Active fragmentation generates extreme statistics of gut bacterial cluster sizes

Brandon H. Schlomann
(Dated: June 2, 2020)

Recent experiments suggest that the gut microbiome can be thought of as a type of living suspension of three-dimensional bacterial clusters, with the degree of aggregation of particular species having important consequences for intestinal population dynamics and host-microbe interactions. However, a theory of how bacterial-scale dynamics determine global cluster size distributions is lacking. Such a theory would enable researchers to infer hidden aspects of microbiome dynamics by measuring this size distribution, which has become increasingly feasible in a variety of systems. Analyzing imaging-derived data on cluster sizes for several different bacterial strains in the larval zebrafish gut, we find a common family of broad size distributions that decay approximately as power law $p(n) \sim n^{-\mu}$ over multiple decades with exponents $\mu \approx 2$, and then become shallower in a strain-dependent manner. We argue that this form of the size distribution arises from a Yule-Simons-type process in which bacteria grow within clusters and can escape from them, drawing an analogy between gut bacteria and evolutionary dynamics. We deduce a model class that robustly generates size distributions consistent with the data by tuning a single, strain-dependent parameter. Together, these results point to the existence of general, biophysical principles governing the spatial organization of the gut microbiome that may be useful for inferring fast-timescale dynamics that are otherwise inaccessible.

I. INTRODUCTION

Trillions of bacteria inhabit the gastrointestinal tracts of humans and other animals, where they make up some of densest and most diverse microbial ecosystems on Earth. Two central problems concerning these so-called gut microbiota are currently the subject of intense investigation: (1) the determination of microbiota composition—why certain species are more prevalent than others—and (2) understanding the pathogenic and beneficial potential these microbes have on their animal hosts. For both of these problems, the spatial organization of gut bacteria is thought to be an important mediator. In both macroecological and non-gut microbial systems, the spatial organization of organisms is well known to impact both intra- and inter-species interactions [REFs]. Further, within the gut, aspects of spatial organization such as proximity of bacteria to the epithelial boundary can determine the strength of host-microbe interactions [REFs]. Despite these important roles, the spatial organization of bacteria within the intestine remains poorly understood.

Recent advances in the ability to image gut microbial communities in model animals have begun to elucidate some features of bacterial spatial organization common to multiple host species [REFs]. Bacteria in the gut appear to exist largely within three-dimensional, multicellular aggregates, often encased in mucus, whose size range can span several orders of magnitude. This type of spatial structure has been observed in mice, fruit flies, zebrafish, and in human fecal samples [REFs]. However, a thorough understanding of how these structures are generated and their size statistics is lacking.

In contrast to these living systems, powerful theories exist for non-living materials such as colloidal suspensions, emulsions, and polymer gels that connect general features of microscopic dynamics to the global distribution of cluster sizes and large-scale material properties [REFs]. One example is Ostwald ripening, found in many different emulsions, where small particles condense onto the surface of larger ones to minimize surface tension, leading asymptotically to a universal stationary size distribution [REF]. This type of relationship can, in some cases, be analyzed in the reverse direction: presented with an experimentally-measured size distribution that resembles the Ostwald ripening distribution, one could infer that condensation dynamics and surface tension are likely key features of the system. In these types of kinetics-statistics relationships, there is often a trade-off between the generality of the relationship and the degree to which it can be inverted. For example, there are many stochastic processes that can generate so-called $1/f$ noise—fluctuations with a $1/f$ frequency distribution—making the relationship quite general, but this limits what one can learn about a system just by observing this noise pattern [REF]. As a converse example, in systems with exponential size distributions, the characteristic size can provide insight into microscopic dynamics, as is the case, for example, in protein aggregates involved in chemotaxis signalling [wingreen paper].

