

Workshop Report
Rules of Life in the Context of Future Mathematical Sciences
November 08-10, 2018

Organizing Committee:

Adriana Dawes, Ohio State University
Marisa Eisenberg, University of Michigan
Padmanabhan Seshaiyer, George Mason University

Executive Summary

The first Workshop on Rules of Life in the Context of Future Mathematical Sciences was held in Alexandria, Virginia on November 8-10, 2018. The workshop **brought together a highly diverse group of individuals in mathematical biology**, including senior and junior faculty members, experts from government labs and funding agencies, as well as postdocs and graduate students. Workshop participants collaborated intensively through brainstorming activities to coalesce onto **six key priority topic areas** for mathematical biology research under the Rules of Life umbrella.

Workshop Goals

- To gather input from researchers in a variety of fields spanning mathematics and biology, focused on refining emerging research challenges and priorities in mathematical biology in response to Understanding the Rules of Life
- To generate key priority topic areas for mathematical biology research under the Rules of Life umbrella, matching the following criteria:
 - a. Key open problems wherein mathematics can be used to discover and apply fundamental biological rules to generate predictions and understand mechanisms.
 - b. Areas spanning spatial, temporal, and biological scales. Cross-scale interactions and effects are a core feature that emerges across biology, which is reflected in many successful mathematical models and approaches.
 - c. Areas where mathematical tools and theory are most needed, and emerging new mathematical applications that show great promise in understanding the rules of life.
 - d. Areas that bridge and connect biological fields, ranging from molecular and cellular biology to epidemiology and ecology.

Recommendations and Summaries of the Priority Topic Areas Identified in the Workshop

These six topic areas are discussed in more detail in Section E (Workshop Results) of the full report.

1. Understanding Rules of Life in Integrative Biological Systems: Transients and Noise.

Transients, noise, and heterogeneity are ubiquitous features of biological systems. However, much of the existing theory in mathematical biology focuses on understanding equilibrium behavior, and does not fully capture the interplay of transient dynamics, noise, and heterogeneity across individuals. Moreover, the concept of an equilibrium doesn't always fully make sense in a biological context—we often characterize a 'healthy' state as an equilibrium and a 'disease' state as a perturbation from that equilibrium, when in reality nearly all states are transient, and there is not a unique healthy state for most species or individuals. New methods and theory are needed to distinguish between equilibrium states, transient or quasi-stationary states, and noise, as well as to examine how individual heterogeneity affects the system dynamics. These methods will help us answer questions such as: What rules of life are critical to a 'normal' state or a healthy homeostasis? How do we build a theory to describe qualitatively different transient states (e.g. health and disease)?

Recommendations

- a) We strongly recommend development of new solicitation(s) or programs that focus on developing a broad dynamical theory for transients and quasi-stationary states in biological systems.
- b) Related to the above, we also strongly recommend support for research examining the relationships between transients, noise, heterogeneity, and the distinct effects of experimental and biological variation, including the high risk/high reward of developing new theory and mathematical approaches to accounting for these different types of variation.

2. The Mathematical Foundations of Data-Inspired Biology: Learning Rules of Life from Data.

Recent data revolutions in biology have brought an enormous range of new opportunities to measure and collect data at scales not seen before. These opportunities bring with them an equally important and challenging need: to develop methods to understand the causal relationships and underlying biological mechanisms embedded in these data. New mathematical concepts, theories, and methods that can make such problems tractable are essential to finding navigable currents through the rapidly expanding sea of biological data.

Recommendations

- a) We highly recommend development of programs or solicitations that support the needed foundational and methodological work in areas such as: model and data dimension reduction methods, rigorous estimation in network analysis, linking models with 'big data', and a unified theoretical framework for parameter estimation, parameter identifiability, and model selection.
- b) Efforts in this area will also critically require close collaboration between data collection and theory, to ensure relevance and impact. We thus strongly recommend a focus on programs that link theoretical predictions with data collection for this priority area.

3. Bridging Scales to Understand New Rules of Life. Emergent behaviors permeate all scales of life, from gene expression at the cellular scale, to maintenance of homeostasis in complex organs, to organism-level interactions shaping populations. To address these multi-level interactions, there is a need for novel multiscale models and modeling approaches that bridge vast differences in space and time scales.

Recommendations

- a) Programs to develop an understanding of how rules of life transition across scales, by building models and approaches that span spatial and temporal scales to uncover consistent biological principles (i.e. rules of life)
- b) Programs to support harnessing biological data across scales and unifying multiscale models between different levels. We recommend that such programs focus on multi-scale phenomena wherein the dynamics at one scale directly affect those at other scales.

4. A New Fitness Landscape for Mathematical Theory: Quantifying Adaptation and Selection in Understanding Fundamental Rules of Life. Evolution is both one of the most fundamental biological processes, and a quintessential multi-scale feedback process: genetic mechanisms take place at the nano-scale, protein expression at the micro-scale, phenotype at the whole organism scale, and selection at the population scale, ultimately feeding back to the genetic scale. Evolution underlies and can inform all areas of biology, including neuroscience, biomechanics, and population and disease dynamics.

Recommendations

- a) Programs or solicitation(s) to support further development of analytical techniques, including modeling trait trade-offs, trait plasticity, and complex traits; and incorporating high dimensional data (e.g. sequencing and gene expression data).

- b) Programs to support efforts addressing important open questions, including: What is the most appropriate way to define fitness? How do we reconstruct phylogenetic evolutionary histories of species with new data? How do we understand community genomics and eco-evolutionary dynamics across species and time scales? Addressing these questions will build the genotype to phenotype map and generate insight into the fundamental rules of life.

5. Mathematics for the Anthropocene Era: Rules of Life in the Context of Human Impact on Natural Systems. Human activity has had and will continue to have significant impact on living systems, raising the important question: what rules of life govern living systems at the ecosystem and even planetary scale, and how do these rules interact with human sociobehavioral and economic systems? Human-natural system interactions exhibit complex dynamic patterns of emergence and tipping points. Expanding the Rules of Life umbrella to consider these systems can generate insight into the management and mitigation of the impact of human activity on living systems and the planet.

Recommendations

- a) Gaining understanding and developing forecasting, prediction, and control methods will require a range of analytical approaches (e.g. dynamical systems, optimization, game/decision theory, etc.), as well as a range of data (satellite, sensors, social media, cell phone data, etc.). We recommend support for programs that integrate or develop mathematical and data-based methods for understanding human-natural system interactions.
- b) We strongly recommend expansion and development of interdisciplinary programs or solicitation(s) that meaningfully draw from the range of disciplines involved in human-natural systems interactions, which span both mathematics and biology, as well as adjacent fields such as behavioral and social sciences, economics, health sciences, and environmental sciences.

6. Broader Impacts: Convergence of Research and Education for the Rules of Life. In the spirit of the NSF Understanding the Rules of Life mission “to train the next generation of researchers to approach scientific inquiry in a way that crosses scales and scientific disciplines,” education and training were identified as key focus areas under Rules of Life. In coordination with the Directorate for Biological Sciences we recommend to create:

Recommendations

- a) Programs for enhancing student research, education, and training at the intersection of biological and mathematical sciences that will help broaden their experience and prepare them for undergraduate, graduate, and post-graduate study and careers at the interface of mathematical and biological sciences along with understanding of evaluation and impact assessment of these programs, leading to increased effectiveness and improved understanding of the Rules of Life;
- b) Programs for faculty development, including longer-term training grants, conferences and workshops to connect scientists, mathematicians, and educational researchers with the goal of bringing together multiple diverse perspectives to focus on finding innovative cross-disciplinary solutions to grand challenge problems at the interface of mathematical and biological sciences.

Common Broader Impact Themes Identified Across All Priority Areas. All six priority areas identified several common themes in terms of recruitment, training, and retention. In particular, **all six groups emphasized that interdisciplinary training is needed** to provide integrated depth in mathematics and biology. **Statistical training and working with data** were noted by multiple groups as a key element of this interdisciplinary training. Multiple groups emphasized the need for programs at all levels (student and faculty), and noted that diverse perspectives are important for deeply interdisciplinary research.

Recommendations: We strongly recommend development and support of interdisciplinary training programs, particularly focused on statistical training and working with data. In particular, we recommend

developing integrated training programs blending mathematics, statistics, and biology. This will be critical for the success of all six topic areas.

The intellectual impact of supporting research in these emerging priority areas is clear: advances at the interface of mathematics and life sciences urgently need rigorous and comprehensive quantitative methods. In addition to furthering our understanding of the Rules of Life, these research efforts will grow convergent research and take advantage of the data revolution.

Workshop Report:

Rules of Life in the Context of Future Mathematical Sciences

The first Collaborative Workshop on Rules of Life in the Context of Future Mathematical Sciences was hosted jointly by George Mason University (GMU), Ohio State University (OSU), and University of Michigan (UM) in Alexandria, Virginia on November 8-10, 2018. This workshop was funded by NSF DMS #1839608 and led by Dr. Padmanabhan Seshaiyer (GMU), Dr. Adriana Dawes (OSU) and Dr. Marisa Eisenberg (UM). The goal of this workshop was to bring together researchers from a variety of fields spanning mathematics and biology, to refine emerging research challenges in mathematical biology in response to Understanding the Rules of Life, one of the NSF's 10 Big Ideas.

