

Strongly Deterministic Population Dynamics in Closed Microbial CommunitiesZak Frentz,^{1,*} Seppe Kuehn,^{1,2} and Stanislas Leibler^{1,3}¹*Laboratory of Living Matter and the Center for Studies in Physics and Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA*²*Department of Physics, Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, USA*³*The Simons Center for Systems Biology and School of Natural Sciences, Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA*

(Received 15 May 2015; published 26 October 2015)

Biological systems are influenced by random processes at all scales, including molecular, demographic, and behavioral fluctuations, as well as by their interactions with a fluctuating environment. We previously established microbial closed ecosystems (CES) as model systems for studying the role of random events and the emergent statistical laws governing population dynamics. Here, we present long-term measurements of population dynamics using replicate digital holographic microscopes that maintain CES under precisely controlled external conditions while automatically measuring abundances of three microbial species via single-cell imaging. With this system, we measure spatiotemporal population dynamics in more than 60 replicate CES over periods of months. In contrast to previous studies, we observe strongly deterministic population dynamics in replicate systems. Furthermore, we show that previously discovered statistical structure in abundance fluctuations across replicate CES is driven by variation in external conditions, such as illumination. In particular, we confirm the existence of stable ecomodes governing the correlations in population abundances of three species. The observation of strongly deterministic dynamics, together with stable structure of correlations in response to external perturbations, points towards a possibility of simple macroscopic laws governing microbial systems despite numerous stochastic events present on microscopic levels.

DOI: [10.1103/PhysRevX.5.041014](https://doi.org/10.1103/PhysRevX.5.041014)

Subject Areas: Biological Physics, Statistical Physics

I. INTRODUCTION

Random processes pervade biological phenomena at all scales, from molecular interactions in cells [1,2] to interactions between organisms in ecosystems [3,4]. These stochastic processes can dramatically alter system dynamics, leading some to argue that historical contingency is an essential part of long-term biological dynamics and that narrative is the appropriate framework to describe biological systems [5]. Dependence on random processes is particularly relevant for ecological dynamics, which include all scales of interactions. When studying ecology in natural conditions, one rarely knows every interacting species, nor can one monitor precisely the chemical and physical environment. In these conditions, population dynamics measurements typically consist of a single, unique, time series [6,7]; thus, it is impossible to separate random fluctuations from trends [8]. In this case, each

change in species abundance is a unique event, whose origins one can only speculate about. In fact, any organism that is part of a complex ecosystem can be subject to random interactions with other individuals of the same or of different species [3,4]. It can also interact with a stochastically varying physical environment, characterized by time-dependent chemical and physical parameters, such as temperature, pressure, chemical potential, or illuminance [9,10]. In light of such pervasive stochastic processes, acting on a variety of temporal and spatial scales, the search for general principles underlying ecological dynamics should follow the example of other fields in biology, such as genetics, and move into the laboratory, where controlled experiments, performed on replicate systems, can bring valuable insights [11,12].

A. Closed ecosystems

In order to assess the role of stochastic events in ecosystems on relatively long time scales, we recently used closed microbial ecosystems composed of three species: the photosynthetic green alga (denoted as species A) *Chlamydomonas reinhardtii*, the bacterium (B) *Escherichia coli*, and the ciliate (C) *Tetrahymena thermophila* [13]. Closed ecosystems (CES) [14–16] are

*zfrentz@rockefeller.edu

Published by the American Physical Society under the terms of the *Creative Commons Attribution 3.0 License*. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

sealed to material exchange but open to heat exchange and incident light. As model systems, CES are ideal for studying population dynamics since they permit the long-term maintenance of microbial communities under controlled external conditions. Replicate CES can be created with low variation in initial conditions (nutrients, species abundances), permitting the study of population dynamics in ensembles of replicate ecosystems. Quantitative, non-invasive methods for measuring population dynamics were developed for characterizing these systems: light-sheet fluorescence microscopy in previous work [13] and, in the present study, digital in-line holographic (DIH) microscopy [17]. Measuring population dynamics in replicate CES permits us to explicitly quantify the role of stochastic events and deterministic processes in system dynamics.

Under constant illumination, these closed ecosystems sustain their three species for years. The detailed ecological interactions in this system are not well known. However, each of the three species is well characterized biologically. The green alga *C. reinhardtii* is widely distributed in soil and water, is phototactic, and presumably acts as a primary producer in our CES by fixing carbon through photosynthesis [18]. *T. thermophila*, a single-celled eukaryote that is common in freshwater ponds, commonly consumes bacteria but can also grow axenically [19]. The bacterium *E. coli*, which is commonly found in mammalian guts and soil [20], acts to some extent as a decomposer in our CES, metabolizing nutrients that are not utilized by the other species. To our knowledge, these three organisms are not commonly found in the same wild populations nor are they isolated from hundreds of other interacting species. Our CES are therefore purely synthetic, and we should not expect that any specific preadaptation between these species has taken place naturally. In ideal laboratory conditions, these organisms' doubling times are as short as 6 h, 20 min, and 2 h for *A*, *B*, and *C*, respectively. However, the doubling time of each species in our CES, which varies over time, is not known.

As demonstrated in a recent study of diatoms interacting with bacteria [21], it is likely that many chemical interactions arise between the three species in our CES. These interactions obscure the simple producer, consumer, decomposer classification described above. In fact, our CES exhibited a wide range of complex phenomena [13]. In the presence of gravity and thermal convection, each CES quickly became an ensemble of local ecological niches. For example, at the bottom of the containers, we observed large clumps of algae living among debris and dense colonies of nonmotile bacteria. Phenotypic transformations occurred spontaneously in many CES including filamentous *E. coli* and, most strikingly, the appearance of unusually large *T. thermophila* (with ≈ 10 -fold bigger volume) capable of ingesting algae [22]. In addition to these phenotypic changes, random processes taking place within each CES included swimming, taxis, predation,

genetic variation, cell division, and death, which are all important sources of biological complexity.

B. Correlation matrix and ecomodes

Rather than attempt to characterize all of the processes taking place in a CES, a previous study using many replicate ecosystems characterized the structure of the variation in the abundance dynamics in time and across replicate CES. In Fig. 1(c), we reproduce, for comparison, the data obtained in Ref. [13]—we plot the geometric mean abundances for all three species ($s = A, B, C$), $\mu_s = \exp(\langle \log N_s(t) \rangle)$. Here, $N_s(t)$ are defined as numbers of individuals in 1 mL, rather than densities, and $\langle \cdot \rangle$ denotes an average across 24 replicate ecosystems.

The shaded regions in Fig. 1(c) define the geometric standard deviations, $\sigma_s(t) = \exp(\sigma[\log N_s(t)])$, where the standard deviation is computed across replicate ecosystems at each point in time t . The fluctuations across replicates, as measured through $\sigma_s(t)$, increased monotonically in time. We defined the correlation matrix,

$$C_{s,s'}(t) = \frac{\text{Cov}[\log N_s(t), \log N_{s'}(t)]}{\sigma[\log N_s(t)]\sigma[\log N_{s'}(t)]}, \quad (1)$$

where the covariance is computed across replicates. Remarkably, after the first 21 days, the fluctuations of abundances maintained a stable correlation structure in time as quantified by the eigensystem of $C(t)$. The eigenvectors of this matrix were dubbed “ecomodes”; the ecomode corresponding to the largest eigenvalue dominated the fluctuations and reflected positively correlated fluctuations between all three species (see Ref. [13], Fig. 4A). Thus, despite the apparent complexity present in CES, measurements of replicate systems revealed a simple structure in the fluctuations of abundances.

The observation of structured, increasing fluctuations across replicate ecosystems raises an important question: Are these fluctuations driven by processes intrinsic to the ecosystem (e.g., phenotypic switching, mutation, demographic fluctuations), or are they a consequence of uncontrolled variation in external conditions across replicate ecosystems (e.g., illumination, temperature)? In other words, do the intrinsic dynamics follow a strongly deterministic trajectory? Or, do stochastic processes, known to be present in the system, drive the variation observed in Fig. 1(c)?

C. Digital in-line holographic microscopes with well-controlled external conditions

In order to address this question, we built a new instrument consisting of ten DIH microscopes [17]. Hermetically sealed 2-mL cuvettes, each containing 1.6 mL of medium and organisms, were positioned horizontally inside the microscopes, and holograms were

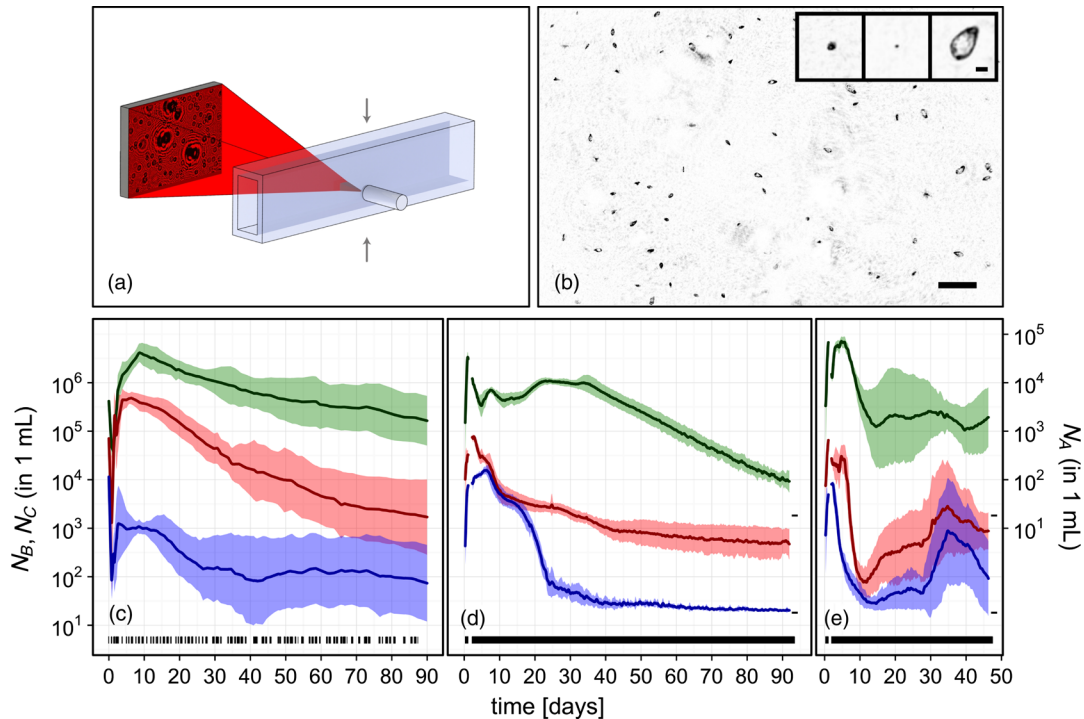


FIG. 1. Closed ecosystem population dynamics by digital holography. (a) Imaging schematic: A red laser beam diverges from a focus at the surface of a gradient-index (GRIN) lens (cylinder on lower right) and illuminates a region within the cuvette containing the microbial ecosystem, forming an interference pattern or hologram, which is recorded by a camera sensor (rectangle, left). Arrows above and below the cuvette indicate the direction of homogeneous white illumination provided by LEDs. (b) A single plane, near the wall of the cuvette facing the camera sensor, is reconstructed computationally from the hologram in (a). Gray scale represents intensity. The scale bar is $200 \mu\text{m}$. Many *T. thermophila* and some *C. reinhardtii* are visible. *E. coli* are not readily distinguished by eye in this image (see Appendix B). Inset: Magnified images of, from left to right, *C. reinhardtii*, *E. coli*, and *T. thermophila* at the same magnification and contrast. The scale bar is $10 \mu\text{m}$. (c) Population dynamics from previous work [13], measured by fluorescence microscopy, in 24 replicate ecosystems. The curve indicates the geometric mean abundance across replicates, and the shaded region includes 1 geometric standard deviation above and below the geometric mean. Vertical black dashes along the bottom indicate measurement times. Green (A—algae, right axis): *C. reinhardtii*; red (B—bacteria): *E. coli*; blue (C—ciliates): *T. thermophila*. (d) Population dynamics measured by digital holography, in 10 replicate ecosystems with $I = I_0$ (1200 lx) constant illumination. Species abundances were measured every 400 s. A gap occurred between 1.1 and 2.3 days, when abundances were too high to measure. The horizontal dashes on the right indicate the offset applied to zero counts to avoid divergence of log transforms. The individual time series used to calculate these statistics are shown in Fig. 2(b). (e) Geometric mean and standard deviation for 10 replicates in $I = 0.25I_0$ illumination [see individual time series in Fig. 2(a)]. Note that for (c–e) the vertical axis on the left of panel (c) applies to N_B and N_C for all three panels, while the vertical axis on the right of panel (e) applies to N_A in all three panels.

