slightly extended timescale of 15 years. Figure 4b-c shows the results of these simulations, which in all cases are consistent with the results presented by Beckmann et al. (2019).

In Figure 4b we introduced a sinusoidal annual cycle of $\pm 5^\circ$C on top of the mean temperature of 15 degrees (red lines). As was the case in a constant environment, the lineage tracking highlights a very high degree of lineage coexistence. Furthermore, while the 20 individuals sampled towards the end of the simulation are broadly distributed in terms of their thermal optima (between 13 and 17°C), they are descended from individuals with a narrower distribution of thermal optima early in the simulation. This is highlighted in Figure 5, which shows the 95th percentiles of the abundance distribution of all individuals throughout the simulation alongside the equivalent percentiles of the lineages sampled towards the end of the simulation. While the abundance distributions show that a significant number of individuals did adapt to the extremes of temperature, the lineage distributions show that very few of these survived to the end of the simulation. Adaptation to the extremes of temperature therefore appear to be an evolutionary dead-end in this simulation, with phenotypes closer to the mean temperature most likely to survive in the long run.

Finally, in panels c and d of Figure 4, we explored the response of the system to an abrupt change in the environmental forcing at t = 5 years. In panel c we instantly increased the average temperature by 5°C, while in panel d we added a rapidly oscillating (square wave) diurnal cycle of $\pm 5^\circ$C. The eco-evolutionary responses to these changes again reflect the findings of Beckmann et al. (2019), with the simulated trait distribution either adapting to the warmer temperature (panel c) or branching into two distinct ecotypes adapted to the warmer and colder extremes of the fluctuating temperature range (panel d).

In both cases, the plotted evolutionary trajectories reveal that the traits of of sampled lineages all began changing towards the new optimal traits *before* the change in environmental conditions. While these changes increased the likelihood of extinction

8

in the old environment, they provided a critical fitness advantage once the conditions changed.

This pattern of evolution is characteristic of a multiple-origin soft selective sweep (Hermisson and Pennings, 2017). When the environment changes (Figure 4b to c), standing phenotypic variation provides multiple seeds by which genetic variation can be carried through to the new environment, easing the severity of population bottlenecks and allowing greater coexistence than might otherwise be predicted from the competitive exclusion principle.

Comparison to a strictly neutral model

To test the degree to which evolution can alleviate the strength of population bottlenecks, we compare the simulated timescales of lineage coalescence (going backwards in time) to predictions of a strictly neutral model (Halley and Iwasa, 2011, and Appendix B). Figure 6a shows that in a constant environment the simulated pattern of coalescence closely follows the theoretical predictions, with over 90% of the lineages remaining distinct through 800 generations.

Coalescence is only slightly accelerated with the introduction of a seasonal temperature cycle (b), but the sudden change in temperature of 5°C introduces a population bottleneck (c), albeit with just under one third of the lineages successfully adapting to the change in conditions (also shown in Figure 4c). The population bottleneck is less severe when speciation is enabled through the introduction of a diel cycle (d). While these latter two experiments do lead to a significant loss of diversity, the introduced environmental perturbations are extremely harsh, with temperature changing by 5-10°C in an instant. In panels (e) and (f) we introduce more realistic (although arguably still severe) changes, adding a 0.5°C per year warming (from the end of year 5) to the experiments with a constant temperature (e) and an annual cycle (f). In the absence of the seasonal cycle this warming term had almost no effect on the pattern of coalescence (panel e). However, when introduced to a simulation with a seasonal cycle, the warming term led to a markedly more rapid loss of diversity. This likely occurs as yearly increases in temperature favour species adapted to the warmest part of the annual cycle over those adapted to the coldest temperatures.

Discussion

Perspectives on microbial life in the ocean are increasingly shaped by the vast amounts of molecular information made available by modern sequencing techniques (Mock et al., 2016). Despite a large and growing number of papers that provide realistic exceptions to the so-called paradox of the plankton (Record et al., 2013), patterns of taxonomic diversity are regularly interpreted through a perspective of competitive exclusion and niche partitioning. A high degree of coexistence is often attributed to (potentially cryptic) niche separation (Louca et al., 2018) - but this strictly requires one hidden niche dimension for every additional coexisting species at equilibrium.

The neutral theory of biodiversity (Hubbell, 2001) provides an alternative view, attributing patterns of taxonomic diversity to the stochastic nature of births and deaths. Clusters of distinct individuals can emerge in the absence of any selective pressures, driven by the random process of ecological drift. However, the importance of stochastic processes is thought to be diminished in extremely abundant populations, for which even relatively small increases in fitness lead to deterministic sweeps of beneficial mutations through the population, resetting genetic divergence to a low level (Louca et al., 2018). Given enough time, even small consistent differences in fitness will lead to competitive exclusion (Hardin, 1960), but the associated timescales can be surprisingly large even for appreciable differences in phenotype. Scheffer and Nes (2006), for example, showed the emergence and coexistence of similar (but not identical) clusters of phenotypes over several thousands of generations. Here we show a similar result, but with timescales of exclusion extended indefinitely as a consequence of a constant input of heritable variation of traits, arising from genetic, heritable epigenetic, and heritable plastic changes.

This pattern of indefinite (albeit stochastic) coexistence can be understood from two perspectives. From a phenotypic perspective, the ecological components of the model point to dominance by a single 'optimal' phenotype under constant environmental conditions (Figure 3b). However, the mutational flux of individuals from better to worse adapted phenotypes effectively flattens the fitness landscape (Figure 3c), allowing unlimited coexistence. Alternatively, from a lineage-based perspective, organisms do not have perfectly fixed traits from one generation to the next, and lineages thus occupy a distribution of traits around the optimal value. Over long periods, the average fitness of different lineages converge to the effectively neutral values, again allowing much longer periods of coexistence (set by population genetic rather than ecological considerations; Figure 6b).

