LEVERAGING INNOVATIONS FROM EVOLUTION

Synthesis Report - September 2023

SPONSORED BY
The National Science Foundation, under grant
MCB-2326865

Table of Contents

Executive Summary

Call to Action

In the 3.7 billion years since the origin of life on Earth, organisms have evolved complex innovative traits in response to myriad environmental challenges. These biological innovations (*e.g*., the ability to survive in extreme conditions, such as subzero temperatures or arid environments; the ability to produce or resist toxins; the ability to 'see' in the dark via echolocation; *etc*.) represent a treasure-trove of bio-solutions that, if properly understood, have the potential to provide unprecedented advances for the bioeconomy (*e.g.*, improve industrial processes, develop new pharmaceutical and agricultural products, produce new energy, and engineer resilient systems for a changing planet).

One inherent challenge when studying innovative traits is the fact that the biological mechanisms responsible for these traits are frequently unknown. Despite exponential growth in genomic resources, the functions of most genes in most genomes remain uncharacterized. There is growing appreciation for non-model organismal research as a means to integrate comparative and evolutionary perspectives into mechanistic studies of gene function. And yet, the simple terminology of 'gene function' belies the extremely complex nature of genotype by environmental interactions that result in dynamic organismal traits. Paradigm changing advances in this area will necessarily cross multiple biological scales (from molecules to ecosystems), and as such require that scientists bridge the cultural and intellectual gaps that currently divide biological subdisciplines.

Convergent evolution, the repeated evolution of similar traits often in response to similar selection pressures, provides an ideal framework to explore the origins, mechanisms, and macroevolution of adaptive traits by acting as a proxy for repeated experiments in evolution. Studying how complex adaptive traits repeatedly evolve has the potential to identify both generalizable and unique solutions to life's common problems. Moreover, studying convergent innovations from a hierarchical perspective at multiple biological (e.g., organismal, organ, cellular, subcellular, molecular) and phylogenetic levels will enable a more detailed and robust understanding of the various evolutionary pathways that have converged on similar solutions despite unique evolutionary contexts and constraints.

Research in the areas of evolutionary innovation and convergent evolution is timely thanks to the explosive growth of high-quality reference genomes from across the tree of life as well as new methodological advances (e.g., single cell sequencing, spatial transcriptomics, ancestral state reconstruction, protein structure modeling, and genome editing). In the face of unprecedented environmental challenges facing human society, now is an opportune time to stimulate basic research leveraging the diversity of life to understand how other organisms have adapted and solved repeated challenges, particularly in the context of biotechnology and the bioeconomy. There is a critical need to develop sustainable partnerships between those studying fundamental questions in biology with those working to harness biological discovery for applied benefit (*e.g.,* engineers, chemists, agriculture and industry experts).

Scoping Session Objectives

The overarching goal of the Leveraging Innovation from Evolution (LIFE) initiative is to bring together diverse scientists to brainstorm how we might take ambitious, transdisciplinary leaps within a convergent evolutionary framework to better understand solutions to life's common problems and effectively engineer new and sustainable technologies. To accomplish this goal, the LIFE leadership team (see LIFE Participants) partnered with Knowinnovation, a consultancy organization focused on accelerating multidisciplinary innovation, to host two Scoping Sessions with the following specific objectives:

- bring together experts working at the forefront of scientific domains necessary for LIFE, including systematics, evolutionary development, molecular biology, biochemistry, biotechnology, and genomics to bridge the cultural and intellectual gaps between their disciplines;
- engage the NSF BIO and BIO-adjacent communities to explore specific research challenges and opportunities pertaining to evolutionary innovation and convergent evolution, including generalizable and unique solutions to life's common challenges;
- articulate needs, strategies, and recommendations to enable transdisciplinary research into convergently-evolved innovations, broadly defined, that incorporate research, infrastructure, and educational priorities; and
- jumpstart ideas, collect input from a large and diverse community, and explore potential areas of research in leveraging evolutionary innovation to support the bioeconomy.

Branches of LIFE

A virtual town hall was held on May 15th, 2023 to raise awareness about the upcoming scoping sessions. During the town hall, participants were assigned to breakout rooms and asked to discuss the following questions: What are you most excited about in your field that is poised to accelerate discoveries in convergent evolution? What are the obstacles to overcome to leverage innovations from evolution? Based in part on participant discussion, the LIFE leadership team identified six specific goal areas, referred to as 'Branches of LIFE'.

- New frontiers in evolutionary convergence (*e.g.*, working across scales, understanding constraints, *etc.*)
- Overcoming challenges associated with non-model organisms
- Tool and technology development
- Collaboration and communication among disciplines

- Leveraging life's diversity for the bioeconomy (*i.e.*, translational, applied, and conservation objectives)
- Fostering the next generation of scientists (*i.e.*, training and broadening participation)

Community Scoping Sessions

The two Scoping Sessions took place in the second half of 2023. An in-person Scoping Session was held August 14th-16th at the Alexander Hotel in downtown Indianapolis, Indiana, and an online Scoping Session was held the week of September 11th hosted on Knowinnovation's in-house virtual meeting and collaboration platform KIStorm.

Participants

In total, 90 participants attended the LIFE Scoping Sessions. A list of the LIFE Participants is provided at the end of this document. Hosting a diverse group of participants was a critical objective of the LIFE Scoping Sessions. Participants represented diverse disciplines and study systems, geographic regions, institution types, and career stages (Figure 1).

Scoping Session Structure and Output

Both LIFE Scoping Sessions took place over the course of three days and were conducted as follows.

Day 1**:** Participants began the Scoping Session getting to know one another and the expertise in the room through a series of orientation activities. Participants then began a Knowinnovation-led clarification process of question finding focusing on the six Branches of LIFE. In the afternoon, participants turned their attention from question finding to proposing solutions in the form of WIBGI (Wouldn't It Be Great If…) ideas. By the end of the first day, participants had written dozens of WIBGIs, each articulating specific approaches and recommendations to address the Branches of LIFE goal areas.

Day 2**:** Participants opened the second day by voting on the most exciting and important WIBGIs identified the previous day. Participants were assigned to working groups such that each group had multiple research expertise areas represented. Groups stewarded the top ranked WIBGIs by further exploring and developing each WIBGI for all participants to better understand what they entail. Lastly, participants formed writing groups of 2-8 people and selected their favorite WIBGI(s) to form the basis of their final recommendations. Writing groups had the remainder of Day 2 to write and gather feedback on their ideas.

Day 3**:** In the morning, participants worked with their writing group to finalize their presentations. Teams presented their final recommendations in the form of 10-minute

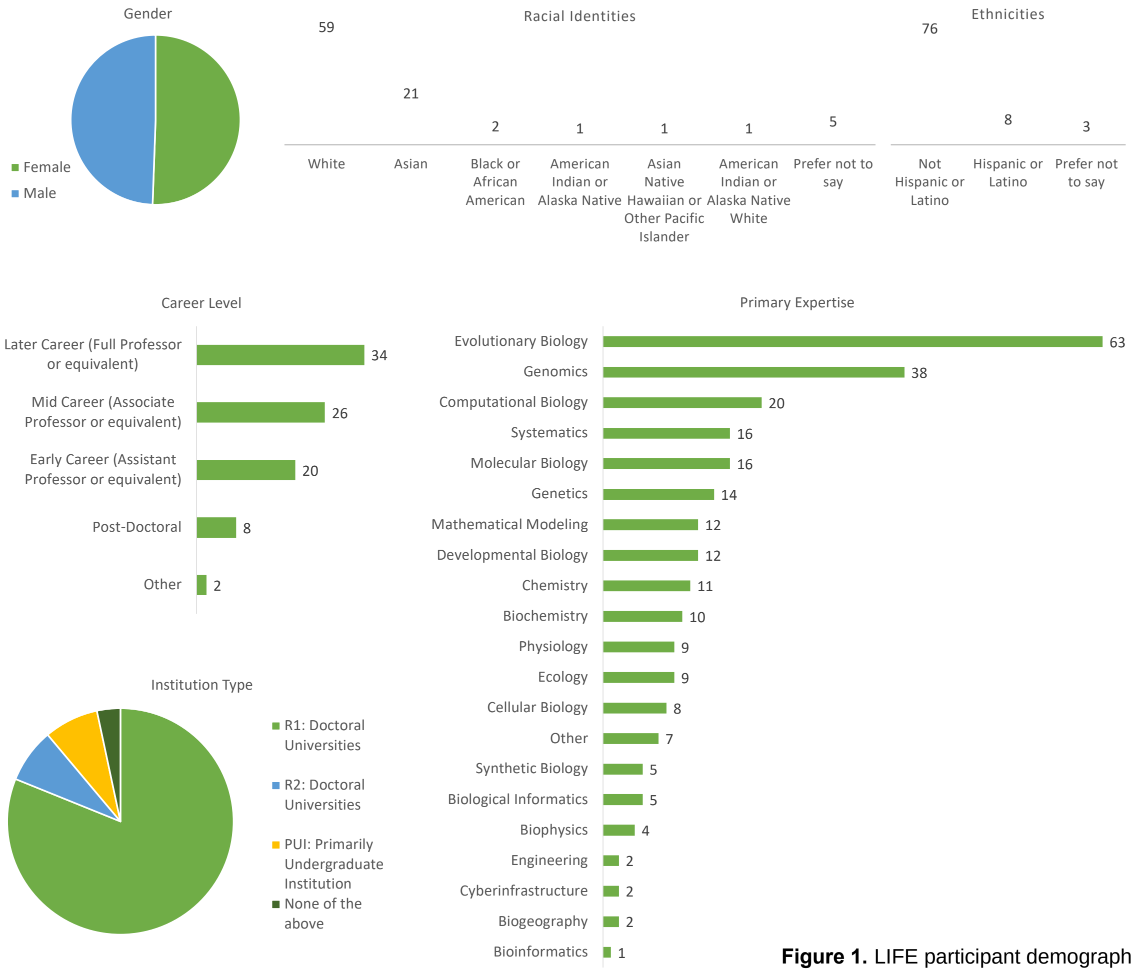

Figure 1. LIFE participant demographics.

presentations to the group. Following the presentations, writing groups continued working on their final Scoping Session Recommendations, which are provided below.

Synthesis Day: Following the close of both Scoping Sessions, a subset of the original participants joined the LIFE leadership team for a Synthesis Day virtual meeting, which was held on September 25th, 2023. During the Synthesis Day, Knowinnovation facilitators led the participants in integrating the feedback and recommendations provided by both Scoping Sessions. Several Cross-Cutting Themes were identified, which are discussed below.

The presentation slides, final recommendations, and this document are all available at the LIFE website https://purdue.ag/innovate-life.

Summary of Recommendations

There were 18 recommendations that emerged from the LIFE Scoping Sessions. These spanned both specific and broad research priorities to address fundamental biological questions and scientific gaps that could inform pressing societal/economic issues. Other recommendations identified new and emerging technology and/or plans for inclusive and integrative training and education. Lastly, many recommendations offered ideas for novel institutional and funding structures that could be utilized to address the priorities identified by LIFE participants. Listed below is an overview of the recommendations, and more details are given in the Scoping Session Recommendations section of this report.

1. PhunCODE: Phenotypes & functions to leverage convergence, diversity, & evolution
   Create phylogenetically informed predictive models that allow inference of intermediate molecular phenotypes across diverse species to improve our ability to traverse the genotype-phenotype map. Specific recommendations include development of the PhyloMatrix to impute predicted phenotypes across the tree of life and a call to incorporate Predict, Test, Learn cycles (adapted from the Design, Build, Test, Learn paradigm in engineering) into research proposals. These recommendations would result in a framework to predict links between genotypes and phenotypes (broadly defined, including organismal and molecular traits) across the tree of life.

2. Learning from phages: Predicting and deciphering host-phage specificity
   We have access to an unprecedented amount of genetic data for phages, but the genetic mechanisms that determine host-virus interaction and specificity are still mostly unknown. Harnessing the incredible amount of data that we can now collect and the experimental tractability of phage and bacterial systems, now is the ideal time to leverage machine learning algorithms to predict which phage can infect which bacterial hosts on a grand scale. These predictions can be used for phage therapy and ecosystem engineering applications. Moreover, feature selection by the algorithm can provide insights into fundamental microbiological processes by revealing the genes and pathways that enable interactions between phages and bacteria. Learning from phage systems would inform our understanding of host-pathogen evolutionary dynamics more generally, including the repeatability of evolution in different conditions/landscapes.

3. LIFE: Evo-inspired engineering of radical phenotypes & emergent traits
   Evolution in nature happens slowly, often resulting in incremental phenotypic changes. We need to create an experimental paradigm that enables saltatory jumps of phenotypes that will enable the rapid origin of new phenotypes and emergent traits. By artificially bringing together traits that have never been expressed together before (by natural processes), we can change the starting point of evolution and widen the range of possible phenotypes that can be selected in novel environments. The ability to rapidly create radical phenotypes will open possibilities for engineering real-time biological solutions in our rapidly changing world.

4. Leveraging diversity in extreme environments
   Leverage convergence across temporal, geographic, and biological scales to uncover mechanisms (e.g., genetic, molecular, cellular, tissue) that allow organisms

to survive and thrive in extreme environments. We propose focusing on convergence to explore solutions to a common challenge (e.g., heat) rather than focusing on a single organism that does something the "best." By establishing NEON-like sites in multiple locations with the same extreme environments, this recommendation will allow long-term collaborative monitoring of species dynamics and adaptive evolution in extreme environments that are formed due to climate change or anthropogenic activities.

5. Convergence in action
   A top-down funding Program Officer to Primary Investigator networking opportunity to foster convergence research across the NSF, NIH, DOE, etc. This initiative is a "PO-driven model" where Program Officers from basic science institutions like the National Science Foundation (NSF) discuss the applications of PIs (Primary Investigators) with equivalent staff from the National Institutes of Health (NIH), Department of Energy (DOE), and other U.S.-based applied research funding agencies. This PO-driven model can help drive innovation in the United States and connect basic and applied research quickly and efficiently. Program officers from these various granting agencies can create/solicit teams that have 'converged' on similar ideas and that might form innovative teams or networks. As program officers across these granting agencies rarely collaborate, the opportunity to cross pollinate ideas across disciplines and move ideas from basic research ideas to applied technologies can be created under this "PO-driven model" and can help drive the "bioeconomy" of the future.

6. PhenBank: Creation of a national center for phenotypic information
   Creation of a structured, centralized, and stably maintained repository of phenotypic information conceptually modeled on GenBank, where data are searchable, linked, and routinely deposited.

7. MC-4-G2P: Modeling frameworks for convergence from genotype-to-phenotype
   Convergent evolution may arise from distinct genetic bases, at different levels of genotype to phenotype map, and across phylogenetically distant taxa, making it challenging to model, and consequently to detect and characterize from empirical data. How do we take the wealth of diverse empirical data and rigorously evaluate hypotheses to integrate across taxa, across scales, across ecologies, and across data types (genomics, pheno-omics, ecological data) to find generalizable principles of convergent evolution? A comprehensive modeling framework for convergent evolution is currently missing in the field. We propose a concerted effort to develop the mathematical and computational tools needed to inspire and facilitate work on convergent evolution and other questions involving the genotype-phenotype map. Such an endeavor would be generalizable across taxa and build mechanistically across layers of biological organization.

8. Leveraging convergence to understand multi-trait evolution across biotic interactions
   Convergent traits are often analyzed in isolation, either as one trait or for one species, and often traits shaped through abiotic selective forces, such as

adaptations to extreme environments. However, interactions among species, such as in pollinator-plant or host-parasite interactions, are important drivers of convergent evolution. These interactions across trophic levels can simultaneously shape multiple traits as they covary under the selective forces created by species interactions. Repeated bouts of multi-trait evolution may be thought of as a trait syndrome, a characteristic suite of covarying traits. To fully understand the convergent multi-trait evolution of complex traits, we must examine how species interactions shape trait syndromes within and across species. The most explanatory studies will examine the evolution of these covarying traits in a phylogenetic and coevolutionary framework (whether mutualistic or antagonistic) and seek to understand the genetic underpinnings of these processes. Understanding multi-trait evolution in the context of species interactions has great translational potential in conservation, agriculture, human health, and wildlife disease management as convergent traits are understood in a predictive framework.

9. PanEDGE: Building the functional tree of life

Creation of national center(s) to enable building a functional tree of life, by delivering “model organism” packages for key nodes. Currently, individual labs or small collaborations are attempting to develop functional genomics tools for emerging model organisms. However, development of stable transformation approaches in non-model organisms is not appropriate for graduate student or postdoc projects because of the high risk and low rewards (even though these tools often lead to high impact papers). Funding of core facilities dedicated to development of transformation tools for non-model organisms would kickstart the next stage of innovation and discovery.