In this spirit, we investigated the distribution of gut bacterial cluster sizes. In this case, we anticipated

that establishing any type of quantitative relationship between bacterial kinetics and global size distributions would be useful: the presence of truly "universal" size distributions might enable predictions of cluster size statistics across diverse host species, including humans, while system-specific size distributions might facilitate the inference of dynamical parameters from static measurements of cluster sizes. Through analysis of data from larval zebrafish, we found a relationship that falls between the two extremes. As detailed below, we identified a minimal kinetic model that robustly generates the size distributions of diverse bacterial strains by tuning a single parameter: the rate at which bacteria escape from clusters. Due to the basic and general nature of the model's assumptions, we predict that this family of cluster size distributions will manifest in essentially all intestines, and encourage quantitative size measurements in host organisms where measurements are possible. Further, because of the tune-ability of the distribution, we believe that the key dynamical parameters can be reliably inferred from cluster size measurements, which would enable insight into bacterial-scale dynamics that are otherwise impossible to observe.

II. BACKGROUND ON THE DATA

Here, we analyze previously-generated data sets of gut bacterial cluster sizes in larval zebrafish [REFs]. In these experiments (performed by some of the authors), zebrafish are reared devoid of any microbes, or "germ-free", and then mono-associated with a single, fluorescently labelled bacterial species. The full intestines of live hosts are then imaged with light sheet fluorescence microscopy [REFs]. The rapid image acquisition possible with this technique enables reliable identification and enumeration of bacterial cells in the gut through computational image analysis [REF]. Single bacterial cells and multicelluar aggregates are identified separately, and then the number of cells per multicellular aggregate are estimated by dividing the total fluorescence intensity of the aggregate by the mean intensity of single cells [REF]. The output of this analysis is then a list of bacterial cluster sizes (we define single cells as clusters of size 1) and positions within each fish.

We created a dataset that combines lists of cluster sizes from 3 different studies [biophysJ,PNAS,Plos2020], containing data on 9 different bacterial strains. These strains and studies are summarized in Table X. Seven of the strains were previously isolated from healthy zebrafish, and the remaining two are genetically engineered knockout mutants of one of the native strains, *Vibrio* ZWU0020, with knockouts in motility (stator protein deletion $\Delta pomA\Delta pomB$) and chemotaxis (kinase deletion $\Delta cheA$), described in [Motaxis]. All strains are of the Proteobacteria phylum [Wiles-mBio].

In all of these studies, it was observed that while the formation of 3D aggregates was common across the different types of bacteria, strains differed in the degree of aggregation, with some retaining larger fractions of individual, planktonic cells, than others. This variation in aggregation is typically quantified by computing metrics like the fraction of the population contained in planktonic cells, the average cluster size, and the total number of clusters. The relationship between aggregation and location along the gut was explored in [biophysJ]. In another study, [ABX], the distribution of clusters sizes was measured for one of the strains in Table X, *Enterobacter* ZOR0014. This study introduced a stochastic simulation model of bacterial cluster kinetics that, when parameterized with separate measurements, quantitatively reproduced the experimentally measured clusters size distribution. However, a theoretical understanding of the form of this distribution was lacking. Furthermore, it was unclear whether this size distribution was unique to *Enterobacter* ZOR0014, or if other strains exhibited similar distributions. Here, we address both of these issues.

III. RESULTS

A. Different bacterial species have qualitatively similar cluster size distributions within the larval zebrafish intestine

We performed a meta-analysis of gut bacterial cluster sizes measured in three previous studies [REFs]. We computed the probability density of cluster sizes, $p(n)$, for each animal using logarithmically-spaced bins. Sizes were rounded up to integer values. Log_{10}-transformed distributions for individual animals are shown as small circles connected by lines in Fig. 1, with each panel corresponding to a different bacterial strain. To estimate an underlying probability distribution for each strain given these replicates, we pooled the clusters sizes from different animals and estimated a normalized probability density using the same bins as for individual animals. This