10

These results are driven by a mutation-selection balance that requires a dependable and relatively high rate of heritable trait changes in comparison to the exclusion timescale. For our simulations, Figure 3 shows that deviations from the thermal optimum of up to $1°C$ are associated with exclusion timescales of over a year - which equates to several hundred generations for the microbial plankton under consideration. Into this system we included heritable trait changes in the thermal optimum as Gaussian noise with a standard deviation of $0.1°C$. While this may seem high, it is worth noting that thermal tolerance is affected by many genetic (Chakravarti et al., 2020) and otherwise heritable factors (McGuigan et al., 2021) and there are thus many potential pathways for this trait to evolve (Schaum et al., 2018). Thermal tolerance is also known to evolve extremely rapidly in response to environmental changes ($\sim$200 generations), even when such changes rely entirely on *de novo* variation and take place in asexual populations (Jin and Agustí, 2018; O'Donnell et al., 2018). Our simulated evolutionary trajectories (Beckmann et al., 2019) are not grossly out of alignment with responses observed in laboratory cultures (O'Donnell et al., 2018) or implied from field observations (Thomas et al., 2012). Further, running simulations with slower mutation rates prevented the model from showing any meaningful evolutionary response at all. Populations either remained unchanged (in response to sinusoidal forcing) or went extinct (in response to sudden temperature changes). Given the sheer size of microbial populations, and the ease with which they may generate the variation required to adapt extremely rapidly in laboratory experiments, the high rates of heritable trait change used in this model are reasonable.

It should however be noted that rarer and more stochastic trait changes might not lead to similar patterns of soft selective sweeps and extended coexistence. If a single large beneficial trait change occurs in isolation, it is likely to displace all other lineages over a timescale related to the associated increase in fitness. For example, we ran several simulations for which mutations occurred 10 times less frequently, but with a standard deviation $\sqrt{10}$ times larger. While this gave an identical expected trait distribution over many generations, the increased stochasticity of the simulation led to harder sweeps and rapid competitive exclusion in response to environmental change. Furthermore, evolution along a single trait axis (in this case thermal tolerance) represents presents a fairly large target for beneficial changes. It remains to be seen what patterns of coalescence will emerge in a model where evolutionary changes occur in multiple trait dimensions simultaneously. In a much larger multidimensional trait space beneficial changes are likely to occur much less predictably, potentially shifting the system towards harder

selective sweeps and stronger competitive exclusion.

These caveats notwithstanding, rapid evolution allows neutrality to emerge through a process of convergent and imperfect evolution and we see the sustained coexistence of phenotypically-similar but genetically-distinct lineages. This is a defining characteristic of functional redundancy (Louca et al., 2016; Louca et al., 2018). The assumptions of our model demonstrate that this phenomenon does not require the existence of additional hidden niche dimensions. Furthermore, our simulations suggest that high numbers of lineages are able to traverse even abrupt changes in environmental conditions (Figure 4), with the adaptive response to environmental changes underpinned by standing phenotypic variation, rather than the emergence of a single beneficial mutation. These patterns of evolution are characteristic of soft selective sweeps, which require either standing variation or a consistent supply of new beneficial mutations - both of which are extremely likely in highly abundant and rapidly reproducing microbial populations. Indeed, we were able to demonstrate the presence of soft sweeps in modelled populations on the order of only one million cells, somewhat less than the estimated 10^{27} *Prochlorococcus* cells currently alive in the ocean, or even the estimated effective population size of 10^{13} in a well-mixed parcel of sea water (Kashtan et al., 2014).

Despite the inclusion of selection and environmental variability, our comparisons to the neutral model of coalescence suggest that strong population bottlenecks are only likely to occur under extremely rapid environmental changes that seem unlikely to occur over large spatial scales in a well-mixed ocean. Several of our simulations remain consistent with a strictly neutral theory (Kingman, 1982; Halley and Iwasa, 2011), which predicts that the expected timescale of diversity loss will be proportional to the effective population size (Equation B.5). For the aforementioned well-mixed population of *Prochlorococcus*, this is much longer than required to explain the observed (Kashtan et al., 2014) millions of years of divergence.

Our findings suggest that rapid evolution likely plays a key role in the coexistence of phenotypically similar but genetically distinct species in microbial communities, with functional redundancy emerging through convergent evolution. Nonetheless, our simulations remain highly idealised, in particular neglecting to account for dispersal and mixing of communities in a three-dimensional environment. Further work is therefore required to explore the significance of soft selective sweeps in a metacommunity context.
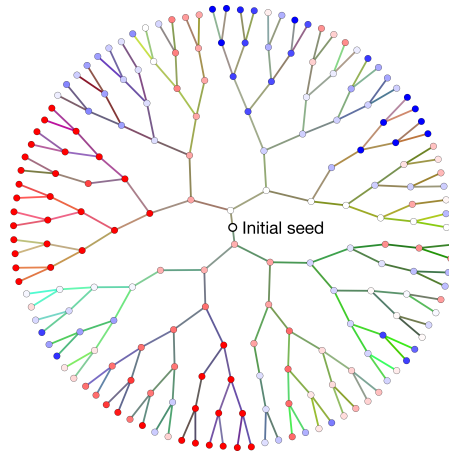
Figures

