

Constraints on global plankton dispersal

B.A. Ward, B.B. Cael, C.R. Young and S. Collins

Ecotypes - adapted to different environments - but also many more genotypes.

How evolutionarily stable are these genotypes? Do genetic differences represent stable niche separation?

Observations (Kashtan 2014) suggest very hundreds of subpopulations with millions of years genetic dispersal - hence ecologically meaningful.

Clear(?) succession of genetically distinct subpopulations through time

Key question is how much gene flow between environments/niches???

“Everything is everywhere, but the environment selects” (?). (Also Beijerinck: one particular species of bacteria found anywhere on Earth, provided environmental requirements are met... ?). Oceanic plankton are sufficiently connected such that all species have the potential to colonise all regions. The community that we see in each location is thus determined by ecological selection.

? : argue that among small organisms (< 1 mm), this connectivity is facilitated by huge population sizes. At local scale, diversity of small species exceeds that of larger species, but at the global scale, diversity of larger species is greater, driven by endemic differences not seen in smaller organisms.

? : “Are there any species smaller than 1 mm?” [Showed empirical trends and used population model. Need to read again!]

Lagrangian studies have been equivocal. ? used an agent-based model to estimate the dispersal and mutation of 100,000 ‘super-individuals’. They found that realistic rates of mutation were enough to sustain several biogeographic provinces, each with a distinct genetic signature, even in the face of realistic ocean mixing. Genotypes found within these provinces were not distributed globally, which they suggest is in conflict with the notion that everything is everywhere. However, it seems likely that the model overestimates the rate at which local diversity is lost by genetic drift (i.e. coalescence?? ?), because the modelled population size (10^5 agents) is so small relative to the true population size ($\sim 10^{27}$). [Fixation expected in 273 years, rather than $\sim 10^{24}$ years].

? used a (biologically inert) particle tracking model in combination with Dijkstra’s ‘shortest-path algorithm’ ? to show that pathways exist to connect the global ocean on timescales of a decade or less. Even though some of the shortest paths are quite unlikely, the huge size of planktonic populations is again invoked

to suggest that the pathways should be followed by at least some individuals. ? do not account explicitly for population size, and it remains to be seen how population size and biogeography might affect the predicted patterns of connectivity.

Stochastic approach that accounts for the enormous and variable size of plankton (meta) populations, the potential for stochastic dispersal, and/or a representation of natural selection.

1 Experiments

To assess the rate of planktonic dispersal across the global ocean, we developed a model that tracks the abundance of different subpopulations in a globally distributed metapopulation (?). Throughout the simulation, each of the 60,646 surface grid boxes supports a predefined but spatially variable carrying capacity N (Figure ??a). This is derived from the estimated global distribution of 0.6 micron diameter *Prochlorococcus* cells (?). At the beginning of the simulation, a resident subpopulation is assumed to have population frequency of 1 throughout the global ocean. However, at each of 94 ‘seed locations’ distributed more or less evenly around the ocean, the resident population is replaced with a taxonomically distinct (but ecologically identical) local subpopulation.

From this initial condition, the model is integrated for 100 years in discrete time. Every six hours, plankton cells are dispersed by the ocean circulation. We initially considered a scenario where cells are transported exclusively within the surface layer. Every 24 hours the metapopulation is assumed to reproduce, with the new metapopulation drawn at random from a multinomial distribution, with probabilities given by the relative abundance of each subpopulation. In regions where a subpopulation is present in high abundance, this stochastic process has no significant effect on the relative abundance of different subpopulations, but it introduces a meaningful chance of local extinction wherever the overall carrying capacity is low (such as at edges of a subpopulations spatial distribution). In the following we discuss the results from simulations with time-invariant temperatures and carrying capacities. We also performed simulations where these variables followed a seasonal cycle (dashed lines in Figure ??), finding that the results were not overly sensitive to the change.

The blue line in Figure ??a shows the timescales over which the 94 seed subpopulations become connected with the rest of the ocean. Largely in agreement with ?, we find that almost 90% of the surface ocean is connected within a decade. **This could be because of increased diffusivities associated with the Eulerian approach used here. But could also be because JW16 discard all indirect shortest paths less than 1 year. Their distribution of ‘raw’ paths is skewed towards zero (but this could be because longer connections are**

missed)???.

