Chapter 15
Reverse Ecology: From Systems to Environments and Back

Roie Levy and Elhanan Borenstein

Abstract The structure of complex biological systems reflects not only their function but also the environments in which they evolved and are adapted to. Reverse Ecology—an emerging new frontier in Evolutionary Systems Biology—aims to extract this information and to obtain novel insights into an organism's ecology. The Reverse Ecology framework facilitates the translation of high-throughput genomic data into large-scale ecological data, and has the potential to transform ecology into a high-throughput field. In this chapter, we describe some of the pioneering work in Reverse Ecology, demonstrating how system-level analysis of complex biological networks can be used to predict the natural habitats of poorly characterized microbial species, their interactions with other species, and universal patterns governing the adaptation of organisms to their environments. We further present several studies that applied Reverse Ecology to elucidate various aspects of microbial ecology, and lay out exciting future directions and potential future applications in biotechnology, biomedicine, and ecological engineering.

1 Introduction

Adaptation is the cornerstone of ecological and evolutionary theory. The various traits that allow an organism to survive successfully in a certain niche evolved over generations of natural selection, shaping the interface between the organism and

R. Levy
Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

E. Borenstein (✉)
Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

Department of Computer Science and Engineering, University of Washington,
Seattle, WA 98195, USA

Santa Fe Institute, Santa Fe, NM 87501, USA
e-mail: elbo@uw.edu

O.S. Soyer (ed.), *Evolutionary Systems Biology*, Advances in Experimental Medicine and Biology 751, DOI 10.1007/978-1-4614-3567-9_15,

the environment. The genomes of the various species in a given habitat are the end result of such selective pressures, collectively encoding the iconic "entangled bank" that Darwin describes in the concluding remarks of his *On the Origin of Species*. However, the relationship between an organism's genome and the manner in which it interacts with its environment, as well as with the myriad species with which it cohabits is dauntingly complex, and, in many cases, extremely challenging to decipher.

Variation among ecological traits within a population is often taken as evidence of adaptation to a specific environmental attribute. Darwin's finches are the classical example: among 13 closely related species, variation in beak morphology is evidence of adaptation to a preferred food in their environment. Given such variation in ecologically relevant phenotypes, ecological genomics seeks to understand the genetic mechanisms that underlie this variation and the adaptive response of species to their environments [1]. For the finches example, it is now known that the expression of bone morphogenetic protein 4 (BMP4) correlates with beak breadth, and that calmodulin expression correlates with beak length [2, 3]. This represents the traditional approach taken to studies relating genetics and ecology: first an ecological adaptive phenotype is identified, then various methodologies are employed to detect causal genetic variation.

This approach, however, can only be applied on a small scale and for relatively well-studied systems. Specifically, it relies on a comprehensive understanding of the organism's ecology and a thorough characterization of its habitat. Such detailed data are often lacking for microbial organisms. With the development of next generation sequencing technologies and environmental genomics, it has now become increasingly common to encounter situations where full genomic information is available for species whose ecology, habitat, and interactions with other species are largely uncharted. Such situations call for a shift in our approach to studying ecological systems and integrating ecology and genomics.

With the advent of functional genomics and systems biology, a new paradigm has risen, termed *Reverse Ecology*. Reverse Ecology aims to infer the ecology of an organism directly from genomic information, with a particular emphasis on microbial ecology. In a recent study, for example, researchers used whole-transcriptome sequencing to identify two ecologically distinct and previously unknown subpopulations of the bread mold *Neurospora crassa* [4]. Specifically, their analysis revealed two genomic islands of high divergence between these subpopulations enriched for genes related to temperature response. Using growth assays, it was confirmed that these subpopulations indeed have significantly different fitness at low temperature, indicating that such divergence is likely a local adaptive response to the different annual temperatures experienced by these subpopulations.

While this population-genomic approach represents an intriguing example of the Reverse Ecology concept, this chapter will discuss a more comprehensive flavor of Reverse Ecology research, one that focuses of the analysis of complex biological systems. In the same way that Systems Biology advocates an emphasis on system-level properties and interactions so does this Reverse Ecology approach argue that much of the information on an organism's ecology is embedded not in the "parts-list" of the system but rather in the way these parts come together and interact. Put differently, system-based Reverse Ecology postulates that as systems become adapted to their environment, their structure, topology, and global properties reflect the environment in which they evolved. Focusing on microorganisms, this research therefore centers on identifying such system-level signatures that can be used to obtain novel insights into the ecology of poorly characterized microbial species. In that sense, Reverse Ecology represents an exciting expansion of Evolutionary Systems Biology: As our understanding of the evolutionary origins of system-level features and the selective pressures acting on biological systems improves, so will our ability to extract valuable information from the structure of these systems and to more reliably predict the ecological context that gave rise to a specific system structure.

