Mathematical Tools for Evolutionary Systems Biology

Laurence Loewe (University of Wisconsin - Madison),
Peter Swain (University of Edinburgh),
Ryan Gutenkunst (University of Arizona)

May 26 - May 31, 2013

Mastering many practical problems in our world demands a deeper and more integrated understanding of biochemical and evolutionary processes than we can currently muster. For example, to fight bacterial infections we need to understand the biochemistry of bacterial phenotypes for designing efficient antibiotics and we also need to understand the population genetics and ecology of these bacteria to design antibiotics usage policies that reduce the speed with which antibiotics resistance evolves [18]. To fight cancer we need to understand the signaling pathways that allow cells to ignore growth control and we also need to understand the population genetics that allow cancer cells to maintain a competitive edge, diversify, migrate and evolve resistance against chemotherapy [9]. Similar challenges of integrating molecular biology and evolutionary biology exist for controlling virus infections, developing bioenergy solutions and many other hot topics. Building bridges between systems biology and evolutionary biology has the potential to provide us with the tools to build and test the hypotheses required for addressing these problems. Developing such an integrated approach is, in a nutshell, the goal of the emerging field of evolutionary systems biology.

Evolutionary systems biology is a broad field, and its key questions and the best ways of attacking them are still being defined. Progress in evolutionary systems biology demands close collaboration between experimentalists, theorists, and computational modelers. This workshop set out to bring together a diverse group of researchers from all three sub-disciplines to share problems and progress, with a focus on connecting data with models.

1 Overview of the Field

In the last decade many have pointed to 'systems biology' as a research program that integrates knowledge, mostly about molecular and cellular systems, to arrive at a deeper understanding. Indeed, after decades of investigating biochemical reactions and molecular processes in isolation, it was time to start putting the pieces of the puzzle together to help see the bigger picture [26]. The development of sophisticated mathematical methods for simulating the dynamics of complex non-linear models [8] provided a formalism which enabled the precise description of integrated models. While there is no accepted definition of 'systems biology' many would say that the repeated refinement of mathematically precise integrated models is fundamental to most systems biology. Such activities have led to much progress in the area of molecular systems biology and other fields that study processes within organisms.

As pointed out repeatedly, the theory of evolution is *the* unifying theory in biology. Since the founding fathers of population genetics started investigating how mutation, selection, genetic drift, recombination and migration work together to shape the evolution of life, biologists have worked towards integrating an ever increasing body of evidence into the theory of evolution, ranging from molecular details to ecosystems. Evolutionary biology, ecology, population genetics and related fields have developed substantial mathemat-

ical expertise in building mechanistic models, analyzing stochastic patterns and evaluating theories (e.g., [5, 16, 27]).

Evolutionary systems biology seeks to develop a unified understanding of the evolution and function of molecular systems [13]. As such, it draws on insights from systems biology to define how molecular systems function to create phenotypes and how those phenotypes are altered by mutations and other perturbations. Mutations that alter organismal phenotypes often also alter organismal fitness, upon which natural selection acts. Models from evolutionary biology can then be leveraged to understand the fate of those mutations. Eventually, evolutionary systems biology promises to enable predictive modeling of molecular systems across both functional and evolutionary scales. Reaching that goal, however, demands addressing several outstanding open problems.

2 Recent Developments and Open Problems

Both focal points of integration have their strengths and weaknesses. Building on its molecular or physiological roots, systems biology can produce extraordinarily elegant mechanistic models of processes that occur within cells or organisms; however it often ignores evolutionary processes or simply assumes that all results of evolution are always optimal. Likewise, population genetics and ecology can produce extraordinarily elegant mechanistic models of evolutionary processes, which can be defined as processes that involve any intergenerational inheritance of hereditary material in self-replicating entities; however, these disciplines often collapse all complexities of molecular biology and physiology into a few numbers like fitness, selection coefficients, survival rates. While such discipline-specific assumptions are very powerful and have led to much progress by separating domains of concern, they have also started to break down in recent years as our models have become more realistic. For example, evolutionary biologists have become increasingly interested in the molecular mechanisms of their study systems. Likewise, systems biologists are realizing that some of the problems they are interested in actually involve inheritance of genomes over many generations and require an understanding of evolutionary factors. Bringing together both areas of biology is set to stimulate many new interesting research questions and insights [23].