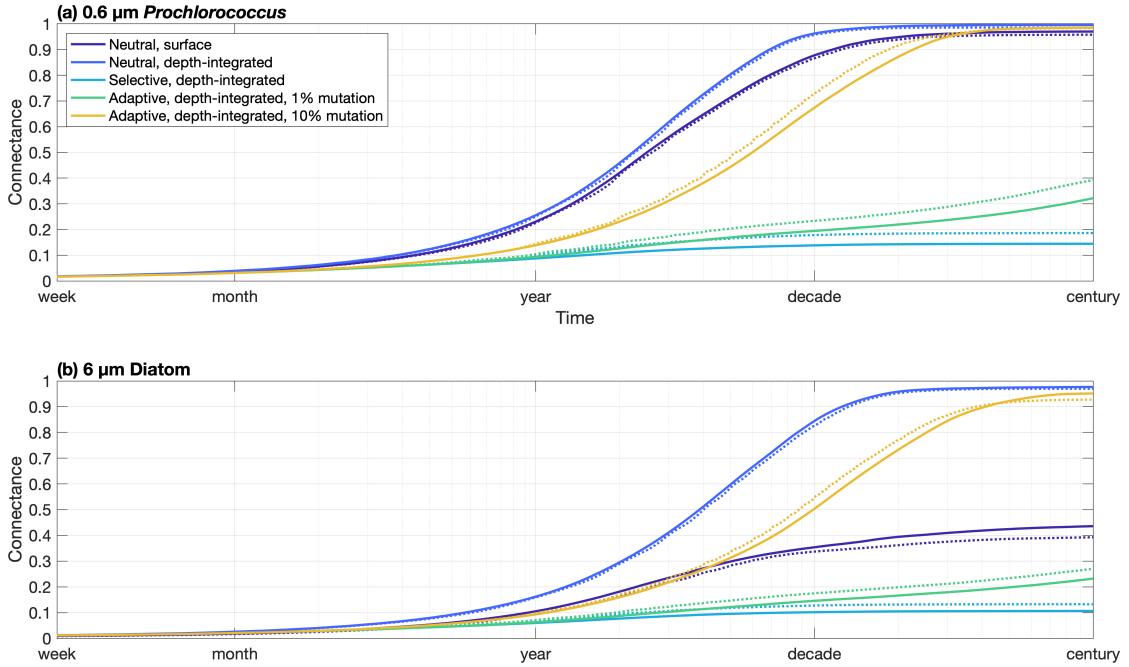


Figure 1: Fraction of connections between the 94 seed locations and the rest of the ocean through time. Solid lines show the results of simulations with time-invariant temperatures and carrying capacities. Dashed lines show the results of simulations with seasonality in these two variables.

Global immigration and emigration times are not spatially homogenous, as indicated in Figure ??a. Immigration times (background colour) suggest that temperate latitudes are generally more easily invaded than the equatorial regions. Conversely, emigration times (coloured dots) suggest that subpopulations initialised at lower latitudes are more rapidly dispersed throughout the ocean than those at from higher latitudes.

These patterns are explained by the surface circulation patterns shown in Figure ??a. The plankton transport vectors are highly divergent in equatorial upwelling regions, driving a consistent efflux of cells that must be topped up to the carrying capacity by reproduction of the local resident population. These regions thus export cells to the rest of the ocean, while remaining resistant to immigration. At higher latitudes, the sub-tropical gyres are characterised by convergent flow, with a consistent influx of cells diluting the local resident populations. These regions are thus easily invaded and are slower to export cells to the rest of the ocean. These patterns are confirmed in Figure ??, which shows immigration times against emigration times for the ‘surface-only’ simulation. Low latitudes (red) act

