National Science Foundation EMERGE Biology Integration Institute

Strategic Plan, Spring 2021

Core Values

The EMERGE Biology Integration Institute is built on the core, interacting values of:

- **Curiosity**: We are united by a passionate curiosity about the mechanistic feedbacks between biological communities (especially microbiomes, due to relative tractability) and the environment. That curiosity leads fundamentally to values of:
- **Diversity:** a breadth of tools, knowledge, and experiences fuel understanding.
- **Integration**: to achieve our scientific goals, we must integrate across disciplines, skillsets, and perspectives. (See Figure 1).
- **Respect**: curiosity, diversity and integration invoke a framework of respect - among Institute members; for the ecosystem we are studying; and for the taxpayers funding this work.
- **Inclusion:** collectively, the above must create a culture of inclusion, so that vital new ideas are heard and all Institute members feel valued and safe to participate whole-heartedly in the work we are pioneering.
- **Rigor**: to integrate disciplines in pursuit of ecosystem feedbacks, driven by curiosity and respect, *requires* scientific rigor.
    - With EMERGE's inclusion of Science of Team Science, that rigor extends not only from the research itself but to the trainings and structures through which we forge our teams.

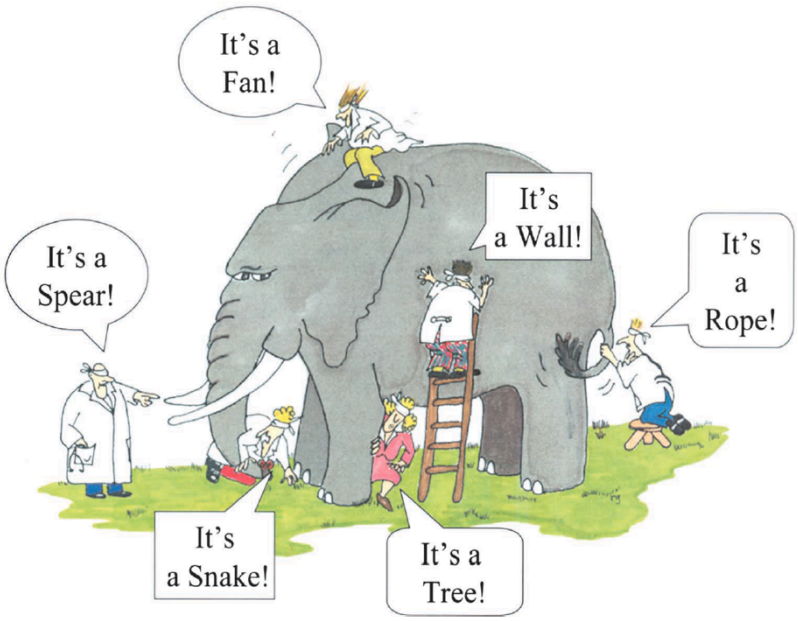

Figure 1: Why EMERGE needs diversity *and* integration. From: P-M Daigneault. 2013. The Blind Men and the Elephant: A Metaphor to Illuminate the Role of Researchers and Reviewers in Social Science. *Methodological Innovations Online*. 8(2): 82-89.

Mission Statement

EMERGE was forged to integrate research and training to characterize EMergent Ecosystem Response to chanGE, in order to better understand, predict, and manage natural ecosystems.

We weave together diverse disciplines - ecosystem science, ecology, evolution, organismal biology, molecular and cell biology, computational science, remote sensing, modeling, biogeochemistry, education, and social science - because that is what is required to understand ecosystems.

We will train early career researchers and in toolkit we develop, which spans not only the integrative science we pursue but also the teamwork structures to help accomplish it.

We will be transparent in our data sharing, and active in our outreach to scientific colleagues and the public.

Decisions are guided by our Mission and our Core Values, and are made via a network organizational structure with modest functional redundancy, which research shows creates resilience within teams[1,2].

Vision Statement

EMERGE aspires to transformatively advance our understanding of ecosystem/change interactions, and to integrate for the first time a mechanistic and predictive understanding of how physiological acclimation, community assembly, and evolutionary adaptation impact climate.

