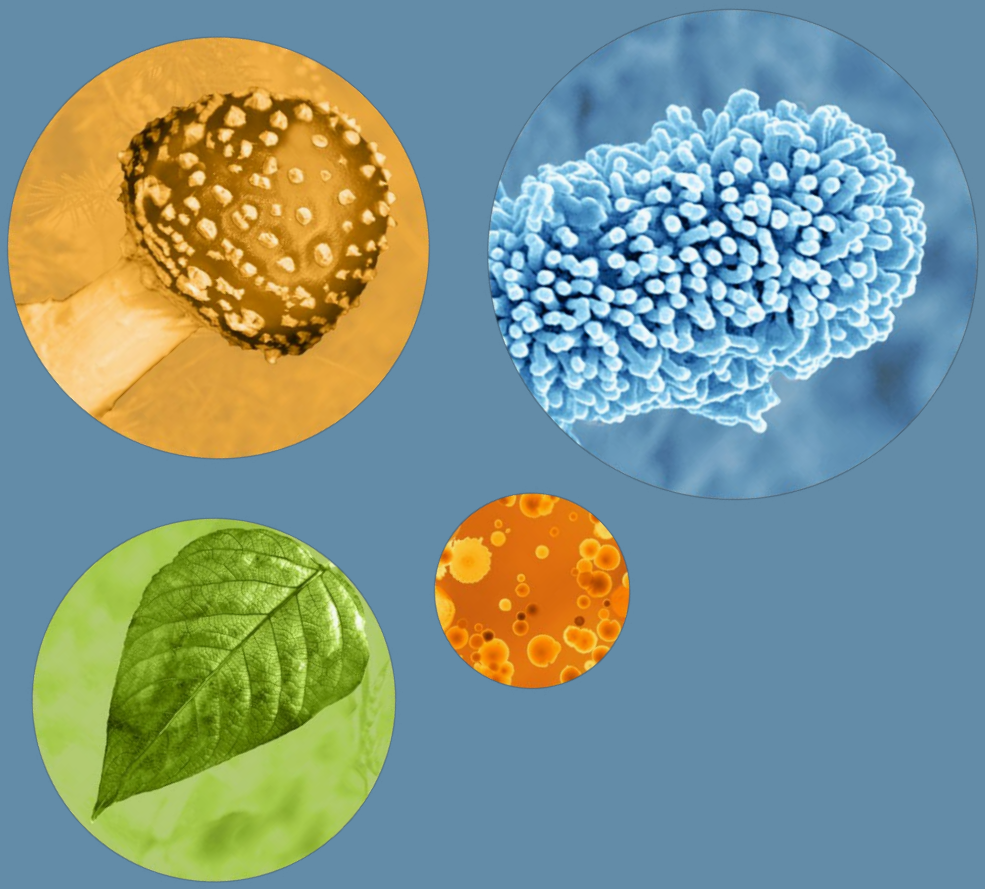
THE 11TH ANNUAL

DOE Joint Genome Institute

Genomics of Energy & Environment Meeting

March 21-24, 2016 • Walnut Creek Marriott • Walnut Creek, California

User Meeting Abstracts



JGI JOINT GENOME INSTITUTE
A DOE OFFICE OF SCIENCE USER FACILITY

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY
Office of Science

All information current as of March 15, 2016

JGI Contact: **Denise Yadon**
DOE Joint Genome Institute
dyadon@lbl.gov

The Joint Genome Institute is a user facility of
the Department of Energy Office of Science

DOE Joint Genome Institute: **www.jgi.doe.gov**
DOE Office of Science: **science.energy.gov**

THE 11TH ANNUAL

DOE Joint Genome Institute

Genomics of Energy & Environment Meeting

March 21-24, 2016 • Walnut Creek Marriott • Walnut Creek, California

User Meeting Abstracts

Sponsored By
U.S. Department of Energy
Office of Science

Table of Contents

Speaker Presentations

Abstracts alphabetical by speaker

The Nature of Scientific Publishing

Bahcall, Orli* (o.bahcall@us.nature.com)

Nature.

Tara Oceans: Eco-Systems Biology at Planetary Scale

Bowler, Chris* (cbowler@biologie.ens.fr); Tara Oceans Consortium

Ecology and Evolutionary Biology Section, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Paris, France.

The ocean is the largest ecosystem on Earth and yet we know very little about it. This is particularly true for the plankton that drift within. Although these organisms are at least as important for the Earth system as the forests on land, most of them are invisible to the naked eye and thus are largely uncharacterized, even though they form the base of marine food webs. To increase our understanding of this underexplored world, a multidisciplinary consortium, Tara Oceans, was formed around the 110-ft research schooner Tara, which sampled plankton at more than 210 sites and multiple depth layers in all the major oceanic regions during expeditions from 2009-2013 (Karsenti et al. Plos Biol., 2011). The seminar will describe the first foundational resources from the project (based on a first data freeze from 579 samples at 75 stations; see Science special issue May 22, 2015) and their initial analyses, illustrating several aspects of the Tara Oceans' eco-systems biology approach. The project provides unique resources for several scientific disciplines, capturing biodiversity of a wide range of organisms that are rarely studied together, exploring interactions between them and integrating them with environmental conditions to further our understanding of life in the ocean and beyond in the context of ongoing climate changes.

Deep Sequencing Methods for Detecting Infectious Agents

Chiu, Charles* (charles.chiu@ucsf.edu)

University of California, San Francisco.

Genetic Basis of Competitive Success and Priority Effect Exhibited by Dominant Nectar Yeast, *Metschnikowia reukaufii*

Dhami, Manpreet* (mdhami@stanford.edu)

Stanford University, Stanford, California, USA.

Floral nectar hosts a complex community of microbes. Floricolous yeast, Metschnikowia reukaufii, is a dominant species in this community, competing with other fungi and bacteria. Specifically, this species exerts a strong priority effect, excluding other microbes and deterministically influencing community structure. Sequencing and analysis of the M. reukaufii genome revealed a putative genetic mechanism by which it specialises in nectar. We found high rates of tandem gene duplication, which were enriched for nitrogen metabolism and transport. The two major high capacity amino acid transport (HCAT) genes, involved in amino acid scavenging when nitrogen is scarce, were present in tandem gene arrays. These additional HCATs are expressed under nectar conditions but repressed if nitrogen is abundant. We propose that M. reukaufii evolved the ability to rapidly and efficiently deplete amino acid from nectar, limiting the growth of subsequently arriving microbes, while protecting itself from osmotic stress and explains the strong priority effect and competitive success exhibited by *M. reukaufii*.

The Divining Root: Understanding How Roots Find Water in a Heterogeneous Environment

Dinneny, Jose* (jdinneny@carnegiescience.edu)

Carnegie Institution for Science.

While plants are sedentary, their bodies often traverse long distances as they explore their local environment in search of resources necessary for growth. The biology of root systems is governed by both micro-scale and systemic signaling that allows the plant to integrate these complex variables into growth and branching decisions that ultimately determine the efficiency with which resources are captured. Research in my lab is aimed at understanding the response of roots to water-limiting conditions and is exploring this process at different organizational scales from the individual cell type to the level of the whole plant.

Mining Genomes for Producing Fuels and Chemicals from Biomass

Donohue, Tim* (tdonohue@bact.wisc.edu)

University of Wisconsin-Madison.

Great Lakes Bioenergy (www.glbrc.org) is one of three Department of Energy funded Bioenergy Research Centers conducting genome-enabled, systems level research to provide the underpinnings for production of fuels and chemicals from the lignocellulosic, or non-edible, fraction of plant biomass. To achieve its mission, GLBRC basic science activities are led by academic and national lab scientists (University of Wisconsin-Madison, Michigan State University, University of British Columbia, Texas A&M

University, Illinois State University, University of New Hampshire, University of Maryland and Pacific Northwest National Lab) who are focused on providing knowledge toward the sustainable production of crops with desirable biofuel traits and energy efficient conversion of biomass into fuels and chemicals. The Center's research portfolio is built on integrated, discovery research activities in Sustainability: to identify factors affecting the ecological attributes of biofuel cropping systems, Plants: to improve the traits and productivity of bioenergy crops, Deconstruction: to improve chemical and enzymatic methods for solubilizing lignocellulosic biomass, and Conversion: to improve microbial and chemical systems for converting materials derived from lignocellulosic biomass to biofuels and chemicals. Since its inception in 2007, Great Lakes Bioenergy has reported findings in >850 papers, produced knowledge that led to filing ~125 patents, licenses for ~3 dozen technologies, and formation of several start-up companies. This talk will report on several projects that have used high-throughput sequencing to mine plant and microbial genomes for the production of cellulosic biofuels.

Microbial Drivers of Cellulose Cycling in Biofuel Crop Soils

Hofmockel, Kirsten S.* (kirsten.hofmockel@pnnl.gov)[1,2]; Erb, Racheal[1]; Bell, Sheryl[1]; Christer, Will[2]; Thompson, Allison[1]; Brown, Joe[1]; Bramer, Lisa[1]; Orr, Galya[2]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA.
[2]Iowa State University, Ames, IA.

We examined the microbial interactions and metabolic functions involved in the degradation of cellulose in enrichment cultures derived from biofuel agroecosystem experiments. To identify key organisms and enzymes involved in root decomposition, we used fluorescently labeled cellulose nanocrystals in combination with metatranscriptomics in controlled enrichment experiments. We demonstrate that the community level responses (OD, respiration, 16S amplicons) were not significantly different when soil communities were grown with cellulose or fluorescently labeled cellulose nanocrystals over a 10d incubation, suggesting our experimental platform can be used to track the fate of cellulose through soil decomposer communities. Our results reveal that Proteobacteria, which represent 10% of the native soil community, dominated the community in our cellulose degrading experiments, comprising 32% of the enrichment culture and 80% of the community by day 10. Pseudomonadales were the most abundant, increasing from 2% of the community to greater than 50% during the cellulose enrichment experiment. Organisms assimilating the fluorescently labeled cellulose have been analyzed in conjunction with metatranscriptomic data to understand the basic ecology regulating cellulose decomposition in diverse soil communities. By targeting extracellular enzymes typically measured in biogeochemical field studies and organisms native to the soil community, we aim to link reduced lab studies to field approaches and predictive models.

This research was supported by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research under Contract No. DESC0010775. This work was part of the FICUS, EMSL-JGI user program.

Pleiotropic and Epistatic Network-Based Discovery: Integrated SNP Correlation, Co-expression and Genome-Wide Association Networks for *Populus trichocarpa*

Jacobson, Daniel* (jacobsonda@ornl.gov)[1,2]; Weighill, Deborah[1,2]; Bleker, Carissa[1,2]; Tuskan, Gerald A.[1]; Muchero, Wellington[1]; Tschaplinski, Timothy[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee.[2]University of Tennessee, Knoxville, Tennessee.

http://bioenergycenter.org/besc/

Biological organisms are complex systems that are composed of pleiotropic functional networks of interacting molecules and macro-molecules. Complex phenotypes are the result of orchestrated, hierarchal, heterogeneous collections of expressed genomic variants. However, the effects of these variants are the result of historic selective pressure and current environmental and epigenetic signals, and, as such, their co-occurrence can be seen as genome-wide correlations in a number of different manners. Biomass recalcitrance (i.e., the resistance of plants to degradation or deconstruction, which ultimately enables access to a plant's sugars) is a complex polygenic phenotype of high importance to DOE's biofuels programs. We are using data derived from the re-sequenced genomes from over 1000 alternate *Populus trichocarpa* genotypes in combination with transcriptomics, metabolomics and phenomics data across this population in order to better understand the molecular interactions involved in recalcitrance. The resulting Genome Wide Association Study networks, integrated with SNP correlation and co-expression networks, are proving to be a powerful approach to determine the pleiotropic and epistatic relationships underlying cellular functions and, as such, the molecular basis for complex phenotypes, such as recalcitrance.

Pleiotropic and Epistatic Network-Based Discovery of Plant Functions Involved in Microbial Interactions: Integrated SNP Correlation, Co-expression and Genome-Wide Association Networks for *Populus trichocarpa*

Jacobson, Daniel* (jacobsonda@ornl.gov)[1,2]; Weighill, Deborah[1,2]; Bleker, Carissa[1,2]; Tuskan, Gerald A.[1]; Muchero, Wellington[1]; Tschaplinski, Timothy[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee.[2]University of Tennessee, Knoxville, Tennessee.

Biological organisms are complex systems that are composed of pleiotropic functional networks of interacting molecules and macro-molecules. Complex phenotypes are the result of orchestrated, hierarchal, heterogeneous collections of expressed genomic variants regulated by and related to biotic and abiotic signals. However, the effects of these variants are the result of historic selective pressure and current environmental as well as epigenetic interactions, and, as such, their co-occurrence can be seen as genome-wide associations in a number of different manners. In this context, a plant's association with its microbiome is a complex set of interactions involving many genes and metabolites. We are using data derived from the re-sequenced genomes from over 1000 alternate *Populus trichocarpa* genotypes in combination with transcriptomics, metabolomics and phenomics data across this population in order to better understand the molecular interactions involved in plant-microbe interfaces. The resulting Genome-Wide Association

Study networks, integrated with SNP correlations and co-expression networks, are proving to be a powerful approach to determine the pleiotropic and epistatic relationships underlying cellular functions and, as such, some of the molecular underpinnings for plant-microbiome associations. The co-occurrence of LRKs and SSPs in a Salicylate-based GWAS and SNP-correlation subnetwork with genes involved in biotic stress and cell wall modification could suggest a model in which a plant senses the presence of microbes (via LRKs), adjusts its own defense responses and attempts to regulate the microbiome chemically (via Salicylates) and by excreting SSPs to regulate gene expression of members of the microbiome itself.

Characterizing Plant Growth Promoting Members of the Duckweed Microbiome

Lebeis, Sarah* (slebeis@utk.edu)[1]; McGuire, Kathleen[1]; Gilbert, Sarah[2]; Acosta, Kenneth[2]; Lam, Eric[2]

[1]University of Tennessee. [2]Rutgers University.

While several mechanisms utilized by beneficial microbes associated with terrestrial plants have been characterized, less is known about how microbes affect aquatic plant growth. We hypothesized that such beneficial relationships also exist for aquatic plants in the family Lemnaceae, commonly known as duckweeds. Duckweed is the fastest growing flowering plant, and holds potential to be used for phytoremediation of wastewater by extracting carbon, nitrogen, and phosphorus. Further, duckweed is easily harvested, and can be used as biomass for biofuel production. We performed massive parallel sequencing of duckweed grown in from a municipal wastewater treatment plant, as well as water controls to characterize the assembly of duckweed microbiomes from their aquatic environment. Further, we have built a collection of isolated endophytic bacteria from natural duckweed ecotypes. In order to determine the potential molecular mechanisms for growth promotion phenotypes, we elected to screen our duckweed-terrestrial plant, Arabidopsis thaliana. Among the duckweed-associated microbes screened, a Micrococcus sp. colonizes the endophytic compartment dramatically altering root morphology of *A. thaliana*, probably via auxin production. We will use the remainder of our isolate collection to define other such growth promotion mechanisms. Finally, we amended the characterized municipal wastewater with our auxin-producing Micrococcus sp. to determine if it altered either duckweed microbiome composition or the ability of duckweed to extract total organic carbon from the water. These studies in duckweed will set the stage for deployment of defined microbial amendments to optimize farming of these aquatic plants for renewable biomass production.

Engineered Metagenomes for Superfund Sites, Subways, and Space Stations

Mason, Christopher* (chm2042@med.cornell.edu)

Weill Cornell Medicine, Cornell University, New York, NY.

Sites with extreme physical dynamics in pH, salinity, toxins, and/or gravity can lead to a bounty of discovery for novel biology, new drugs and unique small molecules. We have been systematically hunting

for and engineering new biological networks with three large-scale projects: the eXtreme Microbiome Project (XMP), the Metagenomics and metadesign of Subways and Urban Biomes (MetaSUB), and work with NASA on the International Space Station (ISS). Among these sampled sites, we have already found that Brooklyn's Gowanus Canal, a EPA SuperFund site, holds a suite of unique and potentially protective microbes, which can absorb toxins (e.gl. toluene and acetone) from the environment. We have been designing artificial sponges to hold these beneficial organisms in the canal during the remediation process, and we are now incorporating these designs into the Canal's rebuilding. This is part of a larger project of urban microbiome monitoring and design, called the Brooklyn Bioreactor, which is a unique collaboration between our laboratory at Weill Cornell Medicine, the landscape architecture firm Nelson Byrd Woltz, the Gowanus Conservancy, and the community laboratory Genspace. We have also formally launched a new research program for medical and graduate students on the study of space genetics and engineered biology, enabling students to learn novel methods in synthetic biology, materials science, nanofabrication, microbiome engineering, and gene drives. Indeed, such multi-disciplinary skills will be needed for continued long-term human space travel and the ever-increasing instances of metagenome, cross-kingdom engineering for health and disease.

Nature's Toolkit: Evolutionary 'Experiments' in Host-Microbe Interactions

McFall-Ngai, Margaret* (KEYNOTE) (mjmcfallngai@wisc.edu)

University of Hawaii.

Towards Perfect Sequence Assembly

Myers, Gene* (myers@mpi-cbg.de)

Max-Planck Institute for Molecular Cell Biology & Genetics.

With the advent of long read sequencers such as the PacBio RS II, the goal of near-perfect *de novo* reconstructions of unknown genomes is once again a realistic possibility. We will explain why, and further give a hypothesis as to why assemblers have improved only marginally since the era of the Human Genome Project circa 2000, namely that it is not about the assembly, but about the artifacts in the reads and the resolution of repeat families, topics that have not received sufficient attention and that are particularly critical issues for long reads.

Therefore we are developing algorithms that carefully analyze a long read shotgun data set before assembly. By efficiently comparing all the data against itself we have developed a computational approach to accurately determine the quality of any stretch of a PacBio read based only on the sequence data itself. These regional QVs allow us to accurately identify low quality regions, chimers, and missed adaptamers. Removing these artifacts with a process we call scrubbing leaves one with reads that assemble without the need for base-level error correction. We further find that we can identify and annotate repetitive sequences prior to assembly, albeit this aspect is still a work in progress.

We will conclude with a number of sequencing projects we are undertaking and on describing what assembly tools are currently available from our lab.

Primary Roles for Secondary Metabolites

Newman, Dianne K.* (KEYNOTE) (dknewman@gmail.com)

California Institute of Technology.

Deciphering the Biomass-Degrading Abilities of Anaerobic Gut Fungi

O'Malley, Michelle A.* (momalley@engineering.ucsb.edu)[1]; Solomon, Kevin V.[1]; Haitjema, Charles H.[1]; Henske, John K.[1]; Gilmore, Sean P.[1]; Borges-Rivera, Diego[2]; Lipzen, Anna[3]; Brewer, Heather M.[4]; Purvine, Samuel O.[4]; Wright, Aaron T.[4]; Theodorou, Michael K.[5]; Grigoriev, Igor V.[3]; Regev, Aviv[2]; Thompson, Dawn[2]

[1]UC-Santa Barbara, Santa Barbara, CA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]DOE JGI, Walnut Creek, CA, USA. [4]PNNL, Richland, WA, USA. [5]Harper Adams University, Newport, Shropshire, United Kingdom.

Renewable chemicals derived from plant biomass (mainly composed of cellulose and lignin) are attractive alternatives to those made from petroleum. To produce chemicals from biomass, enzymes are used to break down cellulose into simple sugars, which are later fermented into value-added products. However, since cellulose is tightly bound within a network of crystalline cellulosic fibers and lignin, existing biomass degrading enzymes are not very efficient. To develop new technologies to break down plant material into sugar, much can be learned by studying how microbes digest lignocellulose in biomass-rich environments, such as the digestive tract of large herbivores. Anaerobic fungi are native to the gut and rumen of these animals, where they have evolved powerful enzymes to degrade plant biomass. Our goal is to develop new experimental tools to engineer anaerobic fungi and anaerobic microbial consortia for lignocellulose breakdown and chemical production. To accomplish this goal, we isolated a panel of anaerobic fungi and associated microbes from different herbivores and screened for their ability to degrade several types of lignin-rich grasses and agricultural waste. By focusing on model anaerobic fungi from the Piromyces, Neocallimastix, and Anaeromyces genera, we have employed next-generation sequencing to discover thousands of new genes, revealing hundreds of novel biomass-degrading enzymes. Additionally, we have characterized key regulatory patterns for these enzymes, which depend on the environment of the fungus. Using this information, we are developing new genetic engineering strategies to manipulate gut fungi at the molecular level, along with 'bottom-up' strategies to synthesize microbial consortia for compartmentalized breakdown and bioproduction.

