DOE/SC–0190

Grand Challenges for Biological and Environmental Research:

Progress and Future Vision

November 2017

A report from the Biological and Environmental Research Advisory Committee

Suggested citation for this report: BERAC. 2017. *Grand Challenges for Biological and Environmental Research: Progress and Future Vision; A Report from the Biological and Environmental Research Advisory Committee,* DOE/SC–0190, BERAC Subcommittee on Grand Research Challenges for Biological and Environmental Research (science.energy.gov/~/media/ber/berac/pdf/Reports/BERAC-2017-Grand-Challenges-Report.pdf).

Cover

Images represent a broad range of natural systems that drive the science supported by the Office of Biological and Environmental Research within the U.S. Department of Energy Office of Science. These systems are not only structurally and spatially complex with many different interacting parts spanning molecular to global scales, but they also are dynamically complex, encompassing processes that occur over time scales ranging from nanoseconds to centuries.

Image credits: Ball-and-stick representation of atoms within a molecule from Pacific Northwest National Laboratory. Ribosome molecular complex image from Ditlev Brodersen, Aarhus University. Microscopic image of fungal hyphae on a root surface, copyright Merton Brown, Visuals Unlimited. Forest ecosystem image from the U.S. National Park Service. Mountain and Earth images from iStockphoto. Cover developed at Oak Ridge National Laboratory.

DOE/SC–0190

Grand Challenges for Biological and Environmental Research: Progress and Future Vision

A Report from the Biological and Environmental Research Advisory Committee

November 2017

Chair
Gary Stacey (University of Missouri)

U.S. Department of Energy

Prepared by the BERAC Subcommittee on
Grand Research Challenges for Biological and Environmental Research

Report available online at
science.energy.gov/~/media/ber/berac/pdf/Reports/BERAC-2017-Grand-Challenges-Report.pdf

Biological and Environmental Research Advisory Committee

Chair
Gary Stacey
University of Missouri

Sarah M. Assmann
Pennsylvania State University

Dennis D. Baldocchi
University of California, Berkeley

Amy M. Brunner
Virginia Polytechnic Institute and State University

James R. Ehleringer
University of Utah

James J. Hack
Oak Ridge National Laboratory

Bruce A. Hungate
Northern Arizona University

Anthony C. Janetos
Boston University

Andrzej Joachimiak
Argonne National Laboratory

Cheryl R. Kuske
Los Alamos National Laboratory

L. Ruby Leung
Pacific Northwest National Laboratory

Gerald A. Meehl
National Center for Atmospheric Research

Jerry M. Melillo
Marine Biological Laboratory

Gloria K. Muday
Wake Forest University

Kristala L. Jones Prather
Massachusetts Institute of Technology

David A. Randall
Colorado State University

James T. Randerson
University of California, Irvine

Karin A. Remington
Computationality, LLC

G. Philip Robertson
Michigan State University

Karen Schlauch
University of Nevada, Reno

Daniel Segrè
Boston University

David A. Stahl
University of Washington

Judy D. Wall
University of Missouri

John P. Weyant
Stanford University

Minghua Zhang
State University of New York at Stony Brook

Huimin Zhao
University of Illinois at Urbana-Champaign

Designated Federal Officer
Tristram West
U.S. Department of Energy Office of Biological and Environmental Research

BERAC Subcommittee on Grand Research Challenges for Biological and Environmental Research

Gary Stacey, BERAC Chair
University of Missouri

James J. Hack
Oak Ridge National Laboratory

Bruce A. Hungate
Northern Arizona University

Anthony C. Janetos
Boston University

Andrzej Joachimiak
Argonne National Laboratory

Cheryl R. Kuske
Los Alamos National Laboratory

Gerald A. Meehl
National Center for Atmospheric Research

Jerry M. Melillo
Marine Biological Laboratory

David A. Randall
Colorado State University

James T. Randerson
University of California, Irvine

Karin A. Remington
Computationality, LLC

G. Philip Robertson
Michigan State University

Daniel Segrè
Boston University

David A. Stahl
University of Washington

John P. Weyant
Stanford University

Minghua Zhang
State University of New York at Stony Brook

Huimin Zhao
University of Illinois at Urbana-Champaign

Contributing Authors

Chapter 2: Grand Challenges in Biological Systems Science

Chair/Lead Writer: Cheryl R. Kuske
Co-Chair/Facilitator: Daniel Segrè
Writers: Mary Dunlop, Andrzej Joachimiak, John Shanklin, and Huimin Zhao

Chapter 3: Grand Challenges in Earth and Environmental Systems Sciences

Chair/Lead Writer: L. Ruby Leung
Co-Chair/Facilitator: Minghua Zhang
Writers: David A. Randall, James T. Randerson, Charles Jackson, Gerald A. Meehl, and Michael Wehner

Chapter 4: Grand Challenges in Microbial to Earth System Pathways

Chair/Lead Writer: Bruce A. Hungate
Co-Chair/Facilitator: David A. Stahl
Writers: Kirsten Hofmockel, Jennifer Pett-Ridge, Richard Phillips, Will Wieder, and Victoria Orphan

Chapter 5: Grand Challenges in Energy Sustainability

Chair/Lead Writer: G. Philip Robertson
Co-Chair/Facilitator: John P. Weyant
Writers: Susan Hubbard, Anthony C. Janetos, James T. Randerson, and David A. Stahl

Chapter 6: Grand Challenges in Data Analytics and Computing

Chair/Lead Writer: Karin A. Remington
Co-Chair/Facilitator: Karen Schlauch
Writers: Shibu Yooseph and Kerstin Kleese van Dam

Chapter 7: User Facilities and Research Infrastructure

Chair/Lead Writer: Andrzej Joachimiak
Co-Chair/Facilitator: Susannah Tringe
Writers: Kirsten Hofmockel, Paul Langan, David A. Randall, Daniel Segrè, and Minghua Zhang

Chapter 8: Emerging Technologies

Chair/Lead Writer: Andrzej Joachimiak
Co-Chair/Facilitator: Susannah Tringe
Writers: Kirsten Hofmockel, Paul Langan, David A. Randall, Daniel Segrè, and Minghua Zhang

Report Preparation

This report was edited and prepared for publication by the Biological and Environmental Research Information System group at Oak Ridge National Laboratory (Kris Christen, Holly Haun, Brett Hopwood, Sheryl Martin, Stacey McCray, Marissa Mills, Judy Wyrick, and Betty Mansfield).

Contents

Executive Summary

The Biological and Environmental Research (BER) program within the U.S. Department of Energy (DOE) Office of Science supports research focusing on the interconnections between energy production and the living environment. This fundamental research, conducted at universities, DOE national laboratories, and research institutions across the country, explores organisms and ecosystems that can influence the U.S. energy system and advances understanding of the relationships between energy and environment from local to global scales.

Research within BER can be categorized into biological systems and Earth and environmental systems. Biological systems research seeks to characterize and predictively understand microbial and plant systems using genomic science, computational analyses, and experimental approaches. Foundational knowledge of the structure and function of these systems underpins the ability to leverage natural processes for energy production, including the sustainable development of biofuels and other bioproducts. Earth and environmental systems research seeks to characterize and predictively understand feedbacks between Earth and energy systems and includes studies on atmospheric physics and chemistry, ecosystem ecology and biogeochemistry, and development and validation of Earth system models extending from regional to global scales. To promote world-class research in these areas, BER supports user facilities that enable observation and measurement of atmospheric, biological, and biogeochemical processes using the latest technologies.

BER regularly solicits input from the scientific community to help guide its programs. The Biological and Environmental Research Advisory Committee (BERAC) is chartered under the Federal Advisory Committee Act to advise BER on its research portfolio and user facilities. To facilitate a synthesis of community input, the director of DOE's Office of Science charged BERAC in March 2016 to review research progress and establish and deliver a revised long-term vision for BER by fall 2017. Questions considered during this process included:

- To what extent has BER successfully met the challenges outlined in the 2010 report, *Grand Challenges for Biological and Environmental Research: A Long-Term Vision*?
- What are the greatest scientific challenges that DOE faces in the long term (20-year horizon), and for which of these should BER take primary responsibility?
- How should DOE position BER to address these challenges?
- What new tools should be developed to integrate and analyze data from different disciplines?
- What unique opportunities exist to partner with, or leverage assets from, other programs within the DOE Office of Science?
- What scientific and technical advances are needed to train the future workforce in integrative science, including complex systems science?

Through a series of BERAC meetings, white papers, and a research community workshop, BERAC addressed these questions, identifying future grand challenges in five areas: biological systems, Earth and environmental systems, microbial to Earth system pathways, energy sustainability, and data analytics and computing. Providing critical support for these challenges are BER user facilities, research infrastructure, and emerging technologies. This report represents a synthesis of these grand challenges and the supporting facilities and technologies.

Grand Challenges in Biological Systems Science

Biological systems naturally transform and distribute energy through photosynthesis and subsequent decomposition of photosynthetic material. Conversion of energy to biomass can occur via various mechanisms, including photosynthetic and chemical pathways with oxygen (i.e., aerobic) and without oxygen (i.e., anaerobic). Greater insights are needed into the regulation of these pathways, the genes responsible for the reactions, and environmental influences on the reactions. This improved understanding is a precursor to enabling changes in pathways that may uncover new or more efficient energy sources.

BER facilitates the understanding of complex biological systems by supporting research that develops and uses high-throughput genomic, proteomic, metabolomic, and computational approaches as well as more traditional advanced experimentation. In the past 10 years, the program has made significant progress in sequencing and synthesizing nucleic acids, rapidly generating high-quality omics datasets, advancing understanding of single-cell metabolism and mechanisms of cellular regulation, and devising approaches to manipulate cellular processes. Research has expanded systems approaches and model organism status to bacteria, fungi, and plant systems increasingly relevant to DOE missions. In these endeavors, BER has partnered with other DOE Office of Science user facilities (e.g., light and neutron sources) and with other federal agencies, including the U.S. Department of Agriculture, National Science Foundation, and National Institutes of Health, to leverage different areas of expertise across the government.

Biological Systems Science Action Items

- Conduct experiments that enhance cooperation among BER-supported user facilities and other DOE user facilities (e.g., DOE Nanoscale Science Research Centers).
- Lead coordinated efforts to improve and validate genomic annotation approaches.
- Improve the performance of metabolomics approaches for BER-relevant science.
- Establish standards across data platforms so investigators can efficiently link genomes with phenotypes.
- Coordinate and align research to understand dynamic linkages and feedbacks between environmental conditions and complex biological systems.

Building on and emerging from these successes are a number of grand challenges for the next decade or longer including the following:

2.1 Understand the biological complexity of plant and microbial metabolism and interfaces across scales spanning molecules to ecosystems.

2.2 Develop technologies to identify DOE mission–relevant metabolic capabilities and engineering possibilities in bacteria, fungi, archaea, viruses, plants, and mixed communities.

2.3 Optimize the use of large datasets that integrate omics surveys with biochemical and biophysical measurements to generate knowledge and identify biological paradigms.

2.4 Understand the links between genotype and phenotype in single but very diverse organisms and in communities of organisms that interact in terrestrial ecosystems.

2.5 Effectively exploit new and emerging technologies in systems biology and physical measurements (e.g., miniaturization) to accelerate biological discoveries.

Grand Challenges in Earth and Environmental Systems Sciences

Research on Earth and environmental systems includes terrestrial and atmospheric systems as they relate to Earth's natural and human-driven energy systems. The study of environmental systems often includes biotic (living) and abiotic (nonliving) variables. Earth and environmental systems science seeks to understand and model the interactions among natural systems, human systems, and biotic and abiotic variables that comprehensively represent observed systems at local to global scales.

This research has advanced greatly over the past decade. Earth system models, of which climate is one component, have benefited from the use of high-performance computing. Moreover, targeted environmental science activities, particularly in the Arctic, have improved the understanding of ecosystem dynamics in traditionally understudied regions, thereby helping to reduce uncertainty in Earth system models. Particular successes include improved high-resolution aerosol and cloud modeling; advances in research on abrupt tipping points (e.g., thawing permafrost and melting sea ice); better coupling of human-based biogeochemical drivers with Earth system models; and quantification of model uncertainties. To make these advances, BER has partnered with other DOE Office of Science programs (e.g., Advanced Scientific Computing Research) and has coordinated research activities with other federal agencies through participation in the U.S. Global Change Research Program.

Earth and Environmental Systems Sciences Action Items

- Develop fine-scale process models and data assimilation capabilities for data-model integration to advance model development and prediction of Earth system processes.
- Advance high-resolution modeling in different simulation and prediction modes supported by exascale computing to improve understanding and prediction of extreme or high-impact events.
- Develop a computational user facility for rapid design, generation, evaluation, and diagnosis of Earth system model simulations, as well as analysis of ensemble predictions and data-model synthesis.
- Develop and maintain a hierarchy of models for hypothesis testing, model development, and uncertainty analysis of the complex human-Earth system.
- Develop and integrate new sensing technologies and optimize field deployments in Atmospheric Radiation Measurement, AmeriFlux Network, Next-Generation Ecosystem Experiments, and other BER programs to explore interactions across different scales of biological organization and biosphere-atmosphere feedbacks.
- Create new integrated field laboratories that target biogeochemical, energy, and water flows between urban areas and surrounding ecosystems.

From this work, new and future challenges include:

3.1 Advance Earth system modeling using a hierarchy of models, from process-resolving coupled models to reduced-order models, to transform understanding of the coupled Earth system and to produce useful and credible simulations and predictions of Earth system behavior at multiple time scales.

3.2 Establish new observational technologies and use them to understand human and Earth system processes, such as land-atmosphere interactions, biogeochemical cycles, and subsurface soils, to estimate critical process parameters using novel analysis methods, such as machine learning and data science, and to quantify model errors.

3.3 Advance basic knowledge and scale-aware simulation capability for Earth system feedbacks associated with aerosols and moist processes to better quantify aerosol forcing, precipitation changes, and extreme events with consequences for energy and water cycles, global distribution of nutrients, and human health.

3.4 Advance modeling and understanding of important ecological, biological, and carbon cycle interactions and feedbacks in the climate system to identify potential tipping points and possible energy strategies.

3.5 Characterize, understand, and model the complex, multiscale water cycle processes in the Earth system including the subsurface to understand and predict water availability and human system response to extremes.

3.6 Understand the time-dependent processes and mechanisms associated with melting glaciers, ice caps, and ice sheets and their contributions to regional sea level rise.

3.7 Quantify the interplay between internally generated climate variability and externally forced response involving anthropogenic and natural factors and their relative roles in the time evolution of regional variability to understand predictability of the Earth system.

3.8 Understand the long-term Earth system stability in response to possible future Earth system outcomes and address the level of confidence and identify emergent constraints for the range of model projections.

Grand Challenges in Microbial to Earth System Pathways

Microbial communities influence soil and plant systems that in turn affect regional and global environments. However, research at one scale does not always translate to other scales because of challenges in data formats, relevance at different scales, and fundamental understanding of the links among scales. The 2010 BERAC Grand Challenges report indicated that BER was well positioned to undertake the grand challenge of coordinating research and analyzing results across scales. Since this time, advanced capabilities have enabled researchers to functionally and rapidly annotate microbial genomes using high-throughput technology and assemble microbial

Microbial to Earth System Pathways Action Items

- Develop and deploy sensors suitable for use in the field that integrate the effects of microbial activity at the ecosystem scale.
- Promote integrated studies that explicitly test predicted microbial network interactions and attempt to assess membership and species-specific and collective functional capabilities within ecologically coherent microbial communities.
- Conduct experiments that help determine the influence of microbial processes at larger, aggregate scales.
- Promote research teams to integrate microbial community dynamics into Earth system models.

communities in the laboratory to represent fundamental biotic interactions and systems-level processes. However, development of more complex model communities is needed to better capture relevant environments and associated biological processes at multiple scales. To address this need, a number of grand challenges were identified:

4.1 Characterize the biogeochemical exchanges driven by food web and plant-microbe interactions and evaluate their process-level impacts, sensitivity to disturbances, and shifting resource availability under changing environmental regimes.

4.2 Define the sphere of influence and key elements of microbial communities in space and time relevant for predicting larger-scale ecosystem phenomena for Earth system understanding.

4.3 Integrate molecular and process data to improve the ability to define ecologically significant traits of individual taxa and communities and use trait-based models to develop predictive links between community dynamics and ecosystem processes.

4.4 Align and deepen connections among conceptual understanding, measurements, and models related to the roles of microbes in determining the rate of transformation, uptake, and loss of chemical elements from ecosystems.

Grand Challenges in Energy Sustainability

Energy production and use are inherently connected to land, air, and water resources. Comprehensively understanding these interactions is therefore important for guiding current and future energy production that will optimize energy availability and environmental quality. In recent years, research on bioenergy conversion and associated environmental considerations has progressed substantially, accompanied by an increased understanding of energy-food-environment tradeoffs and improved characterization of spatial and temporal variabilities of targeted ecosystems. Other significant advances include further development of integrated assessment models, climate models, integrated Earth system models, and the coupling of these models where appropriate to fully address sustainability science questions. Moving forward, four grand challenges will take this research into the next decade and help resolve important questions:

5.1 Further develop the science of coupling energy, water, and land use across different spatial and temporal scales to understand environmental impacts and changing climate and to better predict net biogeochemical fluxes.

5.2 Use observational, experimental, and model-based approaches to explore the sustainability of alternative energy systems, incorporating stability and resilience analysis, uncertainty, transition paths from current

Energy Sustainability Action Items

- Establish a strategically distributed network of energy sustainability testbeds for addressing crucial research questions associated with specific energy strategies and air-water-land feedbacks at multiple scales. The combined testbeds will address several grand challenges while leading to a predictive understanding of couplings between energy systems and natural systems.
- Create an energy sustainability modeling and synthesis center for multidisciplinary teams to address key energy-water-land research challenges. The center will organize multidisciplinary teams to understand impacts (e.g., energy-water-land influences on energy infrastructure), sustainability working groups to resolve short-term research questions, and also facilitate and house integrated data products at resolutions needed for sustainability analyses.

infrastructures, and the use of appropriate common metrics.

5.3 Understand how variability and change in natural systems affect energy system structure and function and determine how best to build this knowledge into models.

5.4 Create new data streams and more effectively use existing observations to ensure the availability of scale-appropriate data, particularly related to high-resolution land use, landscape infrastructure, demographic change, and energy-land-water research.

Grand Challenges in Data Analytics and Computing

Data challenges in BER research programs have increased by orders of magnitude over the past few years. New techniques and services are required to leverage the wealth of research results and transform them into world-leading scientific discoveries. Although the data differ in format, common approaches can be employed among BER programs for archiving, accessing, processing, and generating enhanced data products.

Over the past decade, further progress has been made in developing archives, developing software to handle large amounts of program-specific data, and linking BER researchers with high-performance computing capabilities. However, the growth in data size, complexity, and heterogeneity has far outpaced these efforts, and the challenges are continually escalating with the development of new instruments and sensors that stream data in real time. As a result, future grand challenges are expected to include:

6.1 Develop robust approaches for large-scale data collection, curation, annotation, and maintenance.

6.2 Develop computing and software infrastructure to enable large-scale data (i.e., Big Data) storage and analysis.

6.3 Conduct research to develop suitable algorithms and programming models that can harness current and future computer architectures to effectively model complex coupled systems and analyze extreme-scale data.

6.4 Engineer advanced computational modeling combined with data integration across temporal and spatial scales.

6.5 Conduct research and develop activities that support human understanding of large-scale, multimodal data streams, including the ability to steer experiments in real time.

Supporting Grand Challenges Through the Use of Facilities, Research Infrastructure, and Emerging Technologies

Dedicated facilities and new or improved technologies often can increase the speed and efficiency with which grand challenges are met. DOE, and BER specifically, invests in user facilities and the development and application of technologies to advance

Data Analytics and Computing Action Items

- Prioritize building and maintaining an integrated facility for long-term data storage, archiving, and data analysis.
- Evolve current and develop new ontologies and metadata standards, as well as strengthen compliance in funding announcements and reporting standards.
- Facilitate and encourage increased interactions across disciplines (e.g., computer scientists, statisticians, data analysts, and experimentalists).
- Foster coordination and collaboration within DOE and among other agencies.

scientific understanding. As BER research challenges evolve, so too will capabilities within BER and Office of Science user facilities. Potential opportunities to further enable facilities and emerging technologies to support BER research and identified grand challenges are described here.

Facilities and Research Infrastructure

BER fundamental science is empowered by BER, other Office of Science user facilities, and experimental and observational infrastructure that integrate different technologies to enable state-of-the-art research in plant, microbial, environmental, and Earth system sciences. User facilities are intended to provide unique research capabilities that are beyond the ability of single or small groups of investigators to manage for use by national and international communities. Distributed across the DOE complex, national user facilities provide a concentrated portfolio of research technologies, methodologies, instruments, and computational hardware that would be too expensive or sophisticated to duplicate in multiple locations. These facilities will continue to enable scientific breakthroughs in genomics, imaging, structural characterization, and atmospheric research only if their technologies are cutting edge. Efforts to improve current capabilities are thus critical, along with the continued development of advanced new technologies and their deployment in existing or new user facilities. These resources can enable key experiments needed to understand processes important to BER-aligned research efforts. Periodic assessments of the potential need for new or modified facilities to meet growing and changing BER research requirements are prudent.

Recommended improvements in capabilities and access to BER and Office of Science user facilities include:

7.1 Foster a spirit of collaboration to enable integrative capabilities among BER and Office of Science user facilities, as well as other federal research facilities and infrastructure, thereby promoting a fully interdisciplinary approach to BER-relevant science.

7.2 Solicit input from the BER research community regarding researchers' needs and train them in new experimental, observational, and modeling approaches, thus propagating the knowledge and skills for generating high-impact scientific results.

7.3 Develop innovative enabling technologies and construct and acquire state-of-the-art instruments that exploit the world-leading characteristics of each user facility. This will boost capabilities for basic research in biological systems and Earth and environmental systems science, thereby providing DOE and the nation with leading-edge capabilities for biological and environmental science.

7.4 Develop multimodal imaging and remote-sensing capabilities at user facilities for interrogating length scales ranging from atomic to mesoscale and time scales ranging from nanoseconds to days.

7.5 Build upon existing investments and capabilities at the DOE Office of Science light and neutron science user facilities, continuing to align them with BER missions.

7.6 Further develop the necessary infrastructure at user facilities to study organisms in their natural habitats.

7.7 Develop and adopt technologies to convert genome sequence data into functional understanding at appropriate BER user facilities.

In addition to these recommendations, BERAC reviewed the grand challenges with respect to existing or potential links to BER and Office of Science user facilities and experimental infrastructure. A summary matrix of grand challenges and BER user facilities (see Table ES-1, p. xvii) indicates that (1) BER grand challenges require a diverse set of capabilities within user facilities; (2) coordination and partnerships across facilities are necessary for addressing complex modeling, data management, and data analysis priorities; and (3) some challenges are more readily met by existing user facilities than others, potentially requiring additional consideration for filling existing capability gaps.

Emerging Technologies

Emerging technologies play a key role in state-of-the-art research. Experimental and modeling needs can drive technology or software development. Conversely, existing technologies can be brought together in unique ways to explore existing or new research questions. Suggested development of technologies or capabilities include:

8.1 Characterize the genotype and phenotype of individual cells, including genomics, transcriptomics, proteomics, and metabolomics, to enable high-resolution predictive biology.

8.2 Increase throughput and integration of genomics, transcriptomics, proteomics, and metabolomics to enable improved translation from the molecular to cellular realm.

8.3 Characterize key processes of aerosols, clouds, turbulence, atmospheric convective systems, and their interactions to enable better modeling of water, carbon, and energy cycles of the Earth system and predictions of future climate change and extremes in atmospheric, terrestrial, and watershed events.

8.4 Integrate data covering broad time and length scales—from seconds to years and from Ångströms to the Earth scale—to enable multiscale comprehension and simulation.

8.5 Develop integrative and interpretive computational approaches that can handle large, disparate data types from multiple and heterogeneous sources using advanced and exascale computing.

8.6 Explore the development and application of nanoscience and nanotechnology to further understand biological systems at nanoscales or larger scales (e.g., nanosensors).

8.7 Explore the development and application of microsatellites and unmanned aerial systems as mechanisms for observation of atmospheric and terrestrial variables.

Common Actions

Understanding of the living environment and its association with energy cycles, production, and use continues to develop over time, along with new capabilities and technologies. Since the previous review of BER grand challenges in 2010, biological and environmental sciences have progressed substantially. For example, scientific understanding of single-cell metabolism, microbial communities, and associated systems-level processes has advanced significantly, as has the use of bacteria, fungi, and plants as model systems to address DOE missions in energy production and use. Concurrently, targeted experiments in understudied regions (e.g., the Arctic) have filled knowledge gaps in Earth system models. These models have been coupled with integrated assessment models, and their complexity and resolution have increased through the use of high-performance computing. The interaction between biological and environmental sciences also has fostered progress in bioenergy conversion, energy-environment interactions, and increased capabilities to address sustainability science questions.

With increased understanding of research areas comes increased complexity and the need to integrate and link data streams and model parameters that help build predictive capabilities. This need was reiterated across identified grand challenges, along with other common actions that would facilitate progress on BER grand challenges. Common actions identified among grand challenges include:

- Promote integrative science across BER programs, such that research disciplines benefit more from one another.
- Establish standards for data and metadata across platforms and research disciplines.
- Identify and integrate key components (e.g., molecules, microbial processes, and Earth

Table ES-1. Capability Assessment of Biological and Environmental Research (BER) User Facilities to Address BER Grand Challenges*

Grand Challenge Areas	Grand Challenges**	Atmospheric Radiation Measurement Research Facility***	Environmental Molecular Sciences Laboratory***	Joint Genome Institute***
Biological Systems	2.1 Metabolism	–	●	●
	2.2 Engineering	–	●	●
	2.3 Large data	–	●	●
	2.4 Geno-pheno	–	❍	●
	2.5 Technologies	–	●	●
Earth and Environmental Systems	3.1 Models	❍	–	–
	3.2 Observations	●	❍	–
	3.3 Aerosols	●	❍	–
	3.4 Tipping points	❍	❍	❍
	3.5 Water cycle	❍	❍	–
	3.6 Ice dynamics	–	–	–
	3.7 Forcings	❍	–	–
	3.8 Stability	–	–	–
Microbial to Earth System Pathways	4.1 Biogeochemistry	–	●	●
	4.2 Communities	–	❍	❍
	4.3 Traits	–	❍	❍
	4.4 Microbes	–	●	❍
Energy Sustainability	5.1 Coupling	❍	❍	–
	5.2 Alternatives	–	–	–
	5.3 Variability	–	–	–
	5.4 Scales	–	–	–
Data Analytics and Computing	6.1 Large-scale	●	❍	●
	6.2 Infrastructure	❍	●	●
	6.3 Algorithms	–	❍	❍
	6.4 Integration	❍	❍	–
	6.5 Real-time analysis	–	–	❍

* See Chapter 7 for an expanded table that includes other Office of Science user facilities and other BER research infrastructure.

** See text in respective chapters for descriptions of individual grand challenges.

*** Existing capabilities and partial capabilities that could be leveraged for the grand challenges are denoted by ● and ❍, respectively.

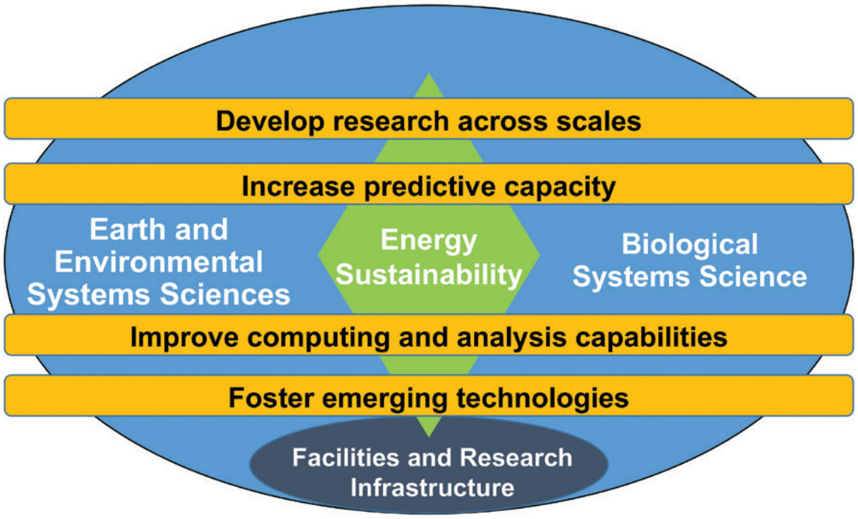


Fig. ES-1. Current Biological and Environmental Research (BER) program research focuses on biological systems science and Earth and environmental systems science, which can be used to address integrative research topics including energy sustainability. In addition to these core research areas, the Biological and Environmental Research Advisory Committee has identified grand challenges in conducting research across scales, developing predictive capacities, and further developing computational and technological capabilities in support of BER research. U.S. Department of Energy facilities and research infrastructure provide a foundation for state-of-the-art, complex research.

system parameters) that are most significant and impactful to DOE missions.

- Identify and develop links among different scales of analysis to obtain more comprehensive representation of individual and system dynamics.
- Increase predictive capabilities of Earth and environmental systems and biological systems research.
- Explore the use of remote sensors at different scales and within different disciplines to promote greater understanding, measurement, and evaluation of research findings.
- Promote interactions among researchers that more strongly link experiments, scientific theories, observation infrastructure, and modeling.

These common actions resonated across science focus areas and therefore were suggested as cross-cutting activities for BER (see Fig. ES-1, this page). The ability to conduct complementary and integrative research across different scales of space and time, along with understanding complex systems for predictive purposes, is a strength that can be found within the combined research and facilities network of DOE's Office of Science. The research challenges and suggested actions documented in this report are intended to help guide BER over the next decade or more. The information provided herein will need to be balanced with DOE mission objectives, research diversity, and available funding to determine which challenges are ultimately met. Also anticipated is that any implementation of these challenges would require additional planning and organization through workshops and research community engagement, along with continued partnerships and coordination with other federal programs.

1 Introduction

The Biological and Environmental Research (BER) program within the U.S. Department of Energy (DOE) Office of Science supports transformative science and scientific user facilities to achieve a predictive understanding of complex biological, Earth, and environmental systems for energy and infrastructure security and resilience. This research, conducted across DOE national laboratories, universities, and research institutions, focuses on interconnections between energy and the environment. The program aims to understand fundamental biological, biogeochemical, and physical principles to be able to predict processes occurring at scales ranging from the molecular and genomics-controlled smallest scales to environmental and ecological processes at the scale of planet Earth.

Starting with the genetic information encoded in organisms' genomes, biological research, housed within BER's Biological Systems Science Division, seeks to discover the principles that guide translation of the genetic code into the functional proteins and metabolic and regulatory networks underlying the systems biology of plants and microbes as they respond to and modify their environments. This predictive understanding will enable the design and re-engineering of microbes and plants for improved energy resilience and sustainability, including advanced biofuels and bioproducts, enhanced carbon storage capabilities, and controlled biological transformation of materials such as nutrients and contaminants in the environment. Earth and environmental systems research, housed within BER's Climate and Environmental Sciences Division, advances fundamental understanding of the dynamic, physical, and biogeochemical processes required to systematically develop and validate Earth system models (ESMs) that integrate across the atmosphere, land masses, oceans, sea ice, and subsurface. These ESMs are required for predictive tools and approaches needed to inform future energy and resource needs. To facilitate world-class research in these areas, BER also supports three user facilities—Joint Genome Institute, Environmental Molecular Sciences Laboratory, and Atmospheric Radiation Measurement Research Facility—that enable observations and measurements of biological, biogeochemical, and atmospheric processes using the latest technologies.

History of BER Science

BER has a long and productive history of research dating back to the Atomic Energy Act passed by Congress in 1946, which established the Atomic Energy Commission (AEC). A mission of AEC, which became the Department of Energy in 1977, was to understand the effects of radioactive material on human health and other living matter. This endeavor led to research on plants, ecosystems, atmosphere, and biology from cellular to organismal levels to understand the impacts of radiation exposure in the early period of the Cold War. From this research, atmospheric transport models evolved, which later developed into climate models and then the ESMs in use today. Likewise, BER research on understanding DNA replication during cell division developed into the international Human Genome Project and launched a new era of modern biotechnology and genomics-based systems biology. The significant and substantial impact that BER has had on the fundamental understanding of biotic and abiotic processes at the organismal to Earth scale has been profound, and it was made possible through long-term planning and investment in fundamental research, as well as coordination with other U.S. governmental agencies, universities, and industry.

Defining New BER Grand Challenges

Input from the science community to the planning process is essential for guiding world-class research. Such scientific input to BER is provided throughout each annual funding cycle via conferences, workshops, principal investigator meetings, program reviews, and proposal panel reviews. BER also receives mission guidance via the Executive Administration's Office of Science and Technology Policy and Congress, either directly or via multiagency administrative working groups and congressional committees.

To synthesize the scientific input for identification of long-term grand research challenges, the Biological and Environmental Research Advisory Committee (BERAC) convened a subset of the community in 2009 to review the status of research related to BER and recommend future research challenges that BER may consider. A report from this effort was published in 2010 (BERAC 2010). BER has accomplished much since then, and the research challenges have changed based on the progression of research and technology. As a result, the director of DOE's Office of Science charged BERAC in March 2016 to review BER's progress and establish a revised long-term vision for delivery in fall 2017 (see Appendix A: Charge Letter, p. 117; Appendix B: Workshop Agenda, p. 119; and Appendix C: Workshop Participants, p. 121).

Questions considered during this process included:

- To what extent has BER successfully met the challenges outlined in the 2010 report, *Grand Challenges for Biological and Environmental Research: A Long-Term Vision*?
- What are the greatest scientific challenges that DOE faces in the long term (20-year horizon), and for which of these should BER take primary responsibility?
- How should DOE position BER to address these challenges?
- What new tools should be developed to integrate and analyze data from different disciplines?
- What unique opportunities exist to partner with, or leverage assets from, other programs within the DOE Office of Science?
- What scientific and technical advances are needed to train the future workforce in integrative science, including complex systems science?

This report represents a synthesis of future grand challenges related to these questions and is organized with respect to the following topic areas: Biological Systems Science, p. 3; Earth and Environmental Systems Sciences, p. 21; Microbial to Earth System Pathways, p. 43; Energy Sustainability, p. 57; Data Analytics and Computing, p. 71; and the role of user facilities and research infrastructure, p. 85, and emerging technologies, p. 95, in supporting these grand challenges.

2 Grand Challenges in Biological Systems Science

Overarching Grand Challenge and 20-Year Vision

Microbial cells, plants, communities, and ecosystems are complex, diverse, and have highly interactive components and, hence, are best studied as systems and at a depth that provides information to support predictive capabilities. Researchers must embrace the complexity of these systems and employ intensive investigation at the interfaces of system components.

Harnessing the Power of Systems Biology for Energy and Environment

Over the past 10 years, the U.S. Department of Energy's (DOE) Biological and Environmental Research program (BER) has led the scientific community in implementing high-throughput approaches to understand the complexity and responses of microbial and plant communities. These approaches are directed toward DOE missions of advancing biofuels generation and understanding terrestrial carbon cycling. Four major themes emerged from the 2017 Grand Challenges workshop that will guide current and future investigations in systems biology over the next 10 years: (1) addressing the challenge of biological complexity, (2) managing and optimizing large-scale data (i.e., Big Data), (3) intensively studying system interfaces, and (4) widening the range of organisms targeted for BER research. These themes are represented in a number of grand challenges and recommended action items for biological systems (see Biological Systems Science Summary of Grand Challenges and Action Items, p. 4).

Embracing Biological Complexity

Although a significant challenge, understanding the complexity of molecular interactions, cellular regulation, and fine-tuning of metabolism in single cells interacting with their environment will be important in developing manageable and tunable cell lines for energy generation and development of a bioeconomy. Future studies will include natural and synthetically engineered systems using combinations of high-throughput sequencing, omics tools, and synthetic biology, together with advanced physical and biochemical measures.

The Biological and Environmental Research Advisory Committee (BERAC) identified biological complexity as a grand challenge in its 2010 Grand Challenges report (BERAC 2010), and biological complexity is expected to remain a challenge over the next 2 decades. One aspect of the tremendous recent progress is that the full extent of biological complexity is coming into sharper focus, primarily resulting from the application of new technologies and increased abilities toward understanding (i.e., comprehending) the large amounts of data being generated. However, even with this progress, capabilities for integrating, synthesizing, and applying this knowledge to enable "predictive biology"—BERAC's overarching vision for systems biology—clearly are still in a very early stage.

Biological Systems Science

Summary of Grand Challenges and Action Items

The overarching theme of BER systems biology efforts is to *enable predictive biology*. To be predictive, not only do major elements (e.g., regulatory molecules, pathways, and keystone organisms) and their functions (i.e., the "parts list") need identification, but also needed is the understanding of their interactions and how those interactions are regulated at the interfaces. These interactions must be understood well enough to enable predictions of their responses under different environmental perturbations (e.g., pH, ionic strength, water, nutrients, and temperature). The immense difficulty with this theme is that environmental conditions are diverse, complex, and interactive in their own right, and they elicit major reactions and responses (often conflicting) from single microbes, microbial populations, communities, and terrestrial ecosystems. With a long-term vision of enabling predictive biology, investigating these biological systems across multiple environmental scenarios and scales is crucial; gleaning only the potential responses to stimuli is not sufficient.

The five BER grand challenges for systems biology, outlined here, are described in more detail in the Grand Challenge Research Recommendations section, p. 9.

2.1. Understand the biological complexity of plant and microbial metabolism and interfaces across scales spanning molecules to ecosystems.

2.2. Develop technologies to identify DOE mission–relevant metabolic capabilities and engineering possibilities in bacteria, fungi, archaea, viruses, plants, and mixed communities.

2.3. Optimize the use of large datasets that integrate omics surveys with biochemical and biophysical measurements to generate knowledge and identify biological paradigms.

2.4. Understand the links between genotype and phenotype in single but very diverse organisms and in communities of organisms that interact in terrestrial ecosystems.

2.5. Effectively exploit new and emerging technologies in systems biology and physical measurements (e.g., miniaturization) to accelerate biological discoveries.

To begin addressing these grand challenges, the community identified five immediate action items.

- Conduct experiments that enhance cooperation among BER-supported user facilities and other DOE user facilities (e.g., DOE Nanoscale Science Research Centers).
- Lead coordinated efforts to improve and validate genomic annotation approaches.
- Improve the performance of metabolomics approaches for BER-relevant science.
- Establish standards across data platforms so investigators can efficiently link genomes with phenotypes.
- Coordinate and align research to understand dynamic linkages and feedbacks between environmental conditions and complex biological systems.

Managing and Optimizing Big Data

Integrated studies resulting in large datasets and computer-intensive analyses are posing a Big Data crisis because generating data is easier than inferring functional activities or deriving knowledge from the integration of disparate datasets. Innovative approaches are needed to organize, decipher, and annotate datasets and make them publicly available to the research community. This effort will require advances in computational capabilities, as well as experiment-based provision of robust and accurate functional inferences that provide the foundation for predicting and assigning molecular functions to genes and their proteins.

The transformation of biology from an observational to a data-intensive predictive and computational science is accelerating, raising numerous challenges. Among them are the processing of Big Data and completeness of the data being generated, as well as their veracity and reproducibility. The utility (e.g., ease of use and cost) of many high-throughput technologies rapidly has increased to the point that they are employed almost universally, with a concomitant increase in required data generation and analysis. However, not all technologies are advancing at the same high rate. Improvement in metabolomics, for example, is needed to ensure that the provided coverage approaches that of other technologies (e.g., transcriptomics). The lack of completeness among the various datasets generated, as well as questions of spatial and temporal scales in biological systems, largely limits the ability to integrate dissimilar data, making them much less useful for generating new knowledge. These are all critical issues with regard to integrating and, ultimately, deploying this information to support BER's mission areas (see also Chapter 6 on Data Analytics and Computing, p. 71.)

Investigating System Interfaces

Understanding complex microbial cells, plants, communities, and ecosystems at a depth that enables predictive capabilities requires intensive investigation at system interfaces. This research spans interfaces across vast scales of interaction: (a) *at the molecular and cellular scale* between types of molecular measurements, between molecules (i.e., genes, proteins, enzymes, and substrates), and between cellular structures; (b) *at the cell-environment and cell-cell scales*, including environmental and interorganism signaling and regulation; and (c) *at the organism-organism scale* through tracking the metabolic interplay among different organisms (e.g., bacteria, fungi, and plants) that regulate cascades of metabolic responses to changing environments. Synthetic biology would be used to generate and test hypotheses based on this data collection to make informed predictions (see sidebar, Synthetic Biology, this page).

Synthetic Biology

As defined in the 2010 Grand Challenges report, synthetic biology is the design and construction of new biological parts, devices, and systems; the redesign of existing, natural biological systems for useful purposes; and the inspiration for new materials and processes derived from the design, functions, and products of living cells. While systems biology generates hypotheses based on data collection over the whole system, it also enables testing of these hypotheses by direct manipulation of the system's parts. Like many of the most exciting advancements in science, synthetic biology arose at the interface of converging disciplines, specifically engineering and biology. Indeed, one aspect of synthetic biology is the application of engineering design principles to biological systems. Although these aspects were the drivers for its development, synthetic biology is quickly outstripping its beginnings and developing into a powerful technology that will transform biology, with profound carry-over effects to society in general.

Expanding Research Beyond Model Organisms

Much information has been generated from a few model organisms over recent years. However, BER research is branching out to multiple, very different organism hosts for biochemical synthesis and process management. The emphasis is turning to bacterial, algal, and fungal hosts from soils and other environments that have innate abilities to withstand harsh chemical environments and have high-efficiency carbon substrate decomposition abilities (e.g., biofuels research). These hosts often are more difficult to grow and manipulate in the laboratory and require design and development of new engineering and metabolic regulation approaches.

In addressing climate change response and feedbacks, filamentous fungi are critical for managing the direction and flow of carbon in the terrestrial environment, but detailed knowledge of their metabolism is lacking. For both these applications, the understanding of plant metabolism, breeding and maintenance of introduced traits, and interactions with bacteria and fungi is critical. Extrapolating metabolic paradigms from model bacterial systems is not possible, and investigations must focus on organisms that are actually in environments of DOE mission interest. At the same time, an emphasis should be placed on leveraging the foundational knowledge that can be gained quickly with laboratory-deployed organisms, maximizing the implementation of this knowledge to accelerate understanding and guide design of experimentation in environmental strains. To maximize the knowledge gained, needs include prioritizing organisms by mission relevance and sharing information across laboratories.

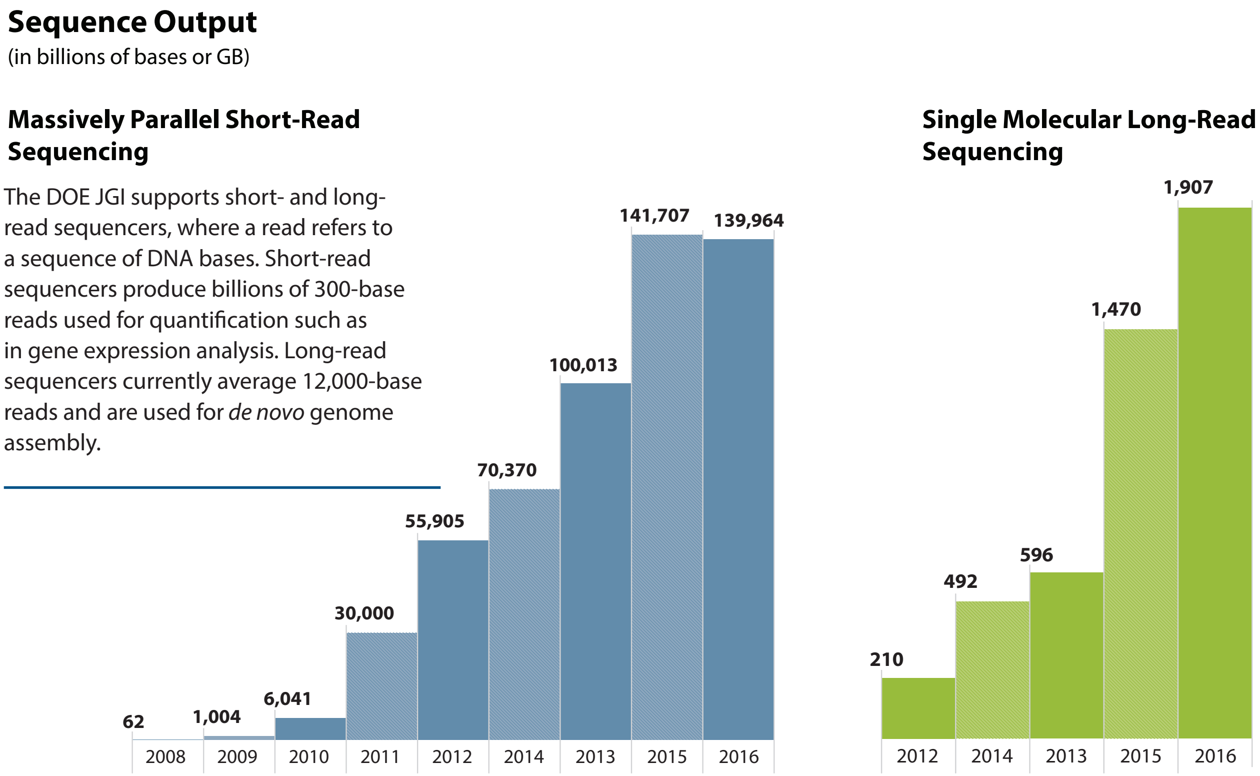
In summary, BER has led and will continue to lead the development and application of systems biology and synthetic biology to address critical needs regarding adaptation to environmental change, enhancing energy sustainability and contributing to the developing bioeconomy. Ultimately, the aim is to contribute significantly to the growth of the bioeconomy. These contributions include defining alternative pathways to energy sustainability for informing the design of future energy systems to answer the question: Can energy pathways be identified that provide optimal economic and environmental outcomes that promote efficient energy production and human well-being?

Building on Progress Related to the 2010 Grand Challenges

Systems biology has seen significant progress in four major areas: (1) sequencing nucleic acids and increasing the ability to rapidly generate high-quality omics datasets; (2) advancing understanding of single-cell metabolism and disentangling the mechanisms of cellular regulation; (3) devising clever schemes to manipulate key cellular processes in single organisms for detection and synthesis (i.e., synthetic biology, see sidebar, p. 5); and (4) expanding systems biology approaches to tackle more DOE mission–relevant bacteria, fungi, and plants as model systems.

Expanded Sequencing and Omics Approaches. Since the 2010 Grand Challenges report, the ability to sequence DNA, RNA, and proteins and generate synthetic molecules has greatly expanded (see also Chapter 8 on Emerging Technologies, p. 95). Data quality and quantity have both increased, and new research approaches and tools have been developed (see Fig. 2.1. Dramatic Increase in Genome Sequencing Throughput, p. 7). Rapidly maturing sequencing subdisciplines are metagenomics, metatranscriptomics, and metaproteomics, with publications now appearing that examine a variety of natural environments, in addition to the well-publicized efforts on animal and human health. In general, however, all these efforts remain at the level of phenomenology, lacking the detailed, mechanistic information needed to develop predictive models of such systems, especially those that link across scales and predict community functions and processes under changing conditions.

Advanced the Understanding of Single-Cell Metabolism and Cellular Regulation. The 2010 report called for research at the single-cell level, and scientists have achieved significant success with a



The DOE JGI supports short- and long-read sequencers, where a read refers to a sequence of DNA bases. Short-read sequencers produce billions of 300-base reads used for quantification such as in gene expression analysis. Long-read sequencers currently average 12,000-base reads and are used for *de novo* genome assembly.

Fig. 2.1. Dramatic Increase in Genome Sequencing Throughput. Over the past decade, new technologies have enabled an exponential increase in the generation, analysis, and comparison of sequencing datasets at the U.S. Department of Energy's Joint Genome Institute (JGI) and other sequencing centers. Today, JGI generates petabytes of high-quality sequence and analysis data. [Image courtesy JGI]

few model systems. Over the past few years, there have been great advances in implementing computationally useful and biologically sound representations of structured metabolic networks in several model organisms at the whole-cell level. In parallel, the growth and increased accuracy of transcriptomic and proteomic datasets have advanced understanding of regulatory networks, potentially adding to the scaffold of genome-scale representations provided by metabolic stoichiometry. However, full integration of metabolic and regulatory networks remains an open and significant challenge, with the incorporation of kinetic information into genome-scale models even more challenging. In addition, incorporating the effects of enzyme promiscuity and cell-to-cell metabolic variability into models remains a key goal.

Moreover, genome annotation (i.e., a detailed characterization of gene function for all genes in the genome) remains a limiting factor in the construction of genome-scale models. Challenges in associating genes with metabolic functions include the capacity to obtain enough metabolomic datasets and interpret their complexity. In addition, molecular complexes are important organizational units in cells, but their structures and functions are not yet understood. The ability to derive initial models based on genomic data has advanced greatly. Yet, the persisting challenge is to integrate multiple data types and multiple levels of biological description (i.e., kinetic, metabolic, and probability models for cellular phenotypes) into unified mathematical frameworks. More effort will be needed to establish pipelines for facilitating data collection and integration toward automated model construction.

Challenges also remain in the area of metabolism, where little progress has been made toward integrating data across scales (e.g., up to the population level). Also lacking is significant progress in the ability to analyze single cells *in situ* in natural environments and to determine *in vivo* kinetics in cells. However, there is currently more recognition of subcellular compartmentalization and cell type variation within a microbial population or within a plant.

Improved Biosystems Design Approaches and Tools. Elegant schemes devised for the discovery and manipulation of regulatory elements are revolutionizing systems biology and enabling the design of synthetic metabolic systems. One such scheme involves clustered regularly interspaced short palindromic repeats, or CRISPRs (see sidebar, A New Era in Precise Genome Editing for Synthetic Biology, this page). Also undergoing great strides is the design of cells and systems as sensors, with a primitive proof of concept indicating the feasibility and reality of metabolic engineering. This area should continue to advance quickly, combining computational protein structure prediction and design algorithms, new molecular engineering techniques, and the collection of in-depth datasets containing cellular performance and processes. In addition to data collection and integration, advances in synthetic biology, such as the ubiquitous implementation of gene editing using CRISPR, are opening new avenues for systems biology research—enabling the engineering of *in silico* and *in vivo* mutants with increasing efficiency and the testing of model capacities to capture the phenotypic effects of genetic perturbations.

Advanced Engineering and Study of Higher-Order Organisms. Expansion of systems biology to a wider variety of DOE mission–relevant organisms is needed to fulfill DOE goals (U.S. DOE 2015). Eukaryotic organisms such as fungi and plants are key players in the DOE missions of biofuels generation and understanding of global carbon cycling. Since the 2010 Grand Challenges report, significant advances have been made in studying and

A New Era in Precise Genome Editing for Synthetic Biology

One of the exceptional benefits of conducting whole-genome sequencing is the potential to discover new regulatory systems in biological organisms that would be missed with single-gene approaches. Since the 2010 Grand Challenges report (BERAC 2010), a technology based on clustered regularly interspaced short palindromic repeats (CRISPRs) has emerged from discovery, through genetic manipulation, to widespread use as a tool for generating organisms with desired metabolic properties (Horvath and Barrangou 2010; Wiedenheft et al. 2012; Lander 2016). CRISPRs originally were identified by their pattern of repeats in metagenomic datasets before their functions were known (Bolotin et al. 2005; Mojica and Montoliu 2016; Pourcel et al. 2005). Associated with CRISPRs is an RNA-guided protein named Cas (CRISPR-associated endonuclease). Cas is a molecular scissor that uses the spacer repeats in CRISPRs to recognize and cut exogenous DNA. CRISPR-Cas systems were first identified in bacterial and archaeal systems where they are now known to function as an acquired immune system against phages. More recently, these systems have been found in eukaryotes. Modifications have led to the CRISPR-Cas9-guide RNA (gRNA) system, which edits genomes at specific sites, inserting a target gene into the genome at a precise location to replace a natural gene (Cong et al. 2013; Mali et al. 2013). Widely applicable, this technology, when multiplexed and democratized, can work successfully with DNA from microorganisms (e.g., bacteria and archaea), model eukaryotic animals, plants, and humans. Since 2013, use of the CRISPR-Cas system has grown exponentially. Its successful implementation in a wide variety of organisms has revolutionized the ability to engineer organisms with desirable metabolic properties. Current applications include microbial engineering, plant and crop breeding, and human and animal gene therapy treatments.

engineering the metabolism of fungi and plants. The cellular and tissue structures inherent in these higher-order organisms add to their complexity, making those structures even more difficult to understand, regulate, and predict than those in bacteria. With baselines now established, disentangling the many levels of metabolism, regulation, and communication within and among eukaryotic cells and tissues may be possible.

Grand Challenge Research Recommendations

Grand Challenge 2.1

Understand the biological complexity of plant and microbial metabolism and interfaces across scales spanning molecules to ecosystems.

Predictive Modeling for Complex Systems

Genomes, transcriptomes, and proteomes contain the full instruction sets for living organisms, but epigenetic processes can modify phenotypic expression. Significant progress over the past 7 years has shown that metabolism in single-cell systems may be understood well enough to enable engineering and manipulation toward production of desired industrial products or phenotypic characteristics (e.g., zip-lignin). Major challenges are to assemble these simple systems into more complex systems more closely resembling reality and to predict their performance by integrating modeling scenarios into biochemical pathways. Genome-scale predictive modeling should be designed for easy extension to other organisms, as well as cellular assemblies and communities of DOE mission relevance (see Fig. 2.2. Overview of the Length and Time Scales Spanning Biological Systems Science, p. 10).

Research Needs and Knowledge Gaps

Dynamically Perturbing Systems and Integrating Results with Measurements, Modeling, and Diverse Data Types. Predictive systems-level modeling will require tight integration of theoretical studies, computer simulations, and experimental measurements that include both omics and imaging metrics. As researchers move from genotype to phenotype, obtaining robust, quantitative data still poses technical challenges. Although genomic, transcriptional, and flux data could provide the basis for genome-scale quantification of an organism's reaction rates, the integration of other layers of biological complexity is needed for accurate description of cellular functions and communication between organisms. The capacity to quantitatively predict functions in biological systems is not limited to static snapshots. To enable accurate understanding of dynamic processes, tools are needed to precisely probe systems in temporally or spatially controlled ways. Tools that introduce perturbations, such as in gene expression or signaling, can be coupled with measurements of outcomes at the transcriptional, translational, and phenotypic levels. Well suited to enable model validation, these methods, with precise perturbations introduced, can test, refine, and revise modeling predictions. Examples of tools that enable perturbations applicable to systems biology include optogenetics, small-molecule inducers, and microfluidics. Overall, the ability to precisely and dynamically perturb a system is invaluable for model validation and will move research beyond a static understanding of cellular processes.