acquired at 400-s intervals [Fig. 1(a)]. Each hologram was reconstructed computationally [17], yielding three-dimensional micrographs of the $\approx 5 \text{ mm}^3$ imaging volume with sufficient resolution and sensitivity to distinguish species morphologically [Fig. 1(b) and Appendix B]. The holographic microscopes allowed us to measure abundances of all three species with temporal resolution of a few minutes, over periods of months. Each microscope provided precise control over temperature and illuminance (20 mK and 2.5% variation across replicates, respectively), making the control of external conditions an order of magnitude more precise than the experiments shown in Fig. 1(c). Using ten dedicated microscopes, we were able to gather a large amount of data for different illumination levels (totaling 9.7 microscope years of data).

II. RESULTS

A. Population dynamics

Figure 1(d) shows the geometric mean dynamics and the geometric standard deviation for ten replicate CES using this instrument under conditions similar but not identical (see Appendix A 2) to those shown in Fig. 1(c). It is immediately clear that the observed dynamics are very different from those measured previously [Fig. 1(c)], in which fluctuations were over threefold for algae and over fivefold for bacteria and ciliates by the end. Under carefully controlled external conditions in our holographic microscopes, fluctuations across replicates no longer increased over the course of the experiment but instead remained small, on the order of 1.5-fold for algae, 1.9-fold for

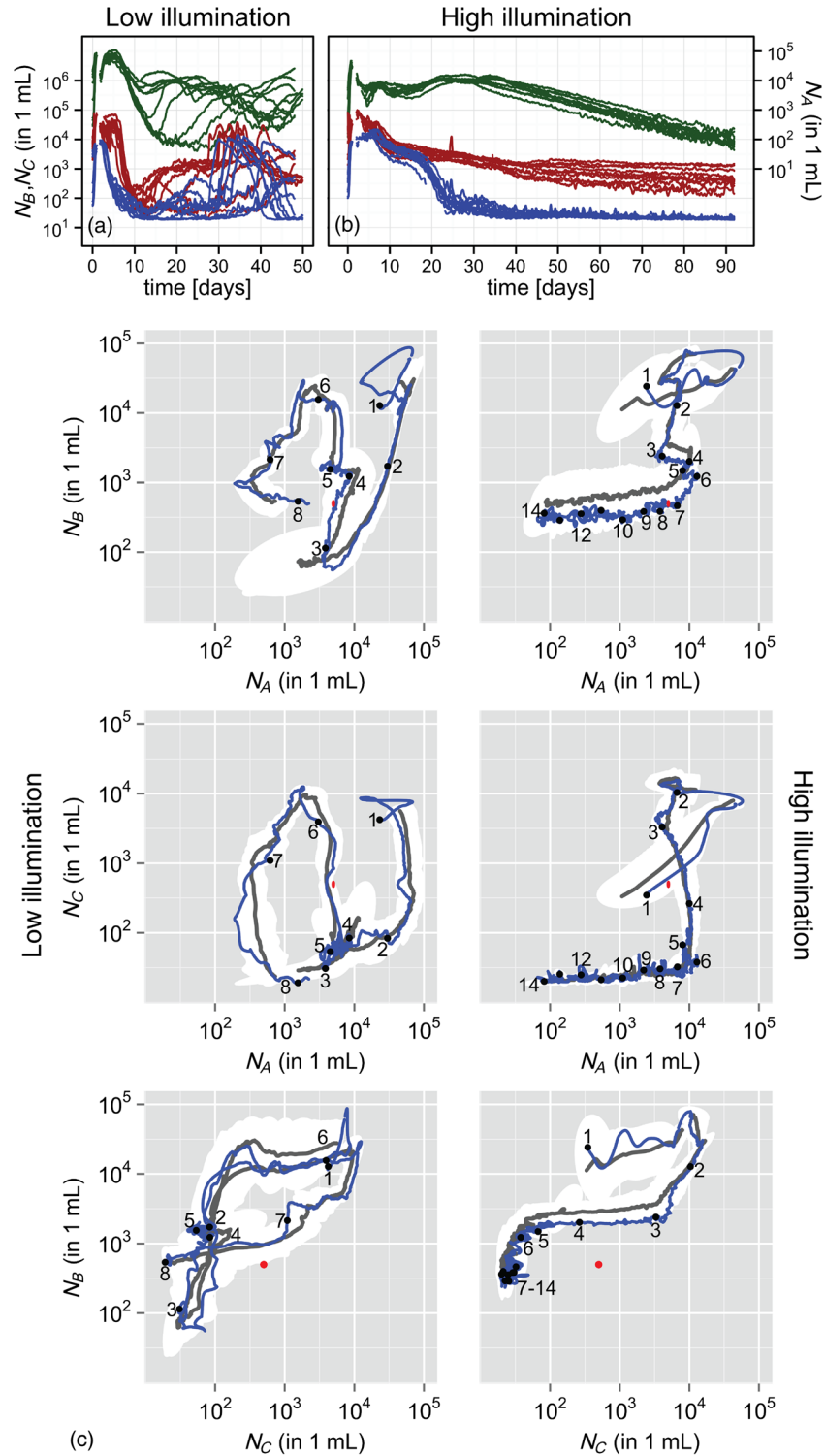


FIG. 2. Population time series and phase portraits at low and high illuminations. (a) Time series of population dynamics for 10 replicate ecosystems at $I = 0.25I_0$ illumination. Green (A—algae, right axis): *C. reinhardtii*; red (B—bacteria): *E. coli*; blue (C—ciliate): *T. thermophila*. (b) Time series for 10 replicates at $I = I_0$, as in (a). (c) Population dynamics are plotted as trajectories in the space of one species abundance versus another, for two illumination conditions: left column, $I = 0.25I_0$ [low illumination, 10 replicates, Figs. 1(e) and 2(a)]; right column, $I = I_0$ [high illumination, 10 replicates, Figs. 1(d) and 2(b)]; top row, *E. coli* vs *C. reinhardtii*; center row, *T. thermophila* vs *C. reinhardtii*; bottom row, *E. coli* vs *T. thermophila*. Gray curve, geometric mean trajectory over replicates; white region, 68% confidence interval of logarithmic abundance over replicates; blue curve, trajectory for a single replicate; red ellipse, 95% confidence interval for initial inoculation abundance; numbered points, timing along trajectory in weeks, starting at the beginning of week 1. The typical time between inoculation and the first labeled point was 20 h.

bacteria, and 1.1-fold for ciliates (these values likely overestimate the intrinsic variations since they also include variations due to differences among positions of the imaged volume within the replicates—see Appendix A 6). Thus, based on the time series in Fig. 1(d), it appears that intrinsic population dynamics in CES were strongly deterministic. This implies that the increasing temporal fluctuations between replicate systems that we observed previously [Fig. 1(c)] were not driven by intrinsic dynamics but rather by extrinsic variations in temperature and illumination.

The illuminance for the CES in Fig. 1(d), $I_0 = 1200$ lx, was constant in time and chosen to approximately match that of our past experiments [Fig. 1(c)]. To explore whether the fluctuations around the mean dynamics always stayed small, we performed another experiment, measuring population dynamics in ten replicate CES with one-quarter the illuminance ($I = 0.25I_0$). The resulting time series are shown in Fig. 1(e), where we can observe both a qualitative change in the geometric means and a substantial increase in the geometric standard deviations relative to the dynamics observed at I_0 . Comparing Figs. 1(d) and 1(e) seems to suggest that reducing the illumination might qualitatively alter the role of random events in the dynamics of the system. In order to scrutinize this observation, we examine the individual time series for all ten replicate ecosystems in the $0.25I_0$ illumination condition [Fig. 2(a)]. It appears that the structure of the dynamics in this condition is very similar across replicates, but there exists variability in the timing of various features characterizing those dynamics. For example, a rapid and nearly 100-fold increase in the *T. thermophila*

abundance is observed in 7 out of 10 replicates, occurring between 29 and 41 days after the start of the experiment. This variability in the timing of strong changes in the abundances leads to large values of the geometric standard deviations compared to the ones observed at $I = I_0$, where individual time series do not exhibit comparable temporal features and are very similar to each other [Fig. 2(b)].

To further quantify this similarity, we construct phase portraits [Fig. 2(c)], i.e., $N_s(t)$ plotted against one another, for $I = 0.25I_0$ and $I = I_0$, and quantify the variability across replicate ecosystems in this space (see Appendix B 5). We find that the replicate-to-replicate variations in the phase space of abundances are similar and small in both illumination conditions [Fig. 2(c), white regions]. Remarkably, in the lower illumination condition ($I = 0.25I_0$), where variability across replicates in time series was large [Fig. 1(e)], replicate CES follow very similar trajectories in the phase portraits.