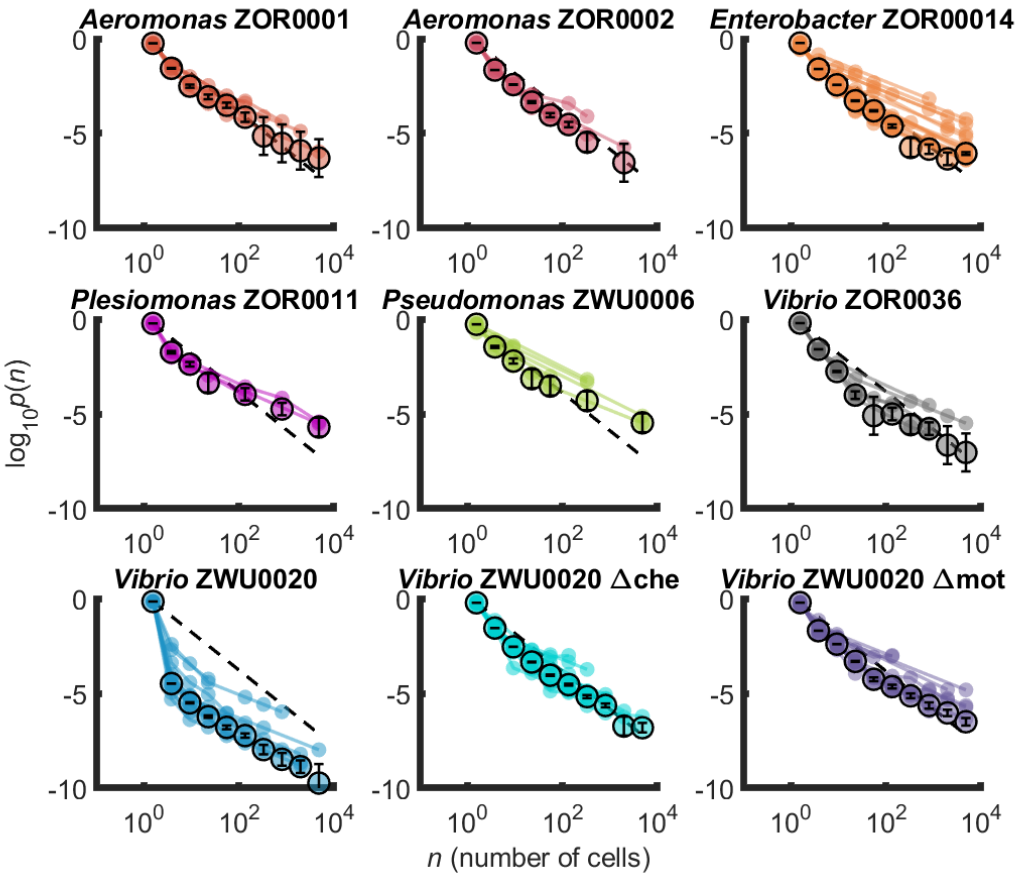

FIG. 1. Caption.
more caption

pooled distribution is shown as the large circles in Fig. 1. Due to the nature of how pooling deals with low-cluster-number systems, the pooled distribution appears below much of the host-host spread (Methods). However, in stochastic simulations of models with known analytic solutions in the large-system limit, we found that pooling more accurately reproduces these results than does averaging individual distributions (Supp Fig). Error bars on the pooled distribution represent Poisson noise of counts within each bin (Methods).

We find broad distributions across all strains known to reliably form aggregates. For comparison, for each strain we overlay a dashed line representing $p(n) \sim n^{-2}$, with the vertical offset fixed to the first bin's probability density. Strains vary in their deviation from this line. Both *Aeromonas* ZOR0001 and *Aeromonas* ZOR0002 follow this line quite closely, while highly aggregated strains, such as *Enterobacter* ZOR0014, have a large-size tail that decays more slowly. At the other end of the spectrum, the notable *Vibrio* ZWU0020, which exists almost entirely as highly motile, planktonic cells [REFs], has a distribution that falls below the n^{-2} line. However, mutant versions of *Vibrio* that lack chemotaxis or motility (Fig 1, bottom row) exhibit enhanced aggregation [Motaxis], resulting in shallower size distributions that closely resemble other, native bacteria, such as *Aeromonas* ZOR0001 and *Aeromonas* ZOR0002.

