ECOLOGY

Ocean currents promote rare species diversity in protists

Paula Villa Martín[1], Aleš Buček[1], Thomas Bourguignon[1,2], Simone Pigolotti[1]*

Oceans host communities of plankton composed of relatively few abundant species and many rare species. The number of rare protist species in these communities, as estimated in metagenomic studies, decays as a steep power law of their abundance. The ecological factors at the origin of this pattern remain elusive. We propose that chaotic advection by oceanic currents affects biodiversity patterns of rare species. To test this hypothesis, we introduce a spatially explicit coalescence model that reconstructs the species diversity of a sample of water. Our model predicts, in the presence of chaotic advection, a steeper power law decay of the species abundance distribution and a steeper increase of the number of observed species with sample size. A comparison of metagenomic studies of planktonic protist communities in oceans and in lakes quantitatively confirms our prediction. Our results support that oceanic currents positively affect the diversity of rare aquatic microbes.

INTRODUCTION

Oceanic plankton can be transported across very large distances by currents. Many planktonic species are cosmopolitan, i.e., they are found virtually everywhere across the global ocean (*1*). These observations suggest that, at first sight, the distribution of planktonic species is not limited by dispersal and, therefore, that niche preference is the predominant factor determining species abundance (*2*). However, in the presence of a limited set of resources, niche theory predicts species-poor communities. In contrast, planktonic communities in the oceans are very diverse (*3*–*6*). This contradiction of the basic principles of niche theory (*7*) has puzzled ecologists for decades (*8*) and has fostered numerous attempts to explain the diversity of plankton (*9*). One proposal is that variable environments offer more possibilities for specialization of ecological traits (*4*, *10*–*14*). Another proposal is that chaotic advection by oceanic currents creates barriers reducing competition among species, therefore promoting species coexistence (*15*, *16*). Quantitative analyses also suggest that oceanic currents play an important role in organizing large-scale diversity patterns (*17*, *18*) and that dispersal limitation contributes, alongside with niche specialization, to the microbial biodiversity of oceans (*19*–*23*). The influence of oceanic currents on biodiversity patterns of planktonic communities can be tested by a comparison of oceans and lakes, in which currents are reduced.

DNA metabarcoding has allowed rapid and extensive measurements of the diversity of aquatic microbial communities, providing new means to study the ecological forces shaping planktonic communities. Metabarcoding studies have revealed that, besides commonly observed species, planktonic communities are characterized by a vast range of rare species. This so-called rare biosphere (*24*, *25*) makes up the majority of planktonic species (*21*, *26*) and is the subject of our study. The diversity of planktonic species can be quantified by the species abundance distribution (SAD), defined as the frequency $P(n)$ of species with abundance n in a sample. SADs of rare marine protists are qualitatively different from those of abundant species (*27*, *28*) and appear to follow a power law distribution

$$P(n) \propto 1/n^{\alpha} \quad (1)$$

The exponent α varies significantly among samples, is weakly correlated with environmental factors, and is significantly larger than 1 on average (*29*). Diversity patterns in other microbial communities, such as that of the human gut (*30*), are well described by a form of SAD following the Fisher log series, $P(n) \propto e^{-cn}/n$ (*31*), as predicted by Hubbell's neutral model (*31*). For large communities, the parameter c is very small, so that the distribution is close to a power law with $\alpha = 1$. Hubbell's neutral model is therefore unable to explain the steep decay of SADs in the rare oceanic biosphere. This steep decay can be obtained with a modified neutral model that takes into account density dependence of growth and death rates (*29*, *32*–*34*). However, the ecological forces determining this density dependence in the oceans are unknown.

Here, we propose that the steep decay of SADs observed in the oceans is caused by the particular way chaotic advection by oceanic currents limits dispersal. Oceanic currents affect distributions of planktonic populations by stirring and mixing. At the submesoscale, oceanic currents also affect ecological interactions, light exposure, and nutrient upwelling (*35*). Previous theoretical studies have shown that currents can affect effective population size (*36*) and provoke counterintuitive effects on fixation times (*37*), particularly in the presence of divergent flows (*38*, *39*). However, these studies did not scrutinize the effects of advection on multispecies communities. To test our hypothesis, we introduce a model that takes into account the role of transport by oceanic currents in determining the genealogy of microbes in a sample. Our model predicts that, in the presence of chaotic advection, SADs are characterized by larger values of the exponent α. It also predicts that chaotic advection causes a sharper increase of species diversity as a function of sample size. To validate these results, we analyze 18*S* ribosomal RNA (rRNA) sequencing data generated from oceanic water samples (*29*). We compare these results with sequencing data from lake protist communities (*40*). The observations quantitatively match our predictions, supporting the idea that chaotic advection by oceanic currents is responsible for the differences in biodiversity patterns between oceans and lakes.

[1]Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan. [2]Faculty of Tropical AgriSciences, Czech University of Life Sciences, Kamýcká 129, CZ-165 00 Prague, Czech Republic.
*Corresponding author. Email: simone.pigolotti@oist.jp

RESULTS

Coalescence model predicts the effect of chaotic advection by oceanic currents on SADs

We introduce a computational model to assess the effect of chaotic advection on the protist species distribution of a water sample. In this model, we assign a Lagrangian tracer (hereafter "tracer") to each individual in the sample (see Fig. 1). Tracers are initially placed in a local area, representing the portion of water where the sample was collected. The spatial coordinates x and y of each tracer move backward in time, following the spatial trajectory of the ancestors of each individual (see Fig. 1). If two tracers are at a sufficiently close distance, then they coalesce into a single tracer with a given probability. This new tracer represents the common ancestor of the two individuals. Last, tracers are assigned at a fixed rate μ to one species. These events represent immigration due to other causes than ocean currents. Assigned tracers are eliminated from the system. At the end of a run, individuals in the original sample are considered conspecific if their corresponding tracers have coalesced to a common ancestor before being eliminated (see Materials and Methods, Fig. 1, and movie S1). This coalescence model can be interpreted as the backward version of an individual-based community model, which includes advection by currents (see fig. S1) (*38*, *41*). The coalescent formulation has the advantage of describing the dynamics of one sample embedded in a larger ecosystem (*42*, *43*).

We simulate the coalescence model with and without oceanic currents. In the latter case, movements of tracers are modeled as a simple diffusion process, taking into account individual movements and small-scale turbulence. In the former case, we superimpose to this diffusion process the effect of large-scale oceanic currents. We model transport by these currents with a kinematic model of a meandering jet, which is a common large-scale structure characterizing oceanic flows (*44*, *45*). Population sizes and parameters characterizing the flow are sampled in a physically realistic range (see Materials and Methods) (*45*). All other parameters characterizing population dynamics are chosen identically in the two cases (see Materials and Methods).