The conference brought together a highly diverse group of senior and junior faculty members, experts from government labs and funding agencies, as well as postdocs and graduate students (see demographics section below). Workshop participants collaborated intensively to coalesce onto a set of six key topic areas for mathematical biology research under the Rules of Life umbrella:

1. Understanding Rules of Life in Integrative Biological Systems: Transients and Noise
2. The Mathematical Foundations of Data-Inspired Biology: Learning Rules of Life from Data
3. Bridging Scales to Understand New Rules of Life
4. A New Fitness Landscape for Mathematical Theory: Quantifying Adaptation and Selection in Understanding Fundamental Rules of Life
5. Mathematics for the Anthropocene Era: Rules of Life in the Context of Human Impact on Natural Systems
6. Broader Impacts: Convergence of Research and Education for the Rules of Life

These six topic areas are discussed in more detail in Section E (Workshop Results) below. Our larger goal is that the resulting white paper report and other deliverables can provide useful insight in planning for the role of mathematical biology in Rules of Life, and can help facilitate the mathematical biology community in driving the field forward.

A. Rationale for the workshop in the context of the Rules of Life program

With the explosion of new, complex problems arising from some of society's biggest challenges, including curing diseases, containing the spread of infectious diseases, and increasing food crop yields, there is a great need to match comprehensive genome/environment-to-phenome theories with predictive capability that can help us to design phenotypes to respond to environmental challenges or lead to next generation state-of-the-art technologies. The emergence of models and generation of massive sets of loosely structured data that require quantitative analysis also presents a great opportunity for experts in mathematical biology.

The Rules of Life in the Context of Future Mathematical Sciences Workshop focused on identifying strategic areas where discoveries can potentially be made that will help us to understand how we predict the phenotype, structure, function, and behavior of an organism, based on what we know about its genes and environment. By identifying some of the basic rules of life across scales of time, space, and complexity, we may be able to predict how cells, brains, bodies, and biomes will respond to changing environments.

To forward this agenda, the workshop focused on developing priority areas in mathematical biology that can provide scientists from the fields of mathematics and biology promising opportunities to collaborate, learn, understand the underlying biology, and to develop efficient mathematical methods for quantitative analysis of fundamental problems in biology. We believe that the priority areas identified in this workshop can provide avenues for mathematical biology research under the Rules of Life NSF Big ideas that will:

- a) enable discoveries that will allow us to better understand complex biological interactions and identify causal, predictive relationships across scales and;
- b) help to identify connections between genomic and phenotypic diversity to encompass biological and environmental processes at levels spanning genomes to ecosystems

Toward this end, we aimed to identify priority areas matching the following criteria:

1. Key open problems wherein mathematics can be used to discover and apply fundamental biological rules to generate predictions and understand mechanisms.
2. Areas spanning spatial, temporal, and biological scales. Cross-scale interactions and effects are a core feature that emerges across biology, which is reflected in many successful mathematical models and approaches.
3. Areas where mathematical tools and theory are most needed, and emerging new mathematical applications that show great promise in understanding the rules of life.
4. Areas that bridge and connect biological fields, ranging from molecular and cellular biology to epidemiology and ecology.

B. Workshop Overview

The meeting was held in Alexandria, Virginia, at the Westin Alexandria Hotel on November 8-10, 2018. The conference included keynote presentations, brainstorming affinity mapping activities, faculty lightning talks, student poster presentations, and expert panel presentations from academia as well as leading mathematical biological institutes, government labs, and funding agencies. In particular, the workshop included talks and panel discussions by representatives from the NSF-Simons Research Centers for Mathematics of Complex Biological Systems. The workshop schedule and more information is available on the workshop website: <https://www.futuremathbio.com>.

C. Application Process and Demographics

We sent targeted invitations to experts from a wide range of fields at the interface of mathematics and biology to participate in the workshop, and also allowed individuals to self-select to apply to participate in the conference. We also announced the workshop through some newsgroups. All applicants (invited and self-selected) completed a pre-survey, including a statement of interest and a short application with information on their research areas and background. In our invitations, we chose individuals based on multiple factors, including:

- Diversity of areas in mathematical biology
- Geographic spread across the US
- Representing diverse backgrounds
- Range of levels of academic career
- Range of academic institution types

The workshop applicants and final group of participants were 50% male and 50% female, and included individuals from a diverse range of backgrounds, institution types (graduate and undergraduate institutions), career stages (including a blend of junior and senior faculty among the faculty participants), and geographic locations (the full list of participants is included on the workshop website: <https://www.futuremathbio.com/conference-details>). They spanned the range of both biology and mathematics, and included researchers in (among many others): biochemistry, biology, ecology, medicine, and public health, as well as mathematics, statistics, and engineering. The group included individuals funded by multiple programs within both NSF Bio and NSF Math. Further details on participant demographics are given in **Appendix H.1**.

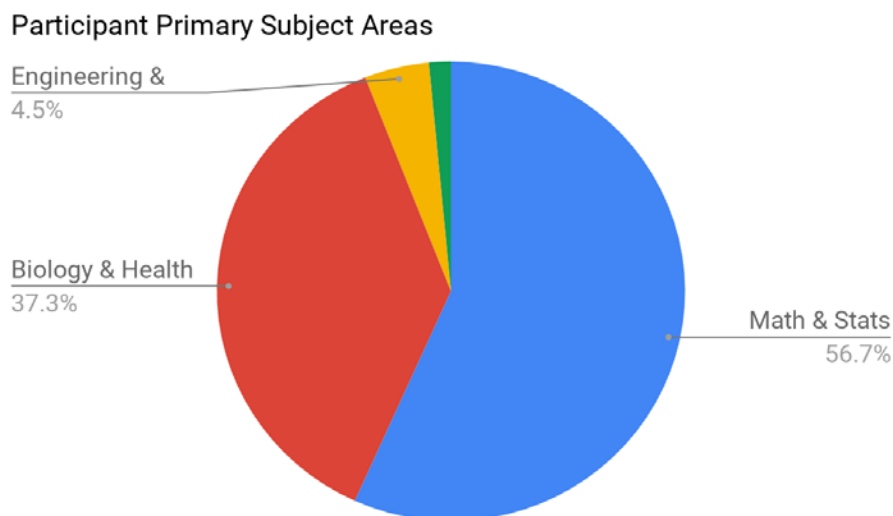


Figure 1. Participant primary subject areas. Participants could choose more than one category.

In their application, participants identified promising directions for mathematical biology, and stated what they hoped to get out of the workshop. We compiled these responses into common themes, which were used to orient the participants when they joined the workshop. The most common themes among the directions for mathematical biology identified by the participants included:

- Machine learning, data science, high dimensional data, computational approaches
- Multiscale modeling in biology
- Data-driven research, linking models with data across scales
- Inferring mechanisms, hypothesis generation and testing
- Building a model-experiment feedback loop
- Building interdisciplinary collaborations
- Climate change & natural disasters
- Educational initiatives that build the next generation of interdisciplinary science
- Engaging with policy-makers

Other themes included single cell biology, systems biology, mathematical neuroscience (specifically focused around Alzheimer's disease, epilepsy, and cognition), infectious diseases, ecology and emerging infectious diseases, personalized medicine, and new mathematical approaches such as algebraic and topological methods.

Common themes among participants goals for the workshop were to:

- Establish new collaborations, networking and exchange research ideas
- Contribute and learn about the future directions, emerging areas, and priorities for mathematical biology
- Learn about the NSF's Big Ideas and the Understanding Rules of Life program
- Determine how national centers can contribute to the research and educational goals in mathematical biology
- Develop educational and outreach ideas for mathematical biology
- Create a sense of 'the big picture' in mathematical biology—as described by one participant, to answer the question: “where do we want to take this field over the next 5, 10, 15 years and how do we get there?”

Additionally, one of the important aspects of the conference was to engage and involve program officers from the National Science Foundation. The program activities included a dedicated panel that brought panelists from both the NSF and NIH, and also included NSF program directors as observers during the affinity mapping brainstorming activities and writing exercises. This interaction helped the Program Officers to get feedback regarding the topics that are important to the community, as represented by the participant sample group.

D. Workshop Structure and Activities

The workshop incorporated a variety of activities, including keynote presentations, brainstorming affinity mapping activities, faculty lightning talks, student poster presentations and expert panel presentations from academia as well as leading mathematical biological institutes, government labs, agencies and the industry. The schedule of activities can be found online at <https://www.futuremathbio.com/workshop-schedule>.

Following opening remarks by the organizers and introductory remarks by Division Director Juan Meza (NSF/MPS/DMS), all participants gave 2-minute lightning talks that consisted of a single slide that described their research interests and areas, and/or problems they felt were particularly promising for mathematical biology. This was followed by introductions to each of the NSF-Simons Centers, describing the center goals and how they relate to the future of mathematical biology and NSF's Rules of Life.

Generation of the research priority areas was accomplished through affinity mapping and focused writing exercises in pairs and small groups, followed by group editing and cross-group commenting.

For affinity mapping, participants gathered at various tables, with each group consisting of members who represented a combination of graduate students, post-doctoral students, junior faculty, senior faculty and administrators. These also included members who had a rich history of funding from a variety of standard research, conference, and workforce programs from the NSF. The Affinity activity went through four different phases including:

- **Phase 1** was *independent brainwriting*. All participants were asked to write specific ideas on a given prompt which was “*What should be strategic priority areas of mathematical biology under NSF's Rules of Life?*” The point here was to create multiple independent ideas and therefore the focus was on quantity.
- **Phase 2** was *table team brainstorming*. This involved the diverse group of people at each table to itemize the independently generated ideas into specific categories, with the challenge of not having more than 4 categories at each table. Each of these categories were then presented with

labels on poster paper that was provided which went up on walls next to each table. The figure below illustrates the participants engaged in the table team activity.

- **Phase 3** included a *gallery walk* first that allowed each table team to compare their ideas with other table teams. Following that we conducted a *group brainstorming* where each group summarized their specific categories. As the participants heard each of the table groups speak, several similarities and differences were noted between table team ideas.
- **Phase 4** included identifying as an entire group the top six priority areas that everyone in the room was vocal about. Most of these priority areas naturally evolved among the table teams.