10. Towards integrated analyses of multiscale, multi-omic, & multivariate data

Leveraging biology to solve societal challenges requires integrating across biological scales. The study of convergent evolution provides one thematic mechanism that ties together common questions across biological scales. However, many questions about convergence (including its frequency, underlying basis, drivers) require multiscale, multi-omic and multivariate data as input to models that can jointly capture evolution, phylogeny, and traits. Currently, the community lacks the tools to perform comparative analyses within integrated phylogenetic models that simultaneously assess data from multiple scales and data types. Two consequences of this big gap are that (i) the richness of many datasets cannot be fully utilized, and (ii) incomplete and inaccurate predictions are being made. To address this shortcoming, we propose to develop phylogenetically aware data-integration models and to coordinate amongst the community engaged in model development and application (via workshops and stakeholder/end-user consultation and collaboration). In addition, we propose the development of an institutionalized web portal and suite of computational resources for hosting, sharing, and disseminating models and analyses, for the use of the broader biological community.

11. Hopeful monsters: Bringing ancestral inferences to life to test evolutionary trajectories

Understanding the evolutionary trajectory of a trait could help make the engineering of said trait easier, for societal benefit. Understanding how phenotypes have responded to past selection could be potentially relevant in the future for applied phenotypes. Here, we propose to synthesize evolutionary intermediates representing probable ancestral states to test alternative paths through inferred fitness landscapes. This approach offers the opportunity to study systems where we have mechanistic understanding of traits (often, morphological traits), but also provides a framework to study less understood traits (such as behavioral, life history, physiological traits). This recommendation has the potential to yield generalizable principles about how developmental constraints and biases shape evolutionary trajectories and allow scientists to move from solely looking backwards (in time) to a forward-looking paradigm.

12. A two-tiered funding mechanism to promote interdisciplinary collaborations
    Current funding mechanisms require more preliminary data than can be generated to shift fields, which is a barrier to transformative research. We recommend the implementation of a two-tier funding program to exploit existing -omics and other big data to understand the mechanisms and consequences of convergent evolution. Tier 1 would enable researchers with different expertise (*e.g.*, computational biology and experimentalists) to work together to exploit existing datasets to generate novel testable hypotheses about convergent evolution. If successful, as measured by the generation of clearly defined and testable hypotheses, teams could move forward to increased funding to test those hypotheses, either through the generation of experimental data and/or the generation and analysis of additional “big” datasets.

13. Informing convergent evolution of chromatin & genome organization
    The way in which genomes are packaged within cells is a critical component in converting a genotype into a phenotype, which has important impacts on how organisms evolve. However, our understanding of how genome packaging evolves is largely unresolved. We now have the capacity to address this important, unresolved question because high-throughput methods are capable of measuring chromatin conformation, 3D organization, and other genomic and epigenomic features in any eukaryote from which we can obtain fresh biological material. This presents us with a tremendous opportunity to study chromatin and 3D genome organization as a nexus point for a wide array of evolutionary processes, including convergence. For example, recent studies have begun to examine how gene expression, 3D organization, and DNA methylation evolve on phylogenies to affect organismal phenotypes. Despite these advances, we lack sufficient sampling of taxa and phenotypes to gain a clear understanding of how genome packaging and phenotypes affect each other and their evolution. We propose to map information on chromatin structure onto phylogenies to use the replicated instances of genome structure to determine convergent molecular mechanisms of gene regulation.

14. Evolution of biotic interactions across scales
    We seek the establishment of funding mechanisms to support the development of research clusters or a center/institute to understand contemporary ecological

interactions through a deep-time reconstruction approach that will elucidate the mechanistic basis of species trait evolution. By studying these interactions across levels of biological organization, we will be able to make strong inferences about the constraints and repeatability of coevolutionary processes in a wide range of taxa. The approach requires knowledge concerning the molecular basis of interactions between species, which leverages phylogenetic information to reconstruct ancestral genes associated with focal traits. Using genome editing technology, we will not only be able to characterize the phenotypes of individuals, but the quantitative interactions between species by using behavioral assays, experimental evolution, and novel modeling platforms that will provide new theories on the evolutionary ecology of species interactions. In addition to integrating historical information into contemporary views on species interactions, several practical applications will emerge from this platform with potential to advance areas of novel drug discovery, climate-adapted transgenic crops, and microbiome-mediated host nutrition.

15. Developing a holistic genotype-phenotype model
    Phenotypes are not just dependent on the genotype of an individual and their environment. Rather, phenotypes are also dependent on the interactions between different parts of the genotype with each other, and these interactions are contingent on the environment and epigenetics factors. Thus, a single genotype is associated not with a single phenotype, but rather associated with a phenotypic space of possibilities. In this recommendation, we propose to define the phenotype space for select species, develop new genotype-phenotype models, and test these models in a range of species showing convergent phenotypes from across the tree of life.

16. Data integration for convergence research using a model-clade approach
    Traditional model species are distantly related and do not represent convergent phenotypes. To spur advances through convergent evolution, we propose resources to support research on clades (groups of related species) encompassing multiple independent origins of the same convergent phenotype. Use of a model-clade approach to generate multi-omic data and database resources would enable convergence research in a manner not currently possible. Community building is a required step for this idea so that researchers with different expertise can create a team to focus on specific clades of interest. This approach will help level the playing field and enable and inspire convergence research for non-model organisms in a phylogenetic framework.

17. Researching multi-scale convergent responses to stress grounded in human context
    Evolutionary convergent responses to human-driven biotic and abiotic environmental change occur across biological scales and must be viewed in context. Marginalized communities disproportionately witness and adapt to environmental stressors; however, rarely are non-biologist local understanding and knowledge included in definitions of context, assessment of biological responses, or development of solutions to environmental change. We propose to identify the environmental context of ecosystems that have been highly impacted by human activities. Identification, study, and biologically based solutions will be done in direct collaboration with

impacted marginalized communities in both rural and urban settings. This approach allows for identification of stressor-resilient organisms, genes, mutations, and mechanisms of adaptation and evolution that can be used to develop solutions to challenges in collaboration with impacted communities.

18. A (social) web to stimulate interdisciplinary collaborations
    Design, development, and launch of a platform and resource to facilitate interdisciplinary and diverse scientific partnerships. This platform would capture elements of social media platforms, including the ability to identify and connect with individuals with commonalities as well as suggestions for new contacts. To design a successful web that is the foundation for meaningful collaborations, we recommend that knowledge and practices from social science researchers be leveraged. Artificial intelligence can be used to power automated recommendations for new contacts and areas that may be the foundation of new collaborations. This pairing would help to stimulate new thinking and overcome existing research silos that impede research potential.

Cross-Cutting Themes

During the Synthesis Session, participants identified and discussed several cross-cutting themes that occurred across multiple recommendations, which are highlighted below.

New approaches to improve biological predictions

Many recommendations stressed the critical need to develop new mathematical models, computational methods, and theoretical frameworks to improve biological predictions. Accurately predicting genotype-phenotype associations requires better methods to collect and leverage disparate data. Several groups called for concerted efforts to expand the conceptual model of the genotype-phenotype map to incorporate data from phylogeny, organismal interactions, and epigenetics. The need for more phylogenetically informed predictive models was emphasized by multiple groups. Work in this area is fundamentally trans-disciplinary, requiring input from multiple domains (computational/statistical, organismal/phylogenetic, molecular/functional). Moreover, progress in this area requires better integration of communities engaged in model development and model application.

A specific type of biological prediction that received a lot of attention during the scoping sessions was that of predicting biotic interactions between organisms. Studying the evolution of biotic interactions as an emergent organismal property has the potential to provide mechanistic insight into these interactions and thereby facilitate the future modification/engineering of these traits that are critical to overall organismal/environmental health. Suites of traits (trait syndromes) often arise together and lead to organisms that perform similar roles in an ecosystem. Understanding how these syndromes evolve will help us manage those roles better.

Participants did not think the current funding environment incentivized this kind of research in a comprehensive way. For example, funding in this area is often limited in scope and/or focused on large institutes or centers. The culture of standard grant review panels needs to be modified to better evaluate high-risk/high-reward projects. One method to achieve this could be through the creation of two-tiered funding structures, where smaller awards are given out in the first stage to risky, exploratory research, and larger awards are provided in the second stage once preliminary data has been collected and successful collaborations have been established. Additional funding-centered recommendations focused on encouraging strategic collaborations between theory and empirical scientists. For example, multiple recommendations included the creation of supplemental grants to incorporate modeling, theory, genetic transformation, *etc.* into existing projects.

Engineer novel or strategic traits

Several recommendations focused on ways that the evolutionary community might take advantage of new molecular tools to engineer radical phenotypes or evolutionary intermediates. Investment in this area will not only allow evolutionary biologists to experimentally test hypotheses and tinker with strategic phenotypes, but it also has the potential to yield new trait combinations that do not already exist. Scientifically informed trait combinations, whether ancestrally inferred or altogether novel, will allow bypassing potential valleys in fitness landscapes and has the potential to produce organisms/systems that may help solve current challenges. Moreover, the ‘resurrection’ of hypothetical ancestors/ancestral traits has the potential to revolutionize the study of evolutionary biology and help with a mechanistic understanding of convergence. Partnering with the new NSF Technology, Innovations and Partnerships (TIP) directorate would bring together academic researchers interested in basic science questions with applications that may appeal to industry.

Critical to the success of these trait engineering goals is strategic selection of taxa/traits for functional investment. Some groups have already constructed centers, institutes, and administrative structures for bringing genetic engineering to diverse non-model communities from support by NSF EDGE and RCNs funding, as well as other short-term mechanisms. However, more long-term funding mechanisms are necessary to keep these efforts alive and democratize genetic engineering for the masses across the tree of life. Substantial knowledge and experience are needed to facilitate democratization of genetic engineering of non-model organisms. This knowledge will require trained technicians in longer-term positions than graduate students or postdoctoral researchers. Sustained resources, such as national centers for transformation, would provide expertise and training for traditionally non-molecular labs to ‘jumpstart’ research that takes advantage of this technology.

Survey the Tree of Life to discover solutions to life’s common problems

Environments around the world are rapidly changing and becoming more extreme. Many of these changes are generating negative consequences for organisms, populations, communities, and whole ecosystems. Therefore, there is an urgent need to understand what makes a system resilient to environmental stressors and changing environments, as well as to predict how it will change in response to ongoing or future stress. If possible, it is also necessary to develop strategies to minimize negative outcomes. One solution put forward during the scoping sessions is to harness convergences across temporal, geographic, and/or biological scales to uncover mechanisms (*e.g.*, genetic, epigenetic/chromatin structure, molecular, cellular, tissue) that allow organisms to survive and thrive in response to environmental stressors/changing environments. It is also necessary to understand the consequences of these mechanisms for higher-level dynamics at the population, community, and/or ecosystem levels. Results from these studies would inform the building of models of how organisms, populations, communities, and/or ecosystems will respond to ongoing and future environmental stressors and changing environments.

Field work and lab work at this scale, combined with human resources, will be expensive. The current model of “silos of funding” isolates social and biological sciences, as well as eco-evolutionary studies in natural ecosystems and those impacting human and environmental health. One potential solution is the development of new funding programs that promote collaboration across disciplines, citizen science, and long-term projects. Funding panels should positively value collaboration with foreign scientists.

This work will also require bringing together experts from across diverse disciplines who might not normally collaborate to obtain biological samples, collect data in a broad range of organisms, and develop statistical/mathematical approaches. There is a critical need for more focused workshops to continue to expand and explore LIFE ideas across disciplines.

This work will also require inclusion of faculty at all stages. However, the current academic tenure system actively disincentivizes junior faculty from engaging in large collaborative projects, instead rewarding only individualized based research for successful tenure promotion. An overhaul of the academic advancement model will be required to ensure all scientific contributions are weighted equally when measuring success. The NSF could provide extra support for junior faculty that is viewed favorably by home institutions, perhaps in the form of a supplemental document in standard grant applications that outlines the role and benefits for junior faculty, similar to what is already required for graduate students and postdoctoral scientists.

Work in these areas will require individuals willing to devote the time necessary to develop new skills and technologies, especially while working as a group. This can be particularly challenging for individuals with greater time restraints (*e.g.*, teaching load). The projects also have a potentially long timescale. Funding panels must realize the logistics and time scales of these projects and support them accordingly.

It can be challenging to coordinate effectively with multiple teams, and often scientists are not trained to do so. This can be further confounded by potential language, technical, and interest barriers (*e.g.*, establishing genuine contacts and collaboration in environmental justice-impacted communities takes time, social science research, and expertise to develop and build meaningful trusting and lasting collaborations). Moreover, inconsistent data collection standards, a dearth of sufficient data depositories, and inconsistent standards for quality checking reduce our capacity to compare data from replicate studies and previously collected data.

Therefore, a critical need exists for centralized leadership that is explicitly responsible for training in logistics and methods that will be used across fields. This would enable the development and enforcement of methods that allow us to compare variables across datasets. This could be implemented through NSF courses or an institution in a grant proposal, with potential benefits for non-R01 institutions to act as training centers.

Novel institutional investments

During the scoping sessions, a significant amount of discussion centered on ideas for novel institutional investments that would enable transdisciplinary breakthroughs in evolutionary innovation and convergent evolution. Of the 18 recommendations included in this report, seven explicitly comment on the dire need for better online databases. The current funding model requires online databases be self-sustaining, which has resulted in the loss of many useful online resources over the years as their funding support has run out. The taxon- or trait-specific nature of many databases was listed as another limitation that prohibits integration across diverse organisms and/or phenotypes. To address these issues, several participants of the scoping sessions advocated for development of a National Center for Phenotypic Information (e.g., NCPI/PhenBank, similar to NCBI/Genbank for genotypic information) to foster convergent research across agencies. Within the larger NCPI organization, a National Center for Convergent Traits could be developed to leverage the knowledgebase of phenotypes that show trait convergence among organisms. Convergent traits of interest could include those that are economically relevant or climate-change associated. To be successful, a National Center for Phenotypic Information would require stable, long-term funding with support for expert curators, data scientists, and staff bioinformaticians to build standard computational tools, database infrastructure, and other enabling technology so that the community can make contributions and grow the resource.

Another recurrent theme that arose within both in-person and virtual scoping sessions was the need for a creative and efficient mechanism, in the form of a science social media application/platform, to connect potential collaborations and promote interdisciplinary networking. Users could join “social” groups associated with specific phenotypes, specific organisms, or overarching goals (*e.g.,* “monster” phenotypes). Use cases of a social media application would be the crowdsourcing of new phenotypes to target for investigation and connecting scientists who might provide general knowledge and data for phenotypes of interest with scientists with expertise on genetic engineering, theoretical modeling, *etc*. This social media platform could be expanded to provide an online meeting venue to host larger discussions on topics requiring transdisciplinary collaboration and input from multiple scientific communities.

Common needs and barriers

In this section, we provide a general overview of several cross-cutting needs and/or barriers that were listed as common challenges across multiple group recommendations.

- **New technologies**
    - Advances in sequencing technology (*e.g.*, chromatin conformation capture) to make it more cost effective and accessible to more research programs and diverse study systems (see recommendations # 2, 13).
    - Advances in sampling approaches (*e.g.*, robotics) to enable comprehensive data acquisition for more organisms and across diverse, often challenging, environments (see recommendations # 1, 4, 8).
    - Advances in technology and support for genetic transformation/manipulation of non-model organisms (see recommendations # 1, 3, 5, 11).
- **Shared resources**
    - Stable and fully supported online databases with consistent ontologies and data standards (see recommendations # 5, 6, 8, 14, 15, 16, 18).
    - Agreed upon community standardized protocols for data acquisition, management, and analysis (see recommendations # 4, 6, 15, 16).
- **People**
    - Tools to facilitate networking and building of collaborative teams (see recommendations # 1, 4, 7, 12, 13, 18).
    - Changes to the academic reward/promotion system to recognize and encourage collaborative research (see recommendations # 4, 7, 9).
    - Community buy-in to support recommended changes and/or new institutional investments (see recommendations # 6, 16, 17, 18).
    - Methods and strategies to address the public perception and ethical concerns associated with genetic engineering (see recommendations # 3, 17).
- **Changes to current funding structures**
    - Mechanisms to fund more high-risk research (see recommendations # 1, 2, 3, 4, 9, 12, 14).
    - Long-term funding support for research sites, databases, and core facilities (see recommendations # 4, 6, 8, 9, 16, 17).
    - Improved coordination across multiple agencies (see recommendations # 2, 5, 6, 17).

- **Data**
    - New theory and/or mathematical models (see recommendations # 1, 2, 6, 7, 8, 11, 13, 14, 15, 16).
    - Strategies to establish consensus on targets/phenotypes to prioritize (see recommendations # 1, 3, 6, 8, 11, 13, 16, 17).
    - Greater taxonomic sampling for strategic phenotypes of interest (see recommendations # 1, 4, 8, 13, 17).
    - Protocols and strategies for integrating across diverse datasets (see recommendations # 1, 11, 14).