SADs predicted by the model present a considerable variability depending on parameters and demographic stochasticity, both in the presence and absence of currents (see Fig. 2, A and B). To characterize individual SAD curves, we fit them with a power law function $P(n) \propto 1/n^{\alpha}$ using maximum likelihood in an optimal range of abundances (see Materials and Methods). For comparison, we also fit an exponential distribution $P(n) \propto e^{-c\,n}$ and a Fisher log series $P(n) \propto e^{-c\,n}/n$ in the same range. In most cases, the power law provides a better fit than the exponential distribution (74 and 77% of samples with and without currents, respectively) and than the Fisher log series (75 and 62% of samples with and without currents, respectively).

Introducing oceanic currents in the model increases, on average, the steepness of SADs (see Fig. 2, A and B). We investigate the physical mechanisms causing this effect. One property of transport by currents is to enhance the effective diffusivity (*46*). We test whether effective diffusivity is responsible for the steepening of SADs by running our model with the effective diffusivity of the kinematic model but without currents. In this case, we find that the distribution of SAD exponent has lower average than in the case with smaller diffusion constant (see fig. S2). This implies that the increase of SAD exponents caused by currents is due to structures created by the flow that cannot be simplified into a diffusion process. We further run our model with a parameter choice yielding currents constant in time (see fig. S3). Neither in this case do we observe the steep SADs as that found in the presence of time-dependent currents.

These results suggest that the time-varying, chaotic nature of oceanic transport is responsible for the steepening of SAD curves.

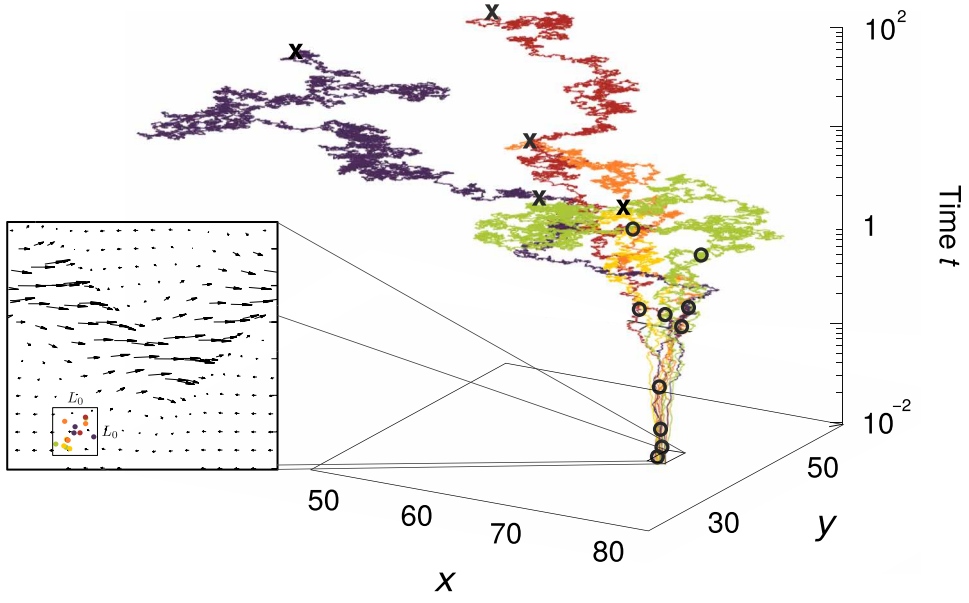

Fig. 1. Genealogy in oceanic currents. (**Left**) The coalescence model predicts the protist species composition in a sample of oceanic water taken from an area of size $L_0 \times L_0$. Different colors represent different species. Arrows represent the velocity field induced by ocean currents. (**Right**) Trajectories of the coalescence model with ocean currents. Individuals are represented by tracers that are transported backward in time and can coalesce with other tracers if they reach a close distance. Coalescence events are marked by open circles; trajectories of individuals that have coalesced are shown in the same color. Tracers are removed from the population at an immigration rate μ (marked by crosses). See also movie S1.

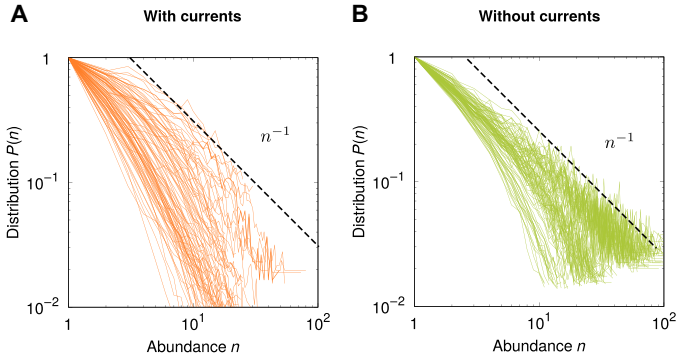

Fig. 2. Coalescence model predicts effect of chaotic advection by oceanic currents on SAD. The two panels show SADs (**A**) in the presence (orange lines) and (**B**) absence (green lines) of oceanic currents for the coalescence model. Here and below, SAD curves are rescaled so that $P(1) = 1$ to ease visualization. Model details and parameters are presented in Materials and Methods. Dashed lines are power laws to guide the eye (see also Fig. 3).

In particular, chaotic transport is often characterized by the presence of barriers limiting diffusion among certain regions of the flow. The presence of these barriers can be detected by finite-size Lyapunov exponents (FSLEs) (*47*). FSLEs quantify the growth rate of a finite separation among two particles advected by a flow. We measure local FLSEs for our model and find that they are significantly correlated with the spatial dependence of the exponent α (see fig. S4). This result supports the view that the long-lived barriers characterizing fluid flows prevent formation of large operational taxonomic units (OTUs) in the model and are thus responsible for the steepening of SAD curves.

Protist SADs are steeper in oceans than in freshwater

To test our predictions, we analyze DNA metabarcoding datasets from two studies of aquatic protists. The first dataset includes oceanic protist DNA sequences of 157 water samples from the TARA ocean expedition (*29*). The second dataset includes protist DNA sequences of 206 freshwater samples taken from lakes (*40*). We calculate SAD for each sample of both datasets using OTUs as proxies for species (see Materials and Methods). Here and in the following, if not stated otherwise, OTUs are built by clustering protist sequences at 97% sequence identity threshold. From now on, we discard "abundant species," defined as those in abundance classes $P(n)$ including less than four species. The remaining "rare species" are the subject of our study. They constitute 93% of all species in ocean samples and 78% of all species in lake samples.