A key challenge for both systems biology and evolutionary biology is the complexity of the systems they study. Indeed much effort in research is dedicated towards finding mechanisms and abstractions that can reduce that complexity when building models without losing either touch with reality or the ability to analytically understand at least some key aspects of a model. Evolutionary systems biology inherits the same challenges, however, it can also offer additional tools. We know that the same systems have to satisfy two constraints: they have to work mechanistically within individuals and they have to satisfy evolutionary constraints. This fact could be used in a similar manner to Occam's razor: use evolutionary considerations to rule out unrealistic biochemical models or use biochemical considerations to rule out unrealistic evolutionary models. Both uses have the same effect that Occam's razor is known for: keep models as simple as possible, but not simpler. This idea is masterfully implemented in flux-balance analyses [15], where valid models (i) have to satisfy the stoichiometry of known biochemical reaction networks and (ii) are then optimized for a property that limits growth (e.g. biomass production). Resulting predicted outcomes of evolution of *E. coli* cells could even be verified in the lab [10]. It is easily conceivable that similar uses of Occam's razor could simplify other modeling scenarios with more complicated dynamics or, for example, by using analyses of quantitative trait loci to determine a credible space of systems biology models.

Evolutionary theory and population genetics give strong predictions about the trajectory of a mutation with a given effect on fitness, but much less is known about the input spectrum of mutational effects [6]. This is because natural selection acts not directly on the genetic mutation itself, but rather through the mutation's effect on phenotype, and the genotype-to-phenotype map is very complex, frequently involving very small mutational effects. Systems biology decomposes the genotype-to-phenotype map by focusing on mechanistic descriptions of the molecular components and interactions that ultimately generate the phenotype. This pathway and network perspective provides a framework for understanding and predicting evolutionarily important properties of mutations, including pleiotropy (the degree to which a single mutation effects multiple phenotypes), epistasis (the degree to which mutations interact), and effect sizes. Integration with systems biology thus offers to make evolutionary biology much more predictive, because mutational effects can be understood much more deeply.

High-throughput sequence determination has developed remarkably fast, and high-throughput phenotyping of individuals must catch up. Recent work in high-throughput yeast morphology phenotyping though imaging is a promising example [28]. Additional progress in dynamically measuring gene expression or protein interaction in large numbers of individuals for large numbers of markers will allow for a much more direct connection between morphological phenotypes and underlying biomolecular changes. This need for high-throughput phenotyping extends all the way to individual molecules. General properties of the effects of random mutations on proteins or binding sites are an important input to quantitative models of system evolution, but they remain to be measured. Developing high-throughput phenotyping assays demands technological innovation, but the payoffs are potentially very large.

Although we now know a great deal about divergence in genetic sequence between species and between individuals within a species, very little is known about divergence in pathways and other molecular functions. A systematic effort to generate such data is critical to progress in evolutionary systems biology. Anecdotal examples suggest that even when a system uses orthologous molecular components, their interactions may differ dramatically [20], but the generality of this phenomenon is unclear. We anticipate that it will be particularly fruitful to study relatives of existing model organisms, because measurement techniques will be more transferable, and new data sets can be integrated with a wealth of existing information. Understanding differences between species will give insight into what features of molecular networks are targeted by natural selection. Extending such measurements to the more subtle differences between individuals will give insight into the system-level variation on which natural selection can act.

It will also be important to incorporate what is known about molecular biology into sequence analysis. For example, existing knowledge of biological pathways and molecular function can provide useful priors for GWAS analysis in connecting phenotypes to genetic changes. This will provide additional information to the important and difficult task of identifying genetic changes that were driven by selection and those that evolved by neutral drift. Further clarifying this distinction will finally answer a longstanding question in evolutionary biology: does evolution proceed more often by changing protein sequence or by changing regulation? The answer to this question may vary, depending on taxonomic class and the selective forces, but it is an essential link to bridge evolutionary and systems biology.

3 Presentation Highlights

Evolutionary systems biology is a very broad field, and the talks at this Workshop were correspondingly diverse. Here we highlight a fraction of the Workshop presentations, in order to convey the breadth of the field. We first highlight presentations on biological problems, then highlight presentations focused on computational methods. *Many presenters shared unpublished results, and to preserve their confidentiality, we do not discuss details of those presentations.*

Metabolic networks, which govern energy and material flows within organisms, are among the best-studied model systems for evolutionary systems biology. Frank Bruggeman gave a wonderful overview of the many selection pressures experiments may impose on the metabolism of microbes, and how those pressures may be modeled. Serial dilution experiments select for fast growth rate, chemostats select (potentially) for affinity for the limiting nutrient, and recent experiments in droplets select for yield given a limiting nutrient. Flux-balance analysis provides a powerful way for modeling steady-state metabolic process, particularly because such models typically require much less information about the system than dynamical models. Consideration of the metabolic benefits of an enzyme and the cost of producing it can then define a fitness landscape, over which the expression of multiple metabolic enzymes can be optimized [2]. Working on a larger scale, Elhanan Borenstein discussed using genome-scale models of metabolic networks to infer the environments that particular microbes exist in, a sort of "reverse ecology" [11]. Given the growing interest in microbial communities, this approach can be broadened to considering the metabolic networks of entire communities, for example, in the human gut [12].