Interestingly, in the $I = 0.25I_0$ condition, we find that the timing of the above-mentioned rapid increase in $N_C(t)$ is strongly correlated with the time of the phenotypic switching between small and large cells of *T. thermophila* that took place in the CES approximately three weeks prior (Fig. 3). We measured this correlation by defining two quantities: τ_p , the time since the beginning of the experiment when the peak in $N_C(t)$ is observed, and τ_m , the time (after an initial decline) when 10% of the *T. thermophila* population is classified as large using a statistical model (Appendix B 6, Fig. 11). We find that τ_m is approximately

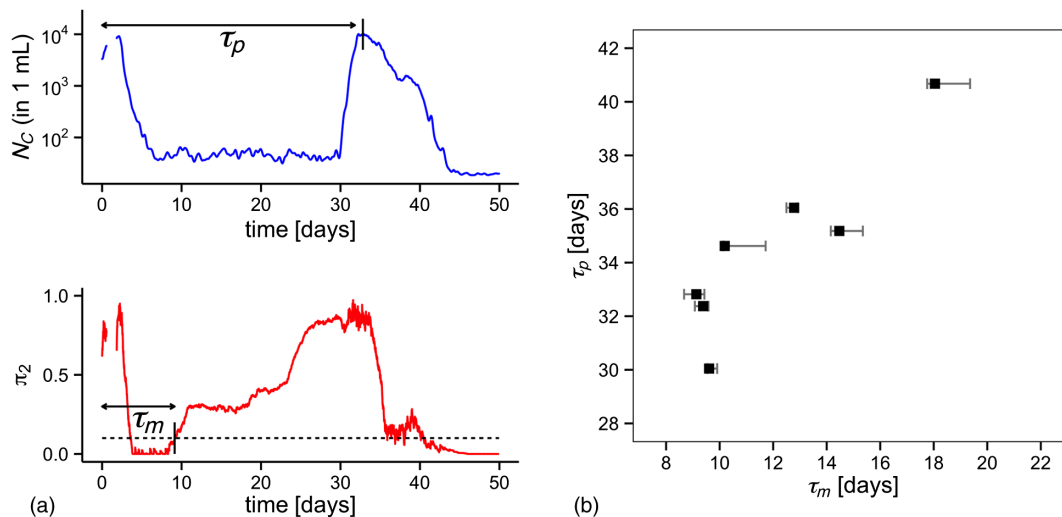


FIG. 3. Morphological dynamics at low illumination. (a) An example of morphological dynamics from a single replicate at low illumination ($I = 0.25I_0$). Upper panel: $N_C(t)$. Lower panel: The fraction of cells at each time point that are classified as large (π_2 , see Appendix B 6). The times τ_p and τ_m are defined as shown by the arrows (see text for details). (b) Scatter plot of τ_p versus τ_m as defined in (a) for 7 CES in the $I = 0.25I_0$ condition. τ_p can only be defined for seven of ten replicates in which the peak in $N_C(t)$ was observed. For the remaining three replicates, the large values of τ_m (not shown) suggest that the peak in $N_C(t)$ would have occurred at times after data acquisition terminated. Error bars in τ_m were computed by varying the threshold applied to π_2 by $\pm 25\%$. The correlation coefficient between τ_m and τ_p is 0.90 ($p = 0.006$).

20 days shorter than τ_p , and the two quantities are strongly correlated [Fig. 3(b)]. In this sense, the apparently stochastic dynamics in each CES are strongly deterministic when conditioned on the timing of the reappearance of the large ciliates, which is variable across replicates.

B. Spatial distributions

We find that low variability across replicate CES extends to the spatial distributions of organisms as well. In Fig. 4, we show these distributions, in the $I = I_0$ illumination condition, for all three species. We observe an asymmetric distribution of algae, potentially driven by a phototactic

response to asymmetry in the illumination due to the gas bubble present in each CES. Bacteria exhibit spatial structure consistent with the pattern of thermal convection that is known to be present in each cuvette (while the temperature is constant in time, gradients have not been suppressed). For the ciliates, roughly half of the population is observed in close proximity to the cuvette walls, where cells move along the boundary. We observe time dependence in all of these spatial structures. A full analysis of this rich set of spatiotemporal data lies beyond the scope of this paper. However, our data already reveal a surprising result: Despite the many stochastic processes taking place in each CES, replicate-to-replicate variation in the time-averaged

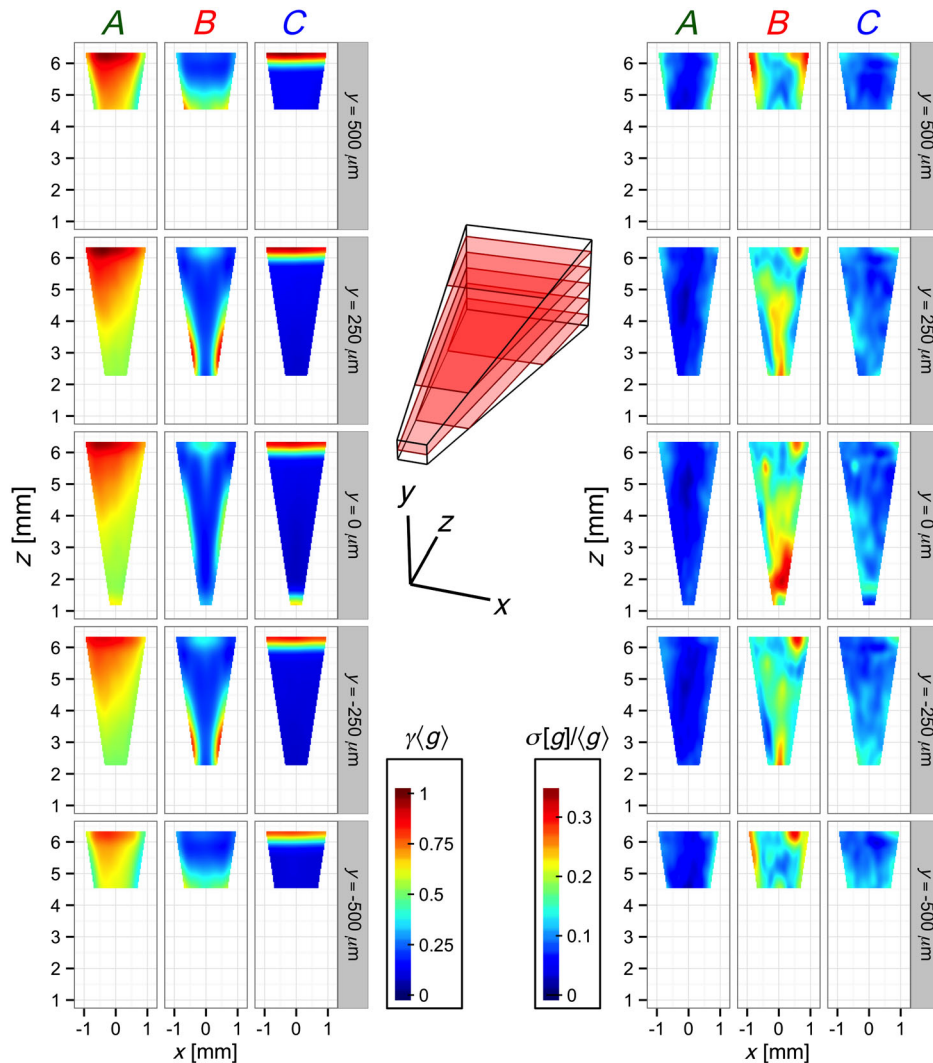


FIG. 4. Spatial distributions of microbes in the imaging volume. Using data from the ten replicates in the $I = I_0$ illumination condition [Figs. 1(d) and 2(b)], the time-averaged spatial density $g_s^{(i)}(x, y, z)$ of each species s and replicate i was calculated as described in Appendix B 4. The mean $\langle g_s(x, y, z) \rangle$ and coefficient of variation, $\sigma[g_s(x, y, z)]/\langle g_s(x, y, z) \rangle$, of the time-averaged spatial density were calculated over replicates. These three-dimensional statistics of spatial density were sliced along five planes of constant y , as shown by the transparent red slices in the central schematic view of the monitored volume [see Fig. 1(a)], and plotted as heat maps within each slice (left panel, normalized mean; right panel, coefficient of variation). Within each panel, species are sorted by column [left column, *C. reinhardtii* (A); center column, *E. coli* (B); right column, *T. thermophila* (C)]. The normalized mean $\gamma_s \langle g_s(x, y, z) \rangle$ is shown, where $\gamma_s^{-1} = \max_{x,y,z} \langle g_s(x, y, z) \rangle$.

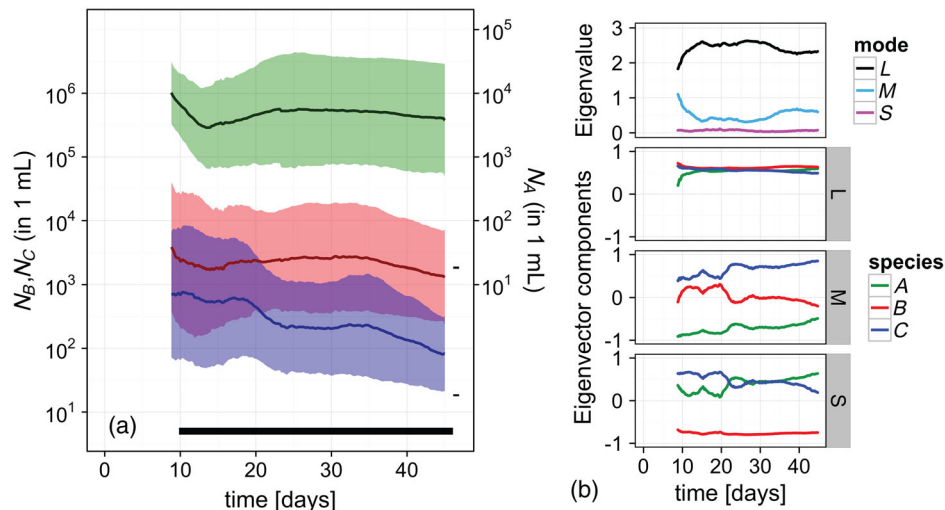


FIG. 5. Structure of fluctuations across ecosystems in a range of conditions. (a) Population dynamics, as in Fig. 1(d), for 63 replicate ecosystems maintained at 7 illumination conditions between $I = 0.2I_0$ and $I = I_0$. The geometric mean (bold curves) and standard deviation (shaded regions) are computed across all 63 replicates and 7 illumination conditions. Data before 9 days are missing in some conditions because of overly high cell abundance. (b) Dynamics of the correlation matrix: The correlation matrix $C(t)$ of logarithmic abundances of the three species was calculated for the data shown in (a). The eigenvalues of $C(t)$ were sorted by size and are plotted as time series in the top panel: L (large, black line); M (medium, cyan line); S (small, magenta line). The composition of the corresponding eigenvectors is plotted in the bottom three panels. Green: *C. reinhardtii* (A), red: *E. coli* (B), blue: *T. thermophila* (C).

spatial distributions of microbes is very low (Fig. 4, right panel); indeed, the coefficient of variation is less than 10% for algae and ciliates, and less than 30% for bacteria. We observe similar variability for the $I = 0.25I_0$ condition (Appendix B 4, Fig. 10).