Figure 1: Genealogy in the IBM after 7 days growth at a constant environmental temperature. Terminal nodes at the perimeter represent live cells that have descended from the initial seed at the centre. Each non-terminal node represents a cell division, with branch lengths linearly proportional to the time between divisions. Nodes are coloured according to the thermal optimum of each dividing cell (red prefers warmer, blue prefers colder). Branch colours correspond to value of the neutral rgb gene. Note that branch colours change gradually along branches, such that related agents have similar colours. Extinct lineages are not shown.
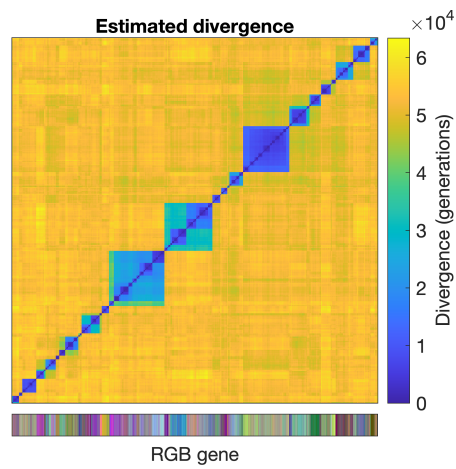
Figure 2: Estimated divergence matrix for 1000 cells sampled at the end of a 1000 year simulation, as derived from from the binary genome. The estimated number of generations since the MRCA is shown according to the right-hand colour scale. The lower colour scale shows each individual's neutral colour trait.
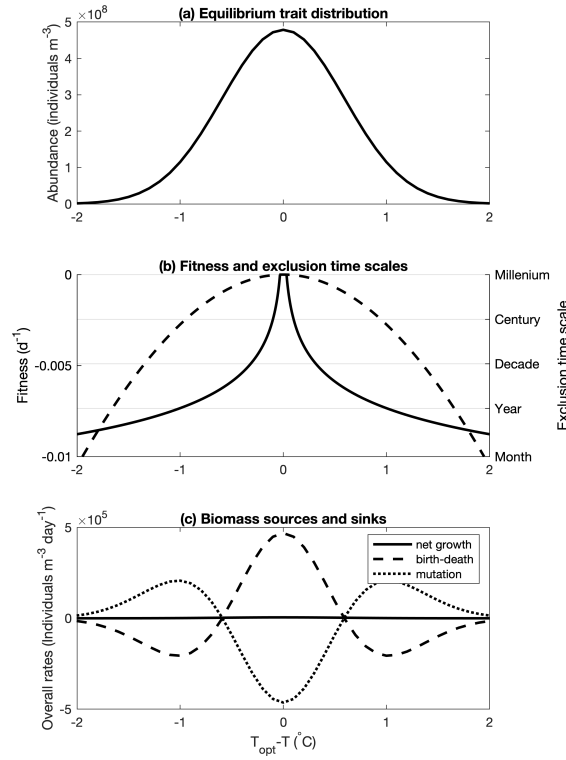
Figure 3: Trait distribution and mechanisms of coexistence. Panel (a) shows the eco-evolutionary equilibrium distribution of phenotypes as a function of the thermal optimum minus the environmental temperature ($T_{opt} - T$). Panel (b) shows the equilibrium net growth rate (or fitness landscape) in the absence of mutations (dashed line) and the associated time scales of competitive exclusion (solid line; calculated under the assumption that limiting nutrients are drawn down to the equilibrium requirement of the best-adapted species). Time scales of competitive exclusion are calculated as the inverse of net growth rate. Panel (c) shows the equilibrium balance of births-deaths vs. mutation. Mutation acts as a sink for the best-adapted phenotypes and as a source for maladapted phenotypes, thus supporting a broad distribution of traits with equal (zero) fitness.
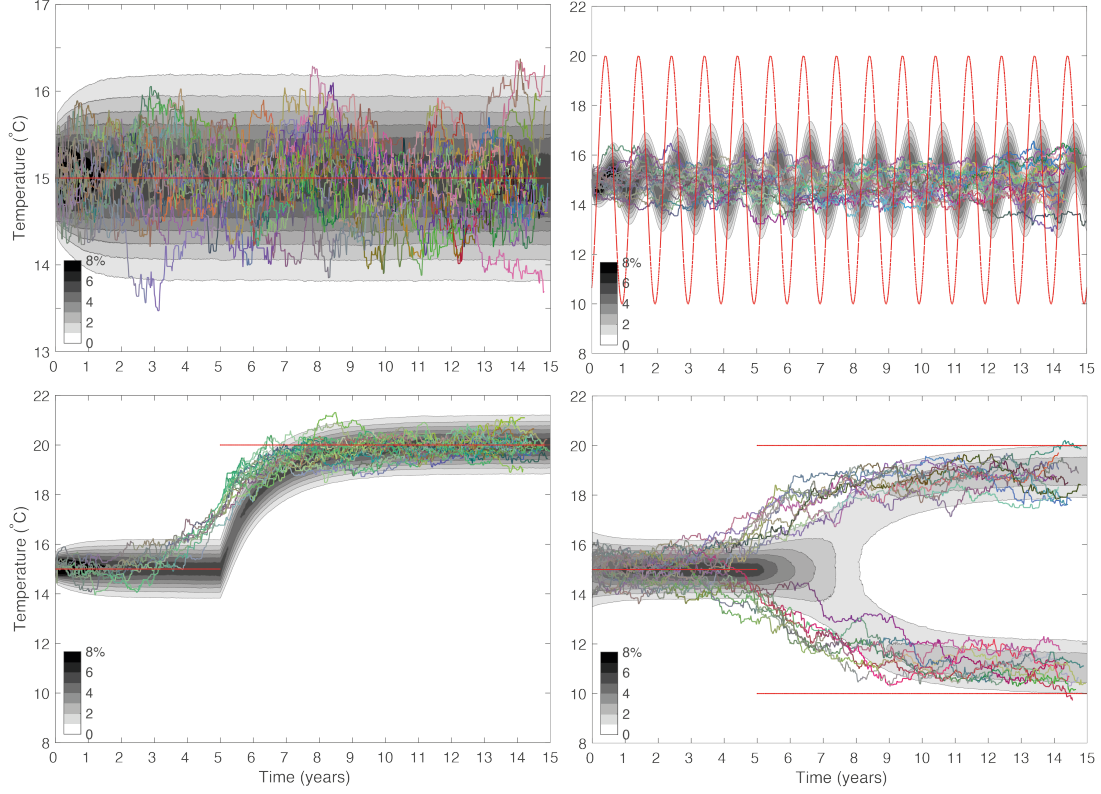
Figure 4: Eco-evolutionary plankton dynamics during three initial 15 year simulations with the IBM. Each simulation was seeded at $t = 0$ with a single cell with a thermal optimum of 15°C. Grayscale contours in each panel show the distribution of individuals among phenotypes through time. The branching lines show the genealogy of 50 cells sampled at random from cells alive during the final year of the simulation. Thermal phenotype is shown with the y coordinate, time of division with the x coordinate. Branch colours correspond to value of the neutral colour trait. The red lines show the range of environmental temperatures throughout each simulation: Panel a - constant temperature; panel b - sinusoidally varying temperature (period 1 year, amplitude 10°C); panel c - constant temperature until $t = 5$ years, switching to a diurnal square wave (period 1 day, amplitude 10°C).
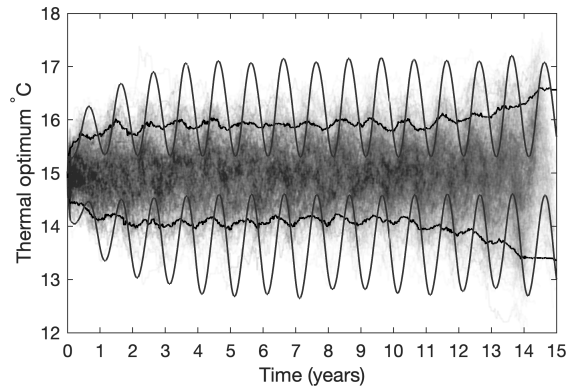
16