In many microbial communities, growth on surfaces and subsequent biofilm production is important for efficiently gathering resources and resisting toxins (such as antibiotics) [3]. Joao Xavier presented his recent experiments on the bacterium *Pseudomonas aeruginosa* [25]. When grown on swarming media, Pseudomonas colonies repeatedly evolve a "hyperswarming" phenotype. These hyperswarmers outcompete the wild type by filling the whole plate, rather than restricting their growth to a branched pattern. Remarkably,

Xavier found that this phenotype was driven by parallel amino acid substitutions in a single gene, *fleN* . These mutations result in the growth of multiple flagella, changing the bacterial movement pattern. This is a remarkable example of convergent evolution, not only in the phenotype in the bacteria, but also in their molecular systems.

Over time, cells accumulate non-genetic damage that may contribute to aging, both in animals like ourselves and in bacteria. In animals, the aging clock is reset upon fertilization of the eggs, but for microbes that divide symmetrically, that clock is never reset. Lin Chao presented his modeling work on the accumulation of the "damage load" of damaged biomolecules in microbes [4]. Chao showed using a model and dynamical systems arguments that below a critical threshold rate of damage accumulation, cells are immortal. Above that threshold, cells must divide asymmetrically, to segregate damaged biomolecules in one "older" cell while providing the "younger" cell with fresh biomolecules. Comparing his model with recent microfluidics experiments on *E. coli*, he showed that they divide almost perfectly symmetrically, suggesting that damage accumulation is slow for this bacteria. This work highlights the interplay between biomolecular effects of damage and evolutionary life-history effects of division asymmetry.

Biomolecular damage may contribute to aging, but molecular mistakes may play an important role in evolution. In particular, Joanna Masel presented work on cryptic sequences. These are regions of the genome that are only rarely expressed [14]. For example, sequences past the normal stop codon in a protein are only translated if there is an error in reading the stop codon. These sequences thus evolve under very low purifying selection and can accumulate diversity. If, however, stressful conditions lead to their expression, that diversity can be harnessed because it may hold adaptive mutations. Recent work from the Masel group has analyzed the trade-offs in this sort of system. Does evolution favor systems in which errors are very rare but have dramatic phenotypic effects, or in which errors are common but have small phenotypic effects? Intriguingly, the answer depends on population size [19]. Particularly important is that the common error scenario enhances evolvability, by increasing the range of possible genotypes and phenotypes a population can access. This again suggests how particular molecular mechanisms (and the errors they create) can have profound impacts on the evolutionary process.

A major challenge for evolutionary systems biology is to map the consequences of specific genomic sequences. A particular challenge is epistasis, in which the phenotypic effect of a mutation depends on the context of other mutations. Ilya Nemenman shared his recent work examining epistatic effects in the *E. coli lac* promoter [17]. Nemenman took advantage of a recent data set that measured the expression driven by roughly 129,000 different mutagenized versions of the 75 basepair promoter. By fitting first-, second-, and higher-order statistical models, Nemenman showed that about two-thirds of the variance in expression is explained by the first-order non-epistatic effects. This was unexpectedly low, suggesting that the phenotypic landscape for this promoter is relatively simple. This is an optimistic message for researchers attempting to understand phenotypic landscapes in more complex systems.

A particularly important complexity in biology is that phenotypes typically arise from the interactions of multiple proteins and the resulting systems often have strongly nonlinear dynamics. Michael Savageau presented his S-system formalism for modeling such systems [22]. An S-system is a system of ordinary differential equations in which the right-hand sides are all sums of terms which are products of power-law functions. This generalization of mass-action kinetics is a good approximation for many biochemical systems, and the special form of the equations allows the development of specialized and powerful numerical and analytical solution techniques. Particularly intriguing is the use of this form to rigorously define biochemical phenotypes [21], by defining the regions in parameter space in which particular terms dominate. Once rigorously defined, the robustness of phenotypes to changes in parameters can be efficiently studied by analytic and numerical approaches. The S-system approach offers great promise for analytically understanding the qualitative phenotypes possible from a biochemical system, but simulation remains a backbone of evolutionary systems biology.