C. Variation of external conditions excites fluctuations along ecomodes

From the analysis of time series of abundances, morphological variation, and spatial distributions, we thus conclude that CES maintained under tight environmental control exhibit strongly deterministic dynamics. It is tempting to presume that the fluctuations observed in previous experiments [Fig. 1(c)] were driven by variations in extrinsic parameters between replicate CES. This explanation assumes that the nature of the interactions inside the CES, leading to the intrinsic dynamics observed here, was not substantially modified in the new experiments. To test this assumption, we studied the nature of correlations between the abundances of the three species and compared them with the ecomodes observed in the previous studies. Fluctuations across replicates within a single illumination condition were typically small and were not stably correlated (Fig. 12). Therefore, to test whether we could excite fluctuations along ecomodes by varying the extrinsic parameters of our CES, we measured population dynamics for 45 days in 8 to 10 CES in each of 7 illumination conditions ranging from $0.2I_0$ to I_0 . Figure 5(a) depicts the dynamics of the geometric mean and standard deviation across all 63 replicates and 7 conditions. It is clear

that varying the illuminance by a factor of 5 induces substantial variation across replicate ecosystems [compare to Fig. 1(d)]. We compute the eigensystem of the correlation matrix $C(t)$, also calculated over 63 replicates and 7 conditions, and the results are shown in Fig. 5(b). Remarkably, the ecomodes in Fig. 5(b) are nearly identical to those we found in the previous experiments. For instance, the ecomode corresponding to the largest eigenvalue of $C(t)$ (L) is identical to that from the previous experiments, and it corresponds to the abundances of all three microbial species fluctuating coherently about the mean. The ranking of the two lower modes is reversed in comparison with the previous experiments [13], but their composition is unchanged. This result strongly supports our hypothesis that the intrinsic dynamics in both experiments are very similar, and it supports the claim that the fluctuations across replicates observed previously, and the structure of these fluctuations, were driven by variation in illumination and possibly other extrinsic variables such as temperature.

III. DISCUSSION

We are thus facing a rather astonishing situation: On microscopic scales, we observe many stochastic processes taking place in each CES. However, on more macroscopic scales, replicate-to-replicate variations in the measured abundances remain small over periods of months. Strongly deterministic dynamics have been observed previously in response to what one could call a “strong perturbation.” For instance, a drastic change of the

environment (e.g., large increase of temperature, salinity, or antibiotic concentration) or of the internal composition of an organism (e.g., deletion of a crucial gene) can induce a quasideterministic change of phenotypes or of species abundances [23]. The CES in our experiments are not, however, subject to such strong perturbations, and we observe long-term strongly deterministic dynamics, rather than rapid transformations.

From the point of view of biology, this situation is reminiscent of the developmental dynamics of a single organism, which can be buffered against perturbations and appear quasideterministic, or subject to “canalization” [24] (also called homeorhesis [25]). In developmental canalization [24], the dynamics and the outcomes are observed to be highly reproducible despite many sources of noise. The analogy of highly deterministic population dynamics observed here with developmental canalization cannot, however, be directly applied. Our closed ecosystems were created with three laboratory strains of microbes, which have lived in isolation for many years in the laboratory, and originated from complex ecosystems with a multitude of components. It is, therefore, difficult to argue that natural selection evolved buffering mechanisms [26] to reduce variation in the population dynamics exhibited by our particular synthetic ecosystems.

Under selection, microbial populations can undergo rapid genetic change [27,28]. The extent of genetic diversification in our ecosystems has not been determined. However, the strongly deterministic dynamics that we observe are unlikely to be wholly due to selection, for at least three reasons. First, the population size is small, at least in the bulk—less than 1×10^5 for algae, less than 1×10^6 for bacteria, and less than 1×10^4 for ciliates. Second, the number of generations during the experiment is likely small. Since our DIH microscopes image an open subvolume within each cuvette [Fig. 1(a)], the abundances we measure arise from division and death as well as migration; thus, we cannot directly resolve the number of divisions. However, in the $I = I_0$ condition, over the first 2 to 3 days, all three species experience rapid growth from their initial abundances [Fig. 1(d)]. From the maximum abundance reached during this period of growth, we can estimate the number of generations that occur for each species, assuming that birth dominates migration during this period. We find that the number of generations is at least 3 for algae, at least 7 for bacteria, and at least 5 for ciliates. We can place a loose upper bound on the number of generations that occurred during the entire experiment by calculating how many generations would transpire if the growth rate inferred early in the experiment were maintained for the full 90 days. From this, we approximate the maximum number of generations to be 160, 230, and 70 for algae, bacteria, and ciliates, respectively. However, given that nutrients are not supplied during the experiment, we expect the actual number of generations to be well below

this estimate. For comparison, similarly sized populations of yeast propagated under rapid growth had very few detectable mutations after 100 generations [27]. Third, the deterministic nature of the dynamics we observe is difficult to reconcile with the random generation of genetic variability through mutation.

From a more general point of view, the observation of deterministic effective dynamics in the presence of noise is reminiscent of equilibrium thermodynamics and hydrodynamics. In both cases, stochastic phenomena taking place on the microscopic level contribute to often-deterministic collective behavior on the macroscopic level, described by simple laws, which use coarse-grained state variables, such as densities or fluxes. The results of our experiments are a strong indicator that complex stochastic phenomena taking place on the microscopic level in a CES could also be coarse grained on more macroscopic scales, where the relevant degrees of freedom are few. What are the collective variables best suited to describe our CES? At present, we cannot provide a definitive answer to this question. However, two points are worth noticing: (i) The trajectories of abundances are strongly deterministic and do not significantly self-intersect, suggesting that the dynamics can be usefully coarse grained to this level; (ii) the first ecomode dominates the externally driven fluctuations, suggesting further dimensional reduction of the space of collective variables. Of course, our CES are much more complex than systems described by thermodynamics or hydrodynamics. In view of many levels of organization, with interdependent components, we cannot invoke simple statistical laws, e.g., the law of large numbers, in order to directly explain the observed strongly deterministic dynamics. The dynamics in thermodynamic or hydrodynamic systems is often driven by temporal or spatial gradients in physical quantities, such as temperature, pressure, or gravitational potential. At present, we do not know what quantities drive dynamics in our CES, or what the appropriate state variables for ecosystems are. However, based on our results, it seems that the view that contingency always dominates long-term biological dynamics, imposing a historical narrative as the appropriate method of study [5], might be too pessimistic.

We propose that coarse-grained regularities, such as the strongly deterministic dynamics and the stable ecomodes observed in this study, might originate in general biochemical [29] and physical constraints, common to many, if not all, microbial ecosystems. They can be explored further by systematically measuring the effects of additional perturbations on population dynamics. Such an approach would be facilitated by the strongly deterministic nature of the dynamics around which one would apply perturbations. Experiments may include both external perturbations, such as time-dependent modulations of temperature or illumination, and internal perturbations, such as modifications of species composition, genomes, gene expression profiles, or

epigenetic markers. In view of the simplicity and low cost of the apparatus used here [17], such systematic studies could, in principle, be performed in hundreds of replicate systems in a wide range of conditions. Beyond explaining the behavior of these particular CES, it is our hope that these studies will establish the basis for a more general, quantitative theory of ecological dynamics.

IV. MATERIALS AND METHODS

A. Media and strains

Three species (“ABC”) microbial communities [13] were constructed using the alga *Chlamydomonas reinhardtii* (A: strain UTEX 2244, mating type +), the bacterium *Escherichia coli* [B: strain MG1655 $\Delta flu \Delta fimA$ HK022 att::(cat P_{AR} dTomato) hsdR], and the ciliate *Tetrahymena thermophila* (C: CU428 mating type VI, Cornell University Culture Collection).

For each experiment, each species was grown axenically as follows. Algae were grown in Sager-Granick medium [30] with Hutner’s trace elements (purchased from [31]) substituted for the standard trace elements. Cells were thawed from a frozen stock and diluted into Sager-Granick media and maintained at 200 rpm shaking, 25°C, and 7500 lx in an incubator. Cultures were grown 7 to 9 days prior to initiating ecosystems resulting in densities of $\approx 1 \times 10^6$ cells/mL. Bacteria were grown from glycerol frozen stocks in the same media used to construct the community, at 30°C with 200 rpm shaking resulting in densities of $\approx 1 \times 10^7$ cells/mL. Ciliates were grown in SPP rich media at room temperature without shaking for approximately 2 days, resulting in densities of 1×10^5 cells/mL. Cultures were initiated from a single long-term soybean stock that was replenished every 6 months [32]. ABC ecosystems were constructed using 1/2x Taub #36 supplemented with 0.03% w/v proteose peptone No. 3 (BD) [33] and a Tris-acetate pH buffer (20 mM Tris, 3 mM acetate).

B. Initiating closed ABC ecosystems

To initiate replicate ecosystems, an axenic culture of each of the three species was grown. Each culture was washed twice in the media used to construct the community. For A and C, the densities of the washed cultures were measured by flow cytometry (BD Biosciences LSR II) using Spherotech AccuCount Fluorescent Particles (Spherotech, ACFP-50-5). For B, culture density after washing was measured by optical density. Three species were then combined into a single culture at densities of 5000 cells/mL, 500 cells/mL, and 500 cells/mL for A, B, and C, respectively. Replicate ecosystems were initiated in a biosafety cabinet by pipetting 1.6 mL of this mixture into sterile and clean quartz cuvettes (StarnaCell, 23-Q-5). Cuvettes were sealed with a sterile, greased (Apiezon M) teflon stopper and epoxy designed to form a hermetic

seal with glass (Epoxy Technology, 905-T). The epoxy was cured for ≈ 20 h upright, without shaking, in an incubator at 25°C with 7500 lx illumination.

Prior to initiating ecosystems, quartz cuvettes were cleaned thoroughly by scrubbing with a detergent (StarnaCell, CellClean), rinsing with deionized (DI) water, sonicating with a mixture of acetone and isopropyl alcohol (IPA), and soaking for 2 h with 50% nitric acid. Finally, cuvettes were rinsed with DI water and IPA, dried under pressurized nitrogen gas, and capped with teflon stoppers. Cuvettes were sterilized by UV light from a transilluminator.

For each of the seven illumination conditions, we performed two independent experiments measuring dynamics in 3 to 6 replicate ecosystems. For each experiment, the entire initiation protocol described here was performed independently.

To test for contamination, we plated axenic cultures used to initiate the communities on four nutrient conditions (Luria-Burtani Broth, Nutrient Broth, Yeast-Peptide-Dextrose, and Thioglycollate broth) and terminated any experiment in which contamination was detected. In the experiments presented here, no contamination was detected by this method.

C. Ecosystem external conditions

As described previously [17], each microscope independently controls the temperature (temporal-fluctuation standard deviation 3 mK, standard deviation across replicates 20 mK) and illumination (standard deviation 2.5%) for each ecosystem. Illumination is supplied symmetrically from above and below for each ecosystem by two 1-W light-emitting diodes (LEDs, Lumileds, nominal color temperature 3000 K). For all experiments presented here, the temperature was held constant at 25°C. The illuminance was constant in time and symmetric from above and below, but this level of illuminance was varied between experiments. Illuminance is measured as a fraction of the maximum calibrated output of the LEDs (I_0 is the maximum output). The maximum output was estimated, using a calibrated photodetector, to be 1200 lx (I_0).

ACKNOWLEDGMENTS

We thank the following authors for discussions, comments, and suggestions: B. Chazelle, J-P. Eckmann, D. Hekstra, D. Huse, J. Lebowitz, A. Libchaber, L. Peliti, T. Plusty, and L-S. Young, as well as current and former members of our laboratory. For technical assistance, we thank J. Chuang, D. Jordan, S. Mazel, and Y. Liu. We thank the following for distribution of genetic constructs and strains: the laboratories of C. D. Allis, the Yale Coli Genetic Stock Center, the National BioResource Project (NIG, Japan), and the Culture Collection of Algae (UT Austin). This research has been partly supported by grants from the

Simons Foundation to S. L. through Rockefeller University (Grant No. 345430) and the Institute for Advanced Study (Grant No. 345801). S. K. acknowledges funding from the Helen Hay Whitney Foundation.