These advances will be captured by a “G2E2G” (genes-to-ecosystems-to-genes) toolkit of conceptual framework, data, tools, models and synthesis. They will be deployed by a trained next generation of diverse leaders and team players. And they will have a broad impact across biology through G2E2G.

While EMERGE focuses primarily on the mechanisms of a model ecosystem’s microbiome due to tractability with current cutting-edge datasets, we endeavor to lay the foundational framework for studies with increasing layers of meta-omic data to map ecosystem integration and environmental feedbacks.

Current Challenges, Opportunities, Strengths, and Weaknesses

A key opportunity of EMERGE is also a key challenge: integrating new scientists with several established teams, who themselves will now be cross-integrating. This expansion brings essential new knowledge and fresh insights, but presents team science challenges. We are working proactively and explicitly with our Team Science core to manage this integration process.

EMERGE is a very large team already, so while there are several skillsets that would further improve the institute (specifically, more mathematics, statistics and theory), with >80 people we cannot afford to grow further until we obtain additional leveraged funding. As we are working with a new combination of concepts and people, our scientific goals are best served by spending our first ~2 years solidifying these ideas and making progress, before we bring in new skills.

Goals, Initiatives and Objectives

EMERGE integrates across 9 Teams (5 Research Activity - "RA" - Teams, and 4 others), that tackle the work of the Institute. Over our first 6 months, the Institute has worked together via all-hands Workshops and team-specific monthly meetings to develop overall goals for the Institute as well as Team-specific goals & objectives; these are delineated at the 5- and 1-year time horizons below. We also developed 6-month deliverables (in October 2020) which are not listed below in the interests of space, but have largely been met (in April 2021). We have combined the Strategic Plan template's sections of Goals with Initiatives and Objectives, since in our launch work they have been integrated.

5-yr goals:

Overall:

- To accomplish a series of parallel interlinked experiments as incubations, bioreactors, and cultures testing the response of the microbiome to differing temperature perturbations, and analyze the resulting datasets in a coordinated way among experiments and Research Aim Teams, to understand a range of community responses to temperature increase, and their interactions.
- To relate the observations of changing gene expression, community composition, genetic response, and system output (metabolomics and carbon gas dynamics) to those observed in the field from multi-omics studies spanning more than a decade.
- To iteratively incorporate insights into models via improved parameterization of currently represented processes and, as needed, inclusion of new processes identified as essential for correctly predicting ecosystem-scale methane dynamics. This specifically involves developing a "G2E2G" version of the ecosystem process model ecosys, aided by distilling the metagenome-assembled genomes from these microbiomes into habitat-specific simplified functional microbiomes using Biocrunch, to test the important of processes currently missing from ecosys.
- To provide and disseminate an "EMERGE Toolkit" to the broader scientific community to perform G2E2G assessments of other model ecosystems; this toolkit spans informatic tools and team science protocols, and will be shared with the community via workshops and meetings, including a Gordon Research Conference co-proposed with several other model ecosystem teams.
- To transparently engage with diverse stakeholders via an integrated outreach program, leveraging the pandemic-required conversion of various team resources to electronic forms, while retaining bricks-and-mortar public engagement through science centers among our member institutions, and including a TEDx event focused on ecosystem-gene reciprocal interactions.
- To train a suite of integrative biologists in the interdisciplinary G2E2G framework through trainees in the Institute itself as well as early career researchers in the broader community via our Summer Institute.

Team RA1: From Expression to Acclimation

- Temperature gradient and ramp peat incubations to test metagenome predictions
- Validate cryo-adaptation and cold-shock as a proxy for acclimation potential; develop ***MetaAcclaim***
- Characterize gene regulatory circuitry components in key microbial lineages; develop ***EcoMONET***
- Adopt new high-resolution molecular tools to complex communities (e.g. meta-RiboSeq, epigenetics and epitranscriptomics)
- Reconstruct the different modes of regulatory mechanisms and attempt to relate them to

community assembly/interactions, adaptation, and system outputs.