For the writing portion of the meeting, we split the participants into groups for each priority area. Within each priority area group, participants split into smaller groups of 2-4 individuals, where each small group was tasked with writing one piece of the priority area report. The small groups then reported back to their priority area group, so that the entire group could comment, edit, and integrate the pieces for review by the whole participant working meeting (and eventually the larger community). To this end, we used a priority area writing template to lend structure, where each small group within a priority area group can work on one piece of the template (template is given in **Appendix H.2**).

Additionally, the workshop included a panel discussing funding, opportunities, and directions in the context of mathematical biology and the Rules of Life program. The panel consisted of:

- Kathy Dickson (NSF/BIO/IOS)
- Jarek Majewski (NSF/BIO/MCB)
- Kimberly Gallagher (NSF/BIO/IOS)
- Theodore Morgan (NSF/BIO/IOS)
- Betsy Von Holle (NSF/BIO/DEB)
- Samuel M. Scheiner (NSF/BIO/DEB)
- David Alexander Rockcliffe (BIO/MCB)
- James Powell (NSF/MPS/DMS)
- Junping Wang (NSF/MPS/DMS)
- Peter Lyster (NIH/NIGMS)
- Grace Peng (NIH/NIBIB)

E. Workshop Results: Key Rules of Life Priority Areas

1. Understanding Rules of Life in Integrative Biological Systems: Transients and Noise

Summary. Heterogeneity is inherent in the systems we are trying to describe and the data we observe; transients and noise are inescapable features of biological systems. It is instinctive to think of natural systems in classical dynamical terms – for example describing health as ‘equilibrium’ and disease as a perturbation. However, in the context of biological systems, ‘equilibrium’ is inherently ill-defined and heterogeneous: a collection of healthy individuals may be described by a collection of attributes (height, weight, hair color, blood pressure) that vary wildly. A diverse group of ‘genotypes’ give rise to the same healthy-phenotype – and the same phenotype may conceal myriad genotypes.

We are calling for the development of a theory of transients, noise, and heterogeneity for biological systems, since classical theory does not adequately address these features. In fact, a state of healthy homeostasis conceals a tremendous amount of heterogeneity, and may even be a transient state in many contexts. How do we characterize such a state? Are there rules or themes that are common across scales – cells to populations to communities – that are essential for the emergence and ‘stability’ of such a state? And finally, how do we characterize excursions or departures from that state? This is essential to understanding ill-health, or disease, or catastrophic events such as extinction. Some argue that life itself is transient. There is a need for theory to describe these healthy ‘baseline’ states and departures from such states, analogous to theory of non-equilibrium thermodynamics. Additionally, improved data collection methods are revealing heterogeneities that we may now effectively study. New quantitative tools are needed to characterize and understand the role of transients and noise in biological systems, and this work must begin with agreed upon common definitions for the states we are studying.

Intellectual Merit. We are proposing the development of theory to explain commonality of transient and noisy phenomena across biological systems. Thus, this research priority cuts to the heart of the Rules of Life concept. There are many key biological areas where transients and noise play a critical role. A significant and important example of a highly heterogeneous system for which we have no theory is the gut microbiome. Microbiome ‘health’ has been demonstrated to be associated with aspects of human health from metabolism to immune function; disorder in the microbiota can lead to human disease outcomes ranging from irritable bowel disease, to Type 2 diabetes, to *C. difficile* infection¹. Thus, a clearer picture of the microbiota is essential to characterizing human homeostasis and disease pathogenesis². But the microbiome itself represents a diverse ecosystem, which is widely variable across individuals; we don’t understand its balance or how excursions from that balance work at the level of the bacterial ecosystem. So, for example, when a *C. diff* infection is treated, it raises the question: does the individual’s microbiome go back to its original ‘healthy’ state? A different ‘healthy’ state? And is healthy only interpreted in terms of its interaction with the human host, or does it have some inherent healthy balance? Infectious diseases provide a range of examples of externally-induced excursions from a ‘heterogeneous but healthy’ state (although they are by no means the only examples). For instance, influenza infects millions of people annually, but symptoms and severity of symptoms vary wildly. One may consider these as inherently heterogeneous excursions from the ‘health’ state. Similarly, following acute infection, hepatitis C infection may spontaneously clear, another example of an excursion from the ‘health’ state,

¹ Bull MJ, Plummer NT. *Part 1: The Human Gut Microbiome in Health and Disease*. Integr Med (Encinitas). 2014 Dec;13(6):17-22.

² Shreiner AB, Kao JY, Young VB. *The gut microbiome in health and in disease*. Curr Opin Gastroenterol. 2015 Jan;31(1):69-75.

or become chronic, leading to a permanent departure from the health-state. An important aspect in describing a 'heterogeneous health equilibrium' is identifying genotypes that may lead to different disease 'phenotypes', such as severity of infection.

These issues also vary depending on the scale considered – at the host level, antibiotic use can alter the in-host microbiota equilibrium. At the between-host level, antibiotic use demonstrably alters the circulating bacterial phenotype, from drug-sensitive to drug-resistant. Stochastic introduction of infectious diseases in disease-free 'healthy' populations represents an external perturbation. What aspect of the population 'phenotype' (e.g. population density, behavioral patterns) will predict outbreak outcomes, in terms of morbidity, mortality, and whether outbreaks result in a new, endemic equilibrium ('departures') or are short in duration ('excursions') from the stable healthy population state. Following the transient excursion, how do we describe the population 'steady' state?

Species invasion is another biological example where transients and noise play a critical role. A closed ecosystem such as that found on an island is considered at equilibrium. When a new species invades, there is an ecological departure from the current equilibrium and a new one is established. How do we understand or predict what will happen? Current conditions may result in the potential for loss of species. But attempts to save them are unpredictable. An example of this equilibrium disruption occurred on Macquarie Island, a subantarctic island. It housed a large population of feral cats that were eradicated to prevent loss of the native bird species. In the absence of cats, rabbits then took over and ate all the vegetation, resulting in the deaths of the birds anyway^{3,4}. Theories to understand heterogeneous 'equilibria' and stochastic transient excursions will permit the design of interventions to preserve ecosystems in predictable ways.

In terms of mathematical innovation, methods which can identify and account for meaningful variability in biological systems are urgently needed. Animal experiments, even among inbred rats and mice, give rise to heterogeneous time series data that is (nearly) effectively continuous. Often the shapes of the time series curves contain important information, e.g. oscillations that are lost if the curves are averaged due to heterogeneity in the phase and period of oscillations. We need new methods that characterize such variability among collections of time series data. Indeed, variability itself can be phenotypic, and model structures need to be developed that account for the extent of variability as well as time- or state-dependent changes in variability. We don't currently have appropriate equilibrium or non-equilibrium-type tools to investigate biological systems that are inherently stochastic, and where transients and noise are an essential aspect of their dynamics. We are also lacking methods to distinguish between new equilibrium states, slow transient or quasi-stationary states, and noise.

These questions have profound consequences for the Rules of Life. To uncover these rules, we must look for collective behaviors across different systems. What rules are critical to a 'normal' state or a healthy homeostasis? Can we identify common rules across systems that exhibit similar transient and equilibrium-like behaviors? What is the effect of variability on near "equilibrium" dynamics and how does variability push a system towards "non-equilibrium"? What are the consequences when these rules governing healthy states are perturbed? This is essential to understanding the emergence and spread of disease, as well as disastrous departures from a state of wellbeing such as extinction events. For a population in a diseased state, what manner of control may be exerted – e.g. policy – to encourage a

³ Bergstrom DM, Lucieer A, Kiefer K, Wasley J, Belbin L, Pedersen TK, Chown SL. *Indirect effects of invasive species removal devastate World Heritage Island*. J App Ecol. 2009 Feb;46:73 – 81.

⁴ Taylor RH. *How the Macquarie Island parakeet became extinct*. NZ J Ecol. 1979 2:42-45.

return to a healthy state? Research on these questions will inform experiments to understand and explore the rules of life in the context of transient and noisy systems.

Broader Impacts. Research on Rules of Life with a focus on understanding the role of transients and noise will have tremendous impacts in STEM workforce development. An interdisciplinary curriculum is needed to provide undergraduate and graduate students with thorough training in the two disciplines. This must include better statistical training for mathematicians and biologists. Integration of emerging data, such as wearables, will likely attract diverse interdisciplinary students and help ensure persistence in the academic pipeline. It also necessary to ensure all practitioners (from undergraduate trainees to senior faculty) understand the methods used and their implicit assumptions.

Challenges and Opportunities. Understanding transients and noise and their role in the Rules of Life presents a number of exciting opportunities for research. The rich existing body of theory for stochastic modeling and dynamical systems can be considered a jumping off point for a dynamical theory of transients, noise, and heterogeneity. Some current standard approaches for modeling stochastic systems make assumptions that are not always obvious, such as the existence of underlying distributions. When developing new models, it is critical to be aware of implicit assumptions of existing techniques. It is also necessary to distinguish between biological and technical variability, such as noisy data when measuring individual transcripts and proteins. New technologies are producing an overwhelming amount of data, and it is not clear how to analyze variability in experimental outcomes at different scales. Other open problems include interpreting variable data that may come from a non-variable source such as experiments involving clonal animal populations, and how to scale the implications or findings of those studies up to populations that are highly diverse. The frontier in this area includes integrating social, behavioral and economic factors in the understanding of a normal or healthy state, and then including those understandings in studying departures from such a state.

