

Selective 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. They 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 Methods

1.1 Plankton as passive tracers

We consider the global abundance of a plankton population with K genotypes within J spatial boxes of an ocean general circulation model (GCM). Local cell abundances within each genotype are thus represented in the $[J \times K]$ population matrix.

Plankton cells are transported between grid boxes using a $[J \times J]$ oceanic ‘transport matrix’ \mathbf{A} that describes the volume transport attributable to advection, diffusion and parameterised sub-grid-scale processes in the GCM [1]. This transport can be written as a linear matrix equation [1], with the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{AX} \quad (1)$$

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 mass flux between source boxes (columns) and recipient boxes (rows). The diagonal elements of the transport matrix should always be nonpositive ($<= 0$), as they describe the mass flux out of the source box to other (nearby) boxes. Similarly, the off-diagonal elements should always be nonnegative ($>= 0$), as they describe the mass flux into the recipient box from other nearby boxes.

This gives the following explicit form of the linear matrix equation

$$\mathbf{X}_{t+1} = \mathbf{DX}_t \quad (2)$$

1.2 Stochastic dispersal

Each grid box supports a predefined carrying capacity of \mathbf{n} individuals. Under strict neutrality, the number of individuals \mathbf{X} surviving at each generation is drawn randomly from the local population (after oceanic transport) with probability equal to the local genotype frequency ($\mathbf{p} = \mathbf{X}\mathbf{n}^{-1}$). Under these assumptions, the expected population size in each generation is given by the multinomial distribution,

$$\mathbf{X} \sim \mathcal{M}(\mathbf{n}, \mathbf{p}) \quad (3)$$

For large \mathbf{N} , equation 3 is reasonably approximated by a normal distribution.

$$\mathbf{X} \approx \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}) \quad (4)$$

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.)

1.3 Connectivity

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.

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.

1.4 Selection

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

$$\mathbf{p} = \frac{\tilde{\mathbf{p}} \circ (1 + \mathbf{s})}{\tilde{\mathbf{p}} \circ (1 + \mathbf{s}) + 1 - \tilde{\mathbf{p}}} \quad (5)$$

[NOT YET GENERALISED FOR MULTINOMIAL]

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?

$$\gamma_T = ae^{bT} e^{\left(-\left(\frac{2(T-z)}{w}\right)^2\right)} \quad (6)$$

2 Results

2.1 Timescales of dispersal

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

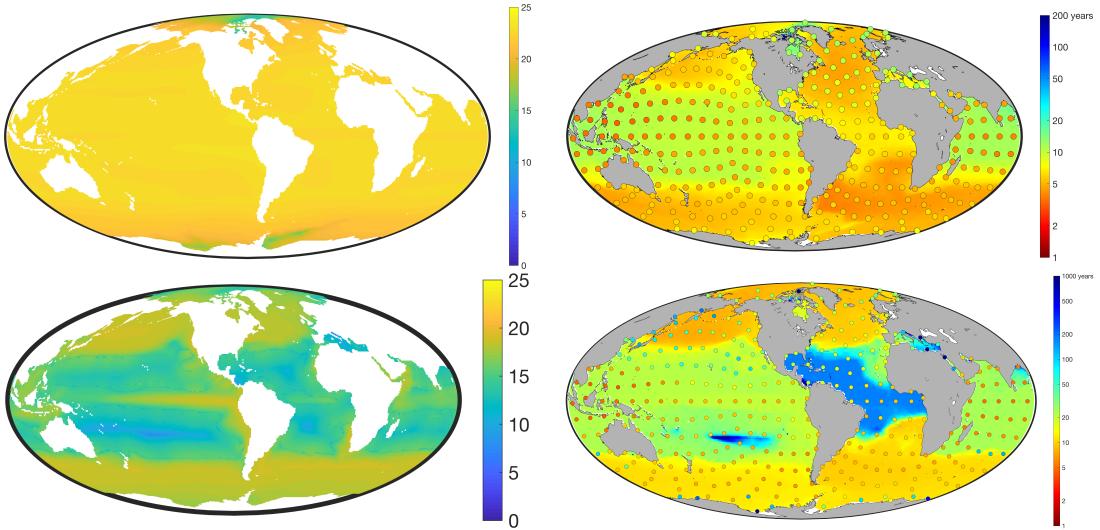


Figure 1: Left-hand side: global abundance distribution of prochlorococcus (top) and diatoms (bottom). Right-hand side: timescales of ocean connectivity for the same two groups. Background colour: shortest timescales over which particles from 95% of the seed locations (indicated by dots) are expected 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).