Stochastic simulation is particular relevant to evolutionary systems biology, because evolution is inherently stochastic, and many of the biochemical systems of interest employ small numbers of molecules. Linda Petzold has long been on the forefront of numerical simulation, both deterministic and stochastic, and she summarized the development of stochastic simulation methods and her group's recent software developments. In particular, approximation methods that improve the efficiency of Gillespie's classic exact algorithm have become increasingly important, be they tau-leaping or hybrid methods that incorporate deterministic simulations. In addition to generating simulations, techniques for capturing rare events and optimizing

parameters are essential. A new frontier is spatial stochastic simulation, in which molecules are tracked in space as well as in time. Petzold's group has recently released StochSS, a cloud-based platform that is incorporating a growing number of sophisticated stochastic simulation tools into an easy-to-use package: `http://www.stochss.org`. This holds great promise for modeling complex dynamic systems.

In addition to developing improved simulators, the theory of stochastic simulation is also improving. In particular, David Anderson present his recent results on using multi-level Monte Carlo to estimate expectations of stochastic processes [1]. This approach offers much faster estimation of these quantities than simply averaging over many exact simulations, yet it avoids many of the biases that can be introduced by approximate simulations. Calculation of such expectations is particularly important for fitting simulations models to data, and a key goal of evolutionary systems biology.

In comparing models with data, a key step is optimizing parameter values to best match that data. Michael Ferris presented an overview of optimization techniques from many fields that have the potential to impact evolutionary systems biology. For example, compressed sensing and the LASSO method aim to minimize the number of parameters needed to explain the data, thus providing a statistical implementation of Occam's Razor. Ferris was also helpful in responding to attendees particular needs for optimization advice.

A particular challenge of systems biology is that the molecular agents that interact are themselves complex objects. James Faeder highlighted how molecular bindings can result in a combinatorial explosion of potential molecular species. This explosion can be handled well by rule-based modeling approaches [7]. These are appealing because they allow for a high level of abstraction and can in principle capture many molecular interactions that are often neglected when modelers study hand-built ODE models. Faeder provided an overview of progress in developing rule-based modeling tools (see `http://bionetgen.org`) and highlighted in particular his group's recent work on Atomizer, a method for extracting the molecular rules that are implied by a given set of differential equations [24]. Applying this approach to existing models reveals ways in which the implicit approximations made by the modelers have resulted in altered behaviors. It also offers the possibility of easily coupling those models onto models of related systems, a process that to date has necessitated careful and laborious manual checking.

4 Outcome of the Meeting

Evolutionary systems biology is a young emerging field, and its practitioners are spread across mathematics, computer science, systems biology, and evolutionary biology. Thus a key goal of this meeting was to bring together researchers with common interests to share problems and ways of attacking them. In that, we think the meeting was a great success. The biology-oriented presentations highlighted a large number of important and intriguing questions that evolutionary systems biology is addressing. The simulation-oriented presentations highlighted both established and new tools for addressing those questions. It was clear from the discussion of the participants that many felt this was an important meeting for creating the evolutionary systems biology community. Many participants bonded both socially and scientifically with colleagues they would never have encountered in their normal circuit of conferences. Forming those connections is, we believe, one of the most important goals for meetings in an emerging field.

We hope to carry the energy and enthusiasm of this meeting forward into future meetings at BIRS. On the last day of this meeting, the participants collectively developed plans for a follow-up meeting in 2015. A goal of that meeting will be to focus on more analytical approaches to understanding biochemical networks. This meeting made it clear that simulation technology is progressing rapidly, but simulating biochemical systems in a whole population of individuals is likely to remain daunting for the foreseeable future. Thus progress in evolutionary systems biology will be greatly enhanced by improved theoretical understanding of the dynamics of biochemical reaction systems. Those theoretical developments will be most powerful, however, if they are guided by experimental insights. Thus that future meeting will also bring together experimentalists, modelers, and theorists.

Evolutionary systems biology is an exciting emerging field in mathematical biology. Integrating a molecular understanding of life with broader evolutionary principles promises to shed new light onto important scientific and societal problems. Because the field sits in between traditional disciplines, independent meetings such as the present are critical to pushing the field forward and nucleating a new community of evolutionary systems biologists.

References

[1] D. F. Anderson and D. J. Higham, Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics, *Multiscale Modeling & Simulation* **10** (2012), 146–179.

[2] J. Berkhout, E. Bosdriesz, E. Nikerel, D. Molenaar, D. de Ridder, B. Teusink, and F. J. Bruggeman, How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism, *Genetics* **194** (2013), 505–512.

[3] K. E. Boyle, S. Heilmann, D. van Ditmarsch, and J. B. Xavier, Exploiting social evolution in biofilms, *Current Opinion in Microbiology* **16** (2013), 207–212.