Figure 5: Evolutionary history of cells sampled in the last year of the simulation compared to abundance distributions throughout the simulation. The smooth black lines show the 95th percentile of the abundance distribution at each point in the simulation. Evolutionary trajectories of 1000 cells sampled during the final year of a simulation are shown as grey lines. The 95th percentiles of this distribution are shown by the jagged black lines. Most of the cells sampled in the last year of the simulation (including those adapted to extremes of the temperature range) are descended from ancestors with thermal optima closer to the mean environmental temperature. Most of the cells that were adapted to extremes of temperature early in the simulation do not have descendants alive at the end of the simulation.
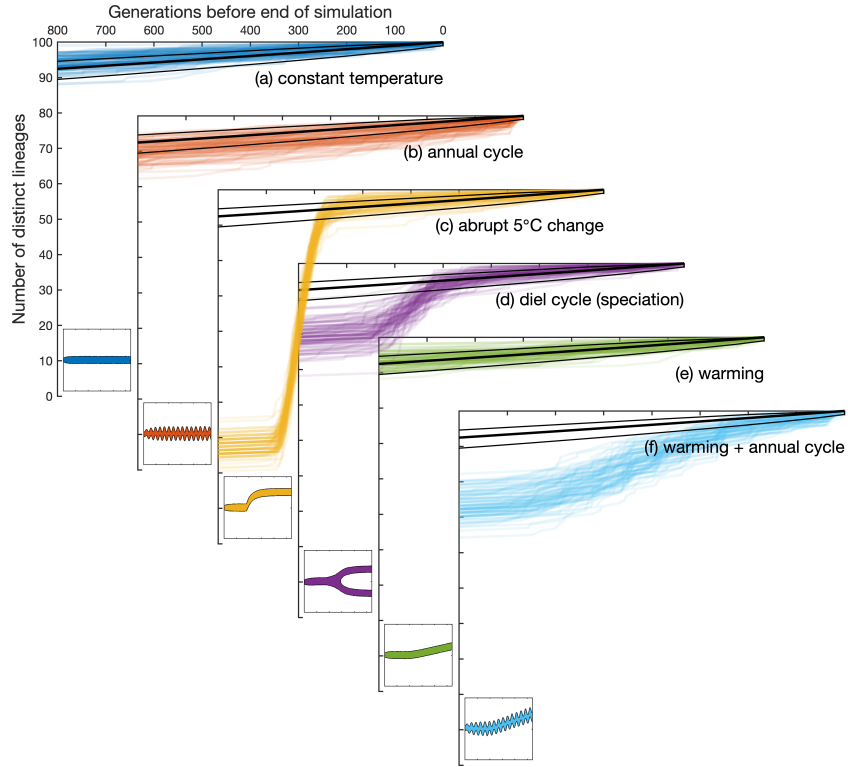
17

Figure 6: Patterns of coalescence under different environmental scenarios. Axes a-d correspond to the experiments shown in Figure 4. Axes e-f show results from two additional experiments: (e) 0.5°C per year warming applied from the end of year 5, (f) as for e, but with an annual cycle of ±5°C. In each case coalescence patterns are shown for 100 randomly selected phylogenies, in comparison the neutral model (black lines, mean±1 s.d.). Inset panels show biomass as a function of time (x axis) and thermal optimum (y axis) for each experiment.

Acknowledgements

References

Beckmann, A., C.-E. Schaum, and I. Hense (2019). "Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems". In: *bioRxiv* 468, pp. 60–71.

Blount, Z. D., R. E. Lenski, and J. B. Losos (2021/10/19 2018). "Contingency and determinism in evolution: Replaying life's tape". In: *Science* 362.6415, eaam5979.

Chakravarti, L. J., P. Buerger, R. A. Levin, and M. J. H. van Oppen (2021/10/19 2020). "Gene regulation underpinning increased thermal tolerance in a laboratory-evolved coral photosymbiont". In: *Molecular Ecology* 29.9, pp. 1684–1703.

Cohen, N. R., M. R. McIlvin, D. M. Moran, et al. (2021). "Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean". In: *Nature Microbiology* 6.2, pp. 173–186.

d'Ovidio, F., S. De Monte, S. Alvain, Y. Dandonneau, and M. Lévy (2010). "Fluid dynamical niches of phytoplankton types". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.43, pp. 18366–183700.

de Vargas, C., S. Audic, N. Henry, et al. (2015). "Eukaryotic plankton diversity in the sunlit ocean". In: *Science* 348.6237, 1261605.

Flombaum, P., J. L. Gallegos, R. A. Gordillo, et al. (2013). "Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*". In: *Proceedings of the National Academy of Sciences of the United States of America* 24.10.1073/pnas.1307701110, pp. 9824–982.

