Regulation of chain length in two diatoms as a growth-fragmentation process

Marco Gherardi,[1,2,3] Alberto Amato,[4] Jean-Pierre Bouly,[3] Soizic Cheminant,[3] Maria Immacolata Ferrante,[4] Maurizio Ribera d'Alcalá,[4] Daniele Iudicone,[4] Angela Falciatore,[3,5] and Marco Cosentino Lagomarsino[3,5,*]

[1]*Dipartimento di Fisica, University of Milano, Via Celoria 16, Milano, Italy*

[2]*INFN, Milano, Italy*

[3]*Sorbonne Universités, UPMC Université Paris 6, UMR 7238, Computational and Quantitative Biology, 15 rue de l'École de Médecine, Paris, France*

[4]*Stazione Zoologica Anton Dohrn, Department of Integrative Marine Ecology, Villa Comunale, Naples, Italy*

[5]*CNRS, UMR 7238, Paris, France*

(Received 11 March 2016; published 23 August 2016)

Chain formation in diatoms is relevant because of several aspects of their adaptation to the ecosystem. However, the tools to quantify the regulation of their assemblage and infer specific mechanisms in a laboratory setting are scarce. To address this problem, we define an approach based on a statistical physics model of chain growth and separation in combination with experimental evaluation of chain-length distributions. Applying this combined analysis to data from *Chaetoceros decipiens* and *Phaeodactylum tricornutum*, we find that cells of the first species control chain separation, likely through a cell-to-cell communication process, while the second species only modulates the separation rate. These results promote quantitative methods for characterizing chain formation in several chain-forming species and in diatoms in particular.

DOI: 10.1103/PhysRevE.94.022418

I. INTRODUCTION

Diatoms are unicellular algae characterized by a rigid external cell wall (the frustule) composed of two siliceous and unequal halves (called the valves), with the bigger one fitting onto the smaller one like a lid fits onto its box. This ecologically key class of algae inhabits all the aquatic environments, from salty waters to brackish and fresh waters. They have a deep impact on oxygen production on Earth, fixing at least one fifth of the total carbon fixed per year on our planet [1]. Additionally, diatoms are the most diverse class of algae, numbering 12 000 described species to date and possibly 8000 unknown species [2].

One of the main features characterizing diatoms is their ability to form variably stiff [3] and differentially shaped chains [4]. Chains are produced only at mitotic division, when two daughter cells do not separate completely but stay together. Usually, intercellular junctions are not accompanied by cytological contact between sibling cells. In diatoms cell division is very peculiar (see Ref. [5] for an extensive review of mitosis and cytokinesis in diatoms), with the daughter cells "growing" within the mother, then separating via an animal-like cleavage furrow [5,6], followed by the synthesis of a new silica cell wall (similar to what happens to the cellulose cell wall in plants). Upon separation, in most species the two daughter cells keep physical contact and produce variably long chains. Chain formation is a species-specific (and possibly ecotype-specific) feature, and the junction modes known to date are quite diverse [7–10], with some species showing cells connecting to each other via frustule processes (e.g., *Skeletonema* [11], *Chaetoceros* [12], the extant genus *Trochosira* [13]), others producing mucus pads that keep the cells together from the apices (e.g., *Asterionellopsis* [14], *Tabellaria* [15], the extant family Rutilariaceae [16]), and others synthesizing chitin threads between adjacent cells (*Thalassiosira* [9]).

The reasons and mechanisms for diatom chain formation are still being debated. Likely chains have ecological importance [17–20], which could range from buoyancy behavior, similar to flocculation in yeast [21,22], to increased resistance to predators. More generally, clustering is sometimes interpreted as a first step towards multicellularity [23], the second being division of labor with the evolution of specialized cells [24], as in the slime mold *Dictyostelium discoideum* [25]. Diatom ability to produce chains has never been interpreted as a sign of nascent multicellularity, although in some cases (*Licmophora*, *Encyonema*, or *Navicula ramosissima*) particularly complex colonies have been hypothesized to contain specialized and behaviorally differentiated cells [26]. In other phytoplankters [27] as well as in yeasts [28], multicellular-like behavior was invoked and experimentally induced.

One prominent question is the extent and mechanisms that different species and ecotypes use to control the range of chain lengths found in a population. A minimal mechanism purely controls the robustness of junctions, with more fragile junctions being more prone to separation (thereby reducing the average chain length of a population). The extent to which separation may be produced by environmental forcing, e.g., strain due to the fluid shear, is still unclear (see, e.g., Ref. [29]). For example, the physical separation rate may be increased in the presence of turbulent flows, while in these conditions it might be more advantageous for a species to maintain or even increase its average chain length. Alternatively, chain separation might be under more tight regulation, involving, e.g., communication between neighboring cells. Such mechanisms may allow chain formers to react to different environmental changes.

Despite the existing knowledge, the process of chain formation is still poorly understood. Most importantly, no widely accepted tools exist to quantify it in controlled settings. In this scenario, even the elementary statistical physics models to understand and quantify data are an open question. Here, we address these issues with a combination of experiments

*marco.cosentino-lagomarsino@upmc.fr

quantitating chain length in steadily growing colonies and mathematical models of chain formation and separation.

The main goal of this work is the definition and application of methods capable of discerning different regulating mechanisms and of quantifying the biases in the underlying processes. Such methods could become general tools for future research. Figure 1 describes our combined approach using theoretical modeling and experimental data. We started from the evidence that the average chain length measures the tendency to form chains but is often not sufficient to reveal the specific processes that regulate the dynamics. The time-dependent distribution of lengths, instead, contains much more detailed information which should be exploited.