Proposed solutions

While many of these perceived obstacles were shared across many recommendations, many of the proposed solutions were unique to a specific recommendation, providing a range of possible action items to explore and potentially implement in the future. In this section, we provide a general overview of several of the proposed solutions listed in the individual recommendations.

- **New funding strategies**
    - Specific solicitation to pair theory and empirical scientists (see recommendation # 7).
    - Supplemental grants to incorporate critical interdisciplinary research need (modeling, theory, genetic transformation, *etc.*) into already funded projects (see recommendations # 7, 9).
    - Two-tiered funding mechanisms to enable high-risk projects (see recommendations # 11, 12, 15).
    - Funding to engage senior and retired natural history experts and connect them with early career scientists (see recommendation # 8).
- **Collaboration and community development**
    - Support for additional scoping sessions and working groups to identify and develop benchmark datasets, models, and taxon sampling (see recommendations # 10, 16).
    - Development of a social network to promote collaboration (see recommendations # 3, 18).
- **New infrastructure**
    - Creation of a stand-alone, permanently funded repository for phenotypic data (see recommendation # 6).
    - Development of new software portals and/or data repositories (see recommendations # 10, 16).

  - Creation of a core facility for developing functional tools for non-model organisms (see recommendation # 9).
  - Creation of research centers to coordinate integrative research efforts on convergent evolution and evolutionary innovation (see recommendations # 10, 14, 15).

Training and broadening participation

Participants of the scoping sessions envision a comprehensive transformation in scientific training that transcends traditional boundaries and promotes collaboration, data-driven evaluation, and creation of accessible resources. Recommendations underscored the importance of engaging various stakeholders and communities, including bioethicists and social scientists, to advance scientific knowledge and prepare the next generation of interdisciplinary scientists and educators. Participants recognized the importance of structured, immersive interdisciplinary training at all educational stages, from K-12 to senior faculty. However, some participants of the synthesis session saw the greatest need to revolutionize scientific training at the undergraduate level. A series of NSF-funded workshops focused on curriculum development was one recommended method to address this critical need. Currently, curriculum development occurs at the university level, which leads to inconsistencies between programs and a lot of redundant effort. A holistic approach to integrative curriculum development would encourage collaboration among universities, span traditional disciplinary boundaries, and establish community consensus on what 21st century life science curricula should entail.

This research on evolutionary innovation is collaborative and needs evolutionary biologists, cell-molecular biologists, bioinformaticians, statisticians/mathematicians, and bioengineers. This interdisciplinary environment is ideal for graduate student training, promising to strengthen data analysis techniques and skills in our emerging workforce. Support for "Grad student sabbaticals" was one proposed mechanism that arose during the synthesis session to enable cross-lab training for graduate students. Currently, student exchanges of this nature are difficult for investigators to fund. One mechanism could be through the creation of a new supplemental funding opportunity through the NSF, which would also support current NSF efforts to provide comprehensive research opportunities for diverse individuals, build training/collaboration across institutions (*e.g.*, between R01s and PUIs), and support scientists from underrepresented backgrounds. Critically, participants advocated for a data-driven approach to determine the most effective training and funding methods and suggested the adoption of predict-test-learn cycles for figuring out what training and funding mechanisms produce best outcomes in terms of producing interdisciplinary scientists.

Additional recommendations to address LIFE training goals include creation of a new NSF postdoctoral fellowship track in complex data integration in phylogenetic context (see recommendation # 10); course-based undergraduate research programs (see recommendations # 2, 8, 15, 16); and multi-institution training grants to encourage cross-silo information sharing (see recommendation # 3).

The creation of biobanks and databases was suggested in many recommendations. In addition to being essential resources for transformative research in evolutionary innovation, they would also serve as accessible resources for the broader scientific community and the public. For example, teachers would have access to online resources of the National Center for Phenotypic Information and could use these resources for K-12 educational module development. These biobanks and databases would also serve to democratize research by making resources freely available, thereby decreasing barriers to low-resource institutions (see recommendation # 6)

Many of the proposed recommendations would benefit significantly from the inclusion of local populations that could contribute to our shared understanding of extreme or changing environments and the organisms that reside in them both currently and historically (see recommendations # 4, 17). For example, marginalized communities disproportionately witness and adapt to environmental stressors; however, only rarely is the understanding and knowledge of local non-biologists included in definitions of context, assessment of biological responses, or development of solutions to environmental change. One way to solve this problem is to directly involve impacted marginalized and/or local communities in the hands-on field research for the proposed project and get their feedback for the biologically based solutions. To facilitate this goal, practical ways for compensating community participants for their participation need to be developed. The incorporation of local populations both in and outside of the US also needs to be “rewarded” by its acknowledgement as a valid broader impact. This approach has the potential to increase the buy-in from local communities that are needed to make real and sustained changes. Biologists could learn from best practices developed by economists as well as social and behavioral scientists who work closely with the communities they study directly.

Bioinspired solutions for the bioeconomy

Participants of the scoping sessions identified numerous ways that investment in fundamental research in evolutionary innovation and convergent evolution could be leveraged for the bioeconomy. Nature has often found common solutions to extreme challenges. By studying these natural solutions, we can derive unifying principles to inform synthetic biology and bioengineering applications, such as the use of cold-adapted enzymes to make industrial processes more environmentally friendly. Improving our ability to predict the functional impact of a particular genetic modification, as well as to generalize from traits in one organism to another, would have significant impact to agriculture with the potential to decouple productivity/tolerance tradeoffs in crop species through the creation of efficient and resistant super-crops. A better understanding of complex microbial communities and how these communities will respond to environmental perturbations could enable engineered microbiomes for enhanced human health, agricultural production, biofuel/bioproducts.

Specific recommendations centered on three main aspects of the bioeconomy.

- **Enabling genetic engineering of diverse organisms and traits**
    - Evolution inspired optimization of economically relevant phenotypes (see recommendations # 1, 11)
    - Engineering novel or economically relevant phenotypes (see recommendation # 3)
    - Adding 'extinct' characteristics back into valuable organisms (see recommendation # 11, 14)
    - Improving gene function predictions to facilitate genetic engineering efforts (see recommendations # 1, 10, 15)
- **Informing biomedical applications**
    - Novel therapeutics (see recommendations # 3, 4, 14)
    - Strategies to combat multi-drug resistant pathogens (see recommendation # 2)
    - Identification of commonalities associated with diverse diseases (see recommendation # 8)
    - Identification of traits associated disease resistance (see recommendation # 8)
    - Engineering evolutionarily stable host microbiomes (see recommendation # 14)
    - Predicting host-parasite interactions (see recommendation # 2)
- **Strategies for a resilient planet**
    - Identification of organisms resistant or resilient to environmental stress (see recommendation # 17)
    - Determining which traits are responsive/resistant to climate change for conservation efforts (see recommendations # 4, 8)

Participants of the synthesis session also advocated for greater support of paid internship programs for graduate or undergraduate students in the basic sciences to interact with industry. These programs would increase student awareness about the bioeconomy and provide insight on how they might contribute to it. Additionally, partnerships with the new NSF TIP directorate could be leveraged to make industry aware of how concepts from basic research may be integrated into their product development pipelines.

Acknowledgements

Many thanks to the LIFE Participants as well as to the Knowinnovation facilitators and workshop producers: Annemarie Boss, Jocelyn Tejeda, Kyle McCarthy, and Zach

Stapleton Jones. Thanks also to Sam Cleveland, Jaime Turner, Jenny Wilson, Sam Pucka, and Greg Wisecaver for their assistance.

This publication is a synthesis of work supported by the National Science Foundation under grant MCB-2326865 (any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation).

Scoping Session Recommendations

This section provides details on the 18 recommendations that emerged from the LIFE Scoping Sessions. The table below illustrates how the recommendations fit into the six Branches of LIFE.

Branches of LIFE

	New frontiers in evolutionary convergence	Overcoming challenges associated with non-model organisms	Tool and technology development	Collaboration and communication among disciplines	Leveraging life's diversity for the bioeconomy	Training and broadening participation
1. PhunCODE: Phenotypes & functions to leverage convergence, diversity, & evolution	X	X	X	X		
2. Learning from phages: Predicting & deciphering host-pathogen specificity	X	X	X	X	X	X
3. LIFE: Evo-inspired engineering of radical phenotypes & emergent traits	X		X	X	X	
4. Leveraging diversity in extreme environments	X	X			X	
5. Convergence in action	X			X		
6. PhenBank: Creation of a national center for phenotypic information	X	X	X	X	X	X
7. MC-4-G2P: Modeling frameworks for convergence from genotype-to-phenotype	X	X	X	X		
8. Leveraging convergence to understand multi-trait evolution across biotic interactions	X	X	X		X	X
9. PanEDGE: Building the functional tree of life		X	X	X		X
10. Towards integrated analyses of multiscale, multi-omic, & multivariate data		X	X		X	
11. Hopeful monsters: Bringing ancestral inferences to life to test evolutionary trajectories	X	X	X	X		
12. A two-tiered funding mechanism to promote interdisciplinary collaborations	X	X	X	X	X	X
13. Informing convergent evolution of chromatin & genome organization	X	X	X	X		
14. Evolution of biotic interactions across scales	X	X		X	X	X
15. Developing a holistic genotype-phenotype model	X	X	X	X		X
16. Data integration for convergence research using a model-clade approach		X	X	X		X
17. Researching multi-scale convergent responses to stress grounded in human context	X				X	X
18. A (social) web to stimulate interdisciplinary collaborations			X	X	X	X

1. PhunCODE: Phenotypes & functions to leverage convergence, diversity, & evolution

What is your Big Idea?

How can we traverse the Genotype - Phenotype map? To be able to leverage convergent evolution to understand organismal adaptation to diverse environments, common responses to climate change, and shared functional innovation across the Tree of Life, we need to improve our ability to predict how phenotype arises from genotype. While genome sequence is readily available or can be readily produced for almost any organism, there are many species for which intermediate molecular phenotypes are challenging or impossible to generate, for technical or ethical reasons. We propose to develop predictive models of intermediate molecular phenotypes (such as chromatin state, transcription, translation, protein structure). These models will enable:

- Inference of intermediate molecular phenotypes across diverse species that are not amenable to laboratory manipulation or destructive sampling
- Researchers to fill gaps in a “Phylo Matrix” (genotype -> molecular -> organismal) by imputing missing data (*e.g.*, infer chromatin state in species without experimental evidence using phylogenetic relationships and other species with experimental data).
- Development of a framework to predict links between genotype and phenotype (organismal traits, molecular traits), across the tree of life

Crucially, these models will likely not be highly accurate to begin with. We propose that using Predict -> Test -> Learn cycle will allow predictive models to rapidly improve in power and accuracy.

Which *Branch of LIFE* does this address?

Our recommendation addresses several branches of Life, including new frontiers in evolutionary convergence; overcoming challenges associated with non-model organisms; tool and technology development; and collaboration and communication among disciplines.

What is required to pursue this?

- Research needs:
    - New tools that are optimizing for phylogenetically informed prediction. These methods will likely develop out of collaborations between machine learning (ML) experts, experts in phylogenetic comparative methods, and molecular and organismal biologists. For example, one new outcome might be understanding how to incorporate phylogeny more directly into predictive/ML models.
    - New assays, in a diversity of organisms, are also required, as we expect that there is currently not sufficiently deep phylogenetically connected datasets to provide the input for ML methods and other predictive approaches.

- Funding mechanism: big enough to support trans-disciplinary teams; including community to encourage disciplines to learn to talk to each other; encourage high-risk high-reward; include predict-test-learn ideas; translational supplements to incentivize applications.
- TIP funding / SBIR-like mechanisms: leveraging diversity for applied applications.

What are barriers to pursuing this now?

- Building transdisciplinary collaborations / teams: this work is fundamentally transdisciplinary and requires collaborative science to rapidly progress.
- Lack of underlying data (this could be a way to “seed” a PhenBank type idea). ML methods and other predictive tools are extremely “data hungry” and the more we can accurately measure, the better our predictions will be.
- Need to know what data to collect to train predictive models. What traits are amenable to prediction in the first place?

What is the expected value and impact?

- Make it possible to understand things that cannot be measured via predictions.
- Improve ability to study across times, scales in systems that are not amenable to experimental testing, e.g. predicting gene expression in response to heat stress in a climate-endangered organism that cannot be destructively sampled.
- Tools: ML-based imputation of molecular phenotypes from genotypes/other phenotypes has wide-scale impact across many fields.
- Translational impact via understanding key economically relevant phenotypes for evolution-inspired optimization.

What are the training and workforce opportunities within this idea?

- Undergraduate curriculum often reflects “traditional” boundaries: this idea could encourage math in biology, phylogenetic thinking in molecular biology; molecular thinking in organismal biology, and all the other sorts of cross-discipline ideas.
- Encourage trans-disciplinary collaboration; incentivize transdisciplinary mentorship.

What is the reasoning, justification, and/or supporting evidence behind this idea?

- Genomes are readily accessible.
- ML / prediction and molecular phenotyping have advanced to the point where integration is feasible, but they often are not integrated in an evolutionary context.
- Now is the time for integration.

Authors

Antonis Rokas	Vanderbilt University
Rob Unckless	Kansas University
Tim Sackton	Harvard University
Chris Todd Hittinger	University of Wisconsin-Madison
Nathan Clark	University of Pittsburgh

2. Learning from phages: Predicting & deciphering host-pathogen specificity

What is your Big Idea?

Genomes of viruses contain all the information required to take control of and replicate themselves using their hosts. We have access to an unprecedented amount of genetic data, but our ability to find the relevant bits that determine host-virus interaction is still mostly unworkable. Harnessing the incredible amount of data we can now collect as well as the experimental tractability available for working with phage and bacteria, we aim to predict which phage can infect which bacterial hosts using machine learning. These predictions can be used for phage therapy and ecosystem engineering applications, and the feature selection by the algorithm can help provide insights into fundamental microbiological processes by revealing the genes and pathways that enable interactions between phage and bacteria. The training features will include information taken at different scales including whole genome sequences, gene network predictions, and predictions of protein-protein interactions. The large datasets that will be collected as part of this work will also be valuable for the study of host-pathogen evolutionary dynamics, including the repeatability of evolution in different conditions/landscapes.

Which *Branch of LIFE* does this address?

- New Frontiers in Evolutionary Convergence, as we learn what convergent mechanisms phylogenetically diverse phages use to infect the same hosts. Models of the processes and model-inspired experiments will help us understand co-evolutionary dynamics, drivers of phage innovation, and under which conditions parallel evolution is most likely.
- Non-Model Organisms, as this work will rely on technologies to study interactions between bacteria and phage that are not culturable.
- Tools and Technology, as this work will lead to new high-throughput technologies to measure phage-bacterial interaction networks, protein structure and interaction modeling, and machine learning approaches. The platform itself, which will capture genome/genome information to make predictions, will be a new technology.
- Collaboration Across Disciplines as we will need the assistance of data scientists, microbiologists, genomicists, and structural biologists.
- Training and Broadening Participation as the project is highly suitable for developing Class-based Undergraduate Research Experiences (CURE).
- Leveraging diversity, as we collect unprecedented amounts of genomic and interaction data from Earth’s two most diverse groups of life forms; bacteria and the viruses that infect them.

What is required to pursue this?

- Technology innovation in creating a high-throughput, sequencing-based approach to measure phage-bacterial interactions in diverse environments.

- Technology innovation in modeling phage and bacterial protein structures and their interactions, ideally accompanied by experimental structural biology to provide templates.
- Resources needed to hire biologists and data scientists.
- Data storage for the massive genomic and interaction network datasets, as well as the computing power to run machine learning algorithms to analyze it.
- Develop new partnerships and strengthen existing collaborations to integrate and coordinate expertise.
- Translational application to bacterial infections will require coordination with clinicians and regulatory agencies (tap into existing efforts).

What are barriers to pursuing this now?

The data and machine-learning concepts needed to tackle this on a small scale exist now. However, many bacteriophage applications require strain-level precision, which will require much more data to make predictions with that specificity. We will need to collect more phage-bacterial interaction data, which creates a combinatorics problem since as the number of phages and bacteria increases, the number of interactions grows exponentially. The technology to capture these datasets does not exist and is a major barrier. Methods have been proposed, but funding to develop and validate them is necessary.

What is the expected value and impact?

- A rapid tool to inform which phages can be used to treat novel multidrug resistant bacterial infections.
- Fundamental understanding of what governs the interactions between the most specious groups of life forms.
- While this is microbial focused, the tools developed should be able to be leveraged to predict other host-parasite interactions (e.g. mammals and their viruses, plants and their fungal pathogens).

What are the training and workforce opportunities within this idea?