Z. F. and S. K. contributed equally to this work.

APPENDIX A: ADDITIONAL EXPERIMENTAL METHODS

1. Seal quality

To test the quality of the teflon stopper-grease-epoxy seal, each cuvette was weighed before and after each experiment. Any reduction in mass was attributed to water evaporation. Changes in mass of <1 mg over the duration of the experiment were below the resolution of the balance used. The mean evaporation rate was 0.098 mg d^{-1} .

2. Differences in conditions from previous experiments

The experiments presented here differed from the measurements of Hekstra and Leibler [13] in two important ways. First, we modified the medium by adding a Tris-acetate pH buffer. This increased pH buffering capacity and resulted in less dramatic changes in pH over the course of the experiment than were observed previously (see below). Second, the incubators used by Hekstra and Leibler for storage of CES during the experiment subjected those systems to $\approx 30\%$ variation in the level of illumination across replicates and greater thermal fluctuations over time (s.d. $\approx 50 \text{ mK}$).

We performed an experiment to test whether the pH buffering capacity of the medium could account for the qualitative changes in the dynamics observed in Figs. 1(c) and 1(d). We measured population dynamics by holography in ecosystems without a Tris-acetate buffer in the medium. The resulting dynamics are shown in Fig. 6. We observe strongly deterministic dynamics even when the buffer is removed from the medium. We conclude that this chemical difference between the experiments of Hekstra and Leibler and the measurements presented here cannot account for the deterministic dynamics we observe. Finally, we note that the large variability in illumination within the incubators used in previous work is consistent with our observation of ecomodes driven by variable illuminance (Fig. 5).

3. Viability and cryopreservation

To test for viability of each species at the end of the experiment, a small sample of each ecosystem was subcultured under selective conditions. For *A*, we used Sager-Granick media which contains no carbon source (other than carbon dioxide). For *B*, viability was confirmed by plating on Luria-Burtani broth. To test for the presence of *C*, we subcultured each ecosystem in fresh experiment medium and allowed for several days of growth. Subcultures were

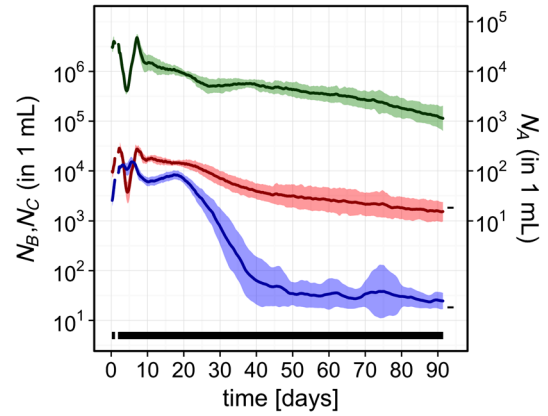


FIG. 6. Effect of pH buffer. Time series for nine CES measured by holography with the Tris-acetate pH buffer omitted from the medium. Geometric mean and standard deviation are shown as in Figs. 1(c)–(e).

then inspected by bright-field microscopy for the presence of *C*. Viability for all systems in all conditions studied here is shown in Fig. 7 (top four panels). For ecosystems in which *A* or *C* was viable, we cryopreserved a sample of the subculture using the appropriate protocol for each species [34].

4. pH

All experiments were initiated in medium at pH 7. At the end of all experiments, pH was measured by color metric pH paper to be 7.4. The observed increase in pH is consistent with the consumption of acetate by *A* over the course of the experiment. The precision of pH measurements was ≈ 0.2 . For experiments with the Tris-acetate buffer removed from the medium, the pH at the end of the experiment was measured to be 5.5.

5. Flow cytometry

At the end of each experiment, a well-mixed sample of each ecosystem was analyzed by flow cytometry (BD LSR II). In all flow cytometry measurements, *C* abundances were too low (1 mL^{-1} to 10 mL^{-1}) to be measured. *B* abundances (dTomato fluorescence) could not be reliably measured because of the presence of aggregates and a large background signal from chlorophyll of *A*. However, abundances of *A* were quantified at the end of each experiment. The results are shown in Fig. 7 (bottom panel).

6. Alignment

In the holographic microscopes, there is some systematic variation in the position of the $5.4 \mu\text{L}$ imaging volume within the cuvette. In particular, because of machining and alignment errors, the distance between the optical axis and the bottom face of the cuvette (*y* axis) is variable. We quantified this variation by imaging a small marker dot on

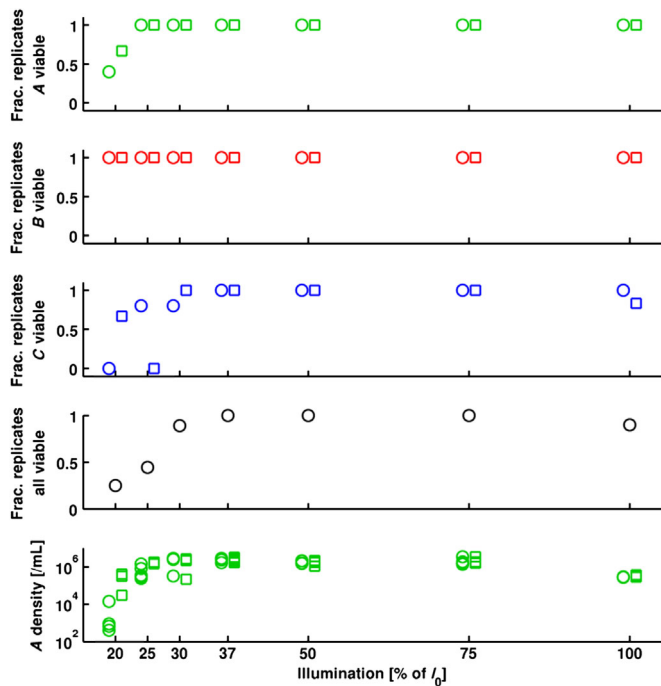


FIG. 7. Final viability and total algal density. Top four panels: Viability of each species, measured as described above, as a function of illumination. Green, red, and blue symbols show the fraction of systems with *A*, *B*, and *C* viable, respectively. Black squares show the fraction of replicates where all three species are viable. The two experiments in each illumination condition are represented by circle and square symbols. Bottom panel: Density of *C. reinhardtii*, as measured by flow cytometry, from a well-mixed sample of each replicate, at the end of each experiment as a function of illumination condition. A single point is plotted for each replicate with symbols denoting different experiments within each illumination condition.

the back face of each cuvette prior to removing it from the microscope at the end of the experiment. The position of the dot was then measured on a separate bright-field microscope with a large field of view. For the 48 ecosystems in which this measurement was performed, we found the distance to be $2560 \mu\text{m} \pm 430 \mu\text{m}$.

Because of the significant spatial heterogeneity we observed in the ecosystems, we investigated whether the variability in abundance over replicates could be explained by their vertical positions. Vertical positions were only measured for the six replicates from one repetition of the $I = I_0$ condition. For these replicates, we found that the relative abundance residuals $(N_s^i(t) - \langle N_s(t) \rangle) / \langle N_s(t) \rangle$ for replicate i and species s stabilized after day 27. We computed the time averages of these quantities from day 27 to day 92 for each of the six replicates in this data set and regressed them on the vertical position of the imaged region within that replicate, y^i . For both algae and bacteria (ciliates were not considered because of low abundance), the relative abundance residual and vertical position were negatively correlated [$\rho_A = -0.86$ ($p=0.014$), $\rho_B = -0.71$ ($p=0.06$)].

After adjusting for the influence of the vertical position on the abundances, the remaining fluctuations are significantly smaller ($\sigma_A = 1.26$ and $\sigma_B = 1.6$), suggesting that better control of the cuvette position should result in the measurement of even less variability across replicates.

APPENDIX B: DATA ANALYSIS

1. Iterative signal removal

The reconstruction of each hologram results in a three-dimensional array of complex amplitudes from within the sample. For our purposes, the phase information is not used, and only the squared modulus, or intensity, is of interest. The focal plane of each object is identified automatically [17].

A major difficulty for automated computer analysis is the low intensity of *E. coli* relative to *C. reinhardtii* and *T. thermophila*. The signal-to-noise ratio for the majority of bacterial cells is sufficient to separate them from the background; however, out-of-focus light from the brighter species complicates segmentation. This out-of-focus light can have significant spatial extent and can exhibit localized regions of intensity above the background. These discrete bright spots do not come into focus and can be distinguished from in-focus objects in three dimensions. However, applying global segmentation results in many false objects. An example of this effect is shown in Fig. 8.

To solve this problem, we exploit the computational nature of digital holography by removing the signal from bright objects before analyzing dimmer objects. By performing an initial segmentation at a high intensity threshold, all bright objects are identified, the focal plane of each bright object is determined, and the morphology is analyzed in that plane. The signal from the object can then be removed from all focal planes by setting all complex amplitudes within the object to zero. The result is propagated back to the hologram plane by inverting the reconstruction procedure, resulting in a hologram that is equivalent to the original with the signal from the bright object removed (Fig. 8). All objects above the intensity threshold are analyzed and removed, and then a second round of segmentation, analysis, and removal is performed at a lower intensity threshold. By using many thresholds, artifacts associated with objects having intensities near the threshold intensity can be avoided. In our implementation, 18 rounds of signal subtraction are used.

2. Species classification

The image analysis of each blob measures the spatial location, spatial moments of the two-dimensional intensity distribution in the focal plane, and statistics describing the number of voxels exceeding various intensity thresholds. A training set of 8967 manually classified objects was used to train a support vector machine [35] to classify objects into four classes (algae, bacteria, ciliates, and background).