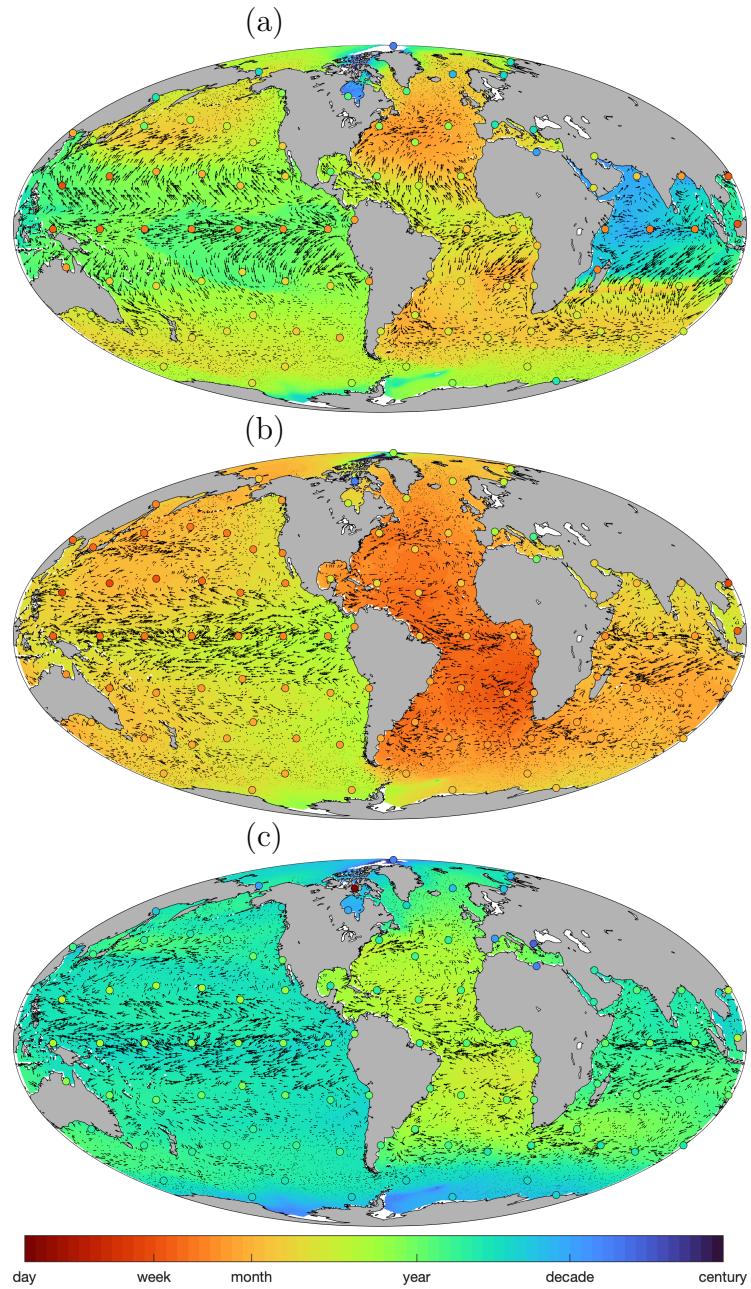


Figure 2: Immigration and emigration timescales (years) for *Prochlorococcus*, given (a) surface only transport, (b) depth-integrated transport and (c) depth-integrated transport with selection. Unique subpopulations were seeded in each of the 94 locations marked with dots. Emigration times, represented by the coloured dots, are defined as the time taken for each seed subpopulation to disperse to 95% of all locations. Immigration times, represented by the background colours, are defined as the time taken for 95% of all seed subpopulations to arrive in each location. Planktonic transport velocities are shown as vectors.

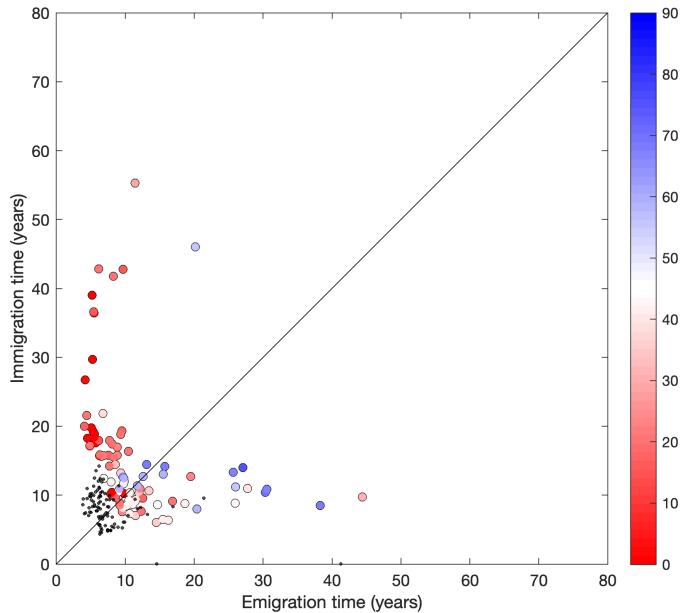


Figure 3: Immigration vs emigration times (years) at the 94 seed locations. The coloured circles show times from the surface-only case. The colour scale indicates absolute latitude, with low latitude regions clearly characterised by fast emigration and slow immigration, with the opposite true at higher latitudes. The black dots show the same time scales in the depth-integrated case.

as source regions, with slow immigration and fast emigration. High latitudes act as sink regions, with fast immigration and slow emigration.