Understanding the Role of Membrane Channels and Transport Proteins in Cellular Function. An additional aspect of biological complexity that must be explicitly introduced in computational descriptions is the hierarchical compartmentalization of systems and subsystems. Biological compartmentalization plays a key role in controlling the flow of molecules and metabolites, whether from outside to inside a prokaryotic cell or between compartments within a eukaryotic cell to form the basis for energy flow within living systems. Though lipids form the matrix of these compartmental barriers, the flow of compounds and ions is mediated mostly by key proteins that vary in nature from simple small-molecule channels to multicomponent protein assemblies. Because integral membrane proteins are difficult to purify and analyze, they are among the least well

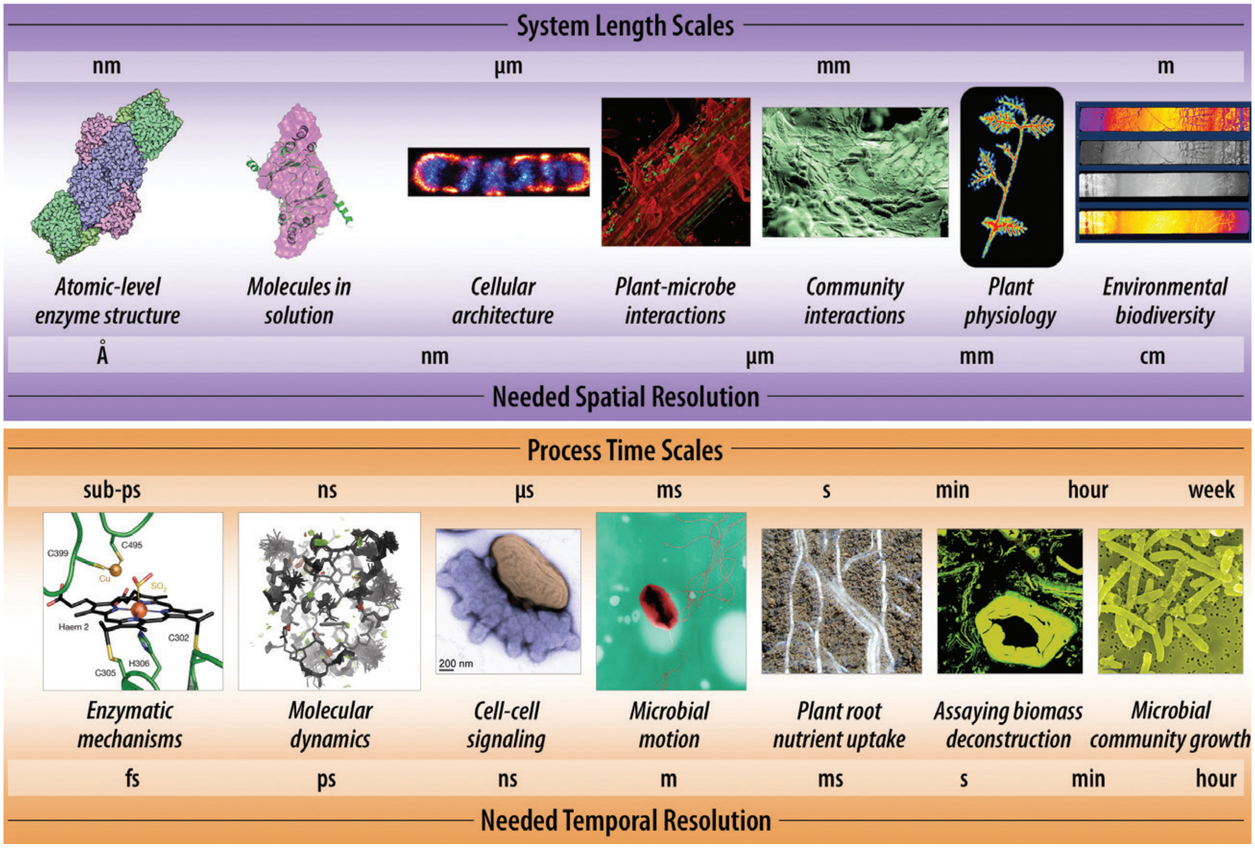


Fig. 2.2. Overview of the Length and Time Scales Spanning Biological Systems Science. Biological systems research includes complex processes that span a size range from Ångströms to a meter and a dynamic range from femtoseconds to a week. Representative processes and their elements are illustrated. **Key:** Å, Ångström; cm, centimeter; fs, femtosecond; m, meter; mm, millimeter; min, minute; µm, micron; µs, microsecond; ms, millisecond; nm, nanometer; ns, nanosecond; ps, picosecond; s, second; sub-ps, subpicosecond. [Figure from BER. 2017. *Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and the Environment*, DOE/SC-0189, U.S. Department of Energy Office of Biological and Environmental Research. See Appendix D, p. 125, to view credits and permissions of individual images.]

characterized protein complements of intra- and intercellular transporters. The serious knowledge gap involving the biochemistry and physiological functioning of these membrane proteins must be filled to create models that can robustly predict cellular behavior. Thus, a key component of this grand challenge is to understand contributions to cellular functions made by membrane channels and transporter proteins.

Expanding Single-Species Experiments and Modeling to Multicellular Systems and Communities. In addition to performing experimental measurements, constructing models, and integrating data and models at the single-species level, expanding the current approaches for these activities will be crucial for studying multicellular systems and microbial communities in complex environments. This expansion will require establishing clear strategies for combining single-organism models into ecosystem-level models and testing predictions at different levels. For example, essential capabilities will include those to measure both intracellular fluxes and metabolic exchanges between different organisms. Because the ecological time scales often overlap with evolutionary time scales, model design should enable simulations of evolutionary processes, from standard mutation and selection dynamics to horizontal gene transfer among species. Extension of current models along these directions will benefit not only from the aforementioned coupling of computer modeling and experimentation, but also from the exploration of completely novel theoretical frameworks. This new knowledge could draw from the experience of diverse quantitative sciences, including theoretical ecology, statistical physics, and applied mathematics.

Grand Challenge 2.2

Develop technologies to identify DOE mission–relevant metabolic capabilities and engineering possibilities in bacteria, fungi, archaea, viruses, plants, and mixed communities.

Leveraging Advances in Synthetic Biology

As a technology, synthetic biology offers dramatic potential to reform the developmental processes of research related to biological systems design and analysis. Up to this point, biology has been limited by what exists in the natural world. For example, expansive sequencing is now providing a reservoir of sequenced genomes, proteins, and regulatory elements, as well as the potential to assemble them in different orders to design new synthesis pathways. However, advances in recoding of microbial genomes, greater understanding of protein structure, and sophisticated gene design tools all have the potential to create innovative, *de novo* pathways, structures, and molecules beyond those currently existing in nature—truly transformational ideas and methods. BER is well positioned to capitalize on these revolutions given its strong background and capabilities in synthetic biology, working with diverse organisms, and exceptional genomic sequencing. Following are descriptions of several specific opportunity areas in which synthetic biology can contribute to BER-relevant research.

Research Needs and Knowledge Gaps

Deploying Synthetic Biology to Efficiently Produce Products. Bacteria, fungi, algae, and plants can be engineered to produce biofuels and bioproducts, offering improvements over conventional, petroleum-based processes in terms of sustainability.

Democratizing Genetic and Omics Technologies for Multiple Species. The goal is to enable genetic manipulation and the use of genomic, transcriptomic, metabolomic, and proteomic analysis methods across a wide and diverse set of species. In the envisioned setting, scientists could manipulate the genomes of environmental isolates and species of BER mission relevance, studying uncultivable microbes *in situ* and broadly applying omics approaches. Achieving this goal requires new technologies that enable tool expansion to reach these targets.

Transforming Recalcitrant Strains. A specific challenge associated with genetic manipulation is the need to transform genetic material in recalcitrant strains. Many methods for transformation exist (e.g., electroporation, chemical approaches, heat shock, and mechanical disruption), but the best approach is often unclear *a priori*. Although technologies like CRISPR have been used in a wide variety of hosts, inserting DNA or proteins into cells is a technical challenge that can limit the use of these technologies. To overcome this recalcitrance, the CRISPR-guide RNA (gRNA) complex potentially can be directly transformed into cells as demonstrated in mammalian cells.

Leveraging tools from nanotechnology will be necessary to further address this challenge, with the subsequent use of high-throughput approaches to screen for successful transformation. For example, microfluidic devices can be used to interrogate the electrical properties required to transform cells in a high-throughput fashion by imposing a gradient of electrical fields on cells (Garcia et al. 2016). High-throughput approaches also may be useful for systematically interrogating an array of different transformation conditions, especially when working with recalcitrant strains. The goal is to enable efficient transformation of genetic material in a wide variety of species, ranging from microbes to plants.

Applying Gene Editing to Diverse Microbial and Plant Species. New methods for genome editing and manipulation hold great promise for use as synthetic biology tools. These methods include CRISPR-associated protein-9 nucleases (Cas9s), transactivation-like effector nucleases (TALENs), zinc finger nucleases (ZFNs), and recombinases (Copeland et al. 2014). These methods enable insertion, deletion, or editing of genomic material. For instance, the technologies can be used to knock out genes, modify specific regions of the genome sequence, or replace promoter regions with alternative sequences. In addition to having DNA-editing capacity, variants on these methods offer the ability to regulate gene expression, such as repressing or activating transcription of a gene of interest (Dominguez et al. 2016).

Workshop participants envisioned applying these methods to diverse microbial and plant species for biological system design. This technological advance will enable precise interrogation of gene function, testing of the role of particular sequence features, and potentially up- and down-regulation of expression in a programmable fashion. Critical for deploying the synthetic biology toolbox to a wide variety of species, these techniques possess the essential ability to manipulate DNA and regulate gene expression.

Accelerating Development of Technologies for the Design-Build-Test-Learn Cycle and High-Throughput Capabilities. Synthetic biology requires the ability to design gene circuits and metabolic networks to achieve production of a desired output such as a bioproduct or biofuel. Engineering these systems and gene circuits is challenging, however, and there may be many different possible designs to consider. For instance, in a microbial biofuel production strain, many genes may encode enzymes required for production, and their order of appearance in a genetic construct and their expression level can dramatically influence final production levels. Therefore, the ability to conduct experiments in a way that enables rapid learning and redesign is critical to the success of synthetic biology for biofuel and biochemical production.

Several technologies can support acceleration of the design-build-test-learn cycle. Technologies that reduce the time or effort involved in conducting experiments are valuable because a particular design round may be only one in a set of many iterations. Examples of technologies that can speed this process include *in vitro* systems for synthetic biology, genome-engineering methods, high-throughput screening or selections, and improvements in methods for cloning or gene synthesis.

In vitro methods such as the cell-free transcription translation (TXTL) system use cell extracts to conduct cell processes (Garamella et al. 2016), but because these approaches can be conducted in a test tube and do not require the use of cells, researchers can avoid steps associated with cell growth and replication. As such, linear DNA obtained through polymerase chain reaction (PCR) or gene synthesis can be used directly as a transcription template. This process reduces the time required to collect data from a few days to a few hours, thus enabling the testing of many design iterations and faster learning of design principles. These *in vitro* methods can be coupled with microfluidics to achieve cell-like conditions, replicating cell division and other processes in a carefully controlled setting (Tayar et al. 2015).

Genome-engineering approaches are powerful tools for synthetic biology and can be used in concert with screening and selection methods to design desired traits (Bao et al. 2016). Examples of successes for metabolic engineering applications include multiplex automated genome engineering (MAGE), CRISPR use for multiplexed genome engineering, and oligo-mediated approaches (Wang et al. 2009; DiCarlo et al. 2013; Ryan et al. 2014; Jakočiūnas et al. 2015). These technologies rapidly introduce genetic variation at the genome level; however, to select or screen for a desired phenotype, they can be coupled to sensor-selector strategies (Dietrich et al. 2013; Yang et al. 2013; Raman et al. 2014). For instance, direct or indirect readout methods can be used with fluorescence-activated cell sorting (FACS) or other high-throughput methods to screen for improved phenotypes. These methods enable rapid engineering of metabolic pathways and are relevant to biofuel and bioproduct production.

Rapidly Synthesizing *De Novo* DNA and Assembling Large DNA Molecules. The ability to synthesize DNA fragments and assemble them into large DNA molecules rapidly and cost-effectively is transforming the way synthetic biology research is conducted (Kosuri and Church 2014; Chao et al. 2015), and BER should continue to encourage efforts in this area. DNA synthesis can greatly advance the ability to perform experiments using genetic material derived from diverse hosts. For example, gene synthesis can optimize codons to enable improved expression in a heterologous host, avoiding codons that are rare in the host of interest. As a second example, DNA synthesis is critical for constructing custom CRISPR single-guide RNAs (sgRNAs), including the use of multiplexed approaches that target multiple genes simultaneously. This area also encompasses a vast quantity of exploitable genome sequence information. Users anywhere in the world can obtain sequence information and synthesized DNA of interest for their particular task.

Grand Challenge 2.3

Optimize the use of large datasets that integrate omics surveys with biochemical and biophysical measurements to generate knowledge and identify biological paradigms.

Maximizing the Wealth of BER Data

Over the past 10 years, DNA sequencing has exploded into multiple platforms for high-throughput, high-quality sequencing from DNA and RNA templates. DOE's Joint Genome Institute (JGI) has taken advantage of these platforms to sequence the genomes of thousands of bacteria and fungi and dozens of plant and algal genomes to broaden phylogenetic references and advance DOE missions in bioenergy and carbon cycling (see also Chapter 7 on User Facilities and Research Infrastructure, p. 85). These technological advances are providing the in-depth datasets needed to disentangle the complexity of individual cellular regulation, metabolic responses to environmental stimuli, and interactions among different microbial communities (i.e., bacteria, fungi, and archaea) and plants in terrestrial ecosystems. Scientists can now test hypotheses about cellular status, community responses, and microbe-plant interactions at resolutions not achievable 10 years ago. Laboratory and field experiments can be interrogated with sufficient technical replication, across time series, and with gradients and combinations of environmental conditions to facilitate statistical analyses. High-throughput sequencing and the ability to synthesize DNA have enabled rapid generation of high-coverage sequence datasets for individual organisms across multiple physiological states, defined mixtures of organisms, natural microbial communities, and microbe-plant interactions.

This expansive sequencing capability presents a new set of challenges for the next 10 years in overcoming the bottleneck of data analysis and interpretation. To maximize the utility of these datasets, three new challenges must be overcome and managed: (1) a simple, managed program of data organization, archiving, and retrieval, both for individual investigative teams and the scientific community as a whole; (2) sequence datasets accompanied by or interfaced with their related metadata, such as chemical and physical measures, cellular conditions, environmental conditions, and references containing the appropriate information; and (3) sequence data that are quality controlled and functionally annotated using appropriate, community-vetted standards and readily available to all (see also Chapter 6 on Data Analytics and Computing, p. 71). The lack of high confidence in annotation is recognized as a significant barrier to enabling predictive biology; however, little progress has been made because of the problem's daunting scale. A potential solution would be to develop capabilities to accelerate discovery and prioritize efforts for providing meaningful, but not particularly onerous, requirements (e.g., veracity of annotation to genes with demonstrated relevance to BER missions). This solution represents an opportunity to integrate breakthroughs in accelerating experimentally grounded gene-function prediction with advanced computational and bioinformatics approaches. Each of these three challenges and potential approaches for achieving success is described in more detail in the following section.

Research Needs and Knowledge Gaps

Organizing, Archiving, and Retrieving Data. With guidance from BERAC's 2010 Grand Challenges report, DOE BER user facilities have begun offering joint proposals. For example, DOE's JGI and Environmental Molecular Sciences Laboratory (EMSL) issued a joint call as part of the Facilities Integrating Collaborations for User Science (FICUS) program, which helps facilitate innovative and efficient use of BER facilities (see Chapter 7 section, Collaborative Research Among Scientists and Facilities, p. 91). Currently, however, users typically partition samples and send them to each facility where they are relabeled according to that facility's convention. This situation exemplifies the critical need for user facility coordination to manage sample integrity and organization. These coordinated datasets need to be housed in a publicly available resource accessible by all DOE researchers and organized in a way that links them to their associated manuscripts. Sophisticated tools must be available with intuitive and easy-to-use web-based interfaces that do not require extensive training in computer science. Moreover, the user facilities could be used as a framework for integrating Big Data approaches across BER mission areas (e.g., biofuels, climate responses, and carbon cycling).

To begin addressing data challenges, BER also might consider supporting a community workshop focusing on the issues associated with the deposition, archiving, retrieval, management, and exploration of large-scale sequencing and related process datasets. Participants, including experts in cell biology and synthetic biology, microbial communities, microbe-plant interactions, and plant metabolism, could devise scenarios to maximize knowledge from existing and future datasets. Critically needed, though, are guidance and a framework for facilitating standard practices across disciplines (see also Chapter 6 on Data Analytics and Computing, p. 71).

Coordinating Sequence Datasets with Their Associated Chemical, Physical, Temporal, and Environmental Treatment Datasets. DOE investigators have studied the responses and feedbacks of terrestrial ecosystems to elevated carbon dioxide, warming, nitrogen deposition, ozone, heavy metals, and organic solvents using the full gamut of BER molecular, chemical, physical, and modeling tools. Over the past 5 years, multiple meta-analyses have emerged in attempts to derive emergent general principles from the wealth of past individual studies. The difficulty has been, and continues to be, that datasets are too sparse and too poorly replicated to identify consistent response trends. This difficulty is due in part to different experimental designs, sampling approaches, reporting requirements, and misidentification of sequences housed in that time's "state-of-the-art" databases. Bringing the datasets from multiple past studies together and providing a robust assessment of their collective findings have been challenging and are resulting in much confusion in the literature. With new technologies in hand, BER can now advance from single-organism investigations across more complex environmental conditions to mixed-community studies toward sustainable production of biofuels and whole-ecosystem assessments of climate change responses and feedbacks. The standardized deposition of all related datasets into public resources is imperative for mining these data efficiently and appropriately and deriving general principles. Specifically needed are new approaches for integrated science designed to aid knowledge mining from omics datasets (e.g., conduct imaging chemistry and physical measurements to phenotype cellular outcomes as necessary). These datasets should be housed in interactive, accessible locations for multiple investigators to enable reanalysis and meta-analysis as desired.

Providing Quality Control and Functional Annotation. The depth of coverage enabled by current and emerging sequencing approaches provides datasets of unprecedentedly high quality. This is true of PCR-based amplicon sequencing, a technology that can interrogate more than 300 samples on a single MiSeq run, and whole-cell or whole-community metagenomic and metatranscriptomic datasets that enable multiplexing of samples to maximize data output. Although current sequencing approaches have multiple technical constraints, the ability to appropriately implement quality control metrics is critical to downstream data analysis. Poor sequences must be culled early in the process to avoid later misinterpretation. Providing quality control standards and approaches is a critical DOE responsibility because most DOE-funded research laboratories can easily generate DNA and RNA sequence datasets. Most universities and national laboratories currently house sequencing facilities. To enable cross analysis of data from different studies and different regions, guidance and standards for all BER-generated data must be provided or at least presented concisely with the datasets. Large-scale experimental efforts are needed to "prime the pump" with respect to computational propagation of functional inference because the quality of sequencing data does not address the lack of appropriate experimentally derived functional annotations of genes that contribute to traits of relevance to BER missions.

Grand Challenge 2.4

Understand the links between genotype and phenotype in single but very diverse organisms and in communities of organisms that interact in terrestrial ecosystems.

Building on Genotyping and Phenotyping Advances for System Prediction

The genome encodes an organism's full potential, but the manifestation of this potential with regard to environmental interactions defines the organism's phenotype. Although the genome is largely fixed (i.e., excluding nonheritable, epigenetic effects), the phenotype can be quite variable, reflecting genotype-by-environment interactions (i.e., varying responses of different genotypes to changes in the environment). Therefore, understanding biological complexity to the point of predicting outcomes requires a complete understanding of both genotype and phenotype. High-throughput sequencing has enabled the full sequencing of various microbial, fungal, and plant genomes, although at different levels of completeness and quality. Continued improvements expected in sequencing technology will, over the next 20 years, provide at least the potential to sequence the full genomes of all living organisms. Applying these approaches to messenger RNA (mRNA) provides a platform to examine gene expression (i.e., transcriptomics), while other applications enable the identification of epigenetic marks (e.g., methylome), chromatin structure (HiC-seq), and protein-DNA interaction (e.g., ChIP-seq). Access to these technologies, coupled with the ability to extract nucleic acids directly from environmental samples, underlies the exciting area of metagenomics, where microorganisms that cannot be cultured in the laboratory can now be identified and investigated.

Although high-throughput nucleic acid approaches have developed at an exponential rate, the ability to query the corresponding phenotypes lags behind, especially when the questions to be answered require *in situ* analysis of organisms (i.e., in their natural habitat). However, exciting new technologies are being applied to these questions. For example, many universities, borrowing from technology deployed in the agricultural biotechnology industry, have installed automated plant growth chambers and greenhouses that provide relatively high throughput analysis of various phenotypic features such as growth rates, photosynthetic efficiency, and biomass. These platforms are useful for basic plant and algal discovery research. However, measurements made in these growth facilities thus far do not translate readily to experiments conducted in the field, hence, arguably contributing only marginally to the ultimate goal of enabling predictive biology in the natural environment.

The good news is that the importance of phenotyping is now readily apparent, in large part due to the stark contrast with the ability to genotype (i.e., by sequencing). Technological advancements now enable measurements of various parameters both in the laboratory and the field, which then can be statistically associated with genotype. For example, one such approach, the genome-wide association study (GWAS), makes use of natural genotypic variation in the population of a given species. Statistical association with specific alleles identified by this variation with measured phenotypic traits enables the identification of new genes controlling those traits. One advantage to this approach is its capacity for use in the field (e.g., with crop plants), providing greater potential for accurately reflecting the phenotypes expressed under natural conditions. An example is the application of GWAS to identify key alleles governing important crop biomass traits for biofuel production (Slavov et al. 2014).

Research Needs and Knowledge Gaps

Applying Multiple Functional Genomics Approaches to Single Samples and Cells. At another level, the application of various functional genomics approaches (e.g., transcriptomics, proteomics, and metabolomics) also defines phenotype when applied to bacteria, fungi, and plants. These measurements can be related to genotype, such as through the use of mutant populations, GWAS, or other approaches. However, in most cases, researchers have applied only one or a few of these experimental approaches (e.g., transcriptomics and proteomics) in a given study. One remaining challenge is to gain the ability to apply a multiomics approach to a single sample, an essential capability for accurate data comparison and integration. Another issue is that of tissue dilution. In many functional genomics approaches, the samples used are multicellular organisms or communities of organisms in which the phenotype may vary from cell to cell in response to a specific environmental change. Hence, if such cell assemblages are used, the data obtained represent an average of cellular responses across all cells, obscuring the fine details of responses at the level of an individual cell. Thus, truly understanding regulatory mechanisms and integrating dissimilar datasets are difficult, even if the measurement data are taken from the same sample. Consequently, as described in the 2010 BERAC Grand Challenges report, there remains great interest in conducting functional genomic measurements (i.e., phenotyping) on single cells, especially if they can be sampled *in situ* without extensive manipulation that could change the phenotypic response (see also Chapter 8 on Emerging Technologies, p. 95).

Developing a Hierarchal Annotation Pipeline that Integrates Experimental and Computational Approaches to Assess Functional Annotation Quality. Given the relative speed with which an organism's genotype can be defined, researchers often explore the DNA sequence to infer phenotype, mining the genome annotation (i.e., description of the genome function) for information such as specific enzymes that encode a pathway of interest. The success of this approach depends on the ability to accurately annotate the genome to provide accurate information that defines the true biochemical function of each gene, regulatory site, and epigenetic mark. The classic tenet that one gene leads to one enzyme is clearly false, with one gene potentially encoding proteins of differing functions (whether via a promiscuous active site, gene splicing, or multifunctional proteins and complexes), adding further complexity to the annotation process. There also is a growing appreciation for the impact of post-translational modifications on protein function or compartmentalization and the myriad forms these impacts can take (i.e., beyond the more-studied role of protein phosphorylation), a discovery which adds even more complexity to gaining a full understanding of the cellular function that a single protein may play. When these issues are considered in the context of the exponential growth in genome sequences deposited in databases, accurate genome annotation, while essential, also clearly becomes an extremely challenging and perhaps impossible task. Currently, the volume of accumulating genomic information mostly enables only *in silico*, automated genome annotation. This work usually is based on comparisons to previously defined annotations that may be incorrect. Hence, a real and growing risk is that the ever-increasing knowledge of genomes is being continually corrupted by incorrect and misleading annotation. At the same time, the means to broadly engineer genomes is advancing more rapidly than a basic understanding of potential engineering targets. The advent of technologies such as the CRISPR-Cas9 system underscores the urgent need for breakthroughs in addressing the gene function knowledge gap and advancing predictive understanding.

How should critical issues related to accurate genome annotation be addressed or at least curtailed? Clearly, experimental biochemistry approaches are the only accurate means to properly define, for example, the enzymatic function of a specific protein. Equally clear, however, is that such approaches, even if conducted in a high-throughput manner, are unlikely to catch up with genome sequencing advances. Experimental biochemistry or physiological approaches will remain largely in the realm of specific studies that require this level of annotation confirmation, but platforms designed to identify genes of importance to achieving BER missions provide a potential approach to prioritizing efforts. In other cases, especially as a prelude to such detailed studies, other available approaches need concentrated development and application.

Computational approaches are the only means currently addressing the full scope of genome annotation. One challenging area that may hold great promise is the use of methods that do not rely on pre-existing knowledge (e.g., Bayesian methods). Such methods are being used to predict protein structure solely from the amino acid sequence (Li et al. 2014). Nevertheless, protein structures predicted by comparison to known, experimentally derived structures are far superior. Hence, Bayesian approaches currently do not provide the level of confidence that comes, for example, from comparison to current databases.

"Perfect" (i.e., experimentally validated) annotation is the ultimate goal, yet perfect should not become the enemy of the good. Therefore, BERAC's recommendation is to develop a rational, hierarchal approach to genome annotation that can provide a continuous gradient of useful information to constrain a given gene's annotation. This approach will provide a starting point for deeper investigation by specialists when needed. For example, transcriptomic approaches can define where, when, and under what conditions a gene is expressed, while imaging approaches can define where in the cell the encoded protein is expressed. Computational and experimental approaches can determine in a high-throughput fashion the proteins with which a specific target protein may interact. High-throughput genetic, biochemical, and structural experimental approaches focused on processes of central importance to BER missions can provide underlying functional inferences to catalyze function prediction and provide a level of confidence to computational predictions. Graphic methods using, for example, mRNA or protein expression data can be used to define co-expression networks, providing a nearest-neighbor approach to further confine the annotation. None of these approaches ultimately leads to a perfect annotation. Yet, concerted BER-led efforts would develop a hierarchal annotation pipeline that fosters breakthroughs in combining experimental and computational approaches to address functional annotation veracity. Such work would greatly increase the success of sequencing-based techniques in understanding BER-relevant processes. The establishment of a database [e.g., DOE's Systems Biology Knowledgebase (KBase)] for storing and efficiently querying this information would greatly support BER research, while also likely contributing a platform for recognizing currently incorporated genome annotation data.

Tackling Genomic "Dark Space" by Characterizing Proteins of Unknown or Uncertain Function. The ability to sequence genomes and quantify a microbe's or plant's transcriptome, proteome, and metabolome during any set of prescribed conditions or stresses is far outpacing current understanding of the transcriptome-encoded proteins (i.e., components of the proteome) and determinants of the metabolome. Available omics data provide invaluable functional insight, but more than half of genes in relevant bioenergy crop genomes encode proteins whose functional annotation is either uncertain or lacking altogether. This critical knowledge gap—the genomic dark space—undermines the ability of BER systems scientists to leverage genomic information from photosynthetic and microbial organisms to meet national energy needs. As stated in the 2010 Grand Challenges document, the uncertainty of functional annotations constitutes a barrier to enabling predictive biology, and this barrier still exists today. Indeed, the means to broadly engineer plant and microbial genomes is advancing more rapidly than protein functional discovery. Most notably, the advent of technologies such as the CRISPR-Cas9 system underscores the urgent need to address this knowledge gap and advance predictive understanding.

Grand Challenge 2.5

Effectively exploit new and emerging technologies in systems biology and physical measurements (e.g., miniaturization) to accelerate biological discoveries.

A number of opportunities exist to accelerate systems biology research using emerging technologies in (1) nanotechnology and microfluidics (see also Chapter 8 on Emerging Technologies, p. 95), (2) imaging, and (3) capabilities that capitalize on foundational knowledge at DOE user facilities (see also Chapter 7 on User Facilities and Research Infrastructure, p. 85).

Research Needs and Knowledge Gaps

Using and Coupling Nanotechnology and Microfluidics. Advances in nanotechnology offer great promise for systems biology research. For instance, these methods can enable single-cell resolution measurements and perturbations and precise recording of interactions between microbes within a community. In addition, devices can be designed to conduct *in situ* experiments, allowing a coupling of the context in which an organism is isolated and the experimental test conditions.

Microfluidic devices also offer great potential because these devices can be used to culture biological samples with precision control over dynamic, time-varying environments. Microfluidic chips often are coupled with imaging to visualize the outcome of experimental perturbations, often at the single-cell level. Their small size affords reduced sample size requirements, smaller volumes of reagents, and the ability to measure many different conditions in parallel or in a repeatable, programmable fashion.

Coupling the microfluidic approaches with nanotechnology could accelerate and automate synthetic biology. For instance, these devices can achieve automated assembly of DNA fragments on a single chip (Shih et al. 2015). When coupled with next-generation cloning methods such as Gibson assembly or Golden Gate (Engler et al. 2008, 2009; Gibson et al. 2009), this technology can be used to rapidly construct prototype assemblies for testing and analysis. Another significant need in this area is the development of additional methods to optically report cellular concentrations and activities [e.g., fluorescent reporters for adenosine triphosphate (ATP) synthesis or glycolysis].

Some potential outcomes of advances in nanotechnology and microfluidics include rapid construction and screening of gene constructs for biofuel production. In addition, researchers can use these methods to assess microbial communities. They can construct synthetic communities using pairs or groups of microbial species and then subject these communities to time-varying environments that allow for slow, moderate, or rapid changes of environmental conditions to assess their impact on community structure. They also can assess spatial organization at this level, especially when coupling spatial organization with imaging methods.

Applying Imaging Technologies. For many years cryo-electron microscopy (cryoEM) was useful for identifying large protein complexes to provide an overall shape and relative orientation with respect to membranes or other subcellular structures. Researchers routinely used single-particle cryoEM analysis in conjunction with gold-labeled antibody binding to map the location of individual proteins within large protein complexes. Recent advances in direct electron detectors, along with the development of new algorithms for data analysis, have revolutionized cryoEM applications so that obtaining near atomic-scale (<3 Ångström) resolution is now possible for large proteins and protein complexes. This technique complements crystallography in probing three-dimensional (3D) structures of macromolecules at the near-atomic level. Additional emerging electron microscopic techniques address sample scales above and below those accessible to single-particle cryoEM. For example, cryo-electron tomography (cryoET) of cells or cell sections <500 nanometers (nm) thick (Villa et al. 2014) produced by focused ion beam (FIB) milling yields 3D images in the 5- to 10-nm resolution range (Diebolder et al. 2012) and can be combined with single-molecule techniques in correlative studies (Hampton et al. 2017). Protein complexes within tomograms can be viewed by single-particle subtomogram averaging, and micro-electron diffraction (microED) from radiation-sensitive micro- and nanocrystals yields high-resolution structures that have proven to be intractable by other means (Shi et al. 2016). This array of electron-based imaging techniques, in addition to a host of emerging X-ray imaging techniques now possible with the coherent beams produced by new-generation synchrotron lattices, will enable new and more detailed visualizations of biological and environmental samples important to BER.

3 Grand Challenges in Earth and Environmental Systems Sciences

Overarching Grand Challenge and 20-Year Vision

Earth system evolution is governed by complex processes and their interactions spanning a significant range of temporal and spatial scales. Credible simulations and predictions require synergistic advances in observation, modeling, data-model integration, and computing, as well as novel experimentation for mechanistic understanding of integrated system behavior and uncertainty.

Improving Predictive Understanding of Earth System Variability and Change

One of the greatest challenges for Earth and environmental systems sciences in the coming decades lies in addressing the need for better information that enables communities to assess, quantify, and plan for climate risks. These risks could affect energy and other engineering infrastructures, human and ecosystem health, food and water supply, conflict, and migration, but the current generation of Earth system models (ESMs) and the measurements that underpin them still are not well suited to inform decisions that need to be made in the context of these changing risks. Needed capabilities include (1) the ability to explore and quantify model aspects that are uncertain, inadequate, or lacking key components; (2) expanded capabilities for observing and measuring not only atmospheric processes, but also terrestrial, coastal, and cryosphere systems; and (3) coordinated research and modeling of ecosystems and processes representing key uncertainties in the coupled human-Earth system.

The U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) has made great strides in addressing some of these challenges through strategic investment in capabilities and facilities that advance understanding and modeling. Building on previous successes and anticipating new opportunities, the Biological and Environmental Research Advisory Committee (BERAC) has identified eight new grand challenges in this report for Earth and environmental systems research (see Earth and Environmental Systems Sciences Summary of Grand Challenges and Action Items, p. 22). Successfully addressing these challenges will transform predictive understanding of Earth system variability and change in the coming decades.

Earth System Models

Current ESMs have a number of limitations including (1) each model is a central estimate, (2) models are not independent, (3) ensemble size is too small to fully sample the changing likelihood of extreme events, and (4) uncertain parameters are undersampled. A significant factor limiting ESM fidelity is model resolution, which is restricted by computing resources; consequently, key Earth system processes cannot be resolved. As exascale computing is realized in the coming decade, researchers have a significant opportunity to improve ESM

Earth and Environmental Systems Sciences Summary of Grand Challenges and Action Items

The overarching goal of BER Earth and environmental systems sciences is to *improve predictive understanding of Earth system variability and change*. BERAC has identified eight grand challenges to advance modeling and observation of the Earth system and its key processes including aerosols and clouds, terrestrial and ocean biogeochemistry, the water cycle, and cryosphere. Addressing these challenges synergistically will improve predictions of water availability, sea level rise, and extreme events, from weather to climate time scales, and will provide an understanding of Earth system stability and low-probability, high-risk outcomes. These grand challenges are listed here and then described in more detail in the Grand Challenge Research Recommendations section, p. 24.

3.1. Advance Earth system modeling using a hierarchy of models, from process-resolving coupled models to reduced-order models, to transform understanding of the coupled Earth system and to produce useful and credible simulations and predictions of Earth system behavior at multiple time scales.

3.2. Establish new observational technologies and use them to understand human and Earth system processes, such as land-atmosphere interactions, biogeochemical cycles, and subsurface soils, to estimate critical process parameters using novel analysis methods, such as machine learning and data science, and to quantify model errors.

3.3. Advance basic knowledge and scale-aware simulation capability for Earth system feedbacks associated with aerosols and moist processes to better quantify aerosol forcing, precipitation changes, and extreme events with consequences for energy and water cycles, global distribution of nutrients, and human health.

3.4. Advance modeling and understanding of important ecological, biological, and carbon cycle interactions and feedbacks in the climate system to identify potential tipping points and possible energy strategies.

3.5. Characterize, understand, and model the complex, multiscale water cycle processes in the Earth system including the subsurface to understand and predict water availability and human system response to extremes.

3.6. Understand the time-dependent processes and mechanisms associated with melting glaciers, ice caps, and ice sheets and their contributions to regional sea level rise.

3.7. Quantify the interplay between internally generated climate variability and externally forced response involving anthropogenic and natural factors and their relative roles in the time evolution of regional variability to understand predictability of the Earth system.

3.8. Understand the long-term Earth system stability in response to possible future Earth system outcomes and address the level of confidence and identify emergent constraints for the range of model projections.

Associated with the grand challenges are six action items recommended for initiating progress toward the challenges:

- Develop fine-scale process models and data assimilation capabilities for data-model integration to advance model development and prediction of Earth system processes.
- Advance high-resolution modeling in different simulation and prediction modes supported by exascale computing to improve understanding and prediction of extreme or high-impact events.
- Develop a computational user facility for rapid design, generation, evaluation, and diagnosis of Earth system model simulations, as well as analysis of ensemble predictions and data-model synthesis.
- Develop and maintain a hierarchy of models for hypothesis testing, model development, and uncertainty analysis of the complex human-Earth system.
- Develop and integrate new sensing technologies and optimize field deployments in Atmospheric Radiation Measurement, AmeriFlux Network, Next-Generation Ecosystem Experiments, and other BER programs to explore interactions across different scales of biological organization and biosphere-atmosphere feedbacks.
- Create new integrated field laboratories that target biogeochemical, energy, and water flows between urban areas and surrounding ecosystems.

spatial resolution for cloud-resolving atmosphere modeling, eddy-resolving ocean modeling coupled with detailed sea ice and ice sheet dynamics, and fine-scale terrestrial modeling of ecosystems and surface and subsurface processes of hydrology and biogeochemistry. DOE is in a leadership position to advance exascale computing and Earth system modeling in tandem.

Supported by computational user facilities, short-term climate predictions can be used as a testbed for accelerating ESM improvement and evaluation and for providing information on which impacts will be felt from the interplay between internal variability and externally forced response. At the same time, a hierarchy of models with different complexities is needed for large-ensemble modeling, hypothesis testing, process understanding, and uncertainty analysis.

Measurements and Observations

To improve understanding and modeling, observational capabilities need to be expanded. The concept and capabilities of BER's Atmospheric Radiation Measurement (ARM) Research Facility are unique and significant, but the ARM Facility focuses on atmospheric measurements, whereas progress in modeling will require measurements for the coupled Earth system. Measurements of terrestrial, coastal, and cryosphere systems have significant gaps. The data challenge for ESMs is to (1) fill these measurement gaps; (2) identify niche areas to integrate models and data; (3) develop data assimilation capabilities for model-data integration; and (4) leverage data analytics to develop new understanding and address key uncertainties in aerosol and cloud processes, terrestrial and ocean biogeochemistry, and water cycle and cryosphere processes.

Coordinated Research and Modeling of Key Uncertainties in the Coupled Human and Earth Systems

Using new variable-resolution capabilities, DOE is uniquely suited to address the significant challenge of regional sea level rise and coastal impacts related to ice sheet dynamics, ocean-ice interactions, ocean thermal expansion, coastal inundation, and other impacts. High-latitude changes and their global influence will continue to be an important target over the next 20 years. DOE is well positioned to address this challenge with coordinated research activities in Earth system modeling, analysis of cryosphere processes, and measurements from BER's Next-Generation Ecosystem Experiments (NGEE)–Arctic project and the ARM supersite on Alaska's North Slope. Lastly, through the development of human–Earth system models, DOE can further develop and use the capability to explore Earth system stability and the potential for low-probability, high-risk outcomes.

Building on Progress Related to the 2010 Grand Challenges

The 2010 BERAC report, *Grand Challenges for Biological and Environmental Research: A Long-Term Vision* (BERAC 2010), posed six grand challenges in climate research. Important progress and some remaining challenges are briefly summarized:

Developed Higher-Resolution Modeling. DOE has contributed to the development of high-resolution versions of the Community Earth System Model (CESM), global variable-resolution models, and super- and ultraparameterizations. Researchers have used high-resolution models to understand the water cycle and extreme events. Efforts are well under way to develop DOE's Energy Exascale Earth System Model (E3SM) as a very high resolution model, but more relevant processes must be integrated into ESMs.

Improved Parameterizations of Aerosols that Affect Clouds. Significant progress has been made through research supported by multiple DOE programs to develop parameterizations of aerosols and aerosol-cloud interactions and perform modeling experiments to understand and quantify aerosol direct and indirect effects. However, indirect aerosol forcing and some basic knowledge of aerosols remain uncertain.

Developed Ecosystem-Observing Systems to Monitor Biogeochemical Cycles. DOE has established ecosystem-observing systems in a few regions, but they are not positioned to monitor biogeochemical cycles on global scales. Ecosystem models have been tested against some observations, but efforts have not yet narrowed enough to estimate critical process parameters.

Advanced Understanding of Important Biological Interactions and Feedbacks. DOE has advanced understanding of potential rates of permafrost carbon losses and the influence of climate and land use change on wildfires. However, understanding regarding the long-term stability of other carbon stores and biological systems is still lacking.

Improved Integration of Anthropogenic Climate Forcings into ESMs. DOE has made important progress in developing an integrated Earth system model (iESM) that couples integrated assessment models (IAMs) with ESMs to represent human-Earth interactions. DOE is well positioned to further enhance integration of anthropogenic forcings into ESMs for improving climate projections and mitigation, as well as impacts, adaptation, and vulnerability research.

Established New Observational Technologies to Quantify Model Errors. Some notable new observational technologies include the ARM supersites, which represent significant progress in providing data critical for improving cloud and aerosol modeling, and DOE's AmeriFlux Network for providing carbon and energy fluxes in different terrestrial ecosystems. Important advances have been made in model-experiment integration as exemplified by (1) the NGEE–Arctic and NGEE–Tropics projects, (2) the development of frameworks for diagnostics and metrics such as the Program for Climate Model Diagnosis and Intercomparison (PCMDI), and (3) the International Land Model Benchmarking (ILAMB) project to comprehensively evaluate and diagnose model errors.

Grand Challenge Research Recommendations

Grand Challenge 3.1

Advance Earth system modeling using a hierarchy of models, from process-resolving coupled models to reduced-order models, to transform understanding of the coupled Earth system and to produce useful and credible simulations and predictions of Earth system behavior at multiple time scales.

Increasing the Predictive Power of Earth System Models

Comprehensive ESMs are the primary tools used in predicting Earth system response to external perturbations. However, large uncertainty and lack of spatial detail have limited the usefulness of ESMs for addressing DOE's energy mission. ESMs could be advanced to produce useful and credible simulations and predictions to support planning of energy infrastructure and energy production and use through (1) the use of a modeling hierarchy, (2) systematic efforts in developing data assimilation systems that advance measurement-to-modeling integration, (3) user facilities that enable rapid development and evaluation of simulations and forecasts, and (4) collaborations among model developers and climate and data analytics scientists.

ESMs include a wide range of physical processes, from large-scale circulations of the atmosphere and oceans to atmospheric convection and mesoscale ocean eddies, to cloud microphysics and radiative transfer, to the biophysics of photosynthesis and respiration on the land surface and in the ocean. ESMs are used to simulate and predict phenomena on time scales ranging from days to centuries, but the results also depend on small-scale processes that have time scales of minutes or less. For these reasons, investigators have always run ESMs on the most powerful computing systems available.

ESMs are able to simulate many important phenomena of the real Earth system, including important modes of internal variability such as El Niño and the system's response to external forcings. However, ESM-based simulations of future anthropogenic climate change are subject to uncertainties caused by insufficient resolution, incomplete understanding of physical processes, and large internal variability that obscures the signals. Models cannot directly simulate processes on scales smaller than their grid spacing, so parameterizations are used to represent subgrid-scale processes. A major challenge in Earth system modeling is reducing uncertainty, an aim that may be achieved by making the grid finer but requires more computing power. Increased spatial resolution has several immediate benefits, including simulation of a wider range of fluid dynamical scales, leading to more realistic representations of a host of processes. Another way to reduce uncertainty is to improve the realism of the parameterizations. Computing power helps with that too, because, as spatial resolution increases, the nature of the unresolved parameterized processes becomes simpler, although the models also can become less constrained as the number of degrees of freedom increases.

Research Needs and Knowledge Gaps

Leverage New Computing Architectures for ESM Evolution. The operational numerical weather prediction (NWP) community already has achieved atmospheric grid spacing decreases from around 100 kilometers (km) to around 10 km, probably achieving within the next few years a convection-resolving grid spacing of 4 km or finer. Decreases in grid spacing, while short of being convection resolving, always provide improvements in short-range forecasts but hardly decreases the errors of longer-range forecasts beyond a few days (Buizza 2010). An interpretation is that the errors in long-range forecasts are due mainly to deficiencies in the physical parameterizations, suggesting that ESMs will benefit less from increased resolution, relative to NWP models. Enabled by computers of the coming decade, the eventual transition to convection-resolving resolution may allow ESMs to benefit from increased resolution. Similar arguments also can be made for eddy-resolving ocean models and process-resolving land surface and sea and land ice models that play important roles in Earth system evolution. A strong collaboration between the Earth system modeling and high-performance computing communities is needed to facilitate a smooth transition in which ESMs evolve to take advantage of new computer architectures (see also Chapter 6 on Data Analytics and Computing, p. 71).

Improve Process Parametrizations. Increased computing power also can improve physical ESM parameterizations. Superparameterization (Randall et al. 2016) and ultraparameterization (Parishani et al. 2017) improve physical representations by embedding a simplified cloud-resolving model and large eddy simulation (LES) model, respectively, within each atmospheric grid column of an ESM. As resolution increases, the model parameterizations become simpler in the sense that fewer processes have to be parameterized. However, even if extreme increases in computing power eventually permit 100-meter grid spacing over the entire globe, ESMs will still need parameterizations of turbulence, radiation, and cloud microphysics including the effects of aerosols. Of these, cloud microphysics and aerosols are likely to be the most problematic (see Grand Challenge 3.3, related to aerosols and moist processes, p. 29).

Develop Fine-Scale, Process-Resolving Models with Data Assimilation Capabilities. ESM parameterizations inevitably involve numerical parameters whose values must be specified based on observations. One approach is to adjust the parameter settings iteratively to reduce the errors, but this approach does not improve understanding of the errors and may reduce errors for the wrong reason. With finer grids, the parameterized grid-scale processes are more amenable to direct comparison with field data, making model evaluation more straightforward. As more processes are explicitly resolved, enabling parameterizations to be increasingly physically based, observations can be used more effectively to estimate the numerical values of model parameters. This applies to models of all Earth system components and is another motivation for increasing model resolution. ARM data are ideally suited for assigning values to parameters associated with atmospheric physical processes. The ARM Research Facility has been exploring use of LES for the weather observed at its Southern Great Plains site. LES data can fill in gaps associated with processes that are not observed or not sufficiently well sampled by the observing system. Assimilating field measurements into LES models can make this strategy more effective. Similarly, fine-scale plant and subsurface hydrology and biogeochemistry models are important tools for linking measurements to models to improve terrestrial system models. A systematic effort to develop fine-scale models with data assimilation capability may support these objectives across multiple DOE programs.

Develop Reduced-Order Models to Understand Earth System Interactions and Responses to Perturbations. As the resolution and complexity of ESMs increase, understanding model behaviors and quantifying the uncertainty in their predictions become more challenging. ESMs of intermediate complexity and reduced-order models are important tools on the other end of the modeling hierarchy for understanding interactions among Earth system components and for unraveling the system response to perturbations and quantifying its uncertainty. Developing and maintaining a hierarchy of models and simulation capabilities should be an important part of a strategy to improve ESMs and advance their use in understanding and predicting Earth system variability and change across all time scales relevant to stakeholder decision making.

Collaborate with Computational User Facilities for Model Testing and Improvement. To further ESM development and utility, computational user facilities need to be expanded for on-demand, rapid generation, evaluation, and diagnosis of ensemble simulations and experimental ensemble predictions. One example is to test ESMs through extensive forecasting on time scales ranging from days (e.g., Palmer et al. 2008) to those that are subseasonal to seasonal, for which observations for model evaluation can be provided in great detail. DOE has taken some steps in this direction (e.g., Phillips et al. 2004), but a more comprehensive and systematic effort would be useful. In addition, the ability to run decade-to-century simulations is important for understanding both fast and slow responses (see Grand Challenge 3.7, related to Earth system predictability, p. 39). The user facilities should engage model developers and computational scientists and also collaborate with scientists in climate research and machine learning, an essential area for turning large-scale data (i.e., Big Data) into knowledge useful for improving understanding and predicting the Earth system. Additionally, increased in-house expertise is needed in physical oceanography to complement DOE's existing in-house expertise in ocean modeling and to improve understanding of the coupled Earth system in which the ocean plays a key role.

Grand Challenge 3.2

Establish new observational technologies and use them to understand human and Earth system processes, such as land-atmosphere interactions, biogeochemical cycles, and subsurface soils, to estimate critical process parameters using novel analysis methods, such as machine learning and data science, and to quantify model errors.

Building on BER Success in Earth System Measurements and Observations

Observations provide the foundational knowledge upon which hypotheses about the function of the complex Earth system and its response to perturbations are developed and tested in combination with modeling. BER's leadership in the development of observing systems and experiments has created a vibrant community of scientists who have provided answers to some of the most urgent questions challenging the ability to develop a predictive understanding of the Earth system. This research has a rich tradition of transforming Earth system science by developing new measurement technologies and deploying them at remote field sites throughout the world. Important examples of these flagship systems include the ARM fixed sites and mobile facilities, AmeriFlux terrestrial ecosystem flux network, Free-Air Carbon Dioxide (CO_2) Enrichment

(FACE) experiments, and NGEE projects (see also Chapter 7 on User Facilities and Research Infrastructure, p. 85). Each of these observing systems operates at a scale that is not possible to sustain with university or private-sector investments. Practitioners of this emerging new field of Earth system data science are harnessing machine learning and other data science approaches to search for patterns and create new process-level understanding of important biological and physical phenomena.

The AmeriFlux Network provides high-quality information about carbon and energy fluxes in different terrestrial ecosystems. Thousands of scientists download AmeriFlux and Fluxnet datasets annually, and publications from this data use have transformed current understanding of the processes regulating photosynthesis, respiration, and environmental factors that influence net ecosystem carbon balance. With 110 registered active sites, the network is arguably the largest and most successful federally funded ecosystem-observing network in the United States. The ARM Research Facility provides measurements of the physical and chemical properties of atmospheric aerosols, clouds and their microphysical compositions, and radiative fluxes, as well as a suite of other meteorological measurements at four fixed sites and two mobile facilities. These measurements represent the most comprehensive observations of aerosols and clouds available in the world. ARM data have been used by thousands of scientists worldwide to understand aerosol and cloud processes, as well as aerosol-cloud-precipitation interactions, and to improve models. The ARM measurement strategies are being adopted by other countries. NGEE–Arctic and NGEE–Tropics have created important new science infrastructures for investigating ecosystem responses to global environmental change in remote and logistically challenging environments. NGEE–Arctic has led to recent breakthroughs in understanding how permafrost influences hydrologic flows and biogeochemical processes in high-latitude ecosystems. NGEE–Tropics, in turn, has provided insight about the sensitivity of tropical forests to drought and other environmental changes associated with El Niño. More recently, BER has developed the concept of integrated field laboratories (IFLs) with the aim of studying interactions along important lateral and vertical gradients that link human population centers with surrounding natural environmental systems. The IFL conceptual design included themes associated with coastal regions, arid lands, mountain watersheds, agriculture, and urban ecosystems.

Building on these BER successes, new observational technologies would further advance the understanding of environments in rapidly changing regions both within and outside the United States. Such understanding is strategically important for energy security, as environmental and societal changes are connected through myriad energy, water, biogeochemistry, and socioeconomic pathways at a broad range of temporal and spatial scales.

Research Needs and Knowledge Gaps

Outfit AmeriFlux with New Technologies for *In Situ* Sequencing and Expanded Measurements. New investment in the AmeriFlux Network has the potential to transform ecosystem science. One challenge is to link variation in ecosystem fluxes with changes in plant gene expression and soil microbial community dynamics to help researchers probe the fundamental mechanisms regulating seasonal and interannual changes in the carbon balance of terrestrial ecosystems. This challenge could be achieved by developing and deploying *in situ* sequencing technologies at several core AmeriFlux sites. New information about the changing composition of fungal and bacterial communities could enable significant advances in modeling ecosystem respiration as well as methane (CH_4) and nitrous oxide (N_2O) trace gas production. Concurrent information on gene expression in plants could revolutionize understanding of cold hardening (and recovery) processes in temperate and high-latitude ecosystems, time-evolving limits to photosynthesis, and production of different volatile organic compounds that create clouds and feedbacks via atmospheric chemistry. Another critical AmeriFlux frontier involves more quantitatively linking carbon and water fluxes to ecosystem demographics, landscape dynamics, land use, and land management. First, new investment in tree and shrub demography at core sites would considerably strengthen the use of AmeriFlux to study climate change impacts on vegetation dynamics and to test next-generation demography models such as the Functionally Assembled Terrestrial Ecosystem Simulator (FATES) model, which is being integrated into the E3SM. Second, systematic deployment of ecosystem-level lidar and remote-sensing measurements have considerable potential for mapping three-dimensional (3D) changes in forest composition before and after climate extremes and other processes that influence ecosystem structure and flux dynamics. Third, new tower clusters along land use and management gradients are essential in developing a quantitative understanding of how humans are modifying ecosystem fluxes at regional and continental scales.

Add ARM Capabilities for Understanding Aerosol-Cloud Interactions, Cloud Radiative Feedbacks, and Extreme Precipitation. New investment in the ARM Facility has potential for characterizing the coupling of aerosols and clouds with atmospheric turbulences and convective systems, which is currently lacking. Such information is essential for understanding aerosol-cloud interactions, cloud radiative feedback, and extreme precipitation. Needed to constrain models are measurements of 3D atmospheric motion fields from remote-sensing instruments and drones across a range of scales from a few hundred kilometers to a few meters, along with concurrent measurements of aerosols and clouds.

Develop New Remote-Sensing Technologies for Atmospheric, Land, and Urban Systems. Another challenge for BER is the design and deployment of new systems to remotely monitor ecosystem properties and environmental variables by exploiting emerging technologies. Over the past decade, several remote-sensing techniques have transformed understanding of Earth system science. Paradigm-changing observing systems include the Argo float network in the oceans, AmeriFlux Network on land, ARM scanning cloud radars, and the National Aeronautics and Space Administration's (NASA) constellation of Earth-observing satellites. The pace of development in the environmental-sensing field continues to accelerate and includes advances in drone, lidar, and subsurface imaging technologies as well as improvements in networking systems. An important science challenge for BER is to develop new systems that provide new environmental information across a range of temporal and spatial scales that are currently inaccessible by NASA or other federal agencies. Development of a strategic plan and the specific targeting of several achievable science objectives in the atmosphere, land, and urban domains are important near-term requirements for BER to maintain its commitment to excellence within this field. New technologies that enable mapping the 3D structure of trace gas, heat, and wind and water flows within cities, for example, may create new opportunities for the design of sustainable energy solutions.

Deploy IFLs, Conduct Regional-Scale Experiments Examining Biosphere–Climate System Feedbacks, and Design Expanded FACE Studies. Three major research needs are identified for future BER field experiments. The first is to develop IFLs that target biogeochemical, energy, and water flows between urban areas and surrounding ecosystems. Important gradients include connections between cities and nearby coastal ecosystems, mountain watersheds, arid lands, and croplands. IFLs should be located in rapidly changing regions or areas that are strategically important for energy security. They also need to exploit existing BER science capabilities across biological and Earth science disciplines (e.g., ARM, AmeriFlux, and DOE's Joint Genome Institute and Environmental Molecular Sciences Laboratory), while simultaneously fostering new ones. The second research need is to create new experiments that capture feedbacks between the biosphere and physical climate system at regional scales. An important example is the analysis of how different land uses modify surface fluxes and atmospheric composition that, in turn, modify regional climate and, ultimately, the sustainability of human enterprise within a region. Existing NGEEs and ARM facilities have not yet embraced these coupling and feedback challenges. Finally, BER's past and current investments in FACE have yielded datasets that have proven essential for developing predictive models of the global carbon cycle. Yet escalating costs have limited the expansion of these experiments into new regions where uncertainties remain extremely high. Considering new breakthroughs in simulating air flows through forest canopies, control systems, and isotopic labeling approaches, BER should revisit whether new FACE designs could be developed that would enable a wider suite of deployments. The direct response of ecosystems to rising atmospheric CO_2 remains one of the most important sources of uncertainty in ESMs, on par with uncertainties associated with aerosol direct and indirect effects and exceeding those associated with climate-carbon feedbacks.