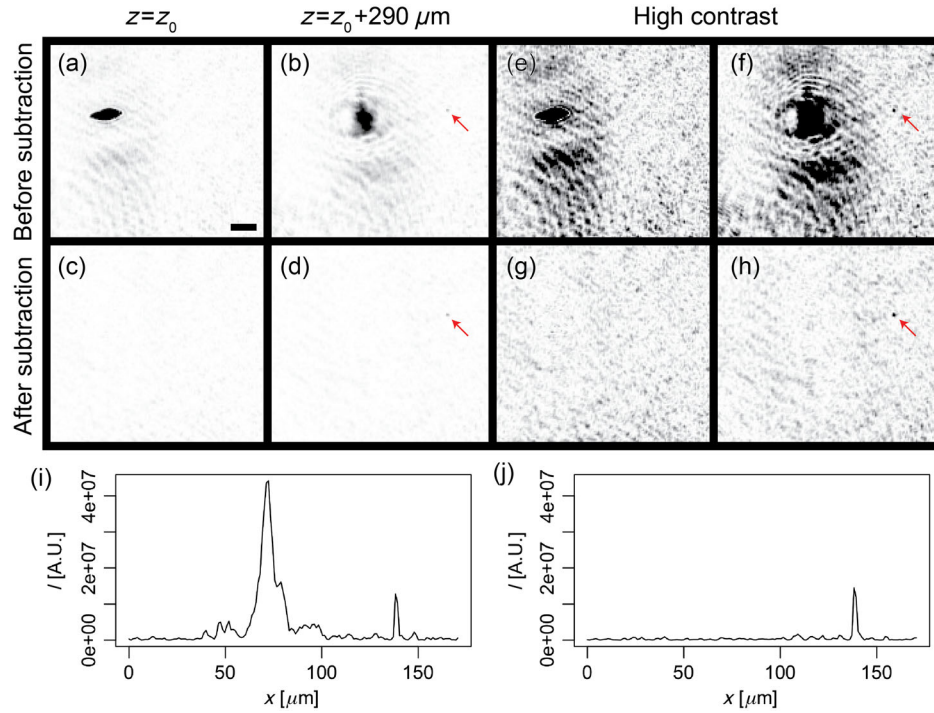


FIG. 8. Signal subtraction detail. (a) Focal plane of reconstructed *T. thermophila*, to the upper left. (b) Same *T. thermophila* from (a), 290 μm out of focus. An *E. coli*, indicated by a red arrow, is barely visible, to the right—see (f) and (h) for better visibility at high contrast. (c) Focal plane from (a), after the signal from the *T. thermophila* has been subtracted. (d) Focal plane from (b), after signal subtraction, showing unperturbed *E. coli*. (e)–(h) Duplicate panels (a)–(d) at high contrast. The scale bar is 20 μm . (i) Intensity profile across the horizontal coordinate in (b) and (f), through the *E. coli* cell, before subtraction. The large high-intensity region around 70 μm is the out-of-focus *T. thermophila*, while the peak at 140 μm is the in-focus *E. coli* cell. (j) Intensity profile through the same row of pixels as in (i), after signal subtraction—see (d) and (h).

A test set of 1825 manually classified objects was used to estimate the classification error rate—the results are tabulated in Table I; there were no detection errors for the test set.

3. Count statistics

Cells leave and enter the imaged region by a combination of swimming and convective flow. Our goal is to filter the noise due to sampling in order to infer an underlying abundance. We can roughly estimate the residence times of each species using measurements of the convective flow and diffusion coefficients for typical swimming behaviors; in each case, the estimated residence time is less than one minute, so we expect little correlation in sampling noise for measurements separated by 400 s.

TABLE I. Classification table.

		Manual classification			
		Algae	Bacteria	Ciliates	Background
SVM classification	Algae	812	19	0	0
	Bacteria	3	618	1	0
	Ciliates	2	0	282	0
	Background	0	1	0	87

We assume that the counts of each species are a temporally uncorrelated Poisson-distributed random variable $A(t)$, whose rate fluctuates in time. This rate is interpreted as a spatial density integrated over the imaging region, from which Poisson-distributed random counts are drawn because of discrete cells entering and leaving the region. The rate is unitless and corresponds to the expected value for the counts as a function of time. It is estimated independently for each species in each replicate. In our experiments, the underlying rate varies over time by at least 3 orders of magnitude, so the variance due to counting noise also varies by several orders of magnitude. In this case, it is clear that a stationary filter is inappropriate, and we specify a probabilistic model for the counts and underlying rate which has superior performance. The quantity of interest is the conditional probability of the underlying logarithmic rate given the counts:

$$\mathfrak{Q} = P(W|\{A(t_i)\}), \quad (\text{B1})$$

where W represents the logarithm of the rate over the time interval of the experiment, $[0, t_m]$, and t_i is the time of measurement i , $i = 1, 2, \dots, m$. Applying Bayes' rule gives

$$\mathfrak{Q} = P(\{A(t_i)\}|W)P(W)/P(A), \quad (\text{B2})$$

where $P(A)$ does not depend on W and so does not affect the estimation of W . From our assumption that $A(t)$ is Poisson and temporally uncorrelated,

$$\begin{aligned} P(\{A(t_i)\}|W) &= \prod_{i=1}^m P(A(t_i) = a_i | W(t_i) = w_i) \\ &= \prod_{i=1}^m \frac{e^{-w_i} e^{w_i a_i}}{a_i!}. \end{aligned} \quad (\text{B3})$$

Previous work suggested that the fluctuations of $W(t)$ across replicates can be well described as a random walk [13]. To derive an expression for $P(W)$ for our filter without imposing additional structure, we assume that $W(t)$ itself is a random walk, with Gaussian increments. Define $\Delta t_i = t_{i+1} - t_i$, and let γ be a parameter that determines the increment variance:

$$\begin{aligned} P(W) &= P(W(t_1) = w_1) \prod_{i=1}^{m-1} P(W_{i+1} = w_{i+1} | W_i = w_i) \\ &= P(W(t_1) = w_1) \prod_{i=1}^{m-1} (2\pi\gamma\Delta t_i)^{-1/2} \\ &\quad \times \exp(-(w_{i+1} - w_i)^2 / 2\gamma\Delta t_i). \end{aligned} \quad (\text{B4})$$

We can now write the portion of the log likelihood that depends on W :

$$\begin{aligned} \log(\mathfrak{Q}P(A)) &= \log P(W(t_1) = w_1) \\ &\quad + \sum_{i=1}^{m-1} \left[-\frac{1}{2} \log 2\pi\gamma\Delta t_i - \frac{(w_{i+1} - w_i)^2}{2\gamma\Delta t_i} - e^{w_i} \right. \\ &\quad \left. + a_i w_i - \log(a_i!) \right] \\ &\quad - e^{w_m} + a_m w_m - \log(a_m!). \end{aligned} \quad (\text{B5})$$

This expression is maximized over w_i by gradient ascent, using a moving average of the logarithm of measured counts as starting values. The parameter γ was chosen by testing the distribution of a_i against a Poisson distribution with rate e^{w_i} . The optimal value was $\gamma = 2.56 \times 10^{-2}$ /day. The distribution of the initial logarithmic rate was assumed to be Gaussian, with mean equal to the average of the logarithm of the first few measurements:

$$P(W(t_1) = w_1) = (2\pi\sigma_0^2)^{-1/2} \exp[-(w_1 - \mu_0)^2 / 2\sigma_0^2], \quad (\text{B6})$$

where $\mu_0 = \frac{1}{n} \sum_{i=1}^n \log a_i$, $\sigma_0 = \mu_0^{-1/2}$ and $n = 5$.

We found that this filter performed very well. The distributions of counts were indistinguishable from Poisson with the inferred rate, and the autocorrelation of the residuals $\delta_i = \log a_i - w_i$ decayed very quickly (Fig. 9), justifying the assumption of independence of consecutive measurements.

4. Spatial density estimation

The analysis of our experiments yields a list of identified organisms, with a location in space, which we denote as $r_{j,s}^{(i)} = (x_j, y_j, z_j)$, where i denotes the experimental replicate, s denotes the species, and j indexes the individual organisms of species s in replicate i . The simplest

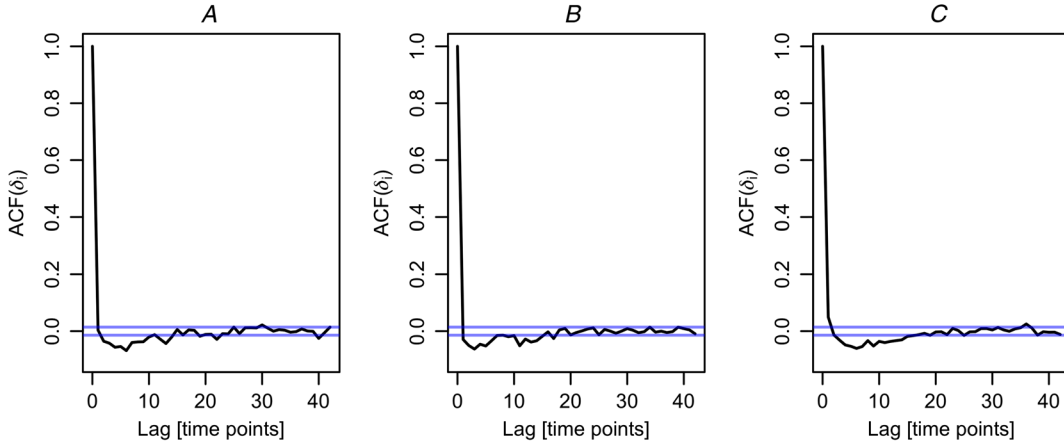


FIG. 9. Autocorrelation of count-smoothing residuals. The autocorrelation function of the residual from the count-smoothing procedure, $\delta_i = \log a_i - w_i$, was calculated for one of the replicates and each species, with lag measured in time points (400 s). The blue line shows the 95% confidence interval. The rapid decay observed here was typical for all replicates.

characterization of spatial structure is the time-averaged spatial density:

$$\Pr(r_s^{(i)} \in V) = \int_V g_s^{(i)}(x, y, z) dx dy dz, \quad (\text{B7})$$

where V is a region in space.

We estimated the spatial density $g_s^{(i)}(x, y, z)$ for each replicate and species via kernel density estimation [36], using cross validation to choose bandwidths. The optimal bandwidths in x , y , and z were $82 \mu\text{m}$, $68 \mu\text{m}$, $204 \mu\text{m}$ for algae; $88 \mu\text{m}$, $72 \mu\text{m}$, $225 \mu\text{m}$ for bacteria; and $107 \mu\text{m}$, $88 \mu\text{m}$, $246 \mu\text{m}$ for ciliates.

Our data are sharply truncated in space, by optical constraints in x and y , and by the cuvette walls in z . To reduce bias in the estimate due to truncations, we supplemented the bounded data set with its reflections across the boundaries [37]. The cuvette wall positions

were determined manually by finding a region near each z extremum where the density of blobs fell close to zero. We estimated the precision of this procedure to be $\pm 5 \mu\text{m}$. The wall locations varied by a small amount, which was accounted for by linearly transforming the data from each replicate along the z axis to coincide with a standard reference.

The confidence interval of the estimate can be calculated by bootstrapping [38]. We calculated 95% confidence intervals with at least 70% coverage, denoted $b_s^{(i)}$, and considered the ratio $b_s^{(i)}/g_s^{(i)}$, averaged over space, to indicate the precision of the spatial density estimates. Typical values over space and replicates were 6% for algae, 10% for bacteria, and 8% for ciliates.