2. The Mathematical Foundations of Data-Inspired Biology: Learning Rules of Life from Data.

Summary. Recent years have seen tremendous growth in the amount of quantitative information available in the biological sciences. New high-throughput experimental technologies have matured and are being combined creatively to explore living systems across all scales, in ways not envisioned previously. However, with the enormous opportunity afforded by these data, comes an equally important and challenging need to develop methods to understand the causal relationships and underlying biological mechanisms embedded in these data. As noted by Dr. Eric Green, Director of the National Human Genome Research Institute in June of 2013, "...generating the data is not the bottleneck, it's analyzing the data, and I will tell you that this is becoming a big focus of the tension at the NIH, recognizing that we need to ready the biomedical research community for a new reality, that it's not transient, this is going to be it forever."

This comment makes it clear that obtaining the large wealth of data is not an end in and of itself — methods are needed to integrate and make sense of different data sets. This has proven to be challenging not only because of the amounts and disparate types of information available across scales ranging from molecular to societal, but also because of the complexity of the underlying systems involved. From a mathematical perspective, analyzing and integrating these data translates to learning mechanisms from experimental data, constructing predictive models that can accurately describe them, and using these models to generate new experimentally-testable hypotheses. Such models, with all of their inputs, outputs, and internal parameters fully elucidated, can be indispensable tools with which to discover the principles governing the behavior of biological systems. Fulfilling this vision at a time when experimental

techniques and the amounts of data they generate are advancing rapidly is an indisputable challenge. New mathematical concepts, theories and techniques that can make such problems tractable are essential to finding navigable currents through the rapidly expanding sea of biological data.

Intellectual Merit. There has been an explosion in the amount of data generated and collected in all areas of biology, and analysis of these data is necessary to make fundamental progress in our understanding of biological systems. Mathematical approaches can provide a bridge between data and mechanism, and have brought successes in many other disciplines. Thus, mathematical approaches are a natural avenue by which foundational and unifying principles at work in biological systems (i.e. rules of life) can be elucidated.

There are several mathematical areas that need to be developed in order to effectively use data to understand the rules of life. Some of the methods currently in use to analyze data require further mathematical foundations in order to provide meaningful results. A cohesive theory to understand and evaluate dimension reduction methods, rigorous estimation in network analysis, and a unified theoretical framework for parameter estimation are examples of areas in which further foundational work is needed. Moreover, some questions require new mathematical methodologies to be developed. For example, inferring model type and properties from a given data set requires novel mathematical methods that can distinguish between stochastic and deterministic processes, or select an appropriate model from a predetermined family. Related to these issues, parameter identifiability analysis is an important focus area, as it can guide new experiments that will generate the necessary data for the development of predictive mathematical models.

Many biological systems are complex, where the number of variables and parameters make the construction of a useful model impractical. For example, a goal of the BRAIN initiative (from the BRAIN 2025 report⁵), is to “produce a dynamic picture of the functioning brain by developing and applying improved methods for large-scale monitoring of neural activity.” To capture this type of system, it may be necessary to model a smaller system, and then scale the results to answer questions about the underlying system in its entirety. Developing a robust mathematical theory about how these models scale is, therefore, critical.

Additionally, biological systems commonly produce noisy data (which may be due to process noise or measurement noise), and current technology generates a wide range of data types. Hence, we need to develop methods to combine heterogeneous, noisy data modalities, along with priors that are consistent with state-of-the-art biological knowledge.

Broader Impacts. Exposure to data is not a traditional component of most mathematicians’ education. Special attention should be made to prepare mathematicians to incorporate working with data as a central element of their training. This may include traditional statistical training as well as training in data science methods (e.g. machine learning approaches). For many of the goals outlined above, new approaches to linking mechanistic models with data are needed, to extract causal relationships from data in addition to describing patterns and making predictions. This may necessitate new approaches to statistical training that focus on linking data with a wider range of mathematical models rather than (or in addition to) working with more standard statistical models (e.g. linear regression).

Shorter events such as workshops or “hackathons” may also be a useful avenue for both recruitment and training in this area. Hackathons are increasingly popular and exciting to students, and can often be a

⁵ <http://www.braininitiative.nih.gov/strategic-planning/brain-2025-report>

productive way to generate new ideas for participants at all levels. Mathematics/biology data hackathons could generate excitement and provide avenues for new ideas to emerge. Similarly, problem-based challenges or collaborative networks that let groups of researchers develop and compare models and methods to solve a particular problem or question can help to cultivate and improve new approaches.

The mathematical tools developed for biological data analysis can be implemented on open-access platforms to make them available to researchers at all levels. We envision that these resources will help train young scholars. Making data sets and appropriate methods available to students as part of collaborative research endeavors, or as part of an energizing model selection competition, could motivate students to enter the field.

Challenges and Opportunities. In part because of the rapidly-developing nature of measurement techniques and data availability, there are a range of challenges and opportunities in this topic area. In many cases the data complexity may exceed current analytical approaches, or the theoretical foundations for linking models with a given type or scale of data may yet not be developed.

For example, a common theme across biological disciplines is interactions between different components, often represented in the form of networks. There are now many methods in use to construct networks from biological data, and often these methods are primarily algorithmic. Further advances in mathematical methods and underlying theory that can aid network construction, modeling, and analysis would be a significant advance and continue to be strongly needed. We specifically need to develop a rigorous statistical framework to describe biological network formation and features (such as centrality measures and community structures). Building a statistical framework that allows for rigorous inference and uncertainty analysis with network-based data is an important focus area.

Similarly, a common methodology with large data sets is to perform dimension reduction. Biological data often have complex phenotypes, and dimension reduction approaches help our understanding by decomposing the data into a few specific variables or combinations of variables that explain most of the variability in the observations. Further development of mathematical approaches in this area and/or a firmer foundation for these methods would certainly provide immediate utility to the biological sciences.

Related to these points, new mathematical results in the areas of parameter estimation and uncertainty will be needed as new data are generated in an increasing range of biological contexts. To study certain biological questions, a range of experiments or measurements may need to be performed. These may generate different types of data informing the same biological phenomenon. For example, in studying Parkinson's, different imaging modalities (e.g. MRI, EEG and PET scans), molecular measurements, and symptom measurements may all be used to study the disease in the same patient. Similarly, understanding infectious disease spread may necessitate incorporating data types ranging from individual-level antibody titer measurements, to population-level disease incidence time series, to population behavioral patterns. Therefore, we need to develop methods for combining heterogeneous, noisy data modalities, given that parameter estimation and model selection can both suffer from unidentifiability and uncertainty, such that there may not be a well-defined optimum parameter set (or sufficiently informative distribution of parameters) or model choice. Estimation and identifiability methods that can provide conclusions about what can and cannot be explained by a given model and data set will be critical in building models and linking them with data to answer specific biological questions. Additionally, scalable model selection and model reduction methods that allow us to transition models across scales will be key to building models and approaches that span data types and scales to uncover consistent biological principles (i.e. rules of life).

Finally, linking models with data also critically requires close collaboration between data collection and theory. Successful collaboration and insight often require an iterative process of model/data generation. This is particularly important as new data generation methods and new analytical approaches are developed. This will require intentionality in improving communication between mathematicians and biologists in order to overcome the high language barrier that still exists between the disciplines.

3. Bridging Scales to Understand New Rules of Life

Summary. Table-top and low-cost comprehensive sequencing has exploded the amount of genetic information available on organisms; and opened new quantitative windows on diversity in real populations. Yet, basic questions about the linkage of genes and their function to the population-level forces that shape gene frequencies remain unanswered. A priority area for Rules of Life should be the promotion of projects that link understanding across scales within their physical and biological environments. Can we predict population level changes from genetic dynamics and conversely? These models have to work in two directions, upward: how does genetic variation affect phenotype and fitness? And downward, how does selection arising from environments, ecosystems and environmental changes act upon genetic variation? Although ecosystems and genes represent two endpoints for this area, bridging the scales involves integrating many levels of biology, from genetics to cell and intracellular behaviors, organization of cells into tissues and physiological organization between tissues.

Intellectual Merit. Emergent behaviors permeate all scales of life, ranging from gene expression at the cellular scale, coordination among nuclei to direct behaviors of syncytial (multinucleate cells) to cellular coordination within biofilms, to maintenance of homeostasis in complex organs, to organism-level interactions shaping populations. To address these multi-level interactions, there is a need for novel multiscale models and modeling approaches that bridge vast differences in space and time scales. At the cellular scale, efforts should focus on linking gene expression to cell behavior, particularly in environments with varying biophysical, biomechanical, or biochemical cues. Rules uncovered via this process will help identify emergent responses from one level to the next, be it for cells to tissues, or from individuals to populations.

There are a wide variety of models that predict behavior at a specific scale. However, there is a need for rigorous approaches that incorporate cross-scale communication once rules of life are identified at each scale. This will require innovative mathematical and statistical methods and techniques enabling inference at one scale based on knowledge obtained from a different scale. These techniques will connect rules discovered at several scales enabling representation via novel mathematical models that capture emergent behaviors across diverse spatial and temporal scales. The era of big data and data-driven modeling also necessitates an effort to integrate machine learning and advanced statistical models to aid in discovery of the key mechanistic or stochastic elements to be included in a multiscale biological model. These issues raise the question, how can machine learning insights be translated into mechanistic insights? Conversely, how can mathematical models inform and improve approaches to data collection and design of experiments?

There is an incredible amount of data being collected across scales, with varying levels of detail and resolution. This data explosion presents a unique opportunity to investigate the effects of changes and variations across diverse spatial and temporal scales, yet requires efforts to unify multiscale models between different levels of the hierarchy of life and between different organisms. Multiscale modeling is an exciting area, with many existing techniques, yet it presents new opportunities for typically underrepresented areas of mathematics to inform innovative approaches to solve challenging biological

problems in the context of Rules of Life. Rapidly increasing computational power and innovative computing architectures can also facilitate development of novel methods for the transitioning and communicating across scales needed to uncover relationships between the rules of life and emergent behaviors.