2 From Systems to Environments

The key premise of Reverse Ecology is that genomic information can be converted into ecological information. This concept is best demonstrated in the context of metabolism. The metabolic activity of a microorganism is directly linked with the biochemical environment in which it is found through various sensing mechanisms and the import of exogenous compounds, and can in turn impact the composition of this environment via secretion of other compounds. Some of the most prominent Reverse Ecology studies to date therefore focus on metabolism. Such studies focus on identifying links between the organization of an organism's metabolic network and its natural environment, devising algorithms to analyze these networks, and obtaining insights into the organism's ecology. These studies are described below.

2.1 The Seed Set Framework: Predicting Exogenously Acquired Compounds

Perhaps, the most straightforward implementation of the Reverse Ecology paradigm is the seed set framework [5]. This framework aims to characterize the biochemical composition of an organism's habitat based on its genome. Specifically, it combines graph-theory-based algorithms with genome-scale metabolic network models to predict the complete set of compounds an organism takes up from its environment.

To this end, full genome sequence data are used to reconstruct the metabolic network of the species under study. The metabolic network is represented as a directed graph where nodes denote metabolites and directed edges connect substrates to products. The seed set of the network is then defined as the minimal set of compounds in the network that allows the synthesis of all other compounds in the network. The seed set can therefore be conceived as the effective biochemical environment of the organism and serve as a simple proxy for its habitat. Formally, under the simple graph representation described above, the seed set is defined as the minimal set of nodes in the graph such that a directed path exists between a node in this set and every other node in the graph. To determine the seed set, the graph is decomposed into its strongly connected components using Kosaraju's algorithm [6]. It is easy to show that any set that includes a single compound from each strongly connected *source* component (i.e., a component that does not have incoming edges) satisfies the above definition (see [5] for a complete description of the algorithm and its justification). Once these source components have been identified, it is therefore possible to enumerate all possible seed set solutions. In practice, however, a single seed set solution is considered, including all the compounds from all the source components. Each such compound is assigned a confidence score (1/size of source component), denoting the likelihood that it is in fact a seed compound.

To examine this seed set framework, Borenstein et al. reconstructed the metabolic networks of 478 species and determined their seed sets [5]. This represents the first large-scale dataset of predicted environments. Using various experimentally validated data on microorganisms' ecologies and environments, it was shown that these computationally determined seed sets accurately describe the organisms' biochemical surroundings and correlate with various attributes of the organisms' habitats. For example, the predicted seed set of *Buchnera aphidicola* was found to be consistent with its lifestyle as an obligate endosymbiont of aphids, relying on the host for various nutrients that it cannot synthesize. On a larger scale, the presence and absence of key compounds (such as amino acids and major vitamins) across the seed sets of a large array of bacterial species were in clear agreement with biological observations of the synthetic capacities of these species.

Studying the obtained seed sets further revealed several intriguing patterns. It was shown that species inhabiting more variable environments have larger seed sets than do those in relatively stable environments. Using the calculated seed sets to reconstruct an ecology-based phylogeny, it was also demonstrated that ecological profiles convey as much information about the evolutionary history of the various species as do full metabolic profiles. Furthermore, marked similarity was observed between the predicted ecology-based phylogeny and a well-established sequence-based phylogeny, suggesting that ecological adaptation plays a key role in genomic evolution. Finally, reconstructing ancestral metabolic networks and using the seed set algorithm to calculate the ecological profiles of ancestral (extinct) species facilitated a systematic evaluation of the dynamics through which metabolites (seeds and non-seeds) are integrated into the metabolic network. The data supported the retrograde model of network evolution according to which metabolic pathways are extended outward toward the periphery of the network [7]. This seed set framework laid the foundation for multiple studies of microbial ecology and spearheaded further Reverse Ecology research.

2.2 The Network Expansion Framework: Predicting Metabolic Capacity

The seed set framework described above predicts the composition of the biochemical environment in which an organism flourishes, inferring the set of compounds that the organism takes up. Yet, a different aspect of an organism's ecology is its persistence and activity in a given environment, and its potential impact on this environment. To address this challenge, Ebenhöh et al. devised a simple framework that has found much use in various Reverse Ecology studies: the *network expansion* algorithm [8]. In this algorithm, an organism is assumed to have some initial set of metabolites (termed the *expansion seed*; not to be confused with the seed set described above) available in the environment. The algorithm aims to predict the total set of metabolites that this organism can synthesize given this expansion seed. This set of produced metabolites is termed *scope*, and is initially assumed to include only the metabolites available in the expansion seed. The algorithm iteratively examines which reactions, from the set of all the metabolic reactions that the organism can potentially catalyze, have all their substrates already included in the scope. These reactions are assumed to occur and their products are added to the expanding scope. As a heuristic, it is assumed that certain cofactors such as ATP are available before they are reached by expansion. The algorithm terminates when no further reactions can be added. The compounds included in the scope once the algorithm terminates are considered producible [9].

This network expansion framework nicely complements the seed set framework, together offering a comprehensive view of the interaction between an organism and its environment. Using the seed set framework and the network expansion framework, both the organism's natural environment and its synthetic capacity in other environments can be determined. These frameworks provide an essential (though yet elementary) toolbox for addressing fundamental questions concerning both the ecology of specific poorly characterized microbial species and universal strategies in microbial ecology. Several such applications of Reverse Ecology will be discussed below.