Tightly Couple Models and Experiments to Test Predictions and Hypotheses and Inform Observational Strategies. Seamless model-experiment integration is critical for measurements to effectively inform model development and hypothesis testing. Models are increasingly capable of resolving the complexity of processes exhibited within observations and interpreting causal pathways. Combining inverse modeling, data assimilation strategies, and novel metrics can help design better observational systems to test relationships important to predictions. The past decade has seen tremendous growth in new capabilities such as machine learning and other data science approaches. Consequently, the time is right to invest in capabilities synthesizing different fields of expertise to create new process-level understanding of many important biological and physical phenomena (see also Chapter 6 on Data Analytics and Computing, p. 71).

Grand Challenge 3.3

Advance basic knowledge and scale-aware simulation capability for Earth system feedbacks associated with aerosols and moist processes to better quantify aerosol forcing, precipitation changes, and extreme events with consequences for energy and water cycles, global distribution of nutrients, and human health.

Resolving Critical Uncertainties for Aerosols, Cloud-Climate Feedbacks, and Precipitation

Aerosols originate from both natural and anthropogenic sources. Atmospheric aerosols directly affect Earth's energy balance and air quality. Depending on the physical and chemical properties of aerosols, direct radiative forcing on the Earth can be negative or positive due to scattering of solar radiation to space and aerosol absorption of solar and infrared radiation. Aerosols can modify the number density concentration of cloud particles and, thus, cloud particle sizes, which impact cloud optical properties. Cloud particle sizes, in turn, affect the precipitation efficiency and lifetime of clouds. These indirect effects of aerosols on Earth's energy budget and precipitation still are not well understood, and their simulations are poorly constrained by observations, so aerosols represent one of the largest uncertainties in current ESMs (Boucher et al. 2013; Nazarenko et al. 2017).

Clouds strongly affect Earth's energy budget by reflecting solar radiation and trapping infrared radiation. Their net radiative impact depends on cloud temperature, hydrometeor particle sizes, and mass of liquid and ice in the clouds. Clouds form as a result of supersaturation by the lifting of moist air in the presence of aerosols and atmospheric flows across a broad spectrum of temporal and spatial scales. How clouds respond to climate change determines whether they positively or negatively feed back to a climate forcing such as atmospheric greenhouse gases (GHGs; Geoffroy et al. 2017).

Current ESMs require parameterizations of many atmospheric processes important to aerosols and clouds. The large uncertainties of current ESMs in simulating the indirect effects of aerosols on Earth's energy budget, cloud-climate feedback processes, and intensity-frequency distribution of precipitation seriously limit the utility of these models in simulating energy and water cycles and projecting their future changes.

Resolving these areas of uncertainty will advance efforts to predictively understand the Earth system and its variability on a number of fronts. For example, deeper insights into aerosols and their indirect effects will improve simulations of aerosol forcing on the Earth system from natural and anthropogenic sources. Meanwhile, increased knowledge of clouds and their interactions with turbulence and the Earth surface would enable researchers to better quantify cloud-climate feedbacks and reduce uncertainties of future climate change projections. Better understanding and an improved ability to model mesoscale convective systems (MCSs) will lead to better simulations of the global water cycle and prediction of extreme weather events that affect the production, transport, and use of human energy systems. Finally, robust, scale-aware physical parameterizations will enable practical applications of ESMs to predict water, energy, and carbon cycles as well as their impacts on human energy systems.

Research Needs and Knowledge Gaps

Improve the Understanding of Biogenic Aerosol Emissions, New Particle Formation, and Aerosol Evolution in the Atmosphere. Knowledge of the chemical and physical properties of aerosols, including their mixing states, is needed to understand their growth and evolution. Aerosol effects on mixed-phase clouds are known to significantly affect cloud lifetime and precipitation, but this knowledge is incomplete and their simulations are highly uncertain. Also poorly understood is the nucleation of ice cloud particles by aerosols. Both laboratory experiments and field measurements are needed to test and improve the fundamental theory of new particle formation (see Fig. 3.1. New Particle Formation, p. 31), secondary organic aerosol (SOA) growth, and aerosol nucleation of ice particles. Process models should be developed based on first principles of physics and chemistry to connect measurements with models. Parameterizations should be developed on the basis of theoretical understanding and process models for use in coarse-resolution ESMs for long-term simulations.

Understand and Accurately Simulate Cloud Interactions with Atmospheric Turbulence and Earth's Surface. Clouds primarily develop in turbulent atmospheric flows, and they, in turn, affect turbulence flows through their effects on radiation and latent heating. Accurately simulating turbulence-cloud coupling is essential for simulating both the aerosol indirect effect and cloud-climate feedback. The ubiquitous low clouds over the eastern oceans in the subtropics exist due to atmospheric turbulence generated by cloud-top radiative and evaporative cooling. Surface heterogeneities, including terrain, vegetation, and soil temperature and moisture, influence atmospheric turbulence and cloud formation. Research is needed to develop a hierarchy of models to describe cloud processes—from single air parcels to single shallow cumulus clouds, to large eddies of several kilometers—to enable ESMs to explain and parameterize the coupling of atmospheric turbulence and clouds and the roles of surface heterogeneities and large-scale atmospheric conditions. New observations should be made to better quantify the interaction among turbulence, clouds, and surface fluxes and to constrain the models.

Characterize Mesoscale Convective Systems and Their Role in Extreme Weather Events. MCSs are responsible for some of the most damaging, high-impact weather events such as tornados, hail, and extreme precipitation. Recent findings suggest that the increased frequency of longer-lasting MCSs is responsible for increased extreme precipitation over the central United States in the past decades (Feng et al. 2016). MCSs self-organize to exhibit a variety of dynamical structures strongly modulated by large-scale atmospheric and surface conditions

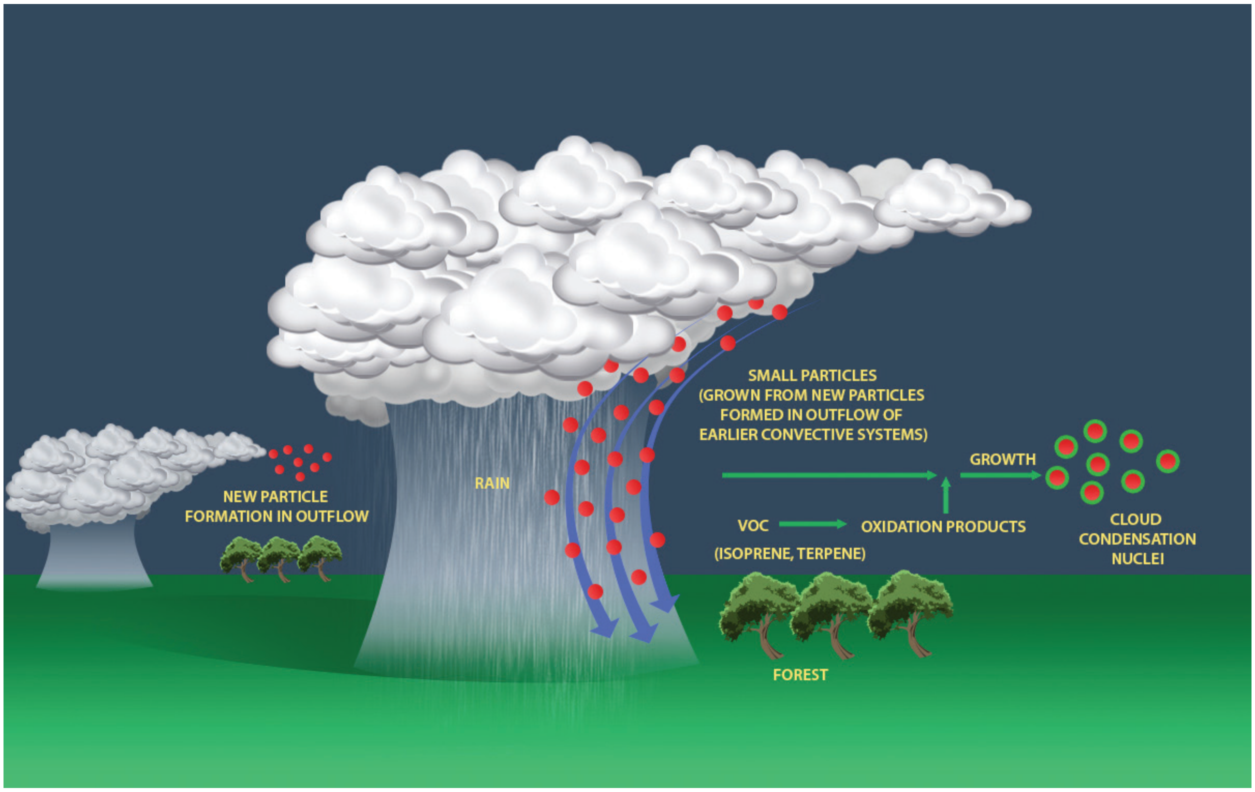


Fig. 3.1. New Particle Formation. New particles form in the outflow region of earlier convective clouds, leading to high concentrations of small particles in the free troposphere. These small particles are injected into the boundary layer by the downward motions in the convective system, where they interact with volatile organic compounds (VOCs) and other gases emitted by vegetation from the Earth's surface to grow and become cloud condensation nuclei. [Image courtesy Luiz Machado, National Institute for Space Research, and Jian Wang, Brookhaven National Laboratory]

(Houze et al. 2015). The initiation, propagation, and decay of MCSs and their structures of dynamics and thermodynamics still are not well understood and are poorly simulated in models. Research is needed to characterize MCSs with simultaneous measurements of cloud microphysics and precipitation, updrafts and downdrafts, cold pools, anvils, gravity waves, and environmental conditions of the atmosphere and surface. Research also is needed to improve theoretical understanding of the organization within MCSs, their propagation, and time evolution under different environmental conditions. A hierarchy of models can be used to describe their mechanisms, leading to their realistic simulation and prediction. Research is needed to understand the predictability limit of MCSs through data assimilation and hindcast and forecast experiments (see also Chapter 8 section, Planet Scale, p. 109).

Develop Scale-Aware Physical Parameterizations, Particularly for Deep Convection. The impact of subgrid-scale processes on resolved-scale dynamics and thermodynamics is represented by physical parameterizations in ESMs. By definition, the subgrid-scale processes change with model resolution, but physical parameterizations in current models are designed for resolutions limited by computational resources. As high-resolution and variable-resolution models become practical, the physical basis and formulation of some existing parameterizations are no longer valid. Research is needed to develop scale-aware physical

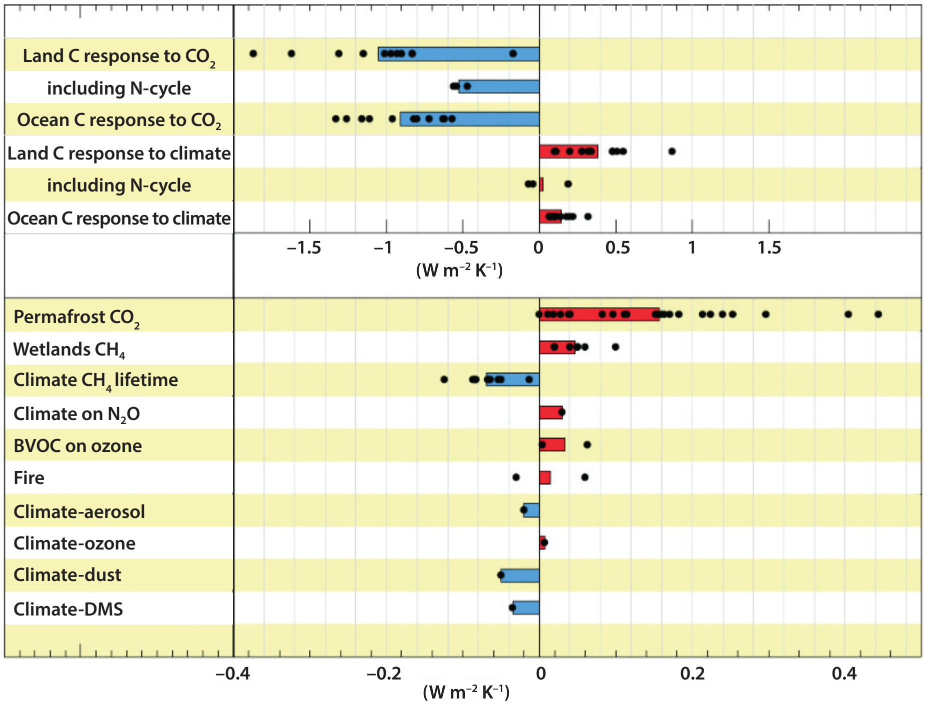


Fig. 3.2. Biogeochemical Feedbacks. A synthesis of the magnitude of biogeochemical feedbacks on climate. **Key:** BVOC, biogenic volatile organic compounds; C, carbon; CH_4, methane; CO_2, carbon dioxide; DMS, dimethyl sulfide; K, Kelvin; m, meter; N, nitrogen; N_2O, nitrous oxide; W, watt. [Image reprinted with permission from Ciais, P., et al. 2013. "Carbon and Other Biogeochemical Cycles." In: *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Eds.: Stocker, T. F., et al.), Cambridge University Press, Cambridge, UK, and New York, NY, USA.]

parameterizations for models across resolutions. A particular challenge is the treatment of deep convection that barely can be resolved by global models but may occupy a large fraction of a model grid cell. As deep convection may organize to MCSs, the inability to resolve the former also limits representation of the latter in climate models. Research on new methodologies is needed to overcome this challenge.

Grand Challenge 3.4

Advance modeling and understanding of important ecological, biological, and carbon cycle interactions and feedbacks in the climate system to identify potential tipping points and possible energy strategies.

Developing a Rigorous Understanding of Ecosystems and the Global Carbon Cycle

Biogeochemical feedbacks have the potential to considerably alter future trajectories of atmospheric composition and climate in the 21st century (see Fig. 3.2. Biogeochemical Feedbacks, this page). Multiple lines of evidence suggest that forest responses to rising atmospheric CO_2 are responsible for a large component of the contemporary terrestrial carbon sink (Pan et al. 2011; Schimel et al. 2015). However, several massive terrestrial carbon reservoirs may be vulnerable to future change. They include carbon stored in permafrost soils within boreal and tundra biomes (Schuur et al. 2015), aboveground biomass in tropical forests (Brienen et al. 2015), carbon preserved under anoxic conditions in boreal and tropical peatlands (Turetsky et al. 2015), and organic matter associated with mineral soils that may be sensitive to changes in temperature or soil moisture (Pries et al. 2017). Within oceans, the net air-sea CO_2 flux will respond to future changes in the solubility of dissolved inorganic carbon and ocean circulation. A weakening of ocean mixing and overturning could reduce the ocean's ability to soak up anthropogenic CO_2 that is emitted into the atmosphere, directly by reducing surface inorganic carbon flows into the interior (Schwinger et al. 2014) and indirectly by reducing the availability of surface nutrients that regulate the strength of the biological pump (Galbraith and Martiny 2015). Combined, the impact and uncertainty associated with these biogeochemistry feedbacks are as large as those associated with clouds and ocean dynamics (Gregory et al. 2009). Thus, an improved understanding of carbon cycle processes is essential for developing predictive models of Earth system dynamics over time scales of years to centuries.

Apart from understanding Earth system feedbacks, fundamental knowledge of ecosystems is essential for understanding their resilience to changes in human and environmental drivers. Accurate predictions of water availability for energy and agriculture, for example, depend on a mechanistic representation of vegetation controls on evapotranspiration in upstream watersheds. Similarly, analysis of the factors that influence U.S. food security can benefit from a mechanistic representation of crops in ESMs, enabling yield predictions that respond simultaneously to changes in multiple environmental factors. In oceans, future changes in net primary production and interior oxygen availability may limit the range of important commercial fisheries (Deutsch et al. 2015). Understanding the mechanisms that regulate the carbon sequestration potential of ecosystems is important to systematically explore future land use scenarios that may limit carbon accumulation in the atmosphere, thus helping to stabilize the Earth system.

Achieving a rigorous understanding of the global carbon cycle and ecosystems is necessary for understanding long-term Earth system consequences such as floods, water availability, wildfire damage, and heat waves in different energy scenarios. Such knowledge also is critical for the design of energy solutions that optimize outcomes across different economic sectors and regions. Moreover, maintaining a robust biosphere is an essential foundation for national and economic growth as well as regional and global food security.

Research Needs and Knowledge Gaps

Predict the Changing Structure and Composition of Ecosystems Across the Global Land Surface. Although considerable progress has been made during the past decade in understanding controls on tree mortality (Anderegg et al. 2015), the mechanistic representation of underlying physiological processes in ESMs remains rudimentary, with advances expected from better simulation of plant-soil hydrologic coupling and stand dynamics and from the use of experimental testbeds for model development and evaluation. Environmental controls on seed dispersal, recruitment, migration, pest outbreaks, wildfires, invasive species, and changing land use patterns represent equally important controls on vegetation dynamics at region-to-continent scales. Recognizing this broader suite of processes and interactions will be necessary for making progress on this challenge. To create a new synthetic understanding of vegetation dynamics, new investments will need to target experimental manipulation, remote ecosystem monitoring, mechanisms coupling human and natural ecosystems, and simulation capacity to rapidly and systematically explore a wide range of future scenarios.

Improve Model Representation of the Terrestrial Carbon Sink and Responses of Key Systems (e.g., Arctic, Tropics, and Peatland) to Change. As ESM simulation of the global carbon cycle is becoming more mature, critical uncertainties remain that limit the ability to predict feedbacks and future atmospheric composition with confidence. Perhaps the most significant uncertainty originates from the CO_2 fertilization effect—specifically, the degree to which the terrestrial biosphere is taking up carbon in response to increasing atmospheric CO_2. Model predictions of the magnitude of the contemporary terrestrial carbon sink vary by more than a factor of two in current ESMs. Contemporary and future responses of high-latitude and tropical ecosystems to different global change drivers remain the largest source of uncertainty in carbon cycle predictions. Thus, model-experiment integration targeting these regions is essential for making future progress. Other carbon cycle challenges will require investments in new areas. In Indonesia, decomposition and burning of peatlands during El Niño events have considerably accelerated annual carbon losses associated with land use change, while creating regional haze events that cause premature deaths (Koplitz et al. 2016). In northern tundra and boreal forest ecosystems, peatland also represents a very large and vulnerable carbon reservoir that has the potential to considerably modify future atmospheric composition (Turetsky et al. 2015). Thus, accurate peatland representation in ESMs represents an important need essential for future carbon cycle and air quality predictions.

Accurately Simulate Atmospheric CH_4 and N_2O. Fully coupled Earth system simulation of biogeochemical cycles other than CO_2 remains nascent, so another important challenge is to simulate the transient dynamics of atmospheric CH_4 and N_2O. The biogeochemical processes regulating these gases have undergone profound modification by human activities, and their future trajectories remain highly uncertain. These gases have the potential to considerably modify emissions scenarios. For CH_4, new investment is needed in land use model development, complementing the existing capability to simulate CH_4 production and oxidation in soils. A fully coupled nitrogen cycle will require similar investments in land use, along with explicit mass-conserving simulation of land-to-ocean nutrient flows and better representation of nitrogen and oxygen dynamics in the ocean. Although state-of-the-art ESMs now simulate the influence of nitrogen limitation on photosynthesis, the model's ability to simulate preindustrial atmospheric N_2O mole fractions and the transient rise of this trace gas over the past 200 years represents an important (and unrealized) mechanistic test of global nitrogen cycle knowledge.

Develop New Approaches to Couple and Simulate Land, Aquatic, and Ocean Biogeochemical Cycles. Coastal ecosystems are vulnerable to changes in human systems and environmental conditions such as extreme events. Developing realistic future scenarios of ocean ecosystem responses to global change requires new approaches for coupling land, aquatic, and ocean biogeochemical cycles, including the flow of freshwater, nutrients, organic carbon, and sediments from terrestrial ecosystems to the river and coastal ocean. Understanding and modeling plant, soil, and hydrologic processes across land, river, and coastal systems are important as these processes exert a fundamental control on biogeochemical interactions. Changes in both the frequency and magnitude of hydrologic extremes, as well as changes in land management, will have significant impacts on nutrient flows and their ecological impacts. A mechanistic understanding of this coupling also is required for predicting nutrient trapping in coastal sediments, and, ultimately, the delivery of phosphorus, nitrogen, and iron to the open ocean, as well as how extreme events such as tropical cyclones and associated storm surge and coastal inundation may influence biogeochemistry in the coupled land-river-ocean system. Anthropogenic modification of river inputs is likely to be considerable over the next several decades with expected increases in global agriculture. Sustained agricultural and urban inputs have the potential to modify the strength of the biological pump, as well as coastal and open ocean interior oxygen concentrations.

Develop a Quantitative, Multiscale Understanding of the Coupling Mechanisms Associated with Reactive Trace Gases and Aerosols. Over the past decade, recognition of reactive trace gases and aerosols as important drivers of land and ocean ecosystem dynamics has increased. Examples of coupling mechanisms include the effects of aerosols on diffuse light; the impact of ozone on photosynthesis and canopy conductance; and the redistribution of nutrients by dust storms, wildfires, and spores. A key challenge within this emerging field is the need to develop a quantitative understanding of the impact of the different coupling mechanisms across a range of spatial and temporal scales. These mechanisms often are obscured in the observational record as a consequence of extreme events modifying multiple aspects of atmospheric composition. For example, large volcanic eruptions can profoundly affect diffuse light, yet these events also influence tropospheric hydroxyl concentrations and iron deposition in remote ocean regions. Similarly, drought extremes in the tropics associated with El Niño trigger massive biomass burning events that simultaneously modify surface energy fluxes, diffuse light, ozone, moisture availability, and nutrient inputs. Continued investment in field campaigns, such as BER's Green Ocean Amazon campaign, and model analysis of biosphere-atmosphere interactions are essential for making further progress in this important frontier.

Grand Challenge 3.5

Characterize, understand, and model the complex, multiscale water cycle processes in the Earth system including the subsurface to understand and predict water availability and human system response to extremes.

Advancing Water Cycle Research

Freshwater is essential to life and supports myriad human activities that in turn alter freshwater availability. Water cycles through the Earth system as water evaporates from the land and ocean surface to reside as water vapor and clouds in the atmosphere. Precipitation from clouds changes soil moisture, snow and ice, groundwater, and runoff over land, and precipitation and river runoff change ocean salinity and stratification and sea level. Terrestrial moisture status and ocean stratification influence evaporation from the land and ocean surface. Human activities such as water extraction, reservoir operation, irrigation and agricultural production, and industrial and municipal water use have significantly altered the water cycle (Postel et al. 1996; Wada et al. 2010; Richey et al. 2015). Processes that are integral to the water cycle vary over a considerable range of spatial and temporal scales, challenging both observation and modeling to understand, quantify, and predict the properties of these processes as well as their variability and change rates. Water cycle processes are major sources of uncertainty in modeling and quantifying climate sensitivity because these processes play a key role in water vapor, cloud, surface albedo, and carbon cycle feedbacks. Uncertainty in predicting water cycle changes also has direct implications for predicting freshwater availability, sea level rise, and extreme events such as floods, droughts, and storm surges, all restraining the ability to assess human system vulnerability to weather and climate extremes.

The highly coupled multiscale, multiphase nature of water cycle processes has long been a major challenge confounding modeling and prediction. Increasing grid resolution has yielded some improvements in modeling the spatial distribution of clouds, precipitation, soil moisture, and snowpack, but biases in simulating water cycle processes in ESMs remain large.

Despite increasing recognition of the role of subsurface water processes in modulating surface fluxes and plant functioning, growth, and mortality, there are gaps in understanding and modeling subsurface water processes, partly because of limited measurements and modeling across local-to-watershed and large river basin scales. Global estimates of groundwater storage are extremely uncertain, varying by an order of magnitude depending on the data and methods used (Famiglietti 2014; Gleeson et al. 2016). Human systems are beginning to be incorporated into terrestrial models, but their interactive role in the integrated water cycle of the human-Earth system is still largely unexplored.

Resolving these water cycle uncertainties cuts across all BER Earth system science research needs because water is integral to all Earth system processes through its indispensable role in connecting the water, energy, and carbon cycles. Quantifying the spatiotemporal distribution and predicting the changes in freshwater supply, as well as understanding the drivers of extreme events and their changes, are critical for supporting DOE's energy mission as energy production and use are intimately linked to water supply and quality through the energy-water nexus. Therefore, an urgent need remains to characterize, understand, and model the complex, multiscale water cycle for understanding and predicting freshwater availability and assessing the vulnerability of human systems to extremes and prolonged climatic stress.

Research Needs and Knowledge Gaps

Leverage Advances in Global Cloud-Resolving Modeling. The multiscale modeling challenge in Earth system modeling is most acute for the water cycle because errors in modeling water cycle processes at small temporal and spatial scales can upscale through water cycle feedbacks and water-energy-carbon interactions to affect larger-scale and longer-term Earth system predictions. Strong evidence suggests that when atmospheric models resolve clouds and convection, significant improvements in model fidelity are possible, such as more realistic precipitation characteristics (Kendon et al. 2012; Kooperman et al. 2016; Gao et al. 2017) and land-atmosphere feedbacks (Hohenegger et al. 2009; Leung and Gao 2016). Global cloud-resolving modeling with grid spacings of 4 km or less is becoming feasible with advances in computing resources for short-term simulations and weather forecasting. Routinely running cloud-resolving models coupled with land surface models in weather forecasting mode at continental and global scales can provide the foundational knowledge needed to improve the models for use in coupled ESMs. Such endeavors should be key activities of the computational user facilities discussed in Grand Challenge 3.1 on Earth system modeling, p. 24.

Develop Multimodel Ensembles of Cloud-Resolving Forecasts. Collaborations with U.S. and international modeling centers on cloud-resolving weather forecasting is highly beneficial to enable comparison and shared expertise and data for rigorous forecast evaluation and diagnosis. Exploring forecast sensitivity to model formulations and parameterizations of unresolved processes can guide model development. Production of a multimodel ensemble of cloud-resolving forecasts of extreme events that are poorly characterized by observations can advance understanding of their large-scale drivers and predictability. With significant gaps in understanding the governing characteristics of extreme events, such as their intensity, duration, size, frequency, and co-occurrence, another research need is to develop dynamical and statistical theories about which factors control these characteristics. The multimodel ensemble forecasts will provide unprecedentedly high resolution data for knowledge discovery through machine learning and analysis of large datasets, combined with the development of theories and novel metrics and diagnostics.

Improve Convection and Circulation Simulations to Better Project Regional Precipitation Changes. Global precipitation changes in a warmer climate are constrained by the global energy budget to be roughly 2% per degree Celsius of warming (Held and Soden 2006). However, large uncertainty remains in projecting future changes in regional precipitation because precipitation changes can be dominated by poorly constrained circulation changes whose interactions with convection are not well understood or simulated by models (Bony et al. 2015). The substantial variability of large-scale circulation (Shepherd 2014) is only beginning to be characterized using large-ensemble simulations (Kay et al. 2015). Computational advances enabling convection-permitting simulations and large-ensemble simulations may facilitate the development of theories to constrain regional precipitation changes in a perturbed climate. Modeling hierarchies may be used to disentangle different physical and dynamical mechanisms and develop emergent constraints on hydrological cycle changes (Klein and Hall 2015).

Improve Subsurface Process Fidelity in Terrestrial Models. Terrestrial subsurface processes influence the partitioning of surface sensible and latent fluxes and, hence, land-atmosphere interactions and local water recycling. They also directly influence groundwater, an important source of freshwater supply. Measurements have provided evidence that plants can access moisture from diverse sources for transpiration (Thompson et al. 2011), and subsurface hydrology can be strongly influenced by geology and lithology, for example, when rock fractures (Salve et al. 2012) and soil micro- and macropores (Beven and Germann 2013) dominate the flow paths. However, land-surface models used in ESMs have limited representations of soil hydrology, plant-root-soil moisture interactions, surface water–groundwater interactions, and terrestrial-aquatic interactions. These limitations hinder skillful simulation of the spatial and temporal variability of evapotranspiration and recharge to groundwater, with implications for modeling flood generation, precipitation-soil moisture feedback, vegetation drought response, and land carbon sinks. Improving representation of such processes in terrestrial models is an ongoing need. Recent efforts to develop global hydrogeological datasets of soil permeability and porosity (Gleeson et al. 2014), soil depth (Pelletier et al. 2016), and lithology (Hartmann and Moosdorf 2012) provide a significant opportunity for global high-resolution modeling of subsurface processes (e.g., Gleeson et al. 2016).

Understand the Effects of Human-Earth System Interactions on the Water Cycle. Human systems such as dams and reservoirs, irrigation, and crop management have large footprints on the regional and global water cycle. However, the mechanistic nature of the footprints and how they may change in the future are not well understood. Efforts in modeling human systems generally focus on global modeling using conceptual representations or regional modeling with detailed, process-based representations that may not be transferable to the continental or global scale for lack of global input data. Human systems data and coupled models must be developed to explore the dynamics of human-Earth system interactions and their impacts on the water cycle. Combined with data assimilation and cloud-resolving modeling, high-resolution terrestrial models coupled with human systems models may provide much improved quantification of hydrological fluxes and storage to better characterize the water cycle and its historical changes and improve understanding of the drivers of water cycle changes.

Grand Challenge 3.6

Understand the time-dependent processes and mechanisms associated with melting glaciers, ice caps, and ice sheets and their contributions to regional sea level rise.

Predicting Changes in Sea Level Rise

Over the past century, sea level has risen roughly 20 centimeters (cm), largely due to increases in ocean heat content and, to a lesser extent, surface melt from the Greenland ice sheet and global mountain glaciers (Hay et al. 2015). By the end of this century, sea level will rise another 50 to 100 cm, and possibly more, from continued warming and melt (IPCC 2013). The larger concern involves potential changes in the force balances that currently limit rates of ice flow through marine outlets. For instance, changes in the delivery of heat to ice shelves could lead to ice shelf thinning, reducing the buttressing effect of confined outlets. This change could result in significant acceleration of ice flows, potentially activating ice streams deep into the interior of their catchments. Moreover, several catchments are known to reside in marine basins with reverse-slope bedrock geometry, a configuration that has been associated with a hypothesized marine ice sheet instability mechanism. The concern is that changes in ice flow dynamics could significantly increase the rate of sea level rise, potentially generating an additional meter per century or more.

The challenge for predicting ice flow changes is that many of the controlling processes are not directly observable. Around the Antarctic Ocean, for instance, heat delivery to ice shelves is mediated by upwelling of relatively warm, salty water masses onto the continental shelf and by ice-shelf boundary layer processes (see Fig. 3.3. Melting Ice Shelf, p. 38). These boundary conditions and processes are poorly observed and resolved in numerical models. Similar observational challenges affect understanding of subglacial hydrologic systems and fracture mechanics.

Changes in sea level potentially are a predictable and consequential outcome of human activity on the environment, with significant implications for the resilience of coastal and offshore energy infrastructures. BER can play a pivotal role in developing advanced modeling tools that can leverage a limited observational record to interpret the processes involved with these changes. However, developing a modeling capability is only part of the challenge. Longer prediction time scales and the limited observational record will make difficult the testing and evaluation of the contributions of new capabilities to improved performance. The development of

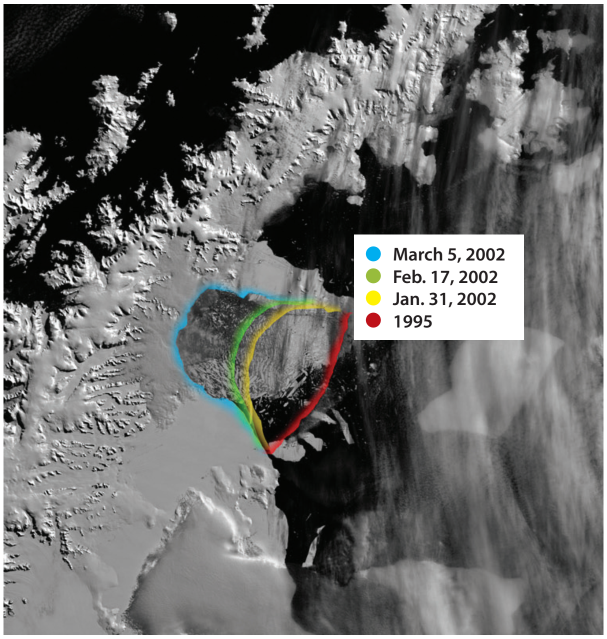


Fig. 3.3. Melting Ice Shelf. The Larsen B ice shelf in Antarctica in 1995 (red) superimposed with its complete collapse in just over a month (yellow, green, and blue). Image captured by the moderate-resolution imaging spectroradiometer (MODIS) on the National Aeronautics and Space Administration's (NASA) Terra satellite between January 31 and March 5, 2002. [Image courtesy NASA]

modeling capabilities within an uncertainty quantification framework is therefore needed. Such a framework can play an important role in synthesizing the modeling and observational components of research endeavors.

Research Needs and Knowledge Gaps

Model Processes and Mechanisms Controlling Ice Melt and Dynamics. The ability to predict decade-to-century changes in global glacial ice inventories and their effect on regional sea level rise depends on understanding and representing within numerical models the time-dependent processes and mechanisms controlling ice melt and ice dynamics. Such predictions also require understanding and modeling the set of gravitational, wind, and ocean current forces that affect the global distribution of ocean mass. The highest-priority model development needs, among others, in this regard are ice-ocean interactions, subglacial hydrology, and ice-fracture mechanics and calving. Because observations of these processes are severely limited, the community will need to adopt a much tighter coupling among observation, theoretical, and modeling activities to advance current understanding and test model predictive capabilities. Polar ice caps integrate and respond to environmental change on all time scales, including decades to centuries. Needed is a better understanding of how ice sheets participate in this variability. Attributing recent changes to particular causes requires recognizing in observations the signature of forced change, relative to unforced variability, coming from ice-ocean interactions, changes in subglacial hydrology, surface melt, or even longer-term adjustments from states established during the Quaternary and Holocene that still reside within the polar ice caps.

Develop Data Assimilation and Inversion Tools to Deal with the Dearth of Observational Data. Many of the scientific aspects to these challenges stem from the limited observational record with which to test hypotheses and advance understanding. A key capacity for dealing with this constraint is to develop tools for data assimilation and data inversion, which can be used to establish initial conditions for near-term predictions of changes in ice mass, as well as plausible model interpretations of uncertain physical properties or uncertain forcings. Such tools need to be refined further to make optimal use of available observations and infer uncertainties. Currently feasible is focusing on a few well-instrumented catchment basins that can serve as testbeds for developing understanding, predictive modeling capability, and future tests to establish model credibility in predictions of sea level rise.

Assess Uncertainties in Data and Modeling Methods and Forge Interagency Collaborations. Pursuing multiple strategies is needed to estimate the boundary conditions and physical processes that affect sea level rise. This endeavor may involve gaining a deeper understanding of the limitations in observational and modeling methods for representing plausible alternatives. In many cases, incorporating statistical methods may be useful for simulating such uncertainties so that their effects may be properly represented in model predictions. Expertise in the observations and products, including ice core and ocean sediment data, radar images of ice sheet stratigraphy, geothermal fluxes, inferences of ocean and ice sheet bed geometry, and subsurface fluid flow, is important to complement modeling capabilities. Given the importance and scarcity of observations, interagency collaborations, particularly with research activities aimed at model-data synthesis and discovery, are highly beneficial.

Grand Challenge 3.7

Quantify the interplay between internally generated climate variability and externally forced response involving anthropogenic and natural factors and their relative roles in the time evolution of regional variability to understand predictability of the Earth system.

Providing Decadal Predictions for Climate–Human System Interactions

Over the next 20 years, policymakers will be planning how to adapt to climate change and effectively mitigate further changes. Climate science will need to provide more accurate climate predictions and more detailed assessments of the overall impacts of changes to Earth and human systems. To achieve this goal, reliable and decision-relevant climate model products with uncertainty information are needed. These products typically are produced by running a model initializing in the mid-1800s. It progresses forward in time, adding observed, time-evolving external forcings such as volcanic eruptions, solar variability, visible air pollution, and increasing GHGs, while looking to the mid-21st century and beyond, where different emissions scenarios provide the major source of uncertainty. For nearer-term time scales, "decadal climate prediction" is an emerging field in climate science that aspires to make near-term predictions for the next 10 to 20 years (Meehl et al. 2014). Credible predictions of climate and human systems, from subseasonal to 10 to 20 years in the future, would support BER in providing a foundational science capability for planning future energy and resource needs. These decadal climate predictions use ESMs that are initialized with a set of climate-system observations at a specific time. This model is run forward for about 10 years, with the objective of simulating both the internally generated variability and response to external forcings. Thus, the challenge for near-term prediction is to quantify the interplay between the internal variability and response to external forcings with the goal of providing climate information to assess impacts.

This scenario can be illustrated by viewing the observational record of 20th century global temperatures as a "rising staircase" (Kosaka and Xie 2016). Even though CO_2 has continuously increased over the 20th and early 21st centuries, the temperature response has not been a continuous increase. Rather, there have been 10- to 20-year periods of accelerated and slower warming (see Fig. 3.4. Recent Slowdown in Global Warming Symptomatic of Decadal Climate Variability, p. 40). Time-evolving changes to external forcings combine with internally generated 10- to 20-year decadal variability to modulate what otherwise would have been a continuous increase of global temperatures due to the steady increase of CO_2. For near-term predictions, the internal variability or noise is a critical part of the system to be predicted, involving the interplay of the internal variability with other externally forced responses.

Research Needs and Knowledge Gaps

Use Hindcast and Multiple Ensemble Approaches to Assess Short-Term and Regional Predictions. For decadal climate prediction, sets of "hindcasts" typically are performed with ESMs to quantify the predictability and reliability of the predictions. These hindcasts involve starting the model at certain times with climate system conditions observed at that time. Application of multiple ensemble methodology has achieved success in NWP for periods of a week or so. Applied to decadal climate prediction, each start date for each hindcast should have multiple ensemble members

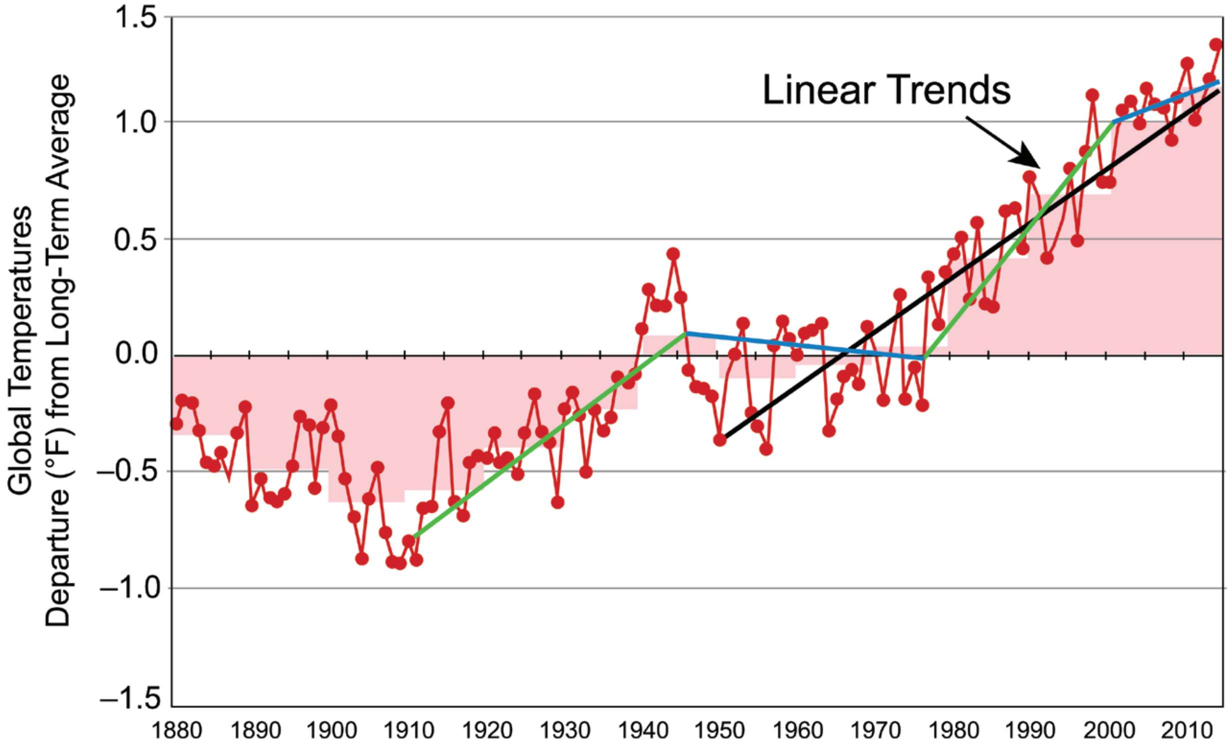


Fig. 3.4. Recent Slowdown in Global Warming Symptomatic of Decadal Climate Variability. Time series of annual mean globally averaged surface temperature anomalies (red dots) based on data from Karl et al. (2015). The black line shows a long-term linear trend, computed from 1950 to 2014, forced mainly by increasing greenhouse gases. Green lines are multidecadal linear trends for positive phases of the interdecadal Pacific oscillation (IPO), blue lines for negative phases of the IPO. Multidecadal trend lines associated with IPO phases generally follow the long-term change in global temperature but differ intermittently depending on the phase. [Image courtesy National Oceanic and Atmospheric Administration. From Meehl, G. A. 2015. "Decadal Climate Variability and the Early-2000s Hiatus," *Variations* **13**(3). U.S. Climate Variability and Predictability Program. usclivar.org/sites/default/files/documents/2015/Variations2015Summer.pdf]

to assess the uncertainty of the predictions on regional scales, as well as for time scales of a season to a decade in advance. More work is needed to explore the ensemble space, especially to provide credible, time-evolving probabilistic regional climate change information.

Develop Best Practices for Initializing ESMs with Observational Data. To improve predictions, a number of actions are needed. When an ESM is initialized with observations at a certain start date, the model starts to drift away from the observed state as a consequence of model biases. Adjustment of these biases is necessary before evaluating the hindcasts and predictions. Improving ESMs would reduce the need for bias adjustments. Additionally, there is no best practice for how to initialize a fully coupled ESM; a more thorough exploration is needed to determine the best way to initialize an ESM with observations of individual or combined states of the Earth system components, with and without data assimilation. Much work to date has focused on predicting ocean temperatures on decadal time scales, since that is where most of the mechanisms thought to produce decadal climate variability reside. More assessments are needed, however, of predictions over land, such as precipitation and temperature, on regional to local spatial scales and from seasonal to 20 years ahead in temporal scales.

Improve Understanding of How the Climate System Works on Decadal Time Scales. Currently, several proposed modes of decadal climate variability exist in the different ocean basins. The fundamental premise of decadal climate prediction is that, if initialized properly, the processes and mechanisms associated with those modes of decadal variability could provide capabilities for near-term regional prediction. Thus, insights into such processes and mechanisms would be essential to understanding how internal climate system variability interacts with the response to external forcing in decadal climate predictions, depending on analyses of observations in conjunction with climate model simulations.

Test Decadal Predictions Using Weather Forecasting and Multimodel Ensembles. The atmospheric components of ESMs used for decadal climate predictions are very similar to those used in operational NWP. Atmospheric ESM components should be tested by using them to perform weather forecasts, as discussed in Grand Challenge 3.1 on Earth system modeling, p. 24. The coupled atmosphere-land-ocean models of the ESMs also can be tested in subseason-to-season (S2S) forecasting mode. Interagency collaborations provide an opportunity for comparison and analysis of a large multimodel ensemble, which may elucidate coupled model behaviors relevant to decadal predictability and prediction skill.

Grand Challenge 3.8

Understand the long-term Earth system stability in response to possible future Earth system outcomes and address the level of confidence and identify emergent constraints for the range of model projections.

Predicting Potential High-Impact Earth System Thresholds and Feedbacks

Current assessments of climate change risks do not adequately account for the high-risk, low-probability outcomes associated with threshold behavior within vulnerable systems. Some examples of vulnerabilities include significant mass wasting of the polar ice caps, appreciable carbon releases from stores within marine sediments, and the possible prolonged droughts or regime shifts in tropical hydroclimates. The potential for threshold behavior within these systems is supported by a large diversity of observational evidence interpreted from paleoclimate archives. The mechanisms and forcings involved with triggering these events are still largely unexplained. Much remains to be understood about potential threshold behavior within terrestrial and marine ecosystems. Current model development practices are not ideally set up to identify such vulnerabilities, much less quantify their likelihood.

As the climate changes, humans will continually adapt their behavior, both consciously and unconsciously, in response. For instance, electrical demand will increase as warmer climate increases the need for air conditioning. While some behavior changes affect carbon emissions in a relatively straightforward manner, some human-climate interactions are far more complex. For example, in many low-stabilization scenarios, substantial bioenergy requirements necessitate that large portions of arable land be devoted to energy production at the possible expense of food production and without consideration of the water supply needed to support production of other crops or energy sources. The impacts on the global agricultural economy and water resources may be significant but are poorly quantified. Such stresses likely will trigger socioeconomic and technical changes that could alter population trends, energy demands, and other societal behaviors relevant to GHG emissions. These feedbacks also could trigger instabilities in the coupled human-Earth system, either through the feedbacks themselves or by pushing the climate system past a critical threshold. Many decision makers require information about high-risk, low-probability outcomes, but current scenario development is based on plausible pathways with little or no estimate of their likelihood. Outlier scenarios typically are outside those generally used to force model simulations of future climate. However, improved understanding of the potential for high-impact Earth system threshold behavior as well as human-Earth system feedbacks will lead to more realistic assessments of the feasibility of future scenarios and risks. This grand challenge will better inform development of appropriate U.S. energy policies intended to achieve specific targets. Just as insurers need to correctly price their policies for high-risk, low-probability events, cost-effective climate change adaptation policies also must be developed with the consideration of such outcomes.

Research Needs and Knowledge Gaps

Investigate Plausible High-Risk Scenarios Exploring Potential Earth System Threshold Behavior. Much of the community's effort is directed toward developing optimal model configurations based on the observational record of the past few decades. This focus has limitations given the value of information about potential extremes within a risk assessment. More lateral thinking is needed concerning methodology to explore and test alternate hypotheses for the physics of climate. As an example, an "abrupt change early-warning system" can be used to prioritize observation, modeling, and theory toward critical Earth system components where the risk of threshold behavior could be imminent (NRC 2013). Analysis of high-risk, low-probability climate scenarios could employ advanced statistical techniques. While more can be done to generate quantitative information about uncertainties in climate model predictions using state-of-the-art ESMs, the limiting factor in making significant advances about low-probability, high-risk scenarios is the community's scientific creativity aided by mathematical and computational wherewithal.

Improve Human-Earth System Modeling Capabilities. To discover how human-climate feedbacks could trigger instabilities, more work is needed to fully couple IAMs with ESMs (e.g., see www.globalchange.umd.edu/iesm/) and use these tools to construct large ensembles of simulation databases to more fully explore possible interactions. However, more research is required to understand when offline IAMs versus coupled human-Earth system models are needed and to develop more robust, offline IAMs for interrogating datasets to look at human feedbacks. Higher resolution should be a target for IAM improvement, as well as comparison with observations and evaluation of parametric and structural model uncertainty using emulators. Also needed is representation of a broader range of interactions between human and Earth system processes. There are three essential areas of model development: (1) human-Earth interactions at fine scales such as processes in urban population centers and coastal environments are not well represented in the current generation of models; (2) more holistic representations of carbon-water interactions are important to account for human and ecosystem response to perturbations and their feedbacks to the Earth system; and (3) representing potential feedbacks involving migration, population, and labor productivity is important to explore their significance in projecting future changes.

Improving human-Earth system modeling capabilities will provide the tools needed to characterize how humans respond to climate change and how those responses affect further climate change, influencing the Earth system trajectory and the realization and stability of future scenarios. There are implications of long-term outcomes for near-term climate, particularly related to the need to understand negative emissions. More attention needs to be paid to reversibility and hysteresis on a variety of temporal and spatial scales. Lastly, closer programmatic coordination via common risk analyses should target areas of DOE interest, including risks to energy and water demands.

4 Grand Challenges in Microbial to Earth System Pathways

Overarching Grand Challenge and 20-Year Vision

Define the levels of biological organization most relevant to scaling from single cells to ecosystems and global cycles; capture how that organization varies in time and space; and identify critical interactions that dictate rates of carbon, nutrient, and energy transformation in different environments.

Determining the Pathways that Link Microbial Activities to the Earth System Scale

Microbes affect atmospheric and climate change, energy production, and remediation of legacy waste. Moreover, they link the carbon, nitrogen, and water cycles by catalyzing ecological processes and biogeochemical transformations that are strongly mediated by water availability and that affect carbon and nutrient cycling and transfer between the biosphere and the atmosphere. Overall, this high-order grand challenge seeks to determine the pathways that link microbial activities to the Earth system scale, identifying (1) the strong interactions and interactors (i.e., the keystone taxa) that influence these pathways, (2) how microbial communities change in response to environmental changes (i.e., as responses and feedbacks), and (3) how these communities might be managed to elicit desired outcomes of Earth system functioning (see Fig. 4.1. Microbial to Earth System Pathways, p. 45).

The long-recognized linkages between Earth system science and microbiology are central to developing an integrative and practical understanding of how energy use affects the environment. Though individual microbial cells have negligible impact, when summed their influence is massive, often dominating elemental transformations at a hierarchy of spatial and temporal scales. These combined impacts range from the blooms of microbial activity around resource "hot spots" to processing and filtering of contaminants across landscapes, to influencing soil structure, to spreading disease, to determining trace gas fluxes that affect the composition of Earth's atmosphere and the global climate. Microbes catalyze major transformations in every elemental cycle relevant to carbon, water, and energy. Not well known, however, is how these activities respond to changing edaphic drivers and how they shape the direction and pace of those changes.

The biological diversity of microorganisms is far deeper and broader than diversity among plants and animals, but knowledge of how this genetic potential is distributed across the globe is growing rapidly. The hackneyed "black box" is now at least partially translucent. Yet, understanding is still lacking regarding whether and where diversity matters for the processes through which microorganisms affect the Earth as an ecological system.

The Biological and Environmental Research Advisory Committee's (BERAC) vision for microbial to

Microbial to Earth System Pathways
Summary of Grand Challenges and Action Items

Microbial communities influence soil and plant systems that affect regional and global environments. However, research at one scale does not always translate to other scales because of challenges in data formats, relevance at different scales, and fundamental understanding of the links among scales. To help determine the pathways that link microbial activities to the Earth system scale, BERAC has identified four grand challenges that are outlined here and then described in more detail in the Grand Challenge Research Recommendations section, p. 46:

4.1. Characterize the biogeochemical exchanges driven by food web and plant-microbe interactions and evaluate their process-level impacts, sensitivity to disturbances, and shifting resource availability under changing environmental regimes.

4.2. Define the sphere of influence and key elements of microbial communities in space and time relevant for predicting larger-scale ecosystem phenomena for Earth system understanding.

4.3. Integrate molecular and process data to improve the ability to define ecologically significant traits of individual taxa and communities and use trait-based models to develop predictive links between community dynamics and ecosystem processes.

4.4. Align and deepen connections among conceptual understanding, measurements, and models related to the roles of microbes in determining the rate of transformation, uptake, and loss of chemical elements from ecosystems.

Associated with the grand challenges are four action items recommended for initiating progress toward the challenges:

- Develop and deploy sensors suitable for use in the field that integrate the effects of microbial activity at the ecosystem scale.
- Promote integrated studies that explicitly test predicted microbial network interactions and attempt to assess membership and species-specific and collective functional capabilities within ecologically coherent microbial communities.
- Conduct experiments that help determine the influence of microbial processes at larger, aggregate scales.
- Promote research teams to integrate microbial community dynamics into Earth system models.

Earth system pathways is a more complete translation between microbial ecology and Earth system modeling for improved prediction and management. To achieve this vision, BERAC has identified four grand challenges (see Microbial to Earth System Pathways Summary of Grand Challenges and Action Items, this page). Successfully addressing these challenges revolves around (1) identifying microbial traits that can be used to organize microbial diversity and enable prediction of how microorganisms function in nature as cells, populations, and communities; (2) recognizing that the functional roles of microorganisms rely on the complex network of ecological interactions that shape their biogeochemical impact; and (3) integrating microbial attributes quantitatively and describing the functional significance of microbial biodiversity for Earth system processes and their representation in Earth system models (ESMs).

Building on Progress Related to the 2010 Grand Challenges

As noted previously, the concept that microbial processes influence the Earth system is not new. Although this focus was not explicit in BERAC's

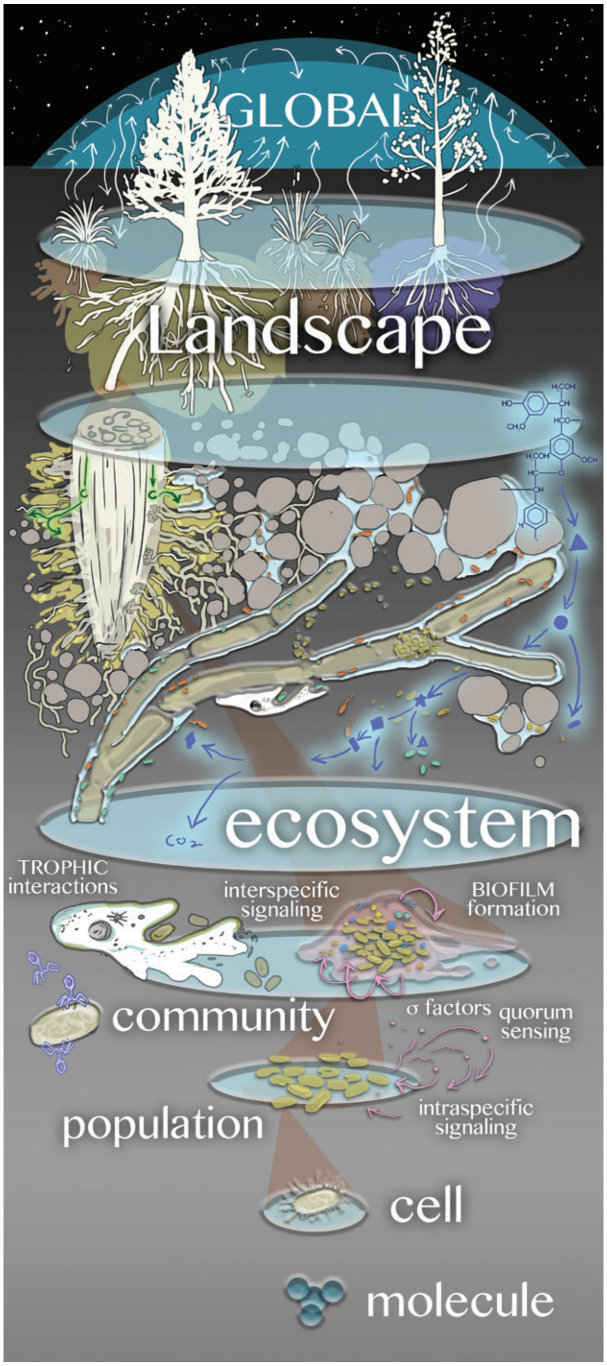


Fig. 4.1. Microbial to Earth System Pathways. This diagram illustrates the range of scales—from molecular to global—through which microorganisms influence the Earth system. [Image courtesy Victor Leshyk]

20-year visioning effort published in the 2010 Grand Challenges report (BERAC 2010), the intention is implicit in many of the challenges that were identified. The topic has been elevated for this current report because connecting the understanding of microbiome function across scales of biological complexity has become increasingly important.

