

# 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”* (Baas-Becking, 1934). (Also Beijerinck: one particular species of bacteria found anywhere on Earth, provided environmental requirements are met... Fenchel and Finlay, 2004) ). 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.

Fenchel and Finlay (2004): 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.

Rossberg et al. (2013): “Are there any species smaller than 1 mm?” [Showed empirical trends and used population model. Need to read again!]

Lagrangian studies have been equivocal. Hellweger et al. (2014) 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?? Kingman, 1982), 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].

Jönsson and Watson (2016) used a (biologically inert) particle tracking model in combination with Dijkstra’s ‘shortest-path algorithm’ Dijkstra (1959) 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. Jönsson and Watson (2016) 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  $N$ . 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, this population is replaced with a unique (but ecologically neutral) local subpopulation at 94 ‘seed locations’ distributed more or less evenly around the ocean.

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. A new metapopulation is drawn at random from a multinomial distribution, with probabilities based on the subpopulation frequencies after dispersal. 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 at the boundaries of each subpopulation, or where the overall carrying capacity is low.

The green line in Figure 1 shows the timescales over which the 94 seed subpopulations become connected with the rest of the ocean. Largely in agreement with Jönsson and Watson (2016), we find that almost 90% of the surface ocean is connected within a decade. However, we find a much lower value for the median connection times, with a value of just 2.4 years. **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)???

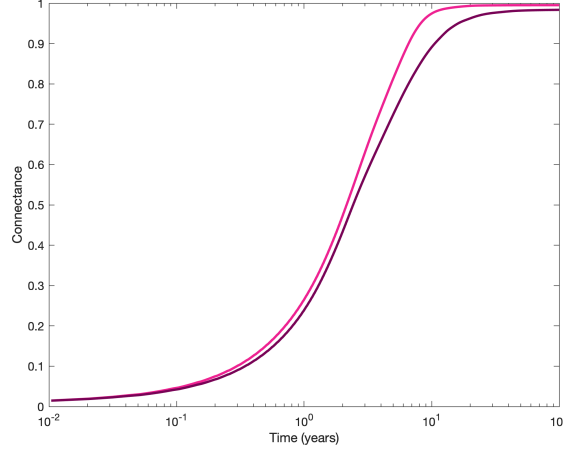


Figure 1: Fraction of connected points between the 94 seed locations and the rest of the ocean through time

Global immigration and emigration times are not spatially homogenous, as indicated in Figure 3a. Immigration times (shown as the background colour) suggest that regions between 20 and 60° N and S are the most easily invaded, whereas the equatorial and polar regions appear to be much more resistant to immigration. Emigration times (shown by the dots) appear to increase either side of a minimum at the equator.

These patterns are consistent with the surface circulation patterns used to drive the simulation. The equatorial regions are characterised by divergent meridional currents and upwelling. With a consistent efflux of cells being topped up to the carrying capacity, these regions can easily export cells to the rest of the ocean, while remaining resistant to immigration.

On the other hand, the sub-tropical gyres are characterised by convergent flow and downwelling. Immigrant cells are continually entrained from bordering regions, with a significant loss to resident cells imposed by the constant carrying capacity. These regions are thus easily invaded and are slower to export cells to the rest of the ocean.

The red dots in Figure 3d show that there are ‘source’ regions with slow immigration and fast emigration, and conversely there are ‘sink’ regions with fast immigration and slow emigration.

These interpretations of the immigration and emigration times shown in Figure 3a are based on the assumption that transport below the surface layer can be neglected. To test the sensitivity of the model to this assumption, we calculated the depth-integrated horizontal transport of cells across all model layers. This is