The assumption that horizontal dispersal of plankton occurs only in the surface layer ignores the potential role of subsurface connectivity. To test the sensitivity of our results to this pathway, we calculated the depth-integrated horizontal transport of cells across the entire water column. This is effectively the depth-integrated horizontal water transport, weighted by the vertical distribution of cellular abundance. After this adjustment to the transport component, and with the carrying capacity increased to the depth-integrated cellular abundance, we repeated the initial experiment in the same way.

The orange line in Figure ?? shows that allowing for sub-surface transport markedly accelerates global plankton dispersal, with more than 97% of the ocean connected in less than a decade. Figure ??b shows that the global distribution of immigration and emigration times has decreased everywhere (see SI for the relative changes). The strongest effects are in the Indian ocean, where the surface-only simulation suggested that the region was highly resistant to immigration. The black dots in Figure ??d show that the differences between source and sink regions largely disappears when the depth-integrated transport is considered.

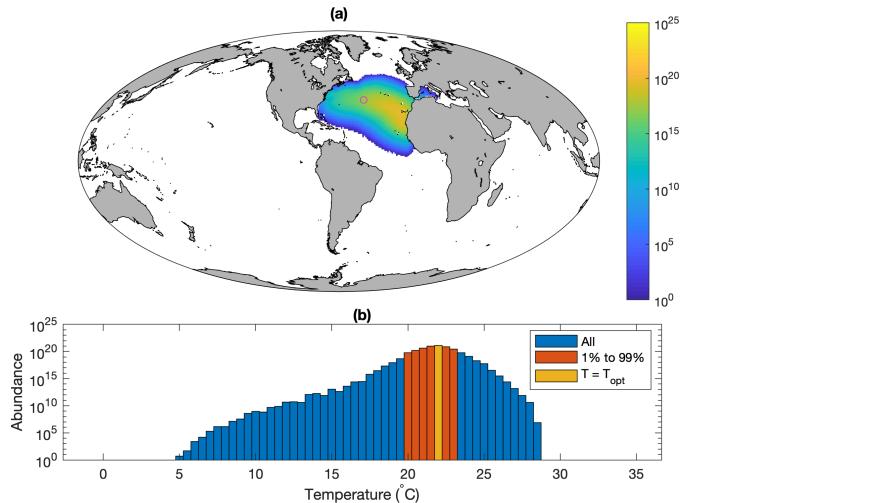


Figure 4: (a) Global distribution of a non-adapting population after 100 years of dispersal. The population was seeded at the circled location, with a thermal optima matching the ambient temperature of 22°C. (b) Abundance as a function of ambient temperature after 100 years of dispersal. The population is completely restricted to waters between 5 and 28°C, with 1st and 99th percentiles of 20 and 23°C.

The previous experiments have assumed that all subpopulations are equally well-adapted to conditions throughout the entire ocean, but we know that phytoplankton populations show distinct thermal optima that closely reflect their local environment (??). As a population is dispersed, it must compete with locally-adapted residents that may well competitively exclude the immigrant population. To test the influence of this effect, we assigned thermal tolerance curves, such that populations are preferentially selected when ambient conditions match their thermal optimum. The global resident population is divided into 77 phenotypes, each with thermal optima matching the annual-mean local water temperature. Each seed population is also assigned a thermal optimum defined by its local environment. The [as yet unplotted] line in Figure ?? shows the global dispersal of the 94 seed populations is severely restricted by temperature-based selection, with global connectivity not rising above XX% in the 100 year simulation.

Figure ?? shows the global distribution of an illustrative seed population added in the central North Atlantic (35°N, 46°W), after 100 years of dispersal. With a thermal optimum matching its initial habitat temperature of 22°C, the population remains restricted to the North Atlantic subtropical gyre, in waters between 5 and 28°C. The population has been unable to disperse beyond its original ocean basin, being outcompeted by resident cold-adapted species at higher latitudes.

Temperature-related selection clearly has the capacity to restrict the dispersal of thermally adapted populations. If populations are to overcome this restriction,

then they must adapt to their new environments (??). We included this capacity in the model by allowing populations to produce mutants with different thermal optima. We divided all of the model subpopulations into 77 distinct phenotypes across a range of thermal optima. We initialised the experiment with each subpopulation optimally adapted to its local temperature, but allowed for a small mutational flux between adjacent phenotypes (??). As populations are transported away from their thermal optima, they will still be outcompeted by the better adapted residents, but their mutated offspring will be able to compete as equals, allowing dispersal to continue (albeit at an attenuated rate). We performed simulations with mutation rates of 1% and 10%.