CURE development. The SEA-PHAGES program already involves undergraduates in collecting, testing, sequencing, and annotating bacteriophage. This program can build on this infrastructure, incorporating additional bacterial strains into the work and designing controls to assess data quality and reliability.

The research is well-suited for undergraduate, graduate, and postdoctoral data scientists, genomicists, structural biologists, and microbiologists who will gain interdisciplinary experience at the boundaries between these fields.

What is the reasoning, justification, and/or supporting evidence behind this idea?

- Pilot trials have been successful with multiple limited datasets, but now is the time to scale up by investing in technological innovations.

- Virus phenotype has been previously successfully modeled using the cutting-edge machine learning tools trained on virus genomes (cf. Hie et al., Science 2021). We will generalize these algorithms to genome-genome interactions.
- The antimicrobial resistance crisis demands new solutions, and this will greatly enhance phage therapeutic effectiveness. Development of phage therapies is an active area (FDA trials are underway), and medical experts who need the information that this research will provide can help advance the work; advocacy groups can help with regulatory hurdles.
- There are other attractive areas in which to deploy bacteriophage, including environmental remediation, agriculture/food supply, biocontrol of contamination, tuning the gut (or other) microbiome.

Authors

Justin Meyer	University of California, San Diego
Luis Zaman	University of Michigan
Amy Keating	Massachusetts Institute of Technology
Alexey Markin	USDA-ARS
Clay Small	University of Oregon
Giovanna Ghirlanda	Arizona State University

3. LIFE: Evo-inspired engineering of radical phenotypes & emergent traits

What is your Big Idea?

Evolution in nature happens slowly, often resulting in incremental phenotypic changes. We need to create an experimental paradigm that enables saltatory jumps that will enable the rapid origin of new phenotypes and emergent traits. By artificially bringing together traits that have never been expressed together before (by natural processes) we can change the starting point of evolution and widen the range of possible phenotypes that can be selected in novel environments. The ability to rapidly create radical phenotypes will open possibilities for engineering real-time biological solutions in our rapidly changing world.

Which *Branch of LIFE* does this address?

New Frontiers in Convergent Evolution, Collaboration Across Disciplines, Tools and Technology

What is required to pursue this?

- Develop new high-risk high-reward pan-directorate grant opportunity/solicitation e.g., similar to TR01- transformative R01 training grant to fund “moonshot” projects
- Encouraging a review panel culture that gathers groups of broad thinkers that can realistically evaluate high risk, interdisciplinary grants
- Multi-institution training grants for trainees (students and postdocs) to learn functional genomics and bioengineering techniques (NRT) to train multidisciplinary diverse thinkers.
- A Research Coordination Network (RCN) grant to integrate disparate ideas/connect unusual research programs
- A Use-inspired Global Center (GC) focused on bio-inspiration to design/build new life and facilitate international collaboration on a broad range of ideas/topics to drive the bioeconomy.
- Infrastructure grants for core facilities that would facilitate creation of organisms with novel trait combinations. (E.g., cores that could generate and/or house transgenic organisms)

What are barriers to pursuing this now?

- Aversion to funding of high-risk research.
- Identifying key convergent traits to mix-and-match by continuing to invest in basic research across the tree of life and ecosystem landscapes.
- Identifying the genetic changes to make that will produce the desired effect using a comparative approach across disciplines.
- Functional misbehavior of heterologous proteins in host intracellular microenvironments is a significant challenge that will require investment in core infrastructure labs to support large-scale experimentation.
- Finding the right balance of ambition and feasibility to provide a reasonable probability of a breakthrough.

- Public perception of and expected resistance to GMOs.

What is the expected value and impact?

- Developing methods for engineering new phenotypes will improve our understanding of how extant phenotypic diversity arose.
- Potential traits that are beneficial could be combined into a single organism, potentially leading to the origin of new emergent traits (that couldn’t be predicted by sequence alone).
- Change the starting point of evolution - we release ourselves from the limitations of starting with the phenotypes that nature has provided, enabling us to terraform new fitness landscapes and discover new boundaries for trait evolution
- Inform a deeper understanding of the physical limits on the evolution of diversity and address questions such as:
    - Does current diversity represent all possible evolutionary outcomes or just the ones that were stochastically selected by past environments. (e.g., what phenotypes can’t evolve? Why? What if we change the environment?)
    - Can we introduce RNA-editing technology into synthesized cell lines to change the rate of protein evolution in that cell or program a cell to self-destruct in the face of certain conditions?
    - Can we create DNA phenotypes that resist deleterious effects of double-stranded breaks by engineering holocentromeres into nuclei?
- Outcomes would include trait-specific transformative impacts that address grand challenges in agriculture, the environment, medicine, and space exploration, and create new opportunities for the bioeconomy. The impact of building photosynthetic cells, for example, that capture energy from sunlight via chloroplasts would be transformative. Potential applications range from (i) improved systems for producing antibodies, vaccines, or high-value metabolites supported by the extensive chloroplast metabolic framework to (ii) addressing environmental challenges of bioremediation and carbon capture to (iii) producing autotrophic livestock that require less feedstock (and have net reduced agricultural emissions) to (iv) producing novel life forms that can survive harsh conditions on Earth or elsewhere in the universe.

What are the training and workforce opportunities within this idea?

- Outreach: bringing ideas/examples of convergent evolution to K-12 students to inspire careers that span the interface of biology and engineering
- Education: There continues to be a lack of cohesive interdisciplinary training opportunities for undergraduate and graduate students. This opportunity will provide in-depth exposure to diverse but complementary research areas founded in functional and comparative genomics, analytical chemistry, mathematical modeling, cell biology, physiology, biochemistry/bioengineering, and evolutionary biology
- Training and mentoring: Create a series of multi-institutional modules to train undergraduates over multiple terms. Student cohorts, supported by graduate and

faculty mentors, from partner institutions would each work on a specific research area, culminating in a symposium to encourage cross-silo information sharing and collaboration.

What is the reasoning, justification, and/or supporting evidence behind this idea?

- We have made significant progress in novel research areas such as assisted evolution, synthetic cell engineering, computational modeling, among others. NSF has invested money and effort into supporting the development of engineered or synthetic cells. Assisted evolution is being employed to save coral reefs. By combining these approaches, we can identify, combine, and study traits that would never come together naturally and create fundamentally new evolutionary trajectories, akin to “plucking fruits from different parts of the tree of life to make the salad.”
- Critical timely challenges require powerful, new, and potentially risky approaches. An enhanced evolutionary capacity will help us tackle some of the most critical timely challenges facing the world today, from climate change to biodiversity loss to human health.

Authors

Leslie Babonis	Cornell University
Dan Hahn	University of Florida
Annie Lindgren	Portland State University
Allen Liu	University of Michigan
Ranjan Srivastava	University of Connecticut
Josh Widhalm	Purdue University

4: Leveraging diversity in extreme environments

What is your Big Idea?

Using convergence across temporal, geographic, and biological scales to uncover the mechanisms (e.g., genetic, molecular, cellular, tissue) that allow organisms to survive and thrive in extreme environments, and the consequences of the specifics of these mechanisms for population and ecosystem levels dynamics. We focus on convergence to explore solutions to a challenge (e.g., heat) rather than focusing on a single organism that does something the "best."

Which *Branch of LIFE* does this address?

Leveraging Life's Diversity, New Frontiers, Non-model organisms

What is required to pursue this?

- Technological innovations to facilitate sampling in extreme environments.
  - Application and development of intelligent robotic platforms that have the ability to observe, capture, and work on organisms in their environments, specifically those that would be difficult for human researchers to safely access, ability to transfer and preserve samples.
- Establishing shared protocols and parameters for data acquisition and analysis (to facilitate cross-scale comparisons).
- Development of a data repository for sharing data collected with established protocols.
- Application of thinking across the biological hierarchy, time, space, or taxonomic groups.

We propose establishing NEON-like sites in multiple sites with the same extreme environments. This will allow long-term collaborative monitoring of species dynamics and adaptive evolution in extreme environments that are forming as a result of climate change/ anthropogenic activities (mining, water drainage, waste disposal, etc.) --> leveraging opportunities for "experimental evolution" in nature (perhaps similar to LTER?)

What are barriers to pursuing this now?

- **Funding**: Field work and lab work at this scale, combined with human resources, will be expensive
- **Fostering cross-discipline collaboration:** This idea will require bringing together experts from across diverse disciplines who might not normally collaborate (do not necessarily attend the same meetings, work in different departments).
- **Inclusion of Faculty at all ranks:** For this idea to be successful it will require inclusion of faculty at all stages. However, the current academic tenure system actively disincentivizes junior faculty from engaging in large collaborative projects, instead rewarding only individualized based research for successful

tenure promotion. An overhaul to the academic advancement model will be required to ensure all scientific contributions are weighted equally when measuring success.

- **Time:** This project will require individuals willing to devote the time necessary to develop new skills and technologies, especially while working as a group. This can be particularly challenging for individuals with greater time restraints (teaching load)
- **Coordination of teams**: It can be challenging to coordinate effectively with multiple teams and often as scientists we are not trained to do so. This can be further confounded by potential language, technical, and interest barriers.
- **Logistics:** Travel to and sampling of biological materials from extreme environments involves significant preparatory compliance and permitting paperwork, especially if students are involved.
- **Import/Export Restrictions**: Achieving replicates of extreme environments will require the ability to move people and samples across country borders which will require compliance with travel as well as export/import restrictions.
- **Lack of standardized methods:** Inconsistent data collection standards, a lack of a data depository, and inconsistent standards for quality checking reduces capacity to compare data from replicate studies and previously collected data.

What is the expected value and impact?

- Potential to inform about the pace of evolution in rapidly changing environments and corresponding potential for discovering novel mechanisms of environmental stress-induced evolution that greatly expand intellectual boundaries of current evolutionary theory; major implications on predicting impacts of anthropogenic acceleration of environmental change
- Potential to generate predictions for how organisms will respond to future extreme environmental challenges.
- Potential to generate paradigms/parameters/methods to standardize this type of work across temporal, taxonomic, and biological scales
- Potential to inform conservation efforts (from knowledge of how organisms respond, what traits are responsive to change, etc.)
- Potential to identify the mechanisms that work for diverse organisms to survive and thrive in extreme environments, and can potentially be developed as therapeutics or other insights into organismal health, including that of agriculturally important species (plants, livestock, birds), companion animals (dogs, cats, birds), and humans.
- Potential for bioinspired solutions for the bioeconomy - applications to biotechnology and industrial processes (e.g. the “greening” of the industrial processes using cold-adapted enzymes)
- Potential to inform processes that could help facilitate the colonization of extreme environments (ocean, space)
- Create a biobank for archiving and sharing resources
- “Unity in diversity” Common solutions to extreme challenges “engineered” by nature can be compared and unifying principles of such engineering (in the level

of materials, structures, and systems) may be derived that can inform synthetic biology and bioengineering.

- Intellectual diversity is fostered by enabling diverse researchers to study the organisms and habitats that they are most passionate about; collaborations may be more fruitful if not everyone works on figuring out the same problem (e.g., cancer) using similar toolkits and approaches whenever they become available.
- Provides a means to extend beyond evolutionary mutant models performed in the lab to capturing natural convergent evolutionary mutant models. A powerful means to capture the pace of evolution and the effects in real time by exploiting multiple species with adaptations to extreme environments.
- Potential to discover mechanisms of speciation
- Ability to address basic science questions about mechanisms of convergent phenotypes by harnessing natural experiments at different scale of replication (e.g., populations, species)
- Recommendation: NEON for extreme environments

What are the training and workforce opportunities within this idea?

- **Research Training across career stages:** Opportunity for research projects that will be suitable for trainees at all career stages and provide broad skill sets across disciplines.
- **Scientific Outreach**: People of all ages are fascinated with extreme environments and projects that explore these habitats can help capture the public's imagination and engagement with research.
- **Early Education Programs:** Provides a great resource for early education programs. Students could engage with projects through webcams setup at research sites. Students could conduct their own studies, for example visiting a natural history museum and identifying the environments that would be considered extreme. Students could then try and identify similar traits between the different species that they observe in these environments (e.g. convergent evolution).
- **Inclusion of Local Populations:** This idea would benefit significantly from the inclusion of local populations that can contribute to our shared understanding of extreme environments and the organisms that reside in them both currently and historically.
- **Job Production:** Opportunity for job production as research facilities are developed at remote sites.

What is the reasoning, justification, and/or supporting evidence behind this idea?

- Increase in extreme climate and weather events and anthropogenic disturbances are becoming more and more common.
- Many (but not all) of the tools needed to complete the proposed project are now broadly available across diverse species, e.g., genomic tools and data, cell-based assays, long-term ecological data from field sites such as NEON etc., museum collections.

- Extreme environments are generally much less species-rich than other environments - meta-omics data are much easier to interpret and discern, single species strains of microbes are easier to derive
- Often many replicates of extreme environments exist, eg salt ponds, different deep sea areas, deserts, etc - so there is potential for looking at convergent solutions that have evolved independently (geographic isolation)
- Extreme environments are forming as a result of anthropogenic activities (salt ponds; acid drainage ponds from mining; anoxic zones, etc) → studying their colonization and the evolution of organisms colonizing them provides a unique opportunity for monitoring and experimentally investigating (convergent) evolution in extreme environments and for assessing impacts of human activities on the “makeup” of the biosphere.
- Studies of organisms inhabiting extreme environments have an
- exceptionally strong track record of high societal impact (e.g. PCR, cryopreservation, food security, biomedical engineering, etc are inspired/ informed by solutions evolved by nature in organisms from extreme environments)
- There are many extreme environments and they are increasing in number as a result of human activities, yet they are understudied because of accessibility/ logistics issues; → Technological progress now more than ever empowers us with the means to study organisms that have evolved solutions to thrive in extreme environments. Scaling up such studies to reach a new level can be expected to generate new inventions and applications/ solutions to societal challenges.

Authors

Andrea Bodnar	Gloucester Marine Genomics Institute
Cynthia Downs	SUNY ESF
Dietmar Kültz	University of California, Davis
Ally Nguyen	Whitehead Institute
Karen Sears	University of California, Los Angeles
Yu She	Purdue University

5. Convergence in action

What is your Big Idea?

A top-down PO to people networking opportunity for coalescence for helping to foster convergence research from NSF, NIH, DOE, etc.

Which *Branch of LIFE* does this address?

New Frontiers in Evolutionary Convergence, Collaboration across disciplines

What is required to pursue this?

Getting POs and directors across NSF, NIH, and other funding agencies to communicate and create teams of PIs who have converged on similar ideas but who may not know of each other. NSF could lead this effort in moving basic science initiatives into applied technologies by linking researchers converging on similar ideas. NSF program officers, and equivalent staff from NIH, DOE and other agencies can meet to discuss the soliciting of primary investigators who would make new teams/networks. Each agency could invest moderate level of funds to support these teams..

What are barriers to pursuing this now?

We are not sure if POs are allowed to divulge who is working on what types of projects (in terms of those asking for funding). But perhaps people with similar grants awarded from different agencies can be connected without revealing details of any particular proposal

What is the expected value and impact?

It would give the basic science research at NSF an applied edge, and the applied work at DOE, NIH, etc. may have an interest in the basic science research that feeds their work.

What are the training and workforce opportunities within this idea?

Workforce from TIP directorate, but also from Broader Impacts to Outreach.

What is the reasoning, justification, and/or supporting evidence behind this idea?

Such a simple idea: Connect convergent ideas on convergence across funding agencies. This PO-driven model can help create teams that drive the new bioeconomy.

Authors

Prosanta Chakrabarty	Louisiana State University
Tomokazu Kawashima	University of Kentucky
Jung-Youn Lee	University of Delaware
Grey Monroe	University of California, Davis
Tilottama Roy	Missouri Western State University

6. PhenBank: Creation of a national center for phenotypic information

What is your Big Idea?

To create a structured, centralized repository of phenotypic information, where data is searchable, linked, and routinely deposited.

Which *Branch of LIFE* does this address?

This recommendation intersects with all branches of LIFE.

What is required to pursue this?

Developing PhenBank database framework. Creation of PhenBank will require the integration of existing phenotypic databases and the development of new infrastructure for supporting the deposition of new data. Interdisciplinary groups representing a wide-range of stakeholders (current database managers, researchers, funding agencies) and data science experts are needed to set data standards, priorities, identify phenotypes to capture, and develop phenotype ontology to enable searching of data.

Computational infrastructure and support. Centralizing existing databases will require new computational infrastructure for long-term storage of data and the development of software for uploading and searching data. Integration of machine learning or AI tools would be valuable to help users search and analyze data. Similar tools could be developed to add existing data into PhenBank. A suggested future extension of PhenBank is the development of phenotyping projects run by PhenBank staff.

Community buy-in and support. PhenBank will require support from PIs, Universities, Museums, NGOs, current database managers, funding agencies, and other members of the scientific community.