Substantial progress has been made for several of the specific goals identified in the 2010 report. Under "enabling predictive biology," for example, the challenge was to develop technologies and models for associating biological activities at the microscale (i.e., dimensions of individual cells) with processes at the macroscale, having the ultimate goal of achieving a more predictive biology. The realization of complete *in situ*, single-cell omics (i.e., for metabolites, transcripts, and proteins) is on the near-term horizon. Recently cultured, novel microbes can rapidly be made genetically tractable and their genomes functionally annotated using high-throughput technology. Advanced spectrometric instruments have greatly enhanced the sensitivity and detection of metabolites and signaling molecules associated with the transformation of carbon and energy in the laboratory and environmental systems. Significant progress has been achieved in assembling simple laboratory microbial communities that recapitulate fundamental biotic interactions and system-level processes. Still a challenge, however, is developing more complex model communities that more fully capture environmentally relevant interaction networks and incorporate the plasticity and resiliency of those networks. Those types of model systems will be critical for (1) identifying the interdependencies most sensitive to environmental change, as well as the general design principles of natural systems (a challenge specified in the 2010 report), and (2) modeling and testing the relevance of those design principles in open environmental systems. Also still urgently needed, however, is the ability to measure gross rates (fluxes) of materials passing between organisms and their environment. Such information will be critical for predicting and manipulating the types and rates of ecosystem processes and response feedbacks influenced by climate change.

Some of these earlier unachieved challenges require rescoping and reformulation to reflect new understanding of system complexities. Clearly, functional metagenomics alone will not enable mass balance closure for biogeochemical cycles or be sufficient to predict organismal fitness, but it must be more fully integrated with modeling and complementary analytical measurements. The more immediate decadal challenge is to improve functional gene annotation methods to achieve a better approximation of phenotypes inferred from genotypes. This chapter outlines an updated set of challenges for (1) defining the different levels of biological organization most relevant to scaling from single cells to ecosystems and global cycles; (2) capturing how that organization varies in time and space; and (3) identifying interactions most critical to controlling the rates of carbon, nutrient, and energy transformations in different environments.

Grand Challenge Research Recommendations

Grand Challenge 4.1

Characterize the biogeochemical exchanges driven by food web and plant-microbe interactions and evaluate their process-level impacts, sensitivity to disturbances, and shifting resource availability under changing environmental regimes.

Microbial Interactions Structuring Earth System Processes

Microorganisms live in the most complex ecological communities on Earth, and their interactions with plants, insects, and animals, while intimate in scale, influence processes at much larger scales. The complexity of microbial communities in taxonomic diversity, biochemical versatility, and evolutionary breadth and depth is both awe-inspiring and an immense grand challenge for any kind of practical synthesis. Understanding the nature, consistency, and organizing principles underlying such interactions is just beginning, yet there is reason to expect that such interactions imprint on ecosystem biogeochemistry in consistent ways, enabling prediction.

Full knowledge of Earth system consequences of interactions in microbial communities will improve the understanding of energy and material flow in ecosystems, as well as the ecological significance of the full range of organisms in microbial communities—bacteria, fungi, protozoa, metazoan, and viruses, including individual taxa or strains that may play keystone roles in such interactions. Identifying these interactions, the environmental influences on them, and their ecosystem-scale consequences should improve prediction by illuminating key pathways to Earth system processes.

Research Needs and Knowledge Gaps

Quantify and Predict Complex Ecological Networks. Within a habitat, organisms interact with each other via flows of energy, matter, and information to form complex ecological networks, including (1) trophic interactions described in food webs (Montoya et al. 2006; Bascompte 2007); (2) signaling networks [e.g., those involved in biofilm formation, quorum sensing (DeAngelis 2016), or the development of symbioses (Neal et al. 2012)]; and (3) cascades of metabolites as communities use a variety of substrates, as well as the resulting waste products, necromass, and persistent soil organic matter (SOM; Kallenbach et al. 2016). These interactions cause nutrient elements and carbon to cycle both into and out of microbial biomass, affecting organic matter stabilization (Kindler et al. 2009; Schweigert et al. 2015), plant nutrient availability, and conversion of microbial cell carbon to carbon dioxide (CO_2; Clarholm 1985; Griffiths 1994; Bonkowski 2004; Fox et al. 2006). The major players in these interactions include plants, bacteria, archaea, fungi, protozoa, metazoans, and likely many phages and viruses not yet discovered. Quantifying and predicting such networks in terms of their composition, structure, dynamics, and biogeochemical impact remain extremely challenging, yet finding generalities in such networks could enable prediction.

Model Elemental Dynamics. Predictive food web models have been developed (Berlow et al. 2009), but the complexity and cryptic nature of soil food webs have hampered understanding (Brose and Scheu 2014). With the advent of more powerful omics and isotopic techniques, the ability to model elemental dynamics in complex ecosystems is improving (Digel et al. 2014), such as identifying unique structural features of soil food webs (Riede et al. 2010) and exploring the effects of temperature and moisture on soil food chains (Binzer et al. 2012; Lang et al. 2012; Lang et al. 2014). Next-generation sequencing approaches, including ever-growing reference sequence collections of mitochondrial and nuclear DNA sequences for higher eukaryote taxa, are making accurate predictions of species composition and richness more feasible for mixed environmental samples (Taberlet et al. 2012; Leray et al. 2013). The next challenge is to identify components of interaction networks that are quantitative and repeatable, and thus predictive of energy and nutrient flow through microbial ecosystems (see also Chapter 8 section, Modeling and Engineering, p. 110).

Detect and Quantify Ecosystem Consequences of Plant-Microbe-Soil Interactions. Interactions between microorganisms and plants have a major influence on plant productivity and biogeochemical fluxes from the ecosystem to the global scale. Carbon flow from plant roots to soil is the primary source of stabilized soil carbon, provides much of the energy that powers soil microbial food webs, drives production and consumption of trace gases, and shapes the nutrient cycles that influence plant growth. Global networks that monitor ecosystem-atmosphere exchange of CO_2 (i.e., eddy flux networks) point to root exudation as a quantitatively significant flux in the carbon cycle, but its effects on nutrient cycling, balance of SOM formation and decomposition (i.e., "priming"), and trace gas production are not quantitatively represented in models because the controls are not well parameterized. Unlike leaf-level processes, which can be scaled using physical and radiative transfer physics, belowground processes are highly variable owing to the soil environment's heterogeneity and the diverse strategies that plants have evolved for resource acquisition (Reich 2014), both of which make linking belowground and aboveground processes challenging at the scale of individual plants, plots, or ecosystems (Kramer-Walter et al. 2016). Details are emerging of how communication among microbes, and between microbes and plants, affects symbiotic associations (Lareen et al. 2016; Rubin et al. 2017) and biofilm formation (Flemming et al. 2016), but translating these quantitatively to ecosystem processes is not yet possible. Thus, new methods are needed for detecting and quantifying the ecosystem consequences of plant-microbe-soil interactions *in situ* (i.e., to minimize biases introduced by isolating plants and microbes in artificial media or under controlled laboratory or greenhouse conditions). Also needed are new conceptual frameworks that enable modeling of these processes at multiple spatial and temporal scales.

Meeting this grand challenge requires a systems approach to microbial to Earth system pathways, where microbial interactions with biogeochemical processes will be observed *in situ* and mechanistically linked to ecosystem-scale consequences. Synthetic systems and enrichment cultures will be used to better understand what is first observed in nature. BERAC recommends several intermediate steps for advancements toward Grand Challenge 4.1:

- Support research that applies stable isotope probing and other labeling techniques to quantitatively trace elemental flows and reveal structuring interactions *in situ*. Such a powerful suite of tools can physically couple elemental flow to biodiversity by measuring isotopes in cells or biomolecules that contain taxonomic information.

- Identify sentinel molecules that indicate (1) the presence, nature, and degree of ecological interactions in complex communities and (2) metabolic status of organisms involved, integrating their significance in ecosystem functioning. These molecules include compounds in both living cells and microbial “remains” [e.g., cell walls, extracellular polymeric substances (EPS), and exoenzymes] that capture the salient components of microbial community, ecology, interactions, and function. Such compounds would help to integrate microbial effects in a targeted way relevant to parameterizing models at higher scales. Examples include indicators of metabolic stresses from environmental change—stresses such as nutrient, moisture, and temperature (e.g., common genetic regulatory stress-response systems, such as sigma factors, as potential integrator signaling molecules).

- Improve quantitative understanding of how root exudates drive specific microbial activities (particularly those related to priming, the deconstruction of exudate compounds versus lignocellulose) and microbial responses that affect plant growth.
- Incorporate local patch-scale dynamics of interkingdom interactions (e.g., plant, fungal, and bacterial) to reconstruct ecological interaction networks and to understand the impact of microbial interactions on SOM formation, mineralization, and stabilization.
- In collaboration with national laboratories and equipment facilities, develop and deploy sensors suitable for field use (i.e., operable without pumps and on solar power) that integrate the effects of microbial activity at the ecosystem scale [e.g., concentrations of hydrogen, nitrous oxide (N_2O), CO_2, methane (CH_4), oxygen, and porewater dissolved organic carbon, as well as the isotopic composition of these analytes].
- Develop new targeted efforts to analyze consortia of interacting microorganisms to isolate and understand the strong interactions detected in nature, including multispecies interactions that span trophic groups such as bacteria, archaea, protists, and viruses.
- Request research proposals to (1) identify the molecular signatures of microbial interactions and physiology that drive and respond to biogeochemical and environmental conditions; (2) quantify *in situ* energy budgets, element fluxes, and population dynamics in microbial ecosystems, identifying the subset of biotic and abiotic interactions that contribute to soil carbon stabilization, pollutant degradation, and their sensitivity to a changing environment (e.g., moisture, temperature, and atmospheric CO_2 concentrations); and (3) develop new tools to determine the roles of specific organisms *in situ*, including tools that enable targeted knockouts of individual organisms (e.g., phage therapy) or specific genes [clustered regularly interspaced short palindromic repeats (CRISPR)] to test how removal of a key taxon or gene affects broader community dynamics.

Grand Challenge 4.2

Define the sphere of influence and key elements of microbial communities in space and time relevant for predicting larger-scale ecosystem phenomena for Earth system understanding.

Microbial Interactions and Scales in Space and Time

The collective activities of interacting microorganisms existing within communities have a significant impact on physicochemical processes in the environment, spanning from biogeochemical transformations of major and minor elements to physical alterations through processes such as biofilm formation, mineral precipitation, and weathering. Yet the ability to recognize the boundaries of a microbial community in space and time is limited. Common descriptions of microbial communities are derived primarily from an artificial construct of microorganisms recovered in a sample rather than the actual ecologically relevant interacting network of microorganisms within a community (see reviews by Konopka 2009 and Vos et al. 2013). Given the small spatial scales and interaction spheres between co-occurring microorganisms, most environmental samples collected for molecular microbial diversity analysis (e.g., sediment, soil, groundwater, or rhizosphere) likely harbor multiple microbial communities and microenvironments. New strategies are essential for enabling more robust and meaningful identification of the true membership, spatial distribution, cellular activities, and interactions within a cohesive microbial community that, in turn, influence an ecosystem's large-scale geochemical and physical processes.

Research Needs and Knowledge Gaps

Develop Reliable Metrics for Investigating Spatiotemporal Interactions. Reliable metrics are urgently needed to visualize and define the spatiotemporal boundaries, membership, and interactions among microbial communities (e.g., in soil). Development of these metrics will enable investigation of several key questions: (1) Are there interdependencies or synergisms among microbial assemblages on discrete soil or sediment particles, or does each particle constitute its own ecological community? (2) If interactions and emergent metabolic processes occur at the multiparticle scale, what is the spatial and temporal extent of these interactions? (3) Can the boundaries of ecologically coherent communities be experimentally tested and modeled?

Assess the Spatial Variation of Microbial Communities. To date, only a handful of well-replicated field investigations have attempted to assess the spatial variation of microbial communities (i.e., diversity) by sampling along well-defined transects spanning millimeters to meters in scale (Martiny et al. 2011). Other studies have documented microscale heterogeneity in metabolic activity on particles [e.g., organic carbon respiration without the corresponding information about the associated microbial diversity (Gonod et al. 2003)]. To understand and define what constitutes a coherent microbial community, information regarding microscale diversity must be connected with data characterizing the variation in microbial spatial distribution, species-specific activities, and emergent community-wide influence on relevant biogeochemical processes.

Geostatistical approaches at the phylotype or gene level show some promise for assessing spatial relationships (e.g., Nunan et al. 2002; Grundmann and Debouzie 2000). Also promising are developments using network modeling of 16S ribosomal RNA (rRNA) amplicon operational taxonomic units (e.g., Barberán et al. 2012; Furhman and Steele 2008) and metagenomic datasets (Lima-Mendez et al. 2015) that offer large-scale assessments of central microbial interaction nodes based on co-occurrence analysis. With the exception of Lima-Mendez et al. (2015), the majority of network studies thus far have not been extended beyond *in silico* predictions. Statistical associations often are inferred from coarse-grained samples, which are collected at the plot or ecosystem scale and are not representative of interactions occurring at the scales of local communities or populations. The assembly of discrete microscale communities likely dictates the extent of interacting microbial assemblages within a microbial ecosystem and, in turn, how interactions among microbial assemblages shift in response to changing environmental conditions. Studying microbial communities in high replication at the local patch scale is necessary to identify ecological interaction networks and to model the functional impact of microbial interactions. To begin to define the boundaries and structure of ecologically coherent communities, the next phase of research should evaluate cross-species metabolite exchange potentials based on genome-informed prediction of nutrient dependencies and community metabolic network models. Experiments should be developed to generate and test network-based hypotheses, employing molecular and analytical visualization methods, stable isotope probing, microfluidic devices, and other complementary methodologies to visualize, analyze, and track interactions among species and assemblages in a spatially explicit manner, including computational statistical analyses that incorporate spatial and temporal dynamics.

Characterize Microbial Community Boundaries and Spatiotemporal Dynamics. Models of major microbial community activities and interactions often produce unreliable predictions of microbial community stability, resiliency, and emergent properties, especially under fluctuating conditions or in highly heterogeneous environments. Current modeling efforts are confounded by even simple attributes (i.e., defining the coherent network of interacting microorganisms that compose an ecologically relevant community in nature, or defining the taxa that are metabolically active versus those that are dormant or represented only by "relic DNA"). Developing new approaches to characterize the boundaries and spatiotemporal dynamics of microbial communities will assist with optimizing the forecasting models of community-level activity and response to environmental change.

BERAC recommends several intermediate steps for advancements toward Grand Challenge 4.2:

- Advance and incorporate methods and approaches that facilitate the study of microscale processes, enabling multiscale investigations that transition from the bulk analysis (i.e., grams to milligrams) of soil to the fine-scale (i.e., millimeters to microns) particle mapping of cellular activity, behavior, and interactions. Independent methodologies are needed to validate current approaches based on 16S rRNA sequence or gene-based network predictions of microbial interactions. Incorporation of single-cell omics techniques; advanced light, electron, X-ray, and ion microscopy; and further tracer development of microbial activity *in situ* (e.g., stable isotope probing or fluorescently labeled substrates) all represent promising directions that will continue to advance the study of microorganisms within the context of their community in complex natural environments.
- Develop new computational statistical tools to incorporate spatial and temporal dynamics and enable interpretation of data-driven microbiome experiments.
- Encourage integrated studies that explicitly test predicted microbial network interactions and attempt to assess membership, species-specific, and collective functional capabilities within ecologically coherent communities.
- Develop a new U.S. Department of Energy (DOE) laboratory service (in collaboration with the National Institute of Standards and Technology), focusing on the development of fluorescent probes and tracers for biomolecule detection and synthesis of isotopically labeled substrates. The availability of well-characterized, isotopically labeled substrates and new fluorescence-based tracers will relieve some of the current limitations of microscopy-based assays for direct quantification of microbial activities, food web interactions, syntrophic associations, and the net flux of carbon and energy within coherent microbial communities.

Grand Challenge 4.3

Integrate molecular and process data to improve the ability to define ecologically significant traits of individual taxa and communities and use trait-based models to develop predictive links between community dynamics and ecosystem processes.

Trait-Based Approaches to Scaling Up

Considering the dynamic nature of microbial interactions may help identify when and where molecular-scale information is important and where minimal reductionist information is sufficient. A trait-based approach provides a tractable and reproducible framework for identifying the physiological characteristics that determine the contribution of microbial communities to a biogeochemical process. This framework, in turn, will enable connection of microbial mechanisms at multiple scales to quantify the integrative effects that govern net effects at the ecosystem scale. As modelers and empiricists collaborate on community phenotypes, empirical studies will generate more appropriate data, and models will improve aggregation of hyperdiverse microbial communities into tractable functional units. As more collaborative data are collected to define the relationship between microbial composition and ecosystem functioning, incorporating microbial interactions into process models will become increasingly valuable.

Research Needs and Knowledge Gaps

Identify Ecologically Significant Traits of Microorganisms and Microbial Consortia. Microbial metabolism cycles elements within and through ecosystems. Representing this influence in predictive models makes sense, and doing so has reduced uncertainty (e.g., Wieder et al. 2013). Efforts are still rudimentary, however, and do not account for microbiome variation among systems or in system-specific responses to changing environmental conditions. Incorporating the spectacular diversity of traits present in the microbiomes of Earth's ecosystems remains challenging. An essential challenge in scaling from microbial (or genomic) diversity to ecosystem function stems from the need to define the ecologically significant traits of microorganisms and microbial consortia, quantify their variation *in situ*, and test whether useful simplifications could reasonably represent the vast diversity while remaining quantitatively tractable. In most cases, diverse soil microorganisms function as networks of interacting organisms, or consortia, rather than as physiologically autonomous cells. Identifying these functional modules, or community phenotypes, that can be mapped to system-level processes may offer a promising approach for enabling prediction of the emergent outcomes of community metabolism.

Defining Traits

Traits broadly encompass the physiological, morphological, and behavioral characteristics of an organism, species group, or functional guild. A microorganism's simplest traits are encoded by just one genetic locus; thus, an organism's genotype matches the potential phenotype. Traits also relate to definitions of a population (i.e., genetically cohesive) and strain (i.e., genetic sequence variation).

Isolated microbial representatives do not represent the full spectrum of traits found in nature (see sidebar, Defining Traits, this page). Conversely, within some microbial groups such as organisms that conduct oxygenic photosynthesis or methanogenesis, traits may exhibit coherence among natural taxonomic groups defined by phylogenetic relationships; they also can map broadly to Earth's different ecosystems (Martiny et al. 2013). Examples include (1) the exclusive provenance of certain marine bacteria, such as the cyanobacterium *Prochlorococcus* and the heterotroph *Pelagibacter*, to the world's oceans and (2) the preference of specific groups of nitrifying microorganisms for either marine or terrestrial systems. Those patterns can be associated with well-defined traits of biogeochemical importance. Photoautotrophy by cyanobacteria, the specialization to mineralization of organic carbon in low-nutrient environments by *Pelagibacter*, and the oxidation of ammonia by both archaea and bacteria are biological processes of fundamental importance to planetary carbon and nitrogen cycles (see Fig. 4.1, p. 45). However, variations in physiological traits influencing the biogeography of major functional guilds—such as adaptive differences among guild members to changing pH, temperature, light, or nutrient concentration—are not captured in these very broad correlations between phylogenetic affiliation and function. For example, different genetically distinct species groups of *Prochlorococcus* vary in basic adaptive traits, including variable resistance to virus attack and optimum conditions of light, temperature, and nitrogen required for growth (Martiny et al. 2009). In turn, these trait differences have been shown to govern oceanic distribution patterns, suggesting that understanding this precision of trait assignment is fundamental to predicting ecosystem response to environmental change.

Link Patterns of Microbial Diversity with Ecosystem Function. Advancing understanding of traits in an ecological context requires conceptual and mathematical models that move beyond genetic potential to characterizing the realized phenotypes (i.e., traits) of microbes and, where possible, the integrated traits of the communities in which they occur

(Allison 2012). For example, DOE's Office of Biological and Environmental Research (BER) supports Next-Generation Ecosystem Experiments (NGEE) in the tropics and the Arctic (see also Chapter 7 on User Facilities and Research Infrastructure, p. 85). These NGEE projects seek to develop process-rich ecosystem models, employing a trait-based modeling approach. Defining key traits is an essential foundation for understanding how traits interact within an organism in response to other organisms (e.g., competition and mutualism) and the environment (e.g., niche partitioning and substrate access) and is a critical objective for identifying and predicting realized phenotypes under field conditions. A trait-based approach could provide an essential framework for linking patterns of diversity with ecosystem function and predicting the response of those systems to environmental change.

Identify Dominant Microbial Interactions that Influence Biogeochemical Cycling. Genomic studies suggest that metabolic complementation regulates biogeochemical pathways in natural microbial communities (Zelezniak et al. 2015). Interactions through metabolic complementarities are common in nature and have potentially large impacts on the function of microbial consortia and, in turn, ecosystem microbiomes. The interactions of microbial consortia also can affect the spatial segregation of consortia among niches, as well as the community's stability, by impacting the potential to buffer environmental change. Therefore, spatially explicit studies that identify the metabolic response of discrete consortia to the local nutritional and physicochemical environment present a tractable approach for identifying the dominant microbial interactions influencing biogeochemical cycling in a diverse range of ecosystems.

Incorporate Microbiome Variation into Earth System Models. Incorporating microbial processes into ESMs is known to improve uncertainty surrounding carbon and nutrient cycling (Wieder et al. 2013), but ESMs have incorporated few of the microbial parameters needed to account for microbiome variation among systems or system response to changing environmental conditions (see also Chapter 3 on Earth and Environmental Systems Sciences, p. 21). Doing so remains a significant challenge because the potential for generating new knowledge from metaomic data still is unfolding. Annotation also remains a key challenge, and sequencing data—even with more robust annotation—represent only microbial potential, not realized phonotypes.

BERAC recommends several intermediate steps for advancements toward Grand Challenge 4.3:

- Solicit proposals that (1) develop models to identify microbial interaction scenarios and thus narrow the set of molecules for targeting in empirical studies; (2) identify chemical signatures (i.e., genes, transcriptomes, metabolites, and isotopic tracers) that link biotic responses to environmental changes with ecosystem feedbacks critical to the carbon and water cycles; (3) characterize the dynamics of microbial interactions and how their plasticity drives organism and ecosystem resiliency in responses to seasonal changes and extreme events; and (4) identify markers for community traits that may scale from cells to systems. To enable their incorporation into ESMs, traits must be considered in new ways—using culture-independent strategies that include isotopes, remote sensing, and changing molecular profiles (e.g., genes and transcripts) across seasons, extreme events, and gradients to identify markers for environmentally significant biological properties. Trait dynamics must be understood, especially for biogeochemistry, capturing how hot spots and "hot moments" drive net effects on ecosystem processes. Identifying traits, both stable and dynamic, will define how plasticity drives organism, community, and ecosystem responses to changes.
- Facilitate cross-disciplinary teams that examine ecosystems across organisms (i.e., plants and microbes), identify important traits across scales (i.e., watershed, ecosystem, community, organism, and molecules), and bring together

scientists, engineers, and modelers. Significant gaps among scientific domains limit the ability for molecular scientists to generate data useful for microbially explicit models. Similarly, modelers are challenged with identifying the important properties and processes governing ecosystem- to watershed-scale biogeochemical processes, which necessitates new funding opportunities encouraging cross-domain investigations to enable studies that develop new and complementary empirical, analytical, and modeling approaches to foster understanding across scales. Enabling cross training and literacy among domains will enable large advances in integrating high-resolution, spatiotemporal data to inform ESMs.

Grand Challenge 4.4

Align and deepen connections among conceptual understanding, measurements, and models related to the roles of microbes in determining the rate of transformation, uptake, and loss of chemical elements from ecosystems.

Microbial Processes in Earth System Models

The persistence and stability of organic carbon in soil was long thought to be dominated by its chemical "quality"—its elemental composition and the nature of chemical bonds affecting its use as a microbial substrate—modulated by factors like temperature and moisture. The environmental and edaphic context is indeed important, but a new view has emerged that emphasizes soil minerals and microbial ecology as important controls of the formation, stabilization, and decomposition of organic matter in soils (see sidebar, Soil Organic Matter, p. 54). This new paradigm has not been implemented in soil biogeochemical models that predict trace gas fluxes of CO_2, CH_4, and N_2O or the persistence of contaminants in the environment. Developing this capacity is imperative, and new initiatives are needed to develop a mechanistic understanding of (1) how the physicochemical protection of SOM determines its long-term fate, compared to its molecular composition (e.g., chemical recalcitrance); (2) how microbial communities, especially those found in the rhizosphere, play a vital role in the formation, stabilization, and decomposition of SOM, including the role of microbial necromass and extracellular products; and (3) how interactions between microbes and minerals at the soil pore scale govern carbon and nutrient retention and loss in ecosystems, trace gas exchange between ecosystems and the atmosphere, and persistence and fate of pollutants in watersheds. Concurrently, special attention must be given to developing numerical tools that tractably transfer these insights to societally relevant scales.

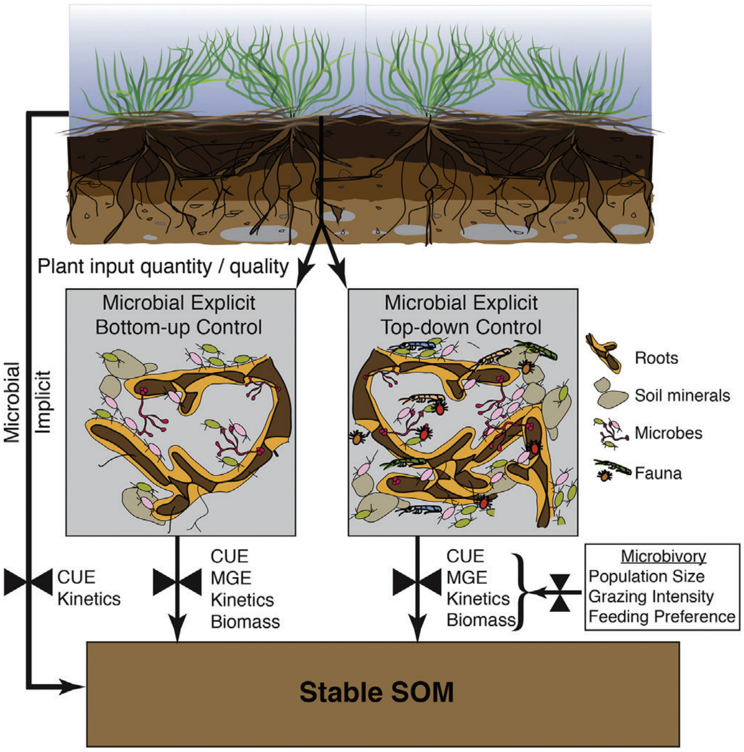
Research Needs and Knowledge Gaps

Develop New Measurement Networks for Collecting Data on Multiple Spatial and Temporal Scales. New data collection and models are needed to address emerging soil paradigms, emphasizing linkages between soil minerals and microbes responsible for SOM stabilization, trace gas production, nutrient availability, and contaminant processing. Integrating these data and formalizing fresh insights into new model structures are critical, and they present simultaneous challenges in identifying the key data and knowledge gaps that can inform and parameterize models yet to be built. Such challenges can be met with new measurement networks that collect various data on multiple spatial and temporal scales. Advancing this research area also requires the development of tools that measure process rates in ways that can be paired with the massive volume of molecular data coming online. Special attention is needed for integrating these measurements to inform, parameterize, and validate models of terrestrial biogeochemistry at ecosystem to Earth system scales.

Determine How Fine-Scale Analyses Map onto Larger-Scale Processes. Advances in high-throughput nucleic acid sequencing have led to new understanding of the extent of biodiversity, its functional potential, and clues about its actual functioning in the environment. Meanwhile, omics

Soil Organic Matter

Soil organic matter (SOM) accumulates in terrestrial ecosystems through the combined activities of photosynthesis by plants, as well as biosynthesis and decomposition by complex mixtures of soil microorganisms (i.e., bacteria, archaea, and fungi). As SOM becomes stabilized through several mechanisms, some of the matter remains in the soil for up to thousands of years, whereas other SOM cycles back to atmospheric carbon dioxide in a matter of days. SOM's role in storing carbon and thereby slowing climate change promotes fertility and crop growth in agriculture, retaining nutrients and contaminants and thus slowing their release to aquatic ecosystems. These processes contributing to SOM stabilization are key to understanding microbial to Earth system pathways. Contributing factors include microbial physiology (e.g., carbon use efficiency, growth rates, and growth efficiency), selective preservation, and mineral stabilization (see figure, Microbial Controls on Soil Organic Matter Stabilization, at right).



Microbial Controls on Soil Organic Matter Stabilization. Three conceptual representations depict different ways models could represent microbes and fauna. **Key:** CUE, carbon use efficiency; MGE, microbial growth efficiency; SOM, soil organic matter. [Reprinted from Grandy, A. S., et al. 2016. "Beyond Microbes: Are Fauna the Next Frontier in Soil Biogeochemical Models?" *Soil Biology and Biochemistry* **102**, 40–44, with permission from Elsevier. © 2016]

Plants also play a major role by providing root-derived (i.e., labile) inputs to soil, which can promote stabilization and destabilization (e.g., the priming effect) in ways not fully understood.

techniques are cheaper than ever, leading to a deluge of data from microbial systems. Yet, there still are no clear connections between these rich data and ESMs for several reasons: (1) relationships between complex microbial community structure and its ecological function have been difficult to establish, even under laboratory conditions, because of inadequate genome annotation; (2) techniques for quantifying gross process rates *in situ* have not kept pace with techniques for describing microbial communities or their putative functions; and (3) representation of soil pore- and plot-scale dynamics at larger spatial and temporal scales remains elusive. Addressing these three knowledge gaps will require consideration of how fine-scale analyses map onto larger-scale processes in ways that better inform predictive models.

Historically, ecosystem theories and biogeochemical models were based on the assumption that environmental conditions change the rate of ecosystem processes, but that biotic responses are consistent across sites (Bradford and Fierer 2012). Emerging evidence, however, suggests that the activities of microbial communities in soils, aquifers, and sediments are shaped by particular local environmental conditions (Fierer et al. 2012; Evans and Wallenstein 2014; Talbot et al. 2015). Accordingly, physiological tradeoffs between acquiring resources versus withstanding environmental stress potentially dictate the activities of belowground communities and their functional response to environmental change.

Better understanding is needed of the extent to which local microbial processes and mechanisms affect projections at Earth system scales. The framework for quantifying these influences is not yet developed. Across scales, multiple processes and drivers regulate ecosystem processes (Bernhardt et al. 2017), but additional insight is needed to understand how these factors integrate to determine emergent Earth system behavior. In particular, theoretical frameworks need investigation and refinement to illuminate broad-scale properties and processes that cause divergent trajectories across environmental gradients and in response to environmental perturbations.

Understand How Factors Controlling Biogeochemical Processes Aggregate Across Scales. Questions related to this fundamental challenge include: Can the mean abundance of a functional gene (or transcript) lead to increased predictive skill of models that operate at scales of square kilometers or larger? Similarly, how may this information transfer to deeper soil horizons or change through time? If new scientific insights are intended to improve predictive models, they also must inform the broad-scale processes and properties that delineate divergent trajectories or threshold behaviors that may not be predictable from current approaches. Predictive modeling necessarily projects ecosystem behavior beyond the domain over which the model has been informed. As in the physical climate system (Knutti and Sedláček 2013), greater fidelity in process representations should increase confidence in modeling terrestrial and aquatic biogeochemical systems (Bradford et al. 2016). Toward this end, knowledge must be developed across environmental gradients, edaphic conditions, and land use practices. Given the challenges and costs of exhaustively sampling diverse subsurface environments, modeling also serves as an important tool to generate hypotheses, refine theory, and highlight key uncertainties in belowground process understanding.

Recent work suggests that trait-environment relationships can be context dependent. For example, the functional form of microbial responses and sensitivities to soil moisture may be constrained by historical environmental conditions and community composition (Hawkes and Keitt 2015; Martiny et al. 2017), suggesting that strategic sampling across biotic and abiotic gradients may be required to define the key ecological boundaries and environmental thresholds. A combination of targeted experiments coupled with multiscale model development can fill key data and knowledge gaps related to prediction of microbial community processes and ecosystem function.

BERAC recommends several new research funding opportunities to fill knowledge gaps needed to link microbial physiology with mineral stabilization and destabilization of SOM and the production and consumption of greenhouse gases. Potential projects that could lead to advancements toward Grand Challenge 4.4 are:

- Identify patterns in microbial functional traits among ecosystems, across edaphic gradients, and in response to perturbations. The physiological traits of microbial communities hold some promise for ways to connect microbial ecology to ESMs (see Grand Challenge 4.3, p. 50). Boiling down data-rich omics information through a functional trait lens will facilitate the identification of key community structure–ecosystem function relationships—namely, informing microbial kinetics, growth rates, growth efficiency, and biochemistry. Similarly, microbial traits may influence microbe-mineral interactions related to the physicochemical stabilization and destabilization of SOM (e.g., through cell chemistry, necromass production, and EPS). These opportunities should enable integration of information at the microbial scale in dimensions tractable for modeling.

- Determine when microbial processes matter and change aggregate behavior. To meaningfully fill knowledge gaps, measurements of biogeochemical process rates (e.g., trace gas fluxes, nutrient cycling, redox conditions, and contaminant fluxes) must be correlated with omics datasets to identify appropriate trait metrics. These insights are critical for understanding how shifts in microbial community composition may drive changes in biogeochemical cycling and, ultimately, increase predictive capacity. Linking process rates with microbial community information and models is critical for larger-scale projections through the assimilation of increasingly autonomous (and diverse) data into models.
- Develop tools to scale microbe-mineral insights from the molecular to ecosystem and global scales. The fine spatial and temporal heterogeneity of the subsurface environment presents enormous challenges in representing microbe- and mineral-scale insights into ecosystem and Earth system models. Models typically average and aggregate over nonlinear processes, but this approach may not be effective at representing the emerging understanding in larger-scale (ecosystem to Earth system) models. Accordingly, new opportunities should focus on identifying the best scale to explicitly model processes and extract emergent properties that can be parameterized at larger-scale models (i.e., as relates to Grand Challenge 4.2, p. 48).

Additionally, BERAC recommends support for opportunities and workshops that increase training, collaborations, and communication across disciplinary boundaries:

- Organize small, recurrent (annual) workshops to bring together microbial ecologists and multiscale modelers. These workshops would implement an integrated modeling-experiment framework called MODEX, with postdoctoral support (housed at a national laboratory) to coordinate group activities.
- Hold workshops at DOE's Joint Genome Institute and Environmental Molecular Sciences Laboratory. Participants would design experiments exploring microbial community structure-function relationships, specifically focusing on rhizosphere priming and plant-microbe interactions, interactions between microbial biomass and mineralogy, nitrification and denitrification, and methanogenesis and methanotrophy.
- Issue new proposal calls to incorporate microbial processes into ESMs. One new program would support the development and deployment of observational networks to challenge models. Large-scale and long-term experiments have been a hallmark of BER research. A second program should focus on building new knowledge and models that bridge BER's Biological Systems Science Division with its Climate and Environmental Sciences Division, capitalizing on synergistic crossover and new frontiers in scaling from molecules to the Earth system.

5 Grand Challenges in Energy Sustainability

Overarching Grand Challenge and 20-Year Vision

Provide a scientific foundation for implementing resilient energy strategies that promote prosperity, security, and human and environmental health by quantifying fundamental interactions between energy and natural system processes over a range of relevant geographic and temporal scales.

Envisioning 20-Year, Resilient Energy Strategies

The need for sustainable energy systems has never been greater. Providing the energy necessary for economic growth and prosperity in the face of a changing planet requires approaches to energy production and delivery that afford both economic and environmental security. This sustainability challenge is formidable, particularly because the energy system is and will remain completely entangled with the use of air, water, and land resources. On the leading edge of Earth system research, the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) is uniquely positioned to advance and integrate the basic science needed to revolutionize a transition to resilient energy strategies (i.e., strategies developed from knowledge of the multiple feedbacks that occur between energy and natural systems as well as system responses to various stressors).

For example, water is essential for growing bioenergy feedstocks, producing hydropower, and cooling thermoelectric and nuclear power generation facilities, but its availability is in part a function of hydroclimate variability and how watersheds respond to these variabilities and are managed. Likewise, the growth potential for wind and solar energy sources is dependent on how the climate system operates on regional scales, land availability, and the technological evolution of the grid itself. Underlying all these factors is the need for sophisticated understanding of the critical infrastructure used for water, energy, transportation, and communication, as well as infrastructure sensitivity to a variety of environmental stresses.

Developing strategies to improve energy system sustainability, therefore, depends on a much more complete understanding of how stressors to land, air, and water resources influence energy strategies and the roles that humans play in these interactions. A systems approach is required to discover fundamental interdependencies, including trigger points and thresholds, and to identify the inevitable tradeoffs that will need to be managed.

Approaches to energy sustainability also must be scale aware, considering interdependencies occurring within the energy strategy's geographic footprint, which could range from local managed ecosystems to regional landscapes to continents (see Fig. 5.1. Multiscalar Energy-Land-Water Interactions, p. 58). Moreover, the details of how national goals intersect with regional and local realities are complicated. Wind power, for example, is growing rapidly within the national grid, but the location of

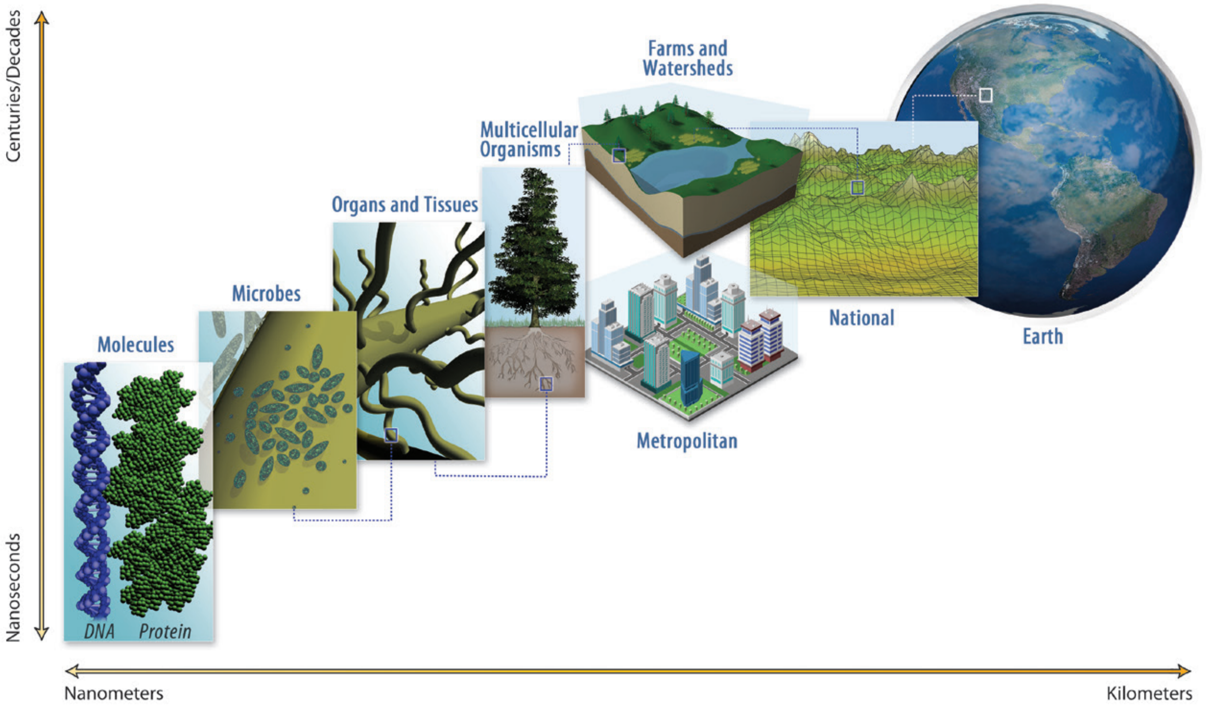


Fig. 5.1. Multiscalar Energy-Land-Water Interactions. Energy sustainability is multiscalar, involving energy-land-water interactions at scales ranging from cellular to global. Interactions within and among scales largely will define the resiliency of different energy sustainability strategies; thus, questions must be scale aware and answers appropriately scalable.

adequate resources is a function of regional climate. Bioenergy's potential expansion depends not only on climatic conditions that support bioenergy crops, but also on the simultaneous availability of healthy soil and water resources at the farm scale, economics of energy demand and supply, and availability of appropriate infrastructure to support this very different energy source. The feedbacks resulting from changes in the current energy portfolio to a much wider dependence on renewables and bioenergy also have the potential to interact with local and regional weather and climate patterns through both biogeochemical and physical processes.

Data-driven approaches, new observations, and modeling innovations will be needed to seek energy pathways that provide optimal economic and environmental outcomes and promote human well-being.

Building on Progress Related to the 2010 Grand Challenges

The 20-year vision of the Biological and Environmental Research Advisory Committee (BERAC) is to enable strategies that guide the implementation of a range of resilient energy systems promoting prosperity, security, and human and environmental health. Such a vision requires important investments in the fundamental sciences central to BER's mission.

BER is uniquely positioned to foster this vision. Assets that support fundamental research regarding basic biology, watershed, climate, and integrated assessment provide a foundation for (1) understanding how very different systems couple to one other and (2) identifying which feedbacks are quantitatively important. BER has a history of addressing complex systems coupling in Earth system sciences [e.g., Earth system models (ESMs)], watershed science (e.g., hydrobiogeochemistry), and bioenergy (i.e., from genes to landscapes). This experience is part of what will be needed to implement an encompassing vision for energy sustainability. BER additionally brings scientific rigor to functional studies of ecological systems, biogeochemical cycling, and the water cycle, as well as the physical climate system—from model development to uncertainty quantification, to new strategies for computational sciences and data analysis.

Considerable progress has been made over the past 6 years to meet the energy sustainability challenges identified in the 2010 Grand Challenges report (BERAC 2010), with many noteworthy examples of success that fall into the report's three broad energy sustainability challenge areas. Descriptions of several successes follow.

Advanced a Fundamental Understanding of the Impacts and Tradeoffs of Alternative Bioenergy Feedstocks and Land Uses on Energy, Climate, and Ecosystem Functioning. BER has positioned itself to meet most of the challenges identified in the 2010 Grand Challenges report through its Bioenergy Research Centers and Genomic Science sustainability programs. Although most of BER's early focus was on energy conversion (i.e., biomass to biofuels), currently receiving more attention are environmental considerations (i.e., climate, biodiversity, and ecosystem functioning including healthy soils), and progress is steady. Breakthroughs are needed to make bioenergy and related bioproducts price competitive and to fully reap environmental benefits through an integrated understanding of the entire field to fuel enterprise.

Improved Understanding of Global Change Effects on Energy and Environmental Systems. BER also is making substantial progress in helping to identify and quantify the impact of energy-food-environment tradeoffs via an understanding of the interactions among individual global change impact sectors, especially those related to energy and water. A recent BER water-energy testbed solicitation is a positive step in the direction of energy sustainability, underpinned by fundamental BER system science and models. Some of the individual and interactive impacts could be large enough to significantly feed back to the evolution of the Earth system as a whole in ways not yet included in ESMs.

Demonstrated Targeted Progress Toward Characterizing the Spatial and Temporal Variabilities of Specific Watersheds and Ecosystems. The Arctic is one of these important ecosystems, where there is now better understanding of ecosystem sensitivity to warming and permafrost thaw, particularly, the responses of microbes, vegetation, and greenhouse gases (GHGs) to climate change. New insights also are emerging about the responses of critical western U.S. river watersheds, such as the Colorado and Columbia rivers, to a range of stressors as well as their associated impact on downgradient water and nutrient delivery with relevance to energy strategies. Very early insights are starting to emerge about other key systems that are especially sensitive to potential state changes, such as tropical forests and human-dominated systems. Next-generation observation platforms—some focused on coincident and autonomous monitoring of above- and belowground interactions (e.g., Dafflon et al. 2017) and integration of multiscale observations through remote-sensing datasets (e.g., Wainwright et al. 2016)—are being tested at key sites (see also Chapter 7 on User Facilities and Research Infrastructure, p. 85). In most cases, researchers are using the observations and process insights to challenge developing multiscale models.

Developed Unifying Models and Frameworks Capable of Testing and Evaluating the Significance of Potential Global Change Drivers, Including Climate, Energy, Land Use, and Water. Programs in high-performance modeling and scaling are unmatched and have contributed substantially to the successful development of integrated assessment models (IAMs), ESMs [e.g., Energy Exascale Earth System Model (E3SM)], and integrated ESMs (iESMs). Other significant progress has been made in developing and applying lifecycle analysis (LCA) and full-cost accounting for ecosystem

Energy Sustainability
Summary of Grand Challenges and Action Items

Energy production and use are inherently connected to land, air, and water resources. Comprehensively understanding these interactions is therefore important for guiding current and future energy production that will optimize energy availability and environmental quality. In recent years, research on bioenergy conversion and associated environmental considerations has progressed substantially, accompanied by an increased understanding of energy-food-environment tradeoffs and improved characterization of spatial and temporal variabilities of targeted ecosystems. Other significant advances include further development of integrated assessment models, climate models, integrated Earth system models, and the coupling of these models where appropriate to fully address sustainability science questions. Moving forward, four grand challenges will take this research into the next decade and help resolve important questions. These challenges are outlined here and then described in more detail in the Grand Challenge Research Recommendations section, p. 61.

5.1. Further develop the science of coupling energy, water, and land use across different spatial and temporal scales to understand environmental impacts and changing climate and to better predict net biogeochemical fluxes.

5.2. Use observational, experimental, and model-based approaches to explore the sustainability of alternative energy systems, incorporating stability and resilience analysis, uncertainty, transition paths from current infrastructures, and the use of appropriate common metrics.

5.3. Understand how variability and change in natural systems affect energy system structure and function and determine how best to build this knowledge into models.

5.4. Create new data streams and more effectively use existing observations to ensure the availability of scale-appropriate data, particularly related to high-resolution land use, landscape infrastructure, demographic change, and energy-land-water research.

Associated with these grand challenges are two action items (see Discussion of Action Items section, p. 67, for more details) that represent recommended first steps toward addressing the challenges:

- Establish a strategically distributed network of energy sustainability testbeds for addressing crucial research questions associated with specific energy strategies and air-water-land feedbacks at multiple scales. The combined testbeds will address several grand challenges while leading to a predictive understanding of couplings between energy systems and natural systems.
- Create an energy sustainability modeling and synthesis center for multidisciplinary teams to address key energy-water-land research challenges. The center will organize multidisciplinary teams to understand impacts (e.g., energy-water-land influences on energy infrastructure), sustainability working groups to resolve short-term research questions, and also facilitate and house integrated data products at resolutions needed for sustainability analyses.

services for bioenergy technologies, especially in the biogeochemistry and water arenas. Progress has been slower, however, for incorporating biodiversity responses and social and policy elements into sustainability research (see also Chapter 3 on Earth and Environmental Systems Sciences, p. 21).

Notably, a concerted push has been made toward more integration of IAMs with ecosystem models, especially through the energy-land-water nexus, where tight linkages can constrain energy efforts. This integration has led to better representation of climate change impacts within IAMs and has highlighted the value of portfolio approaches that effectively link IAMs, ecosystem models, and LCAs. Such linkages likely are to be crucially important for evaluating the complex multiscale effects of potential geoengineering solutions for climate change mitigation such as biomass carbon capture and storage.

Though progress since the 2010 Grand Challenges report has been steady and, in places, noteworthy, many challenges remain. Researchers have gleaned an initial understanding of how to produce enough energy to support people at a higher standard of living, but continuing breakthroughs will be needed to better understand and improve the sustainability of different energy-land-water configurations. Fulfilling these challenges will be essential to achieving BERAC's vision for energy sustainability. Four grand challenges central to achieving this vision are outlined (see Energy Sustainability Summary of Grand Challenges and Action Items, p. 60) and discussed further in the following sections.

Grand Challenge Research Recommendations

Grand Challenge 5.1

Further develop the science of coupling energy, water, and land use across different spatial and temporal scales to understand environmental impacts and changing climate and to better predict net biogeochemical fluxes.

The past and projected future responses of natural and human systems in three key sectors—energy, water, and land—to economic development and changing weather patterns are being studied in more depth than was possible with the data and scientific understanding that existed in 2010, but significant challenges remain. These challenges involve developing a better understanding of critical relationships and data requirements within these key sectors, including a better understanding of important phenomena and drivers at nanometer to global scales. Not all nanoscale processes will be important at higher levels of aggregation for all purposes, but some research is necessary to understand where processes at one scale will have important effects on processes and outcomes at higher aggregation levels. Even as progress is being made in these areas for individual sectors and regions, important interactions and feedbacks among these three sectors and with the Earth system are recognized increasingly as areas experiencing the worst impacts from energy, water, and land development now and likely into the future.

The impacts of changes in climate such as higher temperatures and varying precipitation patterns, as well as those in human systems such as water infrastructure and markets, will have important and interacting consequences for energy, water, and land use systems and markets. Better understanding of these energy-water-land feedbacks and interactions is crucial to assess their integrated impacts on people, infrastructure, and natural systems and to explore tradeoffs among alternative mitigation and adaptation strategies at appropriate spatial and temporal scales.

After farming, electric power generation is the largest user of water. In addition to requiring water for hydropower electricity generation and cooling, thermoelectric plants release heated water into the environment. As in-stream temperatures rise with climate change, the release of heated water can be problematic for aquatic ecologies, as individual power plants can have a significant impact on local in-stream temperatures. In addition, higher temperatures reduce the efficiency of thermoelectric power generation by increasing turbine outlet temperatures. The nation's largest consumptive use of water is irrigated agriculture, but the essential delivery of this water to crops depends on electricity, thereby providing another set of links in the energy-water-land chain. In addition, some electric utilities have been turning to biomass co-firing as a near-term option for generating cleaner power. The transportation sector also has increased its demand for biofuels based on bioenergy crops, placing further demands on water systems and land. There are many other dimensions to the energy-water-land nexus, but these particular components are tightly interconnected and likely to be especially vulnerable to changing climate, while also offering significant opportunities for adaptation. Since the nature of these interactions varies greatly by location, infrastructure, and biophysical constraints, gridded modeling of these nexus components is potentially quite valuable.

Weather variability and climate change are critical drivers of future developments in the energy-water-land system. These impacts need to be studied with future precipitation and temperature projections at the appropriate geographical scale to capture relevant physical linkages, especially for extreme events, which likely will put pressure on existing infrastructure. The link between precipitation and the hydrological system is obvious, but evolving relationships among water use for urban development, power generation, cooling, and agriculture are more complex, particularly when considering the interactions among runoff, reservoir storage, groundwater accumulation and depletion, and in-stream flows. Electricity production is very sensitive to extreme heat, which boosts the demand for air conditioning. Unfortunately, such demand surges also are likely to be closely related to surges in irrigation demand. Recent statistical studies have found that temperature extremes are crucial predictors of yields for corn, soybeans, cotton, rice, and wheat; high temperatures increase the plant's water demand and, at the same time, diminish water availability through evapotranspiration. A fine-grid analysis is necessary to detect these thresholds insofar as a partial or temporal averaging can make detecting them impossible.

Finally, this discussion demonstrates that energy, land, and water models are highly interdependent, with the inputs to each often being some of the outputs from others. Furthermore, these models are linked through both physical relationships and economic market relationships, the latter through trade and transfer agreements.

Energy-land-water research is well-aligned with BER objectives for improvements in three important areas: (1) model representation of energy system transition impacts on society and changes in other sectors, as well as impacts of the whole Earth system on energy system transitions; (2) the research community's ability to include interactions and feedbacks among the energy, water, and land systems and feedbacks in these assessments through integrated impact analyses or IAMs; and (3) the ability of ESMs or IAMs to project the feedbacks from sector changes to the whole Earth system and then back on the energy system (see also Chapter 3 on Earth and Environmental Systems Sciences, p. 21).

Grand Challenge 5.2

Use observational, experimental, and model-based approaches to explore the sustainability of alternative energy systems, incorporating stability and resilience analysis, uncertainty, transition paths from current infrastructures, and the use of appropriate common metrics.

The multiscale, interconnected, and nonlinear nature of energy, water, and land systems, coupled with their vulnerability to multiple stressors, represents a significant challenge for examining the risks and tradeoffs among alternative energy strategies. Transitioning to a future energy grid that is in perpetual, sustainable balance with the natural system's evolution demands several new advances. Identifying pathways to energy system resiliency requires new experimental and observational approaches specifically designed to quantify key interactions influencing coupled energy–natural system behavior. Also required is a new generation of computationally efficient, data-modeling-analysis frameworks that can support scenario analytics, uncertainty quantifications, and well-defined metrics to evaluate a range of outcomes. Development of these new approaches will benefit from integration of a range of developing technologies, including exascale computing, deep-data analytics, networked drones, and *in situ* microbial sensing.

An important challenge for Earth system modeling is to understand the ecological, biological, and carbon cycle interactions and feedbacks in the climate system to identify potential tipping points and mitigation strategies. Energy systems—whether thermoelectrically, hydrologically, wind, solar, or biologically based—provide acute points of interaction that may magnify or attenuate key interactions and feedbacks that affect land use and vegetation dynamics. A follow-on modeling challenge is to stress-test ESMs with energy-related drivers, fully exploring potentials for particular energy systems to enhance rather than diminish the services integral to ecosystem and landscape functioning at different scales.

The new approaches will vary greatly depending on the energy strategy's particular type and scale, as well as assumptions about stressors that prohibit energy sustainability (see Fig. 5.1, p. 58). For example, enabling a sustainable billion-ton bioenergy feedstock supply requires the production of "more with less," or increasing the quality and volume of crop yields on marginal lands challenged by poor soil health, water availability, and nutrient supply. At the scale of individual plants, robust approaches are urgently needed to measure and mechanistically predict plant growth, form, function, and interactions with the surrounding biotic and abiotic environment. New soil and plant sensors and proxies could be developed to identify (1) key controls on tolerance to environmental stresses such as drought, heat, and low soil fertility; (2) how plant resources are partitioned to diverse processes and organs, including respiration, roots, and aboveground biomass; (3) how root and root-microbe processes control the acquisition of water and nutrients; and (4) how the plant microbiome, including rhizosphere communities, mycorrhizal symbioses, and endophytes, contributes to plant performance. Such quantification could, for example, lead to new insights about plant enzymatic activities under water stress or the soil microbial community's ability to access atmospheric nitrogen and soil phosphorous, elements which are abundant but not necessarily bioavailable and could, in turn, contribute to bioenergy sustainability.