Team RA2: From Interactions to Assembly

- Expand DRAM computational capabilities to allow rapid and quantifiable assessment of microbial interactions in the existing and new 'omics data
- Construct an assembly framework that includes taxonomic-based and trait-based data, furthering refining concept of stochasticity and decomposes drivers of abiotic-biotic and biotic-biotic interactions (e.g., pH, temperature, competition).
- In Collaboration with RA4 and RA5, data architecture of RA2 is seamlessly ingestible into modeling and other approaches

Team RA3: From Genetic Variation to Adaptation

To leverage decadal field data, new data generation and analytical capabilities to understand:

- Macrodiversity (inter-population variation) of microbes and dsDNA viruses
- Microdiversity (intra-population variation) of microbes and dsDNA viruses
- Assess macro-/micro-diversity for ssDNA/RNA viruses and mobile elements (plasmids, introns, DGRs)
- Selection and adaptation in all these entities

Team RA4: G2E2G into ecosys and Biocrunch models

By developing new eco-evolutionary theory and computational tools we will

- Advance the *BioCrunch* model to simulate the 3As.
- Develop the ecosys model to create *ecosys+G2E2G*
- Integrate *BioCrunch* and *Ecosys* into a modeling framework from genomes to ecosystem function and dynamics that accounts for the 3 As.
- Use the modeling framework to evaluate the relative contributions of the 3As in the response of each successional stage to warming, and in the transition between stages.

Team RA5: Landscape-scale G2E2G-ecosys

- Develop the protocol to use G2E2G-enabled models at the landscape scale
- Calibrate and test the model structure using EMERGE (and other) observations
- Analyze how ecosystem changes influence biota across scales (**G2E**), feedback to alter gene content and expression over time (**E2G**), and affect ecosystem C cycling.

Team RP: Research Platform

Extension of the sample archive, continued field sampling to capture ongoing rapid thaw via

(a) baseline annual monitoring of the focal sites in YR1-5 (vegetation, peat, porewater, autochambers)

(b) identification of sites across the Mire that span a chronosequence of time-since thaw (75 years of remote sensing data to sample these sites in YR4

(c) collect peat for incubation and bioreactor studies

Team E&T: Education & Training

- Development of a new generation of biologists who can tackle the G2E2G integrative science challenge
- A deeper understanding of the Science of Team Science (SciTS), diversity and inclusion, and science communication for all EMERGE participants

Team O: Outreach

- All EMERGE members will have engaged in at least 1 broader scientific community outreach activity and 1 general public outreach activity.
- EMERGE will partner with other long-term model ecosystems to successfully competitively propose, be awarded, plan, and execute a Gordon Research Conference on cross-site G2E2G comparisons.
- EMERGE will have conducted a TEDx-style meeting to translate Institute science to the public.

Team IDM: Informatics & Data Management

- Informatics: Develop informatics toolkits for gleaning G2E2G insights from multi-omic multi-disciplinary data in RA 1-3.
- Data Management: Expand the IsoGenieDB into a more sophisticated EMERGE-DB with more automation of data import and better adherence to FAIR (Findable, Accessible, Interoperable, Reusable) data principles.

1-yr Goals: **We made these more modest in acknowledgement of the extraordinary time the Institute's launch is occurring during.**

Overall:

- To have formed and populated the 9 Teams, with strategic cross-membership to engage integration immediately, and be holding regular monthly meetings
- To have planned the details of the parallel temperature-perturbation experiments with a consensus across RA Teams to ensure the resulting datasets will meet all teams' needs.
- To have begun regular team science trainings

Team RA1: From Expression to Acclimation

- Establish batch cultures with isolated strains, and also transition to batch-fed and/or continuous culture in chemostats
- Begin characterizing basic physiology of isolates. RA4 needs T opt, pH opt, growth rates, Monod kinetics, etc.
- Complete temp ramp incubations (pending more peat).
- Design validation experiments for biomarker candidates and profiles in defined culture systems.
- Develop EcoMONET v0.1 and MetaAcclaim v1.0

