Shishi Luo – Research statement

Stochastic processes and differential equations provide a rich source of tractable objects for capturing the dynamics of complex systems. In my research, I work closely with mathematicians (Jonathan Mattingly, Michael Reed, and other members of the Duke Math Department) as well as biologists (Katia Koelle's research group) to use these objects to understand important biological questions:

- Evolutionary dynamics under multiple and opposing levels of selection (§1)
  - multiscale stochastic processes
  - measure-valued (Fleming-Viot) processes
  - deterministic and stochastic diffusion limits
- Viral antigenic evolution within and between hosts (§2)
  - inhomogeneous Poisson processes
  - nonlinear differential equations

This has further motivated new stochastic processes approaches to open questions in evolutionary biology (§3.1 and 3.2).

1 Current research: Antagonistic multilevel selection

The evolutionary dynamics of disease-causing pathogens occur both within each infected host and between hosts via transmission. A fast-replicating pathogen can outcompete other pathogens within the same host, but may be less transmissible due to the damage it causes the host [1, 2]. Similarly, in the evolution of cooperation, cheaters outcompete cooperators at the individual level while groups with fewer cheaters/more cooperators are favored at the group level [3, 4]. Although these examples arise in different areas of biology, their fundamentally similar structure – selection acting antagonistically at different levels – suggests there may be a unifying formalism which captures their essential characteristics.

I propose that at a fundamental level, these systems can be described by a stochastic process with evolutionary dynamics, modeled by Moran processes, occurring *concurrently* on two scales. Suppose that type A individuals replicate faster than type a individuals, but that groups with a greater fraction of type a individuals replicate faster than groups with a lower fraction of type a individuals. Type A individuals may be thought of as highly virulent pathogen strains, which, through faster replication, outcompete the less virulent type a individuals at the within-host level. However, type a individuals, by replicating more slowly, inflict less damage upon the host, so hosts that are infected with more type a individuals are more likely to transmit the infection to new hosts than hosts infected with fewer type a individuals.

Assume that a population contains m groups, each containing n individuals (m and n are fixed). Let this be denoted by the stochastic process $X(t) = (X_1(t), X_2(t), \ldots, X_m(t))$, where $X_k(t)$ is the number of type a individuals in group k. I show that by considering the empirical measure, $\mu_t^{m,n} = \frac{1}{m}\sum_{k=1}^{m} \delta_{X_k(t)}$, this stochastic process can be visualized as a ball-and-urn system (Fig. 1). This reformulation significantly simplifies the analysis of the system because events at *different* scales (between and within groups) are mapped to events at the *same* scale (balls moving between urns).

1.1 Implications for the field of stochastic processes

As a mathematical object, $\mu^{m,n}(t)$ is an example of a measure-valued process with interesting properties:

1. As $m \to \infty$, $\mu_t^{m,n}$ converges to μ_t^n, which solves either a system of ordinary differential equations (ODEs) or stochastic differential equations (SDEs).

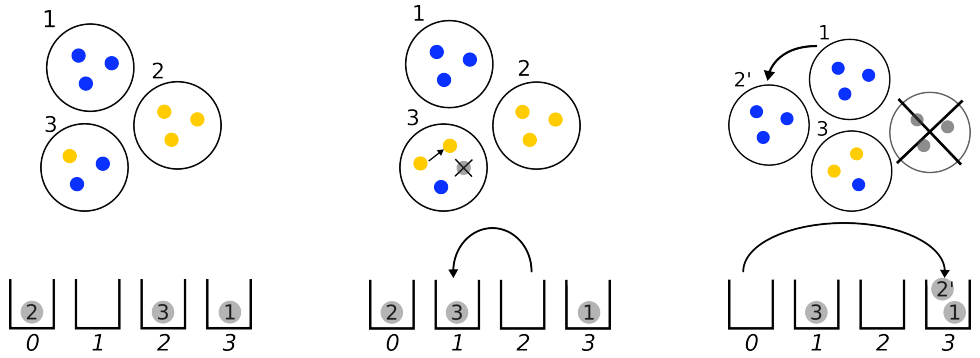

Fig. 1: *Left:* Group 1 has three type a (blue) individuals, represented by ball 1 in urn 3. Similarly, group 2 has zero blue individuals and group 3 has 2 blue individuals. *Middle:* An individual-level replication decreases the number of blue individuals in group 3 from two to one: ball 3 moves to urn 1. *Right:* A group-level replication event removes group 2, which has zero blue individuals, and introduces a group with three blue individuals: ball 2 leaves urn 0 and appears in urn 3 as ball 2'.