As for the model, empirical SAD curves display considerable sample-to-sample variability, both in ocean and in freshwater samples (see Fig. 3). This variability is possibly caused by differences in ecological conditions among sampling sites. Empirical SAD curves are better fitted by a power law than by exponential or Fisher log series in most cases. The exponential distribution provides a better fit than the power law in 13% of lake samples and 13% of oceanic samples, whereas the Fisher log series provides better fits than the power law in 39% of lake samples and 18% of oceanic samples. We obtain similar results with different OTU definitions (95 and 99% instead of 97% similarity) and different thresholds separating abundant from rare species (see fig. S5). Notably, the power law decay of SADs is, on average, steeper in oceans than in lakes (see Fig. 3), as predicted by our coalescence model.

Distribution of the SAD exponent is quantitatively predicted by the coalescence model.

We quantify the agreement between our model and the data by analyzing the distribution of the power law exponent α in Eq. 1. In the presence of currents, the model predicts a value of the exponent significantly larger than one (average $\alpha = 1.70$, SD $\sigma = 0.68$). In the absence of oceanic currents, the model predicts an average $\alpha = 1.26$, ($\sigma = 0.46$), a value compatible with the neutral prediction $\alpha = 1$ in well-mixed systems (*31*) and spatially explicit neutral models (*43*). To verify whether the results are robust to oceanic current models, we also implement a kinematic model of the Adriatic sea and a chaotic Taylor-Green vortex (see fig. S6). In both cases, we obtain qualitatively similar results to that obtained for the meandering jet (see fig. S6), supporting that the observed mechanism is general.

Observations in both oceans and lakes are in excellent agreement with the distributions of exponents predicted by our model (see Fig. 4A). Our analysis confirms that the average exponent α is significantly larger than 1 in the oceans [average $\alpha = 1.79$, $\sigma = 0.52$; see Fig. 4A and (*29*)]. In the lakes, the average exponent is $\alpha = 1.37$ ($\sigma = 0.44$; see Fig. 4A). Adopting a different definition of OTUs (95 and 99% instead of 97%), different thresholds separating abundant from rare species and rarefying oceanic and lake data to the same sample size lead to qualitatively similar results (see fig. S5). In particular, the average exponent α in the oceans is between 4 and 23% larger than that in the lakes, depending on the threshold and the definition of OTUs (P values of Games-Howell test <0.002 in all cases).

The ocean (*29*) and lake (*40*) datasets we analyzed used different polymerase chain reaction (PCR) primers. To verify that this difference does not affect our results, we analyze two further metagenomic datasets, one from oceans (*48*) and one from lakes (*49*), that used the same primer. Also in this case, we find higher average exponent α in the oceans, confirming the robustness of our results (see fig. S7).

In the case of the meandering jet, we find that four parameters characterizing the shape and the mixing level of the jet mostly affect α. The value of the exponent is significantly correlated with the parameters ω, ϵ, and c (see Fig. 4B and fig. S8). In particular, the strong correlation with the forcing frequency ω driving the chaotic motion is a further evidence that the steepening of SAD exponents is caused by chaotic advection.

Chaotic advection by oceanic currents leads to a steeper increase in number of species as a function of sample size

By simulating our model at varying sample size N with and without currents, we predict that currents should significantly increase the number of expected species in each sample (see Fig. 5A). This effect is consistent with the increase of α in the presence of currents: Increasing α suppresses very abundant species and therefore increases the species diversity of the samples. This effect becomes more and more pronounced as N is increased. In the data, we find that samples from oceans contain more species than samples from lakes at similar sample size, which is consistent with our predictions (see Fig. 5A). The observed enrichment is even stronger than predicted by our model.

We now study the increase of number of species with sample size in oceanic and lake water samples individually. In the case of well-mixed populations, the species composition of a given sample is

described by the Ewens sampling formula (*50*), which predicts that the expected number of species in the sample is

$$S = \sum_{j=0}^{N} \frac{\theta}{\theta + j - 1} \quad (2)$$

where $\theta = 2N_{\text{eff}}\mu$ is the fundamental biodiversity number (*31*) and N_{eff} is the effective population size. Alternatively, sample species composition can be empirically described using a power law (*51*).

$$S \propto N^{z} \quad (3)$$

Our model predicts an increase in number of species with sample size, as predicted by the Ewens sampling formula (see Fig. 5, A and B). Both for ocean and freshwater samples, the power law model provides a better fit (see Fig. 5, A and B) with a higher exponent for oceanic samples ($z = 0.73$) compared to lake samples ($z = 0.65$). This result is qualitatively robust with respect to changing the OTU similarity threshold (see fig. 5C). Understanding why the observed number of species seems to depend on the sample size as a power law is an interesting question for future studies.

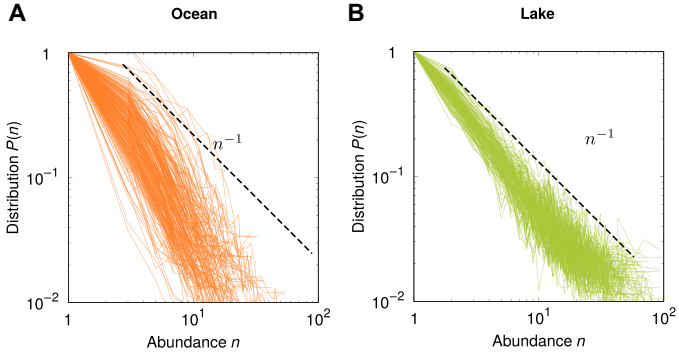

Fig. 3. Rare SADs present a steeper decay with abundance in oceans than in lakes. Continuous lines represent SADs of protist communities from (**A**) 157 oceanic samples (*29*) and (**B**) 206 freshwater samples (*40*). Total numbers of individuals in each sample are in the ranges of (A) (10^3, 10^5) and (B) (10^4, 10^6). In both panels, power laws (dashed lines) are shown to guide the eye.

DISCUSSION

Oceanic currents are known to largely affect plankton distribution at large scale (*15*–*17*). Here, we show that chaotic advection by oceanic currents profoundly affects diversity of rare protist species even at the level of single metagenomic samples. Our coalescence model bridges the gap between large-scale oceanic dynamics and ecological dynamics at the individual level and provides a versatile and powerful tool to validate individual-based ecological models using DNA metabarcoding data. Although we focus on neutral dynamics of rare protists, our approach can be extended to more general ecological settings and to other plankton communities, including animals and prokaryotes. These generalizations, combined with high-throughput sequencing data, will permit to test whether the mechanism described here affects other kingdoms characterized by different population sizes, dispersal, and spatial turnover rates (*52*). These tests can shed light on the main ecological forces determining plankton dynamics and help understanding the difference in empirically observed patterns between abundant and rare species (*24*, *25*).