Together, these observations suggest that generic processes, rather than strain-specific ones, largely govern this aspect of spatial organization within the intestine. To test this idea, we sought to explore the space of kinetic models that contain processes known from time-lapse imaging to occur in the larval zebrafish gut [REFs] and to identify the requirements for generating cluster size distributions consistent with the data.

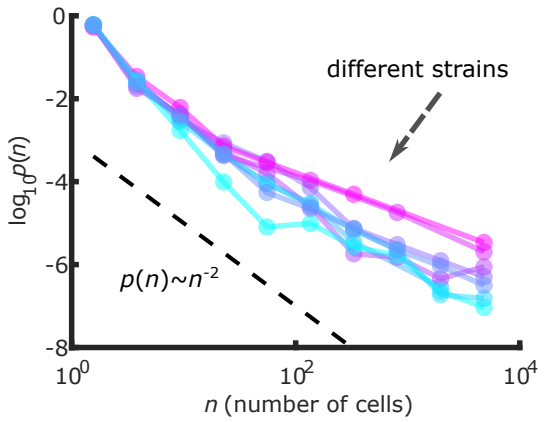

FIG. 2. Caption.
more caption

B. Simplified bacterial cluster dynamics are analogous to classic population genetics models

Previous time-lapse imaging of bacteria in the zebrafish intestine revealed four core processes that can alter bacterial cluster sizes: (1) clusters can increase in size due to cell division, a process we refer to as "growth"; (2) clusters can decrease in size as bacteria break out of them, a process we refer to as "fragmentation" and believe to be linked to the cell division process; (3) clusters can increase in size by joining with another cluster during fluid mixing, a process we refer to as "aggregation"; and (4) clusters can be removed from the system by transiting down and out of the intestine, a process we refer to as "expulsion".

Focusing first on just the growth and fragmentation processes, we identify a mapping to a well-studied stochastic process known as the Yule-Simons process, which is commonly used to understand allele frequency distributions in population genetics [REFs]. In the Yule-Simons process, organisms replicate asexually with rate r, and each replication has a probability ϵ of generating a neutral mutation. Exact calculations [Yule] and well-known heuristic ones [Hallatschek, WalzcakReview] show that at long times the distribution of clone frequencies is stationary with a power-law tail $p(x) \sim x^{-\mu}$ with exponent $\mu = (1 + (1 - \epsilon)^{-1})^{-1}$ that becomes $\mu \approx 2$ for rare mutation.

In the analogy with gut bacteria, the size of mutant clones is analogous to the size of a (physical) bacterial cluster, and the mutation process that generates new clones is analogous to the fragmentation process that generates new clusters (Fig. 3A). In applying this model to gut bacteria, we slightly modify the Yule-Simons process by decoupling fragmentation and growth for each individual cell division event. Instead of cells dividing and then becoming a mutant with probability ϵ, we have separate growth and fragmentation events: cells divide within clusters at a mean rate r and break out of them with a mean rate β. The asymptotic distribution of this process is the same as that for the Yule-Simons process, with $\epsilon = \beta/r$. We performed stochastic simulations of this process using a Poisson τ-leaping algorithm and found a size distribution consistent with the known analytic results (Fig. 3B). The cluster size distributions asymptotically have a power-law form with an exponent of $\mu \approx 2$ for $\beta/r \ll 1$ and then increases as β is increased.