2. As $m, n \to \infty$, $\mu_t^{m,n}$ converges to μ_t, which can be understood as a probability density, characterized either as:
   (a) the solution to the integro-partial differential equation

$$\frac{\partial}{\partial t}\mu_t(x) = \frac{\partial}{\partial x}\left[x(1-x)\mu_t(x)\right] - \rho\left(x - \int_0^1 y\mu_t(y)dy\right) \tag{1}$$

   (b) a Fleming-Viot process which is a solution to the following martingale problem:

$$\langle\phi, \mu_t\rangle - \langle\phi, \mu_0\rangle - \int_0^t \int_0^1 (A\phi)(x)\mu_\tau(dx)d\tau - \alpha\int_0^t \int_0^1 \int_0^1 y\phi(x)Q(\mu_\tau; dx, dy)d\tau \tag{2}$$

   is a martingale with quadratic variation

$$\int_0^t \int_0^1 \int_0^1 \phi(x)\phi(y)Q(\mu_\tau; dx, dy)d\tau \tag{3}$$

   for all $\phi \in \mathcal{D}(A)$. Here, $\langle f, \mu\rangle = \int_0^1 f(x)\mu(dx)$, $A = x(1-x)\left(\frac{d^2}{dx^2} - \sigma\frac{d}{dx}\right)$, and $Q(\mu; dx, dy) = \mu(x)(\delta_x(y) - \mu(dy))$.

Remarkably, equation (1) has a closed-form solution for a uniform initial condition. As for the martingale problem (equations (2) and(3)), $\mu^{m,n}(t)$ represents a case where a Fleming-Viot process – an abstract mathematical object – is directly related to real biological systems. A manuscript containing proofs of the weak convergence of $\mu_t^{n,m}$ to these two objects in $D([0,T], \mathcal{M}([0,1]))$, the space of càdlàg processes on probability measures over $[0,1]$, is in preparation with Jonathan Mattingly. The proof of the deterministic limit uses standard tightness arguments [5] while the proof of the Fleming-Viot limit requires a Girsanov formula to show uniqueness of the solution to the martingale problem [6].

1.2 Implications for the field of evolutionary biology

This ball-and-urn system is a new way to think about multilevel selection. First, it provides an elementary and intuitively appealing visualization of a complex multiscale process. Second, its description can be readily understood by empirical biologists, particularly those who study cooperation in experimental microbial systems, where parameters in the ball-and-urn system

directly correspond to measurable quantities. Finally, the antagonism between different levels of selection can be studied in a systematic and rigorous manner using deterministic and stochastic approximations (by taking $n \to \infty$ and/or $m \to \infty$). Whereas previous theoretical treatments of multilevel selection have been tailored to specific traits or systems, such as pathogen trait evolution or the evolution of cooperation, this ball-and-urn process is a unifying formulation that allows antagonistic multilevel selection to be thought of as a general biological phenomenon, independent of specific biological context. A manuscript describing the formulation, along with key results relevant to evolutionary biology, is currently in preparation.

2 Past research: The impact of within-host dynamics on immune escape

A major challenge in controlling rapidly evolving and acutely infecting viruses, such as influenza, is their ability to escape detection by the immune system by changing their shape via mutation. These escape mutants have been empirically observed to arise in vaccinated mice, but not in unvaccinated mice. This is counterintuitive because more mutations occur in infected unvaccinated mice, due to greater viral replication, than in vaccinated mice. Understanding this observation poses a theoretical challenge. How does one combine within-host viral dynamics, often represented by deterministic differential equations, with mutation dynamics, an inherently stochastic process?

In work published in the Journal of the Royal Society Interface [7], my co-authors and I introduce a model that uses a deterministic characterization of viral dynamics while modeling mutation stochastically. We describe intrahost viral dynamics with a simple system of nonlinear differential equations parameterized to be consistent with empirically measured intrahost influenza dynamics. We then model mutation dynamics as a time-inhomogeneous stochastic Poisson process with an intensity rate proportional to the size of the infecting viral population.