The coalescence model predicts that the chaotic advection is responsible for steeper decay of SAD curves and steeper increase in the number of observed rare species with sample size. Both these predictions are in quantitative agreement with observations, although the exponent of the ocean and lake SADs largely overlap, suggesting that the trend is true globally but not necessarily so locally. The steep decay of SAD distributions in the oceans has been previously explained in terms of density-dependent effects (*29*). Although our study does not preclude this possibility, the comparison with freshwater

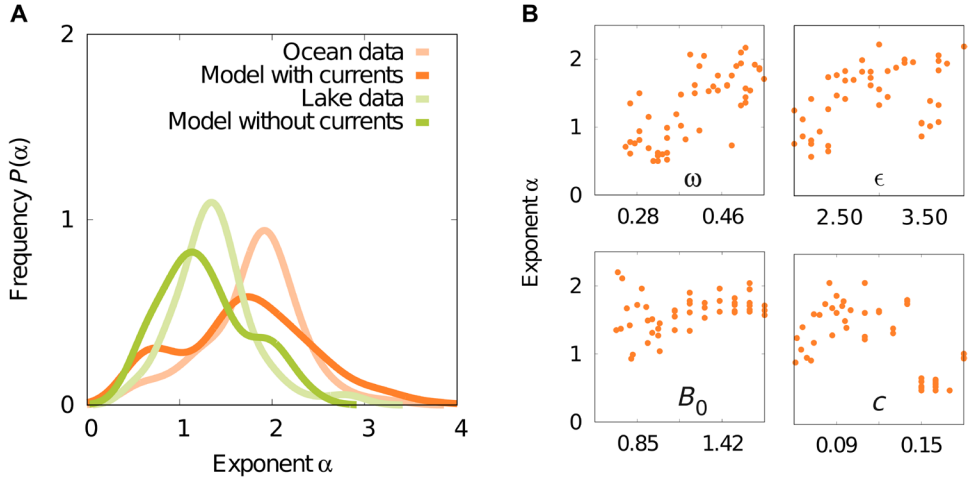

Fig. 4. Power law exponents of SADs. We run our models for different population sizes and different values of flux parameters for ocean samples (see Materials and Methods). We select 157 oceanic samples and 206 freshwater samples as in Fig. 3. We fit the power law exponent α of the SADs to the model and to the data using maximum likelihood. (**A**) Continuous distributions of the exponent obtained by kernel density estimation. (**B**) Dependence of the exponent on four main parameters of the oceanic flow: forcing frequency ω, wave perturbation amplitude ϵ, mean wave amplitude B_0, and phase speed *c*. In each subpanel, other parameters are kept constant (see Materials and Methods). Correlation tests of ω, ϵ, B_0, and *c* with the exponent α yield Pearson coefficients $r_P = 0.72$, 0.48, −0.04, and −0.58 and *P* values $P = 3 \times 10^{-9}$, 6×10^{-4}, 0.77, 10^{-5}, respectively.

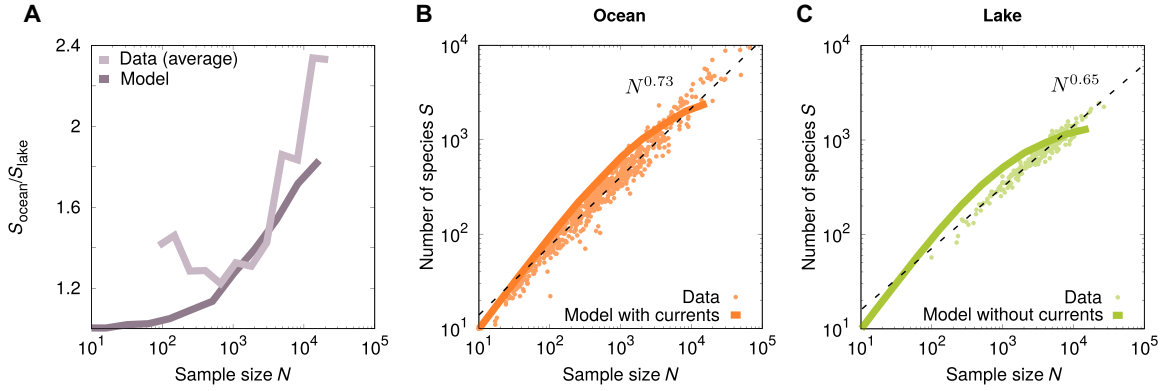

Fig. 5. Chaotic advection by oceanic currents increases species number *S* in a water sample. (**A**) Ratio S_{ocean}/S_{lake} as a function of the sample size *N* for the model and the data. We simulate the model at increasing sample sizes *N* in powers of 2 and obtain continuous curves by interpolation. Other parameters are presented in Materials and Methods. Averaged data are obtained by binning for both oceans and lakes. (**B** and **C**) Number of species *S* in samples of *N* individuals in (B) oceans and (C) lakes. A power law (Eq. 3) fits the data better than the Ewens sampling formula (Eq. 2) for both (B) oceanic (normalized log-likelihood −19.39 versus −449.52) and (C) lake samples (normalized log-likelihood −6.64 versus −117.21). Fitted exponents are $z = 0.73$ and $z = 0.65$ for oceans and lakes, respectively. The results of the coalescence model are shown with and without oceanic currents (orange and green lines, respectively). The Ewens sampling formula provides a better fit than the power law in both cases [normalized log-likelihood −420.95 versus −2131.43 in (B) and −434.66 versus −1902.62 in (C)].

ecosystems strongly suggests that chaotic advection effectively determines this density dependence. The steeper decay of SAD curves predicted by the coalescence model depends on geophysical parameters characterizing mixing. The irregular behavior of the SAD exponent as a function of these parameters (see Fig. 4B) potentially explains the fact that observed values of α appear weakly correlated with other physicochemical measurements at the sampling sites (*29*).

Chaotic flows, such as those considered here, are characterized by areas of strong mixing separated by barriers limiting transport. In our coalescence picture, these barriers reduce the pace of individual coalescence into species and therefore limit the formation of abundant species. Ecologically, this means that competition among individuals at opposite side of a barrier is reduced. This effective isolation prevents formation of very abundant species and therefore of SAD distributions with broader tails. A detailed physical theory of this phenomenon, building on recent advances on describing spatial neutral models (*53*, *43*), remains a challenge for future studies.