Team RA2: From Interactions to Assembly

- Working beta version of DRAM 2.0 with capabilities for 'omics integration and limited abilities to predict interactions between annotated MAGs
- Characterize relationship of Microbial Diversity, functional diversity and Metabolite diversity
- Use the SIP metabolite experiment to infer new pathways
- Assembly analysis of permafrost after thaw from archival data
- Assess annual variability in phylogeny and functional potential across decadal dataset
    - Develop a ranked list/conceptual model of drivers important in assembly from the previous bullets
- Hire students, post-docs and staff on this project and get them up to speed

Team RA3: From Genetic Variation to Adaptation

- Apply these tools (for viruses, DGRs, plasmids, introns) to the longitudinal data from Stordalen

and start finding amazing stories
- Establish strain-resolved datasets with LoriKeet (on existing short-read data?) and VirION2 (adding in long-read and optimized hybrid assemblies) and strategically incorporate these data into a "version 2" MAGs effort
- Start discussions for how to integrate across these streams of data to create a synthesis perspective of adaptation in the Mire

Team RA4: G2E2G into ecosys and Biocrunch models

- BC: cluster genomes into guilds and match to pathways/edges of reaction network
- BC: compress guilds into minimal number of genomes to capture variance in kinetic statistics
- BC: sensitivity analysis of selected process rates to reaction network complexity and guild parameterization
- EcoEvo: dN/dS as signatures of active selection - can we obtain a rank order of traits under selection with RA3
- *ecosys*:
    - Testing of integration of microbial traits derived from BC into ecosys
    - Design conceptual framework for integration of trait selection processes
    - ecosys+G2E2G plan and initial steps

Team RA5: Landscape-scale G2E2G-ecosys

1) Remote Sensing
    a) Drone collection summer 2021 at Stordalen.
        i) Fixed-wing RGB camera collection (same setup as 2014-2019 collection)
        ii) Fly 5 channel 4 rotor drone with upward looking sensor
        iii) Possibly fly hyperspectral/lidar package at specific locations at mire, since complete coverage is not possible
    b) Work on analysis of hyperspectral imagery from 2019 collection in northern Canada
    c) Examine the feasibility of using Structure From Motion to generate DEM from drone collected at Stordalen 2019
    d) Find appropriate graduate student to collect data for 2022
2) Ecosys modeling
    a) Start setting up 2D transect simulations
    b) Preliminary integration of microbial traits from RA4 analysis
    c) Analyze role of vegetation cover change in C cycle
    d) Initial analysis of importance of microbial trait heterogeneity on emergent ecosystem properties and fluxes.
    e) Compare carbon cycle transformation rates with Rick's isotope inversions
3) Field data collection
    a) GPS ground control points tied with vegetation plots
    b) Explore the use of a handheld spectrometer to examine the plot level vegetation and examine collections done in 2014

Team RP: Research Platform

- Autochambers running and supported
- Successful collection of materials for incubations and bioreactors
- Successful collection of peat and geochemical samples at primary field sites
- All documentation in place from sampling

- All samples shipped to final destination

Team E&T: Education & Training

- Summer Institute implemented (Aug 2021) and evaluated
- REU program developed and started recruiting (Starts summer 2022)
- OSU Viromics Workshop (5th one is May 2021) (training folks (80-100 people) in viromics data processing & interpretation; distinct from OSU-CoMS Microbiome Webinar Series)
- EMERGE team has undergone some training in SciTS, DI, Mentoring and SciComm
- EMERGE Trainees and faculty have participated in more than one RA's monthly meeting (ie integration among RAs is occurring via participation)

Team O: Outreach

- Developed video-summary of EMERGE for the public
- Every EMERGE trainee has completed 1 outreach activity; all others have committed to an outreach activity
- Collect members' outreach activities centrally for EMERGE library-of-outreach-resources; suggest expansion or connections or teaming up with others
- The power of genome research and modeling in environmental systems; science of team science; drone work and info you get from it
- Metrics/tracking mechanism for outreach activities