The system of differential equations is sufficiently simple that we can solve for the cumulative viral population size for an infection. This key quantity, along with the inhomogeneous Poisson mutation dynamics, is used to approximate (i) the probability that an escape mutation occurs during an infection and (ii) the expected cumulative viral population of such an escape mutant, as functions of prior host immunity. We find that two main immunological processes (host cell availability and adaptive immunity) create a tradeoff between (i) and (ii). In hosts with weak prior immunity, the initial infection rapidly depletes available host cells; mutations have a high probability of occurring due to the rapid replication, but have no resources on which to increase in number. Conversely, in hosts with strong prior immunity, the initial infection depletes significantly fewer host cells before being cleared by the adaptive immune system; there are sufficient resources for mutants to increase to transmissible levels, but due to low replication, these mutants are unlikely to arise. We may therefore reinterpret the counterintuitive results from mice studies as follows. In the vaccinated mice, viral replication from the initial infection is sufficiently high for mutation to occur but, due to the vaccine-induced adaptive immune response, are not so high as to significantly deplete host cells.

Scaling up this relationship between host immunity and intrahost production of escape mutants to a simple population-level model, we arrive at an alternative hypothesis for the time-inhomogeneous rate of immune escape observed in empirical rates of influenza. Specifically, our model suggests that the production rate of immune escape variants is driven by the accumulation of herd immunity.

3 Future research

3.1 Direct extensions of current research

The general multilevel selection framework (§1) motivates questions regarding specific biological scenarios that are interesting in their own right and require specific mathematical analyses:

1. In the multilevel selection framework described above, only two phenotypes are considered. Incorporating additional phenotypes will be important for studying intermediate trait values or three-strategy games [8, 9]. These formulations will have the same fundamental structure as the two-phenotype case, but will require a higher-dimensional analog of the ball-and-urn system (e.g., urns arranged on a lattice rather than in a line).
2. The snowdrift game is an understudied model of cooperation that has been observed in yeast and is believed to be important in other systems [10, 11]. It is motivated by a situation where two travellers are stuck on either side of a snowdrift. Each traveller can choose whether to shovel snow (cooperate) or save their energy (cheat). Both travellers benefit if at least one of them clears away the snow. Thus, unlike the prisoner's dilemma, it is advantageous to cooperate in the snowdrift game even if the other player cheats. In yeast, this is game is consistent with the secretion of public goods enzymes (invertase). These enzymes do not fully diffuse away from the producer/cooperator yeast; cooperative yeast can therefore invade cheater populations [10]. One interpretation of the snowdrift game in a population can be formulated as the multilevel selection framework above, but with a step function instead of a linear function for group-level fitness. The resulting analysis is nontrivial because whether the same limits can be taken will depend heavily on how this step function scales with the parameters m and n.

3.2 The evolution of multicellularity

The evolution of multicellular organisms is currently an open question in evolutionary biology [12–14]. One current hypothesis [13], motivated by experiments with mat-forming bacteria *Pseudomonas flourescens*, proposes that initially, there is some collective benefit (easier access to oxygen) associated with adhesion, leading to the formation of groups of cooperators (cells that produce a polymer adhesive). However, these groups ultimately collapse due to exploitation by cheaters, which do not produce adhesives but nevertheless benefit from the shared resource by residing on the mat. These cheaters may then be thought of as proto-germ cells because, once the mats collapse, only the non-adhering cheaters survive, and upon mutation to the cooperator type (not uncommon in *P. flourescens*), initiate the formation of an 'offspring' mat.

I have formulated this hypothesis in the language of the ball-and-urn framework (§1), mapping the multiscale process onto a process in one scale, this time on a two-dimensional lattice. Each group (mat) is categorized by the pair (n, k), where n is the number of individuals (cells) in that group and k is the number of type C particles it contains (Fig. 2). I assume that cells which do not adhere to or reside on a mat are in groups of size $n = 1$. Individual-level replication is modeled by birth-death type process to allow group size (n) to vary over time. Cooperator offspring remain in the same group as their parent. Cheater offspring reside on the same mat as their parent, or, if their parent does not reside on a mat, they become their own group of size 1. Group replication does not occur. Instead, I assume that after the fraction (or absolute number) of cheaters in a group exceeds a certain threshold, the group dies, leaving only the cheaters as groups of size 1. These groups will subsequently increase in size if the constituent cheater mutates to become a cooperator, which is assumed to occur at a constant rate.