We selected the chain-forming diatom *Chaetoceros decipiens* as the main experimental model species. Besides producing chains, this species has the advantage of presenting a peculiar feature allowing us to identify the exact spot where a chain will split into two or more subchains. Indeed, *C. decipiens* is characterized by the production of cellular processes called setae. The cells that form part of a chain produce intercalary setae that span perpendicular to the length of the chain up to seven to eight cell radii and are clearly visible in light microscopy. When two cells in a chain are going to separate, they synthesize terminal setae that are parallel to the chain length and are not connected [see Fig. 1(a) and the inset in Fig. 3]. This will ultimately generate two terminal cells whose setae are at the apices of the chains and are usually thicker than the intercalary ones. As a term of comparison, we used the *P. tricornutum*, considering both a wild-type strain, which under specific growth conditions may form short chains in the laboratory (typically two to four cells), and a chain-forming transgenic line (hereafter termed "mutant" but strictly speaking a transformant; see Sec. II), which forms longer chains (up to 50 cells).

As the simplest mechanism for chain formation, we assumed that cells grow and divide, thus increasing the chain length. Dividing cells may separate after division, or the chain may break at random junctions. This constitutes our model (see Sec. II for a precise formulation). In such a model, division and chain separation are treated as independent processes: all cells have the same fixed division rate (per cell) α, and all cell-to-cell junctions have a fixed separation rate (per junction) β. The stochasticity in such a model represents the cell-to-cell and chain-to-chain variability in chain separation. This model predicts that, after some relaxation time, a steady state should be reached for the histogram of the chain length and also predicts the shape of this histogram. We define a clear procedure to compare this model with experimental data from colonies grown in the laboratory. Comparing the (null) model with data gives insight into the non-null process of chain formation. We also considered non-null variants of the model to investigate which extra ingredients in the regulation of growth and chain separation could reproduce the observed data.

II. MATERIALS AND METHODS

A. *Chaetoceros decipiens* and *Phaeodactylum tricornutum* experiments

Diatom strain SZN-Cdec (*Chaetoceros decipiens*) was grown in f/2 diatom-specific culture medium [30]. Cells were grown in tissue culture flasks with no agitation or bubbling at 18 °C, a light:dark (L:D) period of 12L:12D, and a photon fluence rate of about 80 μmol photons m^{-2} s^{-1} produced by cold white light bulbs adapted for phytoplankton growth. Seven replicate experiments were run with *C. decipiens*. Cultures were kept in an active and exponential growth phase for 2 weeks before each experiment by serial dilutions. In order to inoculate cells in new vessels and fresh f/2 medium for the experiments, a cell count was carried out by a Sedgewick-Rafter counting chamber. Initial cell concentration was set at 250 cells mL^{-1} (according to the natural blooming concentration). We preferred low cell concentrations in order to avoid overcrowding and self-shading during the experiments. At the moment of inoculum, cells were in the exponential growth phase. For each experiment, Plexiglas beakers were inoculated with *C. decipiens* cells to reach the aforementioned concentration (in Fig. 1 for *C. decipiens* experiments). Inoculated beakers were put in the experimental conditions and left 24 h to acclimate to the dilution into the fresh medium and to the new vessels. Temperature, photoperiod, and light intensity for the experiment were set as the growth chamber used for maintenance. At the same time of day, a sample was taken from each beaker and fixed with neutralized formaldehyde (final concentration 1% vol/vol) and stored at 4 °C in the dark until analysis. For technical and logistic reasons, three out of the seven replicates lasted 3 d (from T0 to T2), while the other four lasted 4 days (until time point T3). Although *C. decipiens* chains are quite stiff [3], sampling was gently performed by delicately stirring the culture with a 25-mL strippette for 15 s. This sampling procedure did not disturb chain spectra. We remark that a broken chain is easily recognizable, especially for *C. decipiens*, being characterized by the lack of terminal cells at the apices [31].

Cell counts were used both to define chain distribution over time and to calculate growth rates. On average, 405 chains per replicate per time point were counted in the Sedgewick-Rafter counting chambers [32] in a Zeiss Axiophot light microscope at a magnification of 100×. For SZN-Cdec experiments, the number of separating chains (chains presenting at least one separation point identified by two adjacent separating cells within the chain; see Fig. 3 below) was recorded. The chain length of the separated subchains was recorded as well.

Experiments on *Phaeodactylum tricornutum* were performed with the CCMP632 (Pt1) strain obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton. Analyses were performed on wild-type cells showing reduced chain length (maximum of four cells in a chain) in a small percentage (5% of the population) and on a transgenic line showing a chain formation phenotype (up to 50 cells in a chain) under standard conditions: growth in f/2 medium [30] without additional silicic acid and without agitation, temperature of 18 °C and illumination with white fluorescent light in a 24-h light photoperiod. The transgenic line has been obtained as described in [33]. This line contains an antisense fragment for the phytochrome photoreceptor (Phatr2_54330). To date, a direct relationship between the phytochrome deregulation and the observed chain phenotypes has not been established. However, we considered that this line represented a useful tool for the purpose of this work. With light intensity being one of the most important factors