Return to the Sea: the Genome of the Seagrass *Zostera marina*

Olsen, Jeanine* (j.l.olsen@rug.nl)

Groningen Institute of Evolutionary Life Sciences.

Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in *A. thaliana*

Queitsch, Christine* (queitsch@u.washington.edu); Sullivan, Alessandra M.; Bubb, Kerry L.; Arsovski, Andrej A.; Lempe, Janne; Dorrity, Michael; Alexandre, Cristina; Weirauch, Matthew T.; Sabo, Peter J.; Sandstrom, Richard; Thurman, Robert E.; Neph, Shane; Reynolds, Alex P.; Stergachis, Andrew; Vernot, Benjamin; Johnson, Audra K.; Haugen, Eric; Dargiel, Agnieszka; Neri, Jun; Weaver, Molly; Diegel, Morgan; Hughes, Timothy R.; Nemhauser, Jennifer L.; Stamatoyannopoulos, John A.

Department of Genome Sciences, University of Washington, Seattle, WA.

Our understanding of gene regulation in plants is constrained by our limited knowledge of plant cis-regulatory DNA and its dynamics. *Cis*-regulatory elements can be delineated by their characteristic hypersensitivity to the endonuclease DNase I. We present the results of DNase I-seq experiments for the reference plant *A. thaliana*. First, we mapped DNase I hypersensitive sites (DHSs) in *A. thaliana* seedlings and use genomic footprinting to delineate ~700,000 sites of *in vivo* transcription factor (TF) occupancy. These studies reveal novel properties of *A. thaliana* DHSs, including evidence that highly significant GWAS variants are enriched within DHSs and that widespread TF binding within exons may have shaped codon usage patterns. We show that the architecture of *A. thaliana* TF regulatory networks is strikingly similar to that of animals in spite of diverged regulatory repertoires. Second, we present chromatin dynamics in response to heat shock, disclosing thousands of environmentally-sensitive elements. The observed dynamics suggest a provocative hypothesis for future exploration; ongoing technology development for this project will be discussed. Third, we present developmentally dynamic DHSs in seed coat cells during their transition from growth to mucilage secretion. We show that differentially expressed genes are associated with dynamic DHSs and implicate new TFs and candidate genes involved in seed coat differentiation. Lastly, we will present data on the natural variation in chromatin accessibility in five diverse *A. thaliana* accessions. We show that variable DHSs are more polymorphic than static DHSs across the accessions. Deletions account for 15% of variable DHSs, suggesting they are a powerful force in shaping diverse patterns of gene regulation in the accessions.

Adaptation in Plant Genomes: the Importance of Demographic History and Genome Size

Ross-Ibarra, Jeffrey* (rossibarra@ucdavis.edu)

University of California, Davis.

Developing Plant-Inspired Cultured Aromas Using a Foundry for Organism Engineering

Shetty, Reshma* (reshma@ginkgobioworks.com)

Gingko Bioworks, Inc.

There is an emerging demand for sourcing plant-derived extracts (nutraceuticals, flavors, fragrances, sweeteners, etc.) from engineered microbes. While recent advances in synthetic biology and metabolic engineering provide feasible approaches to engineering such organisms, commercial success for developing these "cultured" ingredients presents specific challenges. Unlike biofuels, where efforts can be focused on one particular molecule given the enormous market size, cultured ingredients require developing different organism lines in a rapid and low cost fashion. This requires a scalable solution for bio-manufacturing of organisms, which is provided by our state of the art foundry that continues to grow. I will describe how organism development at Ginkgo leverages our foundry to accelerate the design/build/test using specific examples. In particular, I will highlight the value of combining computer-aided engineering software tools, cheap gene synthesis and high resolution-accurate mass LCMS to develop engineered microbes. Finally, I'll touch on how our improvements in manufacturing organisms lend to opportunities outside of cultured ingredients.

Design and Synthesis of a Minimal Bacterial Genome

Smith, Hamilton* (hsmith@jcvi.org)

Hamilton Smith and the Synthetic Biology Group at the J. Craig Venter Institute, La Jolla, CA and Rockville, MD.

Whole-genome design and complete chemical synthesis were used to minimize the 1079-kb synthetic genome of Mycoplasma mycoides JCVI-syn1.0 (Science, 2010). An initial design, based on collective knowledge of molecular biology plus limited transposon mutagenesis data, failed to produce a viable cell. Improved transposon mutagenesis methods revealed a class of quasi-essential genes needed for robust growth, explaining the failure of our initial design. Three cycles of design, synthesis, and testing, with retention of quasi-essential genes, produced JCVI-syn3.0 (531kb, 438 protein and 35 RNA coding genes), with a genome smaller than that of any autonomously replicating cell found in nature. JCVI-syn3.0 retains almost all genes involved in synthesis and processing of macromolecules. Surprisingly, it also contains 65 genes with unknown functions and 84 genes with imprecisely known biological functions. JCVI-syn3.0 is a versatile platform for investigating the core functions of life, and for exploring whole-genome design.

Searching for Life on Earth and Beyond

Voytek, Mary* (mary.voytek-1@nasa.gov)

NASA.

Life in the Deep Subsurface: What's Shale Got to Do with It?

Wrighton, Kelly* (kwrighton@gmail.com)

Ohio State University.

Poster Presentations

Posters alphabetical by presenting author*

Exploring “Dark Matter Fungi” Using Single-Cell Genomics

Ahrendt, Steven R.* (sahrendt0@lbl.gov)[1, 2], Quandt, C.A.[3], Ciobanu, Doina[1], Clum, Alicia[1], Salamov, Asaf[1], Chang, Jan-Fang[1], Woyke, Tanja[1], James, Timothy Y.[3], Grigoriev, Igor V.[1, 2]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, California. [2]Dept. of Plant and Microbial Biology, University of California, Berkeley, CA. 3Dept. of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI.

Fungi are critical research systems as they impact global carbon cycling and have remarkable potential for sustainable biofuel production processes. Current estimates suggest that under 10% of the estimated 1.5 million species worldwide have been described, and the majority of these undescribed species fall within the earliest diverging lineages of the fungal evolutionary tree. Members of these lineages are distinct from more traditionally recognizable fungal species (e.g. in the Dikarya) by their aquatic lifestyles facilitated by the presence of a motile, flagellated spore. As such, they are often referred to as “zoosporic fungi”. Aquatic ecosystems not only play an important role in the global carbon cycle, but are also necessary for current algal biofuel production processes. While these environments are enriched for zoosporic fungi, and certain zoosporic fungal species are known algal pathogens, the potential impact of such species on these environments and, in particular, this type of biofuel production remains to be explored. Finally, recent research into anaerobic zoosporic fungi inhabiting digestive tracts of ruminant mammals has revealed a potential for efficient degradation of cellulose, currently utilized as a biofuel conversion target.

Zoosporic fungi comprise the Chytridiomycota, Blastocladiomycota, Neocallimastigomycota, microsporidia, and Cryptomycota. The true nature of the phylogenetic relationships between these lineages is poorly resolved at best. However, zoosporic fungi are ubiquitous in the environment, and the diversity of these basal lineages, particularly within the Cryptomycota, is estimated to rival that of the rest of the fungal kingdom. Because they are largely uncultivated and thus understudied, these lineages are sometimes referred to as “Dark Matter Fungi” (DMF). Traditional high-throughput sequencing is poorly suited for the genomic characterization of DMF, since it relies on large amounts of DNA extracted from many cells of the same species. Single-cell genomic sequencing, however, aims to use individual cells to reconstruct genomes of uncultivated organisms directly isolated from the environment. Therefore environmental DMF make exceptional targets for single-cell genomic techniques. Much of the current single-cell genomics work focuses on mammalian, as well as bacterial and archaeal systems, underscoring the need to develop protocols for fungi.

Primary goals for this research include the development of a standard production workflow for fungal single-cell genomes. Such protocols employ the use of pre-selection filtering during environmental isolation, lifestage-based staining for target organism enrichment, fluorescence-activated cell sorting (FACS), cell lysis, strand displacement based whole-genome amplification, single-cell optimized assembly protocols (SPAdes), and the fungal annotation pipeline currently in use for fungal genome sequencing at the JGI. Current fungal single-cell targets include Metschnikowia bicuspidata, Rozella allomycis, Piptocephalis cylindrospora, Dimargaris cristalligena, Syncephalis pseudoplumigaleata, Blyttiomyces

helicus, Caulochytrium protosteloides, and Thamnocephalis sphaerospora, all of which have been sequenced and the majority of draft genome assemblies have been annotated. Comparative genomic analyses are underway to address questions regarding the evolutionary biology of DMF, including aspects of pathogenicity and phylogenetic diversity within these lineages.

Metagenomic Analysis Shows Phylogenetic and Functional Similarities Across Diverse Soil Microhabitats

Starr, Evan P.* (starr.evan@berkeley.edu)[1], Sharrar, Allison M.[1]; Firestone, Mary K.[1]; Banfield, Jillian F.[1]

[1]University of California, Berkeley, Berkeley, California.

Soils are one of the most diverse microbial habitats. Little is known about the structuring of microbial communities through the soil profile and into the subsurface or how these communities regulate carbon and nitrogen cycling. To address these questions, we have obtained genomic information from 26 soil samples taken at the Eel River Critical Zone Observatory in the northern California Angelo Coast Range Reserve. At this study site, the main source of carbon in soil and the subsurface is tree derived carbon from both exudates and dead organic material. Our samples span a variety of environmental parameters, including depth, time since last rainfall, and proximity to root carbon inputs. Six soil samples were obtained from a depth transect going down 1.2 m on a hillslope. Ten subsoil samples are from the same hillslope and were collected by drilling into weathered bedrock at 2 m depth. The holes were then filled with gas concentration samplers for subsequent trace gas analysis. Another ten soil samples were obtained from two sites in a nearby meadow at 20 cm and 40 cm depth, two of which were collected before the first rainfall and the rest at various periods after. To further understand the interface between incoming carbon and the subsurface community, we also sampled from the rhizosphere and bulk (not rhizosphere associated) subsoil of mature Douglas Fir trees (Pseudotsuga menziesii). Metagenomic data from all samples was assembled and draft genomes, including hundreds of partial to near complete genomes, were binned and reconstructed based on time series coverage analysis and tetranucleotide frequency using ggKbase and emergent self-organizing maps. Phylogenetic diversity and overlap through the depth transect and between sampling sites was analyzed using ribosomal protein S3 contig identity and coverage. There was a large amount of phylogenetic overlap across soil depths, with some OTUs being present at every depth, despite numerous environmental differences. There was also a general increase in the relative abundance of nitrogen fixation genes with depth.

Extensive Genomic Reconstructions Provide Insights into Biogeochemical Cycling in a Complex Terrestrial Subsurface Environment

Anantharaman, Karthik* (karthik.a@berkeley.edu)[1]; Brown, Christopher T.[2]; Hug, Laura A.[1]; Sharon, Itai[1]; Castelle, Cindy J.[1]; Probst, Alexander J.[1]; Thomas, Brian C.[1]; Singh, Andrea[1]; Wilkins, Michael J.[3]; Karaoz, Ulas[4]; Brodie, Eoin L.[4]; Williams, Kenneth H.[4]; Hubbard, Susan S.[4]; Banfield, Jillian F.[1, 4]

[1]Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA. [2]Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA. [3]School of Earth Sciences and Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA. [4]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

Microorganisms drive organic compound transformations in the terrestrial subsurface, a major reservoir of carbon on Earth, and mediate linked biogeochemical cycles. Most prior studies have been based on cultivation of a few organisms, profiling of functional genes, and phylogenetic analyses to provide metabolic insights needed for ecosystem modeling. The limitations of these methods arise from large gaps in our knowledge about microorganisms in the subsurface, and because phylogeny does not reliably predict physiology. New approaches that provide genomes from metagenomes, even for highly complex environments, enable direct organism-resolved understanding of biogeochemical processes. For sediment and groundwater-associated bacteria from a shallow sediment-hosted aquifer, we reconstructed 2,540 high-quality and complete genomes from 1,297 distinct species. Remarkably, this single ecosystem yielded genomes from the majority of all known bacterial phyla, as well as from 47 newly discovered phyla. These genomes accounted for up to 36% of all organisms that have ever been detected in this subsurface aquifer environment to date. Analyses of abundance patterns of these organisms over space and time indicated dynamic responses to changing conditions, with newly abundant species rising from an extraordinarily diverse “microbial seed bank”. Based on the combinations of traits associated with carbon, nitrogen, hydrogen and sulfur redox chemistry in specific organisms, we resolved aspects of biogeochemical networks that are connected by metabolic handoffs between organisms. We suggest that generally little understood microorganisms functioning collectively in community context are largely responsible for turning Earth’s subsurface biogeochemical cycles.

This work was supported by Lawrence Berkeley National Laboratory’s Sustainable Systems Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under contract DE-AC02-05CH11231. DNA sequencing was conducted at the DOE Joint Genome Institute, a DOE Office of Science User Facility, via the Community Science Program.

Construction and Annotation of a Representative Community-Based Culture Collection of Sugarcane

Armanhi, Jaderson SL* (jader.armanhi@gmail.com)[1]

Souza, Rafael S.C.[1]; Araújo, Laura M.[1]; Okura, Vagner K.[1]; Mieczkowski, Piotr[2]; Imperial, Juan[3, 4]; Arruda, Paulo[1, 5]

[1]Center for Molecular Biology and Genetic Engineering (Unicamp), Campinas, São Paulo/Brazil. [2]Department of Genetics (UNC), Chapel Hill, North Carolina/USA. [3]Center for Biotechnology and Genomics of Plants (UPM), Madrid/Spain. [4]Superior Council of Scientific Investigations, Serrano, Madrid/Spain. [5]Department of Genetics and Evolution (Unicamp), Campinas, São Paulo/Brazil.

Advances in metagenomics sequencing methods have linked the vast microbial diversity to assorted biological functions in human health, plant development and biotechnological processes. However, isolation, cultivation and identification of microbial candidates are imperative steps for the emerging area of microbial community manipulation and discovery of its biological function. Traditionally, the isolation of representative set of microorganisms from a given environment requires several rounds of

picking and streaking to obtain pure colonies. These methods are time-consuming, costly, may not be appropriate to estimate the recovery of microbial diversity and may result in losses of relevant biological information, such as microorganisms that might depend strictly on microbe-microbe interactions for their growth. Here we present the concept of community-based culture collections (CBC) that is based on the isolation and storage of multiple microorganisms in a single plate well. To annotate the community-based isolates we developed a multiplex strategy sequencing of full-length microbial 16S rRNA amplicons from 96-well plates pooled down to a single tube. Two-step PCR amplifications were designed to add individual barcodes for the identification of each plate and well. This multi-tagging system, coupled to PacBio SMRT sequencing, allows tracing back individual sequences to their original wells regardless of whether the well contains single or multiple microorganisms. The methods were developed and used to annotate a sugarcane CBC. The sugarcane CBC was prepared by picking primary colonies from platings of enriched microbial fractions of rhizosphere, endophytic root, and endophytic stalks. We were able to precisely identify the CBC microorganisms that could then be cross-referenced with metagenomics data for the estimation of microbial community recovery and selection of microbial candidates. Even bypassing the laborious purification steps our strategy allowed accessing axenic cultures, as 60,2% of the wells contained single isolated microbes. The sugarcane CBC contains representatives that sum 17.1%, 17.9% and 62.7% of the total rhizosphere, endophytic root and endophytic stalk bacterial diversity found in sugarcane, respectively. Although we aimed to isolate root and stalk endophytic microorganisms, our collection interestingly contained representatives that sum 58.7% and 83.6% of the exophytic and endophytic leaf diversity, respectively. The CBC construction and annotation methods described here can be used to recover larger fractions of microbiota and preserve putative microbe-microbe interactions. Moreover, the method can be used for construction of CBC from any type of environment as it overcomes major limitation of current microbiome research related to microbiota culture collection.

Development of Plants with Multiple Traits for High Yield of Fermentable Sugars

Aznar, Aude* (aaznar@lbl.gov)[1]; Chalvin, Camille[1, 2]; Shih, Patrick[1]; Birdseye, Devon[1]; Loqué, Dominique[1]; Scheller, Henrik V.[1, 3]

[1]Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, California 94720. [2]Ecole Normale Supérieure de Cachan, 94230 Cachan, France.[3]University of California, Department of Plant and Microbial Biology, Berkeley, CA, 94720.

Second-generation biofuels produced from biomass can help to decrease dependency on fossil fuels, which would have many economical and environmental benefits. To make biomass more suitable for biorefinery use we need a better understanding of plant cell wall biosynthesis. Increasing the ratio of C6 to C5 sugars in the wall is an important target for engineering of plants that are more suitable for downstream processing for second-generation biofuel production. Likewise, decreasing the content of lignin is an important goal. We have studied the basic mechanisms of cell wall biosynthesis and identified genes involved in biosynthesis of pectic galactan including the GALS1 galactan synthase[1] and the URGT1 UDP-galactose transporter[2]. We have applied these findings to engineer plants that have a more suitable biomass composition and have developed synthetic biology and gene stacking tools to achieve this goal. Plants were engineered to have up to three-fold increased content of pectic galactan in stems by expressing GALS1, URGT1 and a UDP-glucose epimerase. Furthermore, the increased galactan was engineered into plants that were already engineered to have low xylan content by restricting xylan

biosynthesis to vessels where this polysaccharide is essential[3]. Finally, the high galactan and low xylan traits were stacked with low lignin obtained by expressing the QsuB gene encoding dehydroshikimate dehydratase[4]. By targeting the transgene expression to specific cell types, we could substantially improve saccharification while avoiding adverse effects on plant growth and development.

1. Liwanag, A.J., et al., Pectin biosynthesis: GALS1 in Arabidopsis thaliana is a beta-1,4-galactan beta-1,4-galactosyltransferase. Plant Cell, 2012. 24(12): p. 5024-36.
2. Rautengarten, C., et al., The Golgi localized bifunctional UDP-rhamnose/UDP-galactose transporter family of Arabidopsis. Proc Natl Acad Sci U S A, 2014. 111(31): p. 11563-8.
3. Petersen, P.D., et al., Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. Biotechnology for Biofuels, 2012. 5(1): p. 84.
4. Eudes, A., et al., Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. Plant Biotechnol J, 2015.

This work was part of the DOE Joint BioEnergy Institute (http:// www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

Engineering RNA Aptamer-Based Gene Circuits for Control of Microbial Interactions

Beliaev, Alex S.* (alex.beliaev@pnnl.gov), McClure, Ryan S.

Pacific Northwest National Laboratory, Richland, WA.

The ability to precisely and accurately control genetic circuits has far reaching implications for the engineering of robust and safe microbial systems with desired outputs. To achieve that goal, we propose to develop a scalable and generalizable pipeline for the development of functional genetic circuits that respond to specific target ligands using RNA-based devices, specifically riboswitches, which induce downstream gene expression as a function of target ligand concentration. Here, previously designed RNA sequences that bind to two small molecule ligands, p-aminophenylalanine (pAF) and theophylline were used as starting material. These sequences, termed aptamers, were cloned upstream of 15 nucleotide random sequences driving expression of a tetracycline resistance (tetA) readout gene. In a subset of these random sequences presence of the ligand causes a structural change in the aptamer such that the ribosome binding site of tetA is available for binding. Duplicate cultures of such aptamer/switch libraries were grown either in the presence or absence of the ligand (pAF, or theophylline). Functional aptamer/switch pairs, riboswitches, were therefore enriched only in cultures containing the ligand. Sequencing of the resulting cultures identified several hundred switch sequences that were > 2-fold more abundant in response to the presence or absence of the ligand. Such changes in abundance represent putative successful riboswitches. These studies are currently being expanded and refined so that enrichment and selection of functional riboswitches can take place within a single community of continuously cultured bacteria. Additional experiments have been aimed at increasing the modularity if riboswitches through the inclusion of fusion proteins containing the gene the riboswitch was originally evolved with. This has resulted in a fusion riboswitch evolved to induce expression of beta-galactosidase modified into a riboswitch designed to induce the expression of tetA. This approach of rapidly and efficiently screening putative riboswitches, as well as increasing their modularity, has several applications as aptamers for

new ligands are developed and synthetic genetic circuits become cornerstones of engineered microbial consortia of bio-industrial importance.

Microfluidic Integration for Genomic Analysis

Blainey, Paul C.* (pblainey@broadinstitute.org); McCarthy, Francis (francism@broadinstitute.org)

Broad Institute of MIT and Harvard, Massachusetts Institute of Technology.