2.3 Applications of Reverse Ecology

While the seed set framework explores the preferred biochemical environment of a species, organisms must also evolve to tolerate suboptimal conditions. Such environmental robustness is also linked with genetic robustness: tolerance to genetic perturbations [10]. Auxotrophy is an example: an organism may tolerate a gene deletion while grown in rich medium but not in minimal medium. Reverse Ecology provides a powerful vehicle for analyzing the evolutionary origins and the environmental specificity of metabolic robustness. For example, Freilich et al. used the seed set framework to determine the species-specific habitats of 487 microbial species, and then applied the network expansion framework to simulate the growth of each species in its natural habitat after systematic deletion of each gene [11]. Specifically, the ability of the species to produce a set of essential metabolites was examined. The fraction of enzymes in each network whose deletion did not hamper the production of these essential metabolites was used as a measure of an organism's ability to buffer genetic perturbations. Conversely, environmental robustness was measured as the fraction of these 487 predicted habitats in which a given species is viable. Finally, conditionally essential reactions were identified by systematically deleting each reaction, and simulating growth in each environment. This analysis allows decoupling metabolic robustness into two distinct components: environmental-dependent vs. environmental-independent robustness. It was shown that environmental-dependent robustness is responsible for more than 20% of the nonessential reactions and correlates with the lifestyle of the species. In contrast, environmental-independent robustness was shown to reflect mostly intrinsic metabolic capacities.

Beyond the analysis of specific genomes, the Reverse Ecology paradigm has been instrumental in generating and validating hypotheses about evolutionary dynamics. For example, to explore the effects that the prehistoric introduction of oxygen had on the evolution of life on earth, Raymond and Segrè applied the network expansion framework to the entire set of known enzymatic reactions [12]. Expansion was performed with two protocols—either allowing or prohibiting inclusion of molecules containing oxygen. Oxic networks had as many as 1,000 more reactions than the largest anoxic networks. Furthermore, the average path length between metabolites decreased notably in the oxic networks, suggesting that the introduction of molecular oxygen to the atmosphere, and thus to metabolism, did involve some metabolic rewiring of the anoxic core. More importantly, however, it allowed the evolution of novel pathways. Strikingly, not all reactions in the oxic pathways explicitly utilize molecular oxygen, indicating that creation of novel pathways, not enzymes, was a salient feature of this transition period in the evolution of life.

3 From Systems to Ecosystems

Just as classical ecology is concerned with both the biotic and the abiotic features of a given species' environment, so is Reverse Ecology. The complex web of competitive and syntrophic interactions between the various microbial species and between some of these species and their hosts is bound to leave marked imprints in the metabolic networks of these species. A natural extension of the Reverse Ecology paradigm described above therefore focuses on predicting such interactions, providing a comprehensive framework for characterizing interspecific effects on a large scale.

3.1 The Biosynthetic Support Framework: Predicting Host–Parasite Interactions

Simple examples of species–species interactions are those between bacterial endosymbionts and their eukaryotic hosts. Such host-associated bacteria face significantly reduced environmental variability and often become dependent on their hosts for nutritional intake [13]. Such species therefore become highly specialized, requiring only a small and well-defined set of nutrients they can acquire from the host environment. Indeed, applying the seed set framework to a large collection of microbial species and coupling the obtained seed sets with data describing the lifestyle of each species, Borenstein and Feldman found that, in general, parasitic bacteria have smaller seed sets than do free-living bacteria [14].

Such dependencies between hosts and parasites are clearly reflected in the organization of their metabolic networks, and can be inferred on a large scale by analyzing these networks and identifying specific signatures of species interactions. To this end, Borenstein and Feldman further expanded the seed set concept and introduced a novel Reverse Ecology measure, aiming to capture the extent to which the nutritional requirement of a symbiont are met by the host [14]. Specifically, given the metabolic networks of a putative parasite and a potential host, the *biosynthetic support score* (BSS) denotes the fraction of seed compounds of the parasite that can be synthesized by the metabolic network of the host. To examine how well this measure reflects host–parasite interactions, three representative eukaryotic species were selected as model hosts: human, fruit fly, and *Arabidopsis*. Considering a large array of microbial species, it was shown that parasitic bacteria have significantly higher biosynthetic support scores than free living bacteria when interacting with these hosts. Furthermore, biosynthetic support scores are higher when the parasite is known to infect phylogenetically similar hosts, facilitating a successful prediction of host–parasite pairs. Reconstructing the metabolic networks of ancestral species among divisions containing many parasites further revealed an evolutionary trend of increase in biosynthetic support score from ancestor to descendant. This supports a model of gradual ecological adaptation of parasites to their hosts, and of increased dependence on specific hosts for subsistence.

3.2 Cohabitation and Metabolic Overlap: Predicting Interspecific Competition

Competition between species is a constant pressure to which a species must adapt. Such competitive interactions play a key role in the assembly of microbial communities and in determining the resilience of these communities to various perturbations [15]. Unfortunately, however, our understanding of competitive interactions and their impact on microbial growth is lacking.

