A minimal model of gut bacterial cluster dynamics

Brandon H. Schlomann

March 6, 2020

We consider a kinetic model of the dynamics of bacterial clusters in the intestine. The model is non-spatial. For small systems, the main object of the model is a set of cluster sizes, $\{n_j\}$, with $j = 1, 2, 3, ..., M(t)$ and $M(t)$ the number of clusters in the system at time t. These small systems are studied via direct stochastic simulation. In the large system limit, we consider the concentration of clusters of size n denoted by $c_n(t)$.

The model has 4 rate processes that govern its dynamics:

1. Aggregation: a cluster of size n and another of size m can aggregate to form a cluster of size $n + m$ according to the aggregation kernel A_{nm}.

2. Fragmentation: a cluster of size $n + m \geq 2$ can break apart into two smaller clusters of size n and m according to the fragmentation kernel F_{nm}.

3. Growth: a cluster of size n can grow to size $n + 1$ due to cell division according to the growth kernel G_n.

4. Expulsion: a cluster of size n is removed from the system, according to the expulsion kernel E_n, modeling intestinal transport.

Assembling these processes, the time evolution of the cluster size distribution is given for large systems by the general master equation

$$\dot{c}_n = \frac{1}{2} \sum_{m=1}^{n} A_{n-m,m} c_{n-m} c_m - c_n \sum_{m=1}^{\infty} A_{n,m} c_m$$
$$+ \sum_{m=n}^{\infty} F_{m-n,n} c_{m-n} - \frac{1}{2} \sum_{m=1}^{n} F_{n-m,m} c_n$$
$$+ G_{n-1} c_{n-1} - G_n c_n - E_n c_n \tag{1}$$

Based on experimental data, we assume the following forms for the rate kernels:

1

- Aggregation: we use a constant kernel

$$A_{n,m} = \alpha. \tag{2}$$

- Fragmentation: we use a "chipping" kernel, in which the only fragmentation process that occurs is the breaking off of single cells ("monomers"). We further use a constant kernel for the overall rate,

$$F_{n,m} = \beta \delta_{m,1} \tag{3}$$

- Growth: we use a logistic growth kernel with global carrying capacity, K,

$$G_n = rn \left(1 - \frac{N}{K} \right) \tag{4}$$

  with $N = \sum_m m c_m$ the total cell density.

- Expulsion: we use a constant kernel

$$E_n = \lambda \tag{5}$$

With these choices, the model master equation becomes

$$\dot{c}_n = \frac{\alpha}{2} \sum_{m=1}^{n} c_{n-m} c_m - \alpha c_n M + \beta(c_{n+1} - c_n) + \beta \delta_{n,1} M$$

$$+ r \left(1 - \frac{N}{K} \right) \left[(n-1)c_{n-1} - nc_n \right] - \lambda c_n. \tag{6}$$

Analytic results for constant kernels

We will now derive some analytic results for the model with constant kernels in the large system limit.

Moment equations

As is often useful in these types of problems, we first derive equations for the zeroth and first moments, which correspond to the density of clusters (M) and the density of total cells (N) respectively. The density of clusters follows

$$\dot{M} = -\frac{\alpha}{2}M^2 + (\beta - \lambda)M - \beta c_1. \tag{7}$$

The equation for M is coupled to c_1, the concentration of single cells, and so is not closed. The total cell (population) density follows

$$\dot{N} = rN\left(1 - \frac{N}{K}\right) - \lambda N \tag{8}$$

The evolution of N is independent of the cluster size distribution. This is contrary to the intuition that the cluster size distribution modulates how expulsion events impact the total population, and therefore may be a sign that the thermodynamic limit is not valid for comparing to our experimental systems.