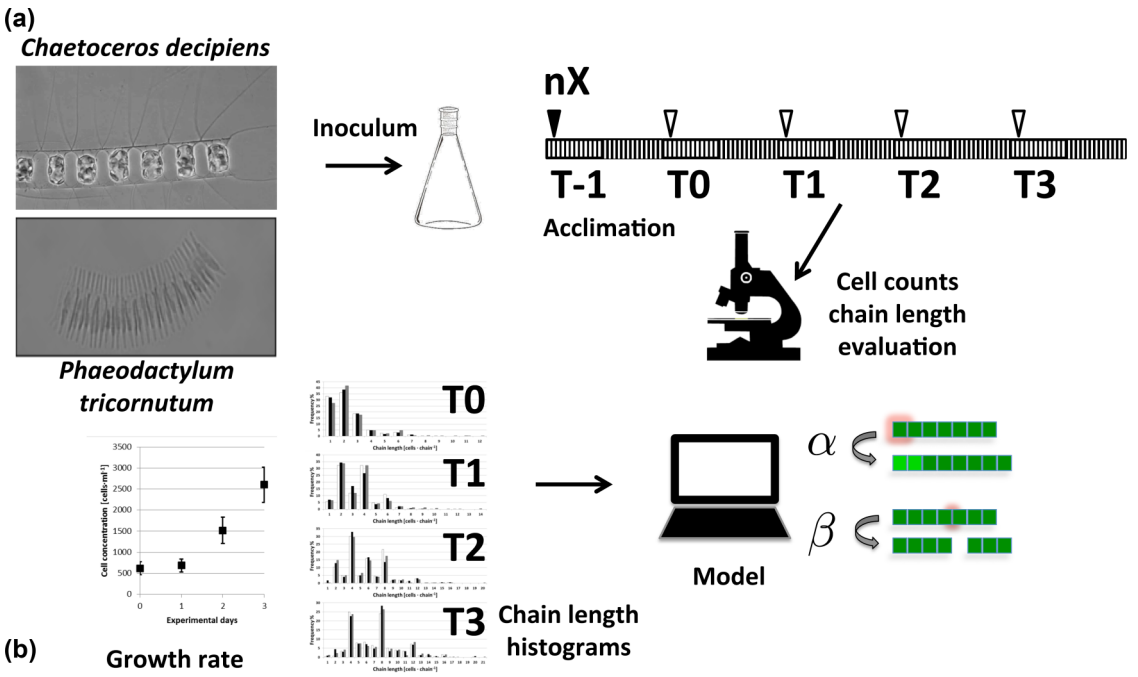

FIG. 1. Workflow of the experiments. (a) Cells were cultivated and kept in exponential growth phase until inoculum. After inoculum into fresh culture medium, cultures were poured into flasks. In the time bar, white boxes identify illuminated periods (12 h), and black boxes identify dark periods (12 h) for experiments with *C. decipiens*; *P. tricornutum* was kept in a 24-h light photoperiod. (b) Daily microscopic inspection led to the growth curves and the chain-length histograms for each replicate for each time point, which were analyzed with the model.

affecting the growth rate, several light intensities have been used. The wild-type cells were grown at seven different light intensities: 200 μmol photons $m^{-2}\,s^{-1}$ ($n = 6$), 75 μmol photons $m^{-2}\,s^{-1}$ ($n = 2$), 50 μmol photons $m^{-2}\,s^{-1}$ ($n = 4$), 25 μmol photons $m^{-2}\,s^{-1}$ ($n = 1$), 15 μmol photons $m^{-2}\,s^{-1}$ ($n = 2$), 10 μmol photons $m^{-2}\,s^{-1}$ ($n = 2$), and 2.5 μmol photons $m^{-2}\,s^{-1}$ ($n = 1$). Moreover, in order to have constant growth rate higher than two divisions per day, wild-type cells were grown in f/2 medium supplemented by 0.5% of tryptone [34] at 200 μmol photons $m^{-2}\,s^{-1}$ ($n = 6$). The transgenic line was grown at seven different light intensities: 75 μmol photons $m^{-2}\,s^{-1}$ ($n = 5$), 50 μmol photons $m^{-2}\,s^{-1}$ ($n = 2$), 25 μmol photons $m^{-2}\,s^{-1}$, 15 μmol photons $m^{-2}\,s^{-1}$ ($n = 2$), 10 μmol photons $m^{-2}\,s^{-1}$ ($n = 7$), 5 μmol photons $m^{-2}\,s^{-1}$ ($n = 2$), and 2.5 μmol photons $m^{-2}\,s^{-1}$ ($n = 1$). For each experiment, cells were adapted to the different light conditions and media for at least 2 weeks and kept at constant growth rate in exponential phase by dilutions every 3 d. Cell concentration was set between 0.5×10^5 and 1×10^5 cells/mL at each dilution. After adaptation, cell number and chain length were directly counted (without fixation) in a Malassez chamber for 3 d (from T0 until T2) at the same time of day.

Raw data are available with the authors.

B. Stochastic model of chain accretion and chain separation

The stochastic model we adopted as a null reference to our data implements duplication (growth and division) of cells and fragmentation of chains as two independent mechanisms. Each cell duplicates with a growth rate α, and duplicate cells remain in chains. Moreover, each interface between two consecutive cells "breaks" with a rate β (see Fig. 1). We will use the terms *fragmentation* and *separation* to refer to the process whereby a chain breaks up into two daughter chains. Dimensionally, α and β are, respectively, rates per cell and per junction, meaning that the actual overall rates of chain separation and accretion (the "propensities" of these processes) depend dynamically on the number of interfaces and cells. Apart from this, the model assumes that the two processes regulated by α and β are independent, and the two rates are fixed. The assumption of a time-independent α is justified by the experimental growth curves, which are well fitted by exponential functions. In particular, for each experiment, growth rates were computed for each time interval (i.e., from T0 to T1, from T1 to T2, etc.), as $\alpha_i = \ln(N_i/N_{i-1})$, where N_i is the total cell count at time point (day) i and ln is the natural logarithm. For all replicates, the variability of α_i for different time segments was less than 15%. The values α for each replicate were obtained by linear fits of the form $\ln N_i = \ln N_0 + i\alpha$. A fixed β is the minimal assumption defining this null model. As we will see, we find that different species may agree or not with the null model and hence may or may not allow us to reject the assumptions.