What are barriers to pursuing this now?

Scope of potential phenotypes to include is huge. From morphology to gene expression to behavior to metabolism, the possible phenotype space to include is vast and presents significant challenges. How will phenotype be defined? Which categories or classes will be included? Possible solution: prioritize phenotypes in existing databases. What range of phenotypic data will be included (e.g. omics to morphometrics)? How might we incorporate quantitative data across many levels of biological organization? How might we deal with phenotypes that are not universal or transferable across systems? How would we “merge” phenotypes from radically different organisms?

Data will be messy, heterogenous, and of variable quality. Phenotype data currently exists in a variety of types (text, images, numerical, video, etc). Which data types will be supported? How could data types be simplified? For example, GenBank benefits from low dimensionality and simple format of data. How is data quality ensured? Possible solution: AI assisted screening and flagging. How are data and databases maintained over time? How will intellectual property rights or other ethical concerns be addressed?

Standard ontology is needed to ensure compatibility and searchability of databases. No standard ontology currently exists. A standard ontology will facilitate creation of data and database structure and will facilitate merging data across many levels and types of organisms. How might we build and create a structured phenotype ontology? Who will be involved in developing ontology? Possible solutions: category-specific ontology, kingdom-specific ontology, build on current phenotype terminology, use a nested structure, use machine-readable standard language. Perhaps machine learning could be used to begin the collation and standardization of phenotypic terms. How might we use ontologies to build/create standards?

The creation of PhenBank will be a significant undertaking. The mechanics of making PhenBank includes challenges beyond those already described above. How will different types of data be synthesized or integrated in one place? Possible solutions: collect current databases into one place, synthesize current databases into one, use AI to assist with database creation and management.
How would we establish data standards (data needs to be machine readable)? What data structure will be used? Possible solution: Build database structure from organism specific ontologies, map ontologies among organisms to enable meta-analyses. How will the historical contingencies of importing data from currently available databases be handled? How would existing databases, such as NCBI, be incorporated? Possible solution: connect through taxonomy ID. What tools would be included (e.g. media viewers as in the Macaulay Library at Cornell Lab of Ornithology)? How might we make PhenBank broad enough and easy enough for people to regularly use?

Long-term funding and maintenance is critical to ensure longevity and stability. Historically, self-sustaining models for databases usually do not work. Who will fund PhenBank? Who will establish PhenBank? Possible solution: change in government priorities, long term NCBI/GenBank-like funding model. Who will maintain PhenBank over time? Possible solutions: partner with journals, libraries, universities (data archive requirements of government funding).

Community buy-in. How might we convince a critical mass of users to make this viable? How might we deal with silo-ed disciplines and turf issues within academia? How do we incentivize researchers to contribute to the database? Possible solution: funding agencies require data archiving in a publicly available and accessible format/place; journals could require upload at publication. How do we incentivize building and contributing to PhenBank within the promotion and reward structure of academia? This type of database is essential but the work required is undervalued. Possible solution: PhenBank is its own entity (like NCBI). How might we leverage the motivation from other efforts? Possible solution: combine with ontology development, combine with human disease data compilation efforts.

What is the expected value and impact?

PhenBank is expected to promote basic, translational, and applied science by sharing and maintaining data for the scientific community, which will promote the bio-economy by incubating new biological paradigms, framework, or theory.

The accessory tools of the PhenBank can include visualization, meta data, and other resources as well. PhenBank has the potential to combine efforts with NCBI for joint queries or cross-referencing data for cross-disciplinary users.

Increased data availability and accessibility. This will lead to new resources for training students, data mining to discover new principles.

Creation of a centralized center for archiving data. This will support current data archiving requirements of funding agencies and publications.

A new resource will incentivize methods development, discoveries, new framework and theory (e.g. many of the things suggested by other working groups).

Development of this resource will increase the opportunities for meetings and interactions of interdisciplinary fields.

Catalyze discoveries by integrating directly (linking) to NCBI and vice versa.

Accelerate comparative work across different species in evolution.

Enable “Phylo-GWAS” (Phylogenetic - Genome Wide Association Study).

An invaluable resource for education and research, benefit the society at all different levels, from K-12 to Primary Undergrad Institutions, Minority serving institutions, to graduate students, postdoctoral trainees and PIs in research institutions, NGOs, government agencies, and industries.

What are the training and workforce opportunities within this idea?

The establishment of PhenBank as a new separate institute would add new jobs to the workforce. These would include jobs directing, managing, and creating the infrastructure of PhenBank, as well as jobs for providing user support. If PhenBank takes an active role in generating high-throughput phenotyping data or model development, this will lead to the creation of jobs for generating this data.

PhenBank has great potential to be a new tool for training students and the workforce. Users will be able to use publicly available databases to develop resources to teach scientific concepts at educational institutions or to increase the scientific literacy of the general public.

Lastly, increasing data accessibility has been shown to decrease barriers to doing science. Therefore the establishment of PhenBank is likely to provide new research and training opportunities for low resource institutions.

What is the reasoning, justification, and/or supporting evidence behind this idea?

Integrative comparative evolutionary methods will require multi-dimensional phenotypic data across a wide range of taxa. Many of the ideas proposed at the LIFE workshop would require such a resource. Accordingly, 72% of participants directly expressed support for the development of PhenBank or a similar resource.

While abundant phenotypic data currently exists, it is poorly integrated across taxa or databases, siloed, and difficult to search.

Curated phenotypic data is contained in specialized databases that were established to address specific needs. Examples include vertebrate morphology (VertNet) and bird multimedia data (Macaulay Library). However, data is rarely linked across phenotypes or taxa, and use of these data requires *a priori* knowledge of these databases.

Uncurated phenotypic data is contained in accessible but poorly searchable databases such as the Dryad Data repository (https://datadryad.org) or Figshare (www.figshare.com). Data in these databases is frequently stored in a wide variety of file formats and metadata describing the data can vary widely in quality. Researchers must typically download the data to their personal computers to view. Assembling different datasets requires time-consuming searching and manual curation. Additionally, papers published prior to the establishment of digital data archives frequently contain phenotypic data in tables or supplemental material, but these data are even more challenging to collate.

Phenotypic data is also routinely collected by researchers but remains inaccessible. Phenotypic data is frequently collected through mechanisms like curated undergraduate research experiences (CURE courses), undergraduate thesis projects, or ancillary data to other research projects (“the file drawer problem”). However, there are few ways to make this data available to the broader scientific community if the data is not included in a publication.

Data archiving is becoming increasingly mandated by funding agencies and journals. Establishing a centralized phenotype data repository would also provide a place for scientists to comply with these requirements while also providing a resource for other scientists.

Authors

Corien Bakermans	Pennsylvania State University- Altoona
Sara E. Miller	University of Missouri-St. Louis
Hongmei Li-Byarlay	Central State University, Wilberforce, OH

7. MC-4-G2P: Modeling frameworks for convergence from genotype-to-phenotype

Supporting comprehensive frameworks for mathematical, computational, and statistical modeling of convergent evolution across levels of the genotype-phenotype map

What is your Big Idea?

Convergent evolution may arise from distinct genetic bases, at different levels of genotype to phenotype map, and across phylogenetically distant taxa, making it challenging to model, and consequently to detect and characterize from empirical data. How do we take the wealth of diverse empirical data and rigorously evaluate hypotheses to integrate across taxa, across scales, across ecologies, and across data types (genomics, pheno-omics, ecological data) to find generalizable principles of convergent evolution?

We propose to make a concerted effort to develop the mathematical and computational tools needed to inspire and facilitate work on convergent evolution and other questions involving the genotype-phenotype map, generating explicit model-based expectations and teaching quantitative reasoning and tools. Modeling based upon theory can provide appropriate null models and expectations for the emergence of phenotypes from genotypic data consistent with different processes that lead to evolutionary convergence. Such an approach is generalizable across taxa and builds mechanistically across layers of biological organization. Yet, a comprehensive modeling framework for convergent evolution is currently missing in the field.

Theory and the analysis/statistical approaches that derive from the mechanistic approach are built on explicit assumptions and expectations of the data that are rigorous and testable. Theory can clarify discussion by explicitly testing verbal models and building a quantitative frameworks for comparing ideas. It can also provide initial answers to questions like, what is our null expectation? How much data is needed? Can we exclude some experiments a priori? Theory can also simplify analysis through approximations that speed analysis, new statistical tests for questions of interest, and by fitting of data to models to determine which is the best fit. Thus by promoting the development of theory and models of convergent evolution, we can accelerate the pace of discovery in this field and in the broader question of how genotype results in phenotype.

Which *Branch of LIFE* does this address?

This primarily impacts Tools and Tech (through new statistical approaches) and New Frontiers in Evolutionary Convergence (new hypotheses), but will also impact Non-model Organisms, and Collaboration Across Disciplines.

What is required to pursue this?

Resources:
Access to high performance computing is necessary.

Interdisciplinary:
Teams/projects combining expertise in physics, biochemistry, molecular biology, population genetics, evolution, applied mathematics, and computational statistics are necessary to successfully actualize proposed research.

Funding Structure:

- Explicit support for theoretical/statistical/modelling development in this field. Perhaps a NOSI. Coordination with DMS might be helpful to promote bringing new mathematics and statistics to convergent evoluion.
- A specific call for teams of an experimmentalist and theorist/modeller (broadly defined) to pursue this question. This would help ensure that the efforts of both synergize. Theorist understands the data possible, knowable; experimentalist gets guidance on design and analysis.
- Supplements/small grants to bring ***existing*** projects together into collaboration (across institutions). Funds to promote getting together, supporting cross fostering students among the labs, and a small amount of funds to pilot collaboratively inspired questions.

What are barriers to pursuing this now?

The field currently lacks a consistent way of thinking about interactions between processes at different scales. Theory can help guide the development of these ways of thinking.

Predicting the effects of genetic variation on various molecular phenotypes is not currently accessible. Research to develop these pieces is necessary.

Interdisciplinary research is hard to evenly incentivize and organize; this objective requires theorists and empiricists to agree on a set of parameters that are both useful for modeling and tractable to measure. Moreover, actually measuring them in the lab/field needs to be incentivized. Further, no one modelling approach is sufficient so collaboration among experts in different approaches to a problem will help find the best path forward. New math may be needed to simplify assumptions, reduce parameterization. This again demands collaborations among mathematicians, modellers, and experimentalists.

What is the expected value and impact?

The generation of this modeling framework enables rigorous scientific evaluation of phenotypic expectations from underlying processes with specific (and changeable assumptions). This will enable a much fuller harnessing of omic data to address a large number of discipline-specific scientific questions.

As noted above, the field lacks a consistent way of thinking about interactions among processes across different scales.

What are the training and workforce opportunities within this idea?

This is a highly interdisciplinary scientific project integrating expertise across a large number of STEM disciplines towards ultimate biological inference. WIth an appropriate framing, aspects of this could be appropriate for summer undergraduate REU projects or vertically integrated projects involving graduate students, postdocs, etc. Collaborations across disciplines and institutions are more accessible than ever.

What is the reasoning, justification, and/or supporting evidence behind this idea?

It is clear that our ability to link genotypic data (that exists in large amounts) to phenotypic data (both predicted and existing) is lacking. Enabling this synthesis will generate a level of understanding of phenotypic evolution that is not possible without it. The use of customizable null models in tracing the evolutionary processes through taxa can help distinguish the likelihood of convergent evolution from the range of outcomes that emerge through sequential random mutations.

Authors

Emily Dolson	Michigan State University
Davorka Gulisija	University of New Mexico
Corbin Jones	University of North Carolina
David Liberles	Temple University
Eduardo Rosa-Molinar	University of Kansas
Chad Westphal	Wabash College

8. Leveraging convergence to understand multi-trait evolution across biotic interactions

What is your Big Idea?

Convergent traits are often analyzed in isolation, either as one trait or for one species, and often traits shaped through abiotic selective forces, such as adaptations to extreme environments. However, interactions among species, such as in pollinator-plant or host-parasite interactions, are important drivers of convergent evolution. These interactions across trophic levels can simultaneously shape multiple traits as they covary under the selective forces created by species interactions. Repeated bouts of multi-trait evolution may be thought of as a trait syndrome, a characteristic suite of covarying traits. To fully understand the convergent multi-trait evolution of complex traits, we must examine how species interactions shape trait syndromes within and across species. The most explanatory studies will examine the evolution of these covarying traits in a phylogenetic and coevolutionary framework (whether mutualistic or antagonistic) and seek to understand the genetic underpinnings of these processes. Understanding multi-trait evolution in the context of species interactions has great translational potential in conservation, agriculture, human health, and wildlife disease management as convergent traits are understood in a predictive framework.

Which *Branch of LIFE* does this address?

Analyses of multi-trait convergence is an overarching effort that leverages the diversity of life, including model and non-model organisms, the development of techniques, and with the potential for training students and professionals in STEM. Branches of life addressed by the effort include:

- Leveraging Diversity
- Non-Model Organisms
- New Frontiers in Evolutionary Convergence
- Tools & Tech
- Training and Broadening Participation

What is required to pursue this?

- Funding the development of new machine learning and statistical techniques and phylogenetic comparative analysis methods
- Creating and supporting high-throughput tools to measure and quantify phenotypic traits across a variety of systems (e.g. plants, animals, prokaryotes)
- Long-term funding of database infrastructure and support. Develop and support nterface access and design. Enable linking of datasets through better integration of multiple databases (GenBank, GloBI)
- Ultimately, large-scale analyses across multiple traits in multiple taxa will be a years-long effort to work toward, requiring new data, approaches, and methods. Funding to support the collection of these data is critical to understanding convergent trait syndromes

What are barriers to pursuing this now?

- Data availability remains a barrier, especially in non-model organisms; many more phenological measurements are needed in a variety of taxa, both closely related and distantly related (to highlight convergence).
- Current coevolutionary analyses are limited in explanatory power, and these methods do not incorporate trait mapping.
- Ancestral state reconstructions are limited to one phylogeny and lose explanatory power across multi-state characters.
- Phylogenetic trait correlations are mainly limited to one pair of traits across one phylogeny.
- We need to define multi-trait conditions more completely.
- The means of distinguishing spurious correlations in traits requires refinement.
- It is challenging to incorporate environmental dependence and spatial heterogeneity in species interactions.

What is the expected value and impact?

This approach and the resulting studies would:

- Build and make accessible data for a large range of non-model organisms, potentially providing insights into understudied and under sampled systems.
- Link data across phenomics, genomics, and phylogenetics.
- Allow for more-informed organismal-based conservation.
- Reveal the traits and pathways that act to maintain interactions, from functional group conservation up to community-wide associations. This can be done in one side of the interaction (i.e. the host) or in both (i.e. host and parasite) using the same methodologies. This provides the potential ability to uncover “new” syndromes.
- Identify targets or commonalities of disease, disease susceptibility and resistance.
- Identify genetic mechanisms underlying convergent coevolutionary interactions (linkage, pleiotropy).
- Allow the assessment of agricultural biotic interactions (pollinators, herbivores, symbionts) and their impacts on yield.
- Provide novel phylogenetic software tools for analysis and new statistical techniques/frameworks that are broadly applicable across fields.
- Enable the prediction of the evolution of ecological interactions and responses to evolution.
- Enable new research directions, particularly in twospecies GxGxExE studies and multi-trait G-matrix theory.

What are the training and workforce opportunities within this idea?

- Data for this idea could be collected through course-based undergraduate research (CURES) and REUs in collaboration with natural history museums and collections. These authentic undergraduate research experiences would produce digital images and/or morphological trait measurements. Existing networks like BCEENET could be develop, host, and promote these resources.

- Engage senior and retired natural history experts, like curators, for both data curation and student mentoring, providing highly specialized organismal training for the next generation of scientists.

What is the reasoning, justification, and/or supporting evidence behind this idea?

- Existing efforts have identified important interactions, such as pollination syndrome, parasitic/host interactions, and covarying traits in selective breeding.
- There is a strong need in the scientific community to consider organismal evolution in the proper context.
- This idea meshes well with other LIFE efforts, such as PhenBank.
- Projects falling under the umbrella of this idea would allow for novel analyses across a widerange of scales (mates v/s ecosystems).
- This work would increase the likelihood of successful conservation of taxa as techniques used for one group are applied to another group that shares traits.
- Research in this area would greatly expand the potential for discovery given counter-examples to shared traits. Why do some species not match expectations, and what may we learn from that?

Authors

Barbara Sharanowski	University of Central Florida, Orlando, FL
Jen Kovacs	Agnes Scott College, Decatur, GA
Zinan Wang	Michigan State University, East Lansing, MI
Nora Mitchell	University of Wisconsin, Eau Claire, WI
Randall Boone	Colorado State University, Fort Collins, CO

9. PanEDGE: Building the functional tree of life

What is your Big Idea?