Spatial distributions for all three species are shown in Fig. 4 for the $I = I_0$ condition and Fig. 10 for the $I = 0.25I_0$ condition. We note that the apparently large $\sigma[g]/\langle g \rangle$ for *E. coli* and *T. thermophila* in the $I = 0.25I_0$ condition is

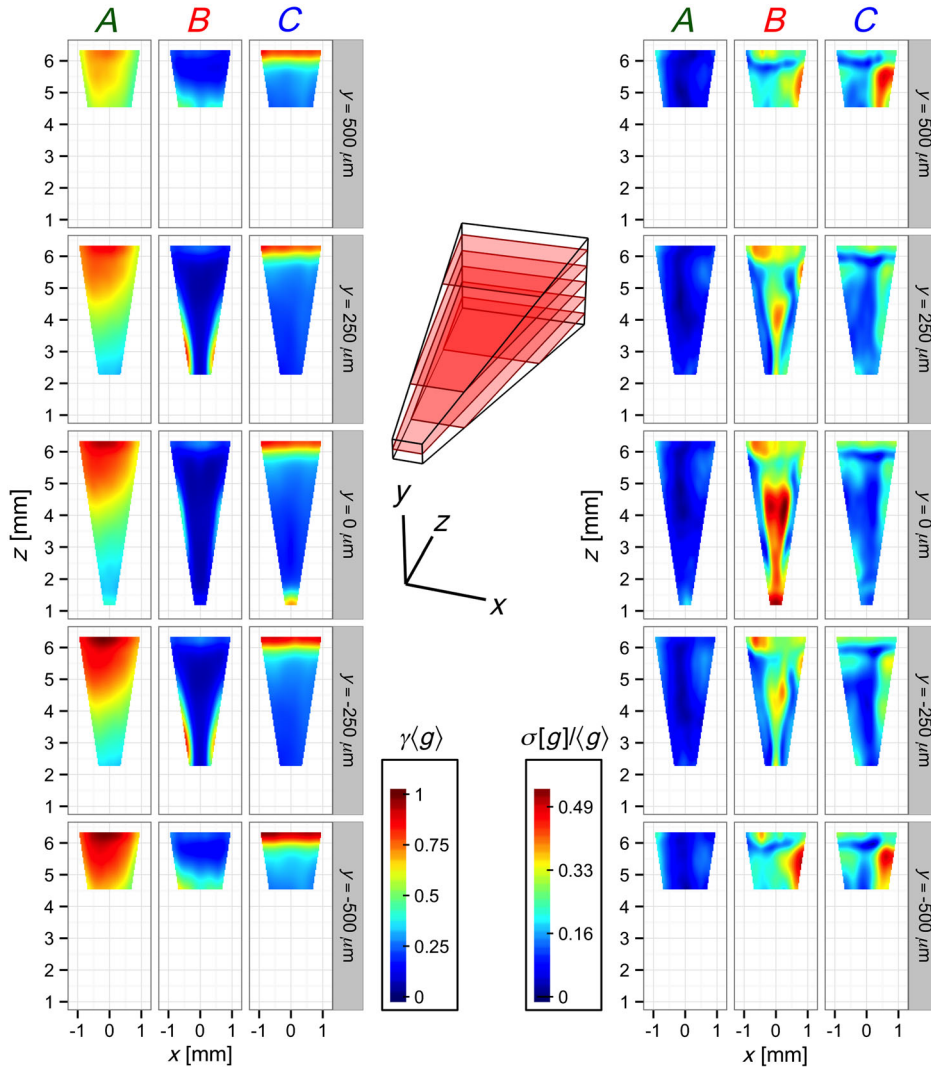


FIG. 10. $I = 0.25I_0$ spatial distributions of microbes in imaging volume. $\gamma(g)$ and $\sigma[g]/\langle g \rangle$ are defined in the caption to Fig. 4 for the $I = I_0$ condition.

due to statistical variation in $\langle g \rangle$ arising from low abundances rather than variation across replicates.

5. Phase portrait statistics

To quantify the stereotypical nature of the trajectories in the $I = 0.25I_0$ and $I = I_0$ illumination conditions, we applied a temporal alignment to the time series of logarithmic densities. We denote the smoothed logarithmic abundance of species s and replicate i by $u_s^{(i)}(t) = w_s^{(i)}(t) - \log(V^{(i)}/1 \text{ mL})$, where $V^{(i)}$ is the volume of the imaged region for replicate i , in units of mL. Note that the abundance $N_s^{(i)}(t) = \exp(u_s^{(i)}(t))$. The goal is to find nondecreasing functions τ_{ij} and τ'_{ij} such that the integrated squared difference upon alignment is minimized:

$$\tau_{ij}, \tau'_{ij} = \arg \min_{\tau, \tau'} \int_0^1 \sum_s (u_s^{(i)}(\tau(l)) - u_s^{(j)}(\tau'(l)))^2 dl, \quad (\text{B8})$$

where \mathfrak{Z} is the space of nondecreasing functions from $[0,1]$ to $[0, T]$, T is the experiment duration, and l parametrizes the alignment functions. The functions τ and τ' can be calculated on a discrete grid by dynamic programming.

Statistics over the aligned trajectories are calculated as follows. First, one of the replicates is chosen as the reference trajectory—this replicate is denoted with the

index i_0 . Aligned versions of each replicate are defined as $\tilde{u}_s^{(j)}(t) = u_s^{(j)}(\tau'_{i_0 j}(\tau_{i_0 j}^{-1}(t)))$ and statistics of aligned trajectories can be calculated for $\tilde{u}^{(j)}(t)$ over replicates j .

The mean of $\tilde{u}_s^{(j)}(t)$ over replicates is calculated and transformed to represent effective population abundance:

$$\tilde{\mu}_s(t) = \exp\left(\frac{1}{n} \sum_{j=1}^n \tilde{u}_s^{(j)}(t)\right). \quad (\text{B9})$$

In Fig. 2(c), the mean aligned densities $\tilde{\mu}_s(t)$ in the $I = I_0$ and $I = 0.25I_0$ illumination conditions are plotted as gray curves, and the covariance matrix for these same data sets is used to compute the 68% confidence interval, plotted as white ellipses. For the $I = 0.25I_0$ data set, it was necessary to adjust the temporal range of alignment determined by T for three replicates that did not exhibit the rapid increase in *T. thermophila* abundance around day 28–42.

6. Morphological dynamics

As discussed above, analysis of the reconstructions yields blobs classified as one of the three microbial species as well as each blob's morphological properties. The morphological dynamics are most significant for species

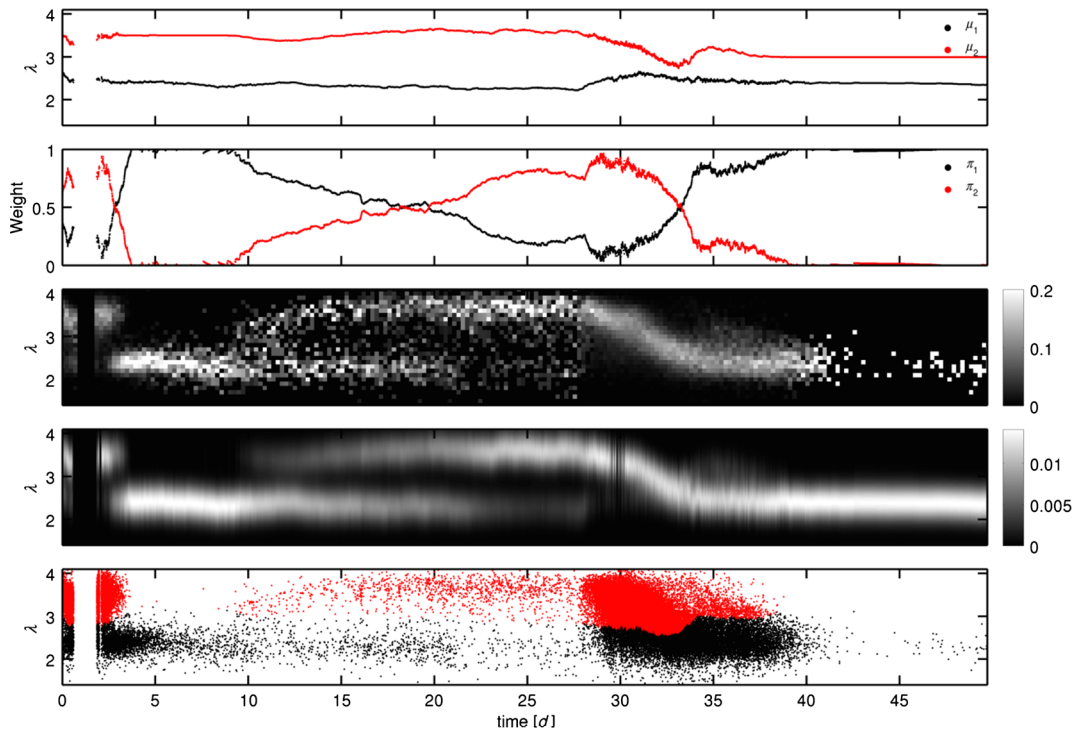


FIG. 11. Gaussian mixture model fit for one replicate in the $I = 0.25I_0$ condition. The upper panel shows the means of the two modes (μ_1 and μ_2), the second panel shows the weights of these modes in the mixture distribution (π_1 and π_2 , respectively), the third panel shows a histogram of the raw data, and the fourth panel shows the inferred probability density at each time point. In both the third and fourth panels, each column (time point) is normalized independently. The bottom panel plots the classification with large cells in red and small cells in black; note that there is no overlap between these classes.

C , so we focus our characterization of the morphological dynamics on this species only.

In particular, we consider the lateral size of each blob, denoted S_j^i for individual j in replicate i , measured in μm . The distribution of lateral size over time and replicates is approximately log-normal, so we consider the logarithm: $\lambda_j^i = \log(S_j^i/1 \mu\text{m})$. Each logarithmic size is associated with the time of measurement, t_j^i , and so we write $\lambda^i(t)$ to indicate time dependence.

We found that distributions of $\lambda^i(t)$ in many temporal windows were bimodal, with a strongly time-dependent structure (Fig. 11). We therefore parametrized the time-dependent distribution of logarithmic size using a Gaussian mixture model (GMM) with two modes. These distributions are estimated independently for each replicate, so we suppress the replicate index i :

$$P(\lambda(t)|\Theta(t)) = \pi_1(t)\phi_1(\lambda(t); \theta_1(t)) + \pi_2(t)\phi_2(\lambda(t); \theta_2(t)), \quad (\text{B10})$$

where $\phi_m(\lambda; \theta_m)$ is Gaussian with $\theta_m = \{\mu_m, \sigma_m^2\}$ and $\Theta = \{\pi_1, \pi_2, \theta_1, \theta_2\}$. We regularize the model for smoothness between time points by maximizing the following objective function:

$$\begin{aligned} O(\Theta(t_i), \Theta(t_{i-1}), \lambda(t_i)) \\ = l(\Theta(t_i); \lambda(t_i)) - \eta_1 \sum_{m=1}^2 (\mu_m(t_i) - \mu_m(t_{i-1}))^2 \\ - \eta_2 (\pi_1(t_i) - \pi_1(t_{i-1}))^2, \end{aligned} \quad (\text{B11})$$

where

$$l(\Theta(t_i); \lambda(t_i)) = \log \left(\prod_k P(\lambda(t_i) = \lambda_k | \Theta(t_i)) \right) \quad (\text{B12})$$

is the log-likelihood of the data observed at time point t_i given the model parameters $\Theta(t_i)$, and the product over k is taken over all cells measured at time t_i . This maximization was performed using a standard nonlinear, constrained optimization routine (Matlab, Mathworks). The η_1 and η_2 set the degree of smoothness between time points and were chosen by cross validation to be $\eta_1 = 450$ and $\eta_2 = 450$. The size variance was set equal for both modes ($\sigma_m^2 = \sigma^2$) for all time points and all replicates. The value $\sigma^2 = 0.08$ was chosen by examining previous GMM fits where σ_m^2 was observed to be stable across time, systems, and modes. Thus, our description of the ciliate morphologies consists of three parameters per time point per system, while capturing the salient features of the morphological dynamics.