2.2 Spatial clustering by timescales of connectivity

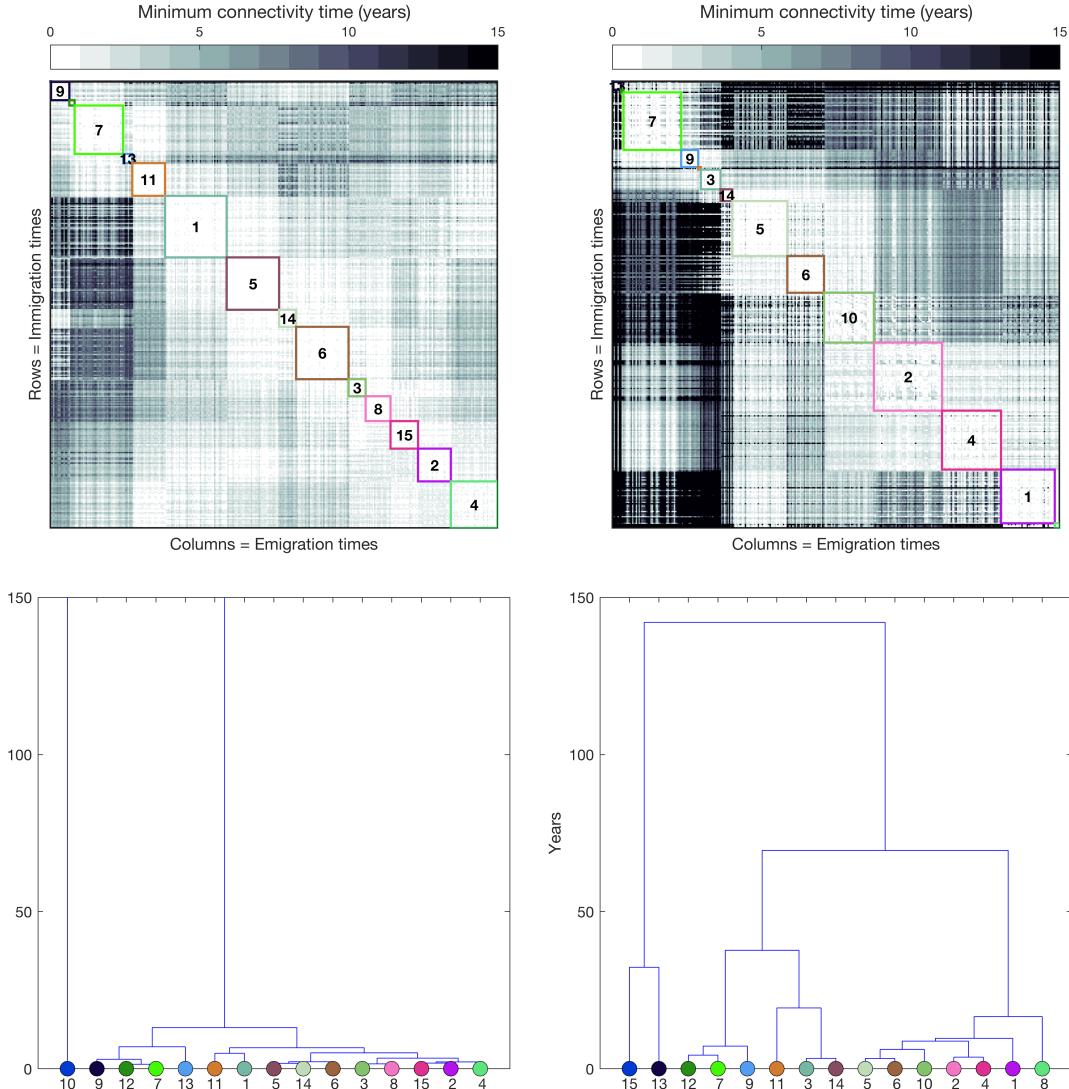


Figure 2: Pairwise global minimum connectivity and clustering. Upper panel: matrix of the connectivity times between seed locations. Columns of the matrix correspond to source locations (such that row values represent emigration timescales), while rows correspond to target locations (such that column values represent immigration timescales). Lower panel: Clustering of the seed locations on the basis of the average of the immigration and emigration timescales between each location. Coloured boxes in the upper panel indicate clusters defined by the average timescales of connectivity (as shown in the lower panel).