Broader Impacts. Broader impact initiatives need to be as scalable, replicable and assessable as core elements of the research project design that they are attached to. The main goal of such educational projects should be the generation of talented researchers in a manner that is cognizant of the current challenges faced by students and faculty at all levels in their academic career.

Particularly, multiscale research needs expertise in many different fields which creates a need for collaborations among different departments and the possibility of experimenting with different educational models that allow for training in several disciplines. There is a need for programs at every level which will allow this cross-educational experience; students, faculty and researchers should be encouraged to make connections between disciplines as early in their careers as possible. Ideally, all constituents in education of the future research workforce should be involved in these projects (i.e. students, teachers, faculty and administrators at all levels).

There is a need for innovative ways to attract historically underrepresented populations in the STEM field. These strategies need to be data and evidence based and in addition to attracting students into science, they need to address the potential exit points in their academic path.

Projects need to include different proactive mentoring structures which follow students through critical transition points. An example of these are mentoring networks to pair K-12 educators with faculty, and students with faculty. These projects need to create new models and standards for outreach and education, that minimize the distance between research and classrooms, and increase the scales of outreach / educational activities. This addresses two current problems with broader impact activities: 1) Although individual faculty may pursue ambitious and effective broader impacts activities, these activities lack momentum and are not sustained after the NSF project ends. 2) Faculty effort may also be wasted in the rediscovery of solutions to broader impact challenges. Projects should include a plan for institutionalization of the new proposed programs to create lasting effects. In addition, these projects should be scalable and replicable by other faculty/institutions, with methods and materials being available for any interested party.

Challenges and potential pitfalls. Multiscale modeling is appropriate to many modeling systems, and resonates with multiple mathematical techniques. Some examples of currently used techniques are rigorous upscaling/mean field models, hybrid-discrete models, and structure-based constitutive laws. Yet these techniques too are quite disparate within mathematics, attracting different mathematical communities. Rigor, which is the usual arbiter for comparing mathematical theories, may be constrained by the fact that the models themselves are evolving, as models are continuously updated in response to evolving empirical data.

Multiscale modeling means bridging disciplinary boundaries within biology. Currently, the different hierarchical levels identified above reside in different areas of biology, that may belong in different departments, with distinct intellectual communities, journals, emphasis areas and vocabularies. E.g. standards of data are very different between ecology and cell biology or biophysics.

“Multiscale” is somewhat of a buzzword in mathematical modeling. A call that targets research in multiscale modeling must avoid being too broad, and must differentiate between deeply multi-scale

phenomena and those that are multi-scale but not necessarily dependent on the links between scales (e.g. a virus that has within-host dynamics, and also rich spatio-temporal transmission dynamics, but for which the dynamics may be separately modeled). Projects must also focus on questions in which hypotheses are offered and quantitatively tested.

We need to create or discover new model biological systems for measuring the impact of genetic variation upon organism fitness. Currently most model systems only bridge a subset of scales. As an example, among fungal models (e.g. yeast, model filamentous fungi such as *Aspergillus sp.* and *Neurospora sp.*), there is a wealth of cell biological and genetic data, but relatively little knowledge of organisms in their ecological context. At the same time, although some systems may lend themselves e.g. to understanding features of organism interaction (such as cooperative dynamics in *Dictyostelium*) the fitness effect of variations, such as cooperation, may not be known. We may have to go to understudied organisms (protists, fungi, bacteria, insects etc.) to find new candidate model systems. Although this may appear to be a challenge primarily to biology, many missing links can be filled out using mathematical models. For example, genetic variation may affect adaptations associated with an organism's dispersal (e.g. spore ejection apparatus in fungi), but the fitness effect of the associated phenotypic variation cannot currently be measured, but might be able to be inferred from models of the physics of spore transport in the atmosphere, abundance of potential habitats etc.

At the structural level, new funding models may be needed for individual investigators or collaborating teams. Currently proposals have to review well in multiple panels for co-funding across NSF directorates. Proposals reviewed by DMS must pass the "substantial new math" test. Much innovation in math biology is synthetic (i.e. bringing established tools to bear on new systems). At the same time, the experimental quantitation needed to test model predictions may seem somewhat routine from a biology perspective. Hence transformative proposals may need their own, specialized review. This may require creation of new opportunities for expert, interdisciplinary review, from panelists trained in multiple disciplines, and able to evaluate work across these disciplines. It also may require new funding conventions, particularly for single investigator projects that have both experimental and theoretical components. As an example of this, Simons Mathematical Modeling of Living Systems Investigator-ships are set up to be of a size to allow a single investigator to run experiments. Yet, there aren't enough awards of this size, arguably, to move the entire community forward. (An unfortunate consequence is this would mean fewer funded grants, and could lead to over-concentration of resource under a few PIs).

In terms of open problems and frontiers, there are several pressing questions that need attention to help bridge the scales to understand the rules of life. These include: how gene expression for individual cells influences cell fate and cell response to environment; how inter-cell communication affects (a) emergent phenomenon at the tissue level (b) individual cellular behavior; how to construct mathematically rigorous models that 'communicate' between many different scales; how do we make predictions on one scale based on data and models in a different scale; how do the RoLs at the gene, cell and cell environment scales affect function and form in multi-tissue organ systems; and how does behavior emerge from collective rules for individual components (e.g. cells or organisms)?

4. A New Fitness Landscape for Mathematical Theory: Quantifying Adaptation and Selection in Understanding Fundamental Rules of Life

Summary. Evolution, which proceeds through mechanisms such as adaptation and selection, is a fundamental biological process and the primary driver in many biological systems that urgently need to be considered: host-pathogen dynamics, understanding the microbiome, evolution in response to human

interventions (antibiotic resistance, adaptation to pesticides), and carcinogenesis. Evolution is a quintessential multi-scale feedback process: genetic mechanisms take place at the nano-scale, protein expression takes place at the micro-scale, phenotype exists at the whole organism scale, and selection takes place at the population scale. These higher scales ultimately feed back to the genetic nano-scale, impacting and driving the occurrence and relations at the lower scales.

New quantitative data is available at each of the scales and mathematical modeling is diversifying and developing to meet these challenges. While evolution is certainly fundamental for understanding how these mechanisms and structures came about, this is perhaps not as explicit as it could be in mathematical models we articulate (or the questions that the mathematical models are designed to address). We recognize that the mathematical theory of evolution has been around since the inception of mathematical biology as a field; but for all this effort, we still do not have a satisfying answer for the question: what is a good mathematical definition for the term “trait”? Or how can we explicitly quantify traits from the nano-scale in such a way that they are expressed in the mathematical models that we develop.

In fact, any mathematical discipline that is confronted with an evolutionary process might recognize aspects of traits that can be expressed in their discipline’s language. This is a good thing! When new mathematical language is introduced to model genotype-to-phenotype maps, this is an opportunity to think about how adaptation and selection acts on these systems. Moreover, one is immediately challenged to ask how to model phenotypic variation, and stochasticity in the manner in which the environment exerts selection pressure on diverse populations.

One example of an opportunity to confront theory with data is between the theory of fitness landscapes with quantitative genetics⁶. Populations must conduct a random walk in genetic space while facing the consequences that exist in phenotype space. How well do existing fitness landscape modeling frameworks capture these dynamics? What new models need to be developed?

We are calling for novel approaches to understand fundamental evolutionary questions and using evolutionary principles to understand their underlying role in other biological systems. There are continual efforts to obtain new data and tools being developed to capture diversity at various scales, but still many perplexing questions in evolution. Can this new data, along with novel ideas, help to better understand the way evolution, the most fundamental Rule of Life, works?

Intellectual merit. Evolution underlies and can inform all areas of biology, including neuroscience, biomechanics, population and disease dynamics (human, animal and plants). Furthermore, evolution is the ultimate multi-scale feedback process: genetic mechanisms take place at the nano-scale, protein expression takes place at the micro-scale, phenotype exists at the whole organism scale, and selection takes place at the population scale. This directly feeds back to the genetic scale, by determining what is sustained over time. This is reflected in many different systems of interest such as:

- Host-pathogen dynamics
- Microbiome
- Evolution in response to human intervention (Antibiotic Resistance, Agriculture and Pesticides)
- Stem cell cancer

⁶ Raghwani J, Thompson RN, Koelle K. *Selection on non-antigenic gene segments of seasonal influenza A virus and its impact on adaptive evolution.* *Virus Evol.* 2017 Nov 9;3(2):vex034.

- Social sciences (such as Allee effects in the spread of ideas where a minimum number of people is required before the adoption of an idea)

Developmental biology must be involved in this research to provide specific perspectives for the genotype-phenotype mapping, such as how complex advantageous traits arise.

Broader impacts. Improved interdisciplinary STEM training is needed to support research in adaptation and selection through quantitative methods.

Challenges & opportunities. Numerous challenges exist in modeling adaptation and selection, integrating novel mathematics/non-traditional mathematics into the study of evolution, and moving beyond what has been done classically. This includes the measure of trait trade-offs and plasticity of traits, and simulating multi-type branching processes. Further challenges include incorporating high dimensional data, such as sequencing and gene expression data, to obtain biological insight into evolution in a variety of contexts. These efforts will work toward building the genotype to phenotype map and confronting mathematical theory with new quantitative data that are available at different biological scales.