The equation for N is closed and has a steady state

$$N = K\left(1 - \frac{\lambda}{r}\right). \tag{9}$$

Asymptotic, steady state cluster size distribution

We wish to compute the form of the steady state cluster size distribution c_n for large sizes $n \gg 1$. To do this, we use a generating function approach (Methods). With the kernels chosen as above, the model builds directly off of a reversible polymerization model considered in (Krapivsky and Redner, 1993), adding growth and expulsion terms. Following their approach, we first construct the generating function $g(z,t) = \sum_n z^n c_n$. This generating function satisfies

$$\partial_t g = \frac{\alpha}{2}g^2 - \alpha gM + \beta\left(\frac{g}{z} - c_1\right) - (\beta + \lambda)g + \beta zM + r\left(1 - \frac{N}{K}\right)z(z-1)\partial_z g. \tag{10}$$

Next, we change variables to the shifted generating function $h(z,t) = g(z,t) - g(1,t) = g(z,t) - M(t)$, which satifies

$$
\begin{aligned}
\dot{h} &= -\dot{M} + \dot{g} \\
&= \frac{\alpha}{2}M^2 - (\beta - \lambda)M + \beta c_1 \\
&\quad + \frac{\alpha}{2}(h+M)^2 - \alpha(h+M)M + \beta\left(\frac{(h+M)}{z} - c_1\right) \\
&\quad - (\beta + \lambda)(h+M) + \beta z M + r\left(1 - \frac{N}{K}\right)z(z-1)\partial_z h \\
\dot{h} &= \frac{\alpha}{2}h^2 + \beta\left[\frac{(1-z)}{z} - \frac{\lambda}{\beta}\right]h + \beta\frac{(1-z)^2}{z}M - r\left(1 - \frac{N}{K}\right)z(1-z)\partial_z h. \qquad (11)
\end{aligned}
$$

Note that the c_1 terms have been eliminated. In the steady state, this reduces to the single ODE

$$
r\left(1 - \frac{N}{K}\right)z(1-z)\frac{dh}{dz} = \frac{\alpha}{2}h^2 + \beta\left[\frac{(1-z)}{z} - \frac{\lambda}{\beta}\right]h + \beta\frac{(1-z)^2}{z}M. \qquad (12)
$$

Using the result from the moment equation for cell density, $r(1 - N/K) = \lambda$, we eliminate the growth rate to obtain

$$
\lambda z(1-z)\frac{dh}{dz} = \frac{\alpha}{2}h^2 + \beta\left[\frac{(1-z)}{z} - \frac{\lambda}{\beta}\right]h + \beta\frac{(1-z)^2}{z}M. \qquad (13)
$$

Following from a fundamental result for generating functions, there is the correspondence between the behavior of $h(z)$ in the limit $z \to 1$ and c_n in the limit $n \to \infty$ (Methods). Specifically, suppose c_n has a power law tail with exponent $\mu > 1$. Then we have the correspondence (Methods)

$$
h(z) \sim (1-z)^{\mu-1} \iff c_n \sim n^{-\mu}. \qquad (14)
$$

Therefore, we wish to solve the steady state equation for h in the limit $z \to 1$, supplemented with the initial condition that $h'(0) = N$, a constraint that follows from the definition of the generating function. Defining $\epsilon = 1 - z$, the differential equation for h becomes

$$
\frac{dh}{d\epsilon} = -\frac{\alpha}{2\lambda\epsilon(1-\epsilon)}h^2 - \left(\frac{\beta}{\lambda}\frac{1}{(1-\epsilon)^2} - \frac{1}{\epsilon(1-\epsilon)}\right)h - \frac{\beta M\epsilon}{\lambda(1-\epsilon)^2}. \qquad (15)
$$

This equation is of the Riccati form, and so it's possible that it can be solved exactly. However, we'll just focus on trying to pull out the large-size asymptotics of the cluster size distribution. This limit corresponds to $\epsilon \to 0$ (see Methods section). By construction, $h \to 0$

as $\epsilon \to 0$; we want to calculate how it goes to zero. The question then, is which terms are relevant in this limit? The leading order term on the right hand side is of order h/ϵ:

$$\frac{dh}{d\epsilon} \approx \frac{h}{\epsilon}, \tag{16}$$

which leads quickly to the solution

$$h(\epsilon) \sim N\epsilon. \tag{17}$$

This then means that the steady state cluster size distribution goes as

$$c_n \sim n^{-2}. \tag{18}$$

Note that none of the rate parameters appear in the leading order differential equation. This suggests that there is only one steady state phase, but it's possible that this breaks down when higher order terms are included in the solution.