More precisely, the system is composed of a time-varying number F_t of chains, each composed of k_i cells. At each time t the state is specified by the numbers k_i, with $i = 1, \ldots, F_t$, representing the number of cells in each chain, i.e., the chain's length. Let $F_t(k)$ be the number of chains of size k at time t, which is the quantity analyzed empirically. Its time evolution is controlled by the two (nonstationary) Poisson processes of cell duplication and chain separation at junctions described above. This model is simple enough to be tractable analytically. One can write an equation for the evolution of $F_t(k)$, taking into

account the different sources of increase or decrease of this quantity, namely, the duplication of a cell in a chain of size $k-1$ and the separation of a chain of size $\bar{k} > k$ (in the two relevant ways) and the two sources of decrease, i.e., the duplication of a cell in a chain of size k and the separation of a chain of size k. The long-time behavior of the chain-length abundance $F_t(k)$ can then be studied by standard techniques; the details are reported in the Appendix. In brief, the model predicts that the chain-length normalized distribution $p_t(k) = F_t(k)/F_t$ reaches a steady state $p_{ss}(k)$ for long times. The steady state is characterized by the fact that the relative proportions of chains of different lengths remain unchanged, while the number of chains keeps increasing. The shape of the distribution $p_{ss}(k)$ is given by the exponential

$$p_{ss}(k) = \frac{\beta}{\alpha}\left(1+\frac{\beta}{\alpha}\right)^{-k}. \quad (1)$$

A practical consequence of Eq. (1) is the following relation between the growth rate α and the mean chain length $< k >$ at the steady state (see the Appendix):

$$\langle k\rangle = \frac{\alpha}{\beta} + 1. \quad (2)$$

To compute the full time dependence of $F_t(k)$ we rely on numerical simulations, performed using the first-reaction method of Gillespie [35].

III. RESULTS

A. Short subchains are suppressed in *C. decipiens*

We first compared the model with the chain-length data from *C. decipiens* (see Sec. II). The model predicts that chain-length histograms should take an exponentially decaying form in the long-time regime. Histograms for this quantity for experimental data, shown in Fig. 2(a), strongly deviate from such prediction (p values of Kolmogorov-Smirnov tests are of order 10^{-14}). For instance, an exponential distribution is monotonically decreasing and has its maximum at 1, which is clearly not the case for the empirical distributions. Assuming the system is not at the steady state, the evolution from the initial distribution (at $t = 0$ d) can be obtained numerically, but again, we found no agreement between the data and the null model, which does not reproduce the observed local maxima in the chain-length distributions.

Therefore, a modified model, taking into account a possible bias in the chain-separation mechanism, was considered. Such a model variant attempts to positively reproduce the data through a modification of the null assumption on the separation rate. Specifically, we assumed that regulation of chain separation is implemented as follows: cells do not separate if a chain boundary is closer along the chain than a fixed number k_0 of cells, thus inhibiting the formation of resulting chains shorter than k_0. In order to estimate the best values of the separation rate β and of the cutoff k_0, we perform extensive simulations starting from the empirical initial condition, varying these parameters and computing the Kolmogorov-Smirnov distance between the simulated and the empirical chain-length distributions [36] at $t = 2$ d. The duplication rate $\alpha = 0.72$ is computed by cell counts. Figure 2 compares this modified model with data. This analysis

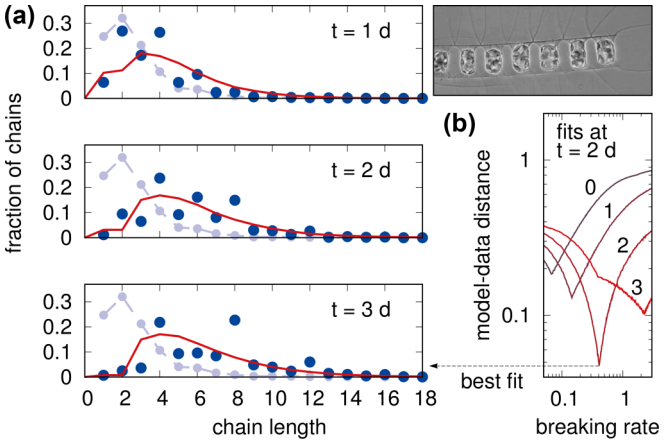

FIG. 2. Biased chain separation in *C. decipiens*. (a) Typical chain-length distributions show that the most common chain is a few cells long, compatible with the presence of a bias suppressing the separation of shorter chains. The three plots correspond to times 1 d apart from each other. Blue dots are experimental data (error bars estimated as Poissonian counting errors are smaller than the points), dashed lines are the initial condition at $t = 0$, and red solid lines are the predictions of the model with bias. (b) Suppression of short-chain formation is manifested by the preference for a cutoff length $k_0 = 2$ in the biased model. Data and simulations are compared through the Kolmogorov-Smirnov distances (vertical axis) between the chain-length distributions at $t = 2$ d. Curves corresponding to different cutoffs $k_0 = 0,1,2,3$ (numbers near the lines) realize the optimum overlap with the data at different values of the separation rate per junction β (horizontal axis, in separations per day).

indicates that separation is inhibited for chains of up to $k_0 = 2$ cells. The fitting procedure also estimates the separation rate, namely, $\beta = 0.41$ separations/d.

Simulations using the best value of β can then be compared to the time evolution of the distributions [Fig. 2(a)], showing good agreement. In particular, the model also correctly reproduces the decay of short-chain abundances from the initial condition to the asymptotic null concentration. Although the parameters are estimated by matching the distributions at $t = 2$ d, the model prediction is in agreement with the data also at $t = 3$ d (for the four replicates that were followed up to that time point). Notice, however, that the data display strong parity effects, with peaks at even chain lengths, that are not captured by the model (see the next section).