Sample preparation is now the most challenging part of many genomic analyses. The complexity and inefficiencies of sample preparation procedures commonly limit the number and types of samples that can be analyzed using sequencing-based approaches. In our JGI ETOP project, we are developing and deploying lab-on-chip microfluidic systems with the capability for complete integration and automation of sample preparation procedures for sequence-based analyses. In other work, we are integrating additional functionality such as cell-level target enrichment into the lab-on-chip workflows and pioneering novel single-cell methods for microbial community analysis.

Insights into the Function and Evolution of Bacterial Microcompartments Prompted by Bacterial Genomes Sequenced by the Joint Genome Institute

Blanchard, Jeffrey, L.* (jeffb@bio.umass.edu); Haas, Kelly; Gann, Eric; Lawson, Shane; Bakhtyari, Ava; Iqbal, Fauzia; Kianfar, Nadia; Mucci, Nicholas; Rottenberg, Jaice; Biddle, Amy

University of Massachusetts Amherst.

Bacterial microcompartments (BMCs) are intracellular organelles composed of a proteinacous shell that encapsulates metabolic products. These structures are icosahedral in shape, akin to bacterial phages and capsid viruses. BMCs are homologous in sequence and structure to carboxysomes found in all cyanobacteria and some chemoautotrophs that encapsulate the enzymes carbonic anhydrase and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), in order to concentrate CO_2 in close proximity to RuBisCO, eliminating the competitive reaction with oxygen. While carboxysomes are present in all cyanobacteria, BMCs are present ~15% of sequenced bacterial genomes and are subject to frequent horizontal gene transfer and loss leading to a sporadic distribution across bacteria. The metabolism of BMCs is more varied, providing for the metabolism of ethanolamine, 1,2-propanediol, 1,2-ethanediol, choline, fucose (via 1,2-propanediol), and rhamnose (via-1,2 propanediol). All of these anaerobic pathways share acetaldehyde as an intermediate and a possible function of the BMC is to prevent cellular exposure to toxic levels of acetaldehyde. The substrates from BMC metabolism are derived from algal, plant and animal cell wall lipids (Phosphatidylethanolamines and Phosphatidylcholine) and carbohydrates (Fucose and Rhamnose) and thus influence carbon and nitrogen cycles in many anaerobic environments.

A novel type of BMC was identified in genomes of isolates from a switchgrass microcosm inoculated with forest soil. The genetic locus for this microcompartment contains several genes not previously identified in BMC loci, as well as the absence of diagnostic key enzymes for known BMC metabolic processes. BMC

genetic loci often contain a large operon consisting of 10-20 genes (among the largest operons known). The rarity of the novel BMC type in bacterial genomic databases coupled with it's limited distribution in environmental samples suggests a recent evolutionary origin, providing insights into the steps in assembling a novel BMC operon and it's subsequence optimization.

Additional genome sequencing efforts by the JGI on isolates from the switchgrass microcosm related to the biofuel catalyst Clostridium phytofermentans ISDg revealed minor differences in gene content between species members, but a notable exception was the absence of a BMC locus hypothesized to be involved in choline metabolism. Clostridium phytofermentans ISDg was able to grow with choline as the sole source of carbon with trimethylamine (TMA) resulting as an end product. TMA is often subsequently metabolized by methanogenic bacteria in marine sediments and animal intestines. Closely related strains of C. phytofermentans ISDg lack this BMC locus and are unable to ferment choline thus providing a "phylogenetic knockout" in support of the function of this pathway in C. phytofermentans ISDg. The unusual presence of three distinct BMCs utilizing choline, ethanolamine, fucose and rhamnose allows C. phytofermentans ISDg to metabolize major cell wall components and suggests complex host interactions.

Evaluation of Unassembled DNA Regions from Illumina and PacBio Technologies and Finishing Microbial Genomes

Brown, Steven D.* (brownsd@ornl.gov)[1, 2]; Utturkar, Sagar M.[1]; Klingeman, Dawn M.[2]; Hurt, Richard A.[2]; Huntemann, Marcel[3]; Pillay, Manoj[3]; Palaniappan, Krishnaveni[3]; Varghese, Neha[3]; Mikhailova, Natalia[3]; Stamatis, Dimitrios[3]; Reddy, T.B.K.[3]; Ngan, Chew Yee[3]; Daum, Chris[3]; Shapiro, Nicole[3]; Markowitz, Victor[3]; Ivanova, Natalia[3]; Kyrpides, Nikos[3]; Woyke, Tanja[3]

[1]Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, Tennessee 37919, USA. [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. [3]DOE Joint Genome Institute, Walnut Creek, California 94598, USA.

Background: Development of next generation sequencing (NGS) technologies has revolutionized genomics research by providing high-throughput, low-cost sequencing methods. Despite extensive sequencing and assembly advances, there are several examples of microbial genomes that remain unfinished by PacBio and Illumina platforms even at the coverage levels estimated in range of 50x to 296x. The aim of the present study was to reveal and characterize regions of DNA which remained unassembled by either by individual or both technologies.

Approach: We sequenced the genomes of Clostridium thermocellum AD2, Clostridium pasteurianum ATCC 6013, Clostridium autoethanogenum DSM 10061, Clostridium paradoxum JW-YL-7, Pelosinus fermentans UFO1, Pelosinus fermentans JBW45, Halomonas sp. KO116 and Bacteroides cellulosolvens DSM 2933 using the Illumina MiSeq and PacBio RS-II platforms. *De novo* and hybrid assemblies were performed using various data combinations and software. Comparison of complete and draft genomes was performed to reveal the nature of the gaps associated with Illumina technology. Two genomes, which could not be automatically finished using either NGS data were manually finished using bioinformatics and PCR/Sanger sequencing. Sanger sequence data was utilized to analyze the gaps within the PacBio assemblies. Genome polishing was performed using quiver, pilon and circulator software.

Results: The ABySS and SPAdes assemblers obtained high quality draft assemblies using only the Illumina data, hybrid methods obtained better assembly statistics while the HGAP method with PacBio only data

obtained the best results (five automatically finished genomes). Manual finishing successfully upgraded the near-finished assemblies of two genomes up to finished status. Assembly comparisons revealed the gaps within Illumina assemblies were mostly corresponding to repetitive DNA elements such as rRNA operons. PacBio gaps were assessed for overall read-depth, read-quality, sample DNA quality, DNA structure and specific biological features of the organisms such as presence of active mobile genetic elements, repetitive transposon elements and plasmid sequences without a single predictor of sequence gaps emerging. Post-assembly polishing steps obtained substantial improvements in overall gene calling accuracy and annotations. For example, the pilon software suggested 314 modifications across four finished genomes of which 183 (58%) have improved gene calling accuracy, 96 (31%) have no change, and 35 (11%) have deteriorated gene calls. The accuracy of 47 random pilon calls was further validated by PCR/Sanger sequencing approach which revealed 40 (85%) pilon calls were accurate. Comparison of draft and final polished finished assemblies revealed substantial number of longer proteins (9 to 342), and several new proteins (1 to 20) in improved genomes.

Conclusion: Using manual finishing, we successfully upgraded near-finished genomes up to finished quality. The importance of assembly polishing steps in terms of gene-calls and consensus accuracy was demonstrated. Despite some limitations, PacBio is currently one of the best available method for microbial genome sequencing. Assembly automation is already becoming integral part of PacBio sequencing and it is highly recommended to perform minimal manual inspection of assemblies which may obtain substantial improvements in overall quality.

Supercomputing and Genome Assembly: Improved Reconstruction of Illumina Sequencing Data

Browne, Daniel R.* (dbrowne.up@gmail.com)[1]; Flowers, Dave[2]; Plott, Christopher[2]; Dickens, Michael[1]; Daum, Chris[3]; Barry, Kerrie[3]; Aramayo, Rodolfo[1]; Schmutz, Jeremy[2]; Devarenne, Timothy P.[1]

[1]Texas A&M University. [2]HudsonAlpha Genome Sequencing Center. [3]US Department of Energy Joint Genome Institute, Walnut Creek, California..

High performance computing hardware was utilized to assemble the genome of Botryococcus braunii race B (a colony-forming, oil-producing green microalga) with the ABySS, CD-HIT, BESST, and REAPR programs. The ABySS algorithm assembles sequence by constructing, processing, and traversing a de Bruijn graph of the k-mer set derived from the input Illumina reads. For this stage of the assembly, we used 500 million 250 bp paired end Illumina reads obtained by sequencing an 800 bp DNA fragment library. With an estimated genome size of 166 Mbp, this library yielded approximately 750X coverage of the genome. Two 150 bp mate pair libraries were also generated, with insert sizes of 1.5 kb and 4 kb, at 261X and 182X coverage respectively. We hypothesize that disparate regions in the genome will assemble optimally at different sizes of k-mer used for de Bruijn graph construction by ABySS. Therefore to obtain a complete assembly, one must compute and consolidate multiple assemblies across the range of k-mer sizes available up to the read length. The complete set of sequences from ABySS was consolidated with CD-HIT. The consolidated contigs were re-assembled with an overlap-layout-consensus (OLC)-type approach using algorithms in the ABySS toolkit. The assembly was then scaffolded using the mate pair libraries and the BESST scaffolding algorithm. Mis-assembled scaffolds were detected and removed with REAPR. This novel approach to genome assembly resulted in a more contiguous and higher quality assembly than any other known method, given

only three Illumina libraries. Furthermore, our work highlights the growing usefulness and importance of supercomputing platforms for *de novo* genome assembly.

Global Ecology and Ecosystem Effects of Marine Viruses

Brum, Jennifer R.* (brum.11@osu.edu); Roux, Simon; Sullivan, Matthew B.

The Ohio State University, Columbus, Ohio, USA.

Marine viruses have important roles in microbial mortality, gene transfer, metabolic reprogramming and biogeochemical cycling. However, methodological limitations have previously prevented a quantitative assessment of their community structure and ecosystem impacts. Recent transformative advances have led to construction of a sample-to-sequence pipeline that enables a quantitative assessment of marine viral communities using metagenomic techniques, and has facilitated a rapid increase in knowledge of marine viral ecology and their impacts on oceanic ecosystem function. Here we focus on recent studies that have greatly advanced our knowledge of marine viruses, including (1) a global-scale assessment of marine viral community structure and the factors that drive this variability, (2) a study that connects this global-scale viral metagenomic data to the roles of viruses in the oceanic carbon cycle, and (3) further methodological advances, including bioinformatic approaches and metaproteomics, that are enabling us to illuminate both taxonomic and functional viral 'dark matter' that dominates environmental viral metagenomes.

Multi-'Omic' Analyses of the Dynamics, Mechanisms, and Pathways for Carbon Turnover in Grassland Soil

Butterfield, Cristina N.* (butterfc@berkeley.edu)[1]; Diamond, Spencer[1]; Li, Zhou[2]; Andeer, Peter[3]; Spaulding, Susan[1]; Burstein, David[1]; Tringe, Susannah G.[5]; Thomas, Brian C.[1]; Hettich, Robert[2]; Northen, Trent R.[3]; Pan, Chongle[3]; Banfield, Jillian F.[1]

[1]University of California, Berkeley, Earth and Planetary Science. [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee. [3]Lawrence Berkeley National Laboratory, Berkeley, California. [4]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, California. [5]US Department of Energy Joint Genome Institute, Walnut Creek, California.

Climate change will alter terrestrial ecosystems. However, the strength and the direction of change will be shaped by feedbacks, most of which will be difficult to predict. Of primary importance in this regard is how the distribution of carbon between the atmosphere and the subsurface will change in response to altered rainfall, temperature and vegetation patterns. Metagenomics, proteomics, transcriptomics, and metabolomics are used to compare the membership and functioning of soil communities at three different depths below the root zone in a grassland that experiences a Mediterranean climate. We track microbial community composition and activity during the period of major carbon turnover in this ecosystem under two rainfall scenarios, identify key carbon currencies released from the soil zone and provide a basis for prediction of how grassland ecosystems will respond to future climate change.

Effects of Size, Season and Salinity on Composition and Function of Estuarine Communities

Campbell, Barbara J.* (bcampb7@clemson.edu)[1]; Castelloe, Emily[1]; Painter, Joseph[1]; Lim, Jean, S.[1]; Gonsior, Michael[2]; Kirchman, David L.[3]

[1]Department of Biological Sciences, Clemson University, Clemson, SC. [2]University of Maryland Center for Environmental Sciences, Chesapeake Biological Laboratory, Solomons, MD. [3]School of Marine Science and Policy, University of Delaware, Lewes, DE.

Bacterial contributions to biogeochemical processes in the environment are not well understood in part because we know little about who is active in the ecosystem. We used a paired metatranscriptomic and metagenomic approach to show how functions change over time and space in the Delaware Bay. Size fractionated (>0.8 µm and <0.8 µm) surface water samples were analyzed from three different salinities (0, 15, 30) and three different seasons (March, August, November). We also analyzed water samples by high resolution mass spectrometry to characterize differences in DOM content. The abundance of different phyla or COG categories were estimated via their relative levels in metagenome and metatranscriptome samples. Bacterial Phyla were also estimated by 16S rRNA tag sequencing. There was an expected gradient of phyla along the salinity gradient, with Actinobacteria dominating at low salinity sites and Bacteroidetes and Proteobacteria dominating at higher salinity sites. By far, the most effect on phyla in the active community (metatranscriptome) was size fraction, where Cyanobacteria, Firmicutes and Verrucomicrobia were more abundant in the >0.8 µm fraction and Proteobacteria were more abundant in the <0.8 µm. Not many large differences were seen between seasons at this phylogenetic resolution. Higher resolution analysis with 16S rRNA tag sequencing showed similar trends. When analyzed by COG category, all metagenomes grouped together and were distinct from the metatranscriptomes. Metatranscriptomes, on the other hand, grouped by size fraction, then season and salinity. Functions overrepresented in the >0.8 µm fraction were Energy production and conversion, Translation, ribosomal structure and biogenesis, Posttranslational modification, protein turnover, chaperones and Signal transduction mechanisms. Functions significantly increased in the <0.8 µm fraction were Amino acid transport and metabolism, Nucleotide transport and metabolism, Coenzyme transport and metabolism, and Cell wall/membrane/envelope biogenesis. DOM content along the estuarine gradient mainly differed by season then salinity, with polyphenolic acids and sulfonates dominating freshwater sites and other sulfur compounds dominating marine sites. Therefore, size, season and salinity matter to composition and function of the microbial communities in the Delaware Bay, indicating differences in microbial contributions to ecosystem services such as carbon and sulfur cycling.

Metagenomic Characterization of Nitrogen-Cycling Microbial Communities Impacting Uranium Release in the Upper Colorado River Basin

Cardarelli, Emily L.* (ecardare@stanford.edu)[1]; Bargar, John R.[2]; Francis, Christopher A.[1]

[1]Department of Earth System Science, Stanford University, Stanford, CA. [2]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA.

Extensive uranium (U) groundwater contamination at DOE's legacy ore processing sites has resisted extensive remediation attempts. The most egregious of these sites lie within the upper Colorado River Basin on siliciclastic floodplains. While these sediments are generally nutrient-poor, large inventories

of nutrients and uranium [as low-solubility U(IV)] are held within low-volume organic-rich fine-grained sediment lenses. Iron sulfide minerals are abundant within these layers, giving rise to the moniker "naturally reduced zones" (NRZs). Sharp biogeochemically active oxic-anoxic transitions reside in contact with the surrounding (sub-)oxic aquifers. There is concern that NRZs are acting as slow-release sources of uranium to the aquifer that could persist for hundreds of years. They are believed to dominate floodplain redox cycling and micronutrient supply.

Nitrate, produced by the decomposition of organic matter coupled to nitrification, is suspected of playing a central role in NRZ-floodplain biogeochemistry by (i) driving heterotrophic carbon cycling in NRZs and (ii) oxidizing the stored U(IV) to relatively mobile U(VI), which makes possible its escape to the surrounding aquifer. Nitrate has a relatively high solubility and diffuses into NRZs fairly easily. Moreover, nitrate and denitrification products are strong oxidants for U(IV). In contrast, molecular oxygen has much lower solubility and reacts strongly with sulfides, preventing penetration into NRZs. The microbial nitrogen cycle is thus believed to be fundamentally important to the mobility and fate of organic carbon and uranium. However, very little is currently known regarding the N-cycling communities within NRZ sediments.

Here, we present clone library and qPCR analyses of key N-cycling functional genes in DNA extracts obtained from 5 DOE-Legacy Management sites (Rifle, Grand Junction, and Naturita, CO; Shiprock, NM, and Riverton, WY). In particular, we have examined the diversity and abundance ammonia-oxidizing and denitrifying organisms, using amoA (encoding the α-subunit of ammonia monooxygenase) and nirK/nirS (encoding dissimilatory nitrite reductase), respectively, as molecular markers, providing the first ever basin-wide insights in N-cycling microbial communities. Interestingly, there appear to be distinct 'ecotypes' associated with different depths and geochemical conditions throughout the sediment cores and among sites. For this CSP project, we plan to combine functional gene approaches, 'deep' 16S rRNA amplicon sequencing, metagenomic and metatranscriptomic sequencing, as well as extensive geochemical data, to acquire novel insights into the phylogenetic, genomic, and functional/metabolic diversity of N-cycling microbial communities within U-contaminated NRZs.

Comparative Genomics of Snow Algae *Chlamydomonas cribrum* and *Chloromonas nivalis*

Castruita, Madeli* (madeli.c@gmail.com)[1]; Leya, Thomas[2]; Merchant, Sabeeha[1, 3]

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, California, USA. [2]Extremophile Research & Biobank CCCryo, Fraunhofer Institute for Cell Therapy and Immunology, Potsdam-Golm, Germany. [3]Institute of Genomics and Proteomics, University of California, Los Angeles, California, USA.

Temperature is one of the most important abiotic determinants of the vitality, viability and distribution of life forms and therefore the biology of cold-adapted organisms is of broad general interest. Although the polar and alpine regions constitute 25% of Earth's land area, the biology of the region which has an impact on CO_2 sequestration and climate change is under –explored. Eukaryotic algae, commonly called 'snow algae' are among the organisms adapted to the harsh conditions in the polar regions. Snow algae represent important primary producers in an extreme environment; very low temperatures combined also with pH extremes, desiccation, low nutrient and high irradiation. We have chosen two species, Chloromonas nivalis and Chlamydomonas cribrum for genomic studies because they are related to the well-studied reference organism Chlamydomonas reinhardtii. Chlo. nivalis and Chla. cribrum

are true psychrophiles, isolated from samples collected near Spitspergen in the Svalbard archipelago. Chlo. nivalis and Chla. cribrum are distinguished morphologically by the presence of a pyrenoid (starch containing intra-chloroplast compartment, which also contains Rubisco and hence the CO_2 assimilation system) in the latter (Buchheim et al., 1997). As the site of CO_2 fixation and the target of the algal carbon concentrating mechanism, there is broad interest in the pyrenoid; the sequence of two genomes may allow us to use a comparative approach for dissection of pyrenoid function and biogenesis. Our goal for generating draft genomes for these organisms is to identify novel pathways and molecules that have evolved in these organisms for dealing with particular stresses, including high light, UV radiation, osmotic changes, dehydration, and poor nutrition.

Oligotyping and Metagenome Analyses Reveal a High-Resolution Recurrence that Frames Potential Ecological Strategies of North Sea Bacterioplankton

Chafee, Meghan* (mchafee@mpi-bremen.de)[1]; Krüger, Karen[1]; Kappelmann, Lennart[1]; Waldman, Jost[1]; Reintjes, Greta[1]; Lucas, Judith[2]; Wichels, Antje[2]; Gerdts, Gunnar[2]; Wiltshire, Karen[2]; Glöckner, F.O.[1]; Fuchs, Bernhard M.[1]; Teeling, Hanno[1]; Amann, Rudolf I.[1]

[1]Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany. [2] Alfred Wegener Institute for Polar and Marine Research, Biologische Anstalt Helgoland, 27483 Helgoland, Germany.

Temperate coastal marine habitats are replete with complex biotic and abiotic interactions that mediate relationships and niche space between resident microbial community members. The dynamics of these interactions are amplified during spring and summer when phytoplankton form massive blooms and heterotrophic bacterioplankton respond to the successional release of high molecular weight dissolved organic matter (DOM, e.g., proteins, polysaccharides, lipids) as algal cells lyse. Dedicated clades of copiotrophic bacterioplankton adapted to high nutrient loads employ a suite of specialized enzymes and transporters for the degradation and uptake of large complex molecules like polysaccharides. Functional profiles of these carbohydrate active enzymes (CAZymes) reveal successions of niche spaces determined through substrate availability during bloom periods. As bacterioplankton process these substrates, algal-derived DOM is effectively incorporated into bacterioplankton biomass where it can be re-assimilated by higher protozoans and thus recycled within the food web or rapidly exported to sediments below. The net result is an enhancement of system-level efficiency and carbon export, respectively.