Mapping to evolutionary dynamics

Power law size distributions with slope -2 also arise in evolutionary dynamics models of neutral mutations in expanding populations. Specifically, cluster size is analogous to allele frequency, and fragmentation is analogous to neutral mutation. Consider a reduced version of the kinetic model that only includes exponential growth and fragmentation:

$$\dot{c}_n = \beta(c_{n+1} - c_n) + \beta\delta_{n,1}M + r\left[(n-1)c_{n-1} - nc_n\right] \tag{19}$$

The evolutionary interpretation of this model is an exponentially growing population where the probability of a mutation is the same for each family, not each individual. While artificial, it turns out that this assumption doesn't change the scaling.

Following the same procedure as above, the zeroth and first moment equations are

$$\dot{M} = \beta(M - c_1), \tag{20}$$

$$\dot{N} = rN. \tag{21}$$

The generating function $g(z,t) = \sum_n z^n c_n$ follows

$$\partial_t g = \beta \left(\frac{g}{z} - c_1 \right) - \beta g + \beta z M + rz(z-1)\partial_z g. \tag{22}$$

The modified generating function $h(z,t) = g(z,t) - g(1,t) = g(z,t) - M(t)$ follows

$$
\begin{aligned}
\partial_t h &= -\partial_t M + \partial_t g \\
&= -\beta(M - c_1) + \beta \left(\frac{(h+M)}{z} - c_1 \right) - \beta(h+M) + \beta z M + rz(z-1)\partial_z h \\
&= \beta \frac{(1-z)}{z} h + \beta \frac{(1-z)^2}{z} M - rz(1-z)\partial_z h.
\end{aligned} \tag{23}
$$

Changing variables to $\epsilon = 1 - z$,

$$\partial_t h = \beta \frac{\epsilon}{1-\epsilon} h + \beta \frac{\epsilon^2}{1-\epsilon} M + r\epsilon(1-\epsilon)\partial_\epsilon h. \tag{24}$$

Since growth is exponential, there is no steady state, but we can solve this PDE in the limit $\epsilon \to 0$. To leading order in ϵ, we just have the derivative term,

$$\partial_t h \approx r\epsilon \partial_\epsilon h. \tag{25}$$

This equation is easily solved by separation of variables: let $h(\epsilon, t) \equiv E(\epsilon)T(t)$. Then, we have

$$\frac{1}{T} \frac{dT}{dt} = \frac{r\epsilon}{E} \frac{dE}{d\epsilon} = \text{const.} \tag{26}$$

Since we know that the population grows exponentially with growth rate r, the separation constant must be r. This leads to an ODE for the ϵ dependence

$$\frac{dE}{d\epsilon} = \frac{E}{\epsilon}, \tag{27}$$

exactly the same as equation as we found for the steady state solution of the full model (equation 16). The solution is

$$E \sim \epsilon, \tag{28}$$

which means that, just like in the full kinetic model above,

$$c_n \sim n^{-2}. \tag{29}$$

From this calculation, we conclude that growth and fragmentation are the fundamental processes that determine the scaling of the cluster size distribution in the full model. The other processes in the full model are still important, however, because they allow for a well defined, non-trivial steady state to be reached. A carrying capacity allows the cell density to reach a steady state. Aggregation allows, in the thermodynamic limit, non-zero concentrations of sizes larger than K, which would otherwise have no mechanism to be generated. It's also possible that, like the non-living counterpart to the present model ($r = \lambda = 0$), there exists a critical aggregation rate above which a steady state no longer exists (Krapivsky and Redner (1993)); this remains unclear however. Finally, expulsion isn't required to reach a steady state, but does allow for transition to extinction, which is experimentally relevant.

Interestingly, the expulsion process can be interpreted in an evolutionary context as the abrupt extinction of a lineage, due, for example, to a rapid fluctuation in environmental conditions that makes a genotype unfit. The cluster model with expulsion can therefore perhaps be interpreted as a model of evolution in a fluctuating environment with a sharp response function. The evolutionary interpretation of aggregation is a bit more bizarre. Aggregation corresponds to a coalescing of lineages, but in real, forward time. In other words, all members of a genotype instantly converting to another.