This model is a good candidate for a minimal process that can be used to understand gut bacterial cluster size statistics because the dynamics are consistent with experimental observations and it offers a robust mechanisms for generating power law distributions with exponent $\mu \approx 2$. However, some of the experimentally measured distributions in Fig. 1 indicate valuse of μ less than 2, which cannot be generated by the Yule-Simons process. Moreover, in this rare fragmentation limit with small values of μ, the Yule-Simons process has no mechanism for generating strain-strain differences in the exponent. Therefore, we sought modifications to the Yule-Simons process that could produce these features.

C. Spatially-restricted fragmentation generates a power law size distribution with a rate-dependent exponent

The first feature of the Yule-Simons process that we revisited was the assumption that all cells in a cluster are equally likely to fragment. We posit that the geometry of gut bacterial clusters makes cells in the center of the cluster less likely to fragment than cells at the edge. We model this confinement with the parameter ν, which sets how the fragmentation rate scales with cluster size, n, with the rate being proportional to n^{ν}. Letting c_n denote the number of clusters of size n, the dynamics of this model are governed by the master equation

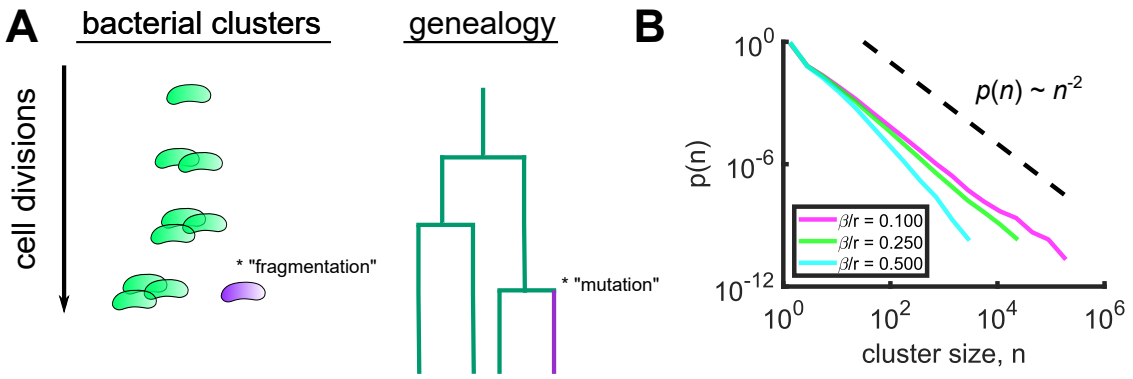

FIG. 3. Caption.
more caption

$$\dot{c}_n = \beta\left((n+1)^\nu c_{n+1} - n^\nu c_n + \delta_{n,1}\sum_m m^\nu c_m\right) + r\left((n-1)c_{n-1} - nc_n\right). \quad (1)$$

A value of $\nu = 2/3$ corresponds to only the surface of a cluster being able to fragment; a value of $\nu = 1$ corresponds to the Yule-Simons process. For simplicity, we study the extreme but analytically tractable case of $\nu = 0$. This choice can be interpreted as approximating a long and thin cylindrical cluster where fragmentation is more likely to happen on the ends. Such clusters are indeed observed in larval zebrafish gut, where they like form through the secretion of mucus radially inward from the intestinal wall.

We computed the long-time, large-size features of the size distribution for this model using a heuristic analytic calculation that is analogous to one used for the Yule-Simons process (Appendix) [REFs]. We found an asymptotic distribution with power law tail, $p(n) \sim n^{-\mu}$, with a rate-dependent exponent that is the solution to the transcedental equation

$$\mu = 1 + \frac{\beta}{r}\left(1 - \frac{1}{\zeta(\mu)}\right), \quad (2)$$

with $\zeta(\mu)$ the Riemann zeta function. Solving for μ numerically (Methods), we find that this analytic results agrees well with stochastic simulations (Fig. X).