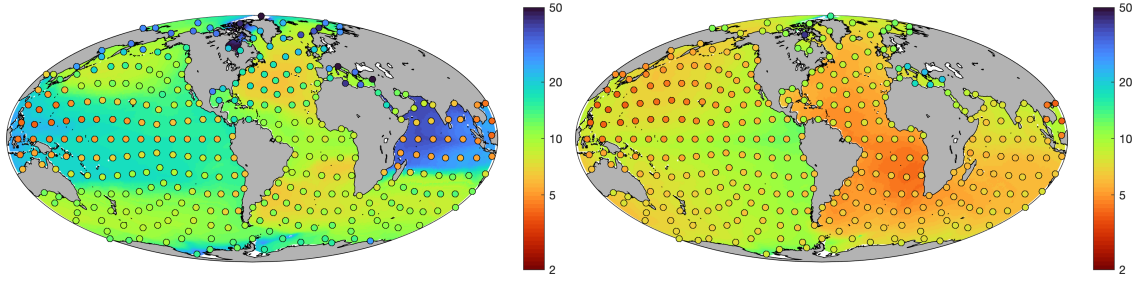


Figure 2

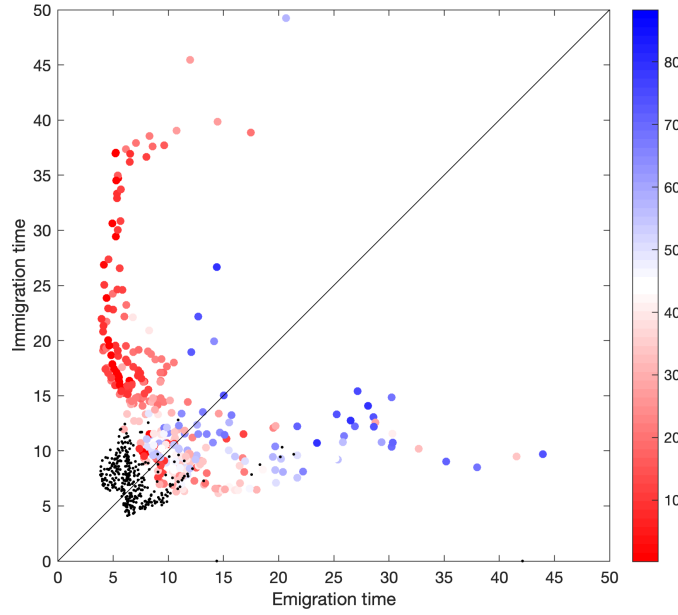


Figure 3

effectively the depth-integrated horizontal water transport, weighted by the vertical distribution of cellular abundance. The same transport could be achieved by using the full depth-resolved transport, while vertically homogenising each water column at every time step. 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 exactly the same way.

The purple line in Figure 1 shows that allowing for sub-surface transport markedly decreases the timescales of global plankton dispersal, with a median connection time of 2.1 years, such that more than 97% of the ocean is connected within a decade. Figure 1 shows the change in immigration and emigration times that occurs with the switch from surface to depth-integrated transport. Figure 3b and c shows that the global distribution of immigration and emigration times has

decreased everywhere, with the strongest effects in the Indian ocean, where the surface-only simulation suggested that the region was highly resistant to immigration. The blue dots in Figure 3d show that the difference between source and sink regions largely disappears when the depth-integrated transport is considered.

We next investigated the effect of carrying capacity on the timescales of global dispersal.

The previous experiments have assumed that all subpopulations are equally well-adapted to conditions throughout the entire ocean. We know however that cells show distinct thermal optima that closely reflect their local environment (?). Such observations suggest that plankton are able to rapidly evolve their thermal tolerance as they are transported across temperature gradients. To account for this, we divided each 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 (?).