Heuristic calculation that gives a different answer

Start from the moment equations:

$$\dot{M} = \beta(M - c_1), \tag{30}$$

$$\dot{N} = rN. \tag{31}$$

Ignoring the c_1 term, we then have

$$N(t) \sim M^{r/\beta} \tag{32}$$

Following the derivation for the Yule-Simons process (see, e.g., arXiv:1907:03891), we note that a single cell that fragmented off of another cluster at time t', will have a size at time t of

$$n(t, t') \sim e^{r(t-t')}, \tag{33}$$

or in terms of a scaled size, or frequency, $n^* = n/e^{rt}$,

$$n^* \sim M(t')^{-r/\beta}. \tag{34}$$

Interpreting M as a rank, we recognize Ziph's law relating frequency and rank, which implies a probability density

$$p(n^*) \sim (n^*)^{-(1+\beta/r)}. \tag{35}$$

Generalization to other forms of chipping

If we generalize to a size-dependent fragmentation rate of the form $F_{nm} = \delta_{m,1}\beta n^\nu$, the equation for the first moment is of the form

$$\dot{M} = \beta(M_\nu - c_1) \tag{36}$$

with fractional moment M_ν defined as

$$M_\nu = \sum_n n^\nu c_n. \tag{37}$$

This form contains as special cases the constant kernel, $\nu = 0$, and the classic Yule-Simons process, $\nu = 1$. For this latter case we have $M \sim N$, but the growth rate of a cluster is reduced by the fragementation rate, since cells can be lost,

$$n(t, t') \sim e^{(r-\beta)(t-t')}, \tag{38}$$

so the scaled size scales as

$$n* \sim M^{1-\beta/r} \tag{39}$$

leading to

$$p(n^*) \sim (n^*)^{-\left(1+\frac{1}{1-\beta/r}\right)}. \tag{40}$$

8

For rare fragmentation events, $\beta \ll r$, this limits to

$$p(n^*) \sim (n^*)^{-2}. \tag{41}$$

Estimates of the extinction transition

Near the extinction transition, the system fluctuations between a state of one cluster and a state of 2 or more clusters. To probe the transition, we will therefore consider a toy model with a state space of just two clusters, which can take on two sizes: "small", representing a single cell, and "large", representing a multicellular cluster capable of fragmenting. We label the state of system with a pair of binary variables encoding the presence/absence of the {small, large} clusters respectively. There are 3 relevant states of the system: $\{1,0\}$, $\{0,1\}$, and $\{1,1\}$, which we labels 1, 2, and 3 respectively. To probe the extinction transition, we compute when the probability of have just one cluster in the system goes to 1, at which point we assume that the system will go extinct for any non-zero λ. Using the same variable conventions as above, the probability distribution over this simplified state space is governed by

$$\dot{P}_1(t) = \lambda P_3 - r P_1 \tag{42}$$
$$\dot{P}_2(t) = r P_1 + (\lambda + \alpha) P_3 - \beta P_2 \tag{43}$$
$$\dot{P}_3(t) = \beta P_2 - (2\lambda + \alpha) P_3 \tag{44}$$

We then solve for the steady state and look at when $P_1 + P_2 \to 1$, or equivalently when $P_3 \to 0$. We find that

$$P_3 = \frac{1}{1 + \lambda/r + (2\lambda + \alpha)/\beta} \tag{45}$$

Using this expression, we can derive simple estimates of the extinction boundary is various limits. For dispersed systems driven to extinction by reducing growth, such as planktonic populations treated with lethal levels of antibiotics, we can ignore the $(2\lambda + \alpha)/\beta$ term and infer that the transition to extinction occurs approximately when

$$\frac{\lambda}{r} \approx 1, \tag{46}$$

or when growth becomes as slow as expulsion. This is equivalent to the prediction of the deterministic theory given above, for which the total population density goes to zero when $r = \lambda$. That the deterministic and stochastic theories agree here reflects the fact that

9

dispersed populations exhibit only small fluctuations in abundance compared to aggregated populations, which can abruptly lose massive clusters.