These results indicate that spatially-restricted fragmentation is a possible mechanism for generating

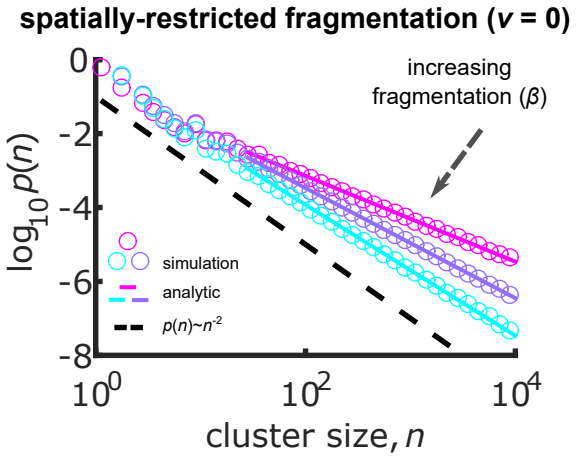

FIG. 4. Caption.
more caption

the range of heavy-tailed size distributions observed in the data. However, the measured distributions are not perfect power laws, but become even shallower for large cluster sizes (Fig. X). Therefore, we constructed a generalized model that includes additional process.

D. Generalized model

Our generalized model, a variant of which was first introduced in [ABX], builds on equation X in 3 ways. First, we add density-dependent growth, $r \to r(1 - N(t)/K)$ with $N(t)$ the total abundance at time t, and K the global carrying capacity. Since we believe that fragmentation is linked to the cell division process, we let this density-dependence also regulate fragmentation, with $\beta \to \beta(1 - N(t)/K)$.