In summary, our study provides a mechanistic theoretical framework to analyze diversity of rare microbial species in aquatic environments at the individual level. This paves the way to quantitatively understand how the chaotic advection by oceanic currents shapes the diversity of planktonic communities.

MATERIALS AND METHODS

Coalescence model

We consider N microbial individuals in an aquatic environment and seek to reconstruct their species identity. Each individual is associated with a Lagrangian tracer with two-dimensional spatial coordinates x and y. Initially, tracers are homogeneously distributed in a sample area $L_0 \times L_0$, representing the place where the sample was collected. The tracers move in space according to the stochastic differential equations

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}x &= -u(x,y,t) + \sqrt{2D}\,\xi_x(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}y &= -v(x,y,t) + \sqrt{2D}\,\xi_y(t) \end{aligned} \tag{4}$$

where u and v are the components of an advecting velocity field representing the effect of oceanic currents. The velocity field has a minus sign since time runs backward: $t = 0$ is the time at which the sample was collected, and positive times correspond to the past history of the tracer. The terms proportional to $\sqrt{2D}$ are diffusion terms modeling individual movement and small-scale turbulence. The quantities $\xi_x(t)$, $\xi_y(t)$ are independent white noise sources satisfying $\langle \xi_i(t) \rangle = 0$, $\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t - t')$ where $\langle \ldots \rangle$ denotes an average and $i, j \in (x, y)$. The advecting field u, v is specified in the next subsection.

To reconstruct the species identity of the tracers, we track their positions backward in time. Tracers at a short distance δ from each other at time $t \geq 0$ can coalesce at a rate r. If this event occurs, then individuals in the sample represented by the two tracers descend from a common ancestor at time t in the past and therefore belong to the same species. We implement immigration events by assigning species at a rate μ. At each time step dt:

1)Each tracer moves from its position (x, y) to $(x + \Delta x, y + \Delta y)$. The increments Δx, Δy are obtained by numerically integrating Eq. 4.

2)Tracers are selected one by one and are removed with probability μ dt (immigration event). Further, each tracer i can coalesce with probability r dt with any other tracer j present in an area of size δ × δ centered at the coordinates of tracer i.

We set $r = 1$, $\mu = 10^{-4}$, and the diffusion constant to $D = 3 \times 10^{-9}$, as further discussed below. The interaction distance δ is chosen to satisfy $D = r\delta^2$, see (*41*). We take the linear size of the sample area on the order of the mean distance traveled by an individual in one generation, $L_0 = 5$ km, estimating a protist lifetime of about 1 day (*54*) and protist movements of about 20 km^2 day^{-1} (*55*). Population size is randomly selected for each run in the range $N \in (10^3, 10^5)$ unless otherwise indicated. For Fig. 4B and movie S1, we set $N = 8192$.

Each simulation is run until all individuals have been assigned to OTUs by coalescence or immigration events. In the range of parameters we explored, the duration of a run is on the order of 100 years. However, most coalescence events occur on a much faster time scale (median of the coalescence times $t \approx 82$ days for the ocean case and $t \approx 67$ days for the freshwater case).

Kinematic model of the oceans

We model large-scale oceanic currents by means of a kinematic model of a meandering jet (*44*, *45*). The velocity field u, v is defined in terms of a stream function. In a fixed reference frame, the stream function reads

$$\psi(x',y',t') = \psi_0\left\{1 - \tanh\left(\frac{y' - A(t')\cos\left[\kappa(x'-c_x t')\right]}{\lambda\sqrt{1+\kappa^2 A^2(t')\sin^2\left[\kappa(x'-c_x t')\right]}}\right)\right\} \quad (5)$$

The stream function is more conveniently written in a dimensionless form

$$\phi(x,y,t) = -\tanh\left(\frac{y - B(t)\cos(kx)}{\sqrt{1+k^2B^2(t)\sin^2(kx)}}\right) + cy \quad (6)$$

being $B(t) = A(t')/\lambda = B_0 + \epsilon\cos(\omega t + \Phi)$, $c = c_x L/\psi_0$, and $k = 2\pi/L$, with L the meander wavelength. The transformation between dimensional and dimensionless units is $x = (x' - c_x t)/\lambda$, $y = y'/\lambda$, and $t = t'\psi_0/\lambda^2$ (*44*). The frame of reference of the dimensionless coordinates moves with the speed c_x of the jet. Given the stream function, the components of the velocity field in dimensionless units are

$$\begin{aligned} u &= -\partial\phi/\partial y \\ v &= \partial\phi/\partial x \end{aligned} \quad (7)$$

We run the simulations using the moving dimensionless coordinates in a virtually infinite system. In the case without currents, this modeling choice is justified a posteriori by the fact that, on the basis of our observations, lake SAD exponents do not present a significant dependence on lake area (see fig. S9). For the ocean simulations, results can be affected by the position of the sample area. For this reason, we place the sample area $L_0 \times L_0$ at random coordinates x_0, $y_0 \in (0,8)$ for each run. For Fig. 5, we fix $x_0 = 7.5$ and $y_0 = 1$.

Parameters of the kinematic model

Realistic parameters of the dimensionless stream function, Eq. 6, are estimated as $L = 7.5$, $c = 0.12$, $B_0 = 1.2$, $\omega = 0.4$, $\epsilon = 0.3$, and $\Phi = \pi/2$ (*45*). We consider parameter ranges based on these values $c \in (0.06, 0.18)$, $B_0 \in (0.7, 1.7)$, and $\omega \in (0.25, 0.55)$ and fix $L = 7.5$, $\Phi = \pi/2$. The value of ϵ has to be larger than a critical value depending on ω to prevent transported particles to remain trapped into long-lived eddies (*45*). To meet this condition while exploring a range of values of ω, we fix $\epsilon \in (2,4)$. For Figs. 4B and 5, we set $c = 0.12$, $B_0 = 1.2$, $\omega = 0.5$, and $\epsilon = 3$.

To convert from dimensionless units to dimensional units, we use the spatial scale $\lambda = 40$ km (*44*) and the stream function scale $\psi_0 = 160$ km^2 day^{-1}. With this choice, the time unit λ^2/ψ_0 is equal to 1 day. The parameter ψ_0/λ represents the maximum velocity in the center of the jet. With our choice of units, the velocity is equal to 40 km day^{-1}, slightly lower than the average velocity of large-scale oceanic currents (about $\psi_0/\lambda \approx 200$ km day^{-1} for the surface Gulf stream and 50 km day^{-1} for the lower thermocline (*44*)).