In order to parse potential niche space in such a dynamic system, we first present diversity data derived from high-resolution temporal sampling (weekly to bi-weekly intervals) using the high-resolution oligotyping method Minimum Entropy Decomposition (MED) to cluster OTUs from 3 replicated sampling years 2010-2012. Differential abundance of MED OTUs >99% similar reveals a granularity of potential niche resolution that would not have been retrieved by traditional OTU clustering. Implementation of MED effectively extracts only those polymorphisms most relevant to in situ environmental selection pressure, essentially providing an ecological fingerprint of resident North Sea bacterioplankton. Despite inter-annual variation in phytoplankton blooms, we see a population of recurrent heterotrophic bacterioplankton during both baseline and bloom periods that is remarkably constrained in composition, relative community abundance and time of appearance each year for both abundant and rare OTUs.

These robustly recurrent patterns reflect the selective power of seasonal forcing in shaping temperate microbial communities with low-frequency temperature-driven seasonal shifts.
Superimposing effects of temperature are higher frequency shifts in OTUs during dynamic bloom events when substrate-induced forcing drives copiotrophic bacterioplankton communities, particularly within the Gammaproteobacteria and Flavobacteriia. Metagenome data from monthly sampling during spring blooms from these same years supports 16S rRNA diversity analyses revealing recurrent groups of specialized bloom taxa with highly constrained repertories of CAZymes. We demonstrate that even though there is substantial inter-annual variation of phytoplankton bloom intensity and taxonomic composition, the accompanying succession of bacterial clades is not a purely stochastic process, but also governed by deterministic principles such as temperature and substrate-induced forcing. The result is a resident bacterioplankton community containing as few as 6-14 dominant taxa each spring responsible for a vast majority of the diversity and thus potential function in a system previously thought to be considerably more stochastic and complex.

Diazotrophic Endophytes in Bioenergy Crops

Chakraborty, Romy* (rchakraborty@lbl.gov)[1]; Schicklberger, Marcus[1]; Masuno, Duy[1]; Jenny, Pang[1]; Jenkins, Stefan[1]; Northen, Trent R.[1]; Loqué, Dominique[1]; Shapiro, Nicole[1, 2]; Woyke, Tanja[1, 2]

[1]Lawrence Berkeley National Lab. [2]US Department of Energy Joint Genome Institute, Walnut Creek, California..

Nitrogen availability is widely recognized to be one of the primary nutrient constraints on plant growth and productivity. Some plants such as legumes have developed unique symbiotic relationship with N-fixing bacteria for a sustainable supply of fixed N for their growth. Free-living N-fixing microorganisms (diazotrophs) associated with plants can form a relationship with non-leguminous plants, but our understanding of the mechanisms involved in such plant-microbe interactions are relatively poorly understood.

Our research focuses on diazotrophic endophytes in bioenergy biomass plants- Switchgrass (Panicum vergatum) and Tobacco (Nicotiana tabacum). We successfully isolated and cultivated phylogenetically diverse endophytic and root-associated diazotrophs from stem, leaves, roots and rhizoplane. N-fixation by the isolates was confirmed by the presence of nifH gene using specific primers, and nitrogenase activity was verified using the well-established acetylene reduction assay. The genomes of some of these strains have now been sequenced by JGI. To validate real-time N-fixation by isolated endophytic diazotrophs in planta, engineered fluorescent reporter strains were designed to locate zones of activity. Re-introducing N-fixing isolate Raoultella terrigena R1Gly into axenic Switchgrass seedlings showed a significant increase in both above ground and below ground plant biomass. In addition to N-fixation, the genome of the strain showed presence of ipdC, the key gene for tryptophan dependent indole-3-acetic acid synthesis, suggesting an additional plant-growth-promoting activity through secreted secondary metabolites. This function was confirmed using biochemical assays in the laboratory. Additionally, genomic analysis indicates that the genome harbors genes suitable for adaptation to an endophytic lifestyle.

Detailed knowledge of key genetic and phenotypic characteristics of endophytic diazotrophs, and understanding of interactions between diazotrophic endophytes and host plant will eventually allow

us to design robust strategies to enhance plant biomass using microbial N-fixation, decreasing our dependency on fertilizers.

eQTL Mapping in *Populus*

Chen, Jay* (chenj@ornl.gov)[1]; Yang, Yongil[1]; Muchero, Wellington[1]; Ranjan, Priya[1]; Weston, David[1]; Yang, Xiaohan[1]; Jawdy, Sara[1]; Gunter, Lee[1]; Lipzen, Anna[2]; Lindquist, Erika[2]; Barry, Kerrie[2]; Schmutz, Jeremy[2, 3]; Tuskan, Gerald A.[1, 2]

[1]Oak Ridge National Laboratory, Oak Ridge, TN. [2]US Department of Energy Joint Genome Institute, Walnut Creek, California.[3]HudsonAlpha Institute for Biotechnology, Huntsville, AL.

Fuels developed from lignocellulosic biomass offer a potential renewable and clean alternative to conventional fossil fuels. Populus is one of DOE's "flagship" plant species that is of special interest as a biofuel feedstock. Hybrids among different Populus species are widely used for biomass production. However, significant biological and technological barriers need to be overcome to achieve cost-effective, sustainable production and conversion of Populus biomass into biofuels. Our understanding of biomass productivity and conversion is limited by the fact that these complex traits require the regulation and coordinated interactions of many genes. However, identifying such genetic networks remains unaccomplished and is urgently needed to inform genetic improvement of feedstocks for biomass production and conversion. This study leverages the combined power of genomic technologies and genetic mapping to analyze the large-scale RNAseq dataset generated by a DOE Joint Genome Institute Community Science Program project. We performed analysis of RNAseq data on a total of 438 biological samples. Among these samples, 124 samples have replicates. In total, the samples represent 312 unique genotypes of the quantitative trait locus (QTL) mapping pedigree and two parents. We found that the transcripts of a total of 15,144 genes were detected (FPKM value ≥1) in the developing xylem of both parents and progenies. We also found that several hundreds of genes showing large (>50-fold) transcript variation within the mapping pedigree. eQTL mapping efforts are being taken to identify cis-genetic and trans-acting genetic elements as well as genetic networks underlying genome-wide transcript variation.

Sequencing Microcosm Communities Active in Methane Oxidation: New Insights into the Physiology and Activities of Major Players

Chistoserdova, Ludmila* (milachis@uw.edu)[1]; Hernandez, Maria E.[2]; Matsen, Janet B.[1]; Beck, David A.[1]

[1]Department of Chemical Engineering, University of Washington, Seattle WA. [2]Biotechnological management of resources network, Institute of Ecology A.C. Xalapa, Veracruz, Mexico.

We will present results from a recent CSP project in which we sequenced a time series of metagenomes of methane-oxidizing microcosm communities incubated under two different partial tensions of oxygen ('high' and 'low'), with four replicates for each condition, resulting in a total of 88 metagenomic datasets, paired with 88 corresponding metatranscriptomic datasets. Not only do these datasets provide insights into the physiology/activities of the major species, but they also provide a more realistic, less biased measure of species relative abundances compared to 16S rRNA gene profiling. What is

instantly obvious from metagenome analysis is that prior community profiling through pyrotagging or iTagging has overlooked some species, due to shallower sampling or due to amplification bias. Based on metagenomic analysis, in each microcosm, in addition to Methylococcaceae and Methylophilaceae, bacteria of the order Burkholderiales were prominently present, typically distributed among multiple families/genera, one of the most frequently identified being Acidovorax. Bacteroidetes were also detected in each microcosm, as in the pyrotag/iTag profiling experiments. However, in addition to Flavobacteriaceae, Cytophagaceae and Sphingobacteriaceae were also detected. Importantly, predatory bacteria were detected in some microcosms, belonging to Bacteriovoraxaceae and Myxococcaceae, and increases in their relative abundances seemed to correlate with drops in abundances of Methylococcaceae, suggesting one additional factor that may potentially control methane-oxidizing species in natural habitats. Given high sequence coverage for the abundant species in the metagenomes, we were able to bin multiple genomes representing specific strains within each major functional category. These include Methylobacter, Methylosarcina, Methylovulum, Methylomonas, Methylotenera, Methylophilus, Acidovorax, Bacteriovorax and others. Metatranscriptomic reads were mapped onto these genome bins as well as onto relevant isolate genomes, to identify major expressed pathways. Not surprisingly, in the Methylococcaceae and Methylophilaceae, the methylotrophy pathways were most highly expressed. The Methylobacter species also feature relatively high expression of the respiratory nitrate reductase genes, while Methylotenera species express genes for the downstream reactions, suggesting that the denitrification pathway may be shared between the two organisms. None of the Methylobacter genomes or genomic bins so far was found to encode the complete denitrification pathway. Denitrification genes were also encoded in the binned genomes of Burkholderiales and were expressed at relatively high levels, suggesting a functional role in denitrification. The Burkholderiales also expressed genes for acetate metabolism, suggesting that this could be their niche in carbon sharing. The Bacteroidetes encode and express pathways for biopolymer degradation, suggesting that they may feed on exopolysaccharides excreted by both Methylococcaceae and Methylophilaceae and likely on biopolymers resulting from cell lysis. The predators of the Bacteriovoraxaceae family likely specifically attack the Methylococcaceae, and the Myxococcaeae may act as generalist predators and biopolymer consumers. Overall, our data suggest that a significant portion of carbon from methane is shared among these major guilds and that methane cycle is interlinked with the nitrogen cycle through a collective action of these species.

Does Long-Term Soil Warming Affect the Rate of Horizontal Gene Transfer in Bacterial Communities?

Chowdhury, Priya Roy* (roychowdhury@umass.edu); Pold, Grace; Rodriguez, William; Blanchard, Jeff: Kristen DeAngelis

Organismic and Evolutionary Biology, University of Massachusetts Amherst, MA.

Horizontal gene transfer (HGT) among bacteria via mobile genetic elements (MGEs) maintains diversity and facilitates population adaptation in changing environmental conditions. MGEs are DNAs that can move within and between genomes allowing their horizontal transfer across different bacterial taxa. These include DNA transfer by conjugation (via small circular DNA or plasmids), transformation (via uptake of environmental DNA or RNA) and by transduction (via bacteriophages). HGT is thought to be the predominant process that contributes to acquisition of antibiotic resistance genes in bacteria in both clinical as well as environmental samples. Sequencing of complete bacterial genomes in recent times have however confirmed that HGT is not limited only to antibiotic resistance genes, but a large

proportion of bacterial genes are acquired through it. With advance in cultivation-independent DNA based methods, combined with sequence information, studies are now able to characterize the distribution and diversity of MGEs in different environmental samples. Although sequence-based studies have found a link between environmental pollution and abundances of MGE in ecosystems, most studies have focused on antibiotic resistance, heavy metal resistance and xenobiotic degradations in bacterial communities. We are interested to see if and how long term increase in temperature affects the genes transferred by MGEs in soil bacterial communities. Understanding the genes transferred via MGEs may provide a mechanism by which soil microbes adapt to global warming, and provide insight into future microbial feedbacks to the climate system.

In this study, we tested whether long-term temperature stress affects the quantity and diversity of viral particles in soil. This will provide a partial indication of how stress affects the relative abundance of MGE and in turn HGT via transduction in soil microbial communities. Our study site is located in the Harvard Forest LTER in Central Massachusetts, where an in-situ soil warming experiment was established in 1991. We report quantitative and qualitative estimation of viral sequences in 16 metagenomes (MetaG) and 16 metatranscriptomes (MetaT) that were isolated from heated and control plots in Prospect Hill (~ 25 years of heating) located within Harvard Forest. Our data showed that overall there were no differences in the number of viral sequences between the heated and the control plots based on sequence homology. However, functional annotation of the MetaG data revealed that MGEs decreased from heated to control plots. The most common viral enzymes in our samples were DNA helicases, methylases and polymerases, all of which play important role in viral replication and reproduction. DNA methylase increased while DNA polymerase decreased from control to heated plots. Further, genes related to phosphate metabolism which are highly sensitive to external stress were found in close proximity with the viral sequences suggesting possible targets of HGT. Although, we didn't see an increase in phage sequences in our heated plots, our study is the first to look at the dynamics of MGE-mediated HGT under chronic long-term stress of temperature elevation, which is one of the pressing environmental problems of current times.

Genome-Wide Identification of Bacterial Plant Colonization Genes

Cole, Benjamin* (bjcole@lbl.gov)[1, 2]; Feltcher, Meghan[3]; Waters, Jordan[2]; Wetmore, Kelly[2]; Ryan, Elizabeth[1, 2]; Wang, Natalie[1, 2]; Ul-Hasan, Sabah[4]; Yoshikuni, Yasuo[1, 2]; Malmstrom, Rex[1, 2]; Deutschbauer, Adam[2]; Dangl, Jeff[2]; Visel, Axel[1, 2, 4]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, California., Lawrence Berkeley National Laboratory, Walnut Creek, CA. [2]Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA. [3]Department of Biology, University of North Carolina, Chapel Hill. [4]Quantitative and Systems Biology Department, University of California, Merced, Merced, CA.

Plant-associated bacteria affect diverse aspects of plant growth and health. A thorough understanding of how these microbes colonize and impact on their plant hosts is critical for targeted bioengineering for agricultural purposes, but the molecular mechanisms underlying the colonization process remain incompletely understood. To systematically investigate genetic loci contributing to root colonization *in vivo*, we employed a transposon-mediated genome-wide saturation mutagenesis strategy (Bar-Seq) to study colonization of Arabidopsis thaliana roots by a well-studied Pseudomonas fluorescens strain known to colonize roots and activate Induced Systemic Resistance to leaf pathogens. Among the 5,500 genes in this P. fluorescens genome, we identified a subset of 120 (2.1 %) whose inactivation reproducibly

diminished the ability of mutants to colonize the *A. thaliana* root system without compromising their ability to grow in rich media. Some of these mutations affected molecular functions known to be required for root colonization, such as flagellar genes required for general motility and genes required for carbohydrate or amino acid metabolism, whereas a sizeable subset encode proteins with hitherto unknown functions or with predicted protein functions unrelated to plant association. To further characterize these genes, we examined the same saturation mutagenesis library under 88 distinct in vitro conditions, linking many of the genes required for *in vivo* root colonization to defined biochemical or other basic functions, including previously uncharacterized motility or stress responses. Taken together, our results demonstrate that sequence-driven saturation mutagenesis can inform high-resolution genome-wide maps of complex bacterial *in vivo* functions important to plant ecosystems, and identify a substantial set of plant root colonization genes, offering a starting point for targeted improvement of the colonization capabilities of beneficial microbes.

Drivers of Diversity in the Laurentian Great Lakes

Coleman, Maureen L.* (mlcoleman@uchicago.edu)[1]; Paver, Sara F.[1]; Vargas, Gabriel[1], Newton, Ryan J.[2]

[1]Department of the Geophysical Sciences, University of Chicago, Chicago, IL USA. [2]School of Freshwater Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI USA.

The Laurentian Great Lakes hold 20% of our planet's surface freshwater and supply drinking water to 40 million people. While they are unquestionably important, they also encompass sharp physical, chemical, and ecological gradients, making this system an ideal natural laboratory for studying microbial community ecology and population-level adaptation and diversification. Moreover, the Great Lakes are experiencing significant environmental change, including rapid warming, blooms of toxin-producing cyanobacteria, and food web transformations by invasive species. Yet there have been only a handful of studies documenting the microbial diversity of this system. Our lab has undertaken the first comprehensive microbial characterization across all five Great Lakes. Samples were collected in spring and summer throughout the water column, from several stations per lake, for the past four years. Taxonomic profiling by 16S rRNA amplicon sequencing revealed distinct communities in each lake despite hydrologic connectivity. Microbial assemblages in spring following ice-off were structured by lake with no significant effects of depth or year, while those from stratified summer samples were strongly depth-dependent. Surface communities differed across lakes, season, and years. While many of the most abundant taxa are typical of other freshwater lakes, including Actinobacteria acI, betaproteobacteria, and alphaproteobacteria group LD12, the Great Lakes also harbor abundant novel taxa, including recently described Chloroflexi. To further explore the genomes of these novel taxa, we assembled six metagenomes and obtained 21 near-complete genome bins. Among the recovered genomes are representatives of the newly recognized phylum Ignavibacteriae, diverse Planctomycetes, and nitrifying Nitrospira. Finally, to complement our sequencing approaches, we have enriched and isolated dozens of cyanobacteria and heterotrophic bacteria. Newly isolated cyanobacteria are closely related to Cyanobium and Synechococcus species that form a sister group to marine Prochlorococcus and Synechococcus. We are now characterizing physiology to explore niche partitioning in the Great Lakes, and using comparative genomics to understand how salinity, temperature, light, nutrients, and other factors have driven diversification across marine and freshwater cyanobacteria.

A Novel Phenazine-Reducing Pyocyanin Demethylase Disrupts *Pseudomonas aeruginosa* Biofilms

Costa, Kyle C.* (costak@caltech.edu)[1]; Glasser, Nathaniel R.[1]; Perry, Elena K.[1]; Conway, Stuart J.[4]; Newman, Dianne K.[1, 2, 3]

[1]Division of Biology and Biological Engineering, California Institute of Technology. [2]Division of Geological and Planetary Sciences, California Institute of Technology. [3]Howard Hughes Medical Institute, California Institute of Technology. [4]Department of Chemistry, Chemistry Research Laboratory, University of Oxford.

Pseudomonas aeruginosa is an opportunistic human pathogen that is responsible for increased morbidity and mortality in both acute and chronic infections. A defining feature of this organism is its production of colorful, redox-active metabolites called phenazines, the best studied of which is the blue phenazine, pyocyanin (PYO). PYO can facilitate iron acquisition, redox homeostasis, energy generation and biofilm development for its producer, yet is often toxic to other cells; the biofilm lifestyle underpins its tolerance of standard antibiotics. Clinically, PYO accumulation correlates with lung function decline in chronic cystic fibrosis infections and is responsible for the blue-green coloration of pus in some acute infections. Despite their importance as virulence factors, therapeutic strategies have not exploited the dependence of P. aeruginosa on phenazines for survival. Here we report the discovery of a PYO oxidative demethylase (PodA), a hitherto hypothetical 17 kDa protein of unknown function. Through functional and X-ray crystallographic characterization, we show that PodA demethylates PYO, maintains robust activity over a range of conditions, and propose a mechanism whereby PodA catalyzes an intramolecular electron transfer from the methyl group to the central aromatic ring of PYO, generating reduced 1-hydroxyphenazine (1-OH-PHZ) and formaldehyde as products. An intramolecular redox rearrangement has not been previously observed for demethylases despite their extensive study. We additionally show that PodA disrupts the development of P. aeruginosa biofilms; this is due, in part, to the removal of a biofilm stimulatory factor, PYO, and generation of a metal chelator, 1-OH-PHZ, which is sufficient to inhibit biofilms when added exogenously. Phenazine modulation via PodA application thus may have therapeutic potential for selectively manipulating P. aeruginosa biofilm formation *in vivo*.

Construction of BAC Vector Library for Comparative Genomics, Elucidation of Gene Functions and Heterologous Expression of Targeted Genes

Crnovcic, Ivana* (icrnovci@scripps.edu)[1]; Cheng, Jan-Fang[2]; Rudolf, Jeffrey D.[1]; Hindra[1]; Yan, Xiaohui[1]; Chang, Chin-Yuan[1]; Yoshikuni, Yasuo[2]; Deutsch, Samuel[2]; Shen, Ben[1]

[1]Department of Chemistry, The Scripps Research Institute, Jupiter, FL 33458, United States. [2]US Department of Energy Joint Genome Institute, Walnut Creek, California., DNA Synthesis Science Program, Walnut Creek, CA 94598, United States.

Genome sequencing has revealed that microorganisms possess immense biosynthetic potential; however, only an estimated 1% of soil dwelling microorganisms have been cultivated in the lab and subsequently explored. This extraordinarily low and discovery-limiting number is attributed to media, incubation times and inoculum sizes traditionally used to isolate and expediently culture rare soil bacteria. Actinomycetales, including the genus Streptomyces, is one of the common classes of soil bacteria yet many of them are difficult to culture and are not amenable to genetic manipulation. Even if these strains can be cultivated, the desired pathways are often silent under laboratory conditions.

Among the known tools, BAC technology has been widely used in basic academic and industrial research. BAC vectors are capable of harboring large DNA inserts (up to 600 kb), are easy to handle, and the clones are stable when maintained in a single copy state. They therefore represent good tools for cataloguing novel genomes and can support the discovery of novel enzymes and natural product biosynthesis. Previously, we successfully developed a strategy for the construction of BAC clones capable of harboring large biosynthetic gene clusters (up to 200 kb) and showed its feasibility by engineered production of two different natural products in heterologous Streptomyces hosts. We isolated BAC clones that harbor a glutarimide containing polyketide (iso-migrostatin) (1) and a non-ribosomal peptide - polyketide (bleomycin) (2) gene cluster.