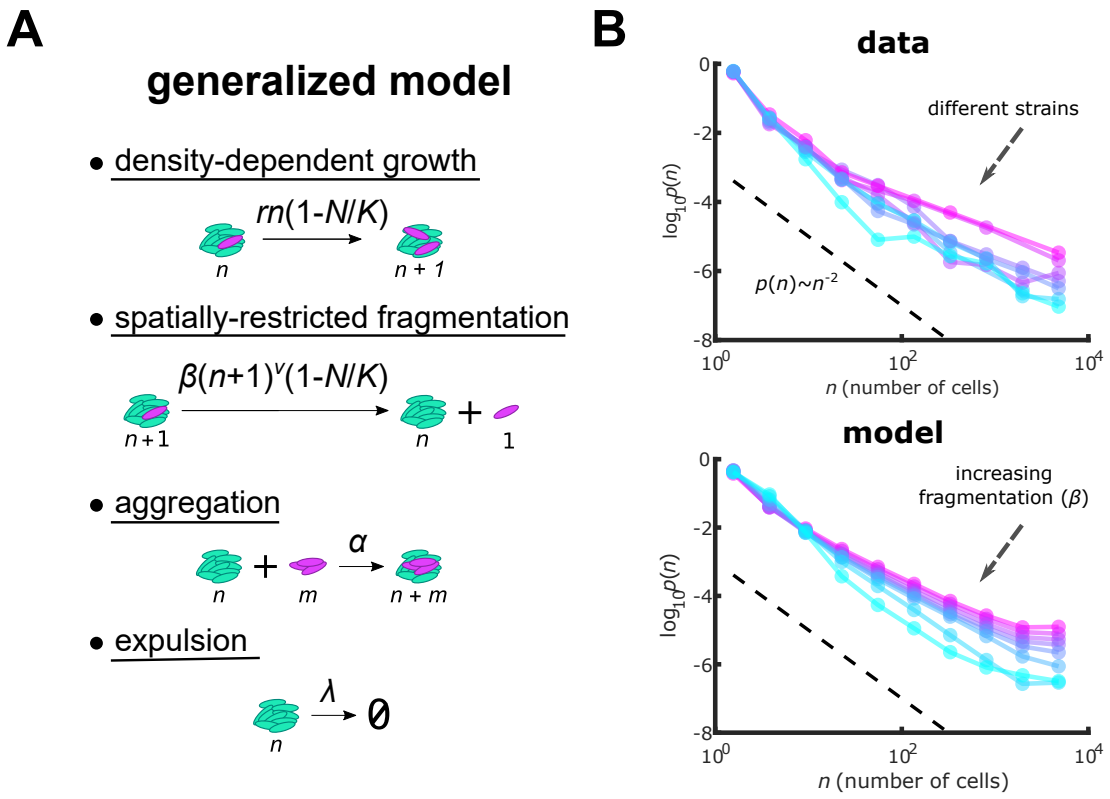

FIG. 5. Caption.
more caption

This modification of the fragmentation rate was not included in the analysis of [ABX], but we feel that it is a more realistic modeling choice. Second, we add an aggregation process, whereby two cluster join together to form a larger cluster. This interaction has been directly observed in zebrafish [ABX]. Since we believe that growth, rather than aggregation, is the dominant process for creating large clusters, we model a weak aggregation process, where the aggregation rate, α, is independent of cluster size [RednerBook]. Finally, we add an expulsion process, by which clusters are removed from the system at a size-independent rate, λ. While this assumption of size-independence is likely incorrect, previous work showed that different choices had little effect on the size distribution; it is simply the rate at which large ($\mathcal{O}(K)$) clusters are expelled that matters. Together, these additions to the model allow for a non-trivial steady state, and we model the experimental data in this limit, allowing us to ignore time as additional parameter.

With these choices, the master equation becomes

$$\dot{c}_n = \frac{\alpha}{2}\sum_{m=1}^{n} c_{n-m}c_m - \alpha c_n M + r\left(1-\frac{N}{K}\right)\left[(n-1)c_{n-1} - nc_n\right] - \lambda c_n$$
$$+\beta\left(1-\frac{N}{K}\right)\left((n+1)^\nu c_{n+1} - n^\nu c_n + \delta_{n,1}\sum_m m^\nu c_m\right).$$

Here, $M = \sum_n c_n$ denotes the total number of clusters, and $N = \sum_n nc_n$ denotes the total number

of cells. As we are primarily interested in finite systems, where stochasticity can be important, we focus on stochastic simulations of this model.

We added weak aggregation and expulsion set by exp. values. We found that the model reproduces the range of measured distributions just by sweeping β. Other supp results.

IV. DISCUSSION

V. METHODS

A. Data

B. Stochastic simulations

The bacterial populations in the larval zebrafish gut whose size statistics are plotted in Fig. 1 vary substantially in the total number of clusters present in a given animal, with numbers ranging from $\mathcal{O}(1)$ to $\mathcal{O}(10^5)$. To efficiently simulate the model in equation (?) for such a wide range of system sizes, we implemented a hybrid algorithm that combines a Langevin growth model with Poisson τ-leaping for the remaining processes [tauRef]. This simulation builds off of a previous implementation [ABX], with minor modifications.

In brief, a simple fixed τ scheme is used to increment time. Within each time step, first all clusters grow according to a logistic growth equation with demographic stochasticity,

$$\frac{dn_i}{dt} = rn_i\left(1-\frac{N}{K}\right) + \sigma\sqrt{rn_i\left(1-\frac{N}{K}\right)}\eta(t). \quad (3)$$

Here, n_i is the size of the i^{th} cluster, $N = \sum_i n_i$ is the total population abundance, and $\eta(t)$ is Gaussian white noise. The parameter σ is used to toggle between deterministic $\sigma = 0$ and stochastic $\sigma = 1$ growth. For stochastic growth, this model is numerically integrated with a straightforward Euler-Marayama scheme with its own timestep Δt:

$$n_i(t+\Delta t) = n_i(t) + \Delta t \cdot rn_i\left(1-\frac{N}{K}\right) + \sqrt{\max\left(\Delta t \cdot rn_i\left(1-\frac{N}{K}\right), 0\right)}\mathcal{N}(0,1), \quad (4)$$

with $\mathcal{N}(0,1)$ a standard Gaussian random number. For deterministic growth, the cluster sizes are updated with the analytic solution of equation (?).

After updating sizes