In physical units, the coalescence rate is equal to $r = 1$ day^{-1}, i.e., about one generation time for protists (*54*). Our choice of the diffusion constant to $D = 3 \times 10^{-9}$ in dimensionless units corresponds to about 6×10^{-5} m^2/s in physical units, which is consistent with observations (*46*).

Species abundance distribution

We compute the distribution $P(n)$ of the species abundances n for each sample. Species with low-to-intermediate abundance appear to follow a different distribution than abundant species, as previously observed (*27*–*29*). For this reason, we filter out species in abundance classes below $P(n) = 4$. To avoid overfitting, we also discard samples with SAD composed of less than 10 points with different frequencies $P(n)$. After this selection, we are left with 157 samples for oceans and 206 for lakes. We compute the SAD $P(n)$ for the coalescent model with and without advection. Each sample is obtained for different flux parameters and population sizes (described above). The resulting distributions $P(n)$ are averaged over up to 10^2 realizations of the model and filtered in the same way as the data samples for consistency.

Data fits

To determine the exponent α, we fit the function

$$P(n) = C/n^{\alpha} \quad (8)$$

in a range of intermediate abundances ($n_{\min}$, $n_{\max}$). The exponent α, the proportionality constant C, and the values of $n_{\min}$ and $n_{\max}$ are simultaneously determined by maximizing the normalized log-likelihood $\ln L = (1/\mathcal{N})\sum_i[n_i \ln P(n_i) - P(n_i) + \ln(n_i!)]$, where $\mathcal{N}$ is the number of nonzero abundance classes for n in the range ($n_{\min}$, $n_{\max}$) and we assumed Poissonian counts. We discard samples for which the range ($n_{\min}$, $n_{\max}$) includes less than 5 points. We also fit an exponential $P(n) = Ce^{-cn}$ and a Fisher log series $P(n) = Ce^{-cn}/n$ with the same method and in the same interval $[n_{\min}, n_{\max}]$ determined with the power law fit. Since all the distributions have the same number of free parameters, we always consider a better fit the distribution characterized by the largest normalized log-likelihood. The percentage of data samples for which a power law fits better than the exponential and Fisher log series and the corresponding exponents is presented in fig. S5.

OTU analysis

We analyze metabarcoding data from marine (*29*) and freshwater (*40*) protist planktonic communities. We retrieve the dataset of oceanic samples from the European Nucleotide Archive (accession ID PRJEB16766) (further referred to as the Ocean dataset). The dataset consists of assembled paired-end Illumina HiSeq2000 sequencing reads of PCR-amplified V9 loop of protist 18*S* rRNA gene obtained from 121 seawater locations distributed worldwide. We trim the primer sites using USEARCH (v.11.0.667) (*56*). Primer sites include 15 and 20 nucleotide sites for the 5- and 3-end, respectively. The trimmed sequences are quality-filtered with USEARCH using the option –fastq_maxee 1.0, which discards sequences with >1 total expected errors in the sequence. The sequences are dereplicated, and singleton sequences (i.e., sequences with single occurrence in the dataset) are removed using VSEARCH (v.2.10.1) (*57*). Chimeric sequences are detected and removed using UCHIME (implemented in VSEARCH) (*58*) and a combination of reference-based (with nonredundant SILVA SSU Ref database ver.132 used as reference) and de novo methods. Sequences are then clustered into OTUs using VSEARCH and the –cluster_size option. We use sequence identity thresholds of 95, 97, and 99%, which provides different levels of taxonomic resolution. To account for different depths of sequencing between samples, the OTU tables are subsampled (rarefied) to depths of 5000 and 10,000 reads per sample using script single_rarefaction.py from QIIME package (*59*). In addition, we analyze a second global oceanic dataset of deep-water ocean samples

(*48*), further referred to as the Pernice ocean dataset. The Pernice ocean dataset consists of 454 pyrosequencing reads of PCR-amplified V4 region of 18*S* rRNA gene obtained from RNA isolated from 27 seawater locations distributed worldwide (*48*). First, primer sites [primer TAReuk454FWD1: CCAGCA(G/C)C(C/T)GCGG-TAATTCC] are trimmed with CUTADAPT (*60*), discarding reads missing the primer site. After trimming, the dataset is analyzed, as described above for the Ocean dataset, with the exception of (i) quality filtering with option –fastq_maxee 2.0, which discards sequences with >2 total expected errors in the sequence, and (ii) taxonomy assignment and taxonomy-based filtering, as described below for the Lake dataset.

We obtain the freshwater dataset, consisting of paired-end Illumina HiSeq2500 reads of amplified genomic region encompassing V9 loop of 18*S* rRNA and ITS1 gene for 217 European freshwater lakes, from the Short Read Archive (Bioproject ID PRJNA414052). First, reads from PCR replicates and sequencing replicates are merged for each lake sample. Next, primer regions are trimmed with CUTADAPT (*60*), discarding reads missing one or both of the primer sites. Forward and reverse reads with a minimal overlap of 70 base pairs and with a maximum of 5 nucleotide differences in the overlapping region are merged with VSEARCH (command –fastq_mergepairs). Next, we extract from the amplified SSU V9 + ITS1 region the SSU V9 region using ITSx (v.1.1.1) (*61*). This step allows the taxonomic resolution of the clustered freshwater planktonic community OTUs to closely resemble the taxonomic community resolution of the marine planktonic community, which is based on sequenced V9 loop regions of 18*S* rRNA genes. The reads with >1 total expected errors in the sequence are discarded, datasets are dereplicated, singletons and chimeras are removed, and the quality-filtered reads are clustered into OTUs and rarefied, as described above for the Ocean dataset. The taxonomy is assigned against SILVA v123 eukaryotic 18*S* subset database. OTUs assigned to Fungi, Metazoa, or Embryophyta (i.e., nonprotist eukaryotes) with at least Bootstrap Support (BS) >0.8 support and OTUs not assigned to kingdom level (BS <0.8) are excluded from the final OTU tables. In addition, we analyze a second lake dataset, further referred to as the Filker lakes dataset (*49*). The Filker lakes dataset consists of merged and quality-filtered Illumina sequencing reads of PCR-amplified V4 region of 18*S* rRNA gene obtained from 13 high-mountain lakes distributed worldwide (*49*). The reads are dereplicated, singletons and chimeras are removed, filtered reads are clustered into OTUs and rarefied, and taxonomy is assigned, as described above for the Lake dataset.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/29/eaaz9037/DC1