We are building a strain collection consisting of 10,000 Actinomycetales (currently ~5000 strains), isolated from unexplored and underexplored ecological niches. Our innovative method for strain prioritization (3) will enable us to identify the most promising strains, for natural product discovery, to target specific pathways for the determination of novel enzymes, or perform a phylogenetic diversity analysis. Construction of a BAC vector library will enable scientists to screen and functionally characterize individual genes and multi-gene pathways revealing novel chemistry and enzymology. Therefore our vectors can be applied in countless ways including DOE JGI submissions of bioenergy and microbe-plant interactions.

(1) Feng Z, Wang L, Rajski SR, Xu Z, Coeffet-LeGal MF, Shen B. Engineered production of iso-migrastatin in heterologous Streptomyces hosts. Bioorg Med Chem, 2009, 17, 2147-2153.
(2) Huang SX, Feng Z, Wang L, Galm U, Wendt-Pienkowski E, Yang D, Tao M, Coughlin JM, Duan Y, Shen B. A designer bleomycin with significantly improved DNA cleavage activity. J Am Chem Soc, 2012, 134, 13501-13509.
(3) Hindra, Huang T, Yang D, Rudolf JD, Xie P, Xie G, Teng Q, Lohman JR, Zhu X, Huang Y, Zhao LX, Jiang Y, Duan Y, Shen B. Strain prioritization for natural product discovery by a high-throughput real-time PCR method. J Nat Prod, 2014, 77, 2296-2303.

Comparative Genomics in the *Salicaceae*: the Case of Sex Determination

DiFazio, Stephen P.* (spdifazio@mail.wvu.edu)[1]; Zhou, Ran[1]; Carlson, Craig[2]; Gouker, Fred[2]; Tang, Haibao[3]; Krishnakumar, Vivek[3]; Town, Christopher[3]; Evans, Luke[1]; Rodgers-Melnick, Eli[2]; Schmutz, Jeremy[5, 6]; Jenkins, Jerry[5, 6]; Smart, Lawrence B.[2]; Tuskan, Gerald A.[4]

[1]West Virginia University. [2]Cornell University. [3]J. Craig Venter Institute. [4]Oak Ridge National Laboratory. [5]US Department of Energy Joint Genome Institute, Walnut Creek, California. [6]Hudson Alpha.

The Salicaceae family, including Populus and Salix, is a powerful model system for studying the influence of genome structure on the evolution of adaptive traits. A comparison of synonymous nucleotide substition rates between Populus and Salix demonstrates that the Salix lineage has accumulated more polymorphisms. This elevated evolutionary rate is recapitulated in higher rates of genome fractionation in Salix following the shared Salicoid genome duplication. The two genera have numerous contrasting phenotypic characterisitics, including growth form, pollination mode, and generation time. Furthermore, although both species are primarily dioecious, they appear to have different sex determination loci. Sex determination in Populus is controlled primarily by loci on chromosome 19 across all species that have been studied thus far. In contrast, it is now clear that sex determination occurs on Chromosome 15 in

Salix purpurea, based on mapping in an F2 intraspecific cross as well as in an association population of unrelated individuals,. This is quite surprising, especially given the high collinearity of the vast majority of the Populus and Salix genomes. Furthermore, examination of the genotypes of sex determination loci in male and female trees suggests that Populus trichocarpa has a predominantly XY sex determination system, while Salix purpurea has a ZW system. A comparative analysis of the sex determination regions of the two genera reveals many shared characteristics, including structural complexity, high repeat density, and suppressed recombination. Ongoing questions include whether there are shared mechanisms of sex determination in the two species, the extent of sex dimorphism, the role of pollinator attraction and defense in sex chromosome evolution, and the presence of sexually antagonistic genes in the sex determination regions.

Novel Anaerobic Fungi for Lignocellulose Degradation

Dollhofer, Veronika* (veronika.dollhofer@lfl.bayern.de)[1]; Callaghan, Tony, Martin[2]; Kinker, Isabella[1]; Bauer, Johann[3]; Griffith, Gareth, Wyn[2]; Sczyrba, Alexander[4]; Lebuhn, Michael[1]

[1]Bavarian State Research Center for Agriculture, Central Department for Quality Assurance and Analytics, Lange Point 6, 85354 Freising, Germany. [2]Institute of Biological Sciences, University of Wales, Aberystwyth SY23 3DA, United Kingdom. [3]Chair of Animal Hygiene, WZW, Technische Universität München, Weihenstephaner Berg 3, 85354 Freising, Germany. [4]Faculty of Technology and Center for Biotechnology, Bielefeld University, Bielefeld, 33615, Germany.

The use of lignocellulosic wastes for renewable energy generation is a favorable strategy to mitigate greenhouse gas emissions and global warming. However, for economical bioenergy production, efficient hydrolysis of lignocellulosic wastes is the issue to be solved before such substrates can be used in industrial pipelines.

Anaerobic fungi are the primary digesters of ingested forage in herbivorous animals and have developed an elaborate machinery to degrade lignocellulosic substrates. They disintegrate the plant material mechanical with their rhizoid growing into the fibers, bursting the lignin coat which protects cellulose and hemicelluloses from degradation. Concomitantly, they attack plant fibers with a multitude of cellulolytic and hemicellulolytic enzymes, catalyzing the conversion of the plant sugars to acids, ethanol, H_2 and CO_2. Some of the enzymes are believed to work synergistically combined in multi enzyme complexes, the cellulosomes. So far, cellulosomes were only known from effective prokaryotic degraders of the family Clostridiaceae. These features make anaerobic fungi favorites for microbial hydrolytic pretreatment in biotechnological processes.

The prerequisite to generate a functioning pretreatment with anaerobic fungi is the isolation of novel strains which are suitable for biotechnological purposes. In total, 21 samples from herbivorous animals (from zoos and from free-living animals) were used to generate an anaerobic fungal strain collection. Isolation of anaerobic fungal strains was conducted strictly anaerobically under a 100 % CO_2 atmosphere. Enrichment media containing wheat straw was inoculated with different amounts of diluted samples and incubated at 39 °C until growth of anaerobic fungi was observed. Pure cultures of anaerobic fungi were obtained through inoculation of agar-containing rolltubes inoculated with zoospore producing enrichment cultures. Rolltubing was repeated until pure anaerobic fungal cultures were obtained. This procedure led to 13 novel anaerobic fungal isolates tentatively identified by 28S rDNA analysis with the newly developed specific primer system AF LSU. The six most interesting ones were accepted for whole-

genome sequencing in the Community Science Programme 2016 of the Joint Genome Institute and will be presented with their phylogenetic placement in the following paragraph and at the conference. Two unique isolates represent bulbous species of which no genome has been sequenced so far.Isolate CaDo 13a was derived from rumen fluid of a free-living alpine goat and is closely related to the genus Caecomyces. The isolate KiDo 2m originates from feces of a domestic yak and is closely related to the genus Cyllamyces, for which only one species has been described until now. Further, we present three isolates KiDo 1h and KiDo 1m from Kiang feces and CaDo 1a from feces of an alpine goat. All of them are forming distinct lineages or clades in phylogenetic analysis and are very likely representing novel genera. In addition, we present the first isolate from an agricultural biogas plant CaDo16a, a novel monoflagellated Piromyces sp.. The phylogenetic placement of the mentioned isolates will be consolidated by analysis of their ITS 1 region as an established marker for fungal phylogeny. A comparison of the 28S rDNA and ITS 1 data will also be presented at the conference.

The Intermediate Wheatgrass Genome: a Resource for Understanding Mechanisms of Perenniality and Accelerating the Development of Perennial Crops

Dorn, Kevin M.* (dorn@ksu.edu)[1]; Kantarski, Traci[1]; Wang, Shuwen[2]; Zhang, Xiaofei[3]; Anderson, Jim[3]; Larson, Steve[4]; Plott, Christopher[5]; Jenkins, Jerry[5]; Grimwood, Jane[5]; Schmutz, Jeremy[5]; DeHaan, Lee[2]; Poland, Jesse[1]

[1]Kansas State University, Manhattan, Kansas USA. [2]The Land Institute, Salina Kansas USA. [3]University of Minnesota, St Paul Minnnesota USA. [4]USDA-ARS Forage and Range Research Lab, Logan, Utah USA. [5]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama USA.

Perennial crop species are central to the overall sustainable intensification of cropping systems for food, feed, fiber, and fuel. Dual-use crops that can serve as a biofuel feedstock in addition to a source of edible grain add additional economic return to the cropping systems that are needed to meet projected demand for both food and fuel. Thinopyrum intermedium (Intermediate wheatgrass - IWG) is a strong perennial species with prolific biomass production, relatively large seeds and high grain yield that has been targeted for domestication and improvement as a dual-use perennial grain and biomass crop. As a perennial crop, IWG can provide essential ecosystem services to current agricultural practices such as stabilizing soil with its deep (>3 meters) root structure, reducing nutrient runoff, and limiting weed growth. For biofuels, IWG produces biomass quantities similar or greater than switchgrass, with the added potential of >1,000 kg/ha of grain that is comparable to wheat in food products. Significant efforts to improve IWG through traditional breeding have focused on traits such as plant height, biomass production, seed size, free threshing seed, and early maturity. While some progress has been made, accelerating the improvement of IWG is likely possible with genomics-assisted breeding. Currently, there are limited molecular breeding and genomic tools for IWG due to its large allohexaploid genome (2n = 6x = 42) and 1C genome size of 12.6 Gb. To facilitate sequencing, we have a haploid line in hand derived from a twin seedling, confirmed by chromosome counting and flow cytometry. Whole genome sequencing of the haploid line has been initiated, starting with 30X coverage shotgun sequencing using 250 bp paired-end reads on the Illumina HiSeq 2500 platform that is being assembled using Meraculous. Additional coverage with both short insert and mate pair reads is currently being generated. A comprehensive reference transcriptome has also been generated from the haploid line that will be used for gene prediction and annotation of the draft genome. To anchor and order the assembly, several F1 populations are being sequenced (POPSEQ). These resources will provide vital information for ongoing

breeding and molecular studies to improve intermediate wheatgrass as a perennial grain species, along with providing a comparative platform for examining the genomic landscape of perenniality across the Pooidae. A series of transcriptome profiling experiments are being conducted between two annual (Brachypodium distachyon and Triticum aestivum) and two perennial species (IWG and B. sylvaticum) to understand the genetic networks underlying the perennial vs. annual growth habit. Collectively, this research will shed light on the complex genetic mechanisms underlying the perennial growth habit while directly accelerating the development of perennial biofuel and grain crops.

Nitrogen Fixation in *Populus*: Identification and Localization of the Key Diazotrophs in Planta

Doty, Sharon L.*(sldoty@uw.edu); Khan, Zareen; Kim, Soo-Hyung

University of Washington, Seattle, WA/USA.

Just as research on the human microbiome has demonstrated the profound importance of our microbiota on human health, plants are also strongly influenced by the ecosystem within them. To improve the environmental and economic sustainability of biomass production, it is essential that the biological interactions between plants and associated beneficial microbiota be more fully understood. Nitrogen is an essential macronutrient for plant growth. The use of N fertilizer for bioenergy plantations, however, is incompatible with the goals of climate change mitigation since its production requires fossil fuels, and soil bacteria convert the excess N fertilizer into nitrous oxide, a potent greenhouse gas. It has long been assumed that only legumes and actinorhizal plants can benefit from symbiotic N_2-fixation; however, it has become clear that N_2-fixation by associated and internal (endophytic) microorganisms can provide substantial N to the host plant without requiring root nodules.

Poplar (Populus) trees are an early successional pioneer plant species able to colonize nutrient-limited, cobble-dominated riparian zones. Native poplar plants have a diverse microbiota including strains that can fix dinitrogen gas, and promote plant growth and health under abiotic stresses including nutrient limitation and drought. Our lab demonstrated using the 15N2 incorporation assay that N_2 is fixed at high levels in wild poplar by endophytes, the microorganisms that live within plants. Although other research groups have reported on endophytes from poplar, no others have demonstrated endophytes with the high levels of N_2-fixation, broad host range, and dramatic growth enhancements under stress that we have seen with endophytes from wild poplar in challenging environments. Building upon the strong body of evidence that endophytes from wild poplar can improve the growth and stress tolerance of a broad range of plant species, this new project seeks to more fully understand the plant-microbe interactions involved in endophytic symbiosis, specifically in regards to N_2-fixation, with the ultimate goal of optimizing this technology towards increasing biomass yields on marginal lands with reduced nutrient and water inputs. Although we have cultured and characterized dozens of N_2-fixing (diazotrophic) endophytes from poplar and willow, it is unknown which strains are the most active in planta, where N-fixation occurs within the plant, and if a specific consortium of strains is required. In order to optimize an inoculum of endophyte strains for poplar plantations for bioenergy, this project will use the strengths of the EMSL and JGI to identify and characterize the key diazotrophic strains from wild poplar. We will research the exchange of N between endophytes and plants. We will identify the key microbial species needed for symbiotic N transfer in Populus and then ultimately test how this plant-microbe interaction impacts plant responses to drought and elevated CO_2 and temperature.

Coupled Metagenomic and Chemical Analyses of Degrading Fungal Necromass and Implications for Microbial Contributions to Stable Soil Organic Carbon

Egerton-Warburton, Louise* (warburton@chicagobotanic.org)[1, 2]; Schreiner, Kathryn M.[3]; Morgan, Benjamin S.T.[1, 2]; Schultz, Jeremy[4]; Blair, Neal E.[2]

[1]Chicago Botanic Garden, Glencoe, IL. [2]Northwestern University, Evanston, IL. [3]University of Minnesota Duluth, Duluth, MN. 4Southern Illinois University, Carbondale, IL.

Fungi comprise a significant portion of total soil biomass, the turnover of which must represent a dominant flux within the soil carbon cycle. Fungal organic carbon (OC) can turn over on time scales of days to months, but this process is poorly understood. Here, we examined temporal changes in the chemical and microbial community composition of fungal necromass during a 2-month decomposition experiment in which Fusarium avenaceum (a common saprophyte) was exposed to a natural soil microbial community. Over the course of the experiment, residual fungal necromass was harvested and analyzed using FTIR and thermochemolysis-GCMS to examine chemical changes in the tissue. In addition, genomic DNA was extracted from tissues, amplified with barcoded ITS primers, and sequenced using the high-throughput Illumina platform to examine changes in microbial community composition. Up to 80% of the fungal necromass turned over in the first week. This rapid degradation phase corresponded to colonization of the necromass by known chitinolytic soil fungi including Mortierella species. Zygomycetes and Ascomycetes were among the dominant fungal groups involved in degradation with very small contributions from Basidiomycetes. At the end of the 2-month degradation, only 15% of the original necromass remained. The residual material was rich in amide and C-O moieties which is consistent with previous work predicting that peptidoglycans are the main residual product from microbial tissue degradation. Straight-chain fatty acids exhibit varying degradation profiles, with some fatty acids (e.g. C16, C18:1) degrading more rapidly than bulk tissue, others maintaining steady concentrations relative to bulk OC (C18), and some increasing in concentration throughout the degradation (C24). These results indicate that the turnover of fungal necromass has the potential to significantly influence a variety of soil OC properties, including C/N ratios, lipid biomarker distributions, and OC turnover times.

Changes in Soil Microbial Diversity Following Amendment with Mesophilic and Thermophilic Digestates During Soil Solarization

Fernandez-Bayo, Jesus D.* (jdfbayo@ucdavis.edu)[1, 2]; Achmon, Ygal[1, 2]; Harrold, Duff R.[2]; McCurry, Dlinka G.[3]; Hernandez, Katie[3]; Dahlquist-Willard, Ruth M.[3]; Stapleton, James J.[4]; Yazdani, Ramin[5]; VanderGheynst, Jean S.[2]; Simmons, Christopher W.[1]

[1]Department of Food Science and Technology, University of California, One Shields Ave., Davis, CA 95616, United States. [2]Department of Biological and Agricultural Engineering, University of California, One Shields Ave., Davis, CA 95616, United States. [3]UC Cooperative Extension, Fresno and Tulare Counties, University of California, Kearney Agricultural Research and Extension Center, Parlier, CA 93648, United States. [4]Statewide Integrated Pest Management Program, University of California, Kearney Agricultural Research and Extension Center, Parlier, CA 93648, United States. [5]Yolo County Planning & Public Works Department, Division of Integrated Waste Management, Woodland, CA 95776, USA.

The use of soil biosolarization is a promising alternative to traditional soil fumigation techniques that use toxic compounds, such as methyl bromide, to control pests. Soil biosolarization involves

amending moist soil with organic matter and then covering with a transparent plastic tarp to induce passive solar heating and microbial production of inhibitory compounds to inactivate soil pests. Soil biosolarization is a relatively new technique and, therefore, its impact on soil microbial diversity is not well known. This disinfestation process likely has a considerable impact on soil microorganisms due to two factors: (i) there is a significant input of new microorganisms in the soil due to the application of organic amendment; (ii) the amended soil is incubated under extreme conditions that are different from the natural condition of the soil. Specifically, the soil reaches significantly higher temperatures, experiences anaerobic conditions due to the presence of the tarp, and may undergo acidification due to microbial activity and the elevated moisture content of the soil. These phenomena may also vary by soil depth. This study analyzed the microbial diversity following 8 days of biosolarization of soil mixed with organic waste from anaerobic digesters operating under thermophilic conditions (TSD) and mesophilic conditions (MSD). The analysis included sequencing of the 16S ribosomal RNA genes from soil microbial communities sampled from different depths (0-7.5, 7.5-15 and 15-22 cm) following biosolarization with digestate amendments. This diversity was compared to that of the original soil community. To delineate the effects of solar heating and soil amendment, the microbial diversity was also measured in amended samples incubated at room temperature. The sequencing data revealed that the taxonomic composition of the archaeal and bacteria communities changed depending on the type of digestate used and the soil depth. As expected, the digestates initially changed the soil microbial diversity because they harbored different anaerobic digestion microbial communities. Despite being initially dissimilar, the bacterial communities in soil amended with either digestate restructured to become more similar in the uppermost soil layer following biosolarization. This may be attributed to the strong temperature pressure near the soil surface, which drove selection of thermophilic bacteria common to both digestates. For the MSD-amended soil, the microbial communities at the lower layers and at room temperature were similar, whereas for the TSD-amended soil these bacterial communities differed in every sample. The differences in the archaeal communities were less significant. Further studies are needed to assess the functional roles of these communities with regards to soil biochemical cycles relevant to biosolarization and agriculture, including VFA production and organic matter turnover.

Assessing the Role of A-Biotic Stress in Determining the Community Composition of the Root Microbiome

Finkel, Omri M.* (ofinkel@live.unc.edu); Castrillo, Gabriel; González, Isai Salas; Dangl, Jeff

Department of Biology, University of North Carolina, Chapel Hill, NC, USA.

The structure and function of host associated microbial communities, microbiomes, is affected by the host phenotype, but it can also drive it. Microorganisms are known to aid plants through processes like nutrient acquisition or protection from pathogens, but the complexity of microbial communities found in association with plants has limited our ability to draw general rules guiding the assembly and maintenance of a functionally robust microbiome. The formulation and empirical testing of these assembly rules are crucial, not only for their intrinsic scientific value, but also for the effort to harness microbes and synthetic microbial consortia to predictably improve crop productivity.

As part of a large scale project to decipher the endophytic root microbiota, a library of ~600 strains was collected from the roots of the model plant Arabidopsis thaliana. A thorough genome sequencing endeavor is underway, with 200 isolate genomes sequenced and annotated so far. Re-colonization of gnotobiotic *A. thaliana* was demonstrated for many of these isolates, either in mono-association or as

members of consortia. High throughput systems are in place for measurement of plant biomass, host transcriptional profiles and nutrient levels as measures of plant productivity and health.
These resources provide an exceptional framework for the design of in planta synthetic community experiments. Synthetic community experiments are a useful experimental tool, as they allow control of stress levels and nutrient flux; as well as close monitoring of how bacterial consortia develop in the plant and their influence on plant biomass and productivity.

We present here a suite of in-silico, in-vitro and in-plantae approaches to designing and testing synthetic communities. We are interested in using a combination of these approaches in order to understand the relative contribution of a-biotic and biotic factors to determining the community composition of bacteria in and around plants. Specifically, we are asking (a) what is the relative contribution of salinity, pH and temperature to the assembly of bacterial communities in plants and (b) is there, along with the direct effect of the a-biotic conditions, an indirect effect on community composition, through modulation of plant exudates.