I can use this formulation and variations thereof to investigate questions on the transition to multicellularity. For example, under what conditions are multicellular-reproduction-like cycles expected to occur? Given that the resulting mathematical object is a stochastic process in two

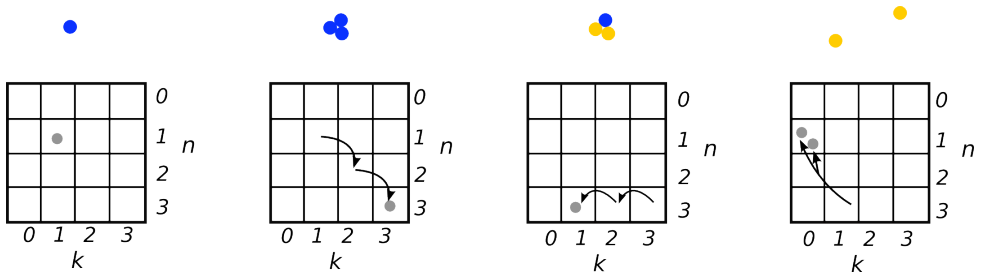

Fig. 2: One possible realization of multicellularity model, *Left to right*: There is initially one cooperator (blue) cell, i.e. a group of size $n = 1$ with $k = 1$ cooperators (grey solid circle in grid at position (1,1)). After two replication events, there is now a group of size $n = 3$ with $k = 3$ cooperators. Over time, mutations to cheaters (yellow) occur, leading to a group of size $n = 3$ with $k = 1$ cooperators. This burden leads to mat collapse, leaving only two cheater cells, each constituting their own group ($n = 1$, $k = 0$ for both).

dimensions, it is not immediately clear that I can apply the same approaches as in §1. I therefore see this project leading to new mathematical approaches as well as addressing an open question in evolutionary biology.

References

[1] Levin, S, Pimentel, D 1981 Selection of intermediate rates of increase in parasite-host systems. *American Naturalist* **117**, 308–315.
[2] Levin, BR, Bull, JJ 1994 Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends in Microbiology* **2**, 76–81.
[3] Okasha, S 2006 *Evolution and the Levels of Selection* (Oxford University Press).
[4] Traulsen, A, Nowak, MA 2006 Evolution of cooperation by multilevel selection. *PNAS* **103**, 10952–10955.
[5] Billingsley, P 1999 *Convergence of Probability Measures* (Wiley-Interscience), 2 edition.
[6] Dawson, D, Maisonneuve, B, Spencer, J, Dawson, D 1993 in *Ecole d'Et de Probabilits de Saint-Flour XXI - 1991*, Lecture Notes in Mathematics (Springer Berlin / Heidelberg) Vol. 1541 1–260.
[7] Luo, S, Reed, M, Mattingly, JC, Koelle, K 2012 The impact of host immune status on the within-host and population dynamics of antigenic immune escape. *Journal of The Royal Society Interface* **9**, 2603–2613.
[8] Kerr, B, Riley, MA, Feldman, MW, Bohannan, BJM 2002 Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**, 171–174.
[9] Semmann, D, Krambeck, HJ, Milinski, M 2003 Volunteering leads to rock-paper-scissors dynamics in a public goods game. *Nature* **425**, 390–393.
[10] Gore, J, Youk, H, van Oudenaarden, A 2009 Snowdrift game dynamics and facultative cheating in yeast. *Nature* **459**, 253–256.
[11] Doebeli, M, Hauert, C 2005 Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecology Letters* **8**, 748–766.
[12] Rainey, PB, Rainey, K 2003 Evolution of cooperation and conflict in experimental bacterial populations. *Nature* **425**, 72–74.
[13] Rainey, PB, Kerr, B 2010 Cheats as first propagules: a new hypothesis for the evolution of individuality during the transition from single cells to multicellularity. *BioEssays* **32**, 872–880.
[14] Ratcliff, WC, Denison, RF, Borrello, M, Travisano, M 2012 Experimental evolution of multicellularity. *PNAS* **109**, 1595–1600.