At the basin scale, hydropower, power plant cooling, and agriculture rely on watersheds to sustainably deliver water and nutrients downgradient. Watersheds experience a range of stressors having unknown ramifications on energy systems. Examples of stressors include floods, droughts, fires, early snowmelt, beetle infestation, and land use change. Developing a predictive understanding of watershed behavior and response to increasingly frequent stressors is challenging mainly for three reasons. First, there is a wide range of hydrological-biogeochemical interactions in a watershed that occur among plants, microorganisms, organic matter, minerals, dissolved constituents, and migrating fluids that influence water, nutrient, carbon, and other key elemental cycles. These interactions often vary significantly, both laterally across a watershed and vertically between its many compartments (e.g., bedrock, groundwater, capillary fringe, vadose zone, soils, land surface, and vegetation).

Second, accurately predicting the cumulative watershed response to a perturbation is particularly challenging because such a prediction must consider the spatial variability of nonlinear responses and their aggregation. Traditional watershed analysis approaches have been developed based on relatively simple watersheds, often with an assumption that historical hydrological trends (e.g., precipitation) are representative of future trends. Recent studies highlight the need for innovative research paradigms that analyze and predict complex and dynamic watershed behavior under future conditions, their impact on energy systems, and other societal benefits. New paradigms will require new networked sensing systems that can jointly (i.e., in many cases, remotely and autonomously) monitor above- and belowground responses to perturbations across scales. Also needed for this challenge are physics-based, scale-aware surface-subsurface watershed models that can accurately predict aggregated hydrological and biogeochemical watershed exports under a range of stressors, as needed, to assess the sustainability of downgradient hydropower and agricultural systems.

Third, the equally complex and understudied domain at the energy-land-water nexus is the energy-intensive urban landscape. BER research historically has focused on the carbon, water, and energy cycles of natural ecosystems and how these processes influence the climate system.

More recently, BER has started to explore water and water-energy systems, but few efforts are focused on quantifying intensively coupled natural-human-energy system interactions, such as those occurring within urban and urbanizing systems. Urban areas, which often rely on a mixture of energy strategies, consume a disproportionately larger percentage of the planet's energy and freshwater, representing over 70% of fossil fuel carbon dioxide (CO_2) emissions, which is nearly triple the net carbon uptake from terrestrial or oceanic carbon sinks (Churkina 2016). Yet, lacking are (1) the fundamental knowledge of hydrological-biogeochemical interactions that occur in urban landscapes and, therefore, (2) sufficient understanding of the drivers, controls, and feedbacks between human and Earth systems to inform the design of more sustainable energy outcomes. Bringing BER resources to bear on these questions will enable forecasting and mitigation against unsustainable outcomes.

Examining the tradeoffs associated with various energy strategies or their mixes—in the context of land-water systems, responding to a range of stressors, and with consideration of the metrics discussed previously—will require new cyberinfrastructure and software tools. These tools must be able to perform scenario analytics and quantify uncertainty while handling diverse models and data associated with a range of strategies, observational datasets, and dynamic models. Scale-adaptive simulation capabilities, which use adaptive mesh refinement methods to "telescope" into subsystems that require finer resolution or different physics, may provide a flexible approach. Both scale-adaptive as well as reduced-order models are expected to be critical for scenario exploration and decision analysis, uncertainty quantification, and visualization of outcomes and tradeoffs (see also Chapter 6 on Data Analytics and Computing, p. 71).

BER is uniquely positioned for assessing, understanding, and predicting interactions among a range of energy and natural system processes needed to underpin sustainable bioenergy, wind, hydropower, solar, thermoelectric, and other strategies. First are BER's existing programs and investments to understand subsurface microbes, plant, and plant-microbe interactions. Key to rapid advances are *in situ* sensing, high-throughput genome analyses, and next-generation visualization technologies (see also Chapter 2 on Biological Systems Science, p. 3). Second are long-standing programs in hydrology and biogeochemistry that have rapidly advanced the understanding of watershed processes across a variety of scales and Earth system compartments, from the subsurface to the troposphere. This expertise could be extended to include interactions between natural and human-dominated landscapes (see also Chapter 3 on Earth and Environmental Systems Sciences, p. 21). Third is BER's extensive data and simulation capacity that enables effective integration of diverse and multiscale processes into watershed, regional, and global contexts.

Grand Challenge 5.3

Understand how variability and change in natural systems affect energy system structure and function and determine how best to build this knowledge into models.

Numerous studies have looked at ways in which different energy systems and technologies ultimately affect the physical environment through their implications for GHG emissions. Considerably less literature is available on the real and potential vulnerabilities of energy systems to change and variability in physical systems, even though concrete demonstrations are beginning to show that these influences are real and, in some cases, substantial. Important to pursue and include in models as critical feedbacks are both empirical assessments of current risk and modeling studies to investigate the potential for future risks to energy systems and infrastructure.

The rationale for incorporating the effects of physical system change and variability on energy system structure and function in the BER portfolio is straightforward. The effects of these first two vulnerabilities on the delivery of energy services are a function of many factors, among them the development and implementation of both current and new technologies. However, when either individual weather events or long-term systemic change in the physical climate system have the potential to interrupt the delivery of energy systems, they reveal a new set of feedbacks that operate from the atmosphere to the built environment, influencing the relationships among land, water, and energy systems.

The most obvious examples for understanding how physical systems influence energy system structure and function are from the repercussions of storm events on energy infrastructure. Well known, for example, is that Hurricane Katrina disrupted the supply of natural gas coming into the port of New Orleans for several months because of the storm's impact on shallow-water oil and gas rigs in the Gulf of Mexico. Not only the storm's intensity, but also the locations of the rigs themselves, established the vulnerability of that energy system component to major events such as hurricanes. Similar phenomena resulted from Superstorm Sandy's impacts on electricity generation and transmission in the New Jersey and New York area. In this case, however, the actual distribution of sensitive infrastructure had scarcely changed since being established roughly a century ago, demonstrating that the spatial distribution of critical infrastructure can "lock in" climate vulnerabilities for a very long time. This effect, not previously sufficiently well understood, has important implications for societal resilience, demonstrating that rebuilding systems with their same degree of vulnerability would not be wise.

More recent examples center around longer-lasting events, especially heat waves and drought that affect the availability of water for cooling thermoelectric and nuclear power plants. The severity of this effect during recent heat wave and drought events in Texas and the southeastern United States required power plants to be derated (i.e., their available power generation capacity was decreased), affecting their ability to provide regional electricity. A third type of vulnerability is implicit in the short-term response of energy demand to climate and weather events. A germane example is the well-known response of electricity demand to periods of high temperature and heat waves. A fourth type of vulnerability to physical systems can develop over a longer-term period, as energy technologies themselves begin to change. In particular, if bioenergy crops penetrate regional energy markets to a great extent, with or without carbon sequestration, their supply of material obviously is dependent on climate conditions and water availability that are favorable to the plants themselves. Thus, longerterm patterns of variability and change in the physical climate system, similar to those already observed for natural systems and phenological change, become important considerations in any assessment of future energy infrastructure and its projected functioning.

Current energy models or IAMs that seek to understand how energy systems change over time do not fully represent energy system feedbacks with the physical climate system. Thus, these models do not represent the potential for either short-term, event-driven effects of the physical climate system on energy or the longer-term effect of systemic changes and variability on energy infrastructure siting and performance. Because the sensitivities are known to exist, however, they must be understood much better from a quantitative perspective to be incorporated into models.

Relevance for BER of this limited feedback in models is twofold. First, from a climate impact perspective, these effects are poorly understood in terms of their frequency and intensity. Efforts like DOE's E3SM development should be able to represent the distribution of climate-related weather events in such a way that they can contribute to a growing body of impact studies that go beyond ecosystems and water resources. Second, from a more comprehensive Earth system perspective, these effects are all examples of feedbacks from the climate system to systems with a human dimension, and thus they have direct consequences for the evolution of those human systems over time, as well as implications for how these systems might be designed to minimize their vulnerability.

Grand Challenge 5.4

Create new data streams and more effectively use existing observations to ensure the availability of scale-appropriate data, particularly related to high-resolution land use, landscape infrastructure, demographic change, and energy-land-water research.

The integration of land use information into ESMs has yielded new insights about contemporary and future changes in the terrestrial carbon budget, forest dynamics, and crop productivity by more realistically constraining model inputs and responses. Key datasets used to develop land use models include time series of cropland area and yield, pasture area, and rates of wood harvesting. Future projections often rely on historical relationships between these variables and regionally aggregated estimates of population, economic wealth, and energy policy. To facilitate the development of next-generation land use models, a new suite of high-resolution datasets describing human and energy systems is a priority. Investment in this area has the potential to be transformative, enabling new insight about the drivers and responses of the global Earth system to changing patterns of energy and land use.

At least five different data classes are needed to create more mechanistic models and to stimulate the development of new conceptual approaches in the field of energy sustainability. These classes include human demography, land use intensification, transportation, economic wealth, and energy systems. For each data class, efforts are under way to develop regional- and global-scale products. However, many of the datasets have limited or no time series information, which is crucial for understanding how human systems (and their impacts) are evolving through time. Nor has there been coordination to create a suite of products that are internally consistent across different modeling domains. The Mauna Loa atmospheric CO_2 time series is perhaps the archetypal example of the critical importance of long-term Earth system monitoring. For other Earth system variables, the National Aeronautics and Space Administration (NASA) and other space agencies have recognized the importance of sustained funding for developing Earth system data records that span different satellite instruments and measurement techniques. Similarly, within BER, long-term funding for the AmeriFlux Network of sites measuring ecosystem CO_2, water, and energy fluxes has enabled the scientific community to understand how terrestrial ecosystems are responding to short-term and secular changes in atmospheric composition and climate, with the value of these time series increasing exponentially with their longevity (see also Chapter 7 on User Facilities and Research Infrastructure, p. 85). In contrast, for human and energy systems, there is less recognition of the importance of long-term data records; as a consequence, data paucity has created a critical bottleneck limiting the growth of the energy sustainability field as a whole.

For human demography, extremely high spatial resolution population information is available from LandScan, which BER already supports. LandScan provides state-of-the-art contemporary global population estimates that are valuable for many applications. However, to connect population to observed changes in land cover, hydrology, and environmental processes such as wildfire dynamics, quantitative time series of changing human populations are needed that possess an internally consistent methodology and for which quantitative uncertainties are available. This type of data is essential for time series analysis and trend detection. More broadly, ancillary data on age structure, wealth (i.e., gross domestic product), and housing density would considerably amplify the value of these data for determining impacts, using the Kaya identity or other conceptual approaches for expressing total CO_2 emissions levels.

Indicators of land use extensification (i.e., clearing new land for human use) may be more easily estimated from remote-sensing observations than indicators of intensification (i.e., higher yields on the same area of land). As a result, arguably some of the greatest achievements in the understanding of land cover and land use change over the last 2 decades have come from the use of high-resolution satellite imagery to quantify patterns of forest cover change, deforestation, and cropland use. However, land use intensification can have equally important impacts on global biogeochemical cycles and disturbance dynamics, and this information is essential for improving many aspects of ESMs. For example, all the prognostic fire models participating in the Fire Model Intercomparison Project (FireMIP) underestimate the declining incidence of fire in grasslands and savannas worldwide over the past 2 decades. A likely contributing factor is that none of the models simulate the density of cattle or other livestock on the landscape. In many areas, cattle density has increased severalfold, reducing fuel amount and continuity that, in turn, limit fire size. Livestock also have profound impacts on the global methane cycle, making these data essential for accurately simulating atmospheric concentrations of this important GHG. Apart from livestock density, other closely related indicators of land use intensification include fertilizer use and water transfers for irrigation. Reliable estimates of nitrogen fertilizer use and irrigation water withdrawals do not even exist for the midwestern United States, yet both are crucial drivers of GHG fluxes, hydrologic flows, and primary productivity in agricultural landscapes, with concomitant effects on energy, water, and carbon interactions.

Transportation databases lack sufficiently resolved time series information for both surface and air travel. Particularly needed are spatially explicit information on road density and the number of cars and trucks for at least the past few decades. For air travel, information about passenger transfers among cities, possibly at a monthly resolution, is critically needed. Such data are important not only for energy use and emissions, but they also can inform models for the spread of invasive species and disease.

Likewise, there are crucial needs for better resolved economic and energy system databases. High-resolution data products extending from at least 1990 forward are needed to sufficiently underpin energy sustainability modeling efforts.

Discussion of Action Items

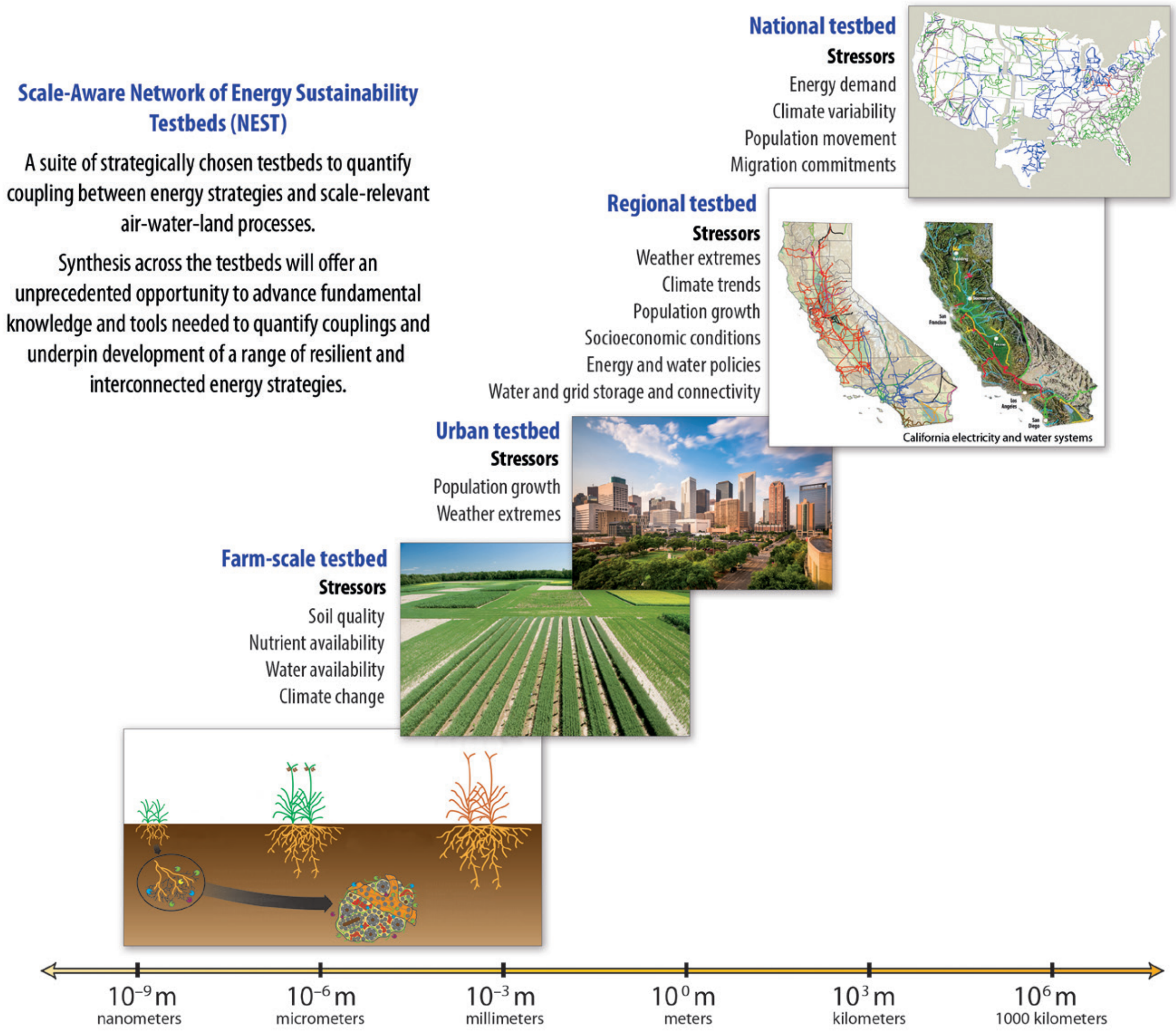
Achieving BERAC's 20-year energy sustainability vision requires advances and integration of many components. Two key components, described in this section, include (1) establishing energy sustainability testbeds for addressing crucial research questions at multiple scales and (2) creating an energy sustainability modeling and synthesis center for multidisciplinary teams to address key energy-water-land research challenges.

Establish a Network of Energy Sustainability Testbeds

BERAC recommends developing a Network of Energy Sustainability Testbeds (NEST) at strategically distributed study sites. A world-unique NEST would comprise a suite of testbeds spanning a range of scales, where each scale is relevant for a particular energy strategy and associated air-water-land forcing or vulnerability. Each testbed in this network would be used to address a specific suite of questions and challenges appropriate for that strategy, but synthesis across the testbeds could offer an unprecedented opportunity to advance fundamental knowledge and tools needed to develop a range of resilient and interconnected energy strategies (see Fig. 5.2. Proposed Network of Energy Sustainability Testbeds, p. 68).

Several characteristics will be important for the success of NEST. Envisioned is a network of about five energy sustainability testbeds, which together would be funded at a level similar to that of a DOE Energy Innovation Hub. The choice of testbed locations should be carefully considered and meet several criteria:

- Distributed testbeds must cover a range of scales matched to the footprint associated with interactions between energy strategies and associated natural system processes. Example



Scale-Aware Network of Energy Sustainability Testbeds (NEST)

A suite of strategically chosen testbeds to quantify coupling between energy strategies and scale-relevant air-water-land processes.

Synthesis across the testbeds will offer an unprecedented opportunity to advance fundamental knowledge and tools needed to quantify couplings and underpin development of a range of resilient and interconnected energy strategies.

Fig. 5.2. Proposed Network of Energy Sustainability Testbeds. These testbeds comprise a suite of strategically distributed study sites chosen to span a range of scales, each relevant to a particular energy strategy and associated air-water-land forcing. Each testbed will be used for experiments, observations, and modeling to address a unique set of questions; synthesis across the testbeds will offer an unprecedented opportunity for advancing the fundamental knowledge and tools needed to develop a range of resilient and interconnected energy strategies. [**Bottom left image:** Soil aggregate illustration to right of black arrow modified from Jastrow, J. D., and R. M. Miller. 1998. "Soil Aggregate Stabilization and Carbon Sequestration: Feedbacks Through Organomineral Associations." In *Soil Processes and the Carbon Cycle*. 207–23. Eds. Lal, R., et al. CRC Press LLC, Boca Raton, Florida. **Farm field** courtesy U.S. Department of Energy's Great Lakes Bioenergy Research Center. **Urban landscape** courtesy iStockphoto. **California map on left** courtesy California Energy Commission. **California map on right** courtesy California Department of Water Resources. **U.S. national power grid map** courtesy Federal Emergency Management Agency]

scales that could be considered include a farm, metropolitan area, river basin, coastal system, multistate region, or even a continent.

- Each testbed should address several of the grand challenges in energy sustainability, leading to a predictive understanding of couplings between energy systems and natural systems (i.e., land, air, and water) relevant to the particular testbed. As illustrated in Fig. 5.2, p. 68, the data, models, energy mix, and variabilities, as well as energy-water-land vulnerabilities, approaches, and science questions, will vary with scale and testbed.
- Network testbeds should be chosen to optimize opportunities for advancing the understanding of how energy-air-water-land systems couple across scales and strategies and how they respond to different stressors. If carefully chosen, synthesis of knowledge, models, and data across the network will offer significant, sum-is-greater-than-the-parts advances toward BERAC's 20-year vision.
- Community acceptance of key testbed characteristics will be a critical component of success. Envisioned are community workshops that involve scientists as well as stakeholders to further refine the network construct and identify priority criteria.

Synthesis across the testbeds will offer an unprecedented opportunity to advance fundamental knowledge needed for developing resilient energy strategies.

Create an Energy Sustainability Modeling and Synthesis Center

A Center for Energy Sustainability Modeling and Synthesis (CESMS) would provide a physical location for teams to tackle complex interdisciplinary problems related to energy sustainability challenges and house a location for the assembly of key databases that would be available both to teams and the larger scientific energy sustainability community.

Energy sustainability research, appropriately integrated, is highly interdisciplinary. Addressing any one of its challenges requires the pursuit of multiple lines of inquiry. Doing this efficiently (i.e., taking advantage of potential synergies and better ensuring integration) will require interdisciplinary teams working in concert to target a specific challenge. Envisioned are two types of teams working at two different temporal scales, funded under the CESMS umbrella:

- **Energy Sustainability Process Teams.** Multidisciplinary teams working across institutional boundaries on long-term problems could be a powerful means to quickly advance energy sustainability science. For example, the formation of a team to understand the impacts of changing climate extremes on energy-water-land interactions and subsequently on the U.S. energy infrastructure might draw on a mix of national laboratory and university scientists, together with industry stakeholders. The team would design observations and experiments, assemble data products, and create the mix of models required to provide an important predictive capacity. The complexity of the problems tackled would require multiyear commitments by all parties.
- ***Ad Hoc* Energy Sustainability Working Groups.** These teams would work on problems of more limited scope that could, with CESMS support, be reasonably expected to resolve a problem over the course of a year or less via multiple working group meetings. *Ad hoc* proposals would be solicited from multidisciplinary groups that would gather at CESMS several times to discuss and work on problems best addressed in a center setting with its cyberinfrastructure and logistical support. Working groups also might be assigned a center-based postdoctoral scientist to facilitate access to databases and other resources to best synthesize existing data and models related to a particular energy sustainability question.

CESMS also will provide a home for assembling integrated data products at appropriate resolutions for energy-land-water resources as elaborated on in the previous discussion on energy sustainability testbeds. Primarily, this includes long-term data for key human and energy system variables not now available in assembled form from any source. BERAC proposes an early workshop to identify community priorities, followed by a funding model [e.g., perhaps similar to NASA's Making Earth Science Data Records for Use in Research Environments (MEaSUREs) Program], which provides the necessary data continuity in an internally consistent way over a period of at least several decades. One path might be to fund one proposal in each of the five variable classes identified previously: human demography, land use intensification, transportation, economic wealth and trade, and energy use.

Overall, CESMS would provide a means for connecting scientists with data in a physical environment designed to facilitate the multidisciplinary approaches crucial for addressing key energy sustainability questions, thereby advancing energy sustainability science quickly and efficiently.

6 Grand Challenges in Data Analytics and Computing

Overarching Grand Challenge and 20-Year Vision

Develop the approaches and computational capabilities to collect, store, and analyze large-scale data across temporal and spatial scales.

Ensuring that Multiscale Scientific Data Support Scientific Rigor, from Collection to Analysis

Research efforts in data analytics and computing continue to focus on processing ever larger amounts of data in a more timely manner. Additional efforts aim to combine and analyze complex data at different spatial and temporal scales to increase understanding and predictability of natural systems across scales. In further developing these capabilities, several objectives are crucial:

- Ensure that scientific data, at whatever scale, are readily accessible to researchers and the research community; well documented and maintained, with identifiable provenance; discoverable; and referenceable in a rapidly evolving commercial technology environment and in the presence of new data collection technologies.
- Ensure that software system design and engineering for data analysis and modeling are focused not only on scalability for today and tomorrow, but also are readily adaptable to emerging computer architectures.
- Develop a cross-disciplinary framework, including the underpinning mathematical, statistical, and computational tools, to facilitate model integration across multiple scales with requisite scientific rigor.
- Build capacity to include “human-in-the-loop” in computational analyses of multimodal streaming data, which will require a continual process of incorporating new ways to combine, visualize, and adjust real-time data streams, along with mechanisms for enabling experimental parameters to be changed on the fly.

In its 2010 Grand Challenges report (BERAC 2010), the Biological and Environmental Research Advisory Committee (BERAC) identified handling and interpreting the explosion of scientific data as a primary challenge. Not surprisingly, six of the seven 2010 recommendations in the chapter, Grand Challenges in Computing for Biological and Environmental Research, were focused on data-related issues, from promoting more standardization in data capture and analysis methods to developing new computing paradigms to harness extreme-scale parallelism for data-intensive analysis applications (see sidebar, Summary of 2010 BERAC Research Recommendations for Computing, p. 72). These challenges have only increased in magnitude since the 2010 report and are continually escalating with the development of new instruments and sensors that stream data in real time.

Also, unsurprisingly, BERAC's 2017 Grand Challenges workshop confirmed and maintained that data are a primary focus of the computational challenges for the U.S. Department of Energy's (DOE) Biological and Environmental Research program (BER). However, since the 2010 report, these challenges have been more clearly defined, moving more toward specific challenges posed by (1) the extreme volume, velocity, and diversity of available scientific data; (2) the need to develop effective means to store, analyze, and interpret extremely large, distributed, and heterogeneous data collections; (3) the rapidly changing computing architectures; and (4) the need for scientific researchers to more immediately visualize and interact with experimental data to steer experiments in the most fruitful directions.

These challenges are not specific to BER; they are well-recognized challenges in the broader scientific community and are active areas of research and community engagement across many scientific disciplines.

Examples of broad research community focus on these issues include:

- A 2017 workshop by the international Human Frontier Science Program (HFSP), which gathered senior data stewards and funding agency representatives to work toward a unified vision and plan for a coordinated international effort to better ensure long-term sustainability and appropriate alignment of funding with scientific impact. The results include a publication that articulates the challenge and also supplies consensus definitions of core principles and goals (www.hfsp.org; Anderson et al. In review).
- In the biomedical research community, a National Institutes of Health (NIH) framework for supporting a "Findable, Accessible, Interoperable, and Reusable" (FAIR) data standard, and establishing a "commons" for tracking and documenting data collected in the course of research (Miller 2016).
- Within the DOE national laboratory system, the Analysis in Motion (AIM) Initiative at Pacific Northwest National Laboratory (PNNL) that was established to address the challenges of streaming data and more fully integrating human-in-the-loop (see Fig. 6.1. AIM Initiative, p. 73). AIM is considered a major laboratory initiative at PNNL.

Summary of 2010 BERAC Research Recommendations for Computing

- Establish a new data management paradigm for data-intensive science with ontologies as a basis for semantic data representations; standards for experimental protocols and data exchange; and an open-access, open-development data management infrastructure.
- Create a new publishing paradigm that credits and rewards researchers for publishing peer-reviewed datasets or analytical methods in addition to conventional peer-reviewed journal articles.
- Develop new computing paradigms capable of meeting the enormous parallel processing and data-intensive analysis needs now emerging for biological, climate, and environmental data.
- Standardize experimental and computational protocols and methods to increase data integration, data usability, and system interoperability to improve research productivity.
- Improve data usability and model accuracy by ensuring that appropriate data quality standards are created and stored with the accompanying data.
- Design and build software solutions that provide researchers with better access to increasingly large, complex, and interrelated datasets.
- Develop virtual laboratories and tools to more fully engage human cognitive faculties and provide richer opportunities for scientific collaborations.

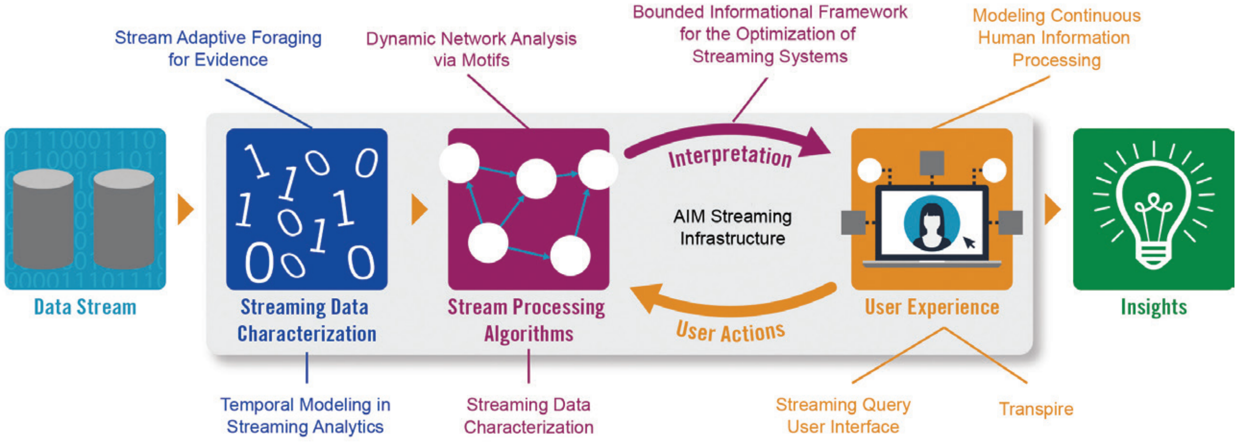


Fig. 6.1. Analysis in Motion (AIM) Initiative. AIM is developing a new analysis paradigm to provide continuous, automated synthesis of new knowledge and enable measurement systems to be steered in response to emerging knowledge, rebalancing the effort between humans and machine (Dasgupta et al. 2017; Crouser et al. 2017). [Image courtesy Mark Greaves, Pacific Northwest National Laboratory]

- Globus cloud research data management services (Chard et al. 2016) that are being applied to data publication and sharing within DOE's Energy Exascale Earth System Model (E3SM), DOE light sources, National Institute of Standards and Technology Materials Data Facility (Blaiszik et al. 2016), NIH Big Data to Knowledge (BD2K) projects, and elsewhere, enabling secure, reliable access to many petabytes of data.

The computer and data science challenges in these areas clearly are not merely stand-alone goals written by and for computer and data scientists; rather, they are driven by the key scientific challenges that researchers face across BER biological, environmental, and climate science studies. Examples include: (1) predictive biology for microbial communities; (2) scaling from microbes to Earth systems; (3) elucidating the coupled Earth system, which includes climate patterns, the oceans, and plants, as well as human impacts that encompass the influence of urban environments and effects of industrial processes; and (4) transforming biology into a quantitative and predictive science. Addressing these scientific challenges requires new types of experiments—with high-resolution imaging, high-throughput data collection, and long-term observations of whole ecosystems (both above and below ground)—while leveraging newly available and ubiquitous observational and sensor systems and the benefits of streaming data from instrumentation. These core BER science areas each are developing new modeling approaches, including integrated models that cross traditional disciplinary boundaries and scales. Addressing these challenges also requires the support of data and computational science advances that are responsive to and keep pace with scientific needs and directions (see sidebar, U.S. Department of Energy User Facility Perspective: Joint Genome Institute, p. 74).

New Sensor Technologies. The 2010 Grand Challenges report (BERAC 2010) anticipated new sensor technologies. The rapid commercialization of these technologies, with tremendous cost decreases, provides great opportunity for broader data collection, but presents the challenge of ensuring that data are

U.S. Department of Energy User Facility Perspective: Joint Genome Institute

The DOE Joint Genome Institute's most pressing data and computing challenges, as indicated in 2017, include:

- Inferring function and gene ontology from large datasets.
- Bringing compute cycles to data rather than data to compute cycles for managing exascale data.
- Leveraging high-performance computing resources in the exascale era.
- Conducting large-scale calculations across large datasets that span both space and time to infer microbial community composition and interactions.
- Ensuring affordable memory through distributed systems to analyze exascale datasets efficiently.
- Combining and building "systems-level" understanding from orthogonal datasets.
- Reducing data to distill biological signals from noise.

collected in a scientifically meaningful form. Raw sensor measurements will serve science well only if they are meaningfully annotated with metadata. Further, such measurements will be most meaningful if they are collected with a predefined scientific methodology that specifically guides which data are collected. Scientists still may make unexpected, serendipitous discoveries by poring through all data collected, but BER investigators should evaluate these new data sources carefully and design controlled experiments to capture the data *and* metadata needed to produce valid scientific results over time.

Disruptive Technologies. Community participants in the 2017 BERAC Grand Challenges workshop expressed their expectation to see continued exponential growth in data volumes over the next 2 decades due to a cycle of continuous technological advances, with an ever-growing gap between large-scale data needs and current capabilities to address those needs. Importantly, workshop participants also identified the emergence of potentially highly disruptive technologies as a major challenge. Some emerging technologies likely will lead to increased growth rates in data volumes and velocity (e.g., whole-genome and very long read sequencing). Others, such as ubiquitous sensors, will result in computational challenges due to their geographic distribution and potentially remote or difficult-to-reach locations where they can be placed (e.g., underground). Advances in computing infrastructures also are likely to lead to disruptive technologies. With the end of Moore's law in sight, a wide variety of novel computing architectures (e.g., exascale, quantum, and neuromorphic computing, as well as supercomputers specialized for deep learning) already are emerging. Effectively assessing and harnessing these new technologies for scientific discovery will create additional challenges for BER science (see Data Analytics and Computing Summary of Grand Challenges and Action Items, p. 75).

Building on Progress Related to the 2010 Grand Challenges

The BER research community has made significant progress in harnessing leading-edge computing technologies [e.g., those provided by DOE's Office of Advanced Scientific Computing Research (ASCR) Leadership Computing Facilities] to advance climate and biological science. In close collaboration with the computer science and applied mathematics research communities, new scalable numerical simulation methods and libraries have been developed and deployed. Climate, environmental, and biological science projects are represented in the recently launched DOE Exascale Computing Project (exascaleproject.org) and are continuing to drive the development of next-generation modeling systems. The general high-performance computing (HPC) skill level among BER researchers is such that while the community faces future hurdles in the

Data Analytics and Computing
Summary of Grand Challenges and Action Items

Data challenges in BER research programs have increased by orders of magnitude over the past few years. New techniques and services are required to leverage the wealth of research results and transform them into world-leading scientific discoveries. Although the data differ in format, common approaches can be employed among BER programs for archiving, accessing, processing, and generating enhanced data products.

Over the past decade, further progress has been made in developing archives, developing software to handle large amounts of program-specific data, and linking BER researchers with high-performance computing capabilities. However, the growth in data size, complexity, and heterogeneity has far outpaced these efforts, and the challenges are continually escalating with the development of new instruments and sensors that stream data in real time. As a result, BERAC has identified five grand challenges outlined here and then described in more detail in the Grand Challenge Research Recommendations section, p. 77:

6.1. Develop robust approaches for large-scale data collection, curation, annotation, and maintenance.

6.2. Develop computing and software infrastructure to enable large-scale data (i.e., Big Data) storage and analysis.

6.3. Conduct research to develop suitable algorithms and programming models that can harness current and future computer architectures to effectively model complex coupled systems and analyze extreme-scale data.

6.4. Engineer advanced computational modeling combined with data integration across temporal and spatial scales.

6.5. Conduct research and develop activities that support human understanding of large-scale, multimodal data streams, including the ability to steer experiments in real time.

Associated with these grand challenges are four action items recommended for initiating progress toward the challenges (see Discussion of Action Items section, p. 80, for more details):

- Prioritize building and maintaining an integrated facility for long-term data storage, archiving, and data analysis.
- Evolve current and develop new ontologies and metadata standards, as well as strengthen compliance in funding announcements and reporting standards.
- Facilitate and encourage increased interactions across disciplines (e.g., computer scientists, statisticians, data analysts, and experimentalists).
- Foster coordination and collaboration within DOE and among other agencies.

required extreme scalability of algorithms, performance, and portability, researchers are well placed to continue making advances if provided with the needed support.

Several recent workshops were dedicated to the data and computing challenges raised in the 2010 Grand Challenges report (BERAC 2010):

- In April 2013, ASCR organized a Data Crosscutting Requirements Review (ASCR 2013a), which highlighted the fact that experimental facilities, especially those across BER and the DOE Offices of Basic Energy Sciences (BES) and High Energy Physics (HEP), are expected to see much more rapid growth in data volumes, velocity, and complexity than previously expected.
- A July 2013 ASCR workshop on Accelerating Scientific Knowledge Discovery, with BER,

BES, and HEP participation, focused on the challenges of enabling effective scientific discovery in an era of extreme-scale data volume, velocity, and diversity. The resulting report (ASCR 2013b) supported research into scalable data storage, sharing, and fusion capabilities.

- In August 2015, BER's Climate and Environmental Sciences Division (CESD) commissioned a workshop of its virtual data working group. The workshop organizers conducted a survey of the broader CESD research community's computing requirements. The top two community needs identified by this survey are (1) an easy way to publish and archive data and (2) the linking and comparison of heterogeneous data types (U.S. DOE 2016).
- In September 2015, an ASCR workshop on the Management, Visualization, and Analysis of Experimental and Observational Data (ASCR 2015), with participation from BER, BES, and HEP, put emerging community challenges into an even starker perspective:
    - All experimental and observational science (EOS) projects represented at this workshop struggle to keep up with the demands and opportunities posed by the increasing data flood.
    - Meeting the challenges of data explosion from EOS projects requires computational platforms, networking, and storage of greater capacity and lower latency, along with software infrastructure suited to project needs. However, existing HPC platforms and their software tools are designed and provisioned for high-concurrency, HPC workloads; single-project data products; and comparatively simpler data needs. The result is a significant gap between EOS project needs and the current state of the art in computational and software capabilities and resources.
- In March 2016, ASCR and BER organized an Exascale Requirements Review. The resulting report (ASCR 2016) again predominantly highlights the communities' data needs for addressing data issues at extreme scales: scalable data processing, data representation and visualization, data transport, and more effective fusion of model simulations with observational data. At this workshop, representatives from BER, BES, HEP, and DOE's Office of Nuclear Physics (NP) again highlighted their critical need for a large-scale data management and analysis capability (both in hardware and software) to meet their core requirements in pursuit of scientific discoveries.
- As mentioned previously, the 2017 international HFSP workshop and its resulting publication (Anderson et al. In review) also highlights the computational challenges faced by the scientific community.

As this broad community engagement demonstrates, BER is not alone in its urgent need for more effective data management and analysis facilities and applications for extreme-scale data. BES, HEP, and NP face similar challenges, and there is great potential for addressing these needs jointly. The *BER Virtual Laboratory: Innovative Framework for Biological and Environmental Grand Challenges* report (BERAC 2013) clearly identified the need for a computational infrastructure to integrate disparate and multiscale measurements, theory, and process understanding into predictive models, ultimately creating new knowledge for developing energy and environmental solutions. This report also highlighted the need to build upon and integrate BER's existing knowledge-discovery infrastructure, such as DOE's Systems Biology Knowledgebase (KBase; kbase.us). In addition, the report recommended close collaboration with ASCR to discover, develop, and deploy computational and networking capabilities for analyzing, modeling, simulating, and predicting complex phenomena important to DOE. This request for a joint data management and analysis capability was renewed and strengthened during the 2016 crosscutting ASCR-BER Exascale Requirements Review (ASCR 2016). Any such solution will certainly build on the experiences and capabilities developed by existing BER infrastructure services such as KBase and the Earth System Grid Federation (ESGF; esgf.llnl.gov).

Grand Challenge Research Recommendations

Grand Challenge 6.1

Develop robust approaches for large-scale data collection, curation, annotation, and maintenance.

As noted previously, BERAC's 2010 recommendation to establish a new data management paradigm for data-intensive science is not unique to BER or DOE, but rather it is experienced across disciplines and federal agencies. This challenge also has not been solved. Closely related to maintaining curated data annotations and augmenting any automated annotation (e.g., genome annotation) are (1) ongoing concern for data preservation (i.e., mirroring data to prevent accidental loss), (2) security (i.e., to prevent intentional or unintentional corruption of data), and (3) data provenance and curation (i.e., whereby updates and corrections to data collections are made in a transparent and traceable way to establish when and why datasets might be altered and corrected). NIH, which has a long history of working toward secure and robust data processes, is moving in this direction by actively working to establish a data commons (see Fig. 6.2. DataMed Ecosystem, p. 78).

Grand Challenge 6.2

Develop computing and software infrastructure to enable large-scale data (i.e., Big Data) storage and analysis.

BER science is highly reliant on access to large-scale computational resources currently provided by ASCR to carry out extreme-scale production runs, as well as midrange computing capabilities for model development and testing. Continued access to these resources remains essential for BER science.

BER science increasingly integrates processes across scales and disciplines, from molecules to the global-system level. One challenge is to integrate data throughout biological lifecycles and across scales, from experimental, observational, and computational sources, spanning domains as diverse as molecular biology, plants, and environmental systems that include land, water, and energy. Making this endeavor difficult is (1) distribution of data across many different heterogeneous archives and databases, (2) creation of data analysis systems that enable scientists and scientific communities to analyze large-scale data and integrate data collections, and (3) provision and ongoing maintenance and development of a suitable set of scalable data analysis tools. Scientists currently spend an inordinate amount of time on discovering, downloading, and integrating data for their analyses. Furthermore, they often cannot use all the available data because they lack the appropriate scalable analysis tools, a challenge that serves as the basis of this recommendation.

At a wide range of recent workshops, BER researchers expressed the need for an integrated data analysis infrastructure [including the 2017 BERAC Grand Challenges workshop and recent ASCR-BER Exascale Requirements Review (ASCR 2016)]. Such an infrastructure should enable easy and unified access to all data, along with the computational capacity and analytical tools to analyze those data.

When designing such a data warehousing and access interface, and the computing and software infrastructure to support it, much can be learned from several existing BER data infrastructure projects. BER's KBase provides a single-analysis environment in which scientists can integrate, analyze, and share data. This system, still in relatively nascent form, provides advanced, scalable analytical tools that users can leverage for their work. However, the current implementation is limited in the variety of

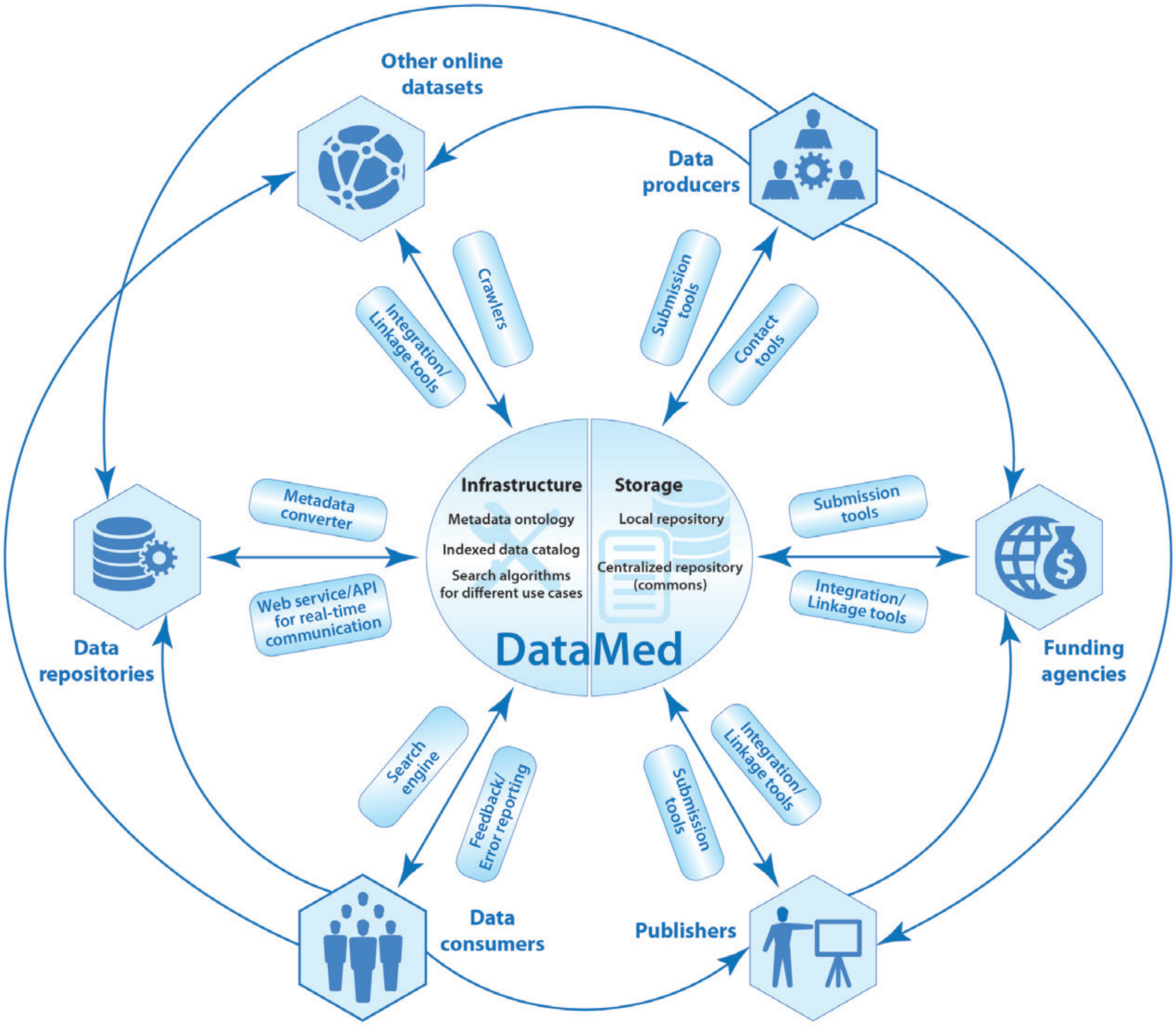


Fig. 6.2. DataMed Ecosystem. Multiple stakeholders interact using components and tools that may or may not yet exist. [Modified with permission from Miller, K. 2016. "Data's Identity Crisis: The Struggle to Name It, Describe It, Find It, and Publish It," *Biomedical Computation Review* (Spring 2016), 22–27.]

available data and overall scale of data that can be analyzed. In a similar vein, ESGF represents a worldwide network of data archives for climate modeling results, with a single search interface that is neutral to the physical location of data resources, and can deliver data from multiple sources to one target destination. However, ESGF does not provide access to data analysis facilities and tools and currently is focused only on climate data. Evaluating solutions from outside BER also would be useful. Examples include the Globus system (globus.org) used in KBase, ESGF, and many other national and international data projects. One of these projects, funded by HEP, ASCR, and the Large Hadron Collider, is the Production and Distributed Analysis (PanDA; news.pandawms.org) project, which manages data, application, and computing access for a worldwide community. Noteworthy is that PanDA helped to analyze 1.5 exabytes of data in 2016.

HPC, and data analysis infrastructure access even more so, is pervasive in today's BER science. Therefore, scientific progress and new discoveries depend on community access not just to "adequate," but to "state-of-the-art" computational and data resources.

Grand Challenge 6.3

Conduct research to develop suitable algorithms and programming models that can harness current and future computer architectures to effectively model complex coupled systems and analyze extreme-scale data.

An important part of the infrastructure needed includes advanced analytics in support of knowledge discovery. Measurements of biological systems using technologies such as high-throughput molecular sequencing (Mardis 2008) and imaging (Kherlopian et al. 2008) routinely generate terabytes of data per experiment. Similarly, high-resolution climate modeling affords spatial resolutions down to the 25-kilometer level while generating hundreds of terabytes of data (Wehner et al. 2014). Collectively, raw data generated by biological and Earth system studies in public databases are quickly approaching exabyte (10^{18}) scales. Integrated analyses of the various data types will afford deeper understanding of system processes and enable knowledge discovery, but they also will require the development and use of advanced algorithms and machine-learning techniques to process the data scale and noise inherent in these measurements.

The large data volumes being generated necessitate the development of indexes and data structures to enable efficient querying of these datasets. The computer science community has developed relevant approaches for indexing large *corpora* of text (Navarro and Mäkinen 2007). Connecting with the computer science community is vital to both learn from the disciplinary advances of its members and engage them in biological research design, so that each study design is well defined in the statistical, numerical, and computational requirements necessary for experimental rigor and success.

Biological and Earth systems datasets are often multidimensional (e.g., corresponding to genes, genomes, metabolites, organisms, and climate model grids). The quantity of the variable definitions alone (i.e., even before data collection) can be daunting, with hundreds or even thousands of variables to measure. However, not all variables may be of equal importance when constructing models of the system being studied. Thus, advanced and automated analytic efforts should include the development of methods for data reduction and variable feature selection (Guyon and Elisseeff 2003) to identify important characteristic dataset features.

Crucial to the accurate study of these systems is the availability of high-quality annotations. In biological datasets, for example, these annotations include those for gene and protein function. The approach to this challenge must couple sensitive and accurate computational methods for homology detection with novel, high-throughput experimental methods for characterizing gene function.

Taken together, advanced analytic approaches remain challenging but are necessary for developing improved predictive and more computationally efficient models at all levels of the system being studied.

Grand Challege 6.4

Engineer advanced computational modeling combined with data integration across temporal and spatial scales.

Advances in the various omics technologies have enabled increasingly expansive studies of biological systems, including in-depth sampling of constituent genomes, transcripts, proteins, and metabolites. Together with the environmental and phenotype metadata collected during the experiments, these omics data types open the door to building multiscale models of the biological system, ranging from gene interaction networks in an organism to multiorganism organization in a community with environmental interactions (see also Chapter 2 on Biological Systems Science, p. 3). Progress has been made on modeling higher-order interactions between organisms in microbial communities using techniques like metabolic flux analysis (Harcombe et al. 2014; Latendresse et al. 2012; Orth et al. 2010), but a substantial gap in continuity still exists between microbial system models and Earth system models, including climate prediction (see also Chapter 3 on Earth and Environmental Systems Sciences and Chapter 4 on Microbial to Earth System Pathways, p. 21 and p. 43, respectively). Relevant to bridging this gap is the need for novel computational techniques, modeling frameworks, and workflows that can effectively handle large volumes of heterogeneous data types.

Grand Challenge 6.5

Conduct research and develop activities that support human understanding of large-scale, multimodal data streams, including the ability to steer experiments in real time.

These activities include visualization of streaming data from various sources and the ability to capture snapshots of experiments in progress, as well as change the direction of these experiments based on immediately available data visible to the researcher or to intelligent algorithms (built to recognize when parameters need to be changed). These capabilities are vital to accelerate progress and reduce wasteful data collection, giving scientists the ability to rapidly tweak experiments while under way, rather than wasting resources to complete and then reset the experiment from scratch. This new paradigm is actively being explored through initiatives such as AIM at PNNL (see Fig. 6.1, p. 73).

Discussion of Action Items

Prioritize Building and Maintaining an Integrated Facility for Long-Term Data Storage, Archiving, and Data Analysis

A new priority should be established, along with the necessary long-term support, to build and maintain an integrated facility for long-term data storage and archival of analytic results. This facility should be secure, mirrored to prevent accidental loss, and usable across current platforms, as well as extensible to technology advances over time.

Evolve Current and Develop New Ontologies and Metadata Standards, as Well as Strengthen Compliance in Funding Announcements and Reporting Standards

A recurring topic at the 2017 BERAC Grand Challenges workshop was the need to make data at all levels available in one common thematic database, from which community users can access the data and to which they can apply their own, or standard, analysis techniques. For example, (1) all climate and climate modeling data should be made accessible in one database (i.e., perhaps similar to KBase), (2) all biological data should be accessible in one database, and (3) linkages between the data resources should be clear and readily accessible. In geosciences research, the EarthCube project (earthcube.org; Gil et al. 2014) consists of a diverse set of scientists to "influence" data collection, analysis, and sharing. The Materials Data Facility (materialsdatafacility.org; Blaiszik et al. 2016) integrates data across dozens of materials datasets, enabling researchers to find, access, and integrate various types of data from different locations. To make data accessible in common thematic databases, the ontologies (see sidebar, What Are Ontologies? p. 81) associated with each theme must be fully developed and compliance enforced. In the genomics realm, the Genomic Standards Consortium (gensc.org) establishes international standards to enable the development of genomic databases, which, in turn, support comparative whole-genomics analyses. In the Earth sciences area, Climate and Forecast (CF) conventions

What Are Ontologies?

Ontologies are formal definitions and names of entities that exist in a domain or field of study. Additionally, ontologies are standard definitions that are accepted across the domain's community and enable clear, precise sharing of data and results within and across domains. A good introduction can be found in the "Ontology Development 101" guide from the protégé project at Stanford University (Noy and McGuinness 2001). Examples of existing ontologies that have been widely adopted are (1) Minimal Information About a Microarray Experiment (MIAME; Functional Genomics Data Society; fged.org/projects/miame/); (2) Minimum Information About a Proteomics Experiment (MIAPE; Proteomics Standards Initiative; psidev.info); (3) Gene Ontology Consortium (geneontology.org); (4) Climate Ontology (National Aeronautics and Space Administration; cds.nccs.nasa.gov/tools-services/ontology/); and (5) Climate and Forecast ontology (W3C Semantic Sensor Network Incubator Group; www.w3.org/2005/Incubator/ssn/ssnx/cf/cf-property).

and metadata (cfconventions.org) are developing metadata standards and best practices. To properly integrate and compare data across biology, climate, and environmental systems, each system must have established and enforced ontologies. Workshop participants unanimously expressed the need for developing standards and ontologies at all BER levels (i.e., climate, environmental systems, and biology), as well as developing an atmosphere that supports and encourages compliance with these standards and ontologies.

Current efforts include the (1) Coupled Model Intercomparison Project (CMIP; World Climate Research Programme; cmip-pcmdi.llnl.gov), a widely used system to provide standard experimental and validation protocols to study Atmosphere-Ocean General Circulation Models (AOGCMs), as well as documentation and data access; (2) Observations for Model Intercomparison Projects (Obs4MIPs; www.earthsystemcog.org/projects/obs4mips/), a repository of well-documented datasets organized by the CMIP5 model output requirements to facilitate model data comparisons; and (3) KBase (kbase.us), a data and software platform for predicting and designing biological function.

As stated in the 2010 Grand Challenges report (BERAC 2010): "Initial focus for standards development should be on the largest and most significant experimental results having the greatest impact on the research community." The same is true for current needs; without the development of strict ontologies and compliance in all components of BER data (i.e., microbiome, systems biology, and environmental systems), research teams will be unable to integrate, analyze, compare, or break down barriers between and across systems. Additionally, metadata and laboratory methods must be clearly documented and available for any publicly deposited experimental data. Without such documentation, consistency within and across laboratories will not be attainable, thus creating insurmountable barricades in integrating multiscalar data across systems.

Facilitate and Encourage Increased Interactions Across Disciplines (e.g., Computer Scientists, Statisticians, Data Analysts, and Experimentalists)

Intra- and interagency collaborations will leverage efforts of ontology development, data deposition, and data integration for specific, targeted system levels, thereby facilitating data exchange and comparisons across these different systems. For example, the National Microbiome Initiative, which was launched in 2016, is a collaborative effort among the White House Office of Science and Technology Policy (OSTP), several federal agencies, and private-sector stakeholders to support the study of microbiomes across different ecosystems (OSTP 2016). Among the primary goals are the facilitation of interdisciplinary research and development of standard technologies, both of which are completely in line with BERAC's recommendations. Similarly, much can be learned from the KBase initiative for biology; this system likely also could be applied to climate and environmental systems, as well as enable easier data exchange, leveraging that was called for at the 2017 BERAC Grand Challenges workshop.

The integration of specific databases developed at other federal agencies [e.g., the Sequence Read Archive at NIH's National Center for Biotechnology Information (NCBI), or the National Science Foundation's (NSF) Biological Databases Initiative] would greatly facilitate integration and comparisons of multiscalar and cross-systems data. Current effort and resources could be reduced notably by more deliberate efforts to coordinate and collaborate among agencies. To conquer the challenge of public data collection ranging from the Earth sciences to microbiome levels, data and tool exchange must be made continuous and seamless across researchers, irrespective of association to federal agency or other funding source. In other words, an agency-agnostic knowledge discovery effort should be implemented.