Guidi, L., S. Chaffron, L. Bittner, et al. (2016). "Plankton networks driving carbon export in the oligotrophic ocean". In: *Nature* 532.7600, pp. 465–470.

Halley, J. M. and Y. Iwasa (2011). "Neutral theory as a predictor of avifaunal extinctions after habitat loss". In: *Proceedings of the National Academy of Sciences* 108.6, pp. 2316–2321.

Hardin, G. (1960). "The competitive exclusion principle". In: *Science* 131, pp. 1292–1297.

Hermisson, J. and P. S. Pennings (2017). "Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation". In: *Methods in Ecology and Evolution* 8.6, pp. 700–716.

Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeographyheory of Biodiversity and Biogeography*. Princeton, New Jersey: Princeton University Press.

Hubbell, S. P. (2006). "Neutral theory and the evolution of ecological equivalence". In: *Ecology* 87.6, pp. 1387–1398.

Huisman, J., P. van Oostveen, and F. J. Weissing (1999). "Critical depth and critical turbulence: Two different mechanisms for the development of phytoplankton blooms". In: *Limnology and Oceanography* 44.7, pp. 1781–1787.

Hutchinson, G. E. (1961). "The paradox of the plankton". In: *American Naturalist* 95.882, pp. 137–145.

Jin, P. and S. Agustí (2018). "Fast adaptation of tropical diatoms to increased warming with trade-offs". In: *Scientific Reports* 8.1, p. 17771.

Jukes, T. H. and C. C. Cantor (1969). "CHAPTER 24 - Evolution of Protein Molecules". In: *Mammalian Protein Metabolism*. Ed. by H. MUNRO. Academic Press, pp. 21–132.

Kashtan, N., S. E. Roggensack, S. Rodrigue, et al. (2014). "Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild Prochlorococcus". In: *Science* 344.6182, pp. 416–420.

Kingman, J. F. C. (1982). "The coalescent". In: *Stochastic properties and their applications* 13, pp. 235–248.

Kremer, C. T. and C. A. Klausmeier (2017). "Species packing in eco-evolutionary models of seasonally fluctuating environments". In: *Ecology Letters* 20.9, pp. 1158–1168.

Litchman, E., C. A. Klausmeier, and K. Yoshiyama (2009). "Contrasting size evolution in marine and freshwater diatoms". In: *Proceedings of the National*

Academy of Sciences of the United States of America 106.doi: 10.1073/pnas.0810891106, pp. 2665–2670.

Loreau, M. (2004). "Does functional redundancy exist?" In: *Oikos* 104.3, pp. 606–611.

Louca, S., L. W. Parfrey, and M. Doebeli (2016). "Decoupling function and taxonomy in the global ocean microbiome". In: *Science* 353.6305, pp. 1272–1277.

Louca, S., M. F. Polz, F. Mazel, et al. (2018). "Function and functional redundancy in microbial systems". In: *Nature Ecology & Evolution* 2.6, pp. 936–943.

Marañón, E., P. Cermeño, M. Latasa, and R. D. Tadonléké (2012). "Temperature, resources, and phytoplankton size structure in the ocean". In: *Limnology and Oceanography* 5, pp. 1266–1278.

McGuigan, K., A. A. Hoffmann, and C. M. Sgrò (2021/10/19 2021). "How is epigenetics predicted to contribute to climate change adaptation? What evidence do we need?" In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1826, p. 20200119.

Menden-Deuer, S., J. Rowlett, M. Nursultanov, S. Collins, and T. Rynearson (Aug. 2021). "Biodiversity of marine microbes is safeguarded by phenotypic heterogeneity in ecological traits". In: *PLOS ONE* 16.8, e0254799–.

Mock, T., S. J. Daines, R. Geider, et al. (2016). "Bridging the gap between omics and earth system science to better understand how environmental change impacts marine microbes". In: *Global Change Biology* 22.1, pp. 61–75.

Monod, J. (1950). "La technique de culture continue, théorie et applications". In: *Annales de l'Institut Pasteur (paris)* 79, pp. 390–410.

O'Donnell, D. R., C. R. Hamman, E. C. Johnson, et al. (2018). "Rapid thermal adaptation in a marine diatom reveals constraints and trade-offs". In: *Global Change Biology* 24.10, pp. 4554–4565.

Record, N. R., A. J. Pershing, and F. Maps (June 2013). "The paradox of the "paradox of the plankton"". In: *ICES Journal of Marine Science* 71.2, pp. 236–240.

Rynearson, T. A. and E. V. Armbrust (2021/10/19 2000). "DNA fingerprinting reveals extensive genetic diversity in a field population of the centric diatom Ditylum brightwellii". In: *Limnology and Oceanography* 45.6, pp. 1329–1340.

Sauterey, B., B. A. Ward, M. J. Follows, C. Bowler, and D. Claessen (Oct. 2014). "When everything is not everywhere but species evolve: an alternative method to model adaptive properties of marine ecosystems". In: *Journal of Plankton Research* 37.1, pp. 28–47.

Schaum, C. .-., A. Buckling, N. Smirnoff, D. J. Studholme, and G. Yvon-Durocher (2018). "Environmental fluctuations accelerate molecular evolution of thermal tolerance in a marine diatom". In: *Nature Communications* 9.1, p. 1719.

Scheffer, M. and E. H. van Nes (2006). "Self-organized similarity, the evolutionary emergence of groups of similar species". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.16, pp. 6230–6235.

Shoresh, N., M. Hegreness, and R. Kishony (2008). "Evolution exacerbates the paradox of the plankton". In: *Proceedings of the National Academy of Sciences* 105.34, pp. 12365–12369.

Sunagawa, S., L. P. Coelho, S. Chaffron, et al. (2015). "Structure and function of the global ocean microbiome". In: *Science* 348.6237. Ed. by E. Boss, C. Bowler, M. Follows, et al.