We aim to answer these questions using 200-member synthetic consortia of genome sequenced isolatesobtained from *A. thaliana* roots. Plants are inoculated with these consortia under varying salinity, pH and temperature conditions. In parallel, in-vitro growth of all members of the community is measured in the presence of plant exudates produced under the same set of a-biotic conditions. Combining plant colonization patterns with in-vitro growth data and genomic features of the community members will provide us with a powerful tool to bridge the gap between observed patterns and mechanism.

Combining Molecular, Genomic, and Isotopic Techniques to Examine the Diversity and Activity of Marine Thaumarchaeota in Monterey Bay and the California Current System

Francis, Christopher A.* (caf@stanford.edu)[1]; Tolar, Bradley B. (btolar1@stanford.edu)[1]; Smith, Jason M.[2]; Chavez, Francisco P.[2]

[1]Department of Earth System Science, Stanford University, Stanford, CA. [2]Biological Oceanography Group, Monterey Bay Aquarium Research Institute, Moss Landing, CA.

Thaumarchaeota are major players in the marine nitrogen cycle with a nearly ubiquitous distribution in the global ocean. They are thought to be the primary ammonia-oxidizing organisms in marine systems, with ammonia-oxidizing archaea (AOA) typically 10-1000 times more abundant than ammonia-oxidizing bacteria in the ocean; however, our understanding of how archaeal ammonia oxidation is influenced by key environmental factors is still quite limited. AOA are traditionally detected through molecular surveys of the 16S rRNA gene as well as the amoA gene encoding the α-subunit of ammonia monooxygenase – the enzyme catalyzing the oxidation of ammonia (the first step of nitrification). Nitrification plays a critical role in the marine nitrogen cycle, altering the distribution of inorganic nitrogen species available to phytoplankton and producing N_2O. Our ongoing CSP project seeks to determine whether variability in the abundance, genomic diversity, or cellular activity of Thaumarchaeota relates to observed changes in nitrification rates, or the prevailing 'metabolic state' of a given water mass. We collected DNA and RNA samples from Monterey Bay and the California Current System, with simultaneous physicochemical and nitrification rate measurements, for metagenomic and metatranscriptomic sequencing. AOA 16S

rRNA gene abundances correlated significantly with changes in nitrification rate with depth, while the relative abundance of genes and transcripts binned to a single AOA was not significantly correlated to nitrification rate. We found a correlation between the relative abundance of Marine Group I (MGI) Thaumarchaeota 16S rRNA reads (as % of total) and the absolute abundance of AOA amoA genes (determined via qPCR). All sequenced metagenomes and metatranscriptomes contain genes relevant to nitrification and other portions of the nitrogen cycle. Further analysis of the sequenced AOA metagenomes and metatranscriptomes will allow us to determine whether nitrification rates correspond to a particular group of AOA, or if other metabolic genes (such as those involved in carbon fixation pathways) are correlated to nitrification.

A Droplet Microfluidic Platform for Automating Genetic Engineering

Gach, Philip, C.* (pcgach@lbl.gov)[1, 2]; Shih, Steve C.C.[1, 2]; Sustarich, Jess[1, 2]; Keasling, Jay D.[3, 4, 5, 6]; Hillson, Nathan J.[1, 3, 4, 7]; Adams, Paul D.[1, 5, 8]; Singh, Anup K.[1, 2]

[1]Technology Division, Joint BioEnergy Institute (JBEI), Emeryville, California 94608, United States. [2]Applied Biosciences and Engineering, Sandia National Laboratories, Livermore, California 94550, United States. [3]Fuels Synthesis Division, Joint BioEnergy Institute (JBEI), Emeryville, California 94608, United States. [4]Biological Systems and Engineering Division, Lawrence Berkeley National Lab, Berkeley, California 94720, United States. [5]Department of Bioengineering, University of California, Berkeley, California 94720, United States. [6]Department of Chemical & Biomolecular Engineering, University of California, Berkeley, California 94720, United States. [7]US Department of Energy Joint Genome Institute, Walnut Creek, California. 94598, United States. [8]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Lab, Berkeley, California 94720, United States.

We present a water-in-oil droplet microfluidic platform for transformation, culture and expression of recombinant proteins in multiple host organisms including bacteria, yeast and fungi. The platform consists of a hybrid digital microfluidic/channel-based droplet chip with integrated temperature control to allow complete automation and integration of plasmid addition, heat-shock transformation, addition of selection medium, culture, and protein expression. The microfluidic format permitted significant reduction in consumption (100-fold) of expensive reagents such as DNA and enzymes compared to the benchtop method. The chip contains a channel to continuously replenish oil to the culture chamber to provide a fresh supply of oxygen to the cells for long-term (~5 days) cell culture. The flow channel also replenished oil lost to evaporation and increased the number of droplets that could be processed and cultured. The platform was validated by transforming several plasmids into Escherichia coli including plasmids containing genes for fluorescent proteins GFP, BFP and RFP; plasmids with selectable markers for ampicillin or kanamycin resistance; and a Golden Gate DNA assembly reaction. We also demonstrate the applicability of this platform for transformation in widely used eukaryotic organisms such as Saccharomyces cerevisiae and Aspergillus niger. Duration and temperatures of the microfluidic heat-shock procedures were optimized to yield transformation efficiencies comparable to those obtained by benchtop methods with a throughput up to 6 droplets/min. The proposed platform offers potential for automation of molecular biology experiments significantly reducing cost, time and variability while improving throughput.

http://pubs.acs.org/doi/abs/10.1021/acssynbio.6b00011

Expression of Lignin-Degrading Enzymes in Yeast *Yarrowia lipolytica*

Gao, Difeng, DG* (difengg@clemson.edu, difengg@gmail.com); Hussain, Murri; Fowler, Jeremy; Blenner, Mark

Clemson University.

Bioconversion of lignocellulose to renewable fuels and chemicals are mainly hindered by the recalcitrant lignin. Degradation and modification of lignin is a key process to the development of biomass-based refinery. Additionally, lignin has potentials to be an abundant and sustainable source for the production of aromatic chemicals. Microbial degradation of lignin usually involves a synergetic set of ligninolytic enzymes, providing an efficient and green route alternative to chemical degradation. In this study, we explored the feasibility of expression of several lignin-degrading enzymes in non-conventional yeast Yarrowia lipolytica. These enzymes including lignin peroxidase, manganese peroxidase, laccase, versatile peroxidase, aryl-alcohol oxidase and glyoxal oxidase are from several fungi. The genes coding for these lignin-degrading enzymes are cloned into Y. lipolytica using constitutive TEF promoter. Both of the native enzyme secretion signal peptides and Y. lipolytica lipase signal peptide are used in directing enzyme secretion.

Identification of Two Genes Required for Long-Chain Alkane Production in Cyanobacteria

Gibbons, Jaimie* (jaimie.gibbons@jacks.sdstate.edu); Gu, Liping; Zhou, Ruanbao

South Dakota State University, Brookings, SD.

Oil reserves worldwide are limited, and as prices have risen, renewable fuels have become increasingly important. Is there a biofactory that can convert carbon dioxide, water, and sunlight into fuels? Yes, several species of cyanobacteria are known to produce alkanes and alkenes from carbon dioxide, water and sunlight. These liquid hydrocarbons are the major component of oil. This provides a more sustainable approach to producing biofuels than using grains or biomass that could compete for food. However, native cyanobacteria do not produce these hydrocarbons in sufficient yield for commercial deployment. This research is to understand how cyanobacteria photosynthetically produce fuels using only atmospheric gases and water. Two putative alkane genes in Anabaena 7120 were knocked out by inserting a gfp-spec cassette within the genes. The engineered genetic sequence in an integration plasmid was homologously integrated into the chromosome of Anabaena 7120. Through a double crossover, the functional alkane genes were replaced by the inactivated gene segment. The GC/MS data revealed that heptadecane (C17H36) produced by the wildtype was not found in the two knock-out mutants, suggesting that one or both of the alkane genes that were inactivated are part of a biosynthetic pathway for the production of heptadecane. The complementation experiment with reintroduction of the functional genes into the two mutants confirmed the role of these two genes. However, recovery of heptadecane production in the complement strain did not reach the same levels as production in the wildtype. It is possible that chromosomal gene expression of the alkane genes in Anabaena 7120 is more stable than plasmid-based gene expression, thus explaining the lower levels of heptadecane obtained in the complement strain.

Engineering Synthetic Systems Inspired by Anaerobic Gut Fungi

Gilmore, Sean P.* (SeanGilmore@umail.ucsb.edu)[1]; O'Malley, Michelle A. (momalley@engineering.ucsb.edu)[1]; Haitjema, Charles H.[1]; Henske, John K.[1]; Sexton, Jessica A.[1]; Solomon, Kevin V.[1]; Brewer, Heather M.[2]; Purvine, Samuel O.[2]; Wright, Aaron T.[2]; Kuo, Alan[3]; Mondo, Stephen[3]; Salamov, Asaf[3]; LaButti, Kurt[3]; Theodorou, Michael K.[4]; Grigoriev, Igor V.[3]

[1]Department of Chemical Engineering, University of California, Santa Barbara. [2]Enivronmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352. [3]US Department of Energy Joint Genome Institute, Walnut Creek, California., 2800 Mitchell Dr., Walnut Creek, CA 94598. [4]Animal Production, Welfare and Veterinary Sciences, Harper Adams University, Newport, Shropshire, TF10 8NB, United Kingdom.

Anaerobic fungi in the hindgut of large herbivores are among the most robust organisms at degrading crude lignocellulose. Their remarkable cellulolytic capabilities have great potential for use in biomass breakdown and biofuel processing. Anaerobic fungi achieve cellulolytic efficiency through the production of large, multi-enzyme complexes called fungal cellulosomes. In isolation, anaerobic fungi metabolize some of the released sugars and convert them into fermentation products. In nature, however, they exist in a community with archaea, bacteria, and protozoa, which drastically alters the behavior of the fungi. By elucidating the parts responsible for efficient biomass degradation at both the protein and cellular level, we seek to replicate this efficiency in synthetic systems.

Fungal cellulosomes are similar to bacterial cellulosomes in that the protein-protein interactions are mediated through parts termed the dockerin and cohesin. However, many differences exist. The dockerin domains exist in tandem repeats and bear no species specificity like those in the bacterial systems. Furthermore, the exact sequence for the cohesin module has yet to be established. Through a combination of –OMICs approaches and traditional biochemical assays, a large putative scaffoldin molecule was identified. The scaffoldin was heterologously expressed and screened for interaction with recombinant dockerin through an ELISA. The KDapp was determined using Equilibrium Surface Plasmon Resonance. A transcriptomic survey of dockerin domain-containing proteins revealed some degree of conservation in dockerin location on classes of CAZymes. Using this observation, the dockerin domains were adapted to thermostable cellulases, demonstrating its applicability as a novel protein scaffolding systems and suggesting the possibility of synthetic cellulosomes for biomass degradation.
Anaerobic fungi have been shown to interact closely with methane producing archaea (methanogens). The methanogens siphon hydrogen and other metabolites from the fungi, allowing the fungi to more efficiently produce energy by increasing the flux through their hydrogenosomes. To further investigate this mechanism, native fungal/methanogen consortia were isolated from herbivore fecal materials. Consortia were maintained together and also separated into monocultures for comparison. Genomic sequencing revealed the presence of one fungus, two methanogens, and one bacterium in one consortium, which was stable under continuous passage for over 20 months. The consortium demonstrated faster and more complete degradation of cellulosic substrates, as well as a wider range of utilized substrates compared to the monocultured fungus alone. By introducing the methanogens into cultures of other well-characterized anaerobic fungi, stable synthetic co-cultures were established. These stable synthetic consortia demonstrated similar efficiency, and suggest a promising option for conversion of crude biomass into sustainable chemicals.

The Pan-Genome of *Brachypodium distachyon*, Capturing the Full Genetic Complement of a Plant Species

Gordon, Sean* (sgordon@lbl.gov)[1]; Contreras-Moreira, Bruno[2, 3]; Des Marais, David L.[4]; Burgess, Diane[5]; Schackwitz, Wendy[1]; Tyler, Ludmila[6]; Martin, Joel[1]; Woods, Daniel[7]; Lipzen, Anna[1]; Shu, Shengqiang[1]; Phillips, Jeremy[1]; Barry, Kerrie[1]; Amasino, Richard[7]; Caicedo, Ana[6]; Mur, Luis[8]; Freeling, Michael[5]; Catalan, Pilar[9]; Vogel, John[1]

[1]DOE Joint Genome Insitute, Walnut Creek, CA. [2]Estación Experimental de Aula Dei (EEAD-CSIC), Zaragoza, Spain. [3]Fundación ARAID, Zaragoza, Spain. [4]Harvard University, Cambridge, MA. [5]University California, Berkeley, Berkeley, CA. [6]University of Massachusetts, Amherst, MA. [7]University of Wisconsin, Madison, WI. [8]Aberystwyth University, UK. [9]University of Zaragoza, Spain.

The genetic diversity of a species is the sum of the diversity found in all individuals of that species. Many studies have attempted to estimate the diversity of plant species by resequencing diverse accessions and aligning the reads to a reference genome. While this approach readily identifies SNPs and small indels with respect to the reference genome, it underestimates total genomic diversity because highly divergent regions align poorly to the reference and, of course, any sequence not found in the reference will be missed entirely. Thus, the true extent of diversity within plant species is largely unknown. De-novo genome assemblies and annotation can be used to more accurately estimate the true genomic diversity within a species. We applied this approach using 54 Brachypodium distachyon accessions to create a pan-genome that contains all the diversity found in the accessions sequenced. Analysis of this data yields a high-confidence B. distachyon pan-genome that includes 13,408 core gene clusters found in all lines, 7,283 soft-core genes cluster absent from a few lines, and 17,195 shell gene clusters found in 3 to 52 lines. We find 30,691 gene clusters represented by the reference genome/reference control and a varying number of other genomes. In addition, we identify 7,135 gene clusters not represented in the reference line or controls but present in multiple divergent lines. We show that non-core genes are expressed at lower levels, have narrower and more variable expression across accessions, are evolving faster, have reduced orthology to related grasses and are less likely to have a homeologous gene retained from the ancient genome duplication in the grass lineage. We evaluate the relationship between the number of sequenced lines and their phylogenetic position in relation to the addition of both genic and non-coding sequence. We describe the physical chromosome position of non-core and non-reference genes and its relation to transposable elements. This analysis suggests possible mechanisms by which dispensable genes are eliminated and also barriers to their removal.

Tropical Microbial Soil Diversity and Carbon Turnover Across a Depth Profile

Green, Elizabeth* (green74@llnl.gov)[1]; Nuccio, Erin[1]; Wollard, Jessica[1]; Nilson, Daniel J.[1]; Heckman, Kate[1]; Campbell, Ashley[1]; Wood, Tana[2]; Pett-Ridge, Jennifer[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA. [2]U.S. Forest Service, International Institute of Tropical Forestry, Rio Piedras Puerto Rico.

Despite tropical systems storing vast amounts of C, relatively little is known about the microbial ecology and carbon stocks of sub-surface tropical soils. While a handful of studies have quantified microbial biomass and taxonomic diversity in tropical surface soils, estimates of the microbial communities in

corresponding sub-surface soils are almost nonexistent. As tropical temperature and rainfall patterns are expected to be altered by climate change, understanding the microbial communities' role in tropical forest soils has never been more pressing. The goal of this study was to measure sub-surface tropical soil C stocks, C turnover time and soil microbial communities. We collected soil from the Luquillo Experimental Forest in Puerto Rico from replicate pits to a depth of ~120 cm. We identified microbial community diversity, quantified C pools and assessed carbon turnover rates with 14C analysis of total C pools, and density fractionated pools.

Bacterial alpha diversity and biomass decreased sharply with depth below 20 cm. Ordination analysis showed a clear separation between shallow (0-20 cm) and deep (>20 cm) microbial community structures. Actinobacteria and Proteobacteria were most abundant in shallow soils (frequently at > 10% abundance) and decreased with depth across all locations. Verrucomicrobia had significant abundance in surface samples, but not below 20 – 30 cm. In contrast, Acidobacteria (class SO35) were particularly abundant at depth, as were Crenarchaeota (mostly Thaumarchaea).

Soil C, N, and root biomass all declined exponentially with depth. In surface soils, total C stocks were 5.5% and decreased to less than 0.5% at 120 cm. Soil organic matter Δ14C was drastically younger (3-1,500 years in 0-20 cm) than deep soil (5,000-40,000 years at 140 cm). In comparison to temperate deciduous forests, these 14C values reflect far older soil C, and organic matter decomposition that highly favors free light C pools, even at depth. While previous work suggests these low C tropical subsoils contain small but metabolically active microbial communities at depths of ~100 cm, these organisms may be extremely organic matter (and O_2) limited, and may preferentially degrade recent inputs.

A Fast, Efficient Approach for Genome Assembly from Metagenomic Data

Green, Richard, E.* (ed@soe.ucsc.edu)[1, 2]; Putnam, N.[1]; Havlak, P.[1]; O'Connell, B.[1, 2]; Blanchette, M.[1]; Stites, J.[1]; Calef, R.[1]; Min, E.[1]

[1]Dovetail Genomics, Santa Cruz, CA 95060. [2]Dept. of Biomolecular Engineering, Univ. of California, Santa Cruz, CA 95064.

High-throughput sequencing allows genetic analysis of the organisms that inhabit a wide variety of environments of biomedical, ecological, or biochemical interest. Shotgun sequencing of environmental samples, which often contain microbes that are refractory to culture, can reveal the genes and biochemical pathways present within the organisms in a given environment. Careful filtering and analysis of these data can also reveal signals of phylogenetic relatedness between reads in the data. However, high-quality *de novo* assembly of these highly complex datasets is generally considered to be intractable.

Here we present a fast and efficient method for *de novo* genome assembly of read data from complex metagenomics datasets. The key aspect of our approach is efficient generation of connectivity data from these complex DNA samples. Adopting an approach that has worked well for *de novo* genome assembly of single organisms, we perform proximity ligation and sequencing of in vitro assembled chromatin from metagenomics DNA samples. We show that this approach generates scaffolding data that uniformly and completely represent the composite species in a metagenomics sample. Combined with shotgun

sequence data, we are able to generate multi-megabase scaffold assemblies of the most abundant organisms in a wide range of metagenomics communities including the human oral and gut microbiome including organisms that are not amenable to growth in culture. This fast and simple approach opens the possibility of high-throughput, culture-free assembly of genomes using only widely available high-throughput sequencing technology.

Improving the Completeness of Metabolic Networks Used for Automatic Generation of Genomic-Scale Constraint-Based Models

Griesemer, Marc* (griesemer1@llnl.gov)[1]; Kimbrel, Jeffrey[1]; D'haeseleer, Patrik[1]; Zhou, Carol[2]; Navid, Ali[1]

[1]Biosciences and Biotechnology Division, Lawrence Livermore National Laboratory, Livermore, CA 94551. [2]Global Security Computing Applications Division, Lawrence Livermore National Laboratory, Livermore, CA 94551.

Reconstruction of genome-scale metabolic models permits analysis using constraint-based methods (like Flux Balance Analysis, FBA) of environmental and genomic influences on organisms, as well as metabolic interactions between different species. Many researchers use automatically generated draft FBA models as a starting template for developing curated, publication quality models. However, typically one-third of genes in a newly sequenced genome lack a proper functional annotation. Thus, improving the completeness of metabolic network bases of the draft models would be very helpful in accelerating the process of developing dependable models. We seek to help this endeavor through development of software tools that can use output of improved genome annotation tools to eliminate gaps in metabolic networks that are used to automatically generate genome-scale models.

Our software pipeline uses the sequence annotation capabilities of the Protein Sequence Annotation Tool (PSAT)1 and parsing by Pathway Tools (PT) as preliminary curation to enhance the model coming out of the U.S. Department of Energy Systems Biology Knowledgebase (KBase)2 model-builder. This also allows us to compare with the Rapid Annotation Subsystem Technology (RAST) annotation and synthesize a list of the common reactions and metabolites using the KEGG, MetaCyc, and SEED databases. Our test case was Phaeobacter gallaeciensis DSM 26640, a bacteria known to have complex symbiotic / opportunistic interactions with the green algae E. huxleyi, and whose genus is highly abundant in our 16S data from N. salina and P. tricornutum algal ponds.