## 2 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 1 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 ??.

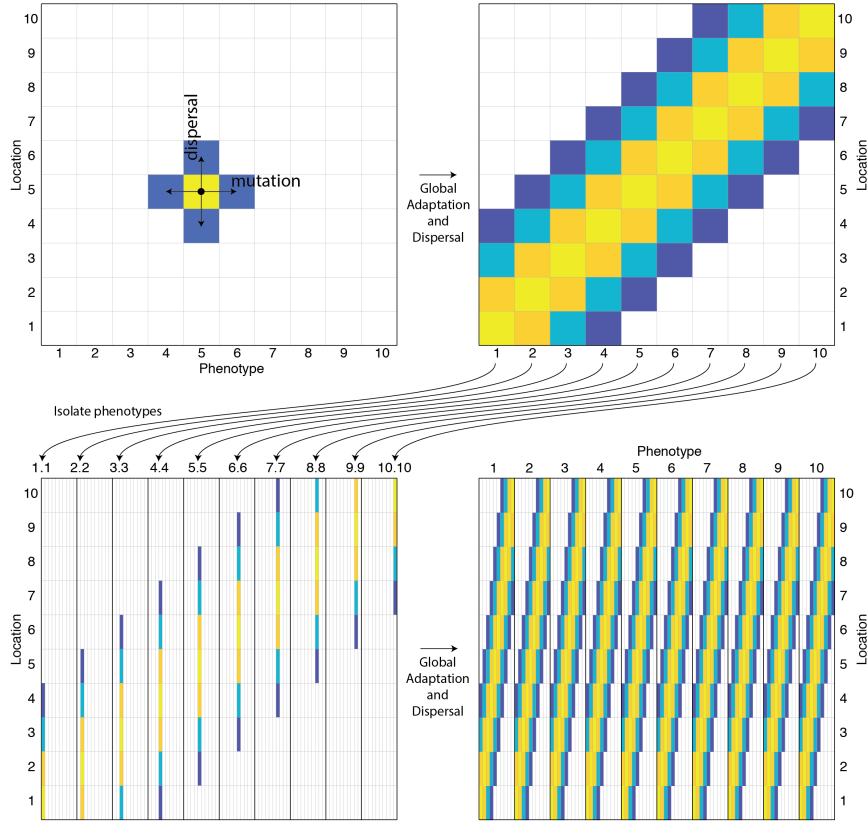


Figure 4

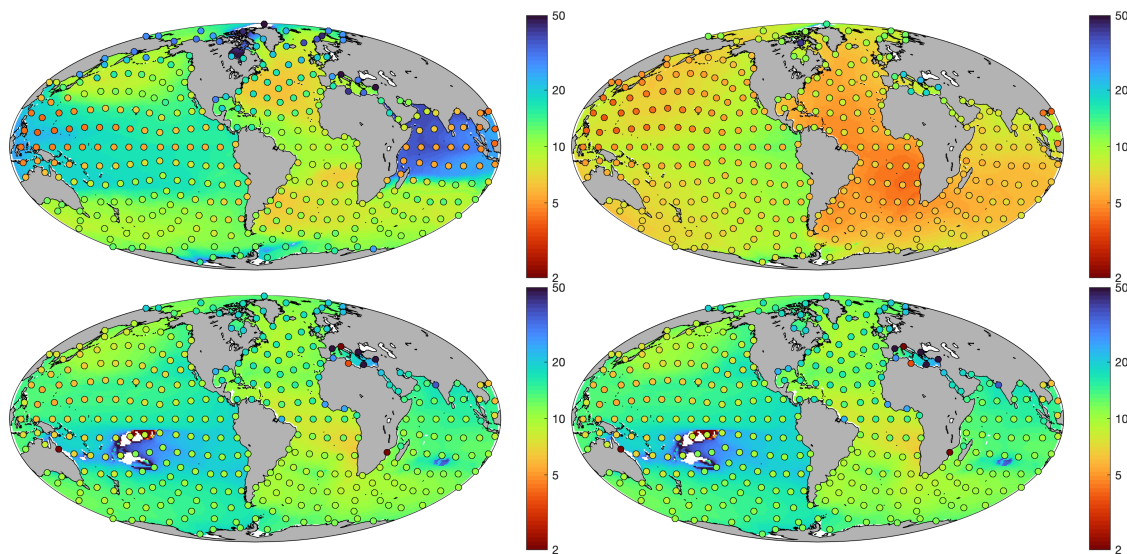


Figure 5

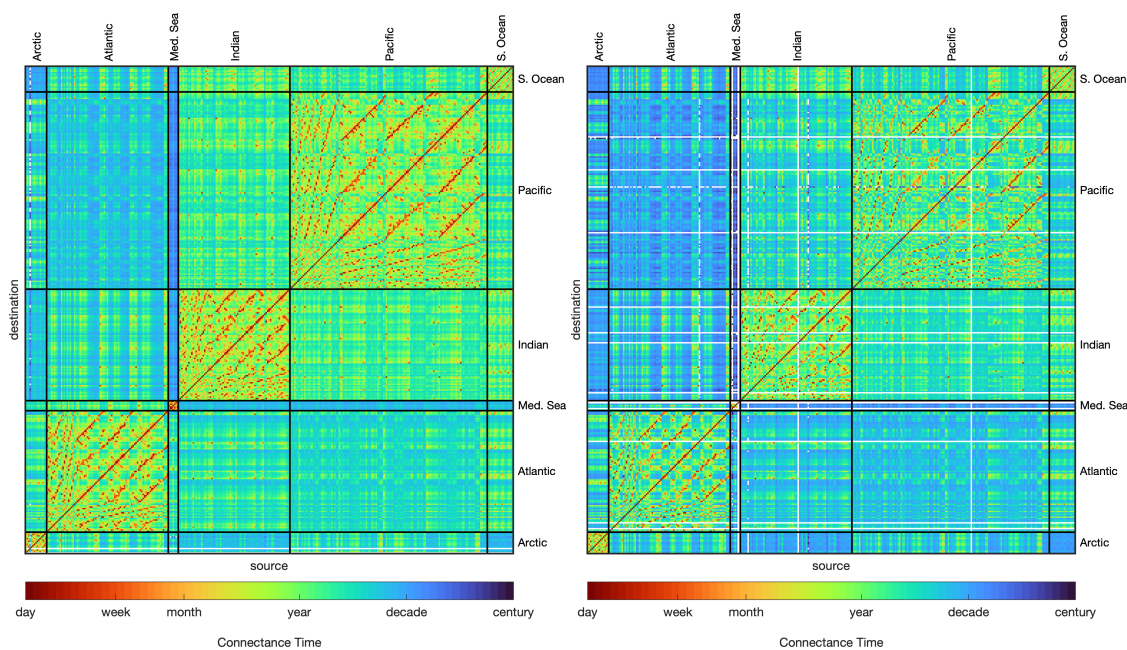


Figure 6

### 3 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 (Khatiwala et al., 2005). This transport can be written as

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

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?



### 3.1 Connectivity

## 4 Results

### 4.1 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 7: 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).

### 4.2 Spatial clustering by timescales of connectivity

Figure 8: 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).

# 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,

$$N_{i,j} = \begin{cases} 0 & \text{if } i = j \text{ or } A_{i,j} \geq 0 \\ A_{i,j} & \text{if } i \neq j \text{ and } A_{i,j} < 0 \end{cases} \quad (6)$$

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

$$\mathbf{B} = \mathbf{A} - \mathbf{N} + (-\mathbf{N}^\top) \quad (7)$$

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.

$$C_{ii} = B_{ii} + \sum_{j=1}^J N_{i,j} + \sum_{j=1}^J N_{j,i} \text{ (for all } i) \quad (8)$$