Thomas, M. K., C. T. Kremer, C. A. Klausmeier, and E. Litchman (Nov. 2012). "A Global Pattern of Thermal Adaptation in Marine Phytoplankton". In: *Science* 338.6110, p. 1085.

Tilman, D. (1980). "A graphical-mechanistic approach to competition and predation". In: *American Naturalist* 116, pp. 362–393.

Tilman, D. (1977). "Resource competition between planktonic algae: an experimental and theoretical approach". In: *Ecology* 58, pp. 338–348.

Ustick, L. J., A. A. Larkin, C. A. Garcia, et al. (2021/10/19 2021). "Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation". In: *Science* 372.6539, pp. 287–291.

Zhang, X.-S. and W. G. Hill (2005). "Genetic variability under mutation selection balance". In: *Trends in Ecology & Evolution* 20.9, pp. 468–470.

Appendix

A   Model Description

A.1   Individual-based Eco-Evo model

The eco-evo model we develop builds upon the individual-based model (IBM) presented by Beckmann et al. (2019). The model represents a closed system, in which individual phytoplankton (b_i) grow as a function of temperature and nutrient availability, divide and die stochastically. Dead phytoplankton enter a detrital pool (D), which is converted back to inorganic nutrient (N) via a linear remineralisation term.

Cellular growth of individual phytoplankton (μ_i) is defined in relation to a theoretical maximum of μ_0 that is modified by temperature- and nutrient-dependent functions ($\mathcal{F}_T$ and $\mathcal{F}_N$).

$$\frac{db_i}{dt} = \mu_i = \mu_0 \cdot b_i \cdot \mathcal{F}_T \cdot \mathcal{F}_N \tag{A.1}$$

Here μ_0 is the maximum doubling rate and b_0 is the reference cellular biomass (Beckmann et al., 2019).

The thermal tolerance function decreases growth rate as the environmental temperature T deviates from a phytoplankton's thermal optimum (T_{opt}). The breadth of the associated thermal niche is given by θ.

$$\mathcal{F}_T(T) = \exp\left[-\left(\frac{T - T_{opt}}{\theta}\right)^2 \right] \tag{A.2}$$

Nutrient limitation is implemented with a Monod (1950) function, with a half-saturation constant of k_N.

$$\mathcal{F}_N(N) = \frac{N}{k_N + N} \tag{A.3}$$

Individuals increase their cellular biomass at a rate set by their physiological traits (μ_0, T_{opt}, etc) and the environmental conditions (T and N). Each cell grows until it reaches or surpasses a division threshold, which is set to twice its minimum viable biomass of b_0. When this point is reached, the cell's biomass is divided equally between two daughter cells.

Mortality is applied stochastically, with cells having a fixed probability of death (γ_0), at each time step. The number of live cells in the model thus changes according to the balance of agent divisions and agent deaths at each time step.

The overall phytoplankton biomass concentration in the model is calculated diagnostically as the sum of the biomass of all live cells.

$$P = \frac{1}{V} \sum_{i=1}^{M} b_i \tag{A.4}$$

where M is the number of live cells and V is the volume of the modelled culture. Note that we regulate the number of cells in the model by controlling the culture volume. We do not use the concept of super-individuals.

In contrast to the phytoplankton, which are treated as a collection of individuals, the nutrient and detrital pools are treated as homogeneous bulk variables. At each time step, uptake from the nutrient pool is taken as the sum of uptake by all individual agents, while production of detritus is taken as the combined biomass of all dying agents. Remineralisation from the detrital pool to the nutrient pool proceeds as a linear function of detrital biomass at each time step, with a mass specific rate of τ.

$$N_{t+1} = N + (\tau D - \sum_{i=1}^{M} \mu_i) \Delta t \tag{A.5}$$

$$D_{t+1} = D + (\sum_{i=i_{\text{die}}} b_i - \tau D) \Delta t \tag{A.6}$$

here i_{die} is the index of all cells dying in a particular time step.

Evolution Trait variation and inheritance are implemented in the IBM by assigning each newly divided agent the thermal optimum of its parent, perturbed by a value drawn from a Gaussian distribution with mean zero and standard deviation σ_M.

$$T'_{opt} = T_{opt} + \sigma_M$$

Changes in the thermal optimum affect the likelihood of survival by increasing or decreasing the agent's growth rate, with better adapted agents more likely to be reproduced in each subsequent generation.

Model parameter	Symbol	Value	Units
Total nutrient load	N_t	5	mmol N m^{-3}
Maximum cellular growth rate	μ_0	$\ln 2$	d^{-1}
Minimum cellular biomass	b_0	5×10^{-10}	mmol N
Nutrient half-saturation	k_N	0.15	mmol N m^{-3}
Linear mortality rate	γ_0	0.1	d^{-1}
Remineralisation rate	τ	0.25	d^{-1}
Thermal optimum	T_{opt}	variable	°C
Breadth of thermal niche	θ	6	°C
Standard deviation of 'mutations'	σ_M	0.1	°C
Time step	Δt	1/24	d
Volume of growth culture	V	10^{-4}	m^3

Table A.1: Standard model parameters.

A.2 Phylogeny

A.2.1 Lineage tracking.

At each cell division we assign the two daughter cells a unique identity number. We also record the thermal optimum, time of division and the identity of the parent cell. This record is purged of extinct lineages at the end of each year in order to maintain the size of the associated files at a manageable level. This approach allows us to reconstruct evolutionary trajectories in the model with complete accuracy, as shown in Figure 4.

A.2.2 Ecologically-neutral colour trait.

While precise, the lineage tracking approach is also very expensive computationally and produces vast amounts of data. As an alternative approach, we added an ecologically-neutral colour trait to identify closely related individuals.