First, we annotated P. gallaeciensis with the RAST tool in KBase, our in-house PSAT functional annotation tools, and Transporter Automatic Annotation Pipeline (TransAAP). Multiple GenBank files were generated to test different approaches to integrate annotations, and then passed to the model-building application in KBase and Pathway Tools (PT) for comparison of methods. Further gapfilling of the model uncovered reactions essential for growth through in silico reaction knockouts. We then compared it to the PSAT annotation tool's metabolic reconstruction and found additional reactions and compounds not detected by RAST, including several marked as essential by gap filling. TransAAP provided additional transporter annotations, and substrate predictions. The combined use of these tools gave us valuable information for reconstructing a draft model. Our project seeks to use these models to understand to study algal-bacterial interactions for enhancement of biofuel production.

While human curation of draft modes will be needed to achieve a complete, quality model suitable for dependable Flux Balance Analysis; through our effort we aim to significantly shorten this process by incorporating additional annotation sources in an automated fashion by these methods.

1. Leung E, Huang A, Cadag E, Montana A, Soliman JL, Zhou CLE (2016) Protein Sequence Annotation Tool (PSAT): a centralized web-based meta-server for high-throughput sequence annotations. BMC Bioinformatics. DOI: 10.1186/s12859-016- 0887-y.

2. Cottingham, Robert, W. 2015. The DOE systems biology knowledgebase (KBase): progress towards a system for collaborative and reproducible inference and modeling of biological function. In Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '15). ACM, New York, NY, USA, 510.

Targeted Sequencing of Genes from Soybean Using NimbleGen SeqCap EZ and PacBio SMRT Sequencing

Gu, Jenny* (jgu@pacb.com)[1]; Eng, Kevin[1]; Sethuraman, Anand[1]; Raterman, Denise[2]; Richmond, Todd[2]; Kujawa, Steve[1]

[1]Pacific Biosciences, Menlo Park, CA USA. [2]Roche NimbleGen, Madison, WI USA.

Full-length gene capture solutions offer opportunities to screen and characterize structural variations and genetic diversity to understand key traits in plants and animals. Through a combined Roche NimbleGen probe capture and SMRT Sequencing strategy, we demonstrate the capability to resolve complex gene structures often observed in plant defense and developmental genes spanning multiple kilobases. The custom panel includes members of the WRKY plant-defense-signaling family, members of the NB-LRR disease-resistance family, and developmental genes important for flowering. The presence of repetitive structures and low-complexity regions makes short-read sequencing of these genes difficult, yet this approach allows researchers to obtain complete sequences for unambiguous resolution of gene models. This strategy has been applied to genomic DNA samples from soybean coupled with barcoding for multiplexing.

Biosolar Synthesis of Myrcene Using CO_2 and H_2O via Engineered N_2-fixing Cyanobacteria

Gu, Liping* (ruanbao.zhou@sdstate.edu); Myrlie, Aldon; Halfmann, Charles; Gibbons, William; Zhou, Ruanbao

Department of Biology and Microbiology, South Dakota State University, Brookings, SD57007, USA.

Biorefineries typically release about one third of the carbohydrate carbon as CO_2 during fermentation. For example, corn ethanol plants in just South Dakota release over 30,000 tons of CO_2 per hour. Ideally, a photosynthetic microorganism, such as N_2-fixing cyanobacteria, could be engineered to directly convert the waste CO_2 back into drop-in, hydrocarbon fuels. Thus, one step of "CO_2 to fuel" model could

eliminate the steps of biomass production, harvest, logistics, and conversion required for conventional biomass-based biofuel production.

Myrcene (C10H16) is a suitable “drop-in” replacement for petrofuels as well as a renewable base chemical for manufacturing fine chemicals. Myrcene is a naturally-occurring terpene alkene, emitted as a volatile compound from many aromatic plants. Myrcene is produced directly from the universal isoprenoid intermediate geranyl diphosphate (GPP) by myrcene synthase (MyrS). Although myrcene is naturally produced by the MEP pathway in some plants, its extraction is uneconomical due to low productivity. Cyanobacteria, through billions of years of evolution, have become finely-tuned biological devices that can efficiently harvest and convert solar energy into chemical energy stored in a variety of reduced carbon compounds. The photosynthetic efficiency of cyanobacteria is approximately 10-100 times higher than most plants. Moever, N_2-fixing cyanobacteria can simultaneously carry out both photosynthetic CO_2-fixation and photosynthetic N_2-fixation. Because of their simplest input requirements for rapid growth and ease of genetic manipulation, as well as suitability for industrialized mass production, N_2-fixing cyanobacteria are particularly attractive photosynthetic organisms for engineering production of excreted long-chain hydrocarbons such as myrcene.

N_2-fixing cyanobacteria, like plants, have native metabolic pathways (Calvin cycle and MEP pathway) to photosynthetically convert CO_2 and water into a variety of reduced carbon compounds, including GPP, the immediate precursor for production of myrcene. The abundant availability of GPP in cyanobacterial cells was confirmed by our limonene work [Halfmann et al., Green Chem. 16 (6), 3175 – 3185]. However, cyanobacteria lack the myrcne synthase that plants use to convert GPP into myrcene. In this research, the myrcene synthase gene from Norway Spruce was fused to a synthetic, dual cyanobacterial Pnir–PpsbA1 promoter and subcloned into a shuttle vector for transformation of an N_2-fixing cyanobacterium Anabaena sp. PCC7120. The first generation of transgenic Anabaena (with a single plant myrcene synthase gene) was capable of photosynthetically, synthesizing and secreting myrcene, as confirmed by GC-MS. Interestingly, the transgenic Anabaena grown with atmospheric N_2 gas as sole nitrogen source (N_2-fixing condition) produced more myrcene than using combined nitrogen (nitrate) as nitrogen source. To further increase myrcene production, a synthetic operon coding for three key enzymes (DXS-IDI-GPPS) from the MEP pathway were over-expressed in Anabaena. This second generation strain produced at least 2-fold more myrcene than the first generation strain using atmospheric N_2 gas as sole nitrogen source.

Populus Deltoides De Novo Assembly

Gunter, Lee E.* (Gunterle@Ornl.gov)[1]; Barry, Kerrie[2]; Schmutz, Jeremy[2, 3]; Plott, Christopher[3]; Jenkins, Jerry[3]; Tuskan, Gerald A.[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830. [2]US Department of Energy Joint Genome Institute, Walnut Creek, California. 94598. [3]HudsonAlpha Institute for Biotechnology, Huntsville AL 35806.

JGI performs sequencing on behalf of the U.S. Department of Energy (DOE) Bioenergy Research Centers (BRC), which have been funded to accelerate basic research in the development of cellulosic ethanol and other biofuels through increased efficiency and diversification of clean energy sources. Populus is one of JGI’s flagship species, and is of special interest as a lignocellulosic feedstock. Populus deltoides provides the genetic background for the poplar transformation pipeline work within the Bioenergy Science Center

(BESC, one of the three BRCs) led by DOE's Oak Ridge National Laboratory. The goal of this project (BRC proposal 1735) is to create a *de novo* assembly of P. deltoides to serve as a reference to the deltoides-specific research within BESC related to genetic modification and gene identification. This resource will be useful to the BRCs in general as well as to the poplar community at large.

The current poplar reference genome is based on P. trichocarpa clone 383-2499 ('Nisqually-1'). Because of the genetic divergence of the two species, the reference genome allows us to identify only ~50% of the whole transcriptome shotgun sequencing (RNA-seq) reads generated from P. deltoides as part of our characterization of poplar transgenics for the BESC Transformation Pipeline and TOP 20 lines. A P. deltoides reference will facilitate our analysis of genetic modifications, RNA-seq, eQTL, proteomic and other omics data.

To date, three Meraculous assemblies have been completed on sequence data derived from 66x depth (coverage) of 400 bp and 49x depth of 800 bp Illumina Hi-seq 2x250 fragment libraries; and from 39x depth of 4 kb and 37x depth of 8 kb Illumina 2x150 mate pair libraries. The best assembly (K=91, BP) produced a contig N50 of 45.3 kb, scaffold N50 of 672.1 kb, and an assembled sequence of 327.1 Mb with 64,820 gene models. Gap patching is underway with PACBIO C6/P4 sequencing. Assembly statistics and plans for comparative analysis with P. trichocarpa will be presented.

Probing the Depths of Microbial Community Structure in Soil

Hahn, Aria, A.* (hahnaria@gmail.com)[1]; Perez, Sarah E.I.[2]; Hanson, Niels W.[2]; Lee, Sangwon[3]; Scofield, Melanie[1]; Durno, Evan W.[2]; Mohn, William W.[1]; Hallam, Steven J.[1, 2]

[1]Department of Microbiology and Immunology, University of British Columbia, Vancouver, Canada; [2]Graduate Program in Bioinformatics, University of British Columbia, Vancouver, Canada; [3]Joint Bioenergy Institute, California, USA.

Accurate description of microbial interactions driving nutrient and energy cycles in terrestrial ecosystems remains challenging given the complexity of soil microbial communities. Co-occurrence analysis provides a statistical framework to chart this complexity. Here we use clustered small subunit ribosomal RNA gene sequences and shotgun metagenomes in unmanaged and harvested boreal forest soils to construct microbial co-occurrence networks and relate these networks in a pathway-centric manner to microbial community metabolism. Network topology and community metabolism were influenced by soil depth and perturbation state. Indeed, metagenomic analysis indicated that harvested surface soil horizons exhibited reduced potential for carbon and nitrogen cycling processes associated with heterotrophic carbon fixation and plant biomass conversion pathways 13 years post harvesting. Furthermore, analysis of shotgun ribosomal gene fragments confirmed the presence of network taxa within the metagenomes. These observations indicate forest management practices can impact metabolic interactions with long-term feedback on ecosystem services including carbon storage potential.

Cellulosomes from Anaerobic Fungi Assemble on Large Non-Catalytic Scaffoldins

Haitjema, Charles H.* (haitjema@engineering.ucsb.edu)[1]; Gilmore, Sean P.[1]; Henske, John K.[1]; de Groot, Randall[1]; Kuo, Alan[2]; Mondo, Stephen[2]; Salamov, Asaf[2]; Grigoriev, Igor V.[2]; Brewer, Heather M.[3]; Purvine, Samuel O.[3]; Wright, Aaron T.[3]; O'Malley, Michelle A.[1]

[1]Department of Chemical Engineering, University of California, Santa Barbara, Santa Barbara, CA, 93106, United States. [2]US Department of Energy Joint Genome Institute, Walnut Creek, CA. 94598, , United States. [3]Environmental Molecular Sciences Laboratory, Fundamental and Computational Sciences, Pacific Northwest National Laboratory, Richland, WA, 99354, United Sates.

Cellulosomes are large multienzyme complexes synthesized by anaerobic microorganisms that deconstruct plant biomass to fermentable sugars. In anaerobic bacteria, cellulolytic enzymes assemble on large non-catalytic scaffoldin molecules that are anchored to the outer membrane. These scaffoldins are decorated with cohesin domains that bind to dockerin domains encoded at the terminal ends of glycoside hydrolases. In the anaerobic gut fungi (phylum Neocallimastigomycota), plant cell wall degrading enzymes also form large macromolecular complexes via modular dockerin domains, but the identification of central dockerin-binding scaffoldin molecules has remained elusive, and this has been a major lack in the understanding of fungal cellulosomes since their discovery more than twenty years ago. Here, we identify large (150 – 700 kDa) proteins that function as scaffoldins in the cellulosomes of three gut fungal genera: Anaeromyces, Neocallimastix and Piromyces. Our results demonstrate that these scaffoldins encode a conserved amino acid sequence repeat that functions in binding to fungal dockerins, and the conservation of this motif allows for interspecies assembly of cellulosome components. Furthermore, the fungal scaffoldin and dockerin sequences contains no homology to their bacterial counterparts, suggesting that the cellulosome-based strategy for degrading plant biomass may have evolved twice among cellulolytic microbes. We anticipate that these fungal scaffoldins and their associated dockerin-encoding proteins will provide new modular components for the general synthesis of multiprotein complexes, including designer cellulosomes for industrial biomass breakdown.

Profiling the Microbiome in Fecal Microbiota Transplantation Using Circular Consensus and Single Molecule, Real-Time Sequencing

Hall, Richard J.* (rhall@pacificbiosciences.com)[1]; Heiner, Cheryl[1]; Stanley, Chris[2]; Hamilton, Matt[2]; Khoruts, Alexander[3]; Sadowsky, Michael[2]

[1]PacBio, Menlo Park, CA. [2]BioTechnology Institute, University of Minnesota, St. Paul, MN. [3]Department of Medicine, University of Minnesota, Minneapolis, MN.

There are many sequencing-based approaches to understanding complex metagenomic communities spanning targeted amplification to whole-sample shotgun sequencing. While targeted approaches provide valuable data at low sequencing depth, they are limited by primer design and PCR. Whole-sample shotgun experiments generally use short-readsequencing, which results in data processing difficulties. For example, reads less than 500bp in length will rarely cover a complete gene or region of interest, and will require assembly. This not only introduces the possibility of incorrectly combining sequence from different community members, it requires a high depth of coverage. As such, rare community members may not be represented in the resulting assembly.

Circular-consensus, single molecule, real-time (SMRT®) Sequencing reads in the 1-3kb range, with >99% accuracy can be efficiently generated for low amounts of input DNA. 10 ng of input DNA sequenced in 4 SMRT Cells on the PacBio RS II would generate >100,000 such reads. While throughput is lower compared to short-read sequencing methods, the reads are a true random sampling of the underlying community since SMRT Sequencing has been shown to have very low sequence-context bias. With reads >1 kb at >99% accuracy it is reasonable to expect a high percentage of reads include gene fragments useful for analysis without the need for *de novo* assembly.

Here we present the results of circular consensus sequencing for an individual's microbiome, before and after undergoing fecal microbiota transplantation (FMT) in order to treat a chronic Clostridium difficile infection. We show that even with relatively low sequencing depth, the long-read, assembly-free, random sampling allows us to profile low abundance community members at the species level. We also show that using shotgun sampling with long reads allows a level of functional insight not possible with classic targeted 16S, or short read sequencing, due to entire genes being covered in single reads.

Carbon and Nitrogen Cycling in Glaciers of the Pacific Northwest: Microbial Biogeochemical Cycling in Face of Rapid Glacial Retreat

Hamilton, Trinity L.* (trinity.hamilton@uc.edu)[1], Havig, Jeff R.[2]

[1]Department of Biological Sciences, University of Cincinnati, Cincinnati, Ohio, 45221. [2]Department of Geology, University of Cincinnati, Cincinnati, Ohio, 45221.

The majority of geomicrobiology research conducted on glacial systems to date have focused on glaciers that override primarily carbonate or granitic bedrock types with little known of the processes that support microbial life in glacial systems overriding volcanic terrains (e.g., basalts, andesites). In addition to being understudied, mid-latitude glaciers such as those found on the flanks of stratovolcanoes in the Cascade Range of Oregon and Washington, are particularly susceptible to rising temperatures due to their size and location, and thus are experiencing rapid retreat. Based on glacier loss rates of ~0.29 km2/yr for Mt. Rainier, the Sisters in Oregon may be glacier free in ~20 years, and Mt. Hood in OR and Mt. Adams in WA in ~80 years. Regardless, primary productivity in supraglacial ecosystems (on the ice and snow surface) appears to be nutrient-limited (by both phosphate and fixed nitrogen) and temperature-limited. Despite this observation, there is mounting evidence that active nitrogen cycling—including nitrogen fixation and nitrification—occurs in supraglacial ecosystems. To better constrain the role of the supraglacial ecosystems in the global carbon and nitrogen cycles, and to gain insight into microbiome composition and function in glacial systems overriding volcanic terrains, we examined the structure of supraglacial microbial communities and primary productivity in alpine glaciers on stratovolcanoes in the Cascade Range of the Pacific Northwest and. Our data indicate that supraglacial ice and snow on these stratovolcanoes host endemic microbial communities and that these assemblages have differing levels of carbon fixation. in situ carbon uptake incubation experiments showed a wide range of carbon assimilation values, from a low of 6 µg C/g biomass/hr at Collier Glacier (N. Sister) to a high of 51 µg C/g biomass/hr at Eliot Glacier (Mt. Hood). Along with total biomass and isotopic information, will be incorporate into spatial models to improve estimates of the contribution of these communities to local and global carbon and nitrogen cycling and are critical to determining ecosystem-wide and global effects of glacial retreat.

Coupling of Belowground Carbon Cycling and Stoichiometry from Organisms to Ecosystems along a Soil C Gradient Under Rice Cultivation

Hartman, Wyatt, H.* (whhartman@lbl.gov)[1]; Ye, Rongzhong[2]; Horwath, William R.[2]; Tringe, Susannah G.[1]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, California. [2]Department of Land, Air and Water Resources, UC Davis, Davis CA.

Ecological stoichiometry is a framework linking biogeochemical cycles to organism functional traits that has been widely applied in aquatic ecosystems, animals and plants, but is poorly explored in soil microbes. We evaluated relationships among soil stoichiometry, carbon (C) cycling, and microbial community structure and function along a soil gradient spanning ~5-25% C in cultivated rice fields with experimental nitrogen (N) amendments. We found rates of soil C turnover were associated with nutrient stoichiometry and phosphorus (P) availability at ecosystem, community, and organism scales. At the ecosystem scale, soil C turnover was highest in mineral soils with lower C content and N:P ratios, and was positively correlated with soil inorganic P. Effects of N fertilization on soil C cycling also appeared to be mediated by soil P availability, while microbial community composition (by 16S rRNA sequencing) was not altered by N addition. Microbial communities varied along the soil C gradient, corresponding with highly covariant soil %C, N:P ratios, C quality, and carbon turnover. In contrast, we observed unambiguous shifts in microbial community function, imputed from taxonomy and directly assessed by shotgun sequenced metagenomes. The abundance of genes for carbohydrate utilization decreased with increasing soil C (and declining C turnover), while genes for aromatic C uptake, N fixation and P scavenging increased along with potential incorporation of C into biomass pools. Ecosystem and community-scale associations between C and nutrient substrate availability were also reflected in patterns of resource allocation among individual genomes (imputed and assembled). Microbes associated with higher rates of soil C turnover harbored more genes for carbohydrate utilization, fewer genes for obtaining energetically costly forms of C, N and P, more ribosomal RNA gene copies, and potentially lower C use efficiency. We suggest genome clustering by functional gene suites might yield simplified guilds related to biogeochemical cycling, even when function is imputed directly from taxonomy. Our findings in a controlled model wetland ecosystem bolster evidence for the role of P in influencing soil C cycling, and our approach could be leveraged to reduce complex microbial data for trait-based modeling of soil C cycling.

Single-Cell Genomics Reveals Co-Metabolic Interactions within Uncultivated Marine Group A Bacteria

Hawley, Alyse, K.* (alysekh@mail.ubc.ca)[1]; Nobu, Masaru K.[2]; Wright, Jody J.[1]; Sage, Brent[1]; Schwientek, Patrick[3]; Swan, Brandon[4]; Rinke, Christian[2, 3]; Liu, Wen-Tso[2]; Stepanauskas, Ramunas[4]; Woyke, Tanja[3]; Hallam, Steven J.[1]

[1]University of British Columbia, Vancouver, BC, Canada. [2]University of Illinois, IL, USA. [3]US Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [4]Bigelow Laboratory for Ocean Sciences, ME, USA.

Marine Group A (MGA) bacteria represent a ubiquitous and abundant candidate phylum enriched in oxygen minimum zones (OMZs). Despite MGA prevalence little is known about their ecology and biogeochemistry. Here we chart metabolic potential of 26 MGA single-cell amplified genomes (SAGs) sourced from different environments spanning ecothermodynamic gradients including surface ocean, OMZs and methanogenic environments. Metagenomic contig recruitment to SAGs combined with tetra-nucleotide frequency resolved nine MGA population genome bins. All population genomes exhibited genomic streamlining with surface ocean MGA being the most reduced. Different strategies for carbohydrate utilization, carbon fixation energy metabolism and respiratory pathways were identified between population genome bins, including various roles in the nitrogen and sulfur cycles. MGA inhabiting OMZ oxyclines encoded genes for partial denitrification with potential to feed into anammox and nitrification as well as a polysulfide reductase with a potential role in the cryptic sulfur cycle. MGA inhabiting anoxic waters, encoded NiFe hydrogenase and nitrous oxide reductase with the potential to complete partial denitrification pathways previously linked to sulfur oxidation in SUP05 bacteria. MGA from methanogenic environments encoded genes mediating syntrophic interactions with fatty acid degraders and methanogens including reverse electron transport. The MGA phylum appears to have evolved alternative metabolic innovations adapting subgroups to occupy specific niches along ecothermodynamic gradients. In turn, differential expression of MGA genes along OMZ oxyclines indicates that these subgroups increasing form co-metabolic interactions under energy limiting conditions promoting a cooperative mode of existence with important implications for coupled C, N and S cycling in the ocean.