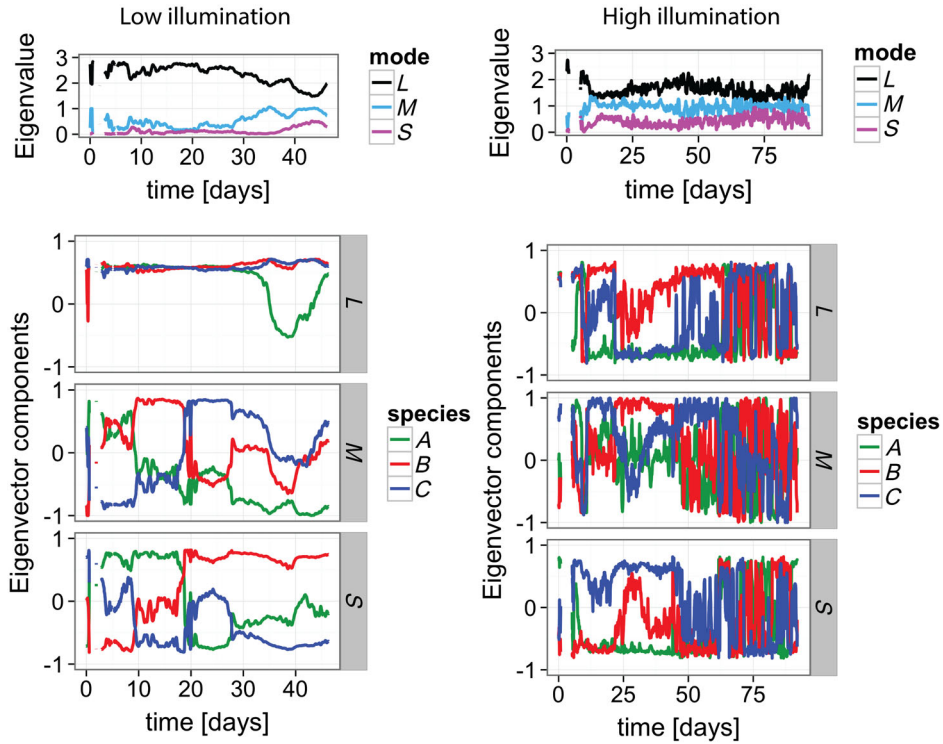


FIG. 12. Ecomodes in individual environmental conditions. Eigensystems of the correlation matrix (as in Fig. 4) for 10 CES in the low-illumination condition ($I = 0.25I_0$, left column) and the high-illumination condition ($I = I_0$, right column). Eigenvalues are plotted in the upper panels, and eigenvector composition in the lower panels [as in Figs. 5(a) and 5(b), respectively].

$\Theta(t_{i-1})$ is not always available because of gaps in the data, which vary in duration from a single acquisition to many days. For small gaps in the data, we wish to maintain the smoothness between nearby time points, while for large gaps, the smoothness penalty, determined by η_1 and η_2 , should be much smaller to permit large changes in the model parameters. Therefore, when $\Theta(t_{i-1})$ is not available because of missing data, we maximize $O(\Theta(t_i), \Theta(t_{i-k}), \lambda(t_i))$, where t_{i-k} is the most recent time point with available data, and we use modified regularization parameters $\eta'_1 = \eta_1/k$, $\eta'_2 = \eta_2/k$.

7. Ecomodes

We computed the eigensystem of the correlation matrix $C(t)$ for ecosystems in a single illumination condition, either high ($I = I_0$) or low ($I = 0.25I_0$), and the results are shown in Fig. 12. We find that in the high-illumination condition, the correlation matrix is dominated by noise. In the low-illumination condition, the largest ecomode has a similar structure to that computed across illumination conditions, but this breaks down after about 30 days. The M and S ecomodes in the $I = 0.25I_0$ condition differ qualitatively from those we computed across all illumination conditions (see Fig. 5).

-
- [1] M. Delbrück, *Statistical Fluctuations in Autocatalytic Reactions*, *J. Chem. Phys.* **8**, 120 (1940).
- [2] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic Gene Expression in a Single Cell*, *Science* **297**, 1183 (2002).
- [3] M. Lieb, *The Establishment of Lysogenicity in Escherichia coli*, *J. Bacteriol.* **65**, 642 (1953).
- [4] J. Sheng, E. Malkiel, J. Katz, J. Adolf, R. Belas, and A. R. Place, *Digital Holographic Microscopy Reveals Prey-Induced Changes in Swimming Behavior of Predatory Dinoflagellates*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17512 (2007).
- [5] S. J. Gould, *Wonderful Life: The Burgess Shale and the Nature of History* (WW Norton & Company, NY, 1990).
- [6] J. Esper, U. Büntgen, D. C. Frank, D. Nievergelt, and A. Liebhold, *1200 Years of Regular Outbreaks in Alpine Insects*, *Proc. R. Soc. B* **274**, 671 (2007).
- [7] M. Dornelas, N. J. Gotelli, B. McGill, H. Shimadzu, F. Moyes, C. Sievers, and A. E. Magurran, *Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss*, *Science* **344**, 296 (2014).
- [8] D. R. Hekstra, S. Cocco, R. Monasson, and S. Leibler, *Trend and Fluctuations: Analysis and Design of Population Dynamics Measurements in Replicate Ecosystems*, *Phys. Rev. E* **88**, 062714 (2013).
- [9] L. W. Alvarez, W. Alvarez, F. Asaro, and H. V. Michel, *Extraterrestrial Cause for the Cretaceous-Tertiary Extinction*, *Science* **208**, 1095 (1980).
- [10] A. M. Michalak, E. J. Anderson, D. Beletsky, S. Boland, N. S. Bosch, T. B. Bridgeman, J. D. Chaffin, K. Cho, R. Confesor, I. Daloğlu *et al.*, *Record-Setting Algal Bloom in Lake Erie Caused by Agricultural and Meteorological Trends Consistent with Expected Future Conditions*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6448 (2013).
- [11] S. E. Luria and M. Delbrück, *Mutations of Bacteria from Virus Sensitivity to Virus Resistance*, *Genetics* **28**, 491 (1943).
- [12] A. H. Sturtevant, H. J. Muller, and C. B. Bridges, *The Mechanism of Mendelian Heredity* (Holt, NY, 1922).
- [13] D. R. Hekstra and S. Leibler, *Contingency and Statistical Laws in Replicate Microbial Closed Ecosystems*, *Cell* **149**, 1164 (2012).
- [14] Z. Kawabata, K. Matsui, K. Okazaki, M. Nasu, N. Nakano, and T. Sugai, *Synthesis of a Species-Defined Microcosm with Protozoa*, *J. Protozool. Res.* **5**, 23 (1995).
- [15] F. B. Taub and A. K. McLaskey, *Oxygen Dynamics of Aquatic Closed Ecological Systems: Comparing the Whole to a Subsystem*, *Ecol. Model.* **293**, 49 (2014).
- [16] E. A. Kearns and C. E. Folsome, *Measurement of Biological Activity in Materially Closed Microbial Ecosystems*, *BioSystems* **14**, 205 (1981).
- [17] Z. Frentz, S. Kuehn, D. Hekstra, and S. Leibler, *Microbial Population Dynamics by Digital In-Line Holographic Microscopy*, *Rev. Sci. Instrum.* **81**, 084301 (2010).
- [18] *The Chlamydomonas Sourcebook*, Vol. 1, edited by E. H. Harris, D. B. Stern, and G. Witman (Academic Press, Waltham, MA, 2009).
- [19] *Tetrahymena thermophila*, *Methods in Cell Biology*, Vol. 109, edited by K. Collins (Academic Press, Waltham, MA, 2012).
- [20] M. A. Savageau, *Escherichia coli Habitats, Cell Types, and Molecular Mechanisms of Gene Control*, *Am. Nat.* **122**, 732 (1983).
- [21] S. A. Amin, L. R. Hmelo, H. M. van Tol, B. P. Durham, L. T. Carlson, K. R. Heal, R. L. Morales, C. T. Berthiaume, M. S. Parker, B. Djunaedi *et al.*, *Interaction and Signalling between a Cosmopolitan Phytoplankton and Associated Bacteria*, *Nature (London)* **522**, 98 (2015).
- [22] A. Sano, M. Watanabe, and T. Nakajima, *Adaptive Characteristics of a Ciliate Tetrahymena thermophila in Endosymbiotic Association with a Green Alga Chlorella vulgaris Derived in a Long-Term Microcosm Culture*, *Symbiosis* **47**, 151 (2009).
- [23] Q. Zhang, G. Lambert, D. Liao, H. Kim, K. Robin, C-k. Tung, N. Pourmand, and R. H. Austin, *Acceleration of Emergence of Bacterial Antibiotic Resistance in Connected Microenvironments*, *Science* **333**, 1764 (2011).
- [24] C. H. Waddington, *Canalization of Development and the Inheritance of Acquired Characters*, *Nature (London)* **150**, 563 (1942).
- [25] C. H. Waddington, *The Strategy of the Genes* (Allen & Unwin, Australia, 1957).
- [26] S. L. Rutherford and S. Lindquist, *Hsp90 as a Capacitor for Morphological Evolution*, *Nature (London)* **396**, 336 (1998).
- [27] G. I. Lang, D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai, *Pervasive Genetic Hitchhiking and Clonal Interference in Forty Evolving Yeast Populations*, *Nature (London)* **500**, 571 (2013).
- [28] M. M. Zambrano, D. A. Siegele, M. Almirón, A. Tormo, and R. Kolter, *Microbial Competition: Escherichia coli*

- Mutants that Take Over Stationary Phase Cultures*, *Science* **259**, 1757 (1993).
- [29] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, *Interdependence of Cell Growth and Gene Expression: Origins and Consequences*, *Science* **330**, 1099 (2010).
- [30] <http://www.chlamy.org/SG.html>.
- [31] <http://www.chlamy.org>.
- [32] M.T. Sweet and C.D. Allis, *Long-Term Storage of Unfrozen Tetrahymena Cultures in Soybean Medium*, *CSH Protocols*, doi: 10.1101/pdb.prot4470 (2006).
- [33] F.B. Taub and A. M. Dollar, *The Nutritional Inadequacy of Chlorella and Chlamydomonas as Food for Daphnia pulex*, *Limnol. Oceanogr.* **13**, 607 (1968).
- [34] D. Cassidy-Hanley, H.R. Smith, and P.J. Bruns, *A Simple, Efficient Technique for Freezing Tetrahymena thermophila*, *J. Euk. Microbiol.* **42**, 510 (1995).
- [35] V.N. Vapnik, *The Nature of Statistical Learning Theory* (Springer Science & Business Media, Berlin, 2013).
- [36] L. Wasserman, *All of Nonparametric Statistics* (Springer Science & Business Media, Berlin, 2006).
- [37] M.G. Albers, Master's thesis, Swiss Federal Institute of Technology Zurich (2012).
- [38] P. Hall and J. Horowitz, *A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions*, *Ann. Stat.* **41**, 1892 (2013).