The green line in Figure ?? shows that even with an extremely high mutation rate of 10%, global dispersal is markedly restricted by selection effects, with 95% connectivity only achieved after approximately 30 years. When the mutation rate is set to 1%, just over 30% of the ocean has been connected within 100 years.

In a final set of experiments, we assessed performed the same set of simulations with a carrying capacity representative of small (6 micron) diatoms (?). Figure ??b shows that in all cases the lower abundances and associated higher probability of extinctions leads to a notable increase in the global timescales of dispersal.

2 Discussion

These results demonstrate that subsurface circulation plays a key role in the global connectivity of (?). Furthermore, the need to adapt across large-scale temperature gradients markedly slows the rate of dispersal. Nonetheless, global connectivity seems to be assured within a matter of decades.

While these timescales are broadly consistent with the estimates of ?, (?) is a cosmopolitan and (extremely) abundant taxon. It seems likely that larger and less abundant taxa with less globally ubiquitous distributions would face much stronger barriers to dispersal. To test this, we repeated the experiments with the carrying capacity changed to the estimated global distribution of diatoms in the 6 micron size class (?).

Figure ??b shows that the equivalent global connectivity of these diatoms for the depth-integrated cases only. It is clear that the decreased abundance and restricted biogeography of this group markedly slows down their global dispersal, in both the neutral and selective cases.

3 Methods

We used the population model to estimate the global dispersal of 341 genotypes, each initialised at unique “seed locations” that were distributed approximately evenly around the surface ocean. We also included one additional tracer representing a globally resident species, with a genotype frequency of $\mathbf{p} = 0$ at all seed locations, and 1 throughout the rest of the surface grid.

Under the assumption that all species have equal fitness, the number of individuals \mathbf{X} surviving at each generation is drawn randomly from the local population (after oceanic transport and mutation) with probability equal to the local subpopulation frequency ($\mathbf{p} = \mathbf{X}\mathbf{n}^{-1}$). Under these assumptions, the expected population size in each generation is given by the multinomial distribution,

$$\mathbf{X} \sim \mathcal{M}(\mathbf{n}, \mathbf{p}) \quad (1)$$

For large \mathbf{N} , equation ?? is reasonably approximated by a normal distribution.

$$\mathbf{X} \approx \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}) \quad (2)$$

with $\boldsymbol{\mu} = \mathbf{n} \circ \mathbf{p}$ and $\boldsymbol{\sigma} = \mathbf{n} \circ \mathbf{p} \circ (1 - \mathbf{p})$. (Here the ‘ \circ ’ symbol denotes element-wise multiplication.)

The Evolutionary Plankton Metacommunity Dynamics (EPMD) model considers the global distribution of an arbitrary number of planktonic subpopulations distributed across a two-dimensional (latitude and longitude) ocean grid. The probability of survival for each subpopulation in each generation is a function of its relative abundance and its thermal tolerance to its current environmental temperature. Each subpopulation can be assigned a particular thermal optimum, or its survival can be made independent of growth. Plankton cells are circulated in physical space according to a realistic ocean circulation model ??.

We estimated minimum connectivity times by integrating the model forward from these initial conditions, noting the time at which each genotype first appeared in each grid cell.

The Evolutionary Plankton Metacommunity Dynamics (EPMD) model considers the global distribution of an arbitrary number of planktonic subpopulations distributed across a two-dimensional (latitude and longitude) ocean grid. The probability of survival for each subpopulation in each generation is a function of its relative abundance and its thermal tolerance to its current environmental temperature. Each subpopulation can be assigned a particular thermal optimum, or its survival can be made independent of growth. Plankton cells are circulated in physical space according to a realistic ocean circulation model ??.