In the opposite limit of strongly aggregated, fast growing populations, we can ignore the λ/r term and infer that the extinction transition occurs when

$$\frac{\beta}{2\lambda + \alpha} \approx 1. \tag{47}$$

This transition is qualitatively intuitive, in the sense that we expect a transition at the crossover between the timescale of generating clusters and the timescale of losing clusters, but cannot be derived from the deterministic theory because the equation of the density of cluster M is not closed.

.... By analogy, test this hypothesis:

$$\frac{2\lambda + \alpha}{\beta} \approx -\ln f \tag{48}$$

....

Open problems

1. In stochastic simulations of the full cluster process, and also in experimental data, we observe the cluster size distribution becoming shallow for more aggregated populations. This is likely a finite size effect—a depletion of mid-sized clusters in favor of a single massive cluster—but is probably also driven by the aggregation process. Can we quantitatively understand this effect?

2. Can we construct a well-controlled mapping from cluster kinetics to total population dynamics, including for finite populations? Specifically, we previously studied a class of models of logistic growth coupled to random catastrophes to describe population dynamics in the gut in the presence of expulsion. This model has 4 parameters: (r, K, λ, f), where the first 3 parameters are the same as in the cluster model and f is the fraction of the population remaining after collapse. Specifically, the model is specified by a stochastic differential equation

$$dX_t = rX_t\left(1 - \frac{X_t}{K}\right) - (1-f)X_{t^-}d\Pi_t(\lambda) \tag{49}$$

where X_t is the population abundance and $\Pi_t(\lambda)$ is a standard Poisson process with rate λ. What is the mapping from cluster model parameters $(\alpha, \beta, r, \lambda, K)$ to collapse fraction, f? In principle, one could try to construct a collapse model with f a random variable characterized by a distribution, $P(f) = P$(a cluster constitutes a fraction $1 - f$ of the population). Perhaps a simpler construction would be to compute a typical collapse fraction, $\langle f \rangle$ or $\langle \ln f \rangle$.

3. With a model of the total abundance dynamics, where is the extinction transition? For collapse model described above, we previously showed that extinction occurs at

$$\frac{\lambda \ln f}{r} = -1. \tag{50}$$

Given a mapping from cluster model parameters to f, does this criterion accurately reproduce numerical abundance phase diagrams?

Methods: asymptotic distribution from generating function

Suppose the cluster size distribution has a power law tail $c_n \sim n^{-\mu}$ with $\mu > 1$ so $\sum_n c_n$ converges. The connection to the generating function $g(z) = \sum_n z^n c_n$ is the following. First, let's move to a continuum approximation and replace the generating function with a Laplace transform:

$$g(s) = \int dn \, c(n) \, e^{-sn} \tag{51}$$

with $z = \exp(-s)$. With this new notation, the limits $z \to 1$ and $s \to 0$ coincide. We re-write this integral as

$$
\begin{aligned}
g(s) &= \int dn \, c(n) \left(1 - \left(1 - e^{-sn}\right)\right) \\
&= \int dn \, c(n) - \int dn \, c(n) \left(1 - e^{-sn}\right) \\
&= M - \int dn \, c(n) \left(1 - e^{-sn}\right).
\end{aligned} \tag{52}
$$

In the second integral, since $s \ll 1$, the integrand will be close to zero approximately until $n \sim 1/s$. At that point, we may use the assumed scaling form of c_n. We then approximate the exponential cutoff function by simply replacing the lower bound of the integral with $1/s$:

$$g(s) - M \sim - \int_{1/s}^{\infty} dn \, n^{-\mu} = s^{\mu-1}. \tag{53}$$

So, using a shifted generating function $h(z) = g(z) - g(1) = g(z) - M$, we see that there is a direct correspondence

$$h(s) \sim s^{\mu-1} \Longleftrightarrow c_n \sim n^{-\mu}, \tag{54}$$

or, in terms of the discrete definition of the generating function,

$$h(z) \sim (1 - z)^{\mu-1} \iff c_n \sim n^{-\mu}. \qquad (55)$$