National center(s) to enable building a functional tree of life, by delivering “model organism” packages for key nodes at the ToL

Which *Branch of LIFE* does this address?

Non-model organisms, tools & tech, collaboration across disciplines, training & broadening participation

What is required to pursue this?

Funding for core facilities dedicated to development of transformation tools for non-model organisms.

What are barriers to pursuing this now?

Currently, individual labs or small collaborations are attempting to develop functional genomics tools for emerging model organisms. However, development of stable transformation approaches in non-model organisms is not appropriate for graduate student or postdoc projects because of the high risk and low rewards (even though these tools often lead to high impact papers). In addition, the labs typically have to build from scratch and suffer from similar trial-and-error journeys. This process can be made more efficient and cost-effective.

What is the expected value and impact?

- Provide a “Kick start” that gets researchers (especially early career) to jump onto the emerging model wagon.
- The resulting functional information for neglected branches of the tree of life improves the ability to reconstruct the trajectory of gene function (functional evolution of genes)
- Allow the researchers to focus on the questions
- Ripple effects: Facilitates the validation of neighboring systems to the one developed.

What are the training and workforce opportunities within this idea?

Training of grads, postdocs and staff in transgenic techniques at the facility, bringing the expertise and protocols back to their home labs.

What is the reasoning, justification, and/or supporting evidence behind this idea?

With the surge in research connecting genotypes and phenotypes, it has become increasingly important to have model organisms to understand and validate gene function. Model systems are few and far between across the tree of life and the absence of funding and expertise hampers their further development in more diverse groups. Having a centralized institute to coordinate and facilitate development of transformation techniques, gene atlases and other resources will kickstart the next stage of innovation and discovery.

Authors

Fay-Wei Li	Boyce Thompson Institute
Jeff Mower	University of Nebraska
Subir Shakya	Harvard University
Verónica S. Di Stilio	University of Washington (Seattle)
Yanni Chen	University of Notre Dame

10. Towards integrated analyses of multiscale, multi-omic, & multivariate data

What is your Big Idea?

Leveraging biology to solve societal challenges requires integrating across biological scales. The study of convergent evolution provides one thematic mechanism that ties together common questions across biological scales. However, many questions about convergence (including its frequency, underlying basis, drivers) require multiscale, multi-omic and multivariate data as input to models that can jointly capture evolution, phylogeny, and traits. Currently, the community lacks the tools to perform comparative analyses within integrated phylogenetic models that simultaneously assess data from multiple scales and data types. Two consequences of this big gap are that (i) the richness of many datasets cannot be fully utilized, and (ii) incomplete and inaccurate predictions are being made.

To address this we propose funding for directed research (including via a dedicated RCN, and funded training mechanism) to develop phylogenetically-aware data-integration models, and to coordinate amongst the community engaged in model development and application (via workshops and stakeholder/end-user consultation and collaboration). In addition, we propose the development of an institutionalized web portal and suite of computational resources for hosting, sharing, and disseminating models and analyses, for the use of the broader biological community.

Which *Branch of LIFE* does this address?

- Tools and Tech
- Leveraging Life’s Diversity
- Non-model organisms

What is required to pursue this?

- **First step.** Gathering and defining specific use-cases, examples, and requirements is the first step in our proposed work. To accomplish this information gathering phase, we propose a scoping meeting that brings researchers with datasets and specific questions together with computational biologists/software developers/theorists. Attendees with data should represent a mix of established, model systems and emerging, quasi-model systems and on the computational/modeling-side researchers should include those with expertise in phylogenetics/comparative methods and scientific computing. Specific outcomes include the identification and development of benchmark datasets, requirements for ensuring model usability, and key research questions.

- **Research infrastructure**. We propose a physical academic research center that houses a centralized scientific leadership team that will coordinate funding, recruit researchers and technical staff, and organize the creation and deployment of dedicated training and software resources. Use community accessible computational power (i.e. CIPRES-like) as a resource to build capacity for running and analyzing comparative models.

- **Interdisciplinary collaborations**. The scientific leadership team will create a multidisciplinary collaboration of researchers across genomics, phenomics, computer science, software development, and computational phylogenetics and evolutionary theory. The collaboration will catalyze data generation and collection, method development and computational implementation, and software deployment.
- **Centralized data and software repository and analysis portal**. For example, the portal will house exemplar datasets and analyses and allow researchers to discover and use appropriate analytical pipelines and upload data and run analyses.
- **Training and outreach opportunities**. Described below.
- **Scientific innovations**. Computational methods will be developed for co-estimating and inferring phylogenetic relationships and the evolution of multiple interacting phenotypic traits using large scale and high-dimensional data.

What are barriers to pursuing this now?

There are three barriers that we identify as important:

- **High entry-level requirements for model development**: Integrating multi-scale data and models needs interdisciplinary training in mathematics, statistics, molecular, and organismal biology.
- **Lack of computational power and access to high-performance computing**: Access to shared and large scale computational analysis to level the field across institutions, PIs, and interdisciplinary groups interested in developing integrative comparative approaches. In addition, the creation of a core database of models, methods, with help targeting end-users aiming to integrate multi-scale data.
- **DEI and the size of comparative methods community:** Diversifying and encouraging new scientists from all backgrounds to enter the field of phylogenetic comparative methods. Currently, there is a small community of comparative methods developers that are in high demand, and make disaggregated individual efforts of outreach via in-person or online workshops nationally and internationally. A cohesive center, training, and teaching strategies, to invite more scientists at all career stages to develop their own comparative approaches and become knowledgeable users are critically needed.

What is the expected value and impact?

- New mathematical and software methods will be developed for analyzing high dimensional and large-scale phenotypic data in a phylogenetic context and deployed via an accessible and flexible cloud platform.

- Community knowledge: The platform will be able to immediately analyze existing data available in national genomic and phenomic databases.
- Analytical methods will be general enough to discover convergence in many phylogenetically-structured, multiscale, and multi-omic datasets.
- Patterns in convergence evolution can be used in a variety of areas important to the bio-economy including improving methods for drug discovery, improving diagnostics for complex/rare diseases, identifying species of conservation concern, and detecting novel pathogens.

What are the training and workforce opportunities within this idea?

The proposal would build a number of opportunities including:

- **Research Coordination Networks:** Encouraging community between methods development labs and experimental/empirical labs and cross-mentoring graduate students
- Targeted funding for postdoctoral/grad support/training. For example, a new track for PRFB on complex data integration in phylogenetic context.
- Support for workshops and long-term research exchanges between computational experts and end-users
- Kaggle-style competitions with high-dimensional data targeted towards HS and undergrad students outside of biology majors.

What is the reasoning, justification, and/or supporting evidence behind this idea?

Prior to the last 10 years, many of multi-scale, multi-omic, and multivariate datasets did not exist, especially for non-model organisms. Conversely, many comparative and phylogenetic methods were developed for relatively low-dimensional, macroscale, datasets.

Now—with the availability of new computational power, emerging algorithms, tools and approaches, and substantial existing comparative data that can be immediately and directly leveraged—is ideal for developing a directed effort to develop, support, and disseminate integrated phylogenetic comparative approaches for complex multi-dimensional data.

Authors

Leonie Moyle	Indiana University
Jeremy Van Cleve	University of Kentucky
Sam Scarpino	Northeastern University
Guin Wogan	Oklahoma State University
Rosana Zenil-Ferguson	University of Kentucky
Jeremy Brown	Louisiana State University

11. Hopeful monsters: Bringing ancestral inferences to life to test evolutionary trajectories

What is your Big Idea?

A new way to explicitly test macroevolutionary hypotheses!

- **Synthesize evolutionary intermediates** that we have inferred as probable ancestral states to test alternative paths through inferred **fitness landscapes**, focusing on the organism level.
    - Ancestral proteins can be reconstructed using experimental lab techniques
    - More complex ancestral phenotypes can be reconstructed using computational models.
- This approach offers the opportunity to study systems where we have mechanistic understanding of traits (often, morphology), but ALSO gives framework to study less understood traits (behavior, life history, physiology), using in-silico resurrections of ancestral phenotypes in adaptive/performance landscapes.
- A DEB focused call similar to EDGE but centered on evolutionary question(s)
    - **Track 1:** Proposals focused on traits for which we don't understand mechanistically well, taxa not yet transformable
    - **Track 2:** Proposals focused on traits for which we understand underlying mechanism. Researchers are ready to synthesize inferred ancestral intermediates

Which *Branch of LIFE* does this address?

New Frontiers in Evolutionary Convergence**;** Non-Model Organisms**;** Tools & Tech**;** Collaboration Across Disciplines

What is required to pursue this?

Genetic engineering in non-model systems

- Potentially biosafety or ethics expertise (could require an extra statement, similar to safe and inclusive fieldwork)

Phylogenetic ancestral state reconstruction of complex traits

Operationalized understanding of complex phenotypes

- Molecular/cellular basis of the trait
- Understanding of performance relevance of the phenotype (performance as partial predictor for fitness)

Incorporation of computational modeling, in-silico testing of phenotypes

What are barriers to pursuing this now?

- Detailed knowledge of genotype-phenotype map and phenotype-performance links

- Widespread comparative knowledge of traits across levels for ancestral reconstruction
    - Uncertainty in ancestral reconstruction (but doesn’t have to be an issue because we can model a distribution and even possibly experimentally recreate multiple phenotypes)
- Lack of general transgenic methods for non-model species (although in-silico modeling addresses this)

What is the expected value and impact?

Derive generalizable principles about how developmental constraints and biases shape evolutionary trajectories. Move from solely looking backwards (in time) to a forward-looking paradigm.

Understanding the evolutionary trajectory could help make the engineering of complex traits (like C_4) easier, for societal benefit

Understanding how phenotypes have responded to past selection could be potentially relevant in the future for applied phenotypes

What are the training and workforce opportunities within this idea?

Educational opportunities include training students in interdisciplinary science, including organismal evolution, genomic tools and transgenics, computational modeling, and ethics.

What is the reasoning, justification, and/or supporting evidence behind this idea?

Emerging ability to transform a wide variety of organisms, combined with better phylogenetic frameworks in which to estimate ancestral states, and high(er) throughput methods for characterizing phenotypes. Excitingly, we are approaching the ability to truly test macroevolutionary hypotheses!

Authors

Jo Wolfe	Harvard University
Erika Edwards	Yale University
Michele Johnson	Trinity University
Kim Hoke	Colorado State University
Matt Stata	Michigan State University
Cecile Ane	University of Wisconsin-Madison
Mark Margres	University of Southern Florida

12. A two-tiered funding mechanism to promote interdisciplinary collaborations

What is your Big Idea?

Initiate a two-tier funding program to exploit existing -omics or other “big” data to understand the mechanisms and consequences of convergent evolution.

Groups will form (2-3 labs) that bring expertise and interest in divergent areas (e.g. computational biology and experimentalists). They will apply in Tier 1 to exploit the existing datasets to generate novel testable hypotheses about convergent evolution. If successful, as measured by the generation of clearly defined and testable hypotheses, the team will move forward to increased funding to test those hypotheses, either through the generation of experimental data and/or the generation and analysis of additional “big” datasets. Many Tier 1 proposals will not move on to Tier 2, but those that do have high-reward impact.

Tier 1: Exploration of existing data in a phylogenetic context and develop novel intersections that inform phenotypic evolution.

Tier 2: Experimental tests of hypotheses emerging from Tier 1, and/or expanded application of methods developed in Tier 1 to additional taxa or data types.

Which *Branch of LIFE* does this address?

The proposed mechanism directly addresses all of these branches.

Program itself:
Focuses on collaborations across disciplines, training, and leveraging diversity.

Outcomes of grants funded by this proposal:
Has strong potential to lead to new tools and tech, advances in non-model organisms, and new frontiers in evolutionary convergence.

What is required to pursue this?

- Need small collaborative groups to form. These likely exist informally (germinated), but have not been nurtured. Could also utilize/synergize with the proposal from Group 8 about building a social network to facilitate collaborations.
- Need NSF buy-in for developing this funding mechanism.
- Need a way to building review panels that are appropriate for this type of mechanism
- We considered Nascent, Eager, EDGE, intBIO granting mechanisms as alternatives. None of these cover these specific needs or have the same potential to provide the impacts described below.

What are barriers to pursuing this now?

- Current funding mechanisms require more preliminary data than can be generated to shift fields. Funding at Tier 1 would allow for dedicated generation of big, ambitious hypotheses for Tier 2.
- Without preliminary data, proposals to build new tools (non-model organisms, new data intersection approaches) can not be justified.
- Current review panels can be over-focused (or too field-focused)

What is the expected value and impact?

- Tier 1 awards would be small enough to launch many high-risk, potentially high-reward collaborations.
- These collaborations would bring novel insights into convergent evolution that could not come from a single group alone.
- Naturally supports the training of future “multilingual” scientists that are so needed right now by bridging computational biologists with experimentalists
- Would drive methods for intersecting different data types.
- Leverage existing data and existing expertise in novel intersecting ways.
- Would feed-forward to generate more of this type of thinking.
- Existing programs, like BII and IntBIO, require mature collaborations. This LIFE mechanism would support nucleation of new ones.

What are the training and workforce opportunities within this idea?

Interdisciplinary training in bioinformatics (computational biology) and experimental biology.

Team science helps facilitate a network mentoring model so that trainees are exposed to multiple disciplines and perspectives, transcending language barriers between subdisciplines.

Training in quantitative, mathematical, statistical, and computational skills in a biological context.

Training in practical skills that are transferable to industry will attract a broader diversity of students and directly impact the bioeconomy.

What is the reasoning, justification, and/or supporting evidence behind this idea?

Dramatic increase of -omics data - there is no shortage of potentially interesting patterns of convergence, but little time/resources to follow through. The emergence of new technologies and tools offer a unique opportunity to tackle important problems in convergent evolution.

People analyzing these data often have insights, but not a mechanism to pursue this in collaboration. People using these data often have ideas of how one might look at those data in different ways, but do not have the ability to do this. This mechanism would nurture those to the next level (Tier 1: rigorously pursue these ideas to generate testable hypotheses) and an even higher level (Tier 2: test and refine these hypotheses to yield novel discoveries around convergent evolution).

Tier 1 would enable not only the discovery of potentially interesting questions of convergence, but could prioritize those that are experimentally feasible.

This relies on investigator-led questions/systems about convergent evolution, so is agnostic to system, approach, etc., but rather encourages big high-risk thinking in small collaborating groups. Allows the goals and group to be nimble in Tier 1 to maximize impact and feasibility of Tier 2. Fundamentally different and complementary to larger NSF-funded initiatives.

The proposed program can accommodate a wide range of organisms and phenotypes. Examples include:

Explaining already appreciated convergent phenotypes:

- thorns in angiosperms (identify novel transcriptional feedback loops)
- self-fertility in nematodes (identify novel intermolecular interactions impacting sex determination)

Discovering previously unrecognized convergence:

- using gene family evolution across mammals to predict novel host antiviral defense
- utilizing proteome-metabolome linkages to discover convergent (or re-recruited) pathways

Authors

Matt Daugherty	University of California, San Diego
Omer Gokcumen	SUNY Buffalo
Maricris Mayes	University of Massachusetts, Dartmouth
Eric Haag	University of Maryland, College Park
Paulyn Cartwright	University of Kansas
Vivian Irish	Yale University

13. Informing convergent evolution of chromatin & genome organization

What is your Big Idea?

Do the molecular features that drive 3D chromatin organization have conserved or diverged genome functions (e.g., gene expression regulation, recombination, repair)?

The way in which genomes are packaged within cells is a critical component in converting a genotype into a phenotype, which has important impacts on how organisms evolve. However, our understanding of how genome packaging—including 3D genome organization and chromatin—evolves is largely unresolved. We now have the capacity to address this important, unresolved question because high throughput methods are capable of measuring chromatin conformation, 3D organization, and other genomic and epigenomic features in any eukaryote from which we can obtain fresh biological material. This presents us with a tremendous opportunity to study chromatin and 3D genome organization as a nexus point for a wide array of evolutionary processes including convergence. For example, recent studies have begun to examine how gene expression, 3D organization, and DNA methylation evolve on phylogenies to affect organismal phenotypes. Despite these advances, we lack sufficient sampling of taxa and phenotypes to gain a clear understanding of how genome packaging and phenotypes affect each other and their evolution. We propose to map information on chromatin structure (epigenetic marks, 3D organization) onto phylogenies to use the replicated instances of genome structure to determine convergent molecular mechanisms of gene regulation.

Which *Branch of LIFE* does this address?

New Frontiers in Evolutionary Convergence, Non-Model Organisms, Tools & Tech, Collaboration Across Disciplines

What is required to pursue this?

We need to broaden our ability to measure and analyze different features of chromatin organization across many different organisms. A major hurdle is forming collaborations among molecular biologists, bioinformaticians and evolutionary biologists in order to perform comparative phylogenetics. Importantly, we need to develop a statistical (phylogenetically informed) method that allows us to compare these variables across datasets. For example, how do we define a single topologically associated domain (TAD) across different species?