Such opportunities include, for example:

- NSF's Advances in Biological Informatics program (www.nsf.gov/funding/pgm_summ.jsp?pims_id=5444), which focuses on the development of new informatics tools for biological data and dissemination of biological knowledge. This program actively supports whole-omics annotation efforts, whole-omics informatics tool development, and microbiome research.

- NSF's iPlant collaborative (now called CyVerse; cyverse.org), which is a community-driven infrastructure to study ecosystems and their interactions, providing HPC services, cloud computing, data storage, and an interactive analytical platform for research in the life sciences.

- Globus research data management services (globus.org), which are deployed extensively across DOE, NSF, NIH, and other centers, to enable secure, reliable, high-speed access to data regardless of location and also to support data sharing, publication, and discovery.

- NSF's Directorate for Biological Sciences, which offers a host of programs to support microbial research (e.g., Center for Microbial Oceanography Research and Education), as well as the National Institute for Mathematical and Biological Synthesis (NIMBioS). NIMBioS is especially important as its focus is to bridge the mathematics-biology interface through broader impacts and training.

- NIH's omics annotation guides, such as the Eukaryotic Genome Annotation Pipeline (www.ncbi.nlm.nih.gov/books/NBK169439/), which are widely accessible. Most standard whole-omics databases are housed at NCBI and are widely accessible.

Foster Coordination and Collaboration Within DOE and Among Other Agencies

New experimental technologies are continuously being developed, enabling a wealth of new scientific discoveries by probing deeper, with more detail, into scientific processes. These new experimental facilities, from light source–based instruments to full-biome observations, will be increasingly complex in their makeup and operation and are poised to create extremely large volumes of data (see also Chapter 7 on User Facilities and Research Infrastructure, p. 85). To fully leverage their scientific potential, instruments and experiments will need to be designed with foresight. Scientists must be able to steer these complex experiments to optimize their scientific outcome, based on rigorously analyzed results, while the data are still being collected through a continual cycle of streaming analysis and steering. Sometimes researchers are not able to orchestrate every step of the experiment due to its speed (i.e., high throughput) or duration (e.g., multiyear biome observations). At these points, autonomous, optimized experimental steering in software may need to take over, while still being guided by the experiment's scientific objectives and investigator oversight. Furthermore, many sensors often are available for use during experiments or observations; however, during the experimental design process, ensuring that statistically relevant measurements are obtained is important, so that findings reflect a robust and rigorous scientific discovery process. Currently, however, instrument and experiment design, as well as experimental data analyses, are treated as distinct and separate activities, often with no overlap in personnel, thus leading to many deficiencies in the experimental process.

An initial workshop would be beneficial to provide an overview of experimental techniques, analytical methods, and experimental planning capabilities, while also providing opportunities for BER researchers, experimentalists, computer scientists, and applied mathematicians and statisticians to share their experiences of sensor selection planning and outcomes, including lessons learned. Further workshops could build on this initial overview, focusing on the design and operation of specific large-scale experiments and providing an opportunity to create an integrated instrument and experimental design process.

Such mechanisms foster stronger multidisciplinary teams, which, in turn, can design better experiments, leading to more rapid discoveries while simultaneously making more efficient use of existing instruments. Over time, the hope is that not only large-scale and complex experiments, but all experiments, will be more optimally designed to support scientific discoveries.

7 User Facilities and Research Infrastructure

Providing Resources for BER Science

The fundamental biological, Earth, and environmental systems science supported by the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) is enabled by user facilities and research infrastructure. Across the DOE complex, national user facilities foster a more interdisciplinary research approach by providing a wide range of technologies, methodologies, and instruments critically important for accelerating the science necessary for BER missions in energy and environment. These capabilities are complex and expensive, requiring dedicated staff for technology maintenance, operation, and broad dissemination to individual researchers and institutions. The constantly evolving user facilities will continue to enable scientific breakthroughs in genomics, imaging, structural characterization of organisms and molecules, and atmospheric measurements only if their technologies remain at the scientific forefront. Improving existing technologies is therefore essential, as is advancing newly developed computational, imaging, remote-sensing, and other characterization capabilities deployed in existing or proposed user facilities (see User Facilities and Research Infrastructure Summary of Key Recommendations, p. 86). These resources will underpin critical experiments and generate the data needed to understand important biological and environmental processes.

User Facilities for Genome Sequencing and Molecular and Atmospheric Sciences

BER is the primary sponsor of three DOE scientific user facilities: Joint Genome Institute (JGI), Environmental Molecular Sciences Laboratory (EMSL), and Atmospheric Radiation Measurement (ARM) Research Facility. JGI (jgi.doe.gov) was founded in 1998 as a high-throughput genome sequencing center tasked with completing the three chromosomes allotted to DOE for the Human Genome Project. After becoming a user facility in 2004, the center opened its unique sequencing capabilities to the energy and environmental research communities. EMSL (www.emsl.pnl.gov), founded in 1997, brings together theoreticians experienced in modeling molecular processes with researchers from the physical and life sciences. Applying advanced experimental and computational resources, they undertake fundamental research on physical, chemical, and biological processes that underpin environmental and energy challenges facing DOE and the nation. The multiplatform ARM Research Facility (www.arm.gov) was founded in 1990 to provide detailed, accurate descriptions of Earth's atmosphere in diverse climate regimes—knowledge needed to resolve uncertainties in climate and Earth system models (ESMs). The ARM Research Facility comprises a network of fixed-location observation sites and mobile observation facilities extensively equipped with instruments to measure atmospheric radiation, aerosols, clouds, other atmospheric properties, and surface fluxes of energy and water. The ARM Facility has a long history of data collection, processing, management, and distribution. In 2017, its archive surpassed 1 petabyte (PB) of data, and the facility recently expanded its computational capabilities to enable the coupling of high-resolution models with ARM observational data.

User Facilities and Research Infrastructure
Summary of Key Recommendations

Eight key recommendations for enhancing BER facilities and infrastructure over the next 20 years have emerged based on an assessment of the developments in these areas since 2010 and discussions about maintaining and advancing needed facility and infrastructure capabilities into the future. These recommendations are combined here with action items from other chapters that may be appropriate for inclusion and integration with user facilities or research infrastructure.

7.1. Foster a spirit of collaboration to enable integrative capabilities among BER and Office of Science user facilities, as well as other federal research facilities and infrastructure, thereby promoting a fully interdisciplinary approach to BER-relevant science.

7.2. Solicit input from the BER research community regarding researchers' needs and train them in new experimental, observational, and modeling approaches, thus propagating the knowledge and skills for generating high-impact scientific results.

7.3. Develop innovative enabling technologies and construct and acquire state-of-the-art instruments that exploit the world-leading characteristics of each user facility. This will boost capabilities for basic research in biological systems and Earth and environmental systems science, thereby providing DOE and the nation with leading-edge capabilities for biological and environmental science.

7.4. Develop multimodal imaging and remote-sensing capabilities at user facilities for interrogating length scales ranging from atomic to mesoscale and time scales ranging from nanoseconds to days.

7.5. Build upon existing investments and capabilities at the DOE Office of Science light and neutron science user facilities, continuing to align them with BER missions.

7.6. Further develop the necessary infrastructure at user facilities to study organisms in their natural habitats.

7.7. Develop and adopt technologies to convert genome sequence data into functional understanding at appropriate BER user facilities.

Large-Scale Ecosystem Observations, Experiments, and Modeling

In addition to user facilities, BER supports research, computing, and data infrastructures. Large-scale field research projects include the Spruce and Peatland Responses Under Changing Environments (SPRUCE) study (mnspruce.ornl.gov); Next-Generation Ecosystem Experiments (NGEE) in the Arctic (ngee-arctic.ornl.gov) and tropics (ngee-tropics.lbl.gov); and the AmeriFlux Network (ameriflux.lbl.gov) for measuring carbon, water, and energy fluxes. The SPRUCE experiment assesses ecological responses to environmental change across multiple spatial scales, including microbial communities, moss populations, various higher-plant types, and some insect groups. SPRUCE provides state-of-the-art infrastructure to manipulate atmospheric carbon dioxide levels and air and belowground temperatures for evaluating critical ecosystem and biological processes and properties. NGEE projects seek to improve the representation of critical environmental processes in ESMs by studying systems that are globally important, climatically sensitive, and understudied or inadequately represented in models. NGEE–Arctic research examines how surface and subsurface processes and properties are interconnected across permafrost-dominated tundra ecosystems, while NGEE–Tropics activities focus on predicting the future of the tropical forest carbon sink and associated energy and water fluxes. The AmeriFlux Network research platform spans the full spectrum of ecosystems and climates, from Amazonian rainforests to Alaska's North Slope. The scientific community uses AmeriFlux measurements to assess responses and feedbacks of terrestrial ecosystems to the environment, including changes in climate, land use, and extreme events such as droughts, storms, or wildfire. In 2012, DOE established the AmeriFlux Management Project to support the broad AmeriFlux community and the AmeriFlux sites.

Other DOE User Facilities Enabling BER Research

The BER research portfolio encompasses a wide range of biological and environmental systems whose cellular ultrastructure and physiology, bioprocesses and bioproducts, and environmental and atmospheric transformations demand understanding. These systems function on femtosecond to month time scales and on Ångström to global length scales. Bridging gaps between these scales is one of the greatest challenges in BER research. To understand complex biological and environmental systems across scales, researchers can leverage world-class light and neutron sources and the Nanoscale Science Research Centers (NSRCs), all constructed and operated by DOE's Office of Basic Energy Sciences (BES). Supporting investigations into the basic structure of matter, BES facilities include the Advanced Light Source at Lawrence Berkeley National Laboratory, Advanced Photon Source at Argonne National Laboratory (ANL), National Synchrotron Light Source II (NSLS-II) at Brookhaven National Laboratory, Stanford Synchrotron Radiation Lightsource at SLAC National Accelerator Laboratory (SLAC), and Spallation Neutron Source (SNS) and High Flux Isotope Reactor (HFIR) at Oak Ridge National Laboratory. The research capabilities offered by these state-of-the-art national user facilities can be brought to bear in many areas of science and technology, including the biological and environmental sciences. For example, BER and the National Institutes of Health have made long-term, coordinated investments in constructing and operating beamlines and instruments for biology applications at the light and neutron sources. This collaborative support has led to significant scientific breakthroughs, discoveries, and growth in broad areas of the biosciences. Now is a well-timed and excellent opportunity to advance BER research by expanding existing DOE investments, aligning BER light and neutron source capabilities, and developing and integrating NSRC and electron source capabilities for biological and environmental research. This approach would enrich and expand BER-funded programs already supported at the facilities and open up new science possibilities relevant to BER's mission.

Data and Computational Capabilities

Data produced by facilities and research projects are not particularly useful without capabilities for data management, analysis, and computation. BER researchers have dedicated access to advanced computational resources through EMSL's Molecular Science Computing capabilities and the DOE Office of Advanced Scientific Computing Research's (ASCR) National Energy Research Scientific Computing Center (NERSC; www.nersc.gov). Not only does NERSC provide critical storage space for archiving BER data, it supplies the high-performance computing capabilities essential for the complex, high-volume calculations required to process and analyze the wealth of data emanating from BER user facilities. Further enabling practical use of the data is the DOE Systems Biology Knowledgebase (KBase; kbase.us), an openly available informatics resource for collaborative computational analysis and modeling of plants, microbes, and their communities. KBase integrates data and tools in a unified graphical interface providing users easy access to sophisticated systems biology workflows. Another BER data asset is the Earth System Grid Federation (ESGF), one of the largest-ever collaborative data efforts in

Earth system science (esgf.llnl.gov). Led by BER, this international multiagency federation develops, deploys, and maintains software to facilitate advancements in geophysical science. ESGF disseminates model simulation, observational, and reanalysis data for research assessments and climate model validation. Currently serving more than 25,000 users, the ESGF archive supports 700,000 datasets from laboratories and universities around the world and contains over 5 PB of Earth system science data from more than 25 research projects and 70 model intercomparison projects (see also Chapter 6 on Data Analytics and Computing, p. 71).

Building on Progress Related to the 2010 Grand Challenges

The key role of user facilities in BER science over the years is expected to continue in the coming decades as facilities adapt to the user community's changing needs. Although BER-supported user facilities have diverse technologies, science drivers, and user communities, they share the same overarching goal of providing access to unique technologies that are otherwise out of reach for individual investigators. Ideally, researchers can come to the user facilities with a question rather than just a technique request, relying on the scientific expertise of facility staff to navigate the technological options. In addition to traditional facilities where users physically conduct experiments, increasingly common virtual facilities are transforming how users interact with instrumentation and resources by enabling scientists to work remotely. Already common for some capabilities, this model potentially could expand to other areas of biological and environmental research.

Expanded EMSL's Suite of Tools for Molecular to Mesoscale Investigations. Since the 2010 Grand Challenges report (BERAC 2010), molecular science facilities have added a number of new capabilities and resources. EMSL, for example, has significantly evolved with the needs of the scientific community, providing a wide range of new molecular to mesoscale experimental capabilities that push the frontiers of instrument resolution and sensitivity. Users can leverage these capabilities to investigate biological and environmental processes under dynamic and native conditions. Additional advancements in EMSL's computational hardware, calibrated codes, and data analytics and storage capabilities enable modeling of biological and environmental processes both at discrete scales and at the systems level across multiple temporal and spatial scales. Also possible now are more comprehensive characterizations of proteins, metabolites, and other biomolecules from natural and engineered biological systems using new mass spectrometry and microfluidic approaches. These technologies include a premier 21 Tesla Fourier transform ion cyclotron resonance mass spectrometer with unmatched resolution for identifying and quantifying molecular compounds from cells and sample mixtures such as soils, biofluids, and aerosols. EMSL also has developed or acquired unique capabilities for dynamic multimodal imaging of cells and environmental materials with submicron to subnanometer resolution using fluorescence, electron, cryo-electron, and ion-based imaging technologies. EMSL users are leveraging molecular ecology tools, including stable isotope probing, to study soil diagenesis and metabolism. Facility capabilities for data and image analytics enable users to identify and interpret chemical signals within and among cells and to study environmental interactions among microbes, microbial communities, plant roots, soil minerals, and aerosol precursor molecules. Together, these new technology platforms enable EMSL scientists and users to test new hypotheses about the functional responses of natural systems and to examine the details of those interactions with increasing specificity.

Increased JGI's Sequencing Output and Range of Analysis Tools. Since 2010, JGI has increased its sequencing capacity by more than an order of magnitude, offering a greatly expanded range of capabilities for functional genomics, epigenomics, and metabolomics in concert with more sensitive and accurate methods for *de novo* shotgun sequencing. These methods have dramatically improved knowledge of genome evolution, stress responses, and interactions among plants, fungi, microbes, and microbial communities. Added as a core capability, DNA synthesis is enabling major scientific breakthroughs in molecular science with implications for bioenergy production and nutrient cycling. When combined with existing metabolomic, genetic, and functional characterization, these capabilities will help scientists understand biological systems on the functional level in their native environments.

Launched New ARM Mobile Facilities and Sites While Expanding Radar, Modeling, and Other ARM Capabilities. As part of its significantly increased capabilities, the ARM Research Facility opened (1) a new observational site in the Azores with a focus on marine clouds and aerosol-cloud interactions; (2) a marine-deployable mobile facility; and (3) a third mobile facility currently studying high-latitude aerosol and cloud processes. ARM researchers have made important progress in understanding the complex micro- and macrophysical processes controlling mixed-phase Arctic clouds. Advances in radar technology at the ARM Facility also have enabled unprecedented three-dimensional (3D) views of cloud and precipitation morphology, updraft dynamics, and microphysical properties. Scanning lidar technology (i.e., light detection and ranging), combined with other ARM sensors, has provided in-depth knowledge of the turbulent structure of the atmospheric boundary layer. Advanced aerosol observations, coupled with laboratory studies, have produced fundamental new insights into formation of new aerosol particles and the formation and evolution of secondary organic aerosol in the atmosphere. ARM observations also have been central to several new radiative transfer, aerosol, and cloud parameterizations developed for ESMs, and the facility recently developed capabilities in large eddy simulation modeling to better couple ARM observations with high-resolution modeling.

Leveraged Advanced Technologies at DOE Light and Neutron Sources for Unprecedented Characterization of Biological Systems. In addition to advancements in BER user facilities, new technologies and capabilities at BES-funded light and neutron sources are enabling deeper characterization of plants, microbes, and their environments. Very small beams at these user facilities enable *in situ* and even *in cellulo* data collection. In many cases, researchers can access a facility's experimental resources for structural biology remotely, leveraging the best instruments from their own laboratories. The spatial and temporal resolutions available from photon and neutron beams enable unprecedented characterization and imaging of plant-microbe-environment interactions at scales ranging from subnanometers to millimeters in length and from femtoseconds to seconds in time. This femtosecond resolution is made possible by extremely bright X-ray free-electron laser (XFEL) sources that are empowering new and dynamic structural biology experiments on the ultrafast time scale. The world's first XFEL, the Linac Coherent Light Source (LCLS) at SLAC, has created a new paradigm for imaging biological samples by developing "diffract-before-destroy" methods for sample probing. Efforts are under way to align these and other technologies at the light and neutron resources and NSRCs as a part of an integrated capabilities portfolio for BER missions.

Applying User Facilities to BER Grand Challenges

There are opportunities for continuing and further expanding the important role of user facilities and research infrastructure in supporting advanced BER research, particularly the grand challenges identified in this report (see Table 7.1. Capability Assesment of DOE Office of Science User Facilities and Research Infrastructure for Addressing BER Grand Challenges and Key Recommendations, p. 90). Although existing facilities and infrastructure are well positioned to address research needs in some areas (e.g., JGI for grand challenges in biological systems), needed capabilities are lacking in others (e.g., resources to conduct energy sustainability research). In some cases, an expansion or refocusing of existing technical experience is necessary for addressing new research issues (e.g., leveraging KBase experience for data analytics and computing grand challenges). Working across DOE Office of Science user facilities, as well as community research infrastructure

Table 7.1. Capability Assessment of DOE Office of Science User Facilities and Research Infrastructure for Addressing BER Grand Challenges and Key Recommendations*

		BER Research Infrastructure							Other Office of Science Research Infrastructure		
Grand Challenge Areas	**Grand Challenges or Key Recommendations***	**ARM Facility**	**EMSL Facility**	**JGI Facility**	**AmeriFlux Network**	**Earth System Grid Federation**	**KBase**	**SPRUCE**	**ASCR Computing Facilities**	**BES Light and Neutron Sources**	**BES Nanoscale Science Research Centers**
Biological Systems *(see p. 4)*	2.1 Metabolism	–	●	●	❍	–	❍	●	–	❍	–
	2.2 Engineering	–	●	●	–	–	❍	–	–	–	❍
	2.3 Large data	–	●	●	–	❍	❍	–	●	–	–
	2.4 Geno-pheno	–	❍	●	–	–	❍	❍	–	–	–
	2.5 Technologies	–	●	●	–	–	–	–	–	❍	❍
Earth and Environmental Systems *(see p. 22)*	3.1 Models	❍	–	–	❍	❍	–	❍	❍	–	–
	3.2 Observations	●	❍	–	●	❍	–	●	❍	–	–
	3.3 Aerosols	●	❍	–	❍	❍	–	–	❍	–	–
	3.4 Tipping points	❍	❍	❍	❍	❍	–	❍	❍	–	–
	3.5 Water cycle	❍	❍	–	❍	❍	–	❍	❍	–	–
	3.6 Ice dynamics	–	–	–	–	❍	–	–	–	–	–
	3.7 Forcings	❍	–	–	❍	❍	–	●	❍	–	–
	3.8 Stability	–	–	–	❍	❍	–	❍	❍	–	–
Microbial to Earth System Pathways *(see p. 44)*	4.1 Biogeochemistry	–	●	●	❍	❍	❍	●	–	–	–
	4.2 Communities	–	❍	❍	❍	❍	❍	●	–	–	–
	4.3 Traits	–	❍	❍	❍	❍	❍	❍	–	–	–
	4.4 Microbes	–	●	❍	❍	❍	❍	●	–	❍	–
Energy Sustainability *(see p. 60)*	5.1 Coupling	❍	❍	–	❍	–	–	–	–	–	–
	5.2 Alternatives	–	–	–	–	–	–	–	❍	–	–
	5.3 Variability	–	–	–	❍	–	–	❍	–	–	–
	5.4 Scales	–	–	–	–	❍	–	❍	–	–	–
Data Analytics and Computing *(see p. 75)*	6.1 Large-scale	●	❍	●	–	●	❍	–	❍	●	❍
	6.2 Infrastructure	❍	●	●	–	●	❍	–	❍	●	❍
	6.3 Algorithms	–	❍	❍	–	❍	❍	–	❍	–	–
	6.4 Integration	❍	❍	–	❍	❍	❍	❍	❍	–	–
	6.5 Real-time analysis	–	–	❍	❍	–	❍	–	–	❍	❍
User Facilities and Research Infrastructure *(see p. 86)*	7.1 Collaboration	●	●	●	●	●	●	●	●	●	●
	7.2 Training	●	●	●	●	●	●	●	●	●	●
	7.3 Technology	●	●	●	●	●	●	●	●	●	●
	7.4 Imaging	●	❍	–	–	–	–	–	–	●	●
	7.5 Light/neutron	–	–	–	–	–	–	–	–	●	–
	7.6 *In situ*	–	●	❍	–	–	–	●	–	❍	❍
	7.7 Genomes	–	❍	●	–	–	❍	–	–	–	–
Emerging Technologies *(see p. 96)*	8.1 Cell characterization	–	●	●	–	–	❍	–	–	❍	❍
	8.2 Throughput	–	●	●	–	–	❍	–	–	❍	–
	8.3 Atmosphere	●	●	–	●	–	–	–	–	–	–
	8.4 Multiscale	–	❍	–	❍	❍	❍	❍	❍	❍	❍
	8.5 Computing	–	❍	–	–	●	❍	–	●	–	–
	8.6 Nanotechnology	–	–	❍	–	–	–	–	–	–	●
	8.7 Remote sensing	❍	❍	–	❍	–	–	–	–	–	❍

* Refer to text in respective chapters for descriptions of individual grand challenges or key recommendations. Existing capabilities and partial capabilities that could be leveraged for the grand challenges and key recommendations are denoted by ● and ❍, respectively.

and programs, will help strengthen and further develop advanced science capabilities within BER.

Maintaining and Advancing Facility and Infrastructure Capabilities

Collaborative Research Among Scientists and Facilities

The BER research community studies increasingly diverse, complex, and challenging biological and Earth and environmental systems science problems whose solutions require a spirit of collaboration among both researchers and facilities and a combination of cutting-edge capabilities leading to impactful discoveries. Collaborations are important to enabling integrative technologies that underpin a fully interdisciplinary approach to BER-relevant science at user facilities. Scientists accustomed to working as individual principal investigators (PIs) may need incentives to build interdisciplinary collaborations and access a range of capabilities to pursue their research questions. The BER Scientific Focus Areas (SFAs) offer one model of effective interaction among researchers.

Achieving these objectives also requires a mechanism allowing scientists to submit a single proposal for access to all relevant capabilities at BER-funded facilities. This process has been started by the Facilities Integrating Collaborations for User Science (FICUS) program, which aims to encourage and enable scientists to more easily incorporate into their research the expertise and capabilities of multiple DOE user facilities. As part of FICUS, user facilities collaboratively develop calls for proposals and coordinate their reviews. So far, JGI and EMSL have issued a joint call, as have EMSL and ARM, and JGI and NERSC. Similar ongoing collaborative initiatives among BES facilities aim to deliver innovative science while expanding the reach and impact of the user facilities. These initiatives include pilot partnerships among SNS, HFIR, and NSLS-II and among JGI, EMSL, and ANL. Another collaborative opportunity to consider is a pilot program that integrates capabilities from JGI, EMSL, light and neutron sources, and NSRCs.

In July 2016, JGI sponsored a FICUS workshop where participants proposed establishing a core steering committee to coordinate facility access efforts. Suggestions for the committee's charge and configuration include:

- Frequent communication (at least monthly) via telecom.
- One representative member from each facility.
- Additional scientists to act as "points of contact" for each technique.
- Creation of a BER science user group to provide advice to the steering committee.
- An outreach process connecting FICUS with major BER programs, such as new Bioenergy Research Centers and SFAs.
- Formulation of steps for developing a web portal to provide information on BER user facility capabilities and access to them.

Training and Outreach

Expanding and educating the BER research community are important for advancing BER science. These efforts start with effectively disseminating information about the capabilities and methodologies available at user facilities. Each facility provides descriptions of its capabilities on its website, which is linked to DOE's site. However, a common web portal with coordinated information about capabilities and user access across the facilities could help scientists navigate the options. Such efforts are beginning to take shape with the recent launch of a web portal describing all BER-supported techniques and facilities for structural biology research ranging from the atomic to cellular scales (www.berstructuralbioportal.org).

Reaching out to members of the BER research community, especially BER Early Career Research

Program grantees, to evaluate their needs and train them in new experimental approaches could pave the way to high-impact scientific results. Workshops and hands-on training would be valuable in this regard, particularly if scheduled in conjunction with BER PI meetings and national conferences focusing on BER-related science and technologies. Facility user meetings should not be overlooked as opportunities to strengthen ties with existing users, engage prospective new users, and highlight BER-relevant research. Leveraging recent developments in remote access and automation could prove useful for training purposes, as well.

The DOE Office of Science Graduate Fellowship program is underused as a means to facilitate cross-technique and cross-facility training. Postdoctoral and visiting fellow programs also could enable PIs to take advantage of unique facility resources and develop expertise for use in their own laboratories, helping to create a self-sustaining user community in BER-related research.

New Methods, Complementary Technologies, and Instruments

Key to future impact and innovation are continued development and pursuit of technologies that build on unique user facility resources. Maintaining high standards for staff expertise in technology development and user support will be critical as well. User facilities will boost capabilities for basic biological and environmental research, thus providing leading-edge science for DOE in these areas.

Other important frontiers to pursue are the development of new instruments enabling multimodal imaging capabilities over length scales from subnanometers, millimeters, and beyond and the integration of existing complementary techniques. The research community can further expand and optimize new LCLS resources by developing and coupling light source technologies that feed into XFEL resources and *vice versa*. Also needed are research and development efforts and capital investment in characterization capabilities for single-particle cryo-electron microscopy (cryoEM) and cryo-electron tomography (cryoET). These techniques, which can bridge the gap between atomic and cellular scales, potentially could revolutionize future protein structural biology and cellular imaging but require significant capital and human investment out of reach for many research institutions. BER researchers need access to cryoEM user facilities for structural characterization of plant and microbial systems. Co-locating cryoEM and cryoET facilities at neutron and light sources or NSRCs would be practical, capitalizing on established open-access user programs and the expertise in operating advanced instrumentation. Additional opportunities exist to leverage and enhance BER-supported facility capabilities through interactions with the NSRCs in imaging and nanofabrication (see also Chapter 8 on Emerging Technologies, p. 95).

In addition to cutting-edge imaging techniques, advanced sequencing technologies and approaches are needed. Such capabilities will be critical to improving genome data quality, gene annotations, and protein structural and functional predictions, ultimately enabling the use of sequence data for functional understanding. As with sequencing technologies, metabolomic and proteomic instrumentation is evolving rapidly, requiring frequent updating to achieve the high spatial resolution goals laid out in this report. Although still nascent, deployable technologies for these applications are potentially transformative.

A central task for BER is to provide research opportunities and scientific user facilities that advance the understanding of complex biological, Earth, and environmental systems important to DOE missions. User facilities provide a wide range of capabilities and technologies that drive research in these fields. Computing resources play a critical role in supporting these facilities and addressing the complexity inherent in BER systems of interest. New capabilities are needed, along with increased coordination and integration among user facilities, to fully examine the different spatial and temporal characteristics of these systems, model them at larger scales, and understand their interactions within and across spatial and temporal scales. This is captured particularly well by the proposed testbeds and synthesis center for energy sustainability research (see Discussion of Action Items section in Chapter 5, p. 67).

Computing and Data Management and Analysis

User facilities are supported by revolutionary changes in advanced computational and artificial intelligence approaches and in capabilities for handling large datasets. Adequate computational capacity is critical for (1) supporting spatial and temporal imaging approaches that improve mechanistic understanding of molecular to mesoscale biogeochemical processes, (2) analyzing advanced radar remote-sensing datasets to better understand microphysical cloud and precipitation processes, and (3) handling the large volumes of sequence and spectral data produced by genomic techniques. Another essential component of facility computing activities is the development of 3D models and high-performance computational methods for disordered systems and spectroscopic characterization. As the complexity of systems studied by BER researchers increases, so too will the need to develop computational tools enabling analysis of datasets from disparate sources over multiple length and time scales. Researcher access to the national laboratories' cutting-edge computing resources for advanced data analysis and modeling could have an important role in developing next-generation capabilities. In all these areas, collaboration and integration among different facilities and techniques are essential for maintaining leadership in energy and environmental science (see also Chapter 6 on Data Analytics and Computing, p. 71).

8 Emerging Technologies

Developing Capabilities to Characterize and Represent Biological and Environmental Processes

The U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) aims to understand fundamental biological, biogeochemical, and physical principles to be able to predict processes occurring at scales ranging from the molecular- and genomics-controlled smallest scales to environmental and ecological processes at the scale of planet Earth. Accomplishing this goal requires a range of capabilities to characterize and represent biological and environmental processes. Some of these capabilities already exist; others require further development and integration to enable complete analytical, imaging, and observational investigations. Further, when integrated at DOE user facilities, these capabilities can be combined and leveraged to foster an interdisciplinary approach to BER-relevant science. Many scientific breakthroughs are precipitated or enabled by technological developments as evidenced by advances in BER science over the past several years. These advances are intertwined with technological breakthroughs in omics, imaging, structural characterization, and atmospheric monitoring.

The technologies, tools, and methods described in this chapter are critical for advancing new, in-depth understanding of the complex, multiscale, multicomponent systems that are central to BER's energy and environment missions (see Emerging Technologies Summary of Key Recommendations, p. 96). As new technologies, tools, instruments, platforms, algorithms, and approaches are developed, however, they must be made readily accessible to the researchers who need them to provide maximum value. Researchers must be able to either import and use these new technologies in their own laboratories or have access to them via user facilities with appropriate advanced expertise and support.

Building on Progress Related to the 2010 Grand Challenges

Many of the grand challenges presented in the 2010 report (BERAC 2010) involved the development and application of new technologies, and significant progress has been made in many BER research areas described below.

Dramatically Increased the Quantity and Quality of Genome and Metagenome Data. This increase is the result of enhanced experimental and computational methods and resources, including less expensive, higher-throughput, and longer-read sequencing technologies; improved methods for nucleic acid extraction and sequencing library preparation; sophisticated sequence analysis software; and enhanced databases and interfaces such as the Integrated Microbial Genomes and Microbiomes (IMG/M) system at DOE's Joint Genome Institute (Chen et al. 2017) and the DOE Systems Biology Knowledgebase (KBase).

Improved Prediction, Expression, and Characterization of Secondary Metabolite Biosynthetic Pathways Involved in Interorganismal Interactions. These improvements include enhanced technologies to define the transcript, protein, and metabolite profiles of any particular organism with high sensitivity and selectivity, as well as mass spectrometry (MS)–based imaging to enable spatially resolved measurement and

Emerging Technologies
Summary of Key Recommendations

In light of BER- and DOE-mediated technological advancements since 2010 and discussions about how emerging technologies will continue to develop into the future, the workshop produced several key recommendations for how BER could facilitate and capitalize on emerging technologies over the next 20 years:

8.1. Characterize the genotype and phenotype of individual cells, including genomics, transcriptomics, proteomics, and metabolomics, to enable high-resolution predictive biology.

8.2. Increase throughput and integration of genomics, transcriptomics, proteomics, and metabolomics to enable improved translation from the molecular to cellular realm.

8.3. Characterize key processes of aerosols, clouds, turbulence, atmospheric convective systems, and their interactions to enable better modeling of water, carbon, and energy cycles of the Earth system and predictions of future climate change and extremes in atmospheric, terrestrial, and watershed events.

8.4. Integrate data covering broad time and length scales—from seconds to years and from Ångströms to the Earth scale—to enable multiscale comprehension and simulation.

8.5. Develop integrative and interpretive computational approaches that can handle large, disparate data types from multiple and heterogeneous sources using advanced and exascale computing.

8.6. Explore the development and application of nanoscience and nanotechnology to further understand biological systems at nanoscales or larger scales (e.g., nanosensors).

8.7. Explore the development and application of microsatellites and unmanned aerial systems as mechanisms for observation of atmospheric and terrestrial variables.

These technological efforts will provide the basis for addressing many of the challenges faced in biological systems science, Earth and environmental systems sciences, microbial to Earth system pathways, and energy sustainability.

visualization of metabolite mobility, flow, transport, and accumulation between cells (Musat et al. 2016). Such methods enable the identification of specific pathways and, ultimately, the modeling of interorganism interactions.

Advanced Methods for Single-Cell Analysis. Only a few single cell–derived genomes were published in 2010 (Podar et al. 2007; Woyke et al. 2009; Woyke et al. 2010). Since then, hundreds of such genomes have been produced, revolutionizing understanding of the evolutionary tree of life (Rinke et al. 2013). Single-cell transcriptomics has grown significantly, with numerous approaches being applied to capture, tag, and analyze messenger RNA (mRNA) from individual cells (Kolodziejczyk et al. 2015). A variety of creative methods are extending single-cell methods to proteins, metabolites, and epigenetics, suggesting that multiomics analysis of individual cells may not be far off (Bock et al. 2016; Tsioris et al. 2014; McGlynn et al. 2015; Rosenthal et al. 2013). At the single-cell level, MS-based imaging is emerging as a viable technology for visualizing a single cell and mapping the image to metabolite information (Pett-Ridge and Weber 2012).

Advanced the Field of Genome Editing. Genome editing is growing at an exceptionally rapid pace and shows astonishing promise in advancing genomic science (Damian and Porteus 2013). A system known as CRISPR (i.e., clustered regularly interspaced short palindromic repeats) is a powerful genome editing tool for inactivating genes, correcting mutated sequences, or inserting intact genes (Cong et al. 2013; Qi et al. 2013) in both prokaryotic and eukaryotic systems. This new genome engineering method is derived from an adaptive immune system that bacteria and archaea use to protect themselves against foreign invasive elements.

Advanced "Pump-Probe" Methods. These advances in structural biology include molecular dynamics pump-probe methods with free-electron lasers (FELs) and serial crystallography (Glownia et al. 2010; Johansson et al. 2013; White et al. 2016). Both X-ray FELs (XFELs) and light sources can perform pump-probe dynamic experiments on the ultrafast time scale from three-dimensional (3D) crystals to enable the study of transient states (Levantino et al. 2015; Young et al. 2016).

Developed Cryo-Electron Microscopy (CryoEM). This major new technology has advanced to the point where high-resolution atomic models can be obtained (without the need for crystals) for ~100 kilodalton (kDa) proteins, viruses, membrane proteins, filaments, and molecular machines (Egelman 2016; Allegretti et al. 2015; Callaway 2015; Li et al. 2013; Wang et al. 2014; Glaeser 2016*)*.

Advanced Imaging and Microscopy Techniques. These advances, including transmission X-ray microscopy, neutron imaging, and confocal and epifluorescence microscopy, are enabling interrogation of difficult samples such as whole organisms, biofilms, and soils, at dramatically improved spatial and temporal resolution (Leis et al. 2009; Larabell and Nugent 2010).

Employed Ultrahigh-Resolution MS for Molecular Characterization. This technique has enabled molecular characterization of the thousands of soil organic molecules previously described in terms of bulk soil carbon:nitrogen:phosphorus ratios (Musat et al. 2016). With this new understanding, refined models of carbon pool diversity can be evaluated across ecosystems, and evidence of potential metabolite pathways can provide much needed context for integrative multiomic approaches.

Advanced Nanoscience and Nanotechnology. New tools have been developed for mapping and characterizing connections at the nanoscale and increasing resolution of imaging technologies (Biteen et al. 2016). Coordinated investments were made through the interagency National Nanoscience Initiative (www.nano.gov), including the establishment of five Nanoscale Science Research Centers (NSRCs), which are national scientific user facilities supported by DOE's Office of Basic Energy Sciences. The five NSRCs are (1) the Center for Functional Nanomaterials at Brookhaven National Laboratory (BNL), (2) Center for Integrated Nanotechnologies at Los Alamos National Laboratory and Sandia National Laboratories, (3) Center for Nanoscale Materials at Argonne National Laboratory (ANL), (4) Center for Nanophase Materials Sciences at Oak Ridge National Laboratory, and (5) Molecular Foundry at Lawrence Berkeley National Laboratory (LBNL).

Deployed New Instruments for Advancing Atmospheric Science. These instruments are measuring clouds and aerosols at fixed and mobile sites supported by DOE's Atmospheric Radiation Measurement (ARM) Research Facility. They include X-band, W-band, and Ka-band scanning cloud radars; C-band polarimetric scanning radar; zenith pointing Doppler lidar; and unmanned aerial systems (UAS) and tethered balloons to increase the sampling frequency of both temporal and spatial cloud measurements.

Enabled Chemical Change Measurements in Live Microbial Specimens. Synchrotron-based broadband Fourier transform infrared (sFTIR) spectroscopy or spectromicroscopy now permits live cell analysis and observation of chemical events in the same cell or sample by avoiding the damage caused by other methods (Holman et al. 2002). As a result, sFTIR spectromicroscopy (Holman et al. 2010; Probst et al. 2013) and time-resolved microfluidic sFTIR spectromicroscopy (Holman et al. 2009; Holman et al. 2010; Loutherback et al. 2016) are now matured techniques for measuring chemical changes in live microbial specimens with diffraction-limited spatial resolution of 2 to 10 microns (μm) and temporal resolution of several to tens of seconds.

Developed Scalable Computational Models for Molecular Transformations. These models were developed in NWChem, high-performance computational chemistry software, to describe molecular transformations in complex environments. Improvements include (1) new, enhanced capabilities to support experiments [nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), photoelectron spectroscopy (PES), electron energy-loss spectroscopy (EELS), and ultraviolet-visible (UV-Vis) spectroscopy] at DOE's Environmental Molecular Sciences Laboratory (EMSL); (2) methods to interpret and support state-of-the-art X-ray light source experiments at user facilities housed at LBNL, SLAC National Accelerator Laboratory (SLAC), ANL, and BNL; (3) optimization methods to study reaction mechanisms; (4) relativistic approaches to support heavy-element chemistry; and (5) new electronic structure methods and multiscale formulations. Significant progress also has been achieved in the development of novel algorithms for emerging multicore architectures [e.g., graphics processing unit (GPU) and Intel Xeon Phi] and expert and machine-learning systems (e.g., EMSL Arrows).

Emerging Technologies and the Importance of Integration Across Scales

Emerging technologies play a key role in state-of-the-art research. Experimental and modeling needs can drive technology or software development. Conversely, existing technologies can be brought together in many unique ways to expand their capabilities or explore new research questions that are essential for complete process understanding at the multiscale perspective.

A central theme at the Biological and Environmental Research Advisory Committee's 2017 Grand Challenges workshop was the concept of scaling, both in terms of the inherent scales of measurements and the importance of cross-scale understanding. At larger scales, observations can be made of emergent phenomena that arise from the interactions of complex and collective processes whose manifestations are not apparent at reduced scales. At smaller scales, and at reduced complexity, mechanistic understanding can be obtained of dominant processes and the underlying mechanisms that give rise to the full phenomenological expression. Nevertheless, these top-down and bottom-up perspectives must be integrated in order to develop quantitative, predictive models. There are various approaches to the development of multiscale models, including hierarchical models where most process understanding is contained at the smallest scales, reduced-order models developed for specific processes at larger scales, and system models of multiple interacting processes. All these models should continue to be pursued as strategies to combine data and knowledge across scales.

The rest of this chapter groups the grand challenges in technology based on the scale at which they apply. While such classifications are imprecise, they provide a context in which potential cross-technology links are more apparent.

Molecular Scale

Structural Biology and Imaging. Many different biophysical measurements can provide spatial and temporal information, but structural imaging offers the most direct means for characterizing a biosystem's organization. Atomic and molecular structure provides the foundation for functional models that bridge larger-scale biological processes to underlying chemistry. Current intuitive understanding of biomolecules at nanoscales remains insufficient to connect the exponential increases in genomic sequence information to form, function, and mechanism. New technological developments have transformed the ability to probe biological phenomena at the nanoscale and relevant temporal scales. These developments include cryoEM and cryo-electron tomography (cryoET), NMR, neutron macromolecular crystallography (NX), small-angle neutron scattering (SANS), small-angle X-ray scattering (SAXS), and macromolecular X-ray crystallography (MX). New advances in experimental methods combined with computation are required to better understand biosystem and ecosystem dynamics such as enzyme function, photosynthetic light capture, and engineering for optimal biofuel synthesis.

FELs, Serial Crystallography, and Electron Microscopy. In X-ray and neutron structural biology, the key to future impact and innovation is continued development of enabling technologies through the construction of state-of-the-art instruments that exploit the world-leading characteristics of each of the light and neutron sources. These instruments will boost capabilities for conducting basic biological research, thereby producing leading-edge science for BER missions in energy and environment. At the frontier are new instruments that will enable multimodal imaging capabilities over length scales from subnanometers to millimeters. Opportunities are available to take further advantage of the new XFEL sources and optimize their use by developing coupling technology at light sources that feeds into XFEL sources and *vice versa*. The commissioning of fourth-generation XFEL light sources such as the Linear Coherent Light Source at SLAC has created the new paradigm of "diffract-before-destroy" imaging (Spence and Hawkes 2008). These femtosecond X-ray sources empower new research in MX and absorption and emission spectroscopy (Kern et al. 2013) by allowing analysis of samples—such as 3D nanocrystals (Aquila et al. 2012), 2D crystals (Frank et al. 2014), and large single-particle complexes (Saldin et al. 2011)—not amenable to analysis at synchrotron light sources.

Molecular Dynamics Pump-Probe Methods. Visualizing protein dynamics relevant to BER missions requires pump-probe methods spanning the femtosecond to second or longer time regimes. From fastest to slowest, typical protein dynamics include electronic and photoinduced transitions, ligand binding, active site motions, loop and helix movement, domain and subunit motions, assembly and disassembly, and molecular and macromolecular diffusion. Continued improvements are needed for stimulating or triggering a reaction or process (i.e., "pumping the system"), in combination with new approaches for capturing time-resolved structural and chemical data (i.e., "probing the system") across the relevant time scales. Fast and ultrafast X-ray approaches (i.e., synchrotron- and XFEL-based, MX, serial crystallography, and spectroscopy) and electron approaches [i.e., dynamic transmission electron microscope (DTEM), ultrafast electron diffraction, and ultrafast electron microscopy] promise new opportunities for interrogating macromolecular complexes in their native geometry and chemical environment. Simultaneous measurements using multiple techniques can provide very new understanding of biological and environmental processes.

Both XFELs and light sources can perform pump-probe dynamics experiments on the ultrafast time scale (tens of femtoseconds to picoseconds to microseconds) from small 3D crystals (Levantino et al. 2015; Young et al. 2016). However, new source developments improving peak brightness, coupled with new methodologies to permit atomic resolution pump-probe imaging of 2D crystals and, ultimately, single-particle macromolecular complexes, would be highly beneficial for understanding the structural dynamics of a wide range of BER-relevant targets in near-native geometries and environments. These techniques potentially could revolutionize future structural biology and cellular imaging, but they require significant capital and human investment currently out of reach for many research institutions.

While these "newer" technologies provide novel insights into more complex systems and at smaller time scales, mature technologies such as MX, SAXS, NMR, NX, and SANS continue to evolve and fulfill new roles in biosystem imaging and analysis. Synchrotron-based MX remains unparalleled in providing high-throughput, high-resolution macromolecular structures, with recent advances extending these results to more challenging and flexible systems. The MX workflow is particularly well adapted for conducting expansive surveys and offers a means for rapidly testing predictions. SAXS also supplies high-throughput structural information (Brunette et al. 2015; Bale et al. 2016; Boyken et al. 2016), while NMR, SANS, and NX allow for altered contrast in heterogeneous materials and provide direct insights into the role of hydrogen atoms that can only be inferred by other techniques (Frederick et al. 2015). In all cases, support for these critical capabilities, as well as their continued innovation, is needed to enhance understanding of biosystem and ecosystem complexity and unravel mechanisms central to biological and environmental processes. Neutron imaging shows great promise for plant biology, revealing plant architecture and internal plant-water dynamics, for example, with pulses of water or heavy water (D_2O) *in situ* and in real time.

Macromolecular Characterization with CryoEM. Recent advances have made possible the study of macromolecules of smaller size and at much higher resolution in vitrified solutions via cryoEM (Merk et al. 2016). These advances include new direct electron detector device cameras with very fast readout (e.g., Gatan Summit Direct Electron Detector), advanced microscopes with very stable optics (e.g., Titan Krios), new modes of data acquisition, control for radiation damage, and correction for sample motion. The ability to collect data on macromolecules in solution in near-native environments without having to grow crystals cannot be underestimated. The relatively modest number of particles needed for high-resolution model reconstruction highlights the power of this technique's recently developed hardware and software. Additional developments in phase plate technology have expanded this powerful technique's applicability to even smaller protein complexes once thought well beyond reach with conventional film or charge-coupled device (CCD) cameras. Observed heterogeneity of molecules in vitrified specimens has enabled exciting discoveries into the conformational landscape of macromolecular complexes (Amunts et al. 2015; Bai et al. 2015; Nguyen et al. 2016). In addition to a cryoEM revolution for going smaller with single-particle structure determination, new methods for micro-electron diffraction (microED) have enabled atomic-resolution structures from protein crystals previously too large for transmission electron microscopy (TEM). Expected improvements in sample preparation, hardware, detectors, data acquisition, and software will enable much wider use and increase the throughput of atomic-resolution structure determination. In addition, emerging TEM capabilities on the nanosecond to millisecond time scale with atomic to near-atomic resolution (Evans and Browning 2013) would extend this approach to spatiotemporal regimes relevant to secondary structure element, domain, and subunit motions central to regulating protein activity. The ability to visualize all conformational states of a given enzyme during a reaction continuum would help advance isozyme engineering for more efficient biofuel production. Continuing development of methods and hardware for single-particle cryoEM, cryoET, TEM, and microED is much needed for enabling high-throughput structural determination and providing broader access to the BER research community.

Advances in NMR and EPR. NMR spectroscopy and EPR have long been used to elucidate protein structures, but they have been hampered by a lack of *in situ* capabilities for interrogating molecule:substrate binding under conditions pertinent to biogeochemical environments (e.g., ambient temperature and pressure). NMR is recognized as an important and often indispensable analytical tool for structure elucidation, molecular identification, measurements of dynamics (on time scales from nanoseconds to hours), and imaging. Many major breakthroughs in biotechnology, medicine, and clinical treatment have been tied to rapid advancement of new and improved NMR tools, but a general lack of sensitivity in NMR typically precludes samples with low analyte concentrations. However, new commercial hardware developments [i.e., 1.7-millimeter (mm) NMR microcryoprobe] have made possible the detection of submicromolar concentrations for sample-limited projects [~30 microliters (μL)] such as those that require fractionation prior to NMR analysis. Coupling specialized sample containment systems that provide environmental control over pressure, temperature, and head space gas will yield information that corresponds to field and *in vivo* conditions. Using these new hardware components will enable biological molecules and their origins and interactions in environmentally relevant samples to be identified and quantified at a level of accuracy that far surpasses anything to date. Another long-standing limitation of NMR spectroscopy, the need for samples with macroscopic dimensions, potentially could be addressed by lab-on-a-chip technology. On a fundamental level, the emerging capability to study individual microsystems such as biological cells using a range of tools will yield understanding of microscopic heterogeneity and interaction dynamics.

Another key component is high-performance computational methods development (e.g., 3D models, disordered systems, spectroscopic characterization, and dynamics). As the complexity of systems under investigation by BER researchers increases, the development of computational tools that enable analysis of datasets from disparate sources over multiple length and time scales is critical. National laboratory resources provide opportunities for scientists to use existing cutting-edge computing capabilities for advanced data analysis and modeling, and this access could have an important role in developing next-generation capabilities. In all these areas, the cross-fertilization between different facilities and techniques is essential for effective national stewardship.

Advances in Imaging and Microscopy. Biology is reducible to a series of highly orchestrated chemical events occurring at defined locations within cells. New methods are needed to image the localization of enzyme reactions and flow of chemicals and macromolecules within and between cells and unravel the complex network of molecular interactions defining cell phenotypes and functions. Substantial improvements in spatial resolution for fluorescence microscopy methods recently have been made with new super-resolution methods and minimization of optical flux needed for detection. These methods now enable nanometer positional accuracy (Balzarotti et al. 2017). Stimulated Raman scattering microscopy, a nonlinear optical approach, also has significantly expanded as a quantitative direct imaging method that allows tracking of cellular dynamics in a label-free manner or by using selective Raman tags.

Additional advances in nonfluorescence optical detection methods include subdiffraction limit measurements using tip-enhanced capabilities and stimulated Raman scattering. Over the last few years, developments in near-field infrared chemical detection technologies have enabled vibrational signature measurements of organic and biological materials at the nanoscale. These new capabilities include infrared scattering–type scanning near-field optical microscopy (IR s-SNOM) and infrared atomic force microscopy (AFM-IR). Efforts are under way to combine IR s-SNOM with the more established tip-enhanced Raman spectroscopy and optical absorption nanospectroscopy to probe the vibrational and chemical signatures of biomolecules at or near the surface of cells in aqueous environments.

Ion microscopy has undergone a number of developments in recent years, from vacuum-compatible microfluidics enabling *in situ* probing of liquid interfaces using time-of-flight secondary ion mass spectrometry (ToF-SIMS) to new frontiers for nanoscale secondary ion mass spectrometry (NanoSIMS) and similar direct imaging approaches. These approaches provide maps of molecular components and can be used to study complex microbiological processes and community interactions.

Genetics and Genomics. Genomic data have been generated from thousands of organisms and environments, enabled by vast increases in sequencing throughput. Yet much of the biological world remains uncharacterized at the molecular level. Sequence data from natural environments still typically lack high-identity matches to reference genomes, and much of the sequence in both reference genomes and metagenomes is unannotated or poorly annotated.

Improving gene function understanding will require advances on multiple fronts.

The continued expansion of the reference genome database through genomic sequencing of both microorganisms and macroorganisms across the tree of life is a critical first step in linking genotype to phenotype and gene to function. These genomic references are critical infrastructure for countless downstream analyses and can be generated not only through sequencing of cultivated, known organisms, but also single-cell and metagenome sequencing of uncultivated, unknown organisms. Improved and higher-throughput sequencing technologies and approaches will accelerate progress toward genomic characterization across the tree of life.

Equally critical to genomic interpretation, but lagging far behind sequence data generation, is functional characterization of genes and genomes. Computational analysis alone will not definitively assign function, yet most methods for functional analysis remain one gene at a time. High-throughput annotation strategies such as transposon mutagenesis and sequencing (Tn-seq) are highly informative but limited in application, so alternatives and improvements should be actively sought.

Other Omics. Transcriptomics, proteomics, and metabolomics are critical molecular-scale techniques, but the expense, technical difficulty, and throughput of these methods still pose impediments, particularly in terms of spatial and temporal resolution. Simplified, high-throughput, and inexpensive methods for these inquiries would dramatically expand the ability to interrogate metabolic and regulatory processes.

Currently, metabolite characterization requires mapping mass spectrometric features to a known database, but, with secondary metabolites, such databases are limiting. Comprehensive public databases are urgently needed, as are new approaches for database-free metabolite characterization.

Direct single-molecule RNA sequencing (e.g., with Oxford Nanopore; Garalde et al. 2016) potentially could open up whole new avenues of investigation.

Reduced-Input Material for Molecular Assays. Analytical instrumentation often vastly outstrips current ability to effectively sample, prepare, and deliver small samples to the instrument. New microfluidic sample preparation–based techniques are enabling high-throughput processing of small samples with very little loss, and structural biology advancements have significantly reduced the amount of material required for structural characterization. XFELs use nanocrystals and microcrystals, and efforts to reduce sample needs and determine structure with a single crystal are common practice at light sources. Under optimal circumstances, cryoEM can produce structures with just several thousand individual particles.

Cellular Metabolite Dynamics. Comprehensive biosystem characterization requires understanding how a system senses environmental perturbations and the mechanisms that cells use to make changes in response to these cues. By developing new bioimaging and spectroscopy capabilities spanning electron, ion, optical, neutron, and X-ray modalities and fusing them with new advances in systems biology, computation, and automation, a deeper understanding of the link between genome to phenome is possible. Exploiting and controlling those linkages would radically enhance the ability to design and harness biosystems for cheaper, more efficient biofuels and bioproducts, increased crop yields, and biological routes to soil remediation. Biological systems rely on chemistry taking place at the correct time and in the right place (Le Gros et al. 2012). Consequently, tracking the location and concentration of molecules over time is a key characterization step for biological investigations. The accurate characterization of small numbers of molecules produced by native or engineered biosystems remains challenging, particularly for single-cell measurements. NMR, FTIR, and MS approaches are well suited to identifying chemical composition. MS imaging (e.g., NanoSIMS) enables part-per-million detection of ions and isotopes with up to ~50 nanometer spatial resolution, although it is a static and destructive method best suited to imaging chemical gradients and cellular assimilation.

Nanotechnology. Nanotechnology enables the manipulation of individual molecules and atoms and has emerged as a dynamic field of innovation (Allhoff et al. 2010). Through nanotechnology, scientists can create novel materials and devices at the nanoscale, and these structures can be used to further BER missions in imaging, monitoring, and understanding of biological and environmental systems to advance energy innovation and development. For example, nanopores can be used to quickly and cheaply identify and characterize small molecules and genomes (Branton et al. 2008); nanosensor networks can be used to detect chemicals and monitor environmental variables potentially in both single and multicellular contexts (Akyildiz and Jornet 2010).

Cellular Scale

Linkage of Genotype to Phenotype for Predictive Biology. Biological organisms are adapted to compete and survive in a vividly diverse range of environments. Despite major advances in genome sequencing and synthetic biology approaches enabling the synthesis of engineered pathways and genes, understanding of an organism's metabolic constituency remains poor and lacks a catalog of the full complement of enzymes encoded within biological organisms. Improved methods are needed to gain a better understanding of the metabolic pathways necessary to leverage advanced genetic engineering methods for designing organisms that synthesize bioproducts with improved or novel properties at synthesis rates compatible with industrial-scale production.

Better understanding of the metabolome of a diverse collection of organisms, along with a mechanistic understanding of the proteins that control metabolite biosynthesis, localization, degradation, and activity, would enable a predictive model of how an organism's total chemical constituency is determined. With this determination of genome sequence and transcriptomic and proteomic datasets, predicting how a cell's chemical environment may change with its environment or vary between genotypes will be possible. Also potentially within reach are predictions that an organism can synthesize a chemical that is never actually detected. Cataloging the genome sequences of diverse organisms may then enable a catalog of the predicted chemical diversity of life that preserves knowledge of biosynthetic pathways present in extinct or rare organisms.

Phenotyping and Profiling Natural Cells from the Environment at the Single-Cell Level. Cells are the functional unit of most biosystems, and their intracellular organization dictates how cellular phenotypes respond to variations in environmental conditions and genetic and chemical modifications.