To address this challenge, Freilich et al. used a combination of the Reverse Ecology framework described above to examine ecological strategies for coping with competition across the microbial tree of life [16]. First, the biochemical environments of 528 species were calculated using the seed set framework. Next, the expected biosynthetic capacity of each species was simulated in every such environment using the network expansion framework. Species were considered viable in a given environment if a set of essential metabolites were producible and found in the scope of the expanded network. From these data, two measures were calculated for each species. First, the environmental scope index (ESI) denotes the fraction of environments in which a species is viable, approximating its environmental flexibility; species with high ESI scores are generalists that can survive in a wide span of environments, whereas species with low ESI scores are specialists. Second, the cohabitation score (CHS) denotes the number of other species which are also viable in each environment in which the species under study is viable, approximating the level of cohabitation the species encounters in each environment. Specifically, the maximal CHS represents the maximal level of competition a species encounters. Comparing these measures to data concerning the doubling-time of each species, it was shown that both ESI and maximal CHS positively correlate with growth rate. These findings suggest that microbial species largely adopt one of two ecological strategies: Generalists that cope with intense competition and grow rapidly, or slow growing specialists that occupy ecological niches with relatively little cohabitation.

The degree to which competitive interactions may impact a given species, however, is not necessarily determined simply by the number of species cohabiting the environment, but rather by the capacity of this species to successfully grow when specific nutrients are competed away by an interacting partner. An additional Reverse Ecology-based metric, the *effective metabolic overlap* (EMO), combines the seed set and network expansion frameworks to predict exactly that [17]. Specifically, to calculate the impact that a certain interacting species may have on another species, the seed sets of both species are determined. Any compound that appears in both sets is removed, and the network expansion framework is used to calculate the set of producible compounds given this smaller expansion seed. The fraction of essential compounds in the obtained scope represents the ability of an organism to tolerate competition by its partner. EMO is then defined as 1 minus this fraction, such that higher EMO values indicate stronger competition. Assaying the EMO within clusters of ecologically co-occurring species, it was found that in general clusters with higher EMO (i.e., fiercer competition) exhibit lower mean growth rates and vice versa, again highlighting a dichotomy in the ecological strategies that microbial species may adopt.

3.3 Stoichiometric Models of Species Interaction

The modeling frameworks and topology-based analyses discussed above are a powerful toolbox for analyzing potential metabolic dependencies between species. In some cases, however, one may wish to predict not only potential interactions but also specific metabolic dynamics in a community of microorganisms and the exact set of metabolites being exchanged actively between the various species. The prediction of specific metabolic fluxes is mostly beyond the scope of topology-based metabolic models and requires a more involved modeling framework such as constraint-based modeling (CBM) [18, 19]. Such models, however, are usually limited in scale and require detailed and manually curated data [20]. Yet, recently, several preliminary studies have demonstrated the use of CBM to construct and study simple multispecies systems and to quantify the extent of metabolic exchange between species.

Stolyar et al. for example, introduced such a multispecies model, comprising two species involved in the anaerobic oxidation of methane, *Desulfovibrio vulgaris* and *Methanococcus maripaludis* [21]. In this model each species was represented as a separate compartment encompassing the species' native metabolic activity. Due to the inherent difficulty of constructing full constraint-based models, only the core metabolism of each species was included. The model correctly identified metabolic transfer, and predicted that hydrogen, not formate, transfer is essential to syntrophic growth.

Continuing the study of metabolite transfer, Wintermute and Silver analyzed the ability of *Escherichia coli* strains with complementary gene deletions to support one-another's growth through syntrophy [22]. Metabolic synergy in such pairs was found to be extremely prevalent. Using stoichiometric models, the benefit and cost of each exchanged metabolite were calculated and a predicted efficiency of cooperation was defined. Comparing this efficiency measure with coculture growth assays demonstrated a strong fitness advantage for efficient cooperators.

Finally, to ascertain the role that the environment plays in determining cooperative interactions, Klitgord and Segrè used genome-scale stoichiometric models to examine all pair-wise combinations of seven species across a large array of media [23]. Of specific interest were media that supported the growth of the two species together but in which one or both species could not grow alone. Surprisingly, media that induce commensalism or mutualism could be found for all species pairs. Furthermore, in general, more media were found that sustain cooperative growth than media that can sustain both species independently, suggesting that nutrient-poor growth environments may be dominated by cooperative species interactions.

4 From Environments to Systems

While the seed set and network expansion frameworks take a somewhat mechanistic Reverse Ecology approach, one can alternatively focus on the phenomenological study of the link between network structure and environmental properties. In this case, rather than inferring a detailed, metabolite-level description of the biochemical environment and of its composition, we wish to study how large-scale, coarse-grained features of the environment impact *global* characteristics of the evolved system.