A number of key open questions would benefit from new mathematical approaches: How can we more mathematically think about traits and plasticity of traits, beyond the classical views based on what was being measured in the past (e.g. morphological landmarks of an individual is one classical view)? What is the most relevant/appropriate way to define fitness? What are the fundamental trade-offs that shape life at all scales? What are limitations on organisms resulting from trade-offs? What role does randomness play in generating diversity? How do we reconstruct phylogenetic evolutionary histories of species with new data? What is the best approach to modeling complex traits: how they develop, how they evolve? Is there a way to reconstruct evolutionary history of complex traits? How do we understand community genomics and eco-evolutionary dynamics across species and time scales?

5. Mathematics for the Anthropocene Era: Rules of Life in the Context of Human Impact on Natural Systems

Summary. Human activity has had and will continue to have significant impact on living systems, raising the important question: what rules of life govern living systems at the ecosystem and even planetary scale, and how do these rules interact with human sociobehavioral and economic systems? Expanding the Rules of Life umbrella to consider systems at the larger scales of populations and landscapes could generate enormous insight into the critical issues of how to manage and mitigate the impact of human activity on living systems and the planet. Questions at the ecosystem and population scales show much of the same potential to benefit from a rules of life approach as do systems at molecular, cellular, and organismal scales—there are enormous new quantities of data (e.g. satellites, sensors, wearables, online data, and cell phone data, as well as a wide range of molecular techniques that can generate -omic data about environments and ecosystems), and similarly large opportunities for analytical approaches to generate insight. The mathematical methods have not yet fully risen to the challenges involved in these systems, many of which relate to the other priority areas in this report, such as dealing with highly multiscale questions, and integrating new data sources. One particularly important challenge is to meaningfully incorporate knowledge and understanding from the range of disciplines involved in understanding human-natural systems interactions, which span both mathematical and biological disciplines, as well as adjacent (and often overlapping) fields such as economics, behavioral and social sciences, health sciences, and environmental sciences.

Intellectual Merit. Human-natural system interactions are highly complex and exhibit complex dynamical patterns of emergence, tipping points, etc. Solving problems and gaining understanding in these areas

will require a range of tools and approaches, and different subareas often require different tools (e.g. fluid dynamics in understanding atmospheric weather patterns vs. for example ODEs or discrete dynamical systems for modeling population dynamics). Further, interventions and control strategies require optimization, economics, and cost effectiveness methods, which can benefit from connections with operations research and game/decision theory to account for issues of logistics, human behavior, and decision making. Finally, because these systems are global in scale and highly complex, model and dimension reduction techniques are needed to tune the granularity of the models we build to the scale of the questions to be answered.

In addition to the need for combining a range of mathematical tools, there are also needs for mathematical tools that are not yet fully developed. Given the complexity of human-natural system interactions, there is a need for tools for adjusting models across different levels of granularity and type, such as transitioning between agent-based and compartmental modeling approaches. From the data side, important areas include working with social media data, wearables, sensor data, and cell phone data—further development is strongly needed in incorporating these data into models and building uncertainty and estimation approaches for these data (e.g. estimation methods for network data; see priority area on The Mathematical Foundations of Data-Inspired Biology: Learning Rules of Life from Data). Finally, developing forecasting, prediction, and control methods that incorporate this range of data and modeling approaches is critical to predicting the implications of human activity on individuals, ecosystems, and the planet as a whole.

Additionally, the coupling between human and natural systems ties into an enormous range of biological areas, including ecology, population biology, resource management, and wildlife conservation. Human activity has biological implications across a wide range of areas, organisms, and scales. Spatial scales for the impact of human activity range from the cellular and organismal (e.g. plastic and contaminant intake by fish, frogs, etc.), to human and animal populations, to entire ecosystems and the globe. Human impact, climate change, and sustainability issues also span temporal scales—from the immediate effects of natural disasters and infectious disease outbreaks to longer term considerations of climate change for the next 10, 50, or 100 years. Further, the global interconnected nature of human activity means that other types of scope or scale are deeply relevant, such as economic scales of socioeconomic status and political scales and spectra. Finally, there is important potential for translational lessons to be learned from successes in modeling interventions at other scales—it may be possible to translate successes in modeling disease treatment at the within-host level to considering human impact at ecosystem scales.

Indeed, this topic area is particularly interdisciplinary—efforts to understand human-natural systems span a broader range of disciplines than mathematics and biology, including environmental and earth sciences, social sciences (such as anthropology and sociology), history, business, and economics (among others). Solving problems in this area will thus require bridging a wide range of disciplinary silos—which represents both a challenge and an opportunity, as mathematics can potentially provide a way to integrate the data and knowledge generated across different fields to yield insights and predictions for the impact of human activity on biological systems.

Broader Impacts. From a recruitment, training, and retention perspective, it will be critical to develop deeply interdisciplinary training that allows and values researchers who can communicate across the range of disciplines needed to address questions in coupled human-natural systems. Incorporating different disciplinary perspectives and an emphasis on team science will be important to making progress in this area. There are also opportunities for internet and social media-based engagement, as well as the potential for locally focused training that uses local, real-world examples and problems as a focus (e.g.

tackling a local pollutant, endangered species, or invasive species). Along these lines, further engagement with other local, state, and federal agencies (e.g. health departments, wildlife/resource management, or national parks) would be highly beneficial.

Challenges and Opportunities. Many of the underlying challenges in this topic area (as well as rules of life more broadly) involve effectively enhancing conversations and collaborations across disciplinary silos. Building teams spanning social, mathematical, and biological disciplines is challenging, but brings tremendous opportunities for knowledge and impact. This cross-disciplinary challenge crops up at many stages, including training, collaboration, faculty evaluation, as well as funding—highly interdisciplinary projects are often difficult to review (and thus to fund) given the standard disciplinary categories for submission (an issue particularly exacerbated for this topic area, as it spans not just mathematics and biology, but sociological processes as well).

More generally, important questions and challenges include: how do we incorporate human responses into our mathematical and biological models? How do we handle complexity in managing/intervening in ecosystems? How do we best integrate the enormous range of human and environmental data to make useful predictions and understand patterns? And finally, how do we do all of the above given the constraints of time and budget that often must be considered in order to be relevant for decision-making?

6. Broader Impacts: Convergence of Research and Education for the Rules of Life

Summary. In 2012, NIH Director Francis Collins told Science Careers, “If I were a senior or first-year graduate student interested in biology, I would migrate as fast as I could into the field of computational biology... There are vast quantities of high-quality data accessible to anybody who has the skills to find the nuggets of truth that are hiding in that information.” Education in mathematical biology is not currently capturing this excitement. Existing educational programs in math biology are developed ad hoc, with sporadic funding. There must be systematic, scientific investigation of how these programs work, and how the behavior of these systems informs and is informed by the Rules of Life. As noted in the other priority areas, understanding the rules of life will necessitate deeply interdisciplinary training. In the spirit of the Understanding the Rules of Life mission “to train the next generation of researchers to approach scientific inquiry in a way that crosses scales and scientific disciplines,”⁷ the following key areas were identified.

Collaboration between existing and new programs within the NSF. Currently there is limited communication between programs offered through Education and Human Resources and Mathematics and Biology directorates. Investing in bridge programs that close the gaps between high school and college, community and four-year colleges, undergraduate and graduate, graduate to post graduate, and transition to tenure would be highly beneficial. In addition, expansion of existing DMS student research experiences and outreach programs that attract students who may not otherwise consider mathematical biology as area of interest would heighten visibility and provide a bridge between the different education levels. Coordination between mathematical biology researchers and educational researchers has tremendous potential to be mutually beneficial, by improving understanding and effectiveness of recruitment, retention, training, and outreach efforts. An NSF solicitation under the Rules of Life rubric, together with NSF INCLUDES, would be timely and impactful. Outreach to institutions with established records of success in math-bio at all levels could be a mechanism for achieving maximal impact.

⁷ https://www.nsf.gov/news/special_reports/big_ideas/life.jsp

Assessment and evaluation of the impact of workforce programs. Existing REU, RTG, MSPRF programs (and former UBM programs) have been successful in recruiting and retaining a diverse population of students. These programs have a need to evaluate their effectiveness and impact and they also have the potential to generate data useful for educational researchers interested in recruitment and retention, diversity and equity, learning theory, and mathematics education. Collaboration with educational researchers can serve to evaluate and enhance existing programs while also advancing educational theory. This approach provides increased opportunities to build communities of scholars. In parallel, study of successful programs can be used to make such programs more portable, broadening impact in the form of improved curriculum and school experiences for students who are choosing their area of interest, thus increasing outreach. In turn, these improvements in student and community outreach would increase the recruitment watershed for REUs, RTGs, and MSPRFs.

Developing educational innovations for teaching and learning. Educational problems create new opportunities to study and apply the rules of life. Increasingly, existing math bio models (big data and compartmental models) are being used to study educational problems (student performance, placement, recruitment, and retention). Jean Piaget, the founder of constructivist educational theory was originally an evolutionary biologist, and the theory is rooted evolutionary principles. Similarities between biological and constructivist educational theory create new, currently unexplored opportunities to create mathematical evolutionary models of student learning. In addition, mathematical models of genotype (hidden) and phenotype (observable) may have currently unexplored applications to educational theories of understanding(hidden) and behavior(observable). Such models offer new venues for future programs and provides an opportunity for NSF support in the area of evidence-based theories of learning in mathematical biology. This could include faculty development, collaborative partnerships, active learning and new student engagement programs.