B. Positions of separating points in chains highlight the biases

The analysis presented above is consolidated by support coming from independent measurements. We used the fact that formation of terminal setae precedes separation of chains in *C. decipiens*. Hence, we could verify whether a bias on future subchain length is detectable in the pattern of chain-separation points. To this aim, we collected data on the positions of the setae along each chain that presented them. This defines a population of chains $n(k)$ (number of chains of length k with at least one terminal seta) and another population

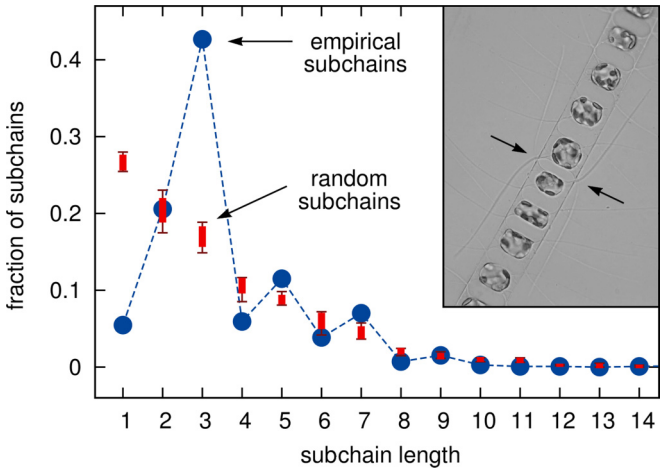

FIG. 3. Formation of setae in *C. decipiens* shows suppression of single-cell subchains and enrichment of odd lengths greater than 1. Subchains are defined as the consecutive cells lying between two terminal setae (or chain boundaries). The proportion of subchains (y axis) with a given number of cells (x axis) is shown both for empirical data (blue circles; the dashed line is a guide to the eye) and for a simulated null model (red) where terminal setae are uncorrelated and randomly distributed along the empirical chains. For empirical data (1113 subchains in total), error bars estimated as Poissonian counting errors are approximately the size of the data points (slightly larger for the data at subchain length 3). For simulations, symbols and boxes denote mean values and standard deviations, respectively, while whiskers are 1st and 99th percentiles. Inset: a *C. decipiens* cell chain showing one separation point. Arrows indicate terminal setae on two adjacent intercalary separating cells, i.e., the junctions where the chain will separate into two subchains.

$n'(k)$, constituted by the subchains lying between consecutive separation points. Let m be the number of separation points. The bias on chain formation can be extracted by comparison with a null mechanism in which all interfaces are equally likely to produce separation points. Starting from the empirical $n(k)$, we simulated 10^4 realizations of this unbiased model by inserting m separation points in randomly chosen interfaces on all chains and compared the resulting subchain populations with the empirical one $n'(k)$.

The results (Fig. 3) highlight the presence of strong bias suppressing the formation of single-cell subchains, which is qualitatively in agreement with the phenomenology observed combining our model and the chain-length distributions. However, this analysis also shows that the positions of the separation points heavily favor the formation of three-cell subchains and, to a lesser extent, are biased towards producing subchains of odd lengths. This parity effect is not due to the finite size of the chains or to the particular lengths they happen to have in our empirical data because those features are taken into account as the starting configuration for the null model considered in this section. Deviations from the null behavior could, in principle, originate from a scenario where chains of different lengths have different probabilities of forming terminal setae. However, we tested this hypothesis on our data by calculating the average number of setae per junction on chains grouped by length and found that setae appear with comparable rates (per junction) on chains of different lengths.

C. Chain formation in *P. tricornutum* is not biased

Next, we considered chain-length data from the set of experiments on *P. tricornutum* at different growth rates controlled by light conditions, as explained in Sec. II.

Let us first focus on the transgenic line. We set out to establish whether chain-length data deviate from the null growth-fragmentation model and from the *C. decipiens* data. In order to compare different species and conditions, we used a relevant property of the steady-state solution of the model: if chain lengths are rescaled (separately for each data set) by the average length $< k >$, then all their distributions should collapse onto the same exponential function [see Eqs. (1) and (2)]. Notice that such a rescaling yields a continuous exponential distribution only in the limit of large $< k >$. For small $< k >$, histograms will present finite-size fluctuations around this law (remark, however, that even in this case the model predicts an exactly exponential distribution of k). These fluctuations are expected to average out when using data sets with many different values of $< k >$, as is the case here. The rescaling is carried out by computing separately for each replicate j the average $< k >_j$. Then each chain length k_i in replicate j is rescaled as $k_i/< k >_j$. All these values are pooled to compute binned histograms.

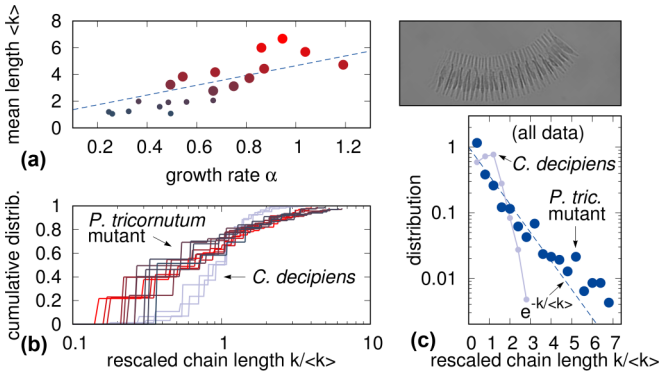

FIG. 4. Null growth-fragmentation behavior in *P. tricornutum* mutants. (a) The mean chain length (y axis) grows linearly with the growth rate (x axis), as predicted by the model, Eq. (2). Circles are colored according to their mean chain length. Small circles are those lying below the threshold $< k > = 2$, which are discarded from the analysis in (b). The dashed line is the model prediction, Eq. (2), with $\beta = 0.27$ fitted from the data. (b) Cumulative distributions of chain lengths for *P. tricornutum* mutants (data at $t = 2$ d for the 11 experiments with $< k > \leqslant 2$) and for *C. decipiens* (data at $t = 1,2,3$ d), rescaled by mean length (separately for each condition), show remarkable collapse for the two different species on two distinct behaviors. (c) The rescaled chain-length histograms (bin width of 0.4) of *P. tricornutum* computed with data pooled from all available conditions [this combination is possible thanks to the agreement shown in panel (a)] follow the exponential form predicted by the model (dashed line).