REFERENCES AND NOTES

1. B. J. Finlay, Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063 (2002).
2. L. G. M. Baas Becking, *Geobiologie of Inleiding Tot De Milieukunde* (WP Van Stockum & Zoon, 1934), pp. 18–19.
3. J. A. Fuhrman, Microbial community structure and its functional implications. *Nature* **459**, 193–199 (2009).
4. M. Stomp, J. Huisman, G. G. Mittelbach, E. Litchman, C. A. Klausmeier, Large-scale biodiversity patterns in freshwater phytoplankton. *Ecology* **92**, 2096–2107 (2011).
5. S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. d'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis; Tara Oceans coordinators, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, P. Bork, Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
6. C. de Vargas, S. Audic, N. Henry, J. Decelle, F. Mahé, R. Logares, E. Lara, C. Berney, N. Le Bescot, I. Probert, M. Carmichael, J. Poulain, S. Romac, S. Colin, J.-M. Aury, L. Bittner, S. Chaffron, M. Dunthorn, S. Engelen, O. Flegontova, L. Guidi, A. Horák, O. Jaillon, G. Lima-Mendez, J. Lukeš, S. Malviya, R. Morard, M. Mulot, E. Scalco, R. Siano, F. Vincent, A. Zingone, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, T. O. Coordinators, S. G. Acinas, P. Bork, C. Bowler, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, F. Not, H. Ogata, S. Pesant, J. Raes, M. E. Sieracki, S. Speich, L. Stemmann, S. Sunagawa, J. Weissenbach, P. Wincker, E. Karsenti, Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).
7. S. A. Levin, Community equilibria and stability, and an extension of the competitive exclusion principle. *Am. Nat.* **104**, 413–423 (1970).
8. G. E. Hutchinson, The paradox of the plankton. *Am. Nat.* **95**, 137–145 (1961).
9. M. Scheffer, S. Rinaldi, J. Huisman, F. J. Weissing, Why plankton communities have no equilibrium: Solutions to the paradox. *Hydrobiologia* **491**, 9–18 (2003).
10. E. Litchman, C. A. Klausmeier, O. M. Schofield, P. G. Falkowski, The role of functional traits and trade-offs in structuring phytoplankton communities: Scaling from cellular to ecosystem level. *Ecol. Lett.* **10**, 1170–1181 (2007).
11. E. Litchman, C. A. Klausmeier, Trait-based community ecology of phytoplankton. *Annu. Rev. Ecol. Evol. Syst.* **39**, 615–639 (2008).
12. J. A. Bonachela, M. Raghib, S. A. Levin, Dynamic model of flexible phytoplankton nutrient uptake. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20633–20638 (2011).
13. C. T. Kremer, C. A. Klausmeier, Coexistence in a variable environment: Eco-evolutionary perspectives. *J. Theor. Biol.* **339**, 14–25 (2013).
14. J. A. Bonachela, C. A. Klausmeier, K. F. Edwards, E. Litchman, S. A. Levin, The role of phytoplankton diversity in the emergent oceanic stoichiometry. *J. Plankton Res.* **38**, 1021–1035 (2015).
15. A. Bracco, A. Provenzale, I. Scheuring, Mesoscale vortices and the paradox of the plankton. *Proc. Biol. Sci.* **267**, 1795–1800 (2000).
16. G. Károlyi, Á. Péntek, I. Scheuring, T. Tél, Z. Toroczkai, Chaotic flow: The physics of species coexistence. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13661–13665 (2000).
17. F. d'Ovidio, S. De Monte, S. Alvain, Y. Dandonneau, M. Lévy, Fluid dynamical niches of phytoplankton types. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18366–18370 (2010).
18. D. J. McGillicuddy Jr., Mechanisms of physical-biological-biogeochemical interaction at the oceanic mesoscale. *Ann. Rev. Mar. Sci.* **8**, 125–159 (2016).
19. J. B. H. Martiny, B. J. M. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Ovreås, A.-L. Reysenbach, V. H. Smith, J. T. Staley, Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).
20. M. C. Urban, M. A. Leibold, P. Amarasekare, L. De Meester, R. Gomulkiewicz, M. E. Hochberg, C. A. Klausmeier, N. Loeuille, C. De Mazancourt, J. Norberg, J. H. Pantel, S. Y. Strauss, M. Vellend, M. J. Wade, The evolutionary ecology of metacommunities. *Trends Ecol. Evol.* **23**, 311–317 (2008).
21. P. E. Galand, E. O. Casamayor, D. L. Kirchman, C. Lovejoy, Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22427–22432 (2009).
22. G. Chust, X. Irigoien, J. Chave, R. P. Harris, Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Glob. Ecol. Biogeogr.* **22**, 531–543 (2013).
23. D. Wilkins, E. van Sebille, S. R. Rintoul, F. M. Lauro, R. Cavicchioli, Advection shapes southern ocean microbial assemblages independent of distance and environment effects. *Nat. Commun.* **4**, 2457 (2013).
24. M. L. Sogin, H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, G. J. Herndl, Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12115–12120 (2006).
25. C. Pedrós-Alió, The rare bacterial biosphere. *Ann. Rev. Mar. Sci.* **4**, 449–466 (2012).
26. M. D. J. Lynch, J. D. Neufeld, Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **13**, 217–229 (2015).
27. A. E. Magurran, P. A. Henderson, Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**, 714–716 (2003).
28. W. Ulrich, M. Ollik, Frequent and occasional species and the shape of relative-abundance distributions. *Divers. Distrib.* **10**, 263–269 (2004).
29. E. Ser-Giacomi, L. Zinger, S. Malviya, C. De Vargas, E. Karsenti, C. Bowler, S. De Monte, Ubiquitous abundance distribution of non-dominant plankton across the global ocean. *Nat. Ecol. Evol.* **2**, 1243–1249 (2018).
30. P. Jeraldo, M. Sipos, N. Chia, J. M. Brulc, A. S. Dhillon, M. E. Konkel, C. L. Larson, K. E. Nelson, A. Qu, L. B. Schook, F. Yang, B. A. White, N. Goldenfeld, Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9692–9698 (2012).

31. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)* (Princeton Univ. Press, 2001).
32. S. Pigolotti, A. Flammini, A. Maritan, Stochastic model for the species abundance problem in an ecological community. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **70**, 011916 (2004).
33. F. He, Deriving a neutral model of species abundance from fundamental mechanisms of population dynamics. *Funct. Ecol.* **19**, 187–193 (2005).
34. S. Azaele, S. Pigolotti, J. R. Banavar, A. Maritan, Dynamical evolution of ecosystems. *Nature* **444**, 926–928 (2006).
35. M. Lévy, P. J. S. Franks, K. S. Smith, The role of submesoscale currents in structuring marine ecosystems. *Nat. Commun.* **9**, 4758 (2018).
36. J. P. Wares, J. M. Pringle, Drift by drift: Effective population size is limited by advection. *BMC Evol. Biol.* **8**, 235 (2008).
37. F. Herrerías-Azcué, V. Pérez-Muñuzuri, T. Galla, Stirring does not make populations well mixed. *Sci. Rep.* **8**, 4068 (2018).
38. S. Pigolotti, R. Benzi, M. H. Jensen, D. R. Nelson, Population genetics in compressible flows. *Phys. Rev. Lett.* **108**, 128102 (2012).
39. A. Plummer, R. Benzi, D. R. Nelson, F. Toschi, Fixation probabilities in weakly compressible fluid flows. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 373–378 (2019).
40. J. Boenigk, S. Wodniok, C. Bock, D. Beisser, C. Hempel, L. Grossmann, A. Lange, M. Jensen, Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *Metabarcod. Metagenom.* **2**, e21519 (2018).
41. S. Pigolotti, R. Benzi, P. Perlekar, M. H. Jensen, F. Toschi, D. R. Nelson, Growth, competition and cooperation in spatial population genetics. *Theor. Popul. Biol.* **84**, 72–86 (2013).
42. J. Rosindell, Y. Wong, R. S. Etienne, A coalescence approach to spatial neutral ecology. *Eco. Inform.* **3**, 259–271 (2008).
43. S. Pigolotti, M. Cencini, D. Molina, M. A. Muñoz, Stochastic spatial models in ecology: A statistical physics approach. *J. Stat. Phys.* **172**, 44–73 (2018).
44. A. S. Bower, A simple kinematic mechanism for mixing fluid parcels across a meandering jet. *J. Phys. Oceanogr.* **21**, 173–180 (1991).
45. M. Cencini, G. Lacorata, A. Vulpiani, E. Zambianchi, Mixing in a meandering jet: A Markovian approximation. *J. Phys. Oceanogr.* **29**, 2578–2594 (1999).
46. A. P. Martin, Phytoplankton patchiness: The role of lateral stirring and mixing. *Prog. Oceanogr.* **57**, 125–174 (2003).
47. G. Boffetta, G. Lacorata, G. Redaelli, A. Vulpiani, Detecting barriers to transport: A review of different techniques. *Physica D* **159**, 58–70 (2001).
48. M. C. Pernice, C. R. Giner, R. Logares, J. Perera-Bel, S. G. Acinas, C. M. Duarte, J. M. Gasol, R. Massana, Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J.* **10**, 945–958 (2016).
49. S. Filker, R. Sommaruga, I. Vila, T. Stoeck, Microbial eukaryote plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns. *Mol. Ecol.* **25**, 2286–2301 (2016).
50. H. Crane, The ubiquitous ewens sampling formula. *Statist. Sci.* **31**, 1–19 (2016).
51. H. Morlon, D. W. Schwilk, J. A. Bryant, P. A. Marquet, A. G. Rebelo, C. Tauss, B. J. M. Bohannan, J. L. Green, Spatial patterns of phylogenetic diversity. *Ecol. Lett.* **14**, 141–149 (2011).
52. E. Villarino, J. R. Watson, B. Jönsson, J. M. Gasol, G. Salazar, S. G. Acinas, M. Estrada, R. Massana, R. Logares, C. R. Giner, M. C. Pernice, M. P. Olivar, L. Citores, J. Corell, N. Rodríguez-Ezpeleta, J. L. Acuña, A. Molina-Ramírez, J. I. González-Gordillo, A. Cózar, E. Martí, J. A. Cuesta, S. Agustí, E. Fraile-Nuez, C. M. Duarte, X. Irigoien, G. Chust, Large-scale ocean connectivity and planktonic body size. *Nat. Commun.* **9**, 142 (2018).
53. P. A. Marquet, G. Espinoza, S. R. Abades, A. Ganz, R. Rebolledo, On the proportional abundance of species: Integrating population genetics and community ecology. *Sci. Rep.* **7**, 16815 (2017).
54. A. J. Milligan, Oceanography: Plankton in an acidified ocean. *Nat. Clim. Change* **2**, 489 (2012).
55. H. Kontoyiannis, D. R. Watts, Observations on the variability of the gulf stream path between 74°W and 70°W. *J. Phys. Oceanogr.* **24**, 1999–2013 (1994).
56. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
57. T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
58. R. C. Edgar, B. J. Haas, J. C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
59. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. Mcdonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, Correspondence QIIME allows analysis of high-throughput community sequencing data intensity normalization improves color calling in SOLiD sequencing. *Nat. Methods.* **7**, 335–336 (2011).
60. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).
61. J. Bengtsson-Palme, M. Ryberg, M. Hartmann, S. Branco, Z. Wang, A. Godhe, P. De Wit, M. Sánchez-García, I. Ebersberger, F. de Sousa, A. S. Amend, A. Jumpponen, M. Unterseher, E. Kristiansson, K. Abarenkov, Y. J. K. Bertrand, K. Sanli, K. M. Eriksson, U. Vik, V. Veldre, R. H. Nilsson, Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* **4**, 914–919 (2013).
62. L. F. Richardson, Atmospheric diffusion shown on a distance-neighbour graph. *Proc. R. Soc. Lond.* **110**, 709–737 (1926).
63. S. Berti, F. A. D. Santos, G. Lacorata, A. Vulpiani, Lagrangian drifter dispersion in the southwestern atlantic ocean. *J. Phys. Oceanogr.* **41**, 1659–1672 (2011).
64. G. Lacorata, E. Aurell, A. Vulpiani, Drifter dispersion in the adriatic sea: Lagrangian data and chaotic model. *Ann. Geophys.* **19**, 121–129 (2001).
65. T. Bohr, M. H. Jensen, G. Paladin, A. Vulpiani, *Dynamical Systems Approach to Turbulence* (Cambridge Univ. Press, 2005).

Acknowledgments: We thank M. Cencini, E. Economo, M. A. Muñoz, and L. Peliti for comments on a preliminary version of the manuscript. **Funding:** We acknowledge support from OIST core funding. **Author contributions:** P.V.M. and S.P. conceived the research; P.V.M. and S.P. wrote the code and performed the numerical simulations; A.B. and T.B. analyzed the metagenomic data; P.V.M. and S.P wrote the first draft, with subsequent input from A.B. and T.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 18 October 2019
Accepted 29 May 2020
Published 15 July 2020
10.1126/sciadv.aaz9037

Citation: P. V. Martín, A. Buček, T. Bourguignon, S. Pigolotti, Ocean currents promote rare species diversity in protists. *Sci. Adv.* **6**, eaaz9037 (2020).