What are barriers to pursuing this now?

- Need for interdisciplinary collaboration for obtaining biological samples, collecting data in a broader range of organisms, and developing statistical approaches to compare across a phylogeny.
- Potential unanticipated challenges in applying unified bioinformatic methods across taxa/convergent traits, in addition to existing technical challenges in sample collection.

What is the expected value and impact?

The statistical approaches developed will be broadly useful for other evolutionary biologists. The projects will bridge molecular, structural, and evolutionary biology. The molecular mechanisms could be used for tech-development (e.g., drug discovery, diagnostics)

Other potential impacts: a better understanding of variation in chromatin / gene regulation among different cell types and how such variation was important for the evolution of multicellularity, and now cell-cell cooperation (cancer).

Chromosome visualization tools including those that integrate into multi-omics workflows/pipelines/toolkits.

What are the training and workforce opportunities within this idea?

This research is very collaborative, needing evolutionary biologists, cell-molecular biologists, bioinformaticians, and statisticians. This interdisciplinary environment is ideal for graduate student training, promising to strengthen data analysis techniques and skills in our emerging workforce.

Could incentivize new investigations/collaborations among institutions, education and development training, to limit bias towards funding research groups that are already well-funded.

What is the reasoning, justification, and/or supporting evidence behind this idea?

With the recent emergence of new technologies (such as Hi-C and ATAC-seq), that allow us to measure multiple features of chromatin structure, we now have the opportunity to study how genomes (and epigenomes) have evolved from the perspective of 3D genome organization. Advances in artificial intelligence methods offer new powerful ways to create and test models in evolutionary and structural biology. However, we don’t have the statistical methodology to perform this analysis with even existing data. And we lack the collaborative teams to perform this research.

A few references/studies that are relevant:

Comparative functional genomics is hard and needs to incorporate phylogeny:
https://www.pnas.org/doi/abs/10.1073/pnas.1707515115

OU model for comparative analysis of RNA-seq data:
https://academic.oup.com/mbe/article/31/1/201/1049642

New method for comparative analysis of gene expression:
https://academic.oup.com/mbe/article/40/5/msad106/7157541

A TAD Skeptic: Is 3D Genome Topology Conserved?
https://www.cell.com/action/showPdf?pii=S0168-9525%2820%2930298-5

A comparative analysis of chromatin accessibility in cattle, pig, and mouse tissues https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-020-07078-9

DNA methylation networks underlying mammalian traits
https://www.science.org/doi/10.1126/science.abq5693

Universal DNA methylation age across mammalian tissues
https://www.nature.com/articles/s43587-023-00462-6

The potential role of DNA methylation in chromatin accessibility (and the “maintenance of heterochromatin inaccessibility”)
https://www.pnas.org/doi/full/10.1073/pnas.2023347118

Authors

Coral Zhou	University of California - Berkeley
Rich Meisel	University of Houston
Mark Margres	University of South Florida
Heather Maughan	Ronin Institute
David Alvarez-Ponce	University of Nevada, Reno
Frederic Chain	University of Massachusetts - Lowell
Suzanne McGaugh	University of Minnesota

14. Evolution of biotic interactions across scales

What is your Big Idea?

We seek the establishment of funding mechanisms to support the development of research clusters or a center/institute to understand contemporary ecological interactions through a deep-time reconstruction approach that will elucidate the mechanistic basis of species trait evolution. By studying these interactions across levels of biological organization, we will be able to make strong inferences about the constraints and repeatability of coevolutionary processes in a wide range of taxa. The approach requires knowledge concerning the molecular basis of interactions between species, which leverages phylogenetic information to reconstruct ancestral genes associated with focal traits. Using genome editing technology, we will not only be able to characterize the phenotypes of individuals, but the quantitative interactions between species, using behavioral assays, experimental evolution, and novel modeling platforms that will provide new theories on the evolutionary ecology of species interactions. In addition to integrating historical information into contemporary views on species interactions, a number of practical applications will emerge from this platform with potential to advance areas of novel drug discovery, climate-adapted transgenic crops, and microbiome-mediated host nutrition.

Which *Branch of LIFE* does this address?

- New frontiers convergence: an explicitly historical experimental approach across deep time scales to understand mechanisms of evolution of biotic interactions
- Contribute to overcoming challenges associated with non-model organisms
- Collaboration and communication among disciplines
- Leveraging life’s diversity for translational, applied, and conservation work
- Fostering the next generation of scientists: training and broadening participation

What is required to pursue this?

- **Stakeholders**: The proposed research should engage a broad range of researchers and the public, who are interested in plant-insect interactions, microbial interactions (mutualistic and pathogenic), as well as innovations in industrial sectors (agriculture, pharmaceuticals, etc.)
- **Resources**: Funding sources that encourage riskier and interdisciplinary initiatives hope for transformative and interdisciplinary pay-off.
- **Disciplines**: Ecology, phylogenetics, biochemistry, behavior, physiology, geneticists, and structural biology.
- **Timeline**: 10-20 years

What are barriers to pursuing this now?

Statistical uncertainty. Ancestral sequence estimation and functional analysis could be a concern. However, improved computational and experimental tools to perform and evaluate ancestral genotype reconstruction on a pathway or even genome-wide scale should enable robust conclusions. Further improvements will arise once more complete omics data are available for focal study groups.

Biochemical uncertainty. In some systems, there may be no discernible activity of recombinant proteins owing to unknown ligands and substrate binding, or perhaps the evolutionary degradation of conserved function. The approach therefore should be initiated with carefully chosen biological systems where there is sufficient information about biochemistry of mechanistic traits with identifiable gene-by-gene interactions.

Community resources. Success will need to be facilitated by well-informed databases centered on systems and networks of diverse collaborators.

What is the expected value and impact?

We expect that the research done under this umbrella will lead to both new insights into the fundamental aspects of evolution, ecology, molecular and structural biology, biochemistry and other fields of basic science, as well as applications in agriculture, health, industry, *etc.*, as outlined below.

Potential applications:

- Synthesis of pharmaceutical compounds. We envision, for example, that the resurrection of some ancient plant/microbial enzymes could lead to the development of novel efficient synthesis of existing compounds with pharmaceutical applications or the development of entirely new compounds.
- New transgenic crops and their symbionts with desired properties. Resurrection of extinct proteins involved in historical aspects of biotic interactions could lead to the identification of molecules that may have had important roles in the past but do not exist today. These could confer valuable characteristics to modern-day organisms even though they were lost during evolution.
- Engineering of evolutionarily stable host microbiome consortia with beneficial properties for the host. Host microbiome engineering will become reality soon, and the work under this call will be necessary for successful engineering of host microbiomes that are evolutionarily stable (i.e., do not easily evolve into pathogenic states or go extinct) or those that interact in desired ways with each other and with the host.

Impact for basic science:

- **Eco-evolutionary integrations**. Our understanding of eco-evolutionary dynamics is currently in its infancy, and we believe this work will make major advances in this direction, with the development of new theory and experimental approaches for reconstructing ancient ecological networks and predicting how ecological networks will evolve in the future. In particular, we will better understand which phenotypes and interactions are a result of “accidental” historical contingencies and what is an almost inevitable outcome of evolution by natural selection.
- **Molecular and structural biology**. We envision that this work could lead us to a better understanding of how protein structures and their interactions are what they are. We might discover whole classes of new protein structures and molecular interactions.

- **Biochemistry**. This work will likely lead to the development of new biosynthetic pathways and to a better understanding of constraints that enzyme structures impose on the current pathways.

What are the training and workforce opportunities within this idea?

- The proposed work will lead to interdisciplinary projects that provide unique integrative training opportunities involving organismal biology, synthetic biology, systems biology, evolution biology, informatics, and mathematics.
- Seminar series on the interdisciplinary fields and topics similar to SynBYSS (Nature Chemical Biology vol. 18, page 353 (2022)) will provide free global education for researchers and the general public.

What is the reasoning, justification, and/or supporting evidence behind this idea?

- While much progress has been made towards understanding how organisms adapt to abiotic conditions via genome evolution, the degree and extent to which biotic interactions have impacted the origins of novel protein functions remains obscure. This is paradoxical since modern-day community ecology studies indicate that ecological interactions are important in driving evolution change.
- Modern-day interactions amongst organisms are mediated by proteins and their products. While the modern-day interactions may be dissected, understanding how the complex interactions were built is often lacking. For instance, protein-ligand interactions as discerned from x-ray crystallography, can indicate modern-day structural elements important for binding. Yet, historically, different residues may have been involved with the origins of novel binding interactions upon which the modern-day interactions were built. Therefore, we seek the historical context upon which modern-day proteins evolved to understand how interactions came to be and can be engineered.
- There is an increasing realization of the importance of ecological communities for human health, agriculture and other human endeavors. But we have a rudimentary understanding of how these communities came to be and why they are structured as they are. To what extent are these communities frozen evolutionary accidents versus products of natural selection, migration, etc. This understanding will form the foundation for future applications in synthetic assembly of communities with desired functions that are robust with respect to mutational perturbations and ecological invasions.

Authors

Matt Anderson	University of Wisconsin-Madison
Todd Barkman	Western Michigan University
Alexandra Deaconescu	Brown University
Bin He	University of Iowa
Sergey Kryazhimskiy	University of California, San Diego
Jay T. Lennon	Indiana University
Tae Seok Moon	Engineering Biology Research Consortium

15. Developing a holistic genotype-phenotype model

What is your Big Idea?

Phenotypes are dependent on the genotype of an individual, their environment, and the interactions between different parts of the genotype with each other, the environment, and is reflected in the epigenotype. Thus, a single genotype is associated not with a single phenotype, but rather associated with a phenotype space of possibilities. In this call-to-action, we will define the phenotype space for select species, develop new genotype-phenotype models, and then test these models in a range of species showing convergent phenotypes from across the tree of life.

Which *Branch of LIFE* does this address?

This call to action targets the following branches of LIFE:

- Tools & Tech. We will develop novel mathematical models linking genotypes to their phenotypic space, which will allow us to predict phenotypes under specific conditions. Also, devising/adjusting genomic approaches for application in non-model organisms.
- Non-model organisms. While initially models will be developed using model organisms, testing of the models will be carried out in non-model systems.
- Collaborations Across Disciplines. To achieve the goals of this project, extensive phenotypic data, genomics and epigenomics data will need to be collected, and the data integrated through machine learning and AI. Thus, this project engages organismal biologists from a range of species, genomics and epigenomics specialists, as well as computational biologists /machine learning specialists.
- New Frontiers in Evolutionary Convergence. This project deliberately uses convergent phenotypes as train and test cases for the models developed to predict phenotypes.
- Training and Broadening Participation. This project entails a wide range of training opportunities for interdisciplinary training working across disciplines. As this project requires the collection of large amounts of phenotypic data, there are many opportunities for individuals at PUIs and other institutions with fewer resources to contribute by engaging undergraduates for phenotyping and sending samples to core facilities.

What is required to pursue this?

This effort requires a coordinated funding call. This might include funding:

- A “coordination center” that would define common approaches/QC/ensure communication across groups
- A possibility of centralized “cores” that would enable unification (and standardization) of data generation, curation, and analyses. This might include sequencing centers that would generate consistent libraries.
- Fund computational and statistical projects aimed at generating models and interacting with experimental groups.

- Fund specific projects from individual groups for organismal experiments, leveraging existing knockout libraries and connect with pre-defined phenotypes (e.g., temperature, salt resistance - broadly applicable across species and has significant applications)
    - Phase 1 - focus on model organisms with set tools already created (e.g. knockout lines, genomes)
    - Phase 2 - opportunity to propose non-model organisms to apply and ground truth the novel methods. Devise new tools

What are barriers to pursuing this now?

Currently, the following barriers prevent solving this problem right now:

- Mathematical/Statistical methods: Significant mathematical/statistical tool development is needed to link individual genotypes to a phenotypic landscape that depends on environment, epigenetics, and interactions.
- Data collection and development of AI: Large collection of data (genotypes, expression data, epigenome profiles, phenotypes) and development of AI that integrates data types collected and utilize training datasets have not been developed.
- Lack of coordination: a mechanism that allows for the coordination of an effort like this that will require dozens of labs and allows for the involvement of individuals from diverse backgrounds and institutions.
- Lack of core facilities (e.g. sequencing centers) to generate and analyze the data.
- Funding (lots)

What is the expected value and impact?

- New tools to predict genotypes or phenotypes.
- Robust G-P “maps” for key phenotypes relevant to environmental challenges (climate and salt tolerance) of value to society and which can be utilized in applied contexts (agriculture, conservation management etc).
- New GWAS methods that can incorporate phenotypic plasticity and GxG effects.
- New AI tools to identify/predict convergence.
- Will facilitate genetic engineering efforts by increasing ability to predict phenotypic effects of edits.

What are the training and workforce opportunities within this idea?

- Collaboration across disciplines
- Coordinating center
- Increase the number of individuals trained in computational methods
- Training and education at the interface of experimental science and quantitative approaches (mathematics, statistics, machine learning)
- Community science (e.g., video games)
- Bring genotype-phenotype studies to the classroom using simple models

What is the reasoning, justification, and/or supporting evidence behind this idea?

- Leverages the natural experiment that is convergent evolution.
- Similar approaches are successfully being used in cancer biology to link tumor forming genes in mice to humans.
- It is widely understood that the concept of a one-to-one genotype-phenotype map is too simplistic and needs to be revised.
- Given the challenges imposed by a changing climate, predicting phenotypes in diverse environments from genotype is becoming an important problem to solve.
- Tools for large-scale phenotyping are becoming more available.
- Costs for epigenome and genome assays are decreasing and more accessible, making this type of large-scale effort possible.
- NSF EDGE program has been successful in addressing G-P relationships but coordinated and standardized efforts are necessary to make progress in addressing additional complexities (e.g., environment) in this sphere.

Authors

Diana Rennison	University of California San Diego
Megan Dennis	University of California, Davis
Nicole Riddle	University of Alabama at Birmingham
Maribet Gamboa	Universidad Catolica de la Santisima Concepcion
Javier Arsuaga	University of California, Davis

16. Data integration for convergence research using a model-clade approach

What is your Big Idea?

Use a model-clade approach* to generate omics data and database resources to enable convergence research. Community building is a required step for this idea so that researchers with different expertise can create a team to focus on particular clades. This approach will help level the playing field and enable and inspire convergence research for non-model organisms in a phylogenetic framework.

*Traditional model species are distantly related (e.g., Drosophila, Mus musculus, C. elegans) and do not represent convergent phenotypes. In order to spur advances through convergent evolution, we propose resources to support research on clades (groups of related species) encompassing multiple independent origins of the same convergent phenotype. Example emerging model clades in convergent evolution research include the genera Heliconius, Mimulus, and Peromyscus, and the family Gasterosteidae.

Which *Branch of LIFE* does this address?

Non-Model Organisms, Tools & Tech, Collaboration Across Disciplines

What is required to pursue this?

- Support for the study of convergence across evolutionary scales (macro, micro, and between populations).
- Capacity-building for predictive modeling for additional organisms.
- Development of metadata repositories to allow discovery of existing data that could be repurposed.
    - Also, evaluation of prior data and resources to determine which need reassessment (e.g. with newer tools).
- Support for working groups to review and evaluate current resources, develop general data, framework standards and practices, data interoperability.
    - Academic societies could convene these, as supported by NSF.
- Dedicated effort to learn from previous standards building and/or enforcing efforts (e.g. genome annotation working groups, iDigBio, CyVerse, NEON management).
    - This should include both successful and less-successful examples.
- Support for technology transfer between existing NSF-funded programs (like EDGE and especially CyVerse) and the broader scientific community.
- Incentives for community engagement efforts (both scientific and general public) (e.g. curation, tool support, software etc.) both in assessing NSF proposals and by universities.
- Ongoing commitment by NSF to fund “enabling” data collection through regular panels that can be used to discover novel examples of convergence e.g. phylogenies, collections, specimen digitization, genome assemblies, trait/behavior/ecological observations.

What are barriers to pursuing this now?

- **Complexity and Scale:** Building the resource to cover all organisms, especially less-studied ones, is a major barrier, in terms of resources and expertise. We therefore propose model clades to address the problem. Inferences could be generated for additional organisms based on output from model clades or taxonomic groups.
- **Interdisciplinary Collaboration:** Interdisciplinary collaboration is crucial, but the slow pace of interdisciplinary research is hindered by disciplinary expertise. Addressing this would require building a common resource to facilitate interdisciplinary training, and to foster cross-disciplinary synergy.
- **Sustainability:** Maintaining databases and tools long-term is a big challenge, therefore significant and sustainable funding and models will be required.
- **Data Relevance and Proofing:** Collecting information that may not be immediately relevant but could be valuable in the future poses a challenge.
- **Community Buy-In:** Model organisms usually have large communities; achieving buy-in from the broader scientific community for a specific non-model organism could be difficult. Incorporating multiple non-model organisms would therefore build a larger community through clade models.