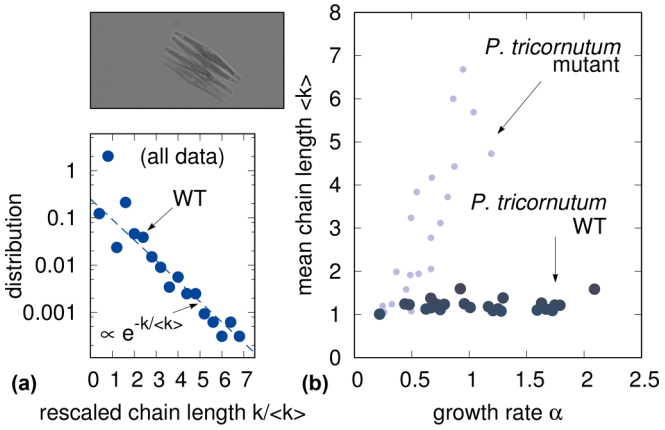

FIG. 5. (a) Wild-type *P. tricornutum* forms short chains, whose lengths are compatible with the null growth-fragmentation model (dashed line); (b) however, the growth rate has undetectable influence on the average length. Data in (a) are rescaled and binned as in Fig. 4. Small gray points in (b) are the same data as in Fig. 4(a); the larger points are the WT data, shaded according to their mean chain length (on the y axis), as in Fig. 4.

Figures 4(b) and 4(c) display the results of this procedure and show that data on *P. tricornutum* obey this collapse property. Additionally, this analysis shows [Fig. 4(c)] that the data on chain length in *P. tricornutum* are in agreement with the null model prediction. Finally, Fig. 4(b) shows that the data on *C. decipiens*, taken at different time points and therefore having different mean lengths, also show collapse of the chain-length distributions but on a nonexponential function [Fig. 4(c)].

Another useful prediction of the model is the relation (2) between the growth rate and the mean chain length at the steady state, which explains the linear scaling observed in Fig. 4(a). Since the *P. tricornutum* data follow the null growth-fragmentation model and growth rate is obtained directly from measurements, this prediction allows us to estimate the separation rate. A fit of the experimental data against this expression yields $\beta = 0.27(3)$ separations/d. Individual variability of the fragmentation rates computed as $\beta = \alpha/(< k > - 1)$ gives an overall range $\beta = 0.1$–0.4 separations/d. Estimation of the fragmentation rate by fitting the empirical distributions against an exponential form gives consistent results. Altogether, the data fully support the validity of a null model of chain formation with no bias, where chain lengths are the outcomes of random duplication and fragmentation.

We now turn to the analysis of the wild type, which formed very short chains, even at large growth rates induced by adding tryptone to the medium. Here, the stochasticity due to small chain size makes it difficult to discern any collapse of the cumulative distributions. However, merging all the data (by normalizing each separate distribution by its mean, as above) again confirms that the overall behavior is robustly exponential [Fig. 5(a)], as predicted by the null model. Interestingly, the mean chain length of the wild-type strain is roughly independent of the growth rate. This is clearly visible Fig. 5(b), which shows a stable mean chain length for a wide range of growth rates. Spearman's correlation test is unable to reject the null hypothesis that the mean chain length for the wild type is constant with growth rate (p value ≈ 0.4). By comparison, the same test on the seven replicate experiments of *C. decipiens* gives a p value of order 10^{-7}.

IV. DISCUSSION AND CONCLUSIONS

The quantitative approach defined and validated here allowed us to explore the mechanisms driving chain formation in two diatom species. Of particular interest is its applicability under the lack of biological information on the underlying processes (e.g., chemical signaling, local microenvironment, cell differentiation, distribution of organic compounds and/or proteins in single cells or chains), possibly allowing us to extend the analysis to many species.

This first investigation already produced several interesting results. First of all, it indicates that systematic measurements of chain length of growing diatoms contain a wealth of information, which is not necessarily accessible by considering the means only. Our results demonstrate that measurement of chain-length distributions can be informative on the collective behavior of diatoms and provide a general method for future investigations.

More specific considerations can be made on the implications of our findings for diatoms. Our approach was capable of discerning different regulating mechanisms by combining data with quantitative growth-fragmentation models. Complementing those with positive models, we were able to quantify the biases in the underlying processes, validating hypotheses and pinpointing specific mechanisms. Indeed, for *C. decipiens*, we firmly reject the null model and establish that characteristic lengths exist, likely due to signaling between cells. The most important of these trends is a constraint preventing the formation of very short chains. We showed that a positive version of the model including this constraint can reproduce satisfactorily the observed chain-length distributions and their dynamics. This result is supported by fits with *positive* models where the constraint is realized.

An independent analysis of *C. decipiens* using terminal setae, where future separations will develop, confirms the existence of a strong negative bias for small chains and suggests the existence of additional biases on even vs odd positions of cells along the chain. Specifically, we have found evidence of different parity effects in our data. Note that this may be the consequence of some degree of synchronization in the separation of cells, but the analysis of the subchains delimited by setae instead suggests that this hypothesis should be ruled out. We do not have a simple explanation for such a complex pattern of biases, other than the notion of a (two-body) interaction between neighboring cells, possibly giving rise to parity effects. We speculate that these may be due to the effect of "morphogenetic" signaling processes inside the chains, giving rise to periodic patterns. Interestingly, cell divisions in chains of the same species have been carefully described by the authors of Ref. [31], who highlighted that the orientation of intercalary setae depends on the position of the cell within the chain, intercalary vs terminal. This reinforces the hypothesis of a possible morphogenetic coordination among the dividing cells.

