Title: How are rules of life organized across the tree of life?

Authors:
Douglas Bartlett, dbartlett@ucsd.edu
Rong Chen, rchen@som.umaryland.edu
Chris Lane, (clane@uri.edu)
Andres Lopez, (jalopez2@alaska.edu)
Geoffrey Poole (gpoole@montana.edu),
Edward H. Snell (esnell@hwi.buffalo.edu)
Sha Sun (shasun@uci.edu)

Summary: Understanding the rules of life in the context of the tree of life presents exciting opportunities but huge challenges. A hierarchical set of rules provides understanding and predictive capability yielding knowledge and opening up whole new technology areas. It is constrained by gaps in knowledge, common language, and capacity in terms of expertise and capability. Exploring this area would build a highly integrated research community as it would require the integration of existing knowledge and the development of new methods from fields ranging from physics to ecology. Building on and extending the knowledge generated by Tree of Life research, implementation of our vision will generate a richer and more detailed description of that tree of life that includes aspects of genealogical and phenotypic diversity.

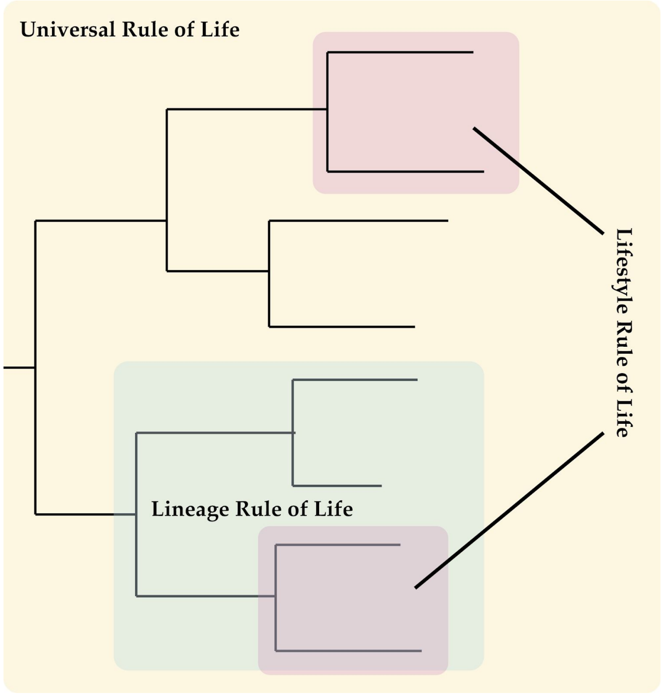
Introduction

At the most fundamental level, rules of life describe matter, energy, and information transmission across the tree of life. Defining and understanding the rules in the context of the tree of life defines how these rules shape evolution across diverse lineages. This enables an exploration of life leading to predictive capacity and the ability to direct evolution to both study and use life to develop novel biological applications. To do this we need to develop a conceptual framework for the rules that link diverse disciplines, scales and timeframes with a common language. This has to be coupled to a transformative experimental effort to understand phenotype diversity across evolution and allow the rules to be tested and verified.

The term ‘Rules of Life’ is used freely yet remains elusive in its definition. Rules provide an operational framework that should be testable computationally, experimentally or both. Like the tree of life, rules are hericirachal starting with basic principles that lead to more complex definitions progressing up the tree. The rules are overlaid on defining characteristics of life - it is ordered, it can take in energy from the environment, it can replicate, increase in size and rebuild and repair. It can reproduce, respond, and adapt to the demands placed on it by the environment.

What is a rule of life?

We envision the rules will be hierarchical, with some being universal and others applying to lineages or lifestyles found across the evolutionary range of organisms (See vision paper 23).



The fundamentals of life, in relation to thermodynamics, cellular requirements and energy flow, are likely to represent the highest order rules. However, a rule need not apply universally, as long as it applies universally under certain criteria (Fig. 1). As an example, there may be rules that apply exclusively to photosynthetic organisms, which appear across the tree and are not monophyletic. Rules may also apply to particular lineages with derived features not found in other organisms.

Figure 1. Hierarchy of possible rules across the tree of life, which could include universal rules (yellow), lineages specific rules (green) and lifestyle specific rules (red).

Opportunities enabled by understanding rules.

Robust knowledge of rules of life provides a system level understanding of life processes across the atomic, organism and ecosystem level. This leads to a groundwork for system level prediction of observable characteristics from the genetic makeup of life and its reaction to perturbations. With a wide enough framework this allows an understanding of life on earth but also opens up possibilities to understand life in alternative environments.

Understanding the rules of life necessarily allows us to identify general principles that underlie characteristics of individuals across the tree of life which elucidates fundamental mechanisms that drive evolution. The rules also help explain robustness and adaptability along all taxa within the tree of life to some extent, informing on the construction of the tree itself.