Team IDM: Informatics & Data Management

- Similar to field data collection templates, make similar templates for incubation and bioreactor data, with the goal that data from different labs can be readily compared and incorporated into DB without endless back-and-forthing with data providers for additional details.
- Incorporate 2021 field data into the EMERGE-DB
- Hold Deep Dive training on DB for Institute.
- Hire web-dev person and get them started on improving web portal

~~~~~~~~~~~~~~~

Roles & Responsibilities:

Due to EMERGE's size, rather than listing all >80 members' Roles and Responsibilities across the 9 Teams here, we direct interested readers to our living EMERGE WhosWho internal google sheet in which we track membership and Roles and Responsibilities across the 9 teams, with responsibilities explicitly linked to 6-month and 1-yr deliverables.
https://docs.google.com/spreadsheets/d/1JETqxdUzaXEB3ajJ7gvE9Z-q5f2bP6_aOy6MITs6YWQ/edit#gid=1949009430

Milestones:

Launch Workshop. Met: October 2020 and April 2021. Due to covid this has been held virtually, spread across six 3-hr Zoom meetings, in two halves: the first in October 2020 (spread over 4 weeks), to coalesce the 9 Teams, establish 6-month, 1-year, and 5-year goals overall and within teams, adopt a common communication platform and engagement norms, and train in Science of Team Science. The second half of the Launch Workshop is in April 2021 (spread over 3 weeks), and focuses on scientific cross-training, authorship guidelines, and mapping synthesis opportunities. The extended framework of the workshop allowed team "homeworks" from one meeting to the next, for example to map
~~~~~~~~~~~~~~~

Individuals's Roles & Responsibilities onto timeline deliverables, and to map out the data stream dependencies for each team and among them. In addition, members were charged with developing a number of prerecorded short talks (some examples include: 2" "paper nuggets" where team members shared 1 figure from a recent paper and articulated its relevance to EMERGE; high-level overviews of elements of the research platform such as "How does the Biocrunch model work"; an introduction to the in-development Fieldwork Safety, Equity and Inclusion Manual) which were then loaded on EMERGE's internal YouTube channel and viewing was required before specific days of the workshop. In this way, we simulated the experience of a rich, multi-day in-person launch workshop as best we could; there were drawbacks still of course, but also benefits.

Establishment of recurring scheduling monthly meetings for each team. Met: January 2021. Due to members' balancing childcare and other responsibilities during this time, scheduling is more challenging across all spheres of life during covid, but in January 2021 we met our goal of getting each Team "up and running" with monthly meetings, on the common Institute calendar and with ongoing agendas tied to our 6-month and 1-yr goals.

Planning and Execution of the 2021 Field Season. Ongoing, will be concluded by August 2021.

Establish regular (~quarterly) "Debrief with the Directors" meetings for the ~22 team leads, separately for each team, with the Co-Directors, to directly communicate about challenges, opportunities, etc. The first occurred in January 2021, the second is currently being scheduled for May 2021.

Initiate and make steady progress (at least one every 2 months) on EMERGE "Deep Dives and Consensus Planning Sessions". Met and ongoing. These 2 hr meetings provide either a deep dive into a topic of key importance to the institute, or a longer focused discussion opportunity to plan an institute activity. We have held Deep Dives on the *ecosys* model and on the EMERGE Microbiome Data Release v.1, and multiple consensus planning sessions on the EMERGE Incubations.

Summer Institute: August 2021. PLanning is well underway for this 2-week intensive course for early career researchers, which will occur in hybrid format with 1 week in person (in Durham NH) and the second week remote.

Overall, we are making good progress in our YR1 Milestones. We will hold the YR2 workshop in fall 2021. We have delayed conducting the EMERGE Incubations until summer/fall 2022 as they require a larger and longer field presence than is feasible at the moment, but this benefits the institute greatly in allowing time to refine plans and allow them to be more fully informed by deep analyses of the existing decade of field data, and of a number of existing incubation and bioreactor datasets. Our clearly-articulated goals, and mapping of roles and responsibilities onto them, will be key to maintaining the forward momentum we have created in this next year.

References:

1. Anklam, P. *Net Work: A Practical Guide to Creating and Sustaining Networks at Work and in the World*. (Routledge, 2007).
2. Lazer, D. & Katz, N. Building Effective Intra-Organizational Networks: The Role of Teams. (2003). at <https://dspace.mit.edu/bitstream/handle/1721.1/55801/?show=full>