One such global property that can be observed in many biological systems is a high degree of modularity [24]. Modularity is obviously expected in designed systems, but its prevalence and origins in biological systems are not clear. Parter et al. hypothesized that increased modularity is linked with the environment in which a system evolved, and compared the level of modularity of genome-scale metabolic networks across 117 bacterial species to the environments these species inhabit [25]. They found a strong association between the level of modularity of these networks and the extent of variability in the environment. This was greatly expanded on by the work of Kreimer et al. who analyzed the evolutionary history of modularity and its ecological context across 325 species, spanning the bacterial tree of life [26]. Host-associated species, whose habitat exhibits relatively low environmental variability, were shown to have less modular metabolic networks compared to, for example, free living bacteria. Interestingly, pathogens whose lifecycles include association with multiple hosts have higher modularity than those associated with a single host, further strengthening the connection between environmental variability and network structure. This analysis also revealed an intriguing decrease in modularity from ancestor to offspring species across the bacterial kingdom—a likely outcome of niche specialization.

Interestingly, Parter et al. came to their hypothesized association between modularity and environmental variability as the result of work performed using evolutionary simulations. Such simulations offer another suite of tools useful in Reverse Ecology: instead of examining genomic-derived models of various species and associating global system-level features with the species' environments, researchers use evolutionary simulations to examine directly how evolution in a certain environmental regime affects the evolved system. While these studies generally employ simplified models that do not accurately represent specific biological processes, they are useful in generating and testing hypotheses pertaining to the relationship between environments and systems. Specifically, in the context of modularity, Kashtan and Alon used evolutionary simulations to demonstrate how evolution in a changing environment can lead to the spontaneous emergence of modularity [27]. To this end, they simulated a population of genomes, each encoding a Boolean network. Each network received inputs from its environment, and its fitness was determined by how closely the network calculated some Boolean function (e.g. (X XOR Y) AND (Z XOR W)). In each generation, the most fit networks were replicated with some mutational frequency. Eventually, the population would evolve a perfect, albeit non-modular solution. The authors next introduced environmental variability: every few generations, the function the network was required to produce would toggle between two states. It was shown that as long as these two target functions could be described as different combinations of the same set of simpler sub-functions (e.g., the expression above and (X XOR Y) OR (Z XOR W)), the evolved networks exhibited a high level of modularity. This lent first support to the hypothesis that biological networks are modular because throughout evolutionary history, organisms have faced such modularly varying environments. Furthermore, it was found that the networks evolved optimal solutions faster in such modularly varying environments, indicating that finding modular solutions to biological functions may facilitate rapid adaptation in changing environments [28].

Additional studies similarly used evolutionary simulations to investigate other global network properties and to identify environmental features that may underlie these properties. Soyer and Pfeiffer, for example, used this approach to associate environmental variability with network robustness [29]. In their simulations, metabolic networks were evolved to convert available metabolites into biomass in various biochemical environments. Environments could either be stable or fluctuate between rich and minimal media. They found that networks evolved in fluctuating environments were more robust to single gene deletions. By analyzing network structure, it was further determined that networks evolved in such varying environments contained more redundant paths, as well as more multifunctioning enzymes. Perhaps most telling, however, is that networks evolved in the fluctuating regime lose their robustness when allowed to evolve in the stable regime for a number of generations. A similar approach, using evolutionary simulations to explore links between selective pressures and an evolving system, was also used to study the transition from a generalist to a specialist ecological strategy and the loss of function that may accompany this process [30].

Such detected associations provide a simple yet powerful Reverse Ecology tool and allow the prediction of various general attributes of the habitat of microbial species. For example, by comparing the modularity of a homology-based metabolic network of a newly sequenced species to the modularity of other, well-studied microbes, one can determine the likely level of variability in the environment of this species, and potentially its lifestyle. Revealing additional associations between global system properties and environmental attributes will further expand our ability to similarly predict various aspects of an organism's habitat.

5 Future Directions and Potential Applications

One of the greatest emerging challenges in life sciences is the analysis bottleneck. Put simply, our ability to generate data greatly outpaces our ability to analyze it and draw information from it. The clearest example is probably the increasing rate at which we are collecting whole genome sequence information from species for which we have little to no ecological understanding [31]. Moreover, even when broad, cross-species ecological data exist, it is still limited in scope or in resolution. Microorganisms, for example, are typically categorized loosely as either “free-living” or “host-associated,” belying the intricacies of their environments and interactions; or take, for example, the work of Freilich et al.: while 486 species were analyzed in their study of ecological strategies, growth rate information and data on environmental complexity were available for only about one-fifth of these species [16].

The Reverse Ecology paradigm offers a unique and promising solution to this problem, generating large-scale ecological insights from high-throughput genomic data. Using the tools outlined in this chapter, we can predict an organism’s biochemical environment, ecological strategy, interactions with other species, and adaptive niche. Accordingly, as genomic coverage of the tree of life improves, so too will our ability to draw ecological and evolutionary inferences.

In this last section, we discuss some of the most promising future directions of Reverse Ecology research, and potential applications. We highlight the role that Reverse Ecology may play in studying microbial communities and specifically the human microbiome, and describe possible extensions of the Reverse Ecology framework.