Interestingly a hierarchical description may challenge current evolutionary structure but in doing so can improve that structure (or offer alternative hypothesis) and guide synthetic biology approaches by overlaying a framework for life that synthetic constructs must abide by. This has broad impact in allowing a process approach to studying life within the context of the complete system and the experimental development of alternative pathways. It provides a framework to harvest highly efficient biological approaches for industrial application.

Understanding the rules of life also guides computational biology and bioinformatics approaches by providing a set of principles that algorithms and models must follow. At the same time, investigating the rules provides opportunities for the exploration of new theories and predictive modeling. A rule breaker implies a gap in knowledge that could yield exciting new biological information.

Understanding the rules has to be an iterative process where some rules will be unknown and other proposed rules will be wrong. These will be refined by the identification of rule breakers which define knowledge areas that when explored will produce new biological knowledge not expected from existing information. This can lead to innovative research topics and new fields of scientific study. Knowing the rules of life also allows us to understand the origin of death - a set of rules that are probably far more simple and fundamental than those for life.

From a workforce perspective research on understanding the rules of life across the tree of life necessarily brings interdisciplinary approaches and critical thinking skills, which provides training and outreach opportunities to future scientists.

Strategies for Developing and Validating the Rules.

We will need to tackle challenging conceptual and methodological issues to identify and map rules of life across the tree of life. Meeting these challenges will require recruiting a community of engaged investigators to coordinate the effort and most importantly, help define conceptual, technical and analytical standards that can serve as a sound scaffold for building this new knowledge. The efforts of this community may be organized around three distinct but interconnected areas: theoretical foundations and alignment of concepts, selecting and developing efficient data production methods, and fostering refinements and extensions of these approaches based on early outcomes.

Foundations and concepts

An essential building block of the effort to determine how different life's principles vary or not on the tree of life is a well-defined and robust set of concepts and theories with details and characteristics that are accepted within the research community. To reach those consensus definitions, the community or a collaborative network of researchers with a stake on this process will need to undertake discussions with the specific goal of arriving at shared terms and operational concepts that can be applied across diverse areas of phenotypic and phylogenetic research.

Beyond a shared conceptual framework, this community will need to produce an updated synthesis of the state of knowledge of the tree of life to serve as the starting road map. This synthesis would serve to delineate important gaps in our knowledge of diversity and evolutionary affinities, and to identify segments of the tree of life that may be particularly promising targets for focused research.

Tools, data production and analyses

An important barrier to overcome on this effort relates to the speed with which phenotypes can be measured and coded. Current approaches to phenotype analysis are relatively slow and tend to focus on anatomical features (See Vision Paper 16). A rules of life research program must include all aspects of the phenotype space (See Vision Paper 2). For this reason, there needs to be a strong investment on the development of high throughput, semi- or completely automated platforms to measure anatomical, physiological and behavioral traits. Fortunately, technology to measure the genotype has seen dramatic advances in the past two decades, which will foster efforts to map discoveries from phenotype-space to their underlying genetic bases.

Efficient methods of measuring the genotype will need to be matched with advances in analytical approaches and tools. These methods should be consistent with and appropriate for the conceptual framework established by the ‘rule of life’ community. Importantly, they need to be implemented in robust computational tools that follow best-practices for reproducibility (e.g. open source and richly documented). Without an accessible ecosystem of analytical tools, efforts to characterize and map rules of life on the tree of life are likely to remain limited in scope and unlikely to serve as foundational to other lines of research.

Synthetic biology and model systems are two approaches that need to be adopted to test and push the rules that are established. A significant investment will also be necessary in transparent Artificial Intelligence to analyze the large amount of data produced and extract meaningful signals.

Interpretation, dissemination and long-term promotion of ‘rules of life’ discoveries

The implementation of the vision outlined here presents an additional important opportunity to lead in the creation and refinement of new models for discovery dissemination that builds and extends open access ideals. For example, this community could build an online platform to organize, document and report ongoing activities in rules of life research. This platform would serve as a near ‘real time’ window on new findings and work in progress. Accomplishing this ambitious goal will require a community that is strongly committed to revolutionizing research culture in a way that promotes faster and freer sharing of samples, data, data products and findings. We envision a platform that leverages outcomes from past and ongoing Tree of Life research and extends those products by contributing new layers of phenotype measurements. An important challenge in the development of this platform will be to create intuitive visual interfaces that allow researchers from across diverse fields to locate, parse and obtain information of interest.

Educational programs and outreach activities on discovering the rules of life

Student training programs will be designed to provide opportunities to undergraduate and graduate students, with the emphasis on interdisciplinary thinking, which will utilize the platforms and knowledge generated by the research effort on understanding the rules of life. Such programs will bring a driving force for the next generation of continuous research innovation (See Vision Paper 13). Community outreach programs will incorporate the conceptual understanding of the rules, which will stimulate thinking and discussion that engage society and bring broader interest from diverse backgrounds. Outreaching activities that bring conceptual understanding of the rules will stimulate thinking and discussion that engage all communities.