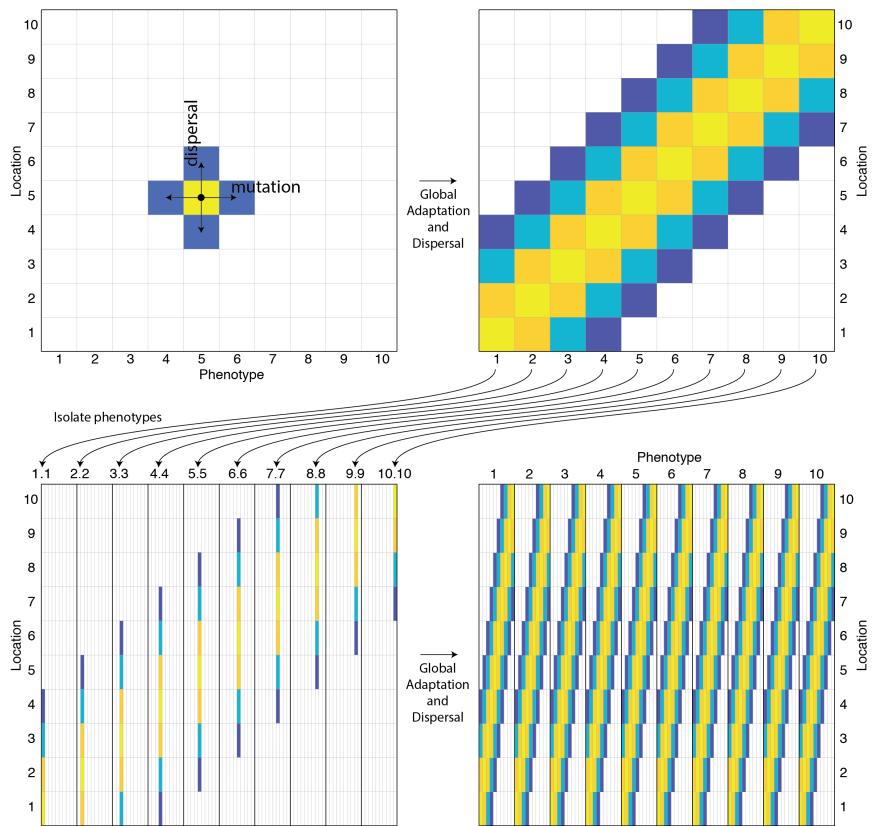


Figure 5

4 Model description

across J spatial boxes of an ocean general circulation model (GCM). The global abundances of all subpopulations are thus represented in the $[J \times K]$ population matrix.

Physical dispersal

Plankton cells are transported between grid boxes using a $[J \times J]$ oceanic ‘transport matrix’ \mathbf{A} that describes the transport of neutrally buoyant cells between points in the ocean grid (?). This transport can be written as

$$\mathbf{T} = \mathbf{Ax}_t \quad (3)$$

Here \mathbf{x} is the $[J \times K]$ matrix of genotype frequencies in each grid box of the GCM. Each element of the transport matrix \mathbf{A} describes the transport of cells between source boxes (columns) and recipient boxes (rows). The transport matrix was derived from the “Estimating the Circulation and Climate of the Ocean” (ECCO) version 4 ocean model (). It represents physical transport attributable to advection, diffusion and parameterised sub-grid-scale processes in the ocean model.

Mutation

Each plankton subpopulation may be assigned a particular phenotype, defined by its thermal optima.

Selection

Selection can be further incorporated through the selection vector \mathbf{s} , that assigns each population in \mathbf{X} a relative fitness of $1 + \mathbf{s}$.

$$\mathbf{p} = \frac{\tilde{\mathbf{p}} \circ (1 + \mathbf{s})}{\tilde{\mathbf{p}} \circ (1 + \mathbf{s}) + 1 - \tilde{\mathbf{p}}} \quad (4)$$

$$s = \exp \left[- \left(\frac{T_{env} - T_{opt}}{w} \right)^2 \right] \quad (5)$$

Thermal niche. Seed with n plankton types, each adjusted to a different optimal temperature. How connected are these phenotypes? How does the connectivity depend on population size? on the shape of the thermal niche? Do results compare well with Thomas 2012 or Righetti 2019?

4.1 Connectivity

5 Results

5.1 Timescales of dispersal

Seed subset of ocean regions with one tracer each. Examine time until globally distributed. Dispersal potential increases with local population size.

Figure 6: Left-hand side: global abundance distribution of prochlorococcus (top) and diatoms (bottom). Right-hand side: timescales of ocean connectivity for the same two groups. Background colour: shortest timescales over which particles from 95% of the seed locations (indicated by dots) are expected to have arrived at each point (i.e. immigration timescales). Dots: timescales over which particles from that seed location are expected to have arrived at 95% of the ocean surface (i.e. emigration timescales).