5.1 Reverse Ecology of Microbial Communities and the Human Microbiome

Interactions between microbial species are often investigated in pair-wise associations, whether experimentally or computationally [21, 23, 32, 33]. Some of the Reverse Ecology studies described above have taken a similar route. Interspecies interactions, however, are often far more complex in nature, as microbial ecosystems regularly comprise hundreds or even thousands of species [34, 35]. The various species that make up each community form tight relationships and establish strong metabolic dependencies [36, 37], which are instrumental to the stability and activity of the community and which have a crucial effect on the interplay of the community as a whole with its environment [38]. These complex and highly diversified communities play a key role in ecosystem dynamics, agriculture, and environmental stewardship. They are essential components of every system of which they are a part, be they environmental communities that recycle organic material into the biosphere or endosymbionts that perform essential functions for their hosts [39, 40]. Arguably the most important set of communities, at least to us humans, are those inhabiting various anatomical sites in the human body. These microbes, collectively known as the human microbiome, are an integral part of many essential processes in the human body and have a tremendous impact on our health [41]. As such, understanding these communities, along with their ecologies and interactions, holds great promise for biomedical applications.

Given the vast importance these communities hold, their study presents a critical application of Reverse Ecology. Moreover, as Reverse Ecology aims to infer ecological insights from genomic information, it presents a tool uniquely suited to studying such largely uncharted communities for which genomic data are now readily available. The challenge lies in developing methods that scale with the complexity of communities. As a first step, it is reasonable to apply Reverse Ecology metrics to all species pairs found in a community. Future work, however, will have to address interactions dependent on the presence of other partners. Environmental cues and attributes can be added to the model, ultimately providing system-level predictive *in silico* models of community metabolism. These will prove invaluable in addressing questions of community perturbation response, robustness, and assembly.

Often, however, due to the strong metabolic dependencies between community members, isolating individual species for experimentation is not feasible [42]. Researchers instead have turned to culture-free methods of studying microbial communities, using, for example, shotgun metagenomics to extract and sequence genomic material directly from the environment. In such studies, the main focus is often the characterization of the community as a whole rather than of any individual member [43], treating the entire community as a single "supra-organism" [44, 45]. Reverse Ecology research of microbial communities can accordingly adopt a similar approach, reconstructing metagenomics-based community-wide metabolic models that totally ignore metabolic compartmentalization imposed by cell boundaries, and study the interface between these community-level supra-organisms and their environments or hosts (see, for example, [46]).

These two approaches, investigating species interactions within the community or using a supra-organism framework, may be thought of as complimentary bottom-up and top-down methods for studying community-level ecology. Taken together, these approaches may reveal fundamental mechanisms responsible for the assembly, function, and dynamics of microbial communities.

5.2 Engineering Species and Communities

The potential applications for the Reverse Ecology framework outreach the work described here. In recent years, bioengineering has shifted from designing simple inaugural devices to system-level rewiring of genetic circuits [47]. To that end, Reverse Ecology offers a novel suite of tools for rational design of such biological interfaces. Janga and Babu, for example, proposed the use of the seed set framework to design drugs which target pathogens without adversely affecting the host [48]. Reverse Ecology can also direct the synthesis of media which promote desirable biosynthetic reactions for the manufacturing of specific byproducts or optimization of bioreactors. Röling et al. have similarly proposed the use of Reverse Ecology to develop media for the targeted isolation and culture of microbes [49]. As interest grows in designing and selecting microbial consortia for medical or technological purposes [50], Reverse Ecology offers uniquely powerful solutions. Hansen et al. for example, have proposed to use Reverse Ecology-based algorithms to determine persistent community transplants in gnotobiotic mice [51]. Ultimately, predictive modeling of community metabolism and rational design of artificial microbiomes may provide an exciting framework for guiding clinical interventions in the human microbiome and the design of transplantable microbiomes with desired metabolic activities [52].

5.3 Beyond Metabolism

The famous physician Arturo Rosenblueth stated once that “the best material model for a cat is another, or preferably the same cat” [53]. Clearly, the metabolic models used throughout the various studies described above represent a massive simplification of the actual metabolic processes and ignore many crucial details concerning microbial metabolism. Better and more accurate models are necessary to advance Reverse Ecology research and to offer researchers a reliable toolbox for predicting an organism’s habitat. Constraint-based models [19], for example, are an intriguing option, and several constraint-based multispecies models have been introduced (as described above). Such models, however, have limited availability and pose many challenges that may hinder large-scale ecological studies. Automated pipelines, such as the Model SEED [54], have the potential to allow researchers to create constraint-based models on a much larger scale, and may ultimately render such models a feasible option for community modeling.

Novel data types are frequently becoming available, thanks to technological advances in instrumentation, many of which pose opportunities to take Reverse Ecology beyond metabolism. While the topology of a metabolic network can determine the breadth of niches available to an organism, reverse ecological analysis of genetic or signaling networks can potentially determine under what conditions an organism occupies which niches. Mahowald et al. for example, performed meta-transcriptomic and meta-proteomic analyses on a simplified gut community of two species [55]. In response to introduction of the invasive partner, each species modified its ecological strategy to exploit resources unavailable to the other. While analysis of these species’ seed sets can indicate that such nonoverlapping strategies are possible, currently available methods are not capable of determining the strategies organisms will take given a set of environmental cues.