[4] L. Chao, A model for damage load and its implications for the evolution of bacterial aging, *PLoS Genetics* **6** (2010), e1001076.

[5] B. Charlesworth and D. Charlesworth, *Elements of Evolutionary Genetics*, Roberts and Company Publishers, Greenwood Village, Colorado (2010).

[6] A. Eyre-Walker, P. D. Keightley The distribution of fitness effects of new mutations, *Nature Reviews Genetics* **8** (2007), 610–618.

[7] J. R. Faeder, M. L. Blinov, W. S. Hlavacek, Rule-Based Modeling of Biochemical Systems with BioNetGen, in *Systems Biology* (I. V. Maly, ed.), volume 500 of *Methods in Molecular Biology*, Humana Press, Springer, New York (2009), 113-167.

[8] D. T. Gillespie, Stochastic simulation of chemical kinetics, *Annual Review of Physical Chemistry* **58** (2007), 35–55.

[9] M. Greaves and C. C. Maley, Clonal evolution in cancer, *Nature* **481** (2012), 306–313.

[10] R. U. Ibarra, J. S. Edwards, and B. O. Palsson, Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth, *Nature* **420** (2002), 186–189.

[11] R. Levy and E. Borenstein, Reverse ecology: from systems to environments and back, in *Evolutionary Systems Biology* (O. S. Soyer, ed.), volume 751 of *Advances in Experimental Medicine and Biology*, Springer, New York (2012), 329–345.

[12] R. Levy and E. Borenstein, Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules, *Proceedings of the National Academy of Sciences U S A* **110** (2013), 12804–12809.

[13] L. Loewe, A framework for evolutionary systems biology, *BMC Systems Biology* **3** (2009), 27.

[14] J. Masel, Cryptic genetic variation is enriched for potential adaptations, *Genetics* **172** (2006), 1985–91.

[15] J. D. Orth, I. Thiele, and B. O. Palsson, What is flux balance analysis?, *Nature Biotechnology* **28** (2010), 245–248.

[16] S. P. Otto and T. Day, *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*, Princeton University Press, Princeton, New Jersey (2007).

[17] J. Otwinowski and I. Nemenman, Genotype to phenotype mapping and the fitness landscape of the E. coli lac promoter., *PloS ONE* **8** (2013), e61570.

[18] R. Pena-Miller, D. Laehnemann, G. Jansen, A. Fuentes-Hernandez, P. Rosenstiel, H. Schulenburg, and R. Beardmore, When the most potent combination of antibiotics selects for the greatest bacterial load: the smile-frown transition, *PLoS Biology* **11** (2013), e1001540.

[19] E. Rajon and J. Masel, Evolution of molecular error rates and the consequences for evolvability, *Proceedings of the National Academy of Sciences U S A* **108** (2011), 1082–1087.

[20] M. Sanetra, G. Begemann, M.-B. Becker, and A. Meyer, Conservation and co-option in developmental programmes: the importance of homology relationships, *Frontiers in Zoology* **2** (2005), 15.

[21] M. A. Savageau, P. M. B. M. Coelho, R. A. Fasani, D. A. Tolla, and A. Salvador, Phenotypes and tolerances in the design space of biochemical systems, *Proceedings of the National Academy of Sciences U S A* **106** (2009), 6435–6440.

[22] M. A. Savageau and E. O. Voit, Recasting nonlinear differential equations as S-systems: a canonical nonlinear form, *Mathematical Biosciences* **87** (1987), 83–115.

[23] O. S. Soyer, ed., *Evolutionary Systems Biology*, Springer, New York (2012).

[24] J. J. Tapia and J. R. Faeder, The Atomizer: Extracting Implicit Molecular Structure from Reaction Network Models, *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics* (2013), 726–727.

[25] D. van Ditmarsch, K. E. Boyle, H. Sakhtah, J. E. Oyler, C. D. Nadell, E. Déziel, L. E. P. Dietrich, and J. B. Xavier, Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria, *Cell Reports* **4** (2013), 697–708.

[26] H. V. Westerhoff and B. O. Palsson, The evolution of molecular biology into systems biology, *Nature Biotechnology* **22** (2004), 1249–1252.

[27] M. C. Whitlock and D. Schluter, *The Analysis of Biological Data*, Roberts and Company Publishers, Greenwood Village, Colorado (2008).

[28] N. Ziv, M. L. Siegal, and D. Gresham, Genetic and nongenetic determinants of cell growth variation assessed by high-throughput microscopy, *Molecular Biology and Evolution* (2013).