Barriers and breaking them.

Understanding the rules of life in the context of the tree of life is compelling on numerous levels as described above. However there are significant barriers to the effort which make it extremely challenging. These barriers are in the areas of defining a rule, gaps in our knowledge that impede the development or validation of a rule, the diverse nature of the tree of life and a common language across all the disciplines that are required to study it, and the breadth of expertise and effort required. Practical questions include managing the data (See Vision Paper 21), organizing a collective international effort, and how to focus the effort for efficiency, i.e. not having to literally study everything.

Defining a rule requires an operational framework that is testable, both computationally and experimentally (See Vision Paper 23). It is highly likely that rules of life are arranged in a hierarchical context whereby foundational rules apply across the tree, but others may apply only in different lineages (not necessarily monophyletic; Fig. 1). Rules of life may be considered as the set of axioms that can be used jointly to logically construct the tree of life. From this perspective, there could be multiple sets of rules which are equivalent in their capability to construct the tree of life. Examples of candidate fundamental rules of life are cell theory, gene theory, evolution, homeostasis, and thermodynamics.

To test and validate rules extensive experimental data is required. There is a knowledge gap that has to be filled in a cost effective and timely manner. High throughput high-capacity approaches of phenotype assessment that provide high information output will need to be developed, potentially as a central service core or cores on a massive scale. This will also require a common set of ways to code phenotype so that the wide community can understand and process the data that is produced.

Many disciplines have their own unique description of different terms. For example “differentiation” and “vector” mean one thing to a mathematician but another to a biologist. A framework or thesaurus to bridge this language gap needs to be agreed upon and established.

A common ontology is needed to describe rules, results, and information across the diverse taxonomic groups and scientific disciplines and allow effective use of this information to discover new rules.

The diversity of life is massive. This leads to an expertise gap in both the skill sets and number of people who can tackle the problem. Automation and algorithm development will help to some extent, but a major goal to succeed in this effort is to build an organized international community to generate and interpret data from the full range of global habitats and across taxonomic lineages. This same community has to have educational goals to expand the workforce and to build the generation of scientists that may be able to use the rules of life in an applied manner.

The issue of diversity is particularly daunting for microbes, including bacteria, archaea and protists. Uncovering novel properties of these organisms will depend on developing scalable high-throughput approaches, especially where they can be used for the most labor-intensive activities. Advances in single-cell methodologies and microcosm incubations under in situ or environmentally relevant conditions hold great promise. The document entitled “Discovering new Microbial Functions: Ex Situ to In Situ” addresses some of these issues.

- Data Management: Effective management of phenotypes and associated metadata will be critical for this effort. Phenotype data generated and used in this effort will come from a diverse set of fields and approaches from physics to ecology. As a result, designing and implementing data coding and annotation structures that can be readily integrated and efficiently accessed by the broader community will be a critical building block toward implementing our vision. Fortunately, a number of partial solutions exist and are being devised to house and distribute specific types of datasets (e.g. Brenda, Savio, Morphobank, NCBI databases). We envision leveraging data management knowledge from those efforts combined with novel technologies for massive-scale database design and operation to create a ‘rules of life on tree of life’ data platform. This component of our effort will necessarily invoke expertise from computer science and engineering.

How does it reintegrate biology

Defining and testing the rules of life is a cross-cutting effort that will require contributions from all fields of biology and other disciplines, including math, physics, chemistry, computer science, and engineering. Interdisciplinary teams will be required to robustly test rule proposals as they arise and determine the bounds of rules that are not universal. Within biology, testing the rules of life across scales will draw from molecular biology through ecology. This fundamental research will influence synthetic biology, medicine, agriculture and climate change science, among others.

DIscussion

Studying the rules of life in the context of the tree of life presents many opportunities but is also constrained by barriers described above. A hierarchical set of rules anticipates concepts that scale across the complexity of systems from cells to whole ecosystems. To focus the study we need to define the window and timeframe that our rules apply to. This in turn directs the research focus to fill the critical knowledge gap that allows us to accurately define those rules. At the extremes we cannot define the origin of life, nor do we expect to define cosmological aspects. We aim to define life as we know it within our own timeframes but have a scalable concept that can be extended beyond this.

We expect the rules to span disciplines and be related to biological, chemical, physical, and mathematical concepts. This requires a workforce that is literate in the overlap between these disciplines and others. We assume that there are universal rules and rules that are very specific to different branches of the tree. While universal rules will be common, specific rules may contradict each other across different branches. It is probable that rules can be combined to yield higher order rules.