Expanding the Reverse Ecology framework to include additional high-throughput data sources, and most notably, metabolomics data, can allow researchers to validate various Reverse Ecology predictions and describe a wider set of ecological attributes. Furthermore, thorough annotation of high-resolution metadata can also be incorporated into new models capable of describing more detailed definitions of adaptive niches.

5.4 The Rise of High-Throughput Ecology

Genomic and metagenomic data coupled with integrative analysis of ecosystems from the molecular to the macroscopic levels are now making ecology into a true high-throughput field. Just as high-throughput methodologies revolutionized molecular biology, allowing researchers to model cellular processes from the genetic to the tissue level, researchers will be able, for the first time, to consider ecosystem dynamics across spatial and temporal scales hitherto unseen. Systems analyses, predictive modeling, and Reverse Ecology tools provide an exciting opportunity to make sense of all these data and study microbial ecosystems on a large scale. As modeling frameworks improve, such tools can in fact go beyond naturally occurring ecologies, and study the microbial *ecosystome*, mapping the contours of the space of possible ecosystems. Studying the ecology of the possible would help put ecological systems in context and could offer a neutral model for investigating ecosystem assembly and dynamics.

Ecology is clearly poised for a major transition in coming years. Genomics' transformation of biological sciences into an information system-level science is apt to be mirrored in the ecological sciences. The Reverse Ecology approach is bound to play a key role in this transformation, allowing the translation of high-throughput genomic data into broad system-level ecological understanding.

References

1. Ungerer MC, Johnson LC, Herman MA (2008) Ecological genomics: understanding gene and genome function in the natural environment. Heredity 100(2):178–183. doi:10.1038/sj.hdy.6800992
2. Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ (2004) Bmp4 and morphological variation of beaks in Darwin's finches. Science 305:1462–1465. doi:10.1126/science.1098095
3. Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature 442: 563–567. doi:10.1038/nature04843
4. Ellison CE et al (2011) Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Proc Natl Acad Sci USA 108:2831–2836. doi:10.1073/pnas.1014971108
5. Borenstein E, Kupiec M, Feldman MW, Ruppin E (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proc Natl Acad Sci USA 105:14482–14487. doi:10.1073/pnas.0806162105
6. Aho A, Hopcroft J, Ullman J (1974) The design and analysis of computer algorithms. Addison-Wesley, Reading, MA
7. Horowitz NH (1945) On the evolution of biochemical syntheses. Proc Natl Acad Sci USA 31:153
8. Ebenhöh O, Handorf T, Heinrich R (2004) Structural analysis of expanding metabolic networks. Genome Inform 15:35–45; International Conference on Genome Informatics
9. Kruse K, Ebenhöh O (2008) Comparing flux balance analysis to network expansion: producibility, sustainability and the scope of compounds. Genome Inform 20:91–101; International Conference on Genome Informatics