A Appendix

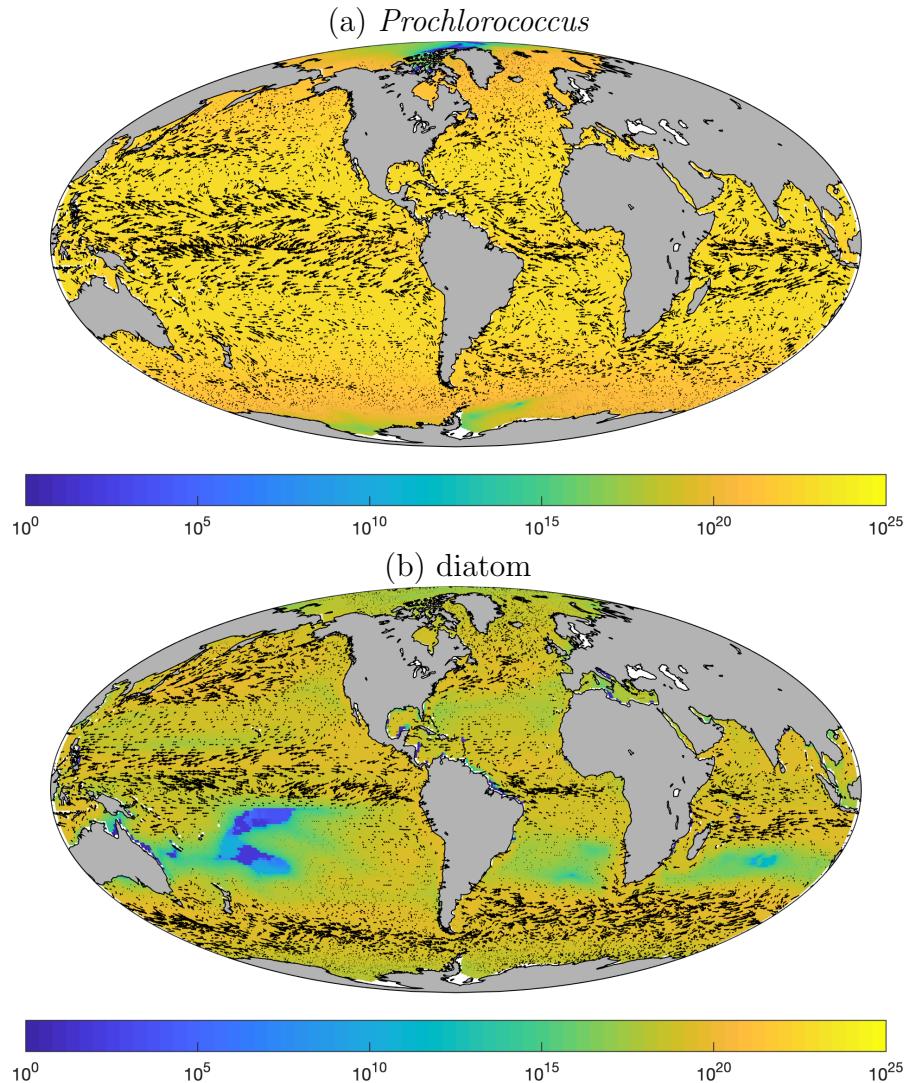


Figure A.1: Global abundance of a 0.6 micron *Prochlorococcus* population (upper panel) and a 6 micron diatom. Distributions are taken from (?).

A.1 Mass conservation correction

Numerical constraints within the GCM mean that some off-diagonal elements may be negative. To remove these artefacts, negative fluxes out of a box are converted to positive fluxes into the box. If we first define the matrix of negative off-diagonal elements as follows,