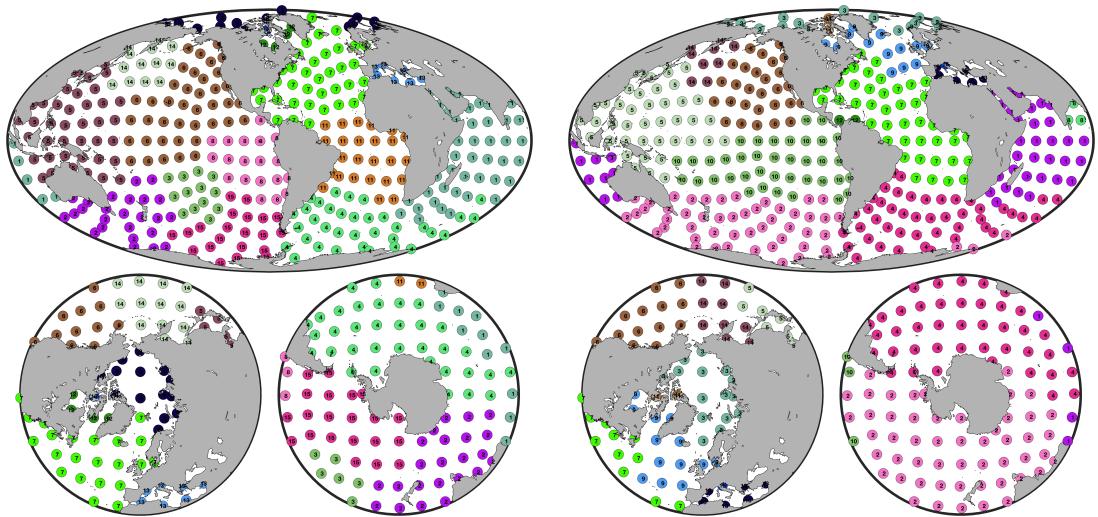


Figure 3: Spatial distribution of the 15 connectivity clusters. Colours and numbers correspond to those used in Figure 2.

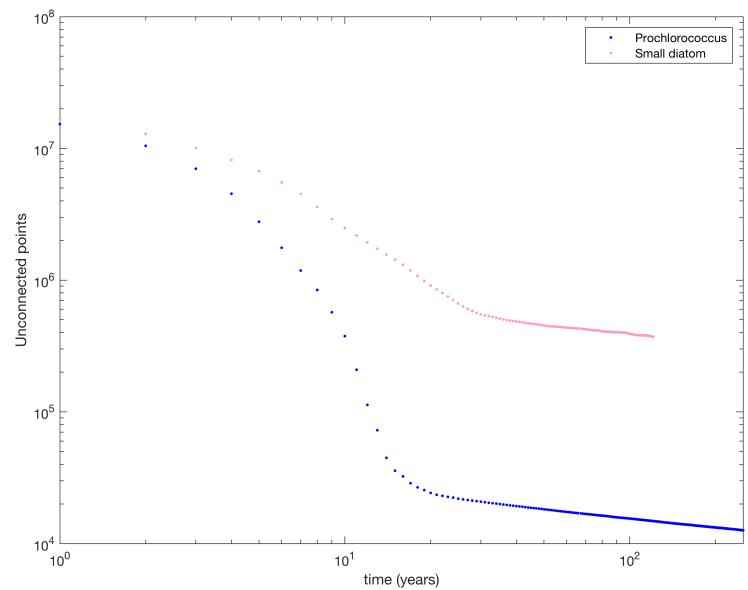


Figure 4: Number of globally unconnected points through time

A Appendix

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,

$$N_{i,j} = \begin{cases} 0 & \text{if } i = j \text{ or } A_{i,j} \geq 0 \\ A_{i,j} & \text{if } i \neq j \text{ and } A_{i,j} < 0 \end{cases} \quad (7)$$

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

$$\mathbf{B} = \mathbf{A} - \mathbf{N} + (-\mathbf{N}^\top) \quad (8)$$

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.

$$C_{ii} = B_{ii} + \sum_{j=1}^J N_{i,j} + \sum_{j=1}^J N_{j,i} \quad (\text{for all } i) \quad (9)$$

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}).

$$\mathbf{D} = \frac{\mathbf{C}}{\mathbf{v}} \times \Delta t + \mathbf{I} \quad (10)$$

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.