Intellectual Merit. Implications for Rules of Life are based on the understanding that (a) educational systems are biological systems, understanding the dynamics of those systems enhances our knowledge of learning and; (b) Understanding human learning and human behavior is critical for understanding the rules of life and interaction between social and biological systems. Key biological areas and scales for educational advancement and cross-disciplinary transmission of ideas include: Genotype and phenotype; Evolution on multiple scales; Contagion processes and; Biological problems that motivate the creation of new mathematics to understand the Rules of Life. Key mathematical areas of emphasis and cross-disciplinary transmission of ideas include Discrete mathematics; Dynamical systems; Big data; Evolutionary models and; Mathematical theories that inform research into the Rules of Life and research in Education. Interdisciplinary training, particularly with a focus on connecting models with data, was identified as an important area. Key educational areas for developing the next-generation mathematical biology workforce include research in Recruitment and retention; Diversity and equity; Learning theory; Mathematics education; Impacting human behavior in biological problems and including international dimensions of Rules of Life problems and; Educational problems that motivate the creation of new mathematical models, informing research into the Rules of Life.

Broader Impacts. One of the key needs for training in the mathematical and biological sciences (and potentially beyond) included: (a) Providing faculty development opportunities through longer term training grants and visits to national centers; (b) Organizing workshops to connect lab scientists, mathematicians, and educational researchers and; (c) Developing a better understanding of current training programs, leading to increased effectiveness and improved understanding of the Rules of Life. Along with this there were also suggestions to successfully recruiting and retaining diverse interdisciplinary cohorts of undergraduate, graduate, and post-graduate scholars calls for persistence and continuity in academic

programs. Discussion on broadening impact was proposed to be achieved through: (a) Technological and curriculum reform programs that cast a broader net and enhance recruitment of diverse students to research programs; (b) Active study of existing funded student research programs that distill successful models into portable components for new programs, broadening impact and; (c) Internship opportunities for undergraduate and high school students, that enhance student interest in understanding the rules of life and provide additional educational research opportunities.

Discussions on how technology can potentially transform the accessibility of ideas and techniques from mathematical biology in areas including: recruitment; curriculum changes; discovery and project-based learning; and textbooks. The need to develop assessments and evaluations for the effectiveness of math biology research experiences as well as broader impact was important. Another broader impact includes how mathematical biology modeling techniques and models including Genotype and phenotype models; Big Data in Education; Compartmental models and; Evolutionary theory, can help inform research and policy education.

F. Recommendations

From on the workshop reports and discussions, several key recommendations emerged. These are summarized as follows, broken down by priority topic area:

1. Understanding Rules of Life in Integrative Biological Systems: Transients and Noise

- a. We strongly recommend development of new solicitation(s) or programs that focus on developing dynamical theory for transients and quasi-stationary states in biological systems. Such theory should be able to reliably distinguish equilibrium states from transient or quasi-stationary states, including in the presence of noise.
- b. Related to the above, we also strongly recommend support for research examining the relationships between transients, noise, heterogeneity, and the distinct effects of experimental and biological variation, including the high risk/high reward of developing new theory and mathematical approaches to accounting for these different types of variation.

2. The Mathematical Foundations of Data-Inspired Biology: Learning Rules of Life from Data

- a. We highly recommend development of programs or solicitations that support the needed foundational and methodological work in areas such as: model and data dimension reduction methods, rigorous estimation in network analysis, linking models with 'big data', statistical inference in mechanistic models, and a unified theoretical framework for parameter estimation, parameter identifiability, and model selection.
- b. Efforts in this area will also critically require close collaboration between data collection and theory, to ensure relevance and impact. We thus strongly recommend a focus on programs that link theoretical predictions with data collection for this priority area. For cross-disciplinary training in this area, we recommend both larger training efforts (e.g. training grants) that provide depth, as well as shorter events such as workshops or "hackathons".

3. Bridging Scales to Understand New Rules of Life

- a. Programs to develop an understanding of how rules of life transition across scales, by building models and approaches that span spatial and temporal scales to uncover consistent biological principles (i.e. rules of life)
- b. Programs to support harnessing biological data across scales and unifying multiscale models between different levels. We recommend that such programs focus on multi-scale phenomena wherein the dynamics at one scale directly affect those at other scales.

4. A New Fitness Landscape for Mathematical Theory: Quantifying Adaptation and Selection in Understanding Fundamental Rules of Life

- a. Programs or solicitation(s) to support further development of analytical techniques, including modeling trait trade-offs, trait plasticity, and complex traits; and incorporating high dimensional data (e.g. sequencing and gene expression data) to better understand the genotype to phenotype map.
- b. Programs to support efforts addressing important open questions, including: What is the most appropriate way to define fitness? How do we reconstruct phylogenetic evolutionary histories of

species with new data? How do we understand community genomics and eco-evolutionary dynamics across species and time scales? We strongly support community-driven efforts to provide consensus and best practices for these active areas. Addressing these questions will build the genotype to phenotype map and generate insight into the fundamental rules of life.

5. Mathematics for the Anthropocene Era: Rules of Life in the Context of Human Impact on Natural Systems

- a. Gaining understanding and developing forecasting, prediction, and control methods will require a range of analytical approaches (e.g. dynamical systems, optimization, game/decision theory, etc.), as well as a range of data (satellite, sensors, social media, cell phone data, etc.). We recommend support for programs that integrate or develop mathematical and data-based methods for understanding human-natural system interactions.
- b. We strongly recommend expansion and development of interdisciplinary programs or solicitation(s) that meaningfully draw from the range of disciplines involved in human-natural systems interactions, which span both mathematics and biology, as well as adjacent fields such as behavioral and social sciences, economics, health sciences, and environmental sciences.

6. Broader Impacts: Convergence of Research and Education for the Rules of Life

- a. Programs for enhancing student research, education and training at the intersection of biological and mathematical sciences that will help broaden their experience and prepare them for undergraduate, graduate and post-graduate study and careers at the interface of mathematical and biological sciences along with understanding of evaluation and impact assessment of these programs, leading to increased effectiveness and improved understanding of the Rules of Life;
- b. Programs for faculty development, including longer-term training grants, conferences and workshops to connect scientists, mathematicians, and educational researchers with the goal of bringing together multiple diverse perspectives to focus on finding innovative cross-disciplinary solutions to grand challenge problems at the interface of mathematical and biological sciences.

Overarching Recommendations

Across all six topic areas, the following recommendations emerged:

- We strongly recommend development and support of interdisciplinary training programs, particularly focused on statistical training and working with data. In particular, we recommend developing integrated training programs blending mathematics, statistics, and biology. This will be critical for the success of all six topic areas.
- Across all six topic areas, a consistent theme was a need for new methodological and theoretical development, combined with meaningful interdisciplinary research that draws from mathematics, statistics, and computation, as well as biology and biology-adjacent (health, social sciences) fields. We thus strongly recommend development and support of solicitations that develop interdisciplinary research and new methods/theory. These efforts will provide a methodological and theoretical foundation for mathematical research in the rules of life umbrella.

G. Overall Broader Impacts of the Workshop

A primary broader impact for the community is likely to be via the generation of the white paper report on the priority areas discussed above. This report will be released to the community, either as a white paper available online, or potentially as a publication in one of the main mathematical biology journals. Additionally, we plan to release several more general overviews of the outcomes of the workshop (we are currently in-progress on a piece for SIAM News, and may also submit to other outlets, such as a letter in the Bulletin of Mathematical Biology), and we are beginning to organize follow-up meetings via minisymposia at major mathematical biology meetings (such as the upcoming SMB meeting in Montreal). Our goal is that the release of this whitepaper may help to inform directions under the Rules of Life umbrella, and spur new questions and collaborations in mathematical biology

Relatedly, a key broader impact of this workshop is the increased collaboration in topics involving mathematical biology to better understand the organizational principles and rules of living systems. Given the interdisciplinary nature of the meeting, we hope that by bringing these diverse groups together we have fostered an exchange of ideas and potential new collaborations on how to tackle new, complex problems in the area of mathematical biology under Rules of Life. Indeed, anecdotally we observed several such conversations between researchers during the workshop.

Additionally, the workshop provided postdoctoral fellows and graduate students in both mathematics and multiple areas of biology an opportunity to present their research and interact with recognized experts in their fields who may offer avenues for future directions for mathematical biology under the Rules of Life. The conference also provided a great opportunity for the next generation STEM workforce, including women and underrepresented minorities, to learn about new research venues that can help them consider pursuing a career in mathematical sciences.

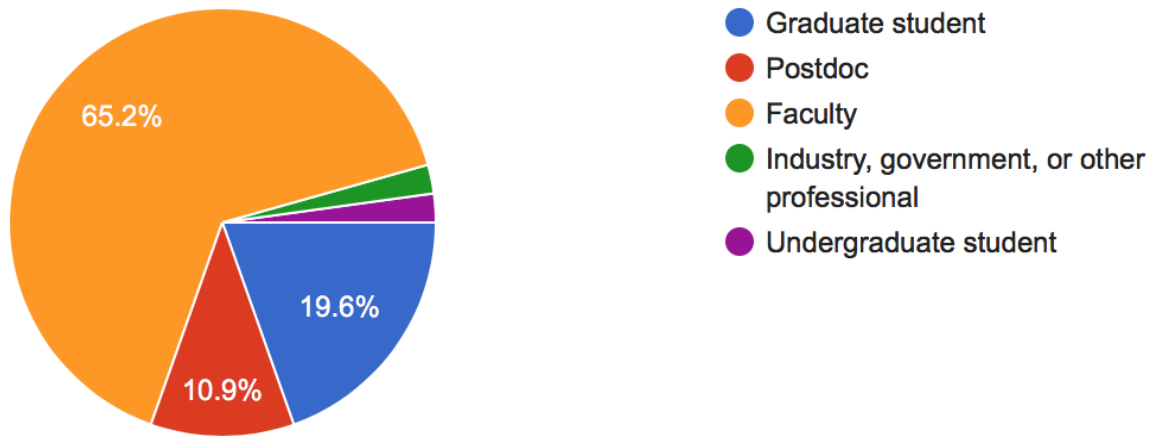
Finally, as part of the workshop, a wide range of useful resources were discussed that are useful for researchers in mathematical biology and for planning for the future of mathematical biology. These resources were collected to a resources page as part of the workshop (available at: <https://docs.google.com/document/d/1gWu5lfx1nVLwESUdmBm0YOTjfxtttNes2bWAYuZknE/>).