To date, most research in single-cell biology has focused on mammalian cells because of their large size, accessibility, and relative ease of imaging. Understanding biology across scales and different environments, however, will require far more sensitive and higher-resolution technologies that also can analyze microbes and microbial communities. Needed are new approaches for handling extremely small samples while interfacing with different instruments to facilitate multiomics measurements. Because of the extremely small amounts of RNA present, proteins and metabolites likely to be recoverable from microbial cells essentially require lossless separation and measurement techniques. Automated, miniaturized, and microfluidic-based methods will be necessary to adapt current multiomics approaches to the single-cell level.

Proteins and metabolites pose a particular challenge because, unlike genomes and transcripts, they cannot be amplified. The large dynamic range of protein and metabolite abundances within the cell will require even more sensitive instrumentation or separation methods. MS-based imaging approaches also hold promise for gaining cellular-scale information on proteins and metabolites in a single cell in addition to morphological and contextual data.

In-Depth Spatiotemporal Understanding of Individual Cells. New methods for *in situ*, dynamic, or cryogenic imaging of whole organisms are needed to determine the highly heterogeneous organization of cells with nanoscale resolution to refine theory and models. Ultrastructural whole-cell imaging has improved significantly through the incorporation of phase plate technologies with electron tomography approaches. This advance has enabled high-resolution, 3D visualization of the cellular ultrastructure, organization, and protein localization of intact cells. With this improved resolution, difficult-to-purify protein complexes in their native cellular context are structurally resolvable using computational tomogram averaging, while dynamic events (e.g., host and phage interactions and phage lifecycle) can be mapped within the cell to reveal previously unknown interactions. The field of visual proteomics also has emerged in recent years and combines whole-cell electron tomography with computational template matching of known structures to visualize, identify, and quantify each cell component, a long-held dream of many biologists. New advances in structural biology, top-down and bottom-up proteomics, high-throughput cryoET, and integration with advanced informatics and high-performance computing (HPC) would further accelerate this field and increase its availability to researchers.

Heterogeneity and stochastic noise are intrinsic constraints on life. Biological processes such as transcription and translation are limited by fluctuations in the amount and localization of the molecules that drive them. Thus, cell regulatory systems have evolved to deal with intrinsic biological noise and sometimes to exploit it (e.g., bet hedging). Understanding biological regulation requires understanding the role and impact of stochastic variations, as well as how variation at the individual cell level translates into population behavior.

In situ measurements using advanced imaging technologies are likely to be crucial for understanding the spatial and compositional heterogeneity of microbial cells, populations, and communities. Although much effort has focused on improving the spatial resolution and chemical detection capabilities of imaging technologies, more attention must be given to their ability to discriminate between different biomolecules. Multispectral fluorescence *in situ* hybridization (FISH) has proven useful for identifying different microbial species in a community. Moreover, combinatorial barcoding of FISH probes, together with super-resolution and single-molecule fluorescence imaging, has shown promise for quantitative multigene expression analyses in intact cells and communities. However, these approaches generally are limited to thin samples with near-ideal optical properties. Improvements need to be made in the robustness of FISH techniques, number of different identifiable molecules, and their use with nonoptical microscopies (e.g., NanoSIMS). Also much needed are fixation and permeabilization methods compatible with downstream nucleic acid sequencing (e.g., on flow-sorted cells).

Multiplexed protein measurements for single cells are increasingly common using either lanthanum-labeled antibodies in conjunction with mass cytometry (i.e., CyTOF) or ion beam imaging (i.e., NanoSIMS). However, the limited availability of microbe-specific antibodies and difficulty of using them to label microbes have restricted this technique primarily to the detection of surface proteins. New approaches are required for multiplexed, *in situ* analysis of protein abundance in bacteria.

Current methods for examining gene expression heterogeneity at the single-cell level mostly use specific antibodies, multiplexed mRNA FISH techniques, gene-tagging approaches using CRISPR techniques, or RNA-Seq (e.g., combined with microfluidic cell separation or Drop-Seq). Although their adaptation for investigating plant and fungal systems is conceptually straightforward, their use in microbial systems is far more daunting. Microbes lack the DNA repair systems required for CRISPR tagging, and their lack of polyA tags and very low transcription burst frequency make single-cell RNA-Seq mostly uninformative (Choi et al. 2008), except in the case of bacteria large enough to maintain significant mRNA pools (e.g., cyanobacteria) or for a few abundant transcripts. Thus, new technologies are needed to determine the gene expression patterns of individual bacterial cells. Because protein abundance represents the integration of both gene expression and cellular phenotype, this parameter is likely to be the most accessible measurement in the near term for defining microbial cell heterogeneity. Needed are improvements in the sensitivity, specificity, and spatial resolution of both microbial proteomics and transcriptomics. These improvements should be developed in conjunction with new assays to measure cell phenotypic (e.g., metabolite) profiles.

Metabolic Pathway Engineering. Engineering metabolic pathways in microbes and plants has seen some success in the reconstruction of pathways necessary for drug synthesis (Lau et al. 2014; Lau and Sattely 2015). Introduction of the morphine pathway into yeast, for example, requires methods not only to express plant enzymes involved in the biosynthesis of this compound, but also to ensure correct localization of the proteins and transport of intermediate substrates between cell compartments (Thodey et al. 2014; Galanie et al. 2015). Thus, understanding how cells subcompartmentalize biosynthetic reactions between cells and cell subcompartments is needed to reconstruct or redesign such synthetic pathways. The ability to localize specific enzymatic steps into these discrete domains of an organism may prove useful in avoiding side reactions that reduce biosynthetic efficiency or lead to toxic byproducts. The ability to engineer alternate hosts with different metabolic capabilities and sensitivities will further expand the range of biosynthesis opportunities.

Cell-free expression platforms have emerged as powerful alternatives to laborious *in vivo* methods of protein synthesis that support the growing demand for easy and cost-effective protein production methods. Cell-free expression platforms permit expression of target proteins in a test tube in a few hours to days depending on protein size and requisite yields. Several cell-free protein expression systems are commercially available based on extracts from *Escherichia coli*, wheat germ, rabbit reticulocyte, *Leishmania tarentolae*, insects, or human cells. Although not currently commercially available, efficient cell-free lysates also can be prepared from tobacco BY-2 cells and Chinese hamster ovary (CHO) and yeast cells. Additional cell-free expression platforms from other source organisms, as well as more efficient pipelines for gene synthesis, expression screening, and purification, will help make these approaches readily accessible to a broader set of researchers.

Community Scale

Interrogation of Community Membership, Functional Potential, and Activity. Microbes play critically important roles in the environment, shaping plant health and productivity, the terrestrial carbon cycle, and environmental remediation. They occupy every biological niche and represent the largest portion of global biodiversity, with estimates indicating that the global protein universe of microorganisms exceeds 10^{12} including novel enzymes. However, present knowledge of microorganisms and their enzymes is based largely on laboratory studies of a small number of pure microbial cultures. The vast majority of microbes within environmental consortia cannot be cultured, leading to a lack of understanding for how microbial communities are structured and how they evolve (Solomon et al. 2014). The ability to identify the organisms involved in environmental processes has improved significantly with modern sequencing technologies, but obstacles such as polymerase chain reaction (PCR) bias and relic DNA result in an incomplete understanding of community behavior. Comprehensive characterization of environmental communities, including macro- and micro-eukaryotes, bacteria, archaea, and viruses, will require moving away from PCR-based methods and developing techniques to target metabolically active organisms regardless of phylogeny.

Microbiome research seeks to define community membership, ecological relationships between organisms, and the roles specific taxa play in system-level chemical and biological processes. Regardless of their habitat, microbiomes comprise many different taxa exploiting an energy source, yet these microbial assemblages often are inherently interdependent and dynamic in both space and time. The advent of high-throughput sequencing, advanced microscopy, and stable isotope tracing techniques has made comprehensive understanding of *in situ* microbiome ecology tantalizingly possible. However, current microbiome studies often are highly descriptive, focused on correlation patterns or simple one-on-one interactions between culturable organisms. A more mechanistic, spatially and temporally resolved understanding of plant-microbe and microbe-microbe interactions is needed to enable sustainable biofuel feedstock growth, increase food security, increase terrestrial carbon sequestration, limit contaminant transport, and enable engineering of microbial traits important to plant communities.

While genomic, transcriptomic, metabolomic, and proteomic methods have advanced considerably, they still are most often applied to single-microbe, cultured systems. Results from these data streams need to be more effectively integrated, and their predictions verified via manipulative experiments, metabolic modeling, and quantitative observations of metabolic fluxes within and between organisms and via imaging or (bio)chemical characterization. Current knowledge of plant metabolites is derived primarily from highly simplified hydroponic systems, but plants naturally grow in a soil matrix where metabolites of interest are likely intercepted by the surrounding microbiome before they can be collected. Thus, new methods are needed for collecting plant exudates and metabolites under realistic conditions.

Genetic Tools for Recalcitrant Organisms. The number of model organisms for which robust genetic tools have been developed is relatively small compared to the number of organisms whose genomes have been sequenced. Although various methods have been developed for delivering DNA into cells, the ability to precisely rewrite sections of a nonmodel organism's genome is only now undergoing rapid development. Advances in this area have been greatly accelerated by the development of CRISPR–CRISPR-associated protein-9 nuclease (Cas9) as a genome editing tool, including approaches (Cong et al. 2013; Qi et al. 2013) targeted to both prokaryotic and eukaryotic systems, now enabling genetic modifications of an expanded set of organisms. However, additional barriers mitigate the effectiveness of even CRISPR systems, including but not limited to cell wall thickness, recombination, and transformation efficiency. Further, these strategies must be developed for systems for which no draft genome is available and for polyploid cells that limit the effectiveness of genetic modification. Ideally, these tools could target individual members of natural consortia for either total ablation or specific genetic manipulation.

Biotic and Abiotic Interactions. The ability to conduct *in situ* interrogations of live microbial community systems, including plant-associated microbes, is critical to understanding carbon-nitrogen and carbon-sulfur cycles in simulated and natural ecosystems, as well as the community roles of individual taxa on community structure and function by sequential identification of reduced gases [hydrogen sulfide (H_2S), nitrous oxide (N_2O), and methane (CH_4)] and shifts in carbon signatures. Measurements of H_2S, sulfur dioxide (SO_2), N_2O, nitric oxide (NO), nitrogen dioxide (NO_2), CH_4, carbon dioxide (CO_2), and other important gases, while characterizing changing pools of dissolved organic carbon, may be accomplished via *in situ* ^1H, ^{13}C, and ^{31}P NMR for microbial communities. Development of environmentally controlled tandem NMR with infrared, gas analyzer probes, and gas chromatography–MS for *in situ* interrogation of live microbial community systems would enable these measurements, resulting in significant progress toward more comprehensive understanding of microbial metabolism coupled to environmental conditions.

Biological systems are constantly changing and adapting to their environment, but most experimental approaches are limited to snapshot-in-time surveys of the system. To fully understand biological system dynamics, sampling must occur over time and space. Stable isotope probing technologies enable this type of temporal sampling. Following labeled molecules over time through space provides understanding of not only the state of the system, but also its nutrient flux, metabolite exchange, and reaction rates. Many technologies exist to measure isotopic enrichment, but improvement in sensitivity and selectivity is necessary to increase the granularity of the studies. Importantly, most isotopic analyses are not high throughput, limiting the application of this approach. A great advantage of isotopic analyses is the ability to probe from the cell to the system. For this impact to be realized, new approaches for data integration and modeling are necessary to combine the data from varied spatial scales into a single biological model.

Laboratory Scale

Model Ecosystems. The interaction of plants with their environment is one of the most complex and highly evolved systems on Earth. This interface or "interaction zone" between biotic and abiotic processes marks the heart of biogeochemistry. Microbes and organic molecules in the belowground environment rarely are randomly distributed; rather, they preferentially interact with specific mineral phases that have varying degrees of crystallinity, mineral size fractions, and redox states. Microbes are known to harvest elemental nutrients and even electrons from their mineral substrates, and organic molecules preferentially associate with mineral phases because of their surface charge at a given pH or the presence of specific functional groups and ligand chemistry. Abiotic parameters such as pH, cation exchange capacity, redox, and water saturation also are heavily influenced by biotic factors during this interplay, resulting in heterogeneity across broad spatial and temporal scales and giving rise to biogeochemical "hot spots" and "hot moments." This dynamic and heterogeneous nature of ecosystems poses major challenges to reproducible scientific investigations in controlled, yet field-relevant, conditions.

One potential means of making laboratory analysis of root growth more relevant to the field is using a transparent medium with actual soil properties. The polymer Nafion is transparent, with a refractive index close to that of water, and is claimed to have physical properties similar to those in soil (Downie et al. 2012). Nafion's primary use as a soil substitute, to date, is to observe root-microbe interactions (Downie et al. 2014).

Additionally, phytotrons (i.e., facilities with plant growth chambers that enable detailed control of environmental conditions such as light, humidity, atmospheric chemistry, sterility, temperature, and pressure) can provide an experimental platform for (1) growing the large number of plant systems required for genome-to-phenome mapping (i.e., from microbes to atmospheric flux); (2) creating replicated experiments; and (3) providing access to identical systems for destructive sampling and monitoring of metabolic function, as well as access to the rhizosphere for microbe and root exudate studies. However, phytotrons lack the inherent complexity of the field conditions where plant communities; mesofauna; and natural soil chemical, mineralogical, and textural heterogeneity interplay with natural hydraulic conditions to create unique spatial and temporal domains from nanometer and microsecond to meter and annual scales. In contrast, ecotrons, in which several cubic meters of natural ecosystems are transported into massive environment-controlled growth chambers, preserve some elements of field conditions and multispecies plant-mesofauna-microbe communities with natural soils permitting minimally disturbed textures and hydrology. Ecotrons provide many of the benefits of phytotron studies but with much greater complexity approaching that of field conditions. Both phytotrons and ecotrons can provide testbeds for laboratory-to-field technology development and sensor calibration.

Laboratory experiments also are crucial for rigorous testing and verification of atmospheric science hypotheses under controlled and systematic conditions. The CLOUD and AIDA chambers in Europe have demonstrated the utility of large, collaborative laboratory experiments that simulate and unravel important aerosol and cloud processes. Establishing a similar facility in the United States could lead to a transformative leap in the understanding of aerosol and cloud formation.

Ecosystem Scale

***In Situ* Measurement Capabilities.** Environmental system function is intimately tied to the biogeochemical cycling of major elements. Improved mechanistic understanding is needed, across multiple levels of spatial and temporal resolution, of the biogeochemical processes that control the transformations and mobility of carbon, nutrients, and contaminants within environmental systems and ecosystems critical to DOE. This improved understanding will enable prediction of environmental system response and, potentially, engineered controls of these systems. In particular, BER seeks to understand the controls that biogeochemical processes exert on nutrient and carbon cycling, sustainable biofuel production, and contaminant mobility from the molecular (Ångström or nanometer) to mesoscale (millimeter or meter), as well as the nonlinear scaling laws that control the effect of these smaller length–scale processes on larger length–scale processes, which, ultimately, affect ecosystem, regional, and Earth system phenomena. There also is a need to illuminate and quantify biogeochemical interactions across multiple spatial scales, from nanometer- to micron-scale cellular metabolism and mineralogy to millimeter- and meter-scale community biogeochemical activity, to ecosystem-scale processes. These investigations necessitate quantified measurements, from nanometer- to meter-length scales, of soil carbon and nutrient flow, related biogeochemical processes, and multiomic characterization of microbial dynamics, along with integration of these data into predictive microbial metabolic and systems-ecology models.

Imaging technologies need to be developed and improved with spatial resolutions from the nanometer- to millimeter-length scales and subsecond temporal resolutions to study biogeochemical processes within environmental systems. These approaches will require a variety of probes with spatial and temporal resolutions relevant to the biogeochemical system being studied. These new and improved imaging technologies must enable researchers to understand and, ultimately, predict the dynamic interplay among the biotic and abiotic components of the environment from the molecular to mesoscale. Moreover, they must be able to use this new understanding to predict regional- and global-scale phenomena.

Hot spots and hot moments occur at a range of scales in response to widely ranging perturbations—such as microbial metabolomic responses to diurnal cycling of root exudation or dissolved organic matter (DOM) inputs to riverine systems from wildfires. Understanding ecosystem responses to large perturbations requires rapid deployment of an array of field measurements. Future possibilities may include not only gathering important samples for later measurement in the laboratory, but also developing the next generation of experimental platforms to enable making those same measurements in the field. Whether using high-resolution mass spectrometers to measure DOM chemistry or an array of oxygen optodes to measure mixing of groundwater and surface waters, there is a need to capture the spatial and temporal variations in ecosystem responses to perturbations that are enabled only by extensive in-the-field measurements.

Environmental Observation. *In situ* observation of human health biomarkers (e.g., 24/7 monitoring of blood glucose levels) is now a reality in biomedicine. Analogous developments are envisioned in ecosystem science, enabling constant monitoring of plant metabolites in leaves and roots, organic matter composition in porewater, or quorum-sensing molecules in lakes. One can imagine, for instance, an array of radiofrequency tags installed across field sites reporting the onset of nutrient limitation or water stress, changes in redox state, or even the occurrence of disease or pest infestation.

Remote-sensing technologies also will play a key role in ecosystem monitoring, and the development of drone-deployable instruments for such applications will be a key direction for future work. These instruments will enable interrogations of both natural and managed environments on a larger scale, with implications for environmental and agricultural research. Such technology includes microsatellites, whose mass ranges from 10 to 100 kilograms (kg) (SpaceWorks 2017). They have become popular in recent years because of their affordability and potential capability for being launched as arrays of microsatellites with a single rocket (Selva and

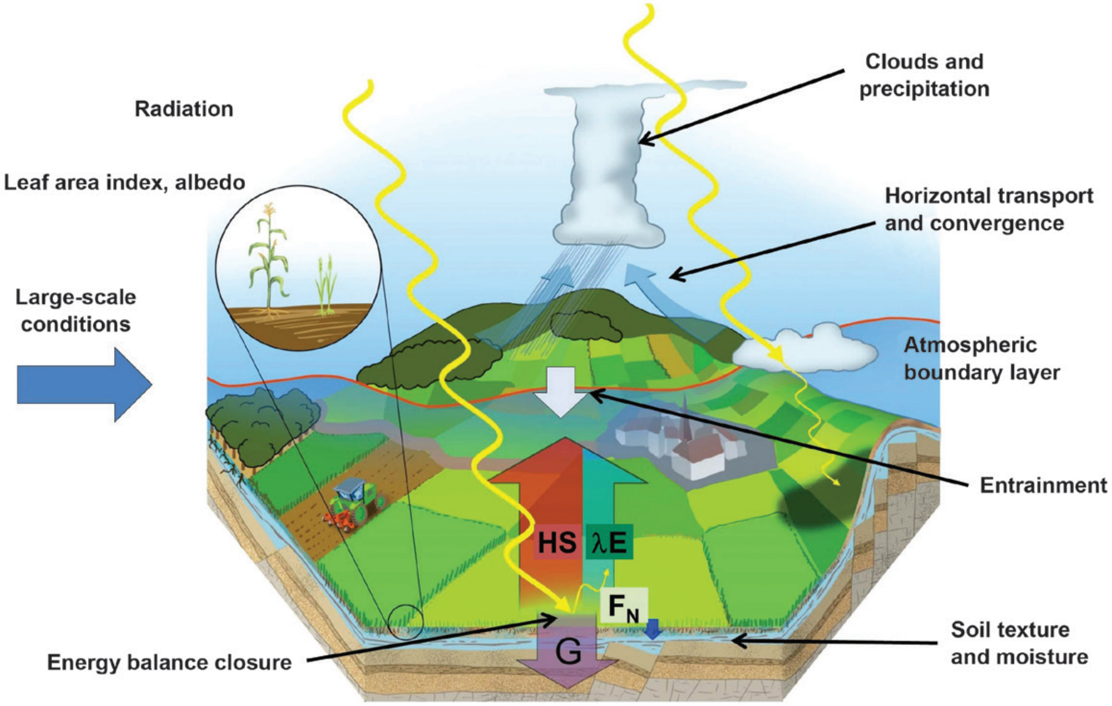


Fig. 8.1. Model Components of Land-Atmosphere Interactions. The evolution of the atmospheric boundary layer and convection initiation are critically dependent on land surface–atmosphere exchanges, including net radiation (F_N), sensible heat flux (HS), latent heat flux (λE), and ground heat flux at the land surface (G). [Reprinted under a Creative Commons Attribution License (CC BY NC-ND) from Wulfmeyer, V., et al. 2015. "A Review of the Remote Sensing of Lower Tropospheric Thermodynamic Profiles and Its Indispensable Role for the Understanding and the Simulation of Water and Energy Cycles," *Reviews of Geophysics* **53,** 819–95. DOI: 10.1002/2014RG000476. © 2015 Wulfmeyer, Hardesty, Turner, Behrendt, Cadeddu, Di Girolamo, Schlüssel, Van Baelen, and Zus]

Krejci 2012). Microsatellites are increasingly used for Earth observation and remote sensing, making them potentially useful for many BER mission areas. Multiple microsatellites with Earth imaging equipment are already in orbit (Belward and Skøien 2015). With projected launch costs ranging from $14,000 to $60,000 per kg, the number of microsatellites launched per year is expected to more than triple between 2016 and 2023 (SpaceWorks 2017). Most of this growth is expected in the Earth observation area (SpaceWorks 2017), representing a unique opportunity for BER science to grow along with these new data streams. Collaboration between BER-funded scientists and the microsatellite development and launch community will be key for making this opportunity a reality.

Planet Scale

Multiscale Process-Level Understanding and Integration into Earth System Models.

Process-level understanding of aerosols, clouds, turbulence, atmospheric convective systems, and their interactions (see Fig. 8.1. Model Components of Land-Atmosphere Interactions, this page) is needed to advance knowledge and modeling of water, carbon, and energy cycles of the Earth system. This understanding also is needed to better predict future climate change and extreme atmospheric events that can significantly impact energy infrastructure. Such understanding could be gleaned from colocated, 3D measurements of the distributions of atmospheric aerosols, clouds, and coupled thermodynamic and dynamical fields, as well as larger-scale surface conditions covering meters for cloud-aerosol interactions to 100 kilometers (km) for mesoscale convective systems (MCSs; see Fig. 8.2a. Coupling of Multiscale Cloud Systems with Atmospheric Circulations and Turbulent Flows, p. 111). Currently lacking are such collocated 3D measurements that potentially could be obtained by several emerging technologies, including photoarray sensors, drones, and high-altitude balloons.

Most clouds are formed in turbulent flows in the presence of aerosols, and they strongly affect the Earth system's radiative energy budget. Clouds, in turn, impact turbulent flows through their radiative impact and latent heating and cooling through the phase change of water. Turbulence is strongly modulated by environmental conditions in the atmosphere and at the surface. Aerosol-cloud-precipitation-turbulence interactions occur at spatial scales ranging from meters at which cloud particles are nucleated, to several hundred meters at which shallow cumulus clouds form, to several kilometers at which cumulus congestus and isolated deep convection occur, to tens of kilometers at which MCSs are organized, to hundreds of kilometers where MCSs are embedded (see Fig. 8.2b., p. 111). Cloud microphysics at the meter scale are known to have a significant impact on MCSs; simulating aerosol-cloud interactions and predicting high-impact atmospheric systems, therefore, require improved measurements and understanding across scales.

The scanning cloud radars used at DOE's ARM Research Facility scan a plane slice of clouds in about 10 minutes. The temporal and spatial resolution of this sampling is not sufficient for capturing the 3D structural and temporal evolution of clouds. Phased-array Doppler radars, however, can take volume scans to significantly improve the sampling of cloud and precipitation systems. These radars have been used to study severe storms (Yoshida et al. 2017). The phased-array radars should include an S-band to measure precipitation particles and an X-band to measure cloud particles to cover domain sizes comparable to the S-band radar coverage.

DOE recently used UASs built by the University of Colorado to take measurements of atmospheric pressure, temperature, humidity, and wind speed at low altitudes under limited meteorological conditions. Also built was a fixed-wing unmanned aerial vehicle with a 22-foot wingspan (ArcticShark) that can reach elevations of 3 to 6 km to take atmospheric measurements under limited meteorological conditions. The UAS capabilities should be significantly enhanced in four aspects to (1) enable measurement operations under all meteorological conditions; (2) provide coordinated measurements among multiple unmanned control systems and ground radars that can map 3D atmospheric motions and thermodynamics within and outside clouds; (3) use miniaturized instruments for measuring aerosols and cloud properties; and (4) enhance surface flux measurements of energy and water, as well as measurements of ecosystem and land surface properties.

DOE recently used tethered balloons at Oliktok Point in Alaska to measure atmospheric thermodynamic states and winds under limited meteorological conditions with altitudes up to 2 km above ground. This capability potentially could be significantly expanded by high-altitude, motor-propelled, solar-powered geostationary balloons in the stratosphere with profiling instruments attached to tethers. Several manufacturers make such balloons (e.g., TCOM, www.tcomlp.com; Raven Aerostar, ravenaerostar.com). Their use for tropospheric measurements of atmospheric moist processes is yet to be explored. Relative to the infrequent launch of radiosondes on weather balloons from the ground, these high-altitude balloons can provide continuous vertical profile sampling of atmospheric winds, temperature, and humidity, along with aerosol and cloud microphysical properties.

Modeling and Engineering

Modeling Single Cells. All mathematical models of cells are essentially models of single cells, with the summation of their range of behaviors giving rise to the population-level phenotype. Thus, a deep

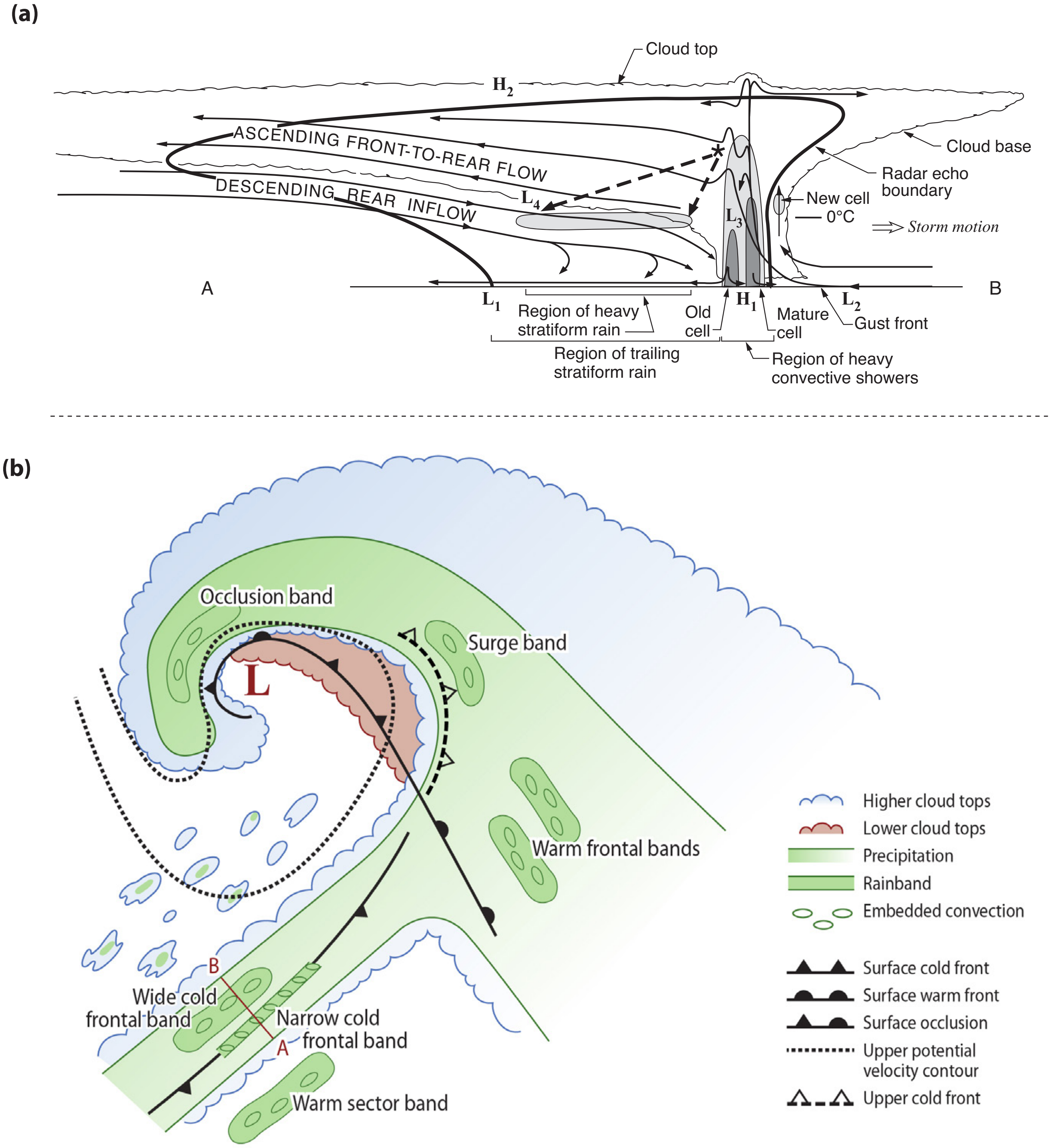


Fig. 8.2. Coupling of Multiscale Cloud Systems with Atmospheric Circulations and Turbulent Flows. (a) Conceptual model of a vertical cross section of a mesoscale convective system (labeled as A to B in panel b). **(b)** Conceptual model of the horizontal structure of a frontal system. [**(a)** Adapted and used with permission from Houze, R. A., Jr., et al. 1989. "Interpretation of Doppler Weather Radar Displays of Midlatitude Mesoscale Convective Systems," *Bulletin of the American Meteorological Society* **70**(6), 608–19. © American Meteorological Society **(b)** Adapted from Houze, R. A., Jr. 2014. "Clouds and Precipitation in Extratropical Cyclones." In: *Cloud Dynamics, Second Edition* **104,** 496 pp., with permission from Elsevier. © 2014 Academic Press]

analysis of cell heterogeneity is required for building predictive models of cell communities. A particular challenge in building mathematical models of microbial cells, however, is the constraint of limiting ribosomal pools and other biosynthetic resources, especially at the low microbial growth rates typical of natural environments. This limitation can result in a pattern of competing reactions that changes with alterations in gene expression profiles. Understanding how microbes regulate biosynthetic resources within a dynamic environment and incorporating this knowledge into whole-cell models will be key to re-engineering their metabolic machinery without compromising their viability. Also important is moving metabolic models away from simple flux-balance models to more biophysically based models that incorporate protein abundance, feedback, and thermodynamics. This effort will provide a foundation for predicting the impact of alterations in gene expression or protein abundance on the cell's physiological state.

Modeling Microbial Communities. Carbon dynamics within soil ecosystems are dominated by a web of biotic transformations mediated by photosynthetic and heterotrophic microbes and metazoans. Recognition of these dynamics has enabled consumer-resource models of food webs to (1) predict the quantitative effects of species removal on the remaining taxa within intertidal food webs (Berlow et al. 2009) and (2) forecast seasonal changes in microbial and metazoan carbon dynamics in a complex aquatic ecosystem (Boit et al. 2012). However, the cryptic nature of soil ecosystems has prevented understanding of soil food webs from developing as rapidly as that of more easily observed aquatic systems (Brose and Scheu 2014). With the advent of more powerful omics and isotopic techniques, the ability to more accurately model soil ecosystem carbon dynamics is rapidly increasing (Jakobsen and Rosendahl 1990; Digel et al. 2014). For example, unique structural aspects of complex soil food webs have been identified (Riede et al. 2010), and the effects of temperature and moisture on the dynamics of simple soil food chains have been explored both theoretically (Binzer et al. 2012) and experimentally (Lang et al. 2012; Lang et al. 2014). The next challenge is to integrate simple components of soil food webs into more realistic and predictive networks by resolving resource flow, incorporating key additional processes, and integrating the ecological and biogeochemical roles of understudied groups such as viruses and microfauna (see also Grand Challenge 4.1, p. 46).

Modeling Interactions. NWChem software contains scalable implementations of several classes of methodologies necessary to describe processes at various molecular scales. For example, computational protocols employing the molecular dynamics module of NWChem are used routinely in studies of small-molecule binding, protein-protein interactions, and processes occurring at protein-mineral interfaces. NWChem's capabilities also have been used in high-throughput calculations for metabolomics and NMR. There is an urgent need for continuing this effort and extending NWChem's capabilities across spatial and temporal scales. Among the most pressing needs are (1) development of methods of lower complexity for mesoscale systems (e.g., semiempirical methods, classical density functional theory, and embedding methods) and (2) development of HPC implementations of near-infinite run times, quantum mechanics (QM), and QM/molecular mechanics methods to model biosystems at unprecedented scale. Theoretical methods and algorithms for extending time scales of dynamics simulations, complex dynamics for describing biomolecule-light interactions to support FEL experiments, methods for simulating long-range transport in biological systems, and algorithms for modeling the thermodynamics of metabolism are also necessary components of future computational frameworks needed to understand biological systems.

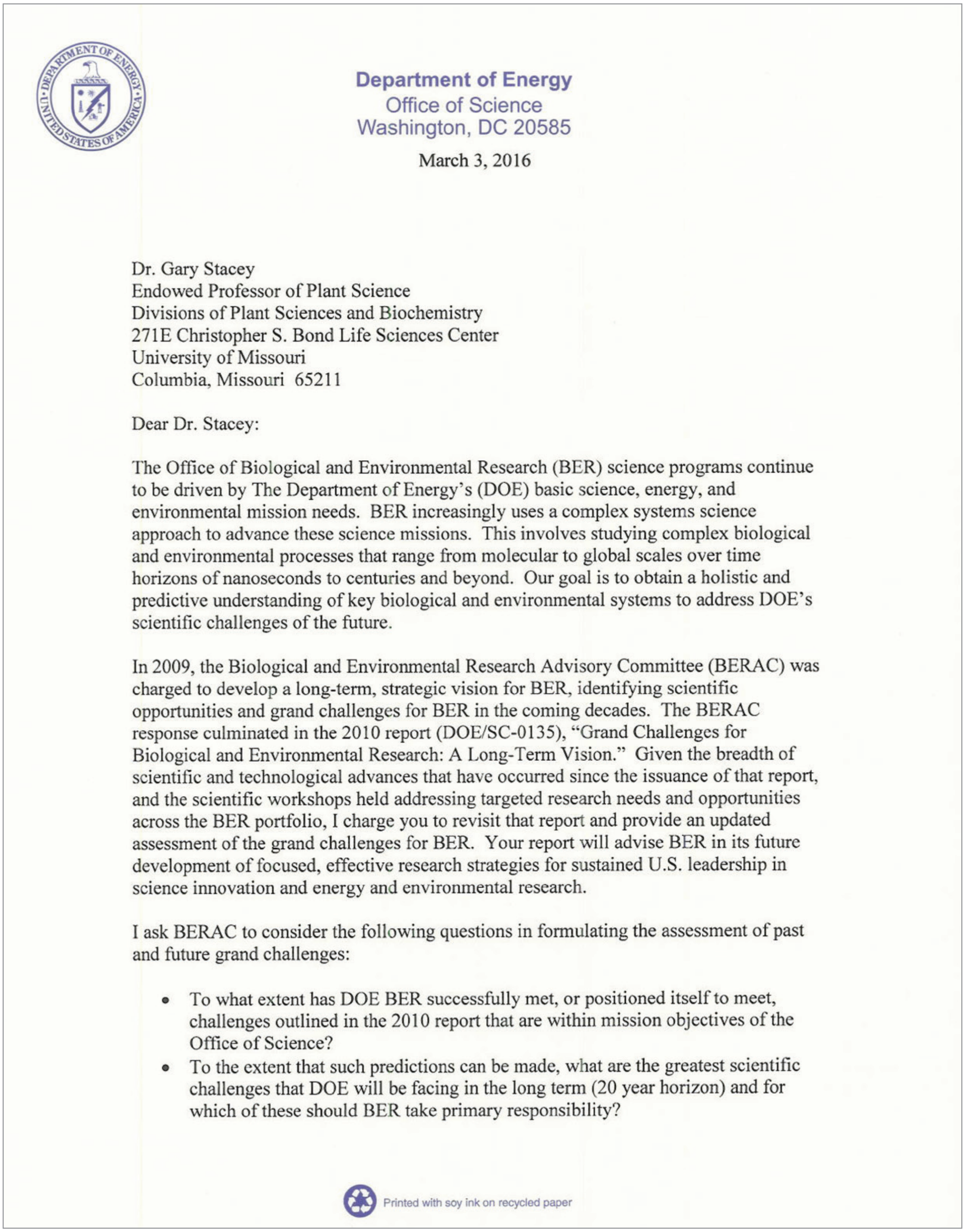
Improved Transport Modeling. Computational methods are critically needed to simulate various aspects of proton-controlled transport in biomass cellular materials, which are tightly coupled with cellular activities in response to biotic and abiotic stresses. Currently, proton-controlled transport simulations can only be performed using standard force fields that lack a fundamental description of proton-transfer processes. Truly predictive modeling of transport processes requires the use of accurate QM methods, adaptive force fields and molecular dynamics, and long-time conformational sampling at time scales of milliseconds.

Biosystems Design and Engineering. Advances in high-throughput biology and biotechnology have led to an array of biological insights in diverse organisms, the environment, medicine, agriculture, and energy supply. The breadth of genomic diversity endows organisms with rich biosynthetic potential and allows them to adapt to diverse environments (Venter et al. 2004; Tringe et al. 2005). Harnessing the potential of species diversity makes biological systems ideal for solving global challenges, such as generating biologically derived chemicals, fuels, and materials to ensure environmental sustainability (Way et al. 2014). Achieving these goals, however, requires not only a thorough understanding of biological systems, but also their safe and programmable control. The ability to modify genetic and biochemical molecules in cells, apply advanced computation to design and analyze engineered biosystems, and isolate engineered cells and communities with desired function remain defining challenges. These advances in biosystems design undoubtedly will play an important role in enabling a predictive understanding of biological systems relevant to energy supply, production of biologically derived molecules, and other DOE mission goals.

Biosystems design may provide a valuable approach for probing, studying, and introducing new functions into biological systems, and it is being fueled by concurrent advances in systems biology and the emerging field of synthetic biology. Synthetic biology combines principles from biology, chemistry, physics, mathematics, and engineering to assemble the biological tools necessary to redesign biological systems. More specifically, synthetic biology employs engineering principles to reduce genetics into DNA “parts,” so that those parts can be understood in isolation and reassembled into new biological parts, devices, and whole systems to build desired or expanded functions in living cells (see also sidebar, Synthetic Biology, p. 5). In many respects, synthetic organic chemistry serves as a model for the nascent field of synthetic biology. This new paradigm for biology is one in which a desired biological function is conceived, designed, and constructed to work as predicted, reliably and robustly using well-defined parts. Given the abundance of recent advances, synthetic biology is now poised both to make fundamental breakthroughs in the understanding of complex biological systems and to actualize an array of impactful applications that address global challenges in food and energy supply, environmental health, and medicine. New tools are required for the application of biosystems design methods to nonmodel organisms key to DOE mission areas, including unusual microbes, environmental isolates, algae, and plants, as well as microbial communities and plant-microbe interfaces. Biosystems design is now capable of producing up to billions of genotypic variants, leading to the need for improved methods to explore extensive phenotype landscapes and link genotype to phenotype. Because the safety of engineered biosystems is a concern, methodologies focusing on biocontainment and genetic isolation also constitute an important need.

Appendices

Appendix A: Charge Letter



Department of Energy
Office of Science
Washington, DC 20585

March 3, 2016

Dr. Gary Stacey
Endowed Professor of Plant Science
Divisions of Plant Sciences and Biochemistry
271E Christopher S. Bond Life Sciences Center
University of Missouri
Columbia, Missouri 65211

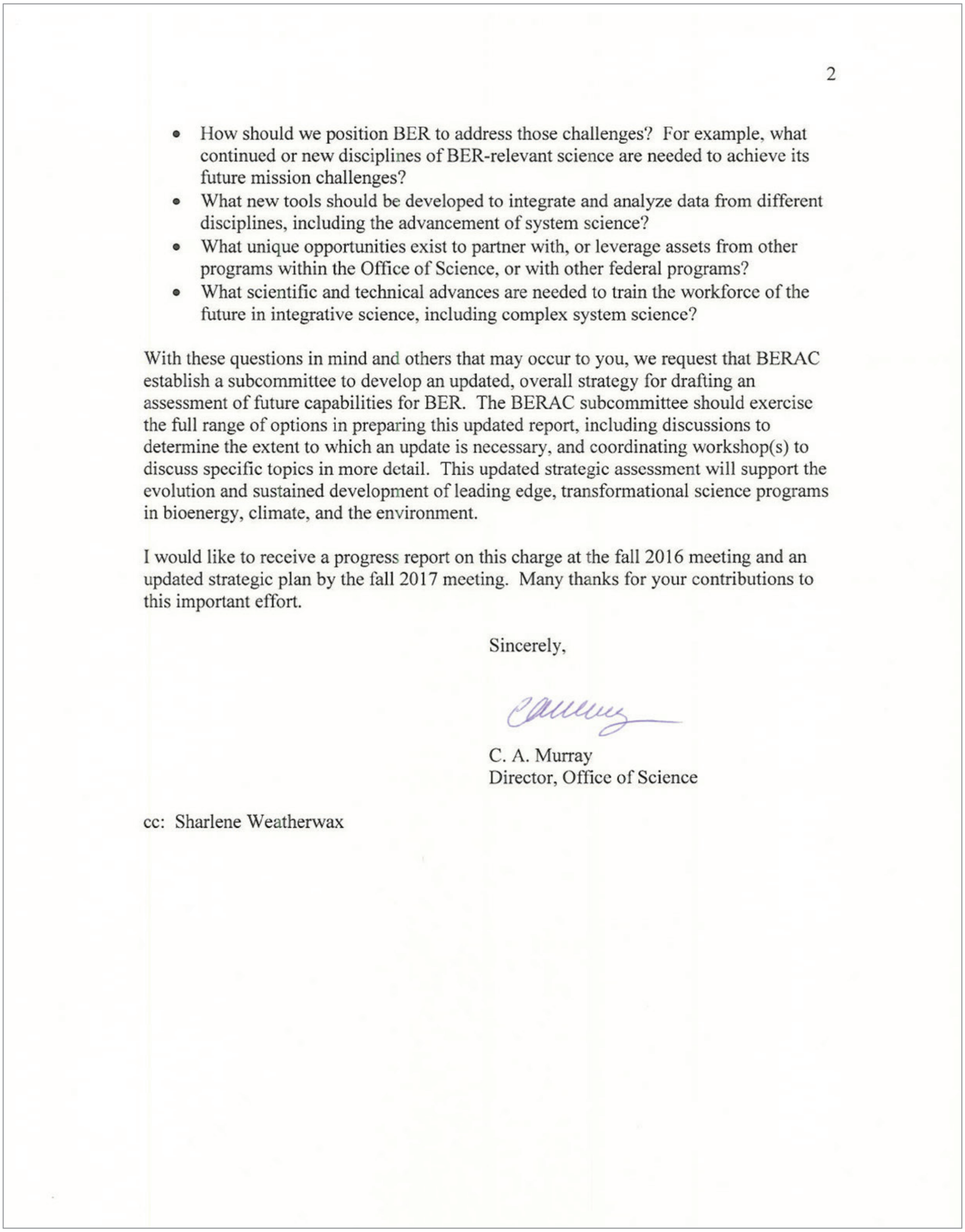
Dear Dr. Stacey:

The Office of Biological and Environmental Research (BER) science programs continue to be driven by The Department of Energy's (DOE) basic science, energy, and environmental mission needs. BER increasingly uses a complex systems science approach to advance these science missions. This involves studying complex biological and environmental processes that range from molecular to global scales over time horizons of nanoseconds to centuries and beyond. Our goal is to obtain a holistic and predictive understanding of key biological and environmental systems to address DOE's scientific challenges of the future.

In 2009, the Biological and Environmental Research Advisory Committee (BERAC) was charged to develop a long-term, strategic vision for BER, identifying scientific opportunities and grand challenges for BER in the coming decades. The BERAC response culminated in the 2010 report (DOE/SC-0135), "Grand Challenges for Biological and Environmental Research: A Long-Term Vision." Given the breadth of scientific and technological advances that have occurred since the issuance of that report, and the scientific workshops held addressing targeted research needs and opportunities across the BER portfolio, I charge you to revisit that report and provide an updated assessment of the grand challenges for BER. Your report will advise BER in its future development of focused, effective research strategies for sustained U.S. leadership in science innovation and energy and environmental research.

I ask BERAC to consider the following questions in formulating the assessment of past and future grand challenges:

- To what extent has DOE BER successfully met, or positioned itself to meet, challenges outlined in the 2010 report that are within mission objectives of the Office of Science?
- To the extent that such predictions can be made, what are the greatest scientific challenges that DOE will be facing in the long term (20 year horizon) and for which of these should BER take primary responsibility?



- How should we position BER to address those challenges? For example, what continued or new disciplines of BER-relevant science are needed to achieve its future mission challenges?
- What new tools should be developed to integrate and analyze data from different disciplines, including the advancement of system science?
- What unique opportunities exist to partner with, or leverage assets from other programs within the Office of Science, or with other federal programs?
- What scientific and technical advances are needed to train the workforce of the future in integrative science, including complex system science?

With these questions in mind and others that may occur to you, we request that BERAC establish a subcommittee to develop an updated, overall strategy for drafting an assessment of future capabilities for BER. The BERAC subcommittee should exercise the full range of options in preparing this updated report, including discussions to determine the extent to which an update is necessary, and coordinating workshop(s) to discuss specific topics in more detail. This updated strategic assessment will support the evolution and sustained development of leading edge, transformational science programs in bioenergy, climate, and the environment.

I would like to receive a progress report on this charge at the fall 2016 meeting and an updated strategic plan by the fall 2017 meeting. Many thanks for your contributions to this important effort.

Sincerely,

C. A. Murray
Director, Office of Science

cc: Sharlene Weatherwax

Appendix B: Grand Challenges Workshop Agenda

Biological and Environmental Research Advisory Committee (BERAC) Grand Challenges II Workshop 2017

Hilton Washington DC/Rockville Hotel
1750 Rockville Pike, Rockville, Maryland

Monday, March 6

8:30 a.m.–9:00 a.m.	**Stephen Binkley**, Acting Director, U.S. Department of Energy (DOE) Office of Science: Overview of DOE Office of Science
9:00 a.m.–9:30 a.m.	**Sharlene Weatherwax,** Associate Director, DOE Office of Biological and Environmental Research; and **Gary Stacey,** BERAC Chair: Description of Meeting Goals and Expected Outcomes
9:30 a.m.–10:00 a.m.	**George Church:** Synthetic Biology [Eisenhower Room]
10:00 a.m.–10:30 a.m.	**Inez Fung:** Earth System Feedbacks [Eisenhower Room]
10:30 a.m.–11:00 a.m.	**Richard Phillips:** Biotic Interactions [Eisenhower Room]
11:00 a.m.–11:15 a.m.	Break
11:15 a.m.–12:30 p.m.	**Breakout Groups I: Discussion**
	1. Earth System Research [Eisenhower Room]
	2. Systems Biology Research [Monroe Room]
	3. Microbial to Earth System Pathways [Jackson Room]
12:30 p.m.–1:30 p.m.	Lunch (on site)
1:30 p.m.–3:30 p.m.	**Breakout Groups I: Discussion and Synthesis**
	1. Earth System Research [Eisenhower Room]
	2. Systems Biology Research [Monroe Room]
	3. Microbial to Earth System Pathways [Jackson Room]
3:30 p.m.–3:45 p.m.	Break
3:45 p.m.–5:00 p.m.	**Report from Breakouts; Compare and Contrast Among Focus Groups**
5:30 p.m.	Adjourn

Tuesday, March 7

9:00 a.m.–9:30 a.m.	**Steven Davis:** Energy Sustainability [Eisenhower Room]
9:30 a.m.–10:00 a.m.	**Shibu Yooseph:** Big Data and Computing [Eisenhower Room]
10:00 a.m.–10:30 a.m.	**Cullen Buie:** Micro- and Nanoengineering [Eisenhower Room]
10:30 a.m.–10:45 a.m.	Break
10:45 a.m.–12:30 p.m.	**Breakout Groups II: Discussion** 1. Energy Sustainability Across Relevant Scales [Monroe Room] 2. Needs and Applications of Big Data and Computing [Jackson Room] 3. Identifying Applications of Emerging Technologies in BER-Relevant Research (e.g., Nanotechnology) [Eisenhower Room]
12:30 p.m.–1:30 p.m.	Lunch (on site)
1:30 p.m.–3:30 p.m.	**Breakout Groups II: Discussion and Synthesis** 1. Energy Sustainability Across Relevant Scales [Monroe Room] 2. Needs and Applications of Big Data and Computing [Jackson Room] 3. Maximizing User Facilities and Experimental Sites [Eisenhower Room]
3:30 p.m.–3:45 p.m.	Break
3:45 p.m.–5:00 p.m.	**Reports from Breakouts; Compare and Contrast Among Focus Groups**
5:30 p.m.	Adjourn

WRITERS' AGENDA
(Selected Writing Team Only)

Tuesday Evening, March 7

Writers' Dinner: Buffet at Rockville Hilton

Discuss Wednesday schedule and writing plan. Review respective breakout reports, meeting notes, and discussion notes to prepare for writing portion.

Wednesday, March 8

9:00 a.m.–10:00 a.m.	Produce outlines for each section.
10:00 a.m.–11:00 a.m.	Compile all outlines and edit to complete report outline. [Eisenhower Room]
11:00 a.m.–12:00 p.m.	Return to groups for completing outline and writing.
12:00 p.m.–1:00 p.m.	Lunch (on site)
1:00 p.m.–4:00 p.m.	Continue writing. Complete full outline, introduction, upcoming challenges, and recommendations for implementation.

Appendix C: Grand Challenges Workshop Participants

Steven Allison
University of California, Irvine

Sarah Assmann
Pennsylvania State University

Inês Azevedo
Carnegie Mellon University

Dennis Baldocchi
University of California, Berkeley

Mark Bradford
Yale University

Amy Brunner
Virginia Polytechnic Institute and State University

Cullen Buie
Massachusetts Institute of Technology

Katherine Calvin
Joint Global Change Research Institute (Pacific Northwest National Laboratory and University of Maryland)

Patrick Chain
Los Alamos National Laboratory

George Church
Harvard University

Elsa Cleland
University of California, San Diego

Steven Davis
University of California, Irvine

Scott Denning
Colorado State University

Adam Deutschbauer
Lawrence Berkeley National Laboratory

Wei Ding
University of Massachusetts, Boston

Mary Dunlop
Boston University

Jim Ehleringer
University of Utah

Graham Feingold
National Oceanic and Atmospheric Administration

Mary Firestone
University of California, Berkeley

Ian Foster
Argonne National Laboratory

Inez Fung
University of California, Berkeley

James Hack
Oak Ridge National Laboratory

Alex Held
Commonwealth Scientific and Industrial Research Organisation (Australia)

Kirsten Hofmockel
Pacific Northwest National Laboratory

Susan Hubbard
Lawrence Berkeley National Laboratory

Bruce Hungate
Northern Arizona University

Charles Jackson
University of Texas

Anthony Janetos
Boston University

Andrzej Joachimiak
Argonne National Laboratory

Ravi Kane
Georgia Institute of Technology

Kerstin Kleese van Dam
Brookhaven National Laboratory

Cheryl Kuske
Los Alamos National Laboratory

Paul Langan
Oak Ridge National Laboratory

Ruby Leung
Pacific Northwest National Laboratory

Jonathan Lynch
University of California, Davis

Costas Maranas
Pennsylvania State University

Christopher Marx
University of Idaho

Sean McSweeney
Brookhaven National Laboratory

Gerald Meehl
National Center for Atmospheric Research

Claire Monteleoni
The George Washington University

Gloria Muday
Wake Forest University

Michelle O'Malley
University of California, Santa Barbara

Victoria Orphan
California Institute of Technology

Jennifer Pett-Ridge
Lawrence Livermore National Laboratory

Richard Phillips
Indiana University Bloomington

Benjamin Preston
Rand Corporation

David Randall
Colorado State University

James Randerson
University of California, Irvine

Patrick Reed
Cornell University

Karin Remington
Computationality, LLC

G. Philip Robertson
Michigan State University

Ramalingam Saravanan
Texas A&M University

Karen Schlauch
University of Nevada, Reno

C. Adam Schlosser
Massachusetts Institute of Technology

Daniel Segrè
Boston University

John Shanklin
Brookhaven National Laboratory

Zengyi Shao
Iowa State University

Gary Stacey
University of Missouri

David Stahl
University of Washington

Dáithí Stone
Lawrence Berkeley National Laboratory

Joao Teixeira
National Aeronautics and Space Administration
Jet Propulsion Laboratory

Kathleen Treseder
University of California, Irvine

Susannah Tringe
U.S. Department of Energy Joint Genome Institute

Keith Tyo
Northwestern University

Akos Vertes
The George Washington University

Judy Wall
University of Missouri

Michael Wehner
Lawrence Berkeley National Laboratory

Jim Werner
Los Alamos National Laboratory

John Weyant
Stanford University

Will Wieder
National Center for Atmospheric Research

Shibu Yooseph
University of Central Florida

Mark Zelinka
Lawrence Livermore National Laboratory

Minghua Zhang
State University of New York at Stony Brook

Huimin Zhao
University of Illinois at Urbana-Champaign

Appendix D: Fig. 2.2 Image Credits and Permissions

Figure from BER. 2017. *Technologies for Characterizing Molecular and Cellular Systems Relevant to Bioenergy and the Environment,* DOE/SC-0189, U.S. Department of Energy Office of Biological and Environmental Research. [science.energy.gov/~/media/ber/pdf/community-resources/Technologies_for_Characterizing_Molecular_and_Cellular_Systems]

Row 1, Image 1. Nitrogenase. Courtesy David S. Goodsell and the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank.

Row 1, Image 2. PF1205. Reprinted by permission from Macmillan Publishers Ltd.: Hura, G. L., et al. 2009. "Robust, High-Throughput Solution Structural Analyses by Small Angle X-Ray Scattering (SAXS)," *Nature Methods* **6**, 606–12. © 2009

Row 1, Image 3. Comprehensive imaging of densely packed transmembrane proteins using photoactivated localization microscopy (PALM). Courtesy Greenfield, D., et al. 2009. "Self-Organization of the *Escherichia coli* Chemotaxis Network Imaged with Super-Resolution Light Microscopy," *PLoS Biology* 7(6), e1000137.

Row 1, Image 4. Confocal microscopy image of bacteria on the surface of poplar roots. Courtesy Jennifer L. Morrell-Falvey, Oak Ridge National Laboratory.

Row 1, Image 5. Scanning electron microscopy (SEM) image of a sulfate-reducing biofilm obtained from a borehole used for long-term (100+ days) acetate injection during biostimulation activities at the U.S. Department of Energy's (DOE) Integrated Field Research Challenge site near Rifle, Colorado. Courtesy Alice C. Dohnalkova, Pacific Northwest National Laboratory. Biofilm material courtesy Kenneth H. Williams, Lawrence Berkeley National Laboratory.

Row 1, Image 6. Positron-emission tomography (PET) image showing uptake of radioactive nitrate in a poplar sapling. Courtesy Richard Ferrieri, University of Missouri.