What is the expected value and impact?

- This endeavor brings convergent evolution - nature’s repeated experiment across millennia - to the front by explicitly focusing on a **model-clade approach**: This will result in unprecedented understanding of how genomes, transcriptomes, and epigenomes have independently evolved to result in similar phenotypes.
- This will also result in community building around a particular clade that has a community-level buy–in.
- This will generate baseline essential omics data (platinum-grade genome with chromosome length assembly, pre-defined tissue-specific transcriptomes, chromatin-regulatory data etc.) taking advantage of the available resources across diverse species in selected model clades. The resulting platform would enable large-scale analysis by diverse researchers.
- Significant potential impacts because of the predictive capacity of the phylogeny - shared ancestry allows prediction on the traits of descent species. This can lead to ancestral state reconstruction, predictive reconstruction, in silico or in vivo.
- If successful, this approach will transform the way we study genetics, evodevo, trait evolution in fundamental ways since it will eliminate basic barriers.
- This work would result in database integration across different taxa in phylogenetic frameworks that work seamlessly, thereby enhancing scientific discovery. Such centralized resources would increase research efficiency.
- There is also the potential discovery of novel forms of convergence.

What are the training and workforce opportunities within this idea?

The broad scope of this project would provide rich opportunities for interdisciplinary training that would span undergraduate and graduate student trainees, along with postdocs. Areas would include systematics, animal physiology, functional and comparative genomics, molecular and cellular biology, along with the computational/data science skill development required for data generation, databases and analysis platforms, and research data management.

In addition to expanding early exposure for trainees in these fields, the expertise required for successful execution would also expand PI's knowledge-base. This work would harness cognitive diversity to support the study of convergent evolution, and to differentiate from homology/parallelism etc. The variety in topics provides a broad basis for recruiting individuals historically under-represented in STEM overall, including to facilitate the development of computational, evolutionary, and molecular skills.

Specific ideas

- Interdisciplinary cross-institutional CURE opportunities.
- Facilitate career readiness for junior career individuals to be competitive in securing industry positions (e.g. agricultural and biomedical companies).
- Outreach to communities using integrated data, e.g. for broad-based education of K-12 and the general public (using accessible and exciting examples).

What is the reasoning, justification, and/or supporting evidence behind this idea?

With genome (and transcriptome, etc.) **sequencing now affordable and accessible** for a broad range of organisms, these data can be leveraged to address important questions in convergent evolution. We are **already seeing the benefits** of resources developed for **new and emerging "model systems"** in understanding genotype-phenotype links, with further depth of understanding of molecular mechanisms, environmental contributors, GxG and GxE interactions, etc. For example, the stickleback genome and subsequent efforts in that community have led to deep understanding of environmental adaptation through repeated evolution. Likewise, many flowering plant communities maintain genomic databases relevant to agricultural uses. Recent advances in applying **gene editing technologies** in non-model systems make new models amenable to experimental follow-up *if* there are interpretable candidates to test. **Ongoing developments in AI/ML** will enable further opportunities to learn from large datasets *if* the data are structured and annotated appropriately and accessible to learning models. With existing resources such as those below and ongoing EXPLOSION of data in new systems, now is the PERFECT time to generate resources that will enable ANY clade with examples of convergence to be a "model clade"!

Authors

Hojun Song	Texas A&M
Wynn Meyer	Lehigh University
Xuan Zhuang	University of Arkansas
Keriayn Smith	University of North Carolina - Chapel Hill
Stephen Cameron	Purdue University
Gavin Conant	North Carolina State University

17. Researching multi-scale convergent responses to stress grounded in human context

What is your Big Idea?

Evolutionary convergent responses to human-driven biotic and abiotic environmental change occur across biological scales and must be viewed in context. Marginalized communities disproportionately witness and adapt to environmental stressors, however, rarely are non-biologist local understanding and knowledge included in definitions of context, assessment of biological responses, or development of solutions to environmental change.

- Identify the environmental context of ecosystems that have been highly impacted by human activities. Identification, study, and biologically based solutions will be done in direct collaboration with impacted marginalized communities in both rural and urban settings.
- To understand how multiscale biological communities (ie. microbial communities, plants, wildlife) adapt and evolve with time in the presence of an environmental stressor (salinity, heavy metals, organic contaminants, draught, agent orange, antibiotics etc.) we will compare disturbed ecosystems to less impacted natural sites and experimental systems replicated in different contexts (scale, climate, region, urban, rural etc.).
- To identify evolutionary adaptations to specific environmental challenges we will obtain representative chemical and biological omic data from (plants, animals, soil, water etc.) in a set of ecosystems that represent high priority community challenges. We will examine specific genotype and phenotype changes as well as the rate of changes (genetic, community composition, abiotic properties) and determine links between abiotic, metagenomic, and transcriptomic genotypic and phenotypic features of the samples.
- Data allows for identification of stressor resilient organisms, genes, mutations, and mechanisms of adaptation and evolution that can be used to develop solutions to challenges in collaboration with impacted communities.

Which *Branch of LIFE* does this address?

- New frontiers in evolutionary convergence
- Leveraging life’s diversity
- Training and broadening participation

What is required to pursue this?

- Long term commitments by funding agencies, researchers and community partners are essential.
- Social science research will be needed to build collaboration with specific environmental justice communities.
- Experimental ecosystem plots across the Anthropocene that are managed (including replicates).
- Genuine collaboration with Environmental Justice impacted communities.

- Practical ways of compensating community participants and facilitating community in making changes initiated by the community.

What are barriers to pursuing this now?

- Long timescale of project, establishing genuine contacts and collaboration in environmental justice impacted communities take time, social science research and expertise to develop and build meaningful trusting and lasting collaborations.
- Experimental ecosystem plots need to be managed and monitored collaboratively for a long time period with ongoing communication between collaborators on progress and outcomes.
- Silos of funding isolate social and biological sciences as well as eco-evolutionary studies in natural ecosystems and those impacting human and environment health.
- A diversity of projects and broad consensus on priority challenges will be difficult.

What is the expected value and impact?

- Co-creation with communities will increase the accessibility and translational relevance of NSF science.
- Understanding the movement of genes in ecosystems leads to predictions and possible manipulations to shape outcomes.
- Molecular adaptations to environmental stressors (such as enzymes) might be engineered to assist adaptation in species of interest for conservation, agriculture, environmental health and mediation and containment of contaminants.
- Capturing biotic changes resulting from abiotic stressors in different abiotic environments will increase understanding on impact of abiotic environment on adaptation.

What are the training and workforce opportunities within this idea?

- Generating meaningful collaborations between scientists and EJ community members will educate scientists in the real context so that their work will be accessible, relevant, and successful.
- Funding to EJ community members who provide the expertise on the EJ challenges will increase participation in a sustainable and long-term way.

What is the reasoning, justification, and/or supporting evidence behind this idea?

Globally there is an increase in environmental justice challenges and impacted communities. The evolutionary responses and biotic to specific abiotic stressors are not well understood. The relative significance and prevalence of specific stressors are not well accounted for by the scientific community, potentially leading to nonoptimal use of resources.

Authors

Rachel J. Whitaker	University of Illinois Urbana Champaign
Nina Goodey	Montclair State University

Ya Yang	University of Minnesota
Valeria Milam	Georgia University of Technology

18. A (social) web to stimulate interdisciplinary collaborations

What is your Big Idea?

Design, development, and launch of a platform and resource to facilitate interdisciplinary and diverse scientific partnerships

Which *Branch of LIFE* does this address?

Tools & resources, Collaboration Across Disciplines, Training & Broadening Participation, Leveraging Diversity

What is required to pursue this?

This platform would capture elements of social media platforms including ease of connecting and communicating, identifying, and connecting with individuals with commonalities, and suggestions for new contacts.

To develop this social web/network to connect prospective collaborators, data centers and accompanying data infrastructure to facilitate data ingestion and storage would be key requirements. Systems could be built upon existing open-source tools and refined for the needs of the scientific community.

To design a successful web that is the foundation for meaningful collaborations, we recommend that knowledge and practices from social science researchers be leveraged. Artificial intelligence can be used to power automated recommendations for new contacts and areas that may be the foundation of new collaborations. This pairing would help to stimulate new thinking and overcome existing research silos that impede research potential.

Skillsets and areas of expertise that would support this project include data engineers, data scientists, social scientists, and a broad spectrum of scientists who would be early adopters and early majority, after the innovative aspects are designed and implemented. Representation from citizen scientist groups could be implemented as an additional arm to this project.

Support and collaboration from various NSF directorates (including but not limited to SBE, CISE, BIO, TIP, EDU) would facilitate more widespread development and user adoption.

What are barriers to pursuing this now?

- Funding, momentum and support to launch this idea
- Development of necessary infrastructure
- Buy-in and adoption
- Current (and longstanding) research silos that currently exist
- Communication barriers across fields (e.g. social and behavioral scientists, evolutionary biologists, computer scientists)

What is the expected value and impact?

One major thrust of this work would be to build the interdisciplinary teams that are required to tackle big unanswered scientific questions. The platform would therefore lower barriers for identifying collaborators and/or complementary or convergent research areas, which may not have been obvious directions.

The platform would therefore support knowledge-sharing and education, as well as network-building and expansion by increasing interdisciplinary opportunities. The social web could also increase research efficiency by reducing barriers to partnerships that would provide necessary expertise.

It can be envisioned that this may have economic impacts by increasing funding outcomes, particularly for emerging/under-developed areas, thereby dispersing funding across broader groups of investigators.

This public forum built by scientists for scientists would also aid in democratizing science. New research ideas may also emerge through public and private discussions on this platform. Support and interest for pursuing emerging research interests could also be gauged in a manner similar to current crowdsourcing efforts, which while common in the public arena, are rare in areas that support scientific research. This platform could strengthen communication channels between scientists and citizen scientists and/or environmental justice communities.

What are the training and workforce opportunities within this idea?

- Broad exposure to many scientific topics for trainees via network building, knowledge sharing, and the public nature of the platform.
- Data science and engineering training and professional development for trainees would also result from this project.
- The design of the platform could facilitate elements of public engagement and education, and could also support citizen science.
- Expanding workforce opportunities to PhD scientists whose career trajectories diverge from traditional academic paths.
- Develop resources for bidirectional learning between traditional academic communities and citizen scientists, and environmental justice community representatives, as an example.

What is the reasoning, justification, and/or supporting evidence behind this idea?

- LIFE scoping workshops!
    - Usage of the jargon board indicates that research would benefit from improved communication and connections across disciplines. The platform would help individual researchers to overcome gaps in language (jargon) and communication to emerge from siloed existence.
- Current social media platforms are too broad in scope to fill this gap for scientists specifically.
- Most scientific societies generally reinforce research silos.

- New technologies and new applications of existing technologies will emerge through connections built in the establishment and operation of this social web.
- Promote inclusion of PhD scientists pursuing non-traditional career paths (e.g., communication, visualization, animation).

Authors

Keriayn Smith	University of North Carolina, Chapel Hill
Wynn Meyer	Lehigh University
Heather Maughan	Ronin Institute
Nina Goodey	Montclair State University

LIFE Participants

In-person Scoping Session

Alexey Markin	USDA-ARS
Allen Liu	University of Michigan
Ally Nguyen	Whitehead Institute
Amy Keating	Massachusetts Institute of Technology
Andrea Bodnar	Gloucester Marine Genomics Institute
Annie Lindgren	Portland State University
Antonis Rokas	Vanderbilt University
Barbara Sharanowski	University of Central Florida, Orlando, FL
Chad Westphal	Wabash College
Chris Todd Hittinger	University of Wisconsin-Madison
Clay Small	University of Oregon
Corbin Jones	University of North Carolina
Corien Bakermans	Pennsylvania State University - Altoona
Cynthia Downs	State University of New York - ESF
Dan Hahn	University of Florida
David Liberles	Temple University
Davorka Gulisija	University of New Mexico
Dietmar Kültz	University of California, Davis
Eduardo Rosa-Molinar	University of Kansas
Emily Dolson	Michigan State University
Fay-Wei Li	Boyce Thompson Institute
Giovanna Ghirlanda	Arizona State University
Grey Monroe	University of California, Davis
Guin Wogan	Oklahoma State University
Hongmei Li-Byarlay	Central State University, Wilberforce, OH
Jeff Mower	University of Nebraska
Jen Kovacs	Agnes Scott College, Decatur, GA
Jeremy Brown	Louisiana State University
Jeremy Van Cleve	University of Kentucky
Josh Widhalm	Purdue University
Jung-Youn Lee	University of Delaware
Justin Meyer	University of California, San Diego
Karen Sears	University of California, Los Angeles
Leonie Moyle	Indiana University

Leslie Babonis	Cornell University
Luis Zaman	University of Michigan
Nathan Clark	University of Pittsburgh
Nora Mitchell	University of Wisconsin, Eau Claire, WI
Prosanta Chakrabarty	Louisiana State University
Randall Boone	Colorado State University, Fort Collins, CO
Ranjan Srivastava	University of Connecticut
Rob Unckless	Kansas University
Rosana Zenil-Ferguson	University of Kentucky
Sam Scarpino	Northeastern University
Sara E. Miller	University of Missouri-St. Louis
Subir Shakya	Harvard University
Tilottama Roy	Missouri Western State University
Tim Sackton	Harvard University
Tomokazu Kawashima	University of Kentucky
Verónica S. Di Stilio	University of Washington
Yanni Chen	University of Notre Dame
Yu She	Purdue University
Zinan Wang	Michigan State University, East Lansing, MI

Virtual Scoping Session

Alexandra Deaconescu	Brown University
Bin He	University of Iowa
Cecile Ane	University of Wisconsin-Madison
Coral Zhou	University of California - Berkeley
Daniel Herschlag	Stanford University
David Alvarez-Ponce	University of Nevada, Reno
Diana Rennison	University of California San Diego
Eric Haag	University of Maryland, College Park
Erika Edwards	Yale University
Frederic Chain	University of Massachusetts - Lowell
Gavin Conant	North Carolina State University
Heather Maughan	Ronin Institute
Hojun Song	Texas A&M
Javier Arsuaga	University of California, Davis
Jay T. Lennon	Indiana University
Jo Wolfe	Harvard University
Keriayn Smith	University of North Carolina - Chapel Hill

Kim Hoke	Colorado State University
Maribet Gamboa	Universidad Catolica de la Santisima Concepcion
Maricris Mayes	University of Massachusetts, Dartmouth
Mark Margres	University of South Florida
Matt Anderson	University of Wisconsin-Madison
Matt Daugherty	University of California, San Diego
Matt Stata	Michigan State University
Megan Dennis	University of California, Davis
Michele Johnson	Trinity University
Nicole Riddle	University of Alabama at Birmingham
Nina Goodey	Montclair State University
Omer Gokcumen	SUNY Buffalo
Paulyn Cartwright	University of Kansas
Rachel J. Whitaker	University of Illinois Urbana Champaign
Rich Meisel	University of Houston
Sergey Kryazhimskiy	University of California, San Diego
Stephen Cameron	Purdue University
Sudhir Kumar	Temple University
Suzanne McGaugh	University of Minnesota
Tae Seok Moon	Engineering Biology Research Consortium
Todd Barkman	Western Michigan University
Valeria Milam	Georgia University of Technology
Vivian Irish	Yale University
Wynn Meyer	Lehigh University
Xuan Zhuang	University of Arkansas
Ya Yang	University of Minnesota

Synthesis Session

Chris Todd Hittinger	University of Wisconsin-Madison
Dan Hahn	University of Florida
Erica Edwards	Yale University
Javier Arsuaga	University of California, Davis
Jennifer Wisecaver	Purdue University
Karen Sears	University of California, Los Angeles
Keriayn Smith	University of North Carolina - Chapel Hill
Leslie Babonis	Cornell University
Matt Daugherty	University of California, San Diego
Matt Stata	Michigan State University

Sudhir Kumar	Temple University
Tae Seok Moon	Engineering Biology Research Consortium
Tim Sackton	Harvard University
Wynn Meyer	Lehigh University

Leadership Team

Anna Allen	NSF, Division of Integrative Organismal Systems (BIO/IOS)
Chris Todd Hittinger	University of Wisconsin-Madison
Erika Edwards	Yale University
Jennifer Wisecaver	Purdue University
Leslie Babonis	Cornell University
Leslie Rissler	NSF, Division of Environmental Biology (BIO/DEB)
Reed Beaman	NSF, Division of Biological Infrastructure (BIO/DBI)
Stephen DiFazio	NSF, Division of Molecular & Cellular Biosciences (BIO/MCB)