Conversely, all the results in the *P. tricornutum* strains may be described by a simple unbiased growth-fragmentation random process. These results suggest that in the *P. tricornutum* strain, the orchestration of chain-length regulation, if present in the wild, does not rely on cell-to-cell communication, only on modulation of the effective separation rate. Notably, while the mutant is fully consistent with a fixed separation rate independent of the conditions, the behavior of the wild type is intriguing. Indeed, its chain-length distributions are in full agreement with the null model, but the average length shows little dependence on the growth rate, in a wide range of growth rates. One possibility is that the fragmentation rate β is so large that the growth rates considered are still in a small-length regime dominated by fluctuations. However, one can also speculate that the remarkably flat appearance of the plot in Fig. 5 is the consequence of an active regulation of the separation rate by the cells in response to variations in the growth rate. In other words, the cells in a chain may increase their separation rate in order to keep the chain length small at larger growth rates. A more refined analysis of the fluctuations on top of the minimum value $k = 1$ would be required in order to fully distinguish between those scenarios. At a qualitative level, we remark that the dispersion of the mean length does not appear to increase with growth rate, thus suggesting an active mechanism. While other explanations may apply, we can surmise the following simple possible interpretation of these trends. Multiplets may form because during the division cells excrete substances, such as polysaccharides, which "glue" them together. If the amount of excreted substance is not tightly regulated, as might be the case in fast-growing cells, which display a wide range of variation in the single division rate, the strength of the gluing effect may vary randomly and with it the separation. A similar argument might apply also to the mutant, with the difference of a more effective gluing and a smaller number of divisions with a weaker link.

To sum up, our results demonstrate that measurement of chain-length distributions can be informative on the collective behavior of diatoms, and we propose this combination as a general tool for future research. For example, similar processes may describe the mechanisms determining the end-of-bloom flocculation and rapid sinking of diatoms [37], when the production of exudates is so large that mixing and relative motions of cell aggregates form larger and larger aggregates. While our experiments and model cannot establish directly the presence of cell specialization, which is one of the drivers of generating multicellular assemblages, our findings raise challenging questions on the ability of diatoms to modulate their multicellular structures. Indeed, the fact that most of the chain-forming diatoms produce specific substructures to form their chains is suggestive of direct regulatory mechanisms.

From an ecological point of view, the role of tuning chain length can impact buoyancy and likely also nutrient uptake, as well as modulating palatability according to the size of the most abundant predators [38,39]. It has also been suggested that chains may favor the encounter with mates [40]. Unveiling the mechanisms of biological regulation of chain formation may also be crucially valuable for biotechnological applications. If diatom aggregation can be controlled (like in *Chlamydomonas* and yeasts), it could be exploited, e.g., for easier collection of cells in bioreactors. Being able to manipulate chain formation in economically interesting species for bioproduction would improve cell gathering and facilitate the use of standard techniques of separation from the medium, e.g., filtering or centrifugation.

ACKNOWLEDGMENTS

G. Dell'Aquila is thanked for cell counts and chain-length measurements for *C. decipiens*. A.A. was funded by EC FP7-People COFUND (Grant Agreement No. 600407) and the MIUR Italian Flagship Project RITMARE (Ricerca Italiana per il MARE).

M.G. and A.A. contributed equally to this work.

APPENDIX: ANALYTICAL DETAILS OF THE MODEL

A mean-field approximate equation for $\Delta_t F_t(k) \equiv F_{t+1}(k) - F_t(k)$ can be written by considering, for each k, the processes affecting the value of $F_t(k)$:

$$\frac{d}{dt}F_t(k) = \alpha(k-1)F_t(k-1) - \alpha k F_t(k) + 2\beta \sum_{\bar{k}=k+1}^{\infty} F_t(\bar{k}) - \beta(k-1)F_t(k). \tag{A1}$$

The first (second) term on the right-hand side of Eq. (A1) expresses the duplication of a cell in a chain of length $k-1$ (k), thus leading to an increase (decrease) of $F_t(k)$. The last term is the loss due to the separation of a chain of length k; the factor $k-1$ is the number of interfaces. The third term counts all the ways that a chain of length k can arise as a consequence of the separation of a longer chain of length $\bar{k}$. Note that the probability of duplication is uniformly distributed on all cells, which are $N_c \equiv \sum_k k F_t(k)$, while the probability of chain separation is uniformly distributed among all interfaces, which are $N_I \equiv \sum_k k F_t(k) - F_t$; the number of chains containing $k-1$ interfaces is, of course, $F_t(k)$. With these definitions, one readily obtains Eq. (A1).

We look for steady-state solutions. Although $F_t(k)$ increases with time, one expects a steady-state solution for the ratio $F_t(k)/F_t$ to exist (this is the normalized histogram of chain lengths, with F_t being the number of chains at time t). The vanishing of the time derivative of $F_t(k)/F_t$ implies (as can be shown by direct computation) that $dF_t(k)/dt = a(t)F_t(k)$, where $a(t) = d\log F_t/dt$ does not depend on k. Then, by taking the finite difference of Eq. (A1), evaluated at $k+1$ and at k, one obtains

$$a(t)[F_t(k+1) - F_t(k)] = \alpha(1-k)F_t(k-1) + (2k\alpha + k\beta - \beta)F_t(k) - (k\alpha + \alpha + k\beta + 2\beta)F_t(k+1). \tag{A2}$$

(This equation is strictly valid only for $k > 1$, as it incorporates the duplication of chains made of $k-1$ cells; however, the assumption of the physical boundary condition $F_t(0) = 0$ extends its validity to $k = 1$.) With the assumption $F_t \propto \exp(\alpha t)$, implying $a(t) = \alpha$, Eq. (A2) is exactly solved by the ansatz $F_t(k) \propto C^{-k}$ if $C = 1 + \beta/\alpha$. The average chain

length $\langle k \rangle$ can then be calculated as

$$\langle k \rangle \equiv \frac{\sum_1^\infty kC^{-k}}{\sum_1^\infty C^{-k}} = \frac{\alpha}{\beta} + 1. \tag{A3}$$