The neutral colour trait is encoded as a heritable three-element vector that corresponds to a unique colour in the red-green-blue (rgb) colour space.

$$\overrightarrow{\text{rgb}} = [\text{r}, \text{g}, \text{b}]$$

The rgb vector is replicated at each cell division and each element then immediately

undergoes a mutation, drawn from the standard normal distribution $(\phi \sim \mathcal{N}[0,1])$.

$$\overrightarrow{\mathrm{rgb}}' = \overrightarrow{\mathrm{rgb}} + \vec{\phi}$$

As the value of the rgb vector has no effect on the fitness of the individual, changes in the rgb genome through generations are mathematically equivalent to a Gaussian random walk in a three-dimensional space. The expected euclidean distance between two rgb vectors $\hat{d}_{\mathrm{rgb}}$ is therefore given as a function of the number of generations (t_{gen}) since their most recent common ancestor

$$\hat{d}_{\mathrm{rgb}} = \sqrt{4t_{\mathrm{gen}}} \cdot c \tag{A.7}$$

With a standard deviation of

$$\sigma_{\hat{d}_{rgb}} = \sqrt{t_{\mathrm{gen}}} \tag{A.8}$$

Here c is a correction factor that accounts for the number of dimensions, n_{rgb}, using the ratio of two gamma functions.

$$c = \frac{\Gamma(\frac{n_{rgb}+1}{2})}{\Gamma(\frac{n_{rgb}}{2})}$$

While the distance between agents in the rgb colour space can be used to estimate the time since their most recent common ancestor, the ratio of equations A.8 and A.7 suggest an expected coefficient of variation (c.v.) of $(2c)^{-1}$. For a three-element rgb vector, the expected euclidean distance will be broadly distributed, with a standard deviation of $\pm 44\%$ of the expected value. Even if the rgb vector is extended to include 50 dimensions, the c.v. only drops to $\pm 10\%$. This is somewhat imprecise (as shown in Figure A.1a), but the colour trait will be useful for identifying closely related individuals (e.g. Figure 1).

A.2.3 Binary genome.

While the neutral colour trait is useful for visualisation, it lacks the precision required to accurately track descent in the model. To achieve this we instead turn to the binary genome, for which each individual in the simulation is assigned a binary string of $L = 2150$ bits.

$$\overrightarrow{\mathrm{bin}} = [\mathrm{bin}_1, \mathrm{bin}_2, ..., \mathrm{bin}_L]$$
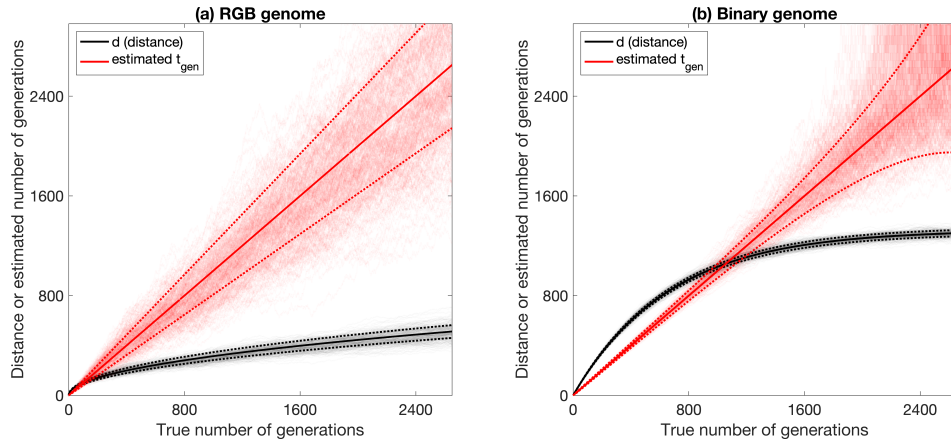
26

Figure A.1: Theoretical and simulated accumulation of differences in the neutral genomes. Panel (a): RGB genome. The black lines shows the expected accumulation of Euclidean distances between the 50-element rgb genomes (dashed lines are ±1 standard deviation). Panel (b): Binary genome. The black lines show the expected rate of accumulation of bitwise differences between 2350 bit binary genomes (±1 standard deviation). In both cases the thick red lines indicate the estimated number of generations for a given distance (±1 standard deviation). Pale lines show the simulated distribution of pairwise distances or estimated divergences among 25 individuals. While both genomes are encoded as a 50-element double precision vectors, it is clear that the binary genome gives a much more precise estimate of the number of generations, as long as the true number of generations is less than approximately half the number of bits.

In practice, the long binary string can be efficiently encoded as a 50-element vector of floating point values, with 53 bits stored in the significand of each double precision element. (We could have stored 64 bits as unsigned integer values, but this was not computationally efficient given our code structure.)

At each generation the binary genome is inherited from the parent cell and undergoes a single random bit-flip with a probability of $p_{\mathrm{mut}} = 1$. The bit to be flipped is drawn from a discrete uniform distribution; $R_{\mathrm{bin}} \sim \mathcal{U}[1, L]$.

$$
\mathrm{bin}'_i = \begin{cases} 1 - \mathrm{bin_i} & \text{if } R_{\mathrm{bin}} = i \\ \mathrm{bin_i} & \text{else} \end{cases}
$$

With one randomly-selected bit flipped at an average rate of once every $1/p_{mut}$ generations, the expected normalised Hamming distance between two binary genomes ($\hat{d}_{bin}$) is given as a function of the number of generations (t_{gen}) since their most recent common ancestor.

$$
\hat{d}_{bin} = \frac{1}{2}[1 - \exp^{(-\frac{4 \cdot t_{\mathrm{gen}} \cdot p_{\mathrm{mut}}}{L})}] \tag{A.9}
$$

with a standard deviation of

$$
\sigma_{\hat{d}_{bin}} = \sqrt{\frac{\hat{d}_{bin}(1 - \hat{d}_{bin})}{L}} \tag{A.10}
$$

These two equations (visualised in Figure A.1b) show that $\hat{d}_{bin}$ increases predictably with the number of generations, saturating at 0.5 as the number of mutations approaches the length of the binary genome (L). The non-linearity of the apparent trend is attributable to unobservable multiple flips of the same bits (homoplasy), as predicted by the two-base Jukes and Cantor model (black line).