H. Appendix

H.1 Participant Demographic Information

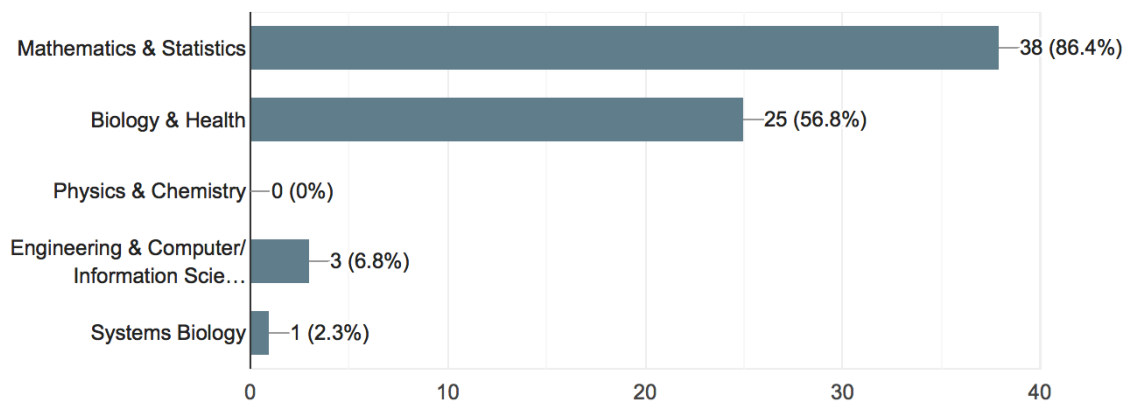
Please note this data includes a small number of individuals who filled out the demographic survey but did not attend the workshop, and it is also missing a small number of individuals who did not fill out the survey.

Workshop participant career stage



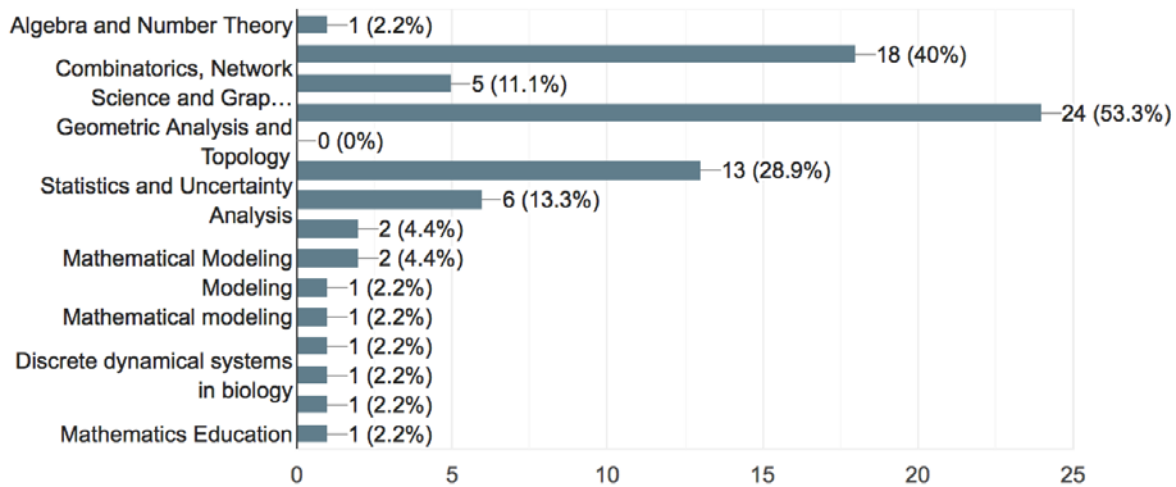
What are your primary subject classification areas? (choose at most two)

44 responses



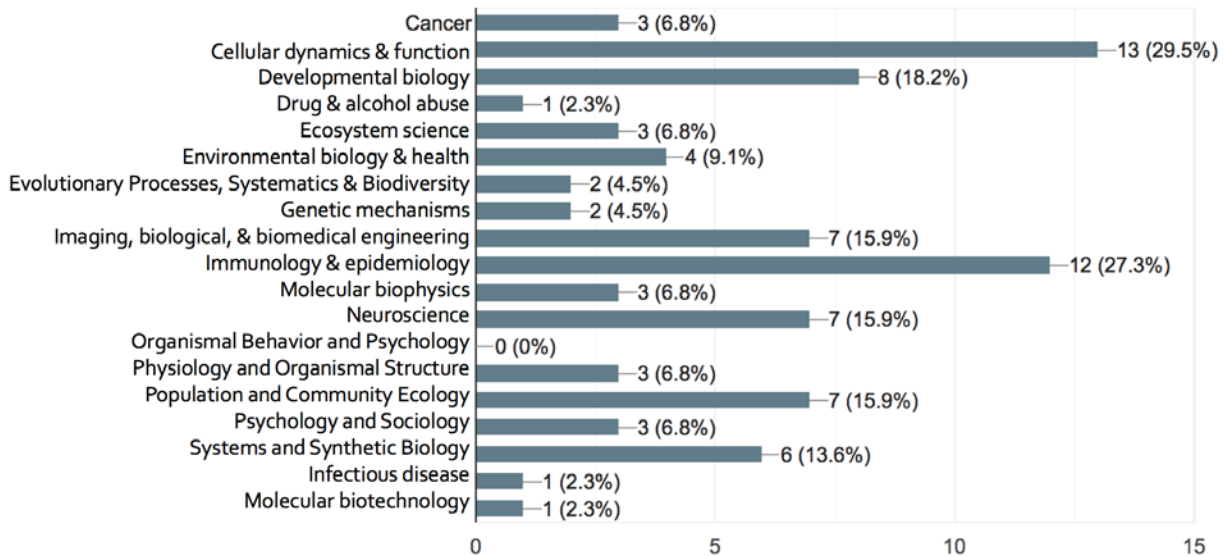
Mathematical research interests (choose at most two)

45 responses



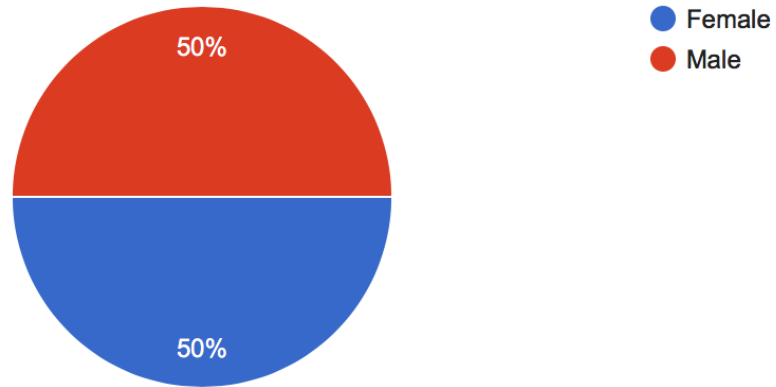
Biological research interests (choose at most two)

44 responses



What is your gender?

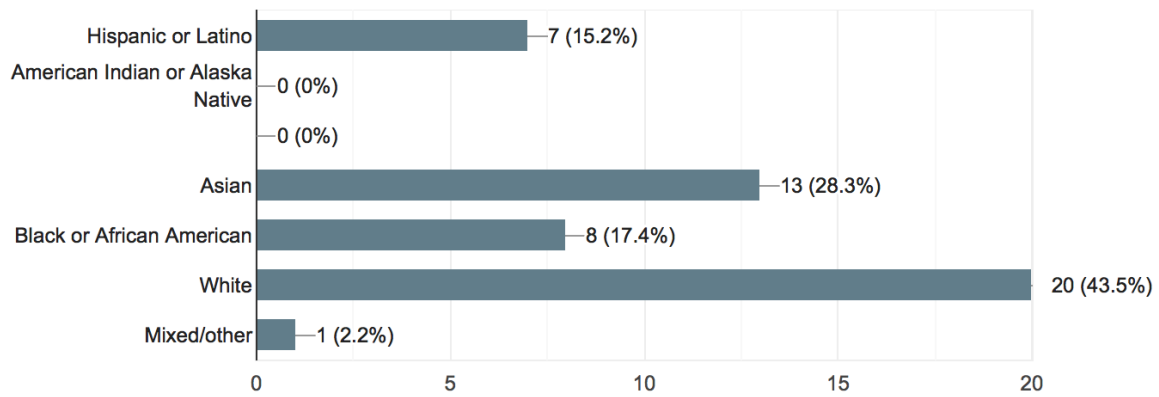
46 responses



Which of the following best represent your racial and ethnic background? Please choose all that apply.



46 responses



H.2 Priority Area Template

The write-up template included the following sections:

1. Priority area summary
2. Intellectual merit
 - a. Key biological areas and scales
 - b. Key mathematical areas
 - c. Implications for rules of life
 - d. Rationale and timeliness (i.e. why is this area exciting, why now)
3. Broader Impacts
 - a. Training - needs for training in the mathematical and biological sciences (and potentially beyond) in order to address this priority area in the field
 - b. Recruitment and retention - how to attract diverse interdisciplinary students and ensure persistence in the academic pipeline
4. Challenges and potential pitfalls
5. Open problems and frontiers