Intuitively, the average length is close to 1 if the fragmentation rate β is large with respect to the growth rate α. Finally, the full normalized distribution at the steady state $p_{ss}(k) = F_t(k)/F_t$ is

$$p_{ss}(k) = \frac{\beta}{\alpha}\left(1 + \frac{\beta}{\alpha}\right)^{-k}. \tag{A4}$$

[1] E. Litchman, C. A. Klausmeier, R. J. Miller, O. M. E. Schofield, and P. G. Falkowski, Biogeoscience **3**, 585 (2006).
[2] M. Guiry, J. Phycol. **48**, 1057 (2012).
[3] H. Nguyen and L. Fauci, J. R. Soc. Interface **11**, 20140314 (2014).
[4] N. I. Hendey, *Ministry of Agriculture, Fisheries and Food. Fishery Investigations Series IV* (HMSO, London, 1964), p. 317.
[5] A. De Martino, A. Amato, and C. Bowler, Bioessays **31**, 874 (2009).
[6] A. Tanaka, A. De Martino, A. Amato, M. A., B. Mathieu, P. Rostaing, L. Tirichine, and C. Bowler, Protist **166**, 506 (2015).
[7] G. R. Hasle, Nova Hedwig. Beiheft **45**, 1 (1974).
[8] G. R. Hasle, Nova Hedwig. Beiheft **45**, 167 (1974).
[9] G. A. Fryxell, Bot. Mar. **21**, 131 (1978).
[10] G. A. Fryxell and W. I. Miller, III, Bacillaria **1**, 113 (1978).
[11] G. R. Hasle, Norw. J. Bot. **20**, 109 (1973).
[12] G. A. Fryxell, J. Phycol. **14**, 62 (1978).
[13] A. E. S. Kemp, A. Davies, and R. B. Pearce, Diatom Res. **30**, 151 (2015).
[14] W. H. C. F. Kooistra, R. Gersonde, L. K. Medlin, and D. G. Mann, in *Evolution of Primary Producers in the Sea*, edited by P. G. Falkowski and A. H. Knoll (Elsevier Academic, Burlington, MA, 2007), pp. 207–250.
[15] B. M. Knudson, Ann. Bot. **17**, 131 (1953).
[16] J. Witkowski, P. A. Sims, and D. M. Harwood, Eur. J. Phycol. **46**, 378 (2011).
[17] M. Pahlow, U. Riebesell, and D. A. Wolf-Gladrow, Limnol. Oceanogr. **42**, 1660 (1997).
[18] F. E. Round, R. M. Crawford, and D. G. Mann, *The Diatoms: Biology and Morphology of the Genera* (Cambridge University Press, Cambridge, 1990), p. 747.
[19] A. Alldredge and C. Gotschalk, Deep Sea Res. **36**, 159 (1989).
[20] V. Smetacek, Mar. Biol. **84**, 239 (1985).
[21] K. J. Verstrepen and F. M. Klis, Mol. Microbiol. **60**, 5 (2006).
[22] K. J. Verstrepen, G. Derdelinckx, H. Verachtert, and F. R. Delvaux, Appl. Microbiol. Biotechnol. **61**, 197 (2003).
[23] S. R. Fairclough, M. J. Dayel, and N. King, Curr. Biol. **20**, R875 (2010).
[24] S. A. Frank, *Foundations of Social Evolution* (Princeton University Press, Princeton, NJ, 1998).
[25] N. Van Driessche, C. Shaw, M. Katoh, T. Morio, R. Sucgang, M. Ibarra, H. Kuwayama, T. Saito, H. Urushihara, M. Maeda, I. Takeuchi, H. Ochiai, W. Eaton, J. Tollett, J. Halter, A. Kuspa, Y. Tanaka, and G. Shaulsky, Development **129**, 1543 (2002).
[26] J. T. Bonner, Integr. Biol. **1**, 27 (1998).
[27] W. C. Ratcliff, M. D. Herron, K. Howell, J. T. Pentz, F. Rosenzweig, and M. Travisano, Nat. Commun. **4**, 2742 (2013).
[28] W. C. Ratcliff, J. T. Pentz, and M. Travisano, Evolution **67**, 1573 (2013).
[29] A. M. Young, L. Karp-Boss, P. Jumars, and E. Landis, Limnol. Oceanogr. **57**, 1789 (2012).
[30] R. Guillard, in *Culture of Marine Invertebrate Animals*, edited by D. R. Smith and M. H. Chanley (Springer, New York, 1975), pp. 29–60.
[31] J. D. Pickett-Heaps, J. Phycol. **34**, 989 (1998).
[32] B. J. McAlice, Limnol. Oceanogr. **16**, 19 (1971).
[33] V. De Riso, R. Raniello, F. Maumus, A. Rogato, C. Bowler, and A. Falciatore, Nucleic Acids Res. **37**, e96 (2009).
[34] M. A. Borowitzka, M. L. Chiappino, and B. E. Volcani, J. Phycol. **13**, 162 (1977).
[35] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[36] I. Chakravarty, J. Roy, and R. Laha, *Handbook of Methods of Applied Statistics* (McGraw-Hill, New York, 1967).
[37] V. Smetacek, C. Klaas, V. H. Strass, P. Assmy, M. Montresor, B. Cisewski, N. Savoye, A. Webb, F. d'Ovidio, J. M. Arrieta, *et al.*, Nature (London) **487**, 313 (2012).
[38] F. Peters, L. Arin, C. Marrasé, E. Berdalet, and M. M. Sala, J. Mar. Syst. **61**, 134 (2006).
[39] J. Bergkvist, P. Thor, H. H. Jakobsen, S.-Å. Wängberg, and E. Selander, Limnol. Oceanogr. **57**, 318 (2012).
[40] V. Botte, M. Ribera d'Alcalà, and M. Montresor, J. Plankton Res. **35**, 914 (2013).