Row 1, Image 7. X-ray computed tomography cross section of an active layer/permafrost soil core from DOE's Office of Biological and Environmental Research Next-Generation Ecosystem Experiments–Arctic project. Courtesy Tim Neafsey, Lawrence Berkeley National Laboratory.

Row 2, Image 1. Haem-copper active site in respiratory enzymes. Reprinted by permission from Macmillan Publishers Ltd.: Hura, G.L., et al. 2009. "The Octahaem MccA is a Haem *c*–Copper Sulfite Reductase," *Nature* **520**, 706–09. © 2015

Row 2, Image 2. Density difference map for the ensemble structure $(mF_{obs} - DF_{model})\exp[i\varphi_{model}]$. Reprinted under a Creative Commons Attribution License (CC BY 4.0) from Burnley, B. T., et al. 2012. "Modelling Dynamics in Protein Crystal Structures by Ensemble Refinement," *eLIFE* **1**, e00311.

Row 2, Image 3. Acetate consumption over time of *Psychrobacter aestuarii* and *Geobacter sulfurreducens* in axenic cultures and co-cultures. Courtesy Alice C. Dohnalkova, Pacific Northwest National Laboratory.

Row 2, Image 4. Transmission electron microscopy (TEM) image of *Escherichia coli* O157:H7. Courtesy Centers for Disease Control and Prevention.

Row 2, Image 5. Plant root nutrient uptake. Courtesy Jennifer Pett-Ridge and Erin Nuccio, Lawrence Livermore National Laboratory.

Row 2, Image 6. Deconstructed biomass. Courtesy National Renewable Energy Laboratory Biomass Structural Characterization Laboratory.

Row 2, Image 7. Engineered bacteria produce biodiesel. Courtesy Joint BioEnergy Institute, Lawrence Berkeley National Laboratory.

Appendix E: References

Akyildiz, I. F., and J. M. Jornet. 2010. "Electromagnetic Wireless Nanosensor Networks," *Nano Communication Networks* **1**(1), 3–19. DOI:10.1016/j.nancom.2010.04.001.

Allegretti, M., et al. 2015. "Horizontal Membrane-Intrinsic α-Helices in the Stator *a*-Subunit of an F-Type ATP Synthase," *Nature* **521**(7551), 237–40. DOI:10.1038/nature14185.

Allhoff, F., et al. 2010. *What is Nanotechnology and Why does it Matter: From Science to Ethics*, Wiley-Blackwell, Chichester, United Kingdom. DOI:10.1002/9781444317992.

Allison, S. D. 2012. "A Trait-Based Approach for Modelling Microbial Litter Decomposition," *Ecology Letters* **15**(9), 1058–70. DOI:10.1111/j.1461-0248.2012.01807.x.

Amunts, A., et al. 2015. "The Structure of the Human Mitochondrial Ribosome," *Science* **348**(6230), 95. DOI:10.1126/science.aaa1193.

Anderegg, W. R., et al. 2015. "Tree Mortality from Drought, Insects, and Their Interactions in a Changing Climate," *New Phytologist* **208**(3), 674–83. DOI:10.1111/nph.13477.

Anderson, W., et al. In review. "Towards Coordinated International Support of Core Data Resources for the Life Sciences," *bioRxiv*. DOI:10.1101/110825.

Aquila, A., et al. 2012. "Time-Resolved Protein Nanocrystallography Using an X-Ray Free-Electron Laser," *Optics Express* **20**(3), 2706–16. DOI:10.1364/OE.20.002706.

ASCR. 2013a. *Data Crosscutting Requirements Review.* PNNL-22639. Office of Advanced Scientific Computing Research, U.S. Department of Energy Office of Science. [science.energy.gov/~/media/ascr/pdf/program-documents/docs/ASCR_DataCrosscutting2_8_28_13.pdf]

ASCR. 2013b. *Accelerating Scientific Knowledge Discovery (ASKD) Working Group Report.* Office of Advanced Scientific Computing Research, U.S. Department of Energy Office of Science. [science.energy.gov/~/media/ascr/pdf/program-documents/docs/ASKD_Report_V1_0.pdf]

ASCR. 2015. *Report on the DOE Workshop on Management, Visualization, and Analysis of Experimental and Observational Data: The Convergence of Data and Computing.* Office of Advanced Scientific Computing Research, U.S. Department of Energy Office of Science. [science.energy.gov/~/media/ascr/pdf/programdocuments/docs/ascr-eod-workshop-2015-report_160524.pdf]

ASCR. 2016. *Advanced Scientific Computing Research; Exascale Requirements Review.* Office of Advanced Scientific Computing Research, U.S. Department of Energy Office of Science. [science.energy.gov/~/media/ascr/pdf/programdocuments/docs/2017/DOE-ExascaleReport-ASCR-Final.pdf]

Bai, X.-c., et al. 2015. "Sampling the Conformational Space of the Catalytic Subunit of Human Γ-Secretase," *eLife* **4**, e11182. DOI:10.7554/eLife.11182.

Bale, J. B., et al. 2016. "Accurate Design of Megadalton-Scale Two-Component Icosahedral Protein Complexes," *Science* **353**(6297), 389–94. DOI:10.1126/science.aaf8818.

Balzarotti, F., et al. 2017. "Nanometer Resolution Imaging and Tracking of Fluorescent Molecules with Minimal Photon Fluxes," *Science* **355**(6325), 606–12. DOI:10.1126/science.aak9913.

Bao, Z., et al. 2016. "Accelerated Genome Engineering Through Multiplexing," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **8**(1), 5–21. DOI:10.1002/wsbm.1319.

Barberán, A., et al. 2012. "Using Network Analysis to Explore Co-Occurrence Patterns in Soil Microbial Communities," *The ISME Journal* **6**(2), 343–51. DOI:10.1038/ismej.2011.119.

Bascompte, J. 2007. "Networks in Ecology," *Basic and Applied Ecology* **8**(6), 485–90. DOI:10.1016/j.baae.2007.06.003.

Belward, A. S., and J. O. Skøien. 2015. "Who Launched What, When and Why; Trends in Global Land-Cover Observation Capacity from Civilian Earth Observation Satellites," *ISPRS Journal of Photogrammetry and Remote Sensing* **103**, 115–28. DOI:10.1016/j.isprsjprs.2014.03.009.

BERAC. 2010. *Grand Challenges for Biological and Environmental Research: A Long-Term Vision; A Report from the Biological and Environmental Research Advisory Committee March 2010 Workshop*. BERAC Steering Committee on Grand Research Challenges for Biological and Environmental Research, U.S. Department of Energy. [science.energy.gov/~/media/ber/pdf/Ber_ltv_report.pdf]

BERAC. 2013. *BER Virtual Laboratory: Innovative Framework for Biological and Environmental Grand Challenges; A Report from the Biological and Environmental Research Advisory Committee*. DOE/SC-0156, U.S. Department of Energy. [science.energy.gov/~/media/ber/berac/pdf/Reports/BER_VirtualLaboratory_finalwebLR.pdf]

Berlow, E. L., et al. 2009. "Simple Prediction of Interaction Strengths in Complex Food Webs," *Proceedings of the National Academy of Sciences USA* **106**(1), 187–91. DOI:10.1073/pnas.0806823106.

Bernhardt, E. S., et al. 2017. "Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept," *Ecosystems* **20**(4), 665–82. DOI:10.1007/s10021-016-0103-y.

Beven, K., and P. Germann. 2013. "Macropores and Water Flow in Soils Revisited," *Water Resources Research* **49**(6), 3071–92. DOI:10.1002/wrcr.20156.

Binzer, A., et al. 2012. "The Dynamics of Food Chains Under Climate Change and Nutrient Enrichment," *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**(1605), 2935–44. DOI:10.1098/rstb.2012.0230.

Biteen, J. S., et al. 2016. "Tools for the Microbiome: Nano and Beyond," *ACS Nano* **10**(1), 6–37. DOI:10.1021/acsnano.5b07826.

Blaiszik, B., et al. 2016. "The Materials Data Facility: Data Services to Advance Materials Science Research," *The Journal of The Minerals, Metals & Materials Society* **68**(8), 2045–52. DOI:10.1007/s11837-016-2001-3.

Bock, C., et al. 2016. "Multi-Omics of Single Cells: Strategies and Applications," *Trends in Biotechnology* **34**(8), 605–08. DOI:10.1016/j.tibtech.2016.04.004.

Boit, A., et al. 2012. "Mechanistic Theory and Modelling of Complex Food-Web Dynamics in Lake Constance," *Ecology Letters* **15**(6), 594–602. DOI:10.1111/j.1461-0248.2012.01777.x.

Bolotin, A., et al. 2005. "Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin," *Microbiology* **151**(8), 2551–61. DOI:10.1099/mic.0.28048-0.

Bonkowski, M. 2004. "Protozoa and Plant Growth: The Microbial Loop in Soil Revisited," *New Phytologist* **162**(3), 617–31. DOI:10.1111/j.1469-8137.2004.01066.x.

Bony, S., et al. 2015. "Clouds, Circulation and Climate Sensitivity," *Nature Geoscience* **8**(4), 261–68. DOI:10.1038/ngeo2398.

Boucher, O., et al. 2013. "Clouds and Aerosols; Supplementary Material," In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. 7SM-1–7SM-4. Eds. T. F. Stocker et al.

Boyken, S. E., et al. 2016. "*De Novo* Design of Protein Homo-Oligomers with Modular Hydrogen-Bond Network–Mediated Specificity," *Science* **352**(6286), 680–87. DOI:10.1126/science.aad8865.

Bradford, M. A., and N. Fierer. 2012. "The Biogeography of Microbial Communities and Ecosystem Processes: Implications for Soil and Ecosystem Models." In *Soil Ecology and Ecosystem Services*, 189-200. Eds. D. H. Wall and R. D. Bardgett. Oxford University Press, United Kingdom.

Bradford, M. A., et al. 2016. "Managing Uncertainty in Soil Carbon Feedbacks to Climate Change," *Nature Climate Change* **6**(8), 751–58. DOI:10.1038/nclimate3071.

Branton, D., et al. 2008. "The Potential and Challenges of Nanopore Sequencing," *Nature Biotechnology* **26**(10), 1146–53. DOI:10.1038/nbt.1495.

Brienen, R. J. W., et al. 2015. "Long-Term Decline of the Amazon Carbon Sink," *Nature* **519**(7543), 344–48. DOI:10.1038/nature14283.

Brose, U., and S. Scheu. 2014. "Into Darkness: Unravelling the Structure of Soil Food Webs," *Oikos* **123**(10), 1153–56. DOI:10.1111/oik.01768.

Brunette, T. J., et al. 2015. "Exploring the Repeat Protein Universe through Computational Protein Design," *Nature* **528**(7583), 580–84. DOI:10.1038/nature16162.

Buizza, R. 2010. "Horizontal Resolution Impact on Short- and Long-Range Forecast Error," *Quarterly Journal of the Royal Meteorological Society* **136**(649), 1020–35. DOI:10.1002/qj.613.

Callaway, E. 2015. "The Revolution Will Not Be Crystallized: A New Method Sweeps Through Structural Biology," *Nature* **525**(7568), 172–74. DOI:10.1038/525172a.

Chao, R., et al. 2015. "Recent Advances in DNA Assembly Technologies," *FEMS Yeast Research* **15**(1), 1–9. DOI:10.1111/1567-1364.12171.

Chard, K., et al. 2016. Globus: Recent Enhancements and Future Plans. *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale*, Miami, Association for Computing Machinery. DOI:10.1145/2949550.2949554.

Chen, S., et al. 2017. "GeMSTONE: Orchestrated Prioritization of Human Germline Mutations in the Cloud," *Nucleic Acids Research* **45**(W1), W207–14. DOI:10.1093/nar/gkx398.

Choi, P. J., et al. 2008. "A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell," *Science* **322**(5900), 442–46. DOI:10.1126/science.1161427.

Churkina, G. 2016. "The Role of Urbanization in the Global Carbon Cycle," *Frontiers in Ecology and Evolution: Urban Ecology* **3**(144). DOI:10.3389/fevo.2015.00144.

Clarholm, M. 1985. "Interactions of Bacteria, Protozoa and Plants Leading to Mineralization of Soil Nitrogen," *Soil Biology and Biochemistry* **17**(2), 181–87. DOI:10.1016/0038-0717(85)90113-0.

Cong, L., et al. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems," *Science* **339**(6121), 819–23. DOI:10.1126/science.1231143.

Copeland, M. F., et al. 2014. "Application of TALEs, CRISPR/Cas and sRNAs as Trans-Acting Regulators in Prokaryotes," *Current Opinion in Biotechnology* **29**, 46–54. DOI:10.1016/j.copbio.2014.02.010.

Crouser, R. J., et al. 2017. "Toward Theoretical Techniques for Measuring the Use of Human Effort in Visual Analytic Systems," *IEEE Transactions on Visualization and Computer Graphics* **23**(1), 121–30. DOI:10.1109/TVCG.2016.2598460.

Dafflon, B., et al. 2017. "Coincident Aboveground and Belowground Autonomous Monitoring to Quantify Covariability in Permafrost, Soil, and Vegetation Properties in Arctic Tundra," *Journal of Geophysical Research: Biogeosciences* **122**(6), 1321–42. DOI:10.1002/2016jg003724.

Damian, M., and M. H. Porteus. 2013. "A Crisper Look at Genome Editing: RNA-Guided Genome Modification," *Molecular Therapy* **21**(4), 720–22. DOI:10.1038/mt.2013.46.

Dasgupta, A., et al. 2017: "Human Factors in Streaming Data Analysis: Challenges and Opportunities for Information Visualization." In *Computer Graphics Forum*.

DeAngelis, K. M. 2016. "Chemical Communication Connects Soil Food Webs," *Soil Biology and Biochemistry* **102**, 48–51. DOI:10.1016/j.soilbio.2016.06.024.

Deutsch, C., et al. 2015. "Climate Change Tightens a Metabolic Constraint on Marine Habitats," *Science* **348**(6239), 1132–35. DOI:10.1126/science.aaa1605.

DiCarlo, J. E., et al. 2013. "Yeast Oligo-Mediated Genome Engineering (YOGE)," *ACS Synthetic Biology* **2**(12), 741–49. DOI:10.1021/sb400117c.

Diebolder, C. A., et al. 2012. "Pushing the Resolution Limits in Cryo Electron Tomography of Biological Structures," *Journal of Microscopy* **248**(1),1–5. DOI:10.1111/j.1365-2818.2012.03627.x.

Dietrich, J. A., et al. 2013. "Transcription Factor-Based Screens and Synthetic Selections for Microbial Small-Molecule Biosynthesis," *ACS Synthetic Biology* **2**(1), 47–58. DOI:10.1021/sb300091d.

Digel, C., et al. 2014. "Unravelling the Complex Structure of Forest Soil Food Webs: Higher Omnivory and More Trophic Levels," *Oikos* **123**(10), 1157–72. DOI:10.1111/oik.00865.

Dominguez, A. A., et al. 2016. "Beyond Editing: Repurposing CRISPR-Cas9 for Precision Genome Regulation and Interrogation," *Nature Reviews: Molecular Cell Biology* **17**(1), 5–15. DOI:10.1038/nrm.2015.2.

Downie, H., et al. 2012. "Transparent Soil for Imaging the Rhizosphere," *PLOS ONE* **7**(9), e44276. DOI:10.1371/journal.pone.0044276.

Downie, H. F., et al. 2014. "Transparent Soil Microcosms Allow 3D Spatial Quantification of Soil Microbiological Processes *In Vivo*," *Plant Signaling and Behavior* **9**(10), e970421. DOI:10.4161/15592316.2014.970421.

Egelman, E. H. 2016. "The Current Revolution in Cryo-EM," *Biophysical Journal* **110**(5), 1008–12. DOI:10.1016/j.bpj.2016.02.001.

Engler, C., et al. 2009. "Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes," *PLOS ONE* **4**(5), e5553. DOI:10.1371/journal.pone.0005553.

Engler, C., et al. 2008. "A One Pot, One Step, Precision Cloning Method with High Throughput Capability," *PLOS ONE* **3**(11), e3647. DOI:10.1371/journal.pone.0003647.

Evans, J. E., and N. D. Browning. 2013. "Enabling Direct Nanoscale Observations of Biological Reactions with Dynamic Tem," *Microscopy* **62**(1), 147–56. DOI:10.1093/jmicro/dfs081.

Evans, S. E., and M. D. Wallenstein. 2014. "Climate Change Alters Ecological Strategies of Soil Bacteria," *Ecology Letters* **17**(2), 155–64. DOI:10.1111/ele.12206.

Famiglietti, J. S. 2014. "The Global Groundwater Crisis," *Nature Climate Change* **4**(11), 945–48. DOI:10.1038/nclimate2425.

Feng, Z., et al. 2016. "More Frequent Intense and Long-Lived Storms Dominate the Springtime Trend in Central U.S. Rainfall," *Nature Communications* **7**, 13429. DOI:10.1038/ncomms13429.

Fierer, N., et al. 2012. "Cross-Biome Metagenomic Analyses of Soil Microbial Communities and Their Functional Attributes," *Proceedings of the National Academy of Sciences USA* **109**(52), 21390–95. DOI:10.1073/pnas.1215210110.

Flemming, H. C., et al. 2016. "Biofilms: An Emergent Form of Bacterial Life," *Nature Reviews: Microbiology* **14**(9), 563–75. DOI:10.1038/nrmicro.2016.94.

Fox, O., et al. 2006. "Soil Fauna Modifies the Recalcitrance-Persistence Relationship of Soil Carbon Pools," *Soil Biology and Biochemistry* **38**(6), 1353–63. DOI:10.1016/j.soilbio.2005.10.014.

Frank, M., et al. 2014. "Femtosecond X-Ray Diffraction from Two-Dimensional Protein Crystals," *IUCrJ* **1**(Pt 2), 95–100. DOI:10.1107/s2052252514001444.

Frederick, K. K., et al. 2015. "Sensitivity-Enhanced NMR Reveals Alterations in Protein Structure by Cellular Milieus," *Cell* **163**(3), 620–28. DOI:10.1016/j.cell.2015.09.024.

Fuhrman, J. A., and J. A. Steele. 2008. "Community Structure of Marine Bacterioplankton: Patterns, Networks, and Relationships to Function," *Aquatic Microbial Ecology* **53**, 69–81. DOI:10.3354/ame01222.

Galanie, S., et al. 2015. "Complete Biosynthesis of Opioids in Yeast," *Science* **349**(6252), 1095–1100. DOI:10.1126/science.aac9373.

Galbraith, E. D., and A. C. Martiny. 2015. "A Simple Nutrient-Dependence Mechanism for Predicting the Stoichiometry of Marine Ecosystems," *Proceedings of the National Academy of Sciences USA* **112**(27), 8199–204. DOI:10.1073/pnas.1423917112.

Gao, Y., et al. 2017. "Sensitivity of U.S. Summer Precipitation to Model Resolution and Convective Parameterizations Across Gray Zone Resolutions," *Journal of Geophysical Research: Atmospheres* **122**(5), 2714–33. DOI:10.1002/2016jd025896.

Garalde, D. R., et al. 2016. "Highly Parallel Direct RNA Sequencing on an Array of Nanopores," *bioRxiv*. DOI:10.1101/068809.

Garamella, J., et al. 2016. "The All *E. coli* TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology," *ACS Synthetic Biology* **5**(4), 344–55. DOI:10.1021/acssynbio.5b00296.

Garcia, P. A., et al. 2016. "Microfluidic Screening of Electric Fields for Electroporation," *Scientific Reports* **6** (21238). DOI:10.1038/srep21238.

Geoffroy, O., et al. 2017. "On the Role of the Stratiform Cloud Scheme in the Inter-Model Spread of Cloud Feedback," *Journal of Advances in Modeling Earth Systems* **9**(1), 423–37. DOI:10.1002/2016ms000846.

Gibson, D. G., et al. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases," *Nature Methods* **6**(5), 343–45. DOI:10.1038/nmeth.1318.

Gil, Y., et al. 2014. Excerpt from "Earthcube: Past, Present, and Future." In EarthCube Project EC-2014–3.

Glaeser, R. M. 2016. "How Good Can Cryo-EM Become," *Nature Methods* **13**(1), 28–32. DOI:10.1038/nmeth.3695.

Gleeson, T., et al. 2016. "The Global Volume and Distribution of Modern Groundwater," *Nature Geoscience* **9**(2), 161–67. DOI:10.1038/ngeo2590.

Gleeson, T., et al. 2014. "A Glimpse beneath Earth's Surface: GLobal HYdrogeology MaPS (GLHYMPS) of Permeability and Porosity," *Geophysical Research Letters* **41**(11), 3891–98. DOI:10.1002/2014GL059856.

Glownia, J. M., et al. 2010. "Time-Resolved Pump-Probe Experiments at the LCLS," *Optics Express* **18**(17), 17620-30. DOI:10.1364/OE.18.017620.

Gonod, L. V., et al. 2003. "Spatial Variability of 2,4-Dichlorophenoxyacetic Acid (2,4-D) Mineralisation Potential at a Millimetre Scale in Soil," *Soil Biology and Biochemistry* **35**(3), 373–82. DOI:10.1016/S0038-0717(02)00287-0.

Gregory, J. M., et al. 2009. "Quantifying Carbon Cycle Feedbacks," *Journal of Climate* **22**(19), 5232–50. DOI:10.1175/2009JCLI2949.1.

Griffiths, B. S. 1994. "Soil Nutrient Flow." In *Soil Protozoa*, 65–91. Ed. J. F. Darbyshire. CAB International.

Grundmann, G. L., and D. Debouzie. 2000. "Geostatistical Analysis of the Distribution of NH_4^+ and NO_2^--Oxidizing Bacteria and Serotypes at the Millimeter Scale Along a Soil Transect," *FEMS Microbiology Ecology* **34**(1), 57–62. DOI:10.1111/j.1574-6941.2000.tb00754.x.

Guyon, I., and A. Elisseeff. 2003. "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research* **3**, 1157–82.

Hampton, C. M., et al. 2017. "Correlated Fluorescence Microscopy and Cryo-Electron Tomography of Virus-Infected or Transfected Mammalian Cells," *Nature Protocols* **12**, 150–67. DOI:10.1038/nprot.2016.168.

Harcombe, W. R., et al. 2014. "Metabolic Resource Allocation in Individual Microbes Determines Ecosystem Interactions and Spatial Dynamics," *Cell Reports* **7**(4), 1104–15. DOI:10.1016/j.celrep.2014.03.070.

Hartmann, J., and N. Moosdorf. 2012. "The New Global Lithological Map Database GLiM: A Representation of Rock Properties at the Earth Surface," *Geochemistry, Geophysics, Geosystems* **13**(12). DOI:10.1029/2012gc004370.

Hawkes, C. V., and T. H. Keitt. 2015. "Resilience vs. Historical Contingency in Microbial Responses to Environmental Change," *Ecology Letters* **18**(7), 612–25. DOI:10.1111/ele.12451.

Hay, C. C., et al. 2015. "Probabilistic Reanalysis of Twentieth-Century Sea-Level Rise," *Nature* **517**(7535), 481–84. DOI:10.1038/nature14093.

Held, I. M., and B. J. Soden. 2006. "Robust Responses of the Hydrological Cycle to Global Warming," *Journal of Climate* **19**(21), 5686–699. DOI:10.1175/jcli3990.1.

Hohenegger, C., et al. 2009. "The Soil Moisture–Precipitation Feedback in Simulations with Explicit and Parameterized Convection," *Journal of Climate* **22**(19), 5003–20. DOI:10.1175/2009jcli2604.1.

Holman, H.-Y. N., et al. 2010. "Synchrotron IR Spectromicroscopy: Chemistry of Living Cells," *Analytical Chemistry* **82**(21), 8757–65. DOI:10.1021/ac100991d.

Holman, H.-Y. N., et al. 2009. "Real-Time Molecular Monitoring of Chemical Environment in Obligate Anaerobes During Oxygen Adaptive Response," *Proceedings of the National Academy of Sciences USA* **106**(31), 12599–604. DOI:10.1073/pnas.0902070106.

Holman, H.-Y., et al. 2002. "Synchrotron Infrared Spectromicroscopy as a Novel Bioanalytical Microprobe for Individual Living Cells: Cytotoxicity Considerations," *Journal of Biomedical Optics* **7**(3), 417–24. DOI:10.1117/1.1485299.

Horvath, P., and R. Barrangou. 2010. "CRISPR/Cas, the Immune System of Bacteria and Archaea," *Science* **327**(5962), 167–70. DOI:10.1126/science.1179555.

Houze, R. A., Jr., et al. 2015. "The Variable Nature of Convection in the Tropics and Subtropics: A Legacy of 16 Years of the Tropical Rainfall Measuring Mission Satellite," *Reviews of Geophysics* **53**(3), 994–1021. DOI:10.1002/2015RG000488.

IPCC. 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, 1535 pp. Cambridge, United Kingdom, and New York, NY, USA.

Jakobsen, I., and L. Rosendahl. 1990. "Carbon Flow into Soil and External Hyphae from Roots of Mycorrhizal Cucumber Plants," *New Phytologist* **115**(1), 77–83. DOI:10.1111/j.1469-8137.1990.tb00924.x.

Jakočiūnas, T., et al. 2015. "Multiplex Metabolic Pathway Engineering Using CRISPR/Cas9 in *Saccharomyces Cerevisiae*," *Metabolic Engineering* **28**, 213–22. DOI:10.1016/j.ymben.2015.01.008.

Johansson, L. C., et al. 2013. "Structure of a Photosynthetic Reaction Centre Determined by Serial Femtosecond Crystallography," *Nature Communications* **4**(2911). DOI:10.1038/ncomms3911.

Kallenbach, C. M., et al. 2016. "Direct Evidence for Microbial-Derived Soil Organic Matter Formation and Its Ecophysiological Controls," *Nature Communications* **7**(13630). DOI:10.1038/ncomms13630.

Karl, T. R., et al. 2015. "Possible Artifacts of Data Biases in the Recent Global Surface Warming Hiatus," *Science* **348**(6242), 1469–72.

Kay, J. E., et al. 2015. "The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability," *Bulletin of the American Meteorological Society* **96**(8), 1333–49. DOI:10.1175/bams-d-13-00255.1.

Kendon, E. J., et al. 2012. "Realism of Rainfall in a Very High-Resolution Regional Climate Model," *Journal of Climate* **25**(17), 5791–806. DOI:10.1175/jcli-d-11-00562.1.

Kern, J., et al. 2013. "Simultaneous Femtosecond X-Ray Spectroscopy and Diffraction of Photosystem II at Room Temperature," *Science* **340**(6131), 491–95. DOI:10.1126/science.1234273.

Kherlopian, A. R., et al. 2008. "A Review of Imaging Techniques for Systems Biology," *BMC Systems Biology* **2**, 74. DOI:10.1186/1752-0509-2-74.

Kindler, R., et al. 2009. "Fate of Bacterial Biomass Derived Fatty Acids in Soil and Their Contribution to Soil Organic Matter," *Organic Geochemistry* **40**(1), 29–37. DOI:10.1016/j.orggeochem.2008.09.005.

Klein, S. A., and A. Hall. 2015. "Emergent Constraints for Cloud Feedbacks," *Current Climate Change Reports* **1**(4), 276–87. DOI:10.1007/s40641-015-0027-1.

Knutti, R., and J. Sedláček. 2013. "Robustness and Uncertainties in the New CMIP5 Climate Model Projections," *Nature Climate Change* **3**(4), 369–73. DOI:10.1038/nclimate1716.

Kolodziejczyk, A. A., et al. 2015. "The Technology and Biology of Single-Cell RNA Sequencing," *Molecular Cell* **58**(4), 610–20. DOI:10.1016/j.molcel.2015.04.005.

Konopka, A. 2009. "What is Microbial Community Ecology?" *The ISME Journal* **3**(11), 1223–30. DOI:10.1038/ismej.2009.88.

Kooperman, G. J., et al. 2016. "Robust Effects of Cloud Superparameterization on Simulated Daily Rainfall Intensity Statistics Across Multiple Versions of the Community Earth System Model," *Journal of Advances in Modeling Earth Systems* **8**(1), 140–65. DOI:10.1002/2015ms000574.

Koplitz, S. N., et al. 2016. "Public Health Impacts of the Severe Haze in Equatorial Asia in September–October 2015: Demonstration of a New Framework for Informing Fire Management Strategies to Reduce Downwind Smoke Exposure," *Environmental Research Letters* **11**(9). DOI:10.1088/1748-9326/11/9/094023.

Kosaka, Y., and S.-P. Xie. 2016. "The Tropical Pacific as a Key Pacemaker of the Variable Rates of Global Warming," *Nature Geoscience* **9**(9), 669–73. DOI:10.1038/ngeo2770.

Kosuri, S., and G. M. Church. 2014. "Large-Scale *De Novo* DNA Synthesis: Technologies and Applications," *Nature Methods* **11**(5), 499–507. DOI:10.1038/nmeth.2918.

Kramer-Walter, K. R., et al. 2016. "Root Traits Are Multidimensional: Specific Root Length Is Independent from Root Tissue Density and the Plant Economic Spectrum," *Journal of Ecology* **104**(5), 1299–1310. DOI:10.1111/1365-2745.12562.

Lander, E. S. 2016. "The Heroes of CRISPR," *Cell* **164**(1–2), 18–28. DOI:10.1016/j.cell.2015.12.041.

Lang, B., et al. 2014. "Effects of Environmental Warming and Drought on Size-Structured Soil Food Webs," *Oikos* **123**(10), 1224–33. DOI:10.1111/j.1600-0706.2013.00894.x.

Lang, B., et al. 2012. "Warming Effects on Consumption and Intraspecific Interference Competition Depend on Predator Metabolism," *Journal of Animal Ecology* **81**(3), 516–23. DOI:10.1111/j.1365-2656.2011.01931.x.

Larabell, C. A., and K. A. Nugent. 2010. "Imaging Cellular Architecture with X-Rays," *Current Opinion in Structural Biology* **20**(5), 623–31. DOI:10.1016/j.sbi.2010.08.008.

Lareen, A., et al. 2016. "Plant Root-Microbe Communication in Shaping Root Microbiomes," *Plant Molecular Biology* **90**(6), 575–87. DOI:10.1007/s11103-015-0417-8.

Latendresse, M., et al. 2012. "Construction and Completion of Flux Balance Models from Pathway Databases," *Bioinformatics* **28**(3), 388–96. DOI:10.1093/bioinformatics/btr681.

Lau, W., and E. S. Sattely. 2015. "Six Enzymes from Mayapple that Complete the Biosynthetic Pathway to the Etoposide Aglycone," *Science* **349**(6253), 1224–28. DOI:10.1126/science.aac7202.

Lau, W., et al. 2014. "Key Applications of Plant Metabolic Engineering," *PLOS Biology* **12**(6), e1001879. DOI:10.1371/journal.pbio.1001879.

Le Gros, M. A., et al. 2012. "Visualizing Sub-Cellular Organization Using Soft X-Ray Tomography." In *Comprehensive Biophysics*. 90-110. Elsevier, Amsterdam.

Leis, A., et al. 2009. "Visualizing Cells at the Nanoscale," *Trends in Biochemical Sciences* **34**(2), 60–70. DOI:10.1016/j.tibs.2008.10.011.

Leray, M., et al. 2013. "A New Versatile Primer Set Targeting a Short Fragment of the Mitochondrial COI Region for Metabarcoding Metazoan Diversity: Application for Characterizing Coral Reef Fish Gut Contents," *Frontiers in Zoology* **10**(34). DOI:10.1186/1742-9994-10-34.

Leung, L. R., and Y. Gao. 2016. "Regional Downscaling of S2S Prediction: Past Lessons and Future Prospects," *US CLIVAR Variations* **14**(4), 13–17. U.S. Climate Variability and Predictability Program.

Levantino, M., et al. 2015. "Using Synchrotrons and XFELS for Time-Resolved X-Ray Crystallography and Solution Scattering Experiments on Biomolecules," *Current Opinion in Structural Biology* **35**, 41–48. DOI:10.1016/j.sbi.2015.07.017.

Li, Q., et al. 2014. "Bayesian Model of Protein Primary Sequence for Secondary Structure Prediction," *PLOS ONE* **9**(10), e109832. DOI:10.1371/journal.pone.0109832.

Li, X., et al. 2013. "Electron Counting and Beam-Induced Motion Correction Enable Near-Atomic-Resolution Single-Particle Cryo-EM," *Nature Methods* **10**(6), 584–90. DOI:10.1038/nmeth.2472.

Lima-Mendez, G., et al. 2015. "Determinants of Community Structure in the Global Plankton Interactome," *Science* **348**(6237). DOI:10.1126/science.1262073.

Loutherback, K., et al. 2016. "Microfluidic Approaches to Synchrotron Radiation-Based Fourier Transform Infrared (Sr-Ftir) Spectral Microscopy of Living Biosystems," *Protein and Peptide Letters* **23**(3), 273–82. DOI:10.2174/0929866523666160106154035.

Mali, P., et al. 2013. "RNA-Guided Human Genome Engineering via Cas9," *Science* **339**(6121), 823–26. DOI:10.1126/science.1232033.

Mardis, E. R. 2008. "Next-Generation DNA Sequencing Methods," *Annual Review of Genomics & Human Genetics* **9**, 387–402. DOI:10.1146/annurev.genom.9.081307.164359.

Martiny, A. C., et al. 2013. "Phylogenetic Conservatism of Functional Traits in Microorganisms," *The ISME Journal* **7**(4), 830–38. DOI:10.1038/ismej.2012.160.

Martiny, A. C., et al. 2009. "Taxonomic Resolution, Ecotypes and the Biogeography of *Prochlorococcus*," *Environmental Microbiology* **11**(4), 823–32. DOI:10.1111/j.1462-2920.2008.01803.x.

Martiny, J. B. H., et al. 2011. "Drivers of Bacterial β-Diversity Depend on Spatial Scale," *Proceedings of the National Academy of Sciences USA* **108**(19), 7850–54. DOI:10.1073/pnas.1016308108.

Martiny, J. B. H., et al. 2017. "Microbial Legacies Alter Decomposition in Response to Simulated Global Change," *The ISME Journal* **11**(2), 490–99. DOI:10.1038/ismej.2016.122.

McGlynn, S. E., et al. 2015. "Single Cell Activity Reveals Direct Electron Transfer in Methanotrophic Consortia," *Nature* **526**(7574), 531–35. DOI:10.1038/nature15512.

Meehl, G. A., et al. 2014. "Decadal Climate Prediction: An Update from the Trenches," *Bulletin of the American Meteorological Society* **95**(2), 243–67. DOI:10.1175/bams-d-12-00241.1.

Merk, A., et al. 2016. "Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery," *Cell* **165**(7), 1698–1707. DOI:10.1016/j.cell.2016.05.040.

Miller, K. 2016. "Data's Identity Crisis: The Struggle to Name It, Describe It, Find It, and Publish It," *Biomedical Computation Review* (Spring 2016), 22–27. biomedicalcomputationreview.org.

Mojica , F. J. M., and L. Montoliu. 2016. "On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals," *Trends in Microbiology* **24**(10), 811–20. DOI:10.1016/j.tim.2016.06.005.

Montoya, J. M., et al. 2006. "Ecological Networks and Their Fragility," *Nature* **442**(7100), 259–64. DOI:10.1038/nature04927.

Musat, N., et al. 2016. "Tracking Microbial Interactions with NanoSIMS," *Current Opinion in Biotechnology* **41**, 114–21. DOI:10.1016/j.copbio.2016.06.007.

Navarro, G., and V. Mäkinen. 2007. "Compressed Full-Text Indexes," *ACM Computing Surveys* **39**(1), article 2. DOI:10.1145/1216370.1216372.

Nazarenko, L., et al. 2017. "Interactive Nature of Climate Change and Aerosol Forcing," *Journal of Geophysical Research: Atmospheres* **122**(6), 3457–80. DOI:10.1002/2016jd025809.

Neal, A. L., et al. 2012. "Benzoxazinoids in Root Exudates of Maize Attract *Pseudomonas putida* to the Rhizosphere," *PLOS ONE* **7**(4), e35498. DOI:10.1371/journal.pone.0035498.

Nguyen, T. H. D., et al. 2016. "Cryo-EM Structure of the Yeast U4/U6.U5 tri-snRNP at 3.7 Å Resolution," *Nature* **530**(7590), 298–302. DOI:10.1038/nature16940.

Noy, N. F., and D. L. McGuinness: *Ontology Development 101: A Guide to Creating Your First Ontology*. Accessed: 2017. [protege.stanford.edu/publications/ontology_development/ontology101.pdf]

NRC. 2013. *Abrupt Impacts of Climate Change: Anticipating Surprises.* National Academies Press, 222 pp. Washington, D.C. [www.nap.edu/catalog/18373/abrupt-impacts-of-climate-change-anticipating-surprises]

Nunan, N., et al. 2002. "*In Situ* Patterns of Soil Bacterial Populations, Mapped at Multiple Scales, in an Arable Soil," *Microbial Ecology* **44**(4), 296–305. DOI:10.1007/s00248-002-2021-0.

OSTP. 2016. "Fact Sheet: Announcing the National Microbiome Initiative." Office of Science and Technology Policy [obamawhitehouse.archives.gov/the-press-office/2016/05/12/fact-sheet-announcing-national-microbiome-initiative]

Orth, J. D., et al. 2010. "What Is Flux Balance Analysis?" *Nature Biotechnology* **28**(3), 245–48. DOI:10.1038/nbt.1614.

Palmer, T. N., et al. 2008. "Toward Seamless Prediction: Calibration of Climate Change Projections Using Seasonal Forecasts," *Bulletin of the American Meteorological Society* **89**(4), 459–70. DOI:10.1175/bams-89-4-459.

Pan, Y., et al. 2011. "A Large and Persistent Carbon Sink in the World's Forests," *Science* **333**(6045), 988–93. DOI:10.1126/science.1201609.

Parishani, H., et al. 2017. "Toward Low-Cloud-Permitting Cloud Superparameterization with Explicit Boundary Layer Turbulence," *Journal of Advances in Modeling Earth Systems* **9**(3), 1542–71. DOI:10.1002/2017ms000968.

Pelletier, J. D., et al. 2016. "A Gridded Global Data Set of Soil, Intact Regolith, and Sedimentary Deposit Thicknesses for Regional and Global Land Surface Modeling," *Journal of Advances in Modeling Earth Systems* **8**(1), 41–65. DOI:10.1002/2015ms000526.

Pett-Ridge, J., and P. K. Weber. 2012. "NanoSIP: NanoSIMS Applications for Microbial Biology." In *Microbial Systems Biology: Methods and Protocols*, 375–408. Ed. A. Navid. Humana Press, New York, N.Y. [link.springer.com/bookseries/7651]

Phillips, T. J., et al. 2004. "Evaluating Parameterizations in General Circulation Models: Climate Simulation Meets Weather Prediction," *Bulletin of the American Meteorological Society* **85**(12), 1903–15. DOI:10.1175/bams-85-12-1903.

Podar, M., et al. 2007. "Targeted Access to the Genomes of Low-Abundance Organisms in Complex Microbial Communities," *ASM Applied and Environmental Microbiology* **73**(10), 3205–14. DOI:10.1128/AEM.02985-06.

Postel, S. L., et al. 1996. "Human Appropriation of Renewable Fresh Water," *Science* **271**(5250), 785–88. DOI:10.1126/science.271.5250.785.

Pourcel, C., et al. 2005. "CRISPR Elements in *Yersinia pestis* Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies," *Microbiology* **151**, 653–63, DOI:10.1099/mic.0.27437-0.

Pries, C. E. H., et al. 2017. "The Whole-Soil Carbon Flux in Response to Warming," *Science* **355**(6332), 1420–23. DOI:10.1126/science.aal1319.

Probst, A. J., et al. 2013. "Tackling the Minority: Sulfate-Reducing Bacteria in an Archaea-Dominated Subsurface Biofilm," *The ISME Journal* **7**(3), 635–51. DOI:10.1038/ismej.2012.133.

Qi, L. S., et al. 2013. "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression," *Cell* **152**(5), 1173–83. DOI:10.1016/j.cell.2013.02.022.

Raman, S., et al. 2014. "Evolution-Guided Optimization of Biosynthetic Pathways," *Proceedings of the National Academy of Sciences USA* **111**(50), 17803–08. DOI:10.1073/pnas.1409523111.

Randall, D., et al. 2016. "Chapter 15. Simulations of the Tropical General Circulation with a Multiscale Global Model." In *Multiscale Convection-Coupled Systems in the Tropics: A Tribute to Dr. Michio Yanai*, 15.11-15.15. Eds. R. G. Fovell, and W.-W. Tung. American Meteorological Society. DOI:10.1175/AMSMONOGRAPHS-D-15-0016.1.

Reich, P. B. 2014. "The World-Wide 'Fast–Slow' Plant Economics Spectrum: A Traits Manifesto," *Journal of Ecology* **102**(2), 275–301. DOI:10.1111/1365–2745.12211.

Richey, A. S., et al. 2015. "Quantifying Renewable Groundwater Stress with GRACE," *Water Resources Research* **51**(7), 5217–38. DOI:10.1002/2015WR017349.

Riede, J. O., et al. 2010. "Scaling of Food-Web Properties with Diversity and Complexity Across Ecosystems," *Advances in Ecological Research* **42**, 139–70. DOI:10.1016/B978-0-12-381363-3.00003-4.

Rinke, C., et al. 2013. "Insights into the Phylogeny and Coding Potential of Microbial Dark Matter," *Nature* **499**, 431–37. DOI:10.1038/nature12352.

Rosenthal, A. Z., et al. 2013. "Localizing Transcripts to Single Cells Suggests an Important Role of Uncultured Deltaproteobacteria in the Termite Gut Hydrogen Economy," *Proceedings of the National Academy of Sciences USA* **110**(40), 16163–68. DOI:10.1073/pnas.1307876110.

Rubin, R. L., et al. 2017. "Plant Growth Promoting Rhizobacteria Are More Effective Under Drought: A Meta-Analysis," *Plant and Soil* **416**(1–2), 309–23. DOI:10.1007/s11104-017-3199-8.

Ryan, O. W., et al. 2014. "Selection of Chromosomal DNA Libraries Using a Multiplex CRISPR System," *eLife*, e03703. DOI:10.7554/eLife.03703.

Saldin, D. K., et al. 2011. "Reconstructing an Icosahedral Virus from Single-Particle Diffraction Experiments," *Optics Express* **19**(18), 17318–35. DOI:10.1364/oe.19.017318.

Salve, R., et al. 2012. "Rain, Rock Moisture Dynamics, and the Rapid Response of Perched Groundwater in Weathered, Fractured Argillite Underlying a Steep Hillslope," *Water Resources Research* **48**(11). DOI:10.1029/2012wr012583.

Schimel, D., et al. 2015. "Effect of Increasing CO_2 on the Terrestrial Carbon Cycle," *Proceedings of the National Academy of Sciences USA* **112**(2), 436–41. DOI:10.1073/pnas.1407302112.

Schuur, E., et al. 2015. "Climate Change and the Permafrost Carbon Feedback," *Nature* **520**(7546), 171–79. DOI:10.1038/nature14338.

Schweigert, M., et al. 2015. "Fate of Ectomycorrhizal Fungal Biomass in a Soil Bioreactor System and Its Contribution to Soil Organic Matter Formation," *Soil Biology and Biochemistry* **88**, 120–27. DOI:10.1016/j.soilbio.2015.05.012.

Schwinger, J., et al. 2014. "Nonlinearity of Ocean Carbon Cycle Feedbacks in CMIP5 Earth System Models," *Journal of Climate* **27**(11), 3869–88. DOI:10.1175/JCLI-D-13-00452.1.

Selva, D., and D. Krejci. 2012. "A Survey and Assessment of the Capabilities of Cubesats for Earth Observation," *Acta Astronautica* **74**, 50–68. DOI:10.1016/j.actaastro.2011.12.014.

Shepherd, T. G. 2014. "Atmospheric Circulation as a Source of Uncertainty in Climate Change Projections," *Nature Geoscience* 7(10), 703–08. DOI:10.1038/ngeo2253.

Shi, D., et al. 2016. "The Collection of Microed Data for Macromolecular Crystallography," *Nature Protocols* **11**, 895–904. DOI: 10.1038/nprot.2016.046.

Shih, S. C., et al. 2015. "A Versatile Microfluidic Device for Automating Synthetic Biology," *ACS Synthetic Biology* **4**(10), 1151–64. DOI:10.1021/acssynbio.5b00062.

Slavov, G. T., et al. 2014. "Genome-Wide Association Studies and Prediction of 17 Traits Related to Phenology, Biomass and Cell Wall Composition in the Energy Grass *Miscanthus sinensis*," *New Phytologist* **201**(4), 1227–39. DOI:10.1111/nph.12621.

Solomon, K. V., et al. 2014. "Extracting Data from the Muck: Deriving Biological Insight from Complex Microbial Communities and Non-Model Organisms with Next Generation Sequencing," *Current Opinion in Biotechnology* **28**, 103–10. DOI:10.1016/j.copbio.2014.01.007.

SpaceWorks. 2017. *Small Satellite Report. Trends and Market Observations*. spaceworksforecast.com.

Spence, J. C., and P. W. Hawkes. 2008. "Diffract-and-Destroy: Can X-Ray Lasers 'Solve' the Radiation Damage Problem?" *Ultramicroscopy* **108**(12), 1502–03. DOI:10.1016/j.ultramic.2008.05.003.

Taberlet, P., et al. 2012. "Towards Next-Generation Biodiversity Assessment Using DNA Metabarcoding," *Molecular Ecology* **21**(8), 2045–50. DOI:10.1111/j.1365-294X.2012.05470.x.

Talbot, J. M., et al. 2015. "Functional Guild Classification Predicts the Enzymatic Role of Fungi in Litter and Soil Biogeochemistry," *Soil Biology and Biochemistry* **88**, 441–56. DOI:10.1016/j.soilbio.2015.05.006.

Tayar, A. M., et al. 2015. "Propagating Gene Expression Fronts in a One-Dimensional Coupled System of Artificial Cells," *Nature Physics* **11**(12), 1037–41. DOI:10.1038/nphys3469.

Thodey, K., et al. 2014. "A Microbial Biomanufacturing Platform for Natural and Semisynthetic Opioids," *Nature Chemical Biology* **10**(10), 83744. DOI:10.1038/nchembio.1613.

Thompson, S. E., et al. 2011. "Comparative Hydrology Across AmeriFlux Sites: The Variable Roles of Climate, Vegetation, and Groundwater," *Water Resources Research* **47**(10). DOI:10.1029/2010wr009797.

Tringe, S. G., et al. 2005. "Comparative Metagenomics of Microbial Communities," *Science* **308**(5721), 55457. DOI:10.1126/science.1107851.

Tsioris, K., et al. 2014. "A New Toolbox for Assessing Single Cells," *Annual Review of Chemical and Biomolecular Engineering* **5**, 455–77. DOI:10.1146/annurev-chembioeng-060713-035958.

Turetsky, M. R., et al. 2015. "Global Vulnerability of Peatlands to Fire and Carbon Loss," *Nature Geoscience* **8**(1), 11–14. DOI:10.1038/ngeo2325.

U.S. DOE. 2016. *Working Group on Virtual Data Integration: A Report from the August 13–14, 2015, Workshop.* DOE/SC-0180, U.S. Department of Energy Office of Science. DOI:10.2172/1227017.

U.S. DOE. 2015. *Biological Systems Science Division Strategic Plan.* U.S. Department of Energy Office of Biological and Environmental Research. [genomicscience.energy.gov/pubs/BSSDStrategicPlan.pdf]

Venter, J. C., et al. 2004. "Environmental Genome Shotgun Sequencing of the Sargasso Sea," *Science* **304**(5667), 6674. DOI:10.1126/science.1093857.

Villa, E., et al. 2014. "Structural Cell Biology: Preparing Specimens for Cryo-Electron Tomography Using Focused-Ion-Beam Milling," *Microscopy and Microanalysis* **20**, 1222–23. DOI:10.1017/S1431927614007843.

Vos, M., et al. 2013. "Micro-Scale Determinants of Bacterial Diversity in Soil," *FEMS Microbiology Reviews* **37**(6), 936–54. DOI:10.1111/1574-6976.12023.

Wada, Y., et al. 2010. "Global Depletion of Groundwater Resources," *Geophysical Research Letters* **37**(20), n/an/a. DOI:10.1029/2010GL044571.

Wainwright, H. M., et al. 2016. "Hierarchical Bayesian Method for Mapping Biogeochemical Hot Spots Using Induced Polarization Imaging," *Water Resources Research* **52**(1), 533–51. DOI:10.1002/2015WR017763.

Wang, H. H., et al. 2009. "Programming Cells by Multiplex Genome Engineering and Accelerated Evolution," *Nature* **460**(7257), 894–98. DOI:10.1038/nature08187.

Wang, Z., et al. 2014. "An Atomic Model of Brome Mosaic Virus Using Direct Electron Detection and Real-Space Optimization," *Nature Communications* **5**(4808). DOI:10.1038/ncomms5808.

Way, J. C., et al. 2014. "Integrating Biological Redesign: Where Synthetic Biology Came From and Where It Needs to Go," *Cell* **157**(1), 151–61. DOI:10.1016/j.cell.2014.02.039.

Wehner, M. F., et al. 2014. "The Effect of Horizontal Resolution on Simulation Quality in the Community Atmospheric Model, CAM5.1," *Journal of Advances in Modeling Earth Systems* **6**(4), 980–97. DOI:10.1002/2013ms000276.

White, T. A., et al. 2016. "Serial Femtosecond Crystallography Datasets from G Protein-Coupled Receptors," *Scientific Data* **3**(160057). DOI:10.1038/sdata.2016.57.

Wiedenheft, B., et al. 2012. "RNA-Guided Genetic Silencing Systems in Bacteria and Archaea," *Nature* **482**(7385), 331–38. DOI:10.1038/nature10886.

Wieder, W. R., et al. 2013. "Global Soil Carbon Projections Are Improved by Modelling Microbial Processes," *Nature Climate Change* **3**(10), 909–12. DOI:10.1038/nclimate1951.

Woyke, T., et al. 2009. "Assembling the Marine Metagenome, One Cell at a Time," *PLOS ONE* **4**(4), e5299. DOI:10.1371/Journal.Pone.0005299.

Woyke, T., et al. 2010. "One Bacterial Cell, One Complete Genome," *PLOS ONE* **5**(4), e10314. DOI:10.1371/Journal.Pone.0010314.

Yang, J., et al. 2013. "Synthetic RNA Devices to Expedite the Evolution of Metabolite-Producing Microbes," *Nature Communications* **4**, 1413. DOI:10.1038/ncomms2404.

Yoshida, S., et al. 2017. "Relationship Between Thunderstorm Electrification and Storm Kinetics Revealed by Phased Array Weather Radar," *Journal of Geophysical Research: Atmospheres* **122**(7), 3821–36. DOI:10.1002/2016jd025947.

Young, I. D., et al. 2016. "Structure of Photosystem II and Substrate Binding at Room Temperature," *Nature* **540**(7633), 453–57. DOI:10.1038/nature20161.

Zelezniak, A., et al. 2015. "Metabolic Dependencies Drive Species Co-Occurrence in Diverse Microbial Communities," *Proceedings of the National Academy of Sciences USA* **112**(20), 6449–54. DOI:10.1073/pnas.1421834112.

Appendix F: Acronyms and Abbreviations

2D, 3D	two-dimensional, three-dimensional
AEC	Atomic Energy Commission
AIM	PNNL Analysis in Motion Initiative
ARM	DOE Atmospheric Radiation Measurement Research Facility
ASCR	DOE Office of Advanced Scientific Computing Research
BER	DOE Office of Biological and Environmental Research
BERAC	Biological and Environmental Research Advisory Committee
BES	DOE Office of Basic Energy Sciences
BNL	Brookhaven National Laboratory
Cas	CRISPR-associated endonuclease
Cas9	CRISPR-associated protein-9 nuclease
CESD	BER Climate and Environmental Sciences Division
CESM	Community Earth System Model
CESMS	Center for Energy Sustainability Modeling and Synthesis
CF	Climate and Forecast ontology
CH_4	methane
cm	centimeter
CMIP	Coupled Model Intercomparison Project
CO_2	carbon dioxide
CRISPR	clustered regularly interspaced short palindromic repeats
cryoEM	cryo-electron microscopy
cryoET	cryo-electron tomography
DOE	U.S. Department of Energy
DOM	dissolved organic matter
E3SM	Energy Exascale Earth System Model
EMSL	DOE Environmental Molecular Sciences Laboratory
EOS	experimental and observational science
EPR	electron paramagnetic resonance
EPS	extracellular polymeric substance
ESGF	Earth System Grid Federation
ESM	Earth system model
FACE	Free-Air CO_2 Enrichment site
FEL	free-electron laser
FICUS	Facilities Integrating Collaborations for User Science program
FISH	fluorescence *in situ* hybridization
FTIR	Fourier transform infrared spectroscopy
GHG	greenhouse gas
gRNA	guide RNA
GWAS	genome-wide association study
H_2S	hydrogen sulfide
HEP	DOE Office of High Energy Physics
HFIR	ORNL High Flux Isotope Reactor
HFSP	Human Frontier Science Program
HPC	high-performance computing
IAM	integrated assessment model
iESM	integrated ESM
IFL	integrated field laboratory
IPO	interdecadal Pacific oscillation
IR s-SNOM	infrared scattering-type scanning near-field optical microscopy
JGI	DOE Joint Genome Institute
KBase	DOE Systems Biology Knowledgebase
kg	kilogram
km	kilometer
LCA	lifecycle analysis
LCLS	SLAC Linac Coherent Light Source
LES	large eddy simulation
MCS	mesoscale convective system
microED	micro-electron diffraction
mRNA	messenger RNA
MS	mass spectrometry
MX	macromolecular X-ray crystallography
N_2O	nitrous oxide
NanoSIMS	nanoscale secondary ion mass spectrometry
NASA	National Aeronautics and Space Administration
NCBI	NIH National Center for Biotechnology Information
NERSC	National Energy Research Scientific Computing Center
NEST	Network of Energy Sustainability Testbeds
NGEE	Next-Generation Ecosystem Experiments
NIH	National Institutes of Health
NIMBioS	National Institute for Mathematical and Biological Synthesis
NMR	nuclear magnetic resonance
NP	DOE Office of Nuclear Physics
NSF	National Science Foundation
NSLS-II	BNL National Synchrotron Light Source II
NSRC	DOE Nanoscale Science Research Center
NWP	numerical weather prediction
NX	neutron macromolecular crystallography

ORNL	Oak Ridge National Laboratory
PanDA	Production and Distributed Analysis project
PB	petabyte
PCR	polymerase chain reaction
PI	principal investigator
PNNL	Pacific Northwest National Laboratory
QM	quantum mechanics
rRNA	ribosomal RNA
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering
SFA	BER Scientific Focus Area
sgRNA	single-guide RNA
SLAC	SLAC National Accelerator Laboratory
SNS	ORNL Spallation Neutron Source
SOA	secondary organic aerosol
SOM	soil organic matter
SPRUCE	Spruce and Peatland Responses Under Changing Environments
TEM	transmission electron microscopy
UAS	unmanned aerial system
µm	micron
XFEL	X-ray FEL

DISCLAIMER