It is also clear that $\hat{d}_{bin}$ increases in a much more predictable way than $\hat{d}_{rgb}$ (as long as the number of mutations remains less than approximately half the number of bits in the binary genome). This makes it a much better candidate for use as a molecular clock.

Accordingly, the estimated number of generations, $\hat{t}_{gen}$ since the divergence of any two lineages can be estimated from the simulated Hamming distance, d, between their binary genomes.

$$
\hat{t}_{\mathrm{gen}} = -\frac{1}{4p_{\mathrm{mut}}} \ln(1 - 2d) \tag{A.11}
$$

28

with standard deviation

$$\sigma_{\hat{t}_{\text{gen}}} = \frac{1}{p_{\text{mut}}} \sqrt{\frac{d(1-d)}{4L(1-2d)^2}} \tag{A.12}$$

Equation A.12 and Figure A.1b demonstrate that the binary genome can be used to estimate divergence with a high degree of precision, as long as $t_{gen} < L/2$ (N.B. longer simulations can be resolved by decreasing the probability (p_{mut}) of a bit flip at each generation).

The demonstrated precision of the binary clock (Figure A.1) allows us to reconstruct the simulated phylogeny without the expense of recording every single agent division. The basic principles of the binary clock are shown in Figure A.2. The dendrogram in Panel a shows the estimated phylogenetic tree for 100 agents sampled from a simulation similar to the one shown in Figure 4d (but with a much smaller population size of ∼5,000 to allow a more structured tree). Panel b shows the first 128-bits of the corresponding binary genomes (one row for each of agent). The known distance matrix from the lineage tracking is compared to the equivalent distance matrix estimated from the binary genome in Figure A.3.
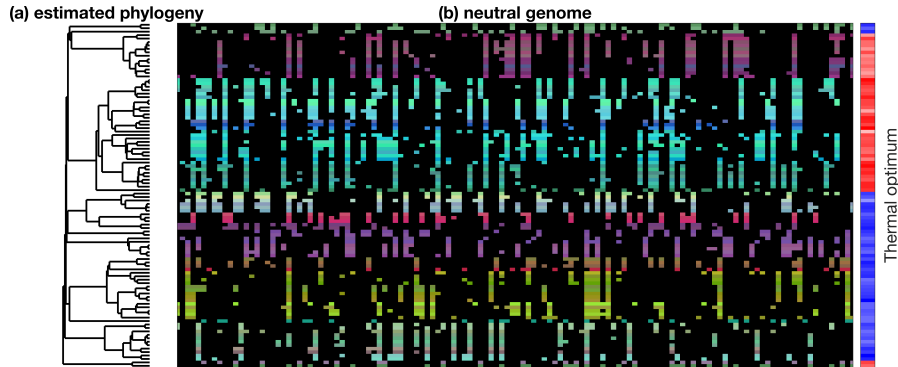

Figure A.2: Phylogeny and neutral genome of 100 individuals sampled at the end of a simulation similar to the one shown in Figure 4d (but with a population of only ∼5,000). The dendrogram in panel (a) represents an estimated phylogeny derived from the binary genomes. Panel (b) shows the first 128 bits of the associated binary genomes. Zeros are black, while ones shown with their neutral colour trait. The right-hand colour scale show the thermal optima of each individual (red = hot, blue = cold).
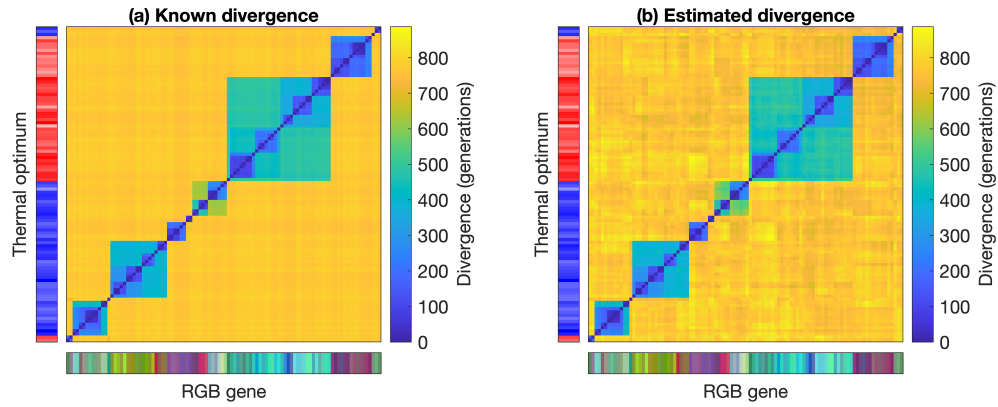
Figure A.3: Heat maps showing pairwise distance matrices for the same 100 cells presented in Figure A.2. Panel (a) shows known distances based on the lineage tracking. Panel (b) shows distances estimated from the neutral binary genomes. In each panel the left-hand colour scale shows the thermal optima of the sampled cells (red/blue = warm/cold adapted). The right-hand colour scale shows the genealogical divergence in generations. The bottom colour scale shows the neutral colour traits of the sampled cells.

B   Neutral model of lineage coalescence

Simulated rates of coalescence through time in Figure 6 are compared to predictions of a neutral theory model (Kingman, 1982; Halley and Iwasa, 2011). Going backwards in time, for a population of N individuals the per generation probability of a single coalescence event among k lineages is given by

$$p = \frac{k(k-1)}{2N} \tag{B.1}$$

This gives an expected waiting time for coalescence T (in generations) of

$$\mu_T = p^{-1} \tag{B.2}$$

with a standard deviation of

$$\sigma_T = \sqrt{\frac{1-p}{p^2}} \tag{B.3}$$

The expected number of lineages can also be expressed as a function of time t (in generations),

$$k(t) = \frac{k_0}{1 + \frac{t}{t_{\text{half}}}} \tag{B.4}$$

where k_0 is the number of sampled lineages at $t = 0$ and t_{half} is given by

$$t_{\text{half}} = \frac{2N}{k_0} \tag{B.5}$$