10. De Visser J et al (2003) Perspective: evolution and detection of genetic robustness. Evol, Int J Org Evol 57:1959–1972
11. Freilich S et al (2010) Decoupling environment-dependent and independent genetic robustness across bacterial species. PLoS Comp Biol 6:e1000690. doi:10.1371/journal.pcbi.1000690
12. Raymond J, Segrè D (2006) The effect of oxygen on biochemical networks and the evolution of complex life. Science 311:1764–1767. doi:10.1126/science.1118439
13. Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. Cell 126(3):453–465. doi:10.1016/j.cell.2006.07.014
14. Borenstein E, Feldman MW (2009) Topological signatures of species interactions in metabolic networks. J Comput Biol 16:191–200. doi:10.1089/cmb.2008.06TT
15. Trosvik P et al (2010) Web of ecological interactions in an experimental gut microbiota. Environ Microbiol 12(10):2677–2687. doi:10.1111/j.1462-2920.2010.02236.x
16. Freilich S et al (2009) Metabolic-network-driven analysis of bacterial ecological strategies. Genome Biol 10:R61. doi:10.1186/gb-2009-10-6-r61
17. Freilich S et al (2010) The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Res 38:3857–3868. doi:10.1093/nar/gkq118
18. Edwards JS, Palsson BO (2000) Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. BMC Bioinformatics 1:1
19. Reed J, Palsson BØ (2003) Thirteen years of building constraint-based in silico models of Escherichia coli. J Bacteriol 185:2692–2699. doi:10.1128/JB.185.9.2692
20. Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protocol 5:93–121. doi:10.1038/nprot.2009.203
21. Stolyar S et al (2007) Metabolic modeling of a mutualistic microbial community. Mol Syst Biol 3:92. doi:10.1038/msb4100131
22. Wintermute EH, Silver PA (2010) Emergent cooperation in microbial metabolism. Mol Syst Biol 6:407. doi:10.1038/msb.2010.66
23. Klitgord N, Segrè D (2010) Environments that induce synthetic microbial ecosystems. PLoS Comp Biol 6:e1001002. doi:10.1371/journal.pcbi.1001002
24. Hartwell LH et al (1999) From molecular to modular cell biology. Nature 402:6761. doi:10.1038/35011540
25. Parter M, Kashtan N, Alon U (2007) Environmental variability and modularity of bacterial metabolic networks. BMC Evol Biol 7:169. doi:10.1186/1471-2148-7-169
26. Kreimer A, Borenstein E, Gophna U, Ruppin E (2008) The evolution of modularity in bacterial metabolic networks. Proc Natl Acad Sci USA 105:6976–6981. doi:10.1073/pnas.0712149105
27. Kashtan N, Alon U (2005) Spontaneous evolution of modularity and network motifs. Proc Natl Acad Sci USA 102:13773–13778. doi:10.1073/pnas.0503610102
28. Kashtan N et al (2009) An analytically solvable model for rapid evolution of modular structure. PLoS Comp Biol 5:e1000355 doi:10.1371/journal.pcbi.1000355
29. Soyer OS, Pfeiffer T (2010) Evolution under fluctuating environments explains observed robustness in metabolic networks. PLoS Comp Biol 6:8. doi:10.1371/journal.pcbi.1000907
30. Ostrowski E, Ofria C, Lenski RE (2007) Ecological specialization and adaptive decay in digital organisms. Am Nat 169:E1–E20
31. Kyrpides NC (2009) Fifteen years of microbial genomics: meeting the challenges and fulfilling the dream. Nat Biotechnol 27:627–632
32. Chalmers NI, Palmer RJ, Cisar JO, Kolenbrander PE (2008) Characterization of a Streptococcus sp.-Veillonella sp. community micromanipulated from dental plaque. J Bacteriol 190:8145–8154. doi:10.1128/JB.00983-08
33. Shou W, Ram S, Vilar JMG (2007) Synthetic cooperation in engineered yeast populations. Proc Natl Acad Sci USA 104:1877–1882. doi:10.1073/pnas.0610575104
34. Torsvik V, Øvreås L, Thingstad TF (2002) Prokaryotic diversity-magnitude, dynamics, and controlling factors. Science 296:1064–1066. doi:10.1126/science.1071698
35. Schloss PD, Handelsman J (2005) Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biol 6:229

36. Schink B, Stams AJM (2006) Syntrophism among prokaryotes. In: Dworkin M et al (ed) The prokaryotes: an evolving electronic resource for the microbiological community, vol 2. Springer, New York
37. Little AEF, Robinson CJ, Peterson SB, Raffa KF, Handelsman J (2008) Rules of engagement: interspecies interactions that regulate microbial communities. Annu Rev Microbiol 62: 375–401
38. McInerney MJ, Sieber JR, Gunsalus RP (2009) Syntrophy in anaerobic global carbon cycles. Curr Opin Biotechnol 20:623–632
39. Douglas A (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. Ann Rev Entomol 43:17–37
40. Lodwig E, Poole P (2003) Metabolism of Rhizobium bacteroids. Crit Rev Plant Sci 22:37–78
41. Turnbaugh PJ et al (2007) The human microbiome project. Nature 449:804–810. doi:10.1038/nature06244
42. Vartoukian SR, Palmer RM, Wade WG (2010) Strategies for culture of 'unculturable' bacteria. FEMS Microbiol Lett 309(1):1–7. doi:10.1111/j.1574-6968.2010.02000.x
43. Tyson GW et al (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43
44. Gordon JI, Klaenhammer TR (2011) A rendezvous with our microbes. Proc Natl Acad Sci USA 108 Suppl 1:4513–4515. doi:10.1073/pnas.1101958108
45. Lederberg J (2000) Infectious history. Science 288:287–293. doi:10.1126/science.288.5464.287
46. Greenblum S, Turnbaugh PJ, Borenstein E (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci USA 109:594–599. doi:10.1073/pnas.1116053109
47. Khalil AS, Collins JJ (2010) Synthetic biology: applications come of age. Nat Rev Genet 11(5):367–379. doi:10.1038/nrg2775
48. Janga SC, Babu MM (2008) Network-based approaches for linking metabolism with environment. Genome Biol 9:239. doi:10.1186/gb-2008-9-11-239
49. Röling WFM, Ferrer M, Golyshin PN (2010) Systems approaches to microbial communities and their functioning. Curr Opin Biotechnol 21:532–538. doi:10.1016/j.copbio.2010.06.007
50. Brenner K, You L, Arnold FH (2008) Engineering microbial consortia: a new frontier in synthetic biology. Trends Biotechnol 26(9):483–489. doi:10.1016/j.tibtech.2008.05.004
51. Hansen EE et al (2011) Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. Proc Natl Acad Sci USA 108 Suppl 1: 4599–4606. doi:10.1073/pnas.1000071108
52. Khoruts A et al (2010). Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. J Clin Gastroenterol 44(5):354–360
53. Rosenblueth A, Wiener N (1945) The role of models in science. Phil Sci 12:316–321
54. Henry CS et al (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28:969–974
55. Mahowald MA et al (2009) Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc Natl Acad Sci USA 106:5859–5864. doi:10.1073/pnas.0901529106