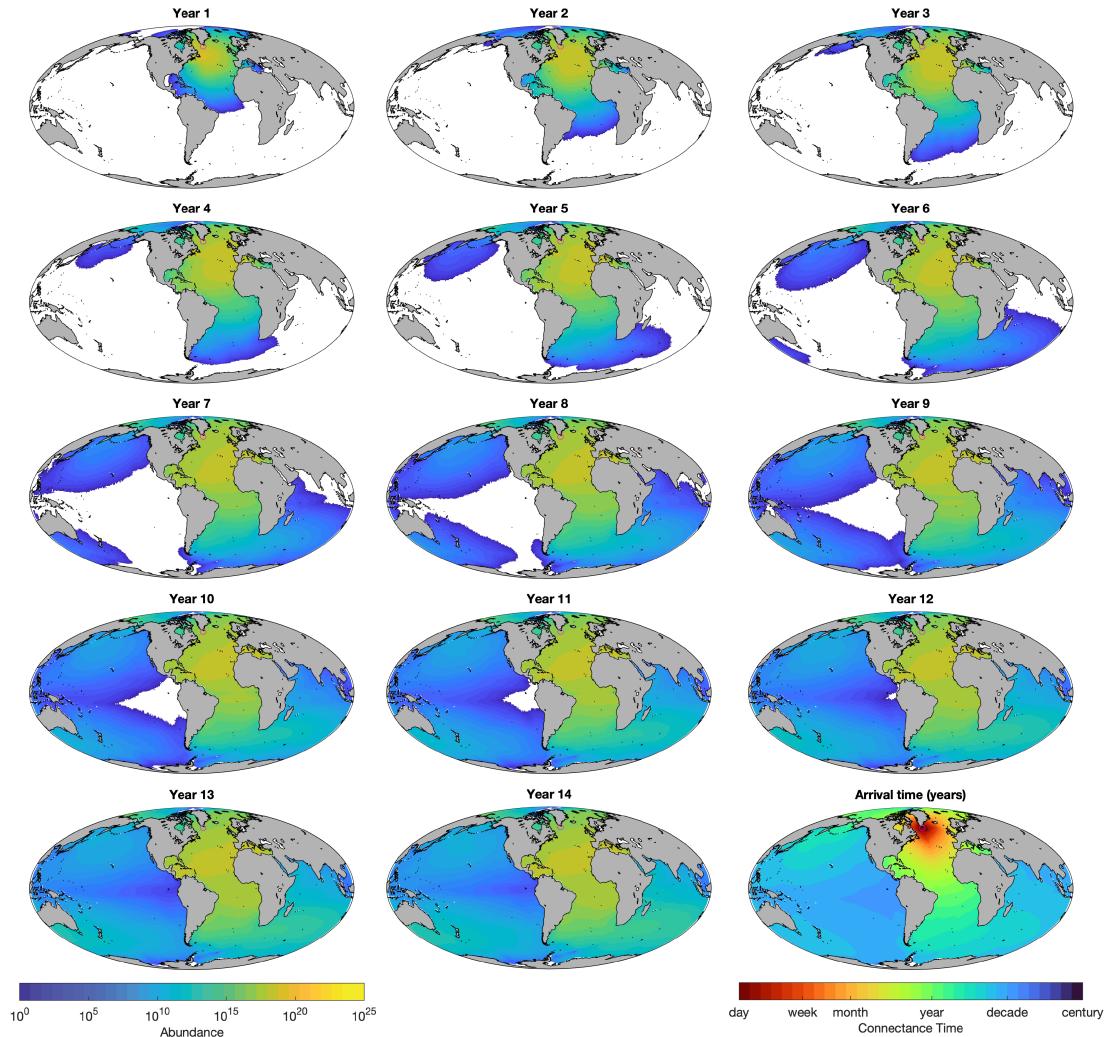


Figure A.2: Neutral dispersal of a *Prochlorococcus* subpopulation seeded off the southern tip of Greenland (magenta circle). Panels 1 to 14 show the global abundance distribution at the end of each year. The final panel shows the arrival times of the subpopulation as it spreads globally.

$$N_{i,j} = \begin{cases} 0 & \text{if } i = j \text{ or } A_{i,j} \geq 0 \\ A_{i,j} & \text{if } i \neq j \text{ and } A_{i,j} < 0 \end{cases} \quad (\text{A.1})$$

We first move the off-diagonal negatives to their transpose, changing the sign.

$$\mathbf{B} = \mathbf{A} - \mathbf{N} + (-\mathbf{N}^\top) \quad (\text{A.2})$$

To conserve mass, changes in the off-diagonals must be compensated by an equivalent change on the diagonals. This is achieved by adding both the row and column sums of the negative off-diagonals to the diagonal.

$$C_{ii} = B_{ii} + \sum_{j=1}^J N_{i,j} + \sum_{j=1}^J N_{j,i} \text{ (for all } i\text{)} \quad (\text{A.3})$$

The resultant matrix is positive-definite on the off-diagonals, and conserves mass.

Finally, the transport matrix is converted from a volume flux per time ($\text{m}^3 \text{ d}^{-1}$) to a unitless operator; first dividing through by the volume of the recipient cells (\mathbf{v}), then multiplying by the selected time step (Δt), and adding the identity matrix (\mathbf{I}).

$$\mathbf{D} = \frac{\mathbf{C}}{\mathbf{v}} \times \Delta t + \mathbf{I} \quad (\text{A.4})$$

A.2 Seed locations

Seed locations were identified by iteratively projecting a subdivided rectangular grid onto the surface of a sphere. Surface ocean coordinates from the GCM grid were then mapped onto these points (by shortest euclidean distance in cartesian coordinates). Finally, the 341 unique coordinates were mapped back to their nearest point on the GCM grid. The model was initialised with one tracer for each of these points, with a genotype frequency of $\mathbf{p} = 1$ at the seed location, and zero everywhere else.