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 ( $\Delta t$ ), and adding the identity matrix ( $\mathbf{I}$ ).

$$\mathbf{D} = \frac{\mathbf{C}}{\mathbf{v}} \times \Delta t + \mathbf{I} \quad (9)$$

## 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.

## References

- Baas-Becking, L. G. M., 1934. Geobiologie of Inleiding Tot de Milieukunde. Van Stockum & Zoon, The Hague.
- Dijkstra, E. W., 1959. A note on two problems in connexion with graph. *Numerische Mathematik* 1, 269–271.
- Fenchel, T., Finlay, B. J., 2004. The ubiquity of small species: Patterns of local and global diversity. *BioScience* 54 (8), 777–784.
- Hellweger, F. L., van Sebille, E., Fredrick, N. D., 2014. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* 345 (6202), 1346–1349.
- Jönsson, B. F., Watson, J. R., 2016. The timescales of global surface-ocean connectivity. *Nature Communications* 7 (11239), DOI: 10.1038/ncomms11239.
- Khatiwala, S., Visbeck, M., Cane, M. A., 2005. Accelerated simulation of passive tracers in ocean circulation models. *Ocean Modelling* 9, 51–69.
- Kingman, J. F. C., 1982. The coalescent. *Stochastic properties and their applications* 13, 235–248.
- Rossberg, A. G., Rogers, T., McKane, A. J., 2013. Are there species smaller than 1 mm? *Proceedings of the Royal Society of London B* 280 (20131248), <http://dx.doi.org/10.1098/rspb.2013.1248>.