Minimization of Chimeras and Substitution Errors in Full-Length 16S PCR Amplification

Heiner, Cheryl* (cheiner@pacificbiosciences.com); Oh, Steve; Hall, Richard J.; Hon, Lawrence

PacBio, Menlo Park, CA.

The constituents and intra-communal interactions of microbial populations have been shown to play an important role in many aquatic and terrestrial ecosystems, and in the healthiness of various plants and animals. One popular, efficient method of profiling microbial communities is to amplify and sequence the evolutionarily conserved 16S rRNA sequence. Currently, most targeted amplification focuses on short, hypervariable regions of the 16S sequence. Distinguishing information not spanned by the targeted region is lost and species-level classification is often not possible.

SMRT Sequencing easily spans the entire 1.5 kb 16S gene, and in combination with highly-accurate single-molecule sequences, can improve the identification of individual species in a metapopulation. However, when amplifying a mixture of sequences with close similarities, the products may contain chimeras at rates as high as 20-30%. These PCR artifacts make it difficult to identify novel species, and reduce the amount of productive sequences.

We investigated multiple factors that have been hypothesized to contribute to chimera formation, such as template damage, denaturing time before and during cycling, polymerase extension time, and reaction volume. Of the factors tested, we found two major related contributors to chimera formation: the amount of input template into the PCR reaction and the number of PCR cycles.

Sequence errors generated during amplification and sequencing can also confound the analysis of complex populations. Circular Consensus Sequencing (CCS) can generate single-molecule reads with >99% accuracy, and the SMRT Analysis software provides filtering of these reads to >99.99% accuracies. Remaining substitution errors in these highly-filtered reads are likely dominated by mis-incorporations during amplification. We compared predicted vs empirical accuracy of full-length 16S CCS reads from amplicons generated with several commercially-available PCR kits, and determined the types of errors found at different levels of predicted accuracy.

We show results of our experiments and describe an optimized protocol for full-length 16S amplification for SMRT Sequencing. These optimizations have broader implications for other applications in which amplicons are sequenced to phase variations across targeted regions or to generate highly accurate reference sequences.

Engineering Anaerobic Gut Fungi for Lignocellulose Breakdown

Henske, John K.* (johnhenske@engineering.ucsb.edu)[1]; Gilmore, Sean P.[1]; Solomon, Kevin V.[1]; Theodorou, Michael K.[2]; Brewer, Heather M.[3]; Purvine, Samuel O.[3]; Wright, Aaron T.[3]; Kuo, Alan[4]; Mondo, Stephen[4]; Salamov, Asaf[4]; LaButti, Kurt[4]; Thompson, Dawn[5]; Regev, Aviv[5]; Grigoriev, Igor V.[4]; O'Malley, Michelle A.[1]

[1]Department of Chemical Engineering, University of California, Santa Barbara. [2]Animal Production, Welfare, and Veterinary Sciences, Harper Adams University, Newport Shropshire, TF10 8NB, United Kingdom. [3]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352. [4]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598. [5]Broad Institute of MIT and Harvard, Cambridge, MA 02142.

To support renewable technologies, it is necessary to develop more efficient methods to extract sugars from crude plant biomass (lignocellulose). While plants contain cellulose that depolymerizes into fermentable sugars for microbial biofuel production, it is trapped within lignin, hemicellulose and other biopolymers that complicate its hydrolysis. To address this issue, one can turn to nature, particularly to microbes that routinely degrade plant biomass. Many large herbivores, such as cows and horses, harbor a consortium of microbes in their digestive tracts that convert recalcitrant biomass into sugars. Within this consortium, anaerobic gut fungi are the primary colonizers of plant material, and represent a rich source of biomass degrading enzymes. We have used transcriptomics to identify biomass degrading enzymes produced by the gut fungi and, moreover, have examined regulation patterns to determine how the suite of enzymes is tailored to the carbon source present.

By providing a pulse of a simple sugar during growth on un--pretreated biomass we have studied how the transcriptome remodels to match the perturbation. This reveals how quickly gut fungi respond to alter the expression of biomass degrading enzymes when they are no longer necessary. Additionally, patterns of co--regulated transcripts may provide insight into the putative function of transcripts with unknown function. Cluster analysis of transcriptomic regulation data reveals co--regulated groups enriched in biomass degrading function, primarily consisting of cellulolytic and/or hemicellulolytic enzymes. Further work examining the transcriptome on a variety of substrates ranging in complexity, including glucose, cellobiose, cellulose, and biomass provides further insight to how these genes are regulated. Gene set enrichment analysis reveals not only how the expression of biomass degrading genes changes across these substrates, but also how the variety of biomass degrading enzymes changes. In the

case of two isolated strains, Neocallimastix californiae and Piromyces finnis, there is a gradual increase in variety of biomass degrading enzymes with the increase in complexity of the carbon source. However, in the case of a third isolated strain, Anaeromyces robustus, GSEA reveals that cellobiose triggers the increase in expression of a wide variety of enzymes, not isolated to those involved in hydrolyzing cellobiose and cellulose. Further, we can examine the changes in expression of core metabolic proteins to determine how the fungal cells tailor their metabolism across these different conditions.

Recently, in partnership with the JGI, we have acquired genomic information for each of these isolates. Combining the regulation information with genomic localization, we can begin to identify potential promoters that control expression of these important genes. By investigating the DNA regions upstream of genes of similar functions and similar regulation patterns, we are currently working to identify consensus sequences that may be controlled by the same transcription factors. Together, this approach will enhance our efforts to develop new tools to genetically engineer the anaerobic fungi.

A Complex System for a Complex Phenotype: Synthetic Communities and the Plant Microbiome

Herrera Paredes, Sur* (sur@email.unc.edu)[1, 2]; Castrillo Molina, Gabriel[2]; Lebeis, Sarah[2, 3]; Esteban, Jose Macalino[2]; Pereira Lima Teixeira, Paulo José[2]; Jones, Corbin D.[1, 2]; Dangl, Jeff[1, 2, 4]

[1]Bioinformatics and Computational Biology Curriculum, The University of North Carolina at Chapel Hill, North Carolina, USA. [2]Department of Biology, The University of North Carolina at Chapel Hill, North Carolina, USA. [3]Department of Microbiology, The University of Tennessee Knoxville, Tennessee, USA. [4]Howard Hughes Medical Institute, USA.

Plants and microbes may form any ecological interaction, from antagonism to mutualism. Research focusing on binary interactions (when only one plant and one microbial species are considered) has revealed an impressive amount of molecular details that explain the mechanisms behind these interactions. However, in nature plant-microbe interactions occur in the background of a complex biotic environment. With this in mind, we have implemented an experimental system that allows us to simultaneously quantify dozens of bacterial isolates in the context of a plant host.

We use these synthetic communities to study plant-microbe interactions in a realistic setting. We have combined results from synthetic community experiments with observations on natural soil, and we identified a gradient of enrichments of bacterial functions in close proximity to plant roots; these gradient is dominated by ABC transporters and metabolic genes. We developed an experimental design approach, termed bacterial combinatorics, that systematically varies the bacterial community encountered by the plant, while retaining enough statistical power to identify bacteria that alter specific plant phenotypes. We coupled this approach with image-based plant phenotyping, and showed that we can identify bacteria that alter plant size and color; furthermore, we quantified the degree of bacterial additivity on these phenotypes, and found that plant size differences are explained mainly by additive bacterial effects; in contrast, interactions are needed to explain differences in plant color. A variant of these approach, termed bacterial rolling circle, demonstrated that data from binary association experiment (one plant and one bacteria species only) informs synthetic community designs to alter phosphate accumulation in plants; in line with our results for plant size, we found that bacterial additive

effects are sufficient to explain most differences. Plant transcriptomes have shown that bacterial communities can alter the plant transcriptional response to phosphate starvation, and will indicate what plant genes and pathways mediate the bacterial effects on plant phenotypes.

Synthetic communities combine the best of systems-level and reductionist approaches and will lead to genetic and molecular mechanisms that are representative of processes that occur in the complexity of natural environments.

Comparative Genomics of N_2O-Producing Fungi Reveals Extensive Denitrification Gene Modularity and Cryptic Nitrate Utilization Genotypes

Higgins, Steven A.* (shiggin8@vols.utk.edu)[1]; Schadt, Christopher W.[1, 2, 3]; Löffler, Frank A.[1, 2, 3, 4, 5]

[1]Department of Microbiology, University of Tennessee, Knoxville, TN, USA. [2]University of Tennessee and Oak Ridge National Laboratory (UT-ORNL) Joint Institute for Biological Sciences (JIBS), Oak Ridge, Tennessee, USA. [3]Biosciences Division, Oak Ridge, Tennessee, USA; [4]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, USA. [5]Center for Environmental Biotechnology, University of Tennessee, Knoxville, TN, USA.

The nitrogen (N) cycle is Earth's most disrupted biogeochemical cycle. Human technology and activity have altered the N-cycle due to the industrial fixation of nitrogen gas for use in agricultural fertilizers. Increased N input enhances microbial activity responsible for formation of nitrous oxide (N_2O), a potent greenhouse gas. Denitrification (production of N_2O or N_2 from nitrate or nitrite) is implicated in increased N_2O emissions from soils. While bacterial N_2O production has been extensively studied, fungi are increasingly recognized for their role in N_2O formation in soils, especially agroecosystems where N input is high. Despite this recognition, key physiological phenomena related to fungal N metabolism remain unclear. Specifically, closely related soil fungi display conflicting phenotypes related to nitrate (NO_3-) utilization and N_2O formation, but precise mechanistic explanations for these observations are lacking. Variation in NO_3- utilization is of particular importance, because fertilizer type and application may have a large effect on fungal N_2O production in situ. We cultivated dozens of N_2O-producing fungi from agricultural soils in the Midwest Cornbelt belonging to the phylum Ascomycota. Laboratory incubations revealed that some isolates produced N_2O when grown in the presence of NO_3-; however, not all isolates of the same genus (e.g., Fusarium, Trichoderma) share this phenotype. To begin to explore potential genetic determinants underlying these contrasting phenotypes, we sequenced the genomes of four fungal isolates using Illumina MiSeq v3 chemistry (yielding ~62.7 million reads) and queried the draft genome assemblies for genes implicated in NO_3- utilization and N_2O production. Application of a hybrid De Bruijn graph and consensus assembly protocol resulted in 2,190 – 4,721 contigs with genome N50 values between 11,276 to 61,351 bp. Preliminary genome size estimates ranged from 26.5 to 47.7 million bp with 52 to 127X sequence coverage. Alignment of reference proteins (FunGene database) catalyzing key steps in the denitrification pathway (NapA, NarG, NirK, NirS, NorB, P450nor, and NosZ) revealed translated genomic regions with similarity only to NapA, NirK, and P450nor. Protein-protein alignments with sequences from 536 JGI fungal genomes revealed that 48 contain homologs of napA (encoding periplasmic nitrate reductase), 39 contain nirK (copper-containing nitrite reductase), and 95 possess p450nor (nitric oxide reductase). Interestingly, only one genome (Aspergillus zonatus) contained a single copy of narG (encoding membrane-bound nitrate reductase), whereas all other Aspergillus species possessed napA homologs. No matches to nirS (encoding cytochrome cd1-containing nitrite

reductase) or nosZ (encoding nitrous oxide reductase) were identified. Of the 48 napA-containing fungal genomes, 20 also possess p450nor, suggesting the potential for nitrate reduction is extensive, but why many fungi do not produce N_2O from added nitrate remains unclear. Importantly, the detection of p450nor in more fungal genomes than nirK suggests p450nor presence is a better indicator of fungal N_2O production potential. The dissimilatory reduction of N-oxyanions in fungi results in N_2O as the final end product of denitrification. Hence, an improved understanding of how N cycle pathways, genes, and their interactions results in greenhouse gas production in fungi is key to mitigating fungal N_2O production, especially in N rich agroecosystems.

The Yeast Biodesign Library: Leveraging DNA Synthesis to Assess and Harness Genes from Diverse Organisms

Hittinger, Chris Todd* (cthittinger@wisc.edu)[1, 2]; Alexander, William G.[1, 2]; Rovinskiy, Nikolay[1, 2]; Lopes, Mariana[2, 3]; Peris, David[1, 2]; Kang, Jin[1, 2]; Tarver, Angela[4]; Harmon-Smith, Miranda[4]; Cheng, Jan-Fang[4]; Deutsch, Samuel[4]; Piotrowski, Jeff S.[1]; Sato, Trey K.[1]; Gasch, Audrey P.[1,2]

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI. [2]Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI; [3]Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [4]US Department of Energy Joint Genome Institute, Walnut Creek, CA.

Two major issues encountered by biofuel researchers in the optimization of Saccharomyces cerevisiae cellulosic ethanol strains are the tolerance to toxins in hydrolysates derived from the lignin in plant biomass and the conversion of pentoses, disaccharides, and other disfavored sugars. The latter is especially problematic because the vast majority of S. cerevisiae strains are unable to ferment the pentose xylose, the major component of hemicellulose, without genetic modification. Several yeasts that belong to a large group known as the “CUG clade” are capable of fermenting xylose and cellobiose readily; unfortunately, these yeasts use an alternative genetic code, which complicates the evaluation of their genes in S. cerevisiae. To address these difficulties, we designed 266 synthetic gene cassettes called the Yeast Biodesign Library. Most of these genes were predicted from the genome sequences of the xylose-fermenting species Spathaspora passalidarum, as well as several other recently discovered species of Spathaspora. Each cassette is driven by a medium-strength constitutive S. cerevisiae promoter/terminator pair and contains an open reading frame that has been cleansed of problematic restriction enzyme sites and incompatible codons. To date, we have successfully used the Yeast Biodesign Library to generate synthetic multigene pathways and to implicate several novel genes in xylose fermentation. Barcode sequencing experiments are in progress to quantify the effect of each cassette on growth in media containing xylose and in ligonocellulosic hydrolysates.

Building Better Biosensors for Biocatalyst Discovery from Metagenomic Libraries

Ho, Joe C.H.* (joeho776@gmail.com)[1, 2]; Pawar, Sandip V.[3, 4]; Strachan, Cameron R.[3]; Yadav, Vikramaditya G.[2, 4]; Hallam, Steven J.[1, 2, 3, 5]

[1]Department of Microbiology and Immunology, The University of British Columbia, Vancouver, BC, Canada V6T 1Z3. [2]Ecosystem Services, Commercialization Platforms and Entrepreneurship

(ECOSCOPE) Program, The University of British Columbia, Vancouver, BC, Canada V6T 1Z3. [3]MetaMixis, Inc., Vancouver, BC, Canada V6T 1Z3. [4]Department of Chemical & Biological Engineering, The University of British Columbia, 2360 East Mall, Vancouver, BC, Canada V6T 1Z3. [5]Graduate Program in Bioinformatics, The University of British Columbia, Vancouver, BC, Canada V5Z 4S6.

One-component transcriptional regulators and their cognate promoters can be harnessed in the development of transcriptional fusions driving the expression of reporters like GFP or luciferase. These biosensors can in turn be used to discover biocatalysts in functional screening of metagenomic libraries. Despite their adaptability, however, current biosensors typically have narrow detection ranges, low sensitivity, and non-linear responses to different substrate concentrations. These limitations can restrict the use of biosensors as search functions in metagenomic screens. Indeed, without a sensitive and selective biosensor, encoded activities conferred by environmental DNA cannot accurately be reported in a high-throughput manner. With this in mind, we have made generalizable improvements to a lignin transformation biosensor, PemrR:GFP, previously identified from an E.coli GFP promoter-trap library. We develop a mathematical model describing the behaviour of PemrR:GFP and use this model to optimize and reprogram the biosensor based on expression of the emrR transcriptional regulator. Dynamic control of emrR expression under increasing constitutive promoter strength improved the dynamic range of fluorescence output across different substrate concentrations. This response was directly related to promoter copy number leading to a four-fold linear increase in PemrR:GFP sensitivity to lignin transformation products including vanillin and syringaldehyde. Using the improved biosensor, we employed microplate-based co-culture screening of lignin transformation activities associated with 24 previously identified fosmid clones grown with hard-wood craft lignin. Our results confirmed the capacity of these clones to activate the improved biosensor with increased sensitivity and improved dynamic range. Biosensor-based detection of lignin transformation products provides a powerful screening paradigm that can be used across a wide range of feedstocks to recover genes or gene cassettes conferring differential lignin transformation profiles. By tuning sensors to different mono-aromatic substrates it becomes possible to detect combinations of environmental clones that synergize in combination to produce specific product profiles of interest and utility to the biorefining community.

Measuring Patterns by Geographical Locations in Marine Metagenome Data Using Newly Adopted Genotyping by Sequencing

Hoon Je Seong* (s4645@naver.com)[1]; Chung Yeon Hwang[2]; Hong Hee Won[3]; Woo Jun Sul[1]

[1]Department of Systems Biotechnology of Chung-Ang University, Anseong, Korea. [2]Division of Polar Biology and Ocean Sciences, Korea Polar Research Institute, Incheon, Korea. [3]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America.

As handling of the huge quantity of metagenomic data is difficult and time-consuming, we proposed more effective metagenomic tool to analyze particular sequences that are related to restriction enzyme sites for targeting reduced numbers of genes rather than to estimate whole genome sequences. This is adapted from a novel approach, Genotyping-by-sequencing (GBS) procedure in plants for analyzing marine metagenome. In this study, general whole genome shotgun sequencing method was compared to metagenomic analysis through GBS procedure for verifying the efficiency. Additionally single nucleotide polymorphisms (SNPs) in restriction enzyme associated sequences by environmental changes were compared by gene's functional categories. Our GBS–metagenome procedure was demonstrated with marine epipelagic samples collected from 10 sites during the voyage from East Sea to Bering (~8,000

km) in July, 2013. These samples were processed by a restriction enzyme “ApeKI” and then sequenced by Illumina HiSeq by yielding 7.31 billion. In addition to environmental study, just 4 same site samples were sequenced by Illumina MiSeq for whole genome shotgun sequencing with 9.55 billion reads to compare GBS’s efficiency. Metagenomic reads were analyzed basically using the HMP Unified Metabolic Analysis Network (HUMAnN) and metagenomic phylogenetic analysis2 (MetaPhlAn2) pipeline to predict abundance of microbial pathways and taxonomic profiles in the community. Microbial pathway composition and abundance patterns were compared between GBS method and general metagenomic method. MiSeq metagenomic sequences were also cut at “CGWCG” artificially recognized by “ApeKI” then were compared with GBS procedure reads. These restriction enzyme associated sequences were grouped (clustered), aligned, and annotated using UCLUST, MUSCLE, and tBLASTX respectively. Entropy in sequence variation among clusters sorted by functional category was measured to evaluate their rates of SNPs variations. Among 10 sites, there were differences of entropy scores in clusters assigned to rhodopsin, ABC transporter, and metabolic genes between East Sea (west side of Japan) and Western Pacific Ocean (East side of Japan). Higher entropy values in Western Pacific Ocean represent the genetic differences than those in East Sea abided by environmental alteration, geographical location and the influences. Therefore GBS-metagenome is very helpful to understand the community quickly resolving budget and time-consuming issues. GBS-metagenome can also identify nucleotide variations in each functional groups through comparing each site and more.

Rapid Sulfur Cycling in Prairie Pothole Lake Sediments Drives Extensive Carbon Mineralization

Hoyt, David W.* (david.hoyt@pnnl.gov)[1]; Dalcin Martins, Paula[2]; Johnston, Michael D.[3]; Tfaily, Malak[1]; Wilkins, Michael J.[2, 3]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory Richland, WA, 99352. [2]Department of Microbiology, The Ohio State University, Columbus, OH, 43210. [3]School of Earth Sciences, The Ohio State University, Columbus, OH, 43210.

The fate of organic carbon in terrestrial inland waters (e.g., lakes and wetlands) is one of the most poorly constrained components in the global carbon cycle, mainly due to the heterogeneity of such aquatic systems. In the prairie pothole region (PPR) of North America, millions of small freshwater lakes are scattered across a 750,000-km2 region of the upper Midwest and southern Canada. These lake sediments contain some of the highest dissolved organic carbon (DOC) concentrations ever measured in freshwater systems, in addition to abundant concentrations of both oxidized and reduced sulfur species. As such, ongoing research is assessing the importance of these environments in carbon mineralization, and associated biogeochemical processes such as methane generation. Radiotracer (35SO42-) measurements have revealed extremely high rates of sulfate reduction in prairie pothole lake (PPL) sediments, generating significant quantities of reduced sulfur species that likely react abiotically with DOC to generate reduced organic CHOS species. We hypothesize that substrates for sulfate reduction are supplied via rapid fermentation of DOC and the reoxidation of reduced S species at the sediment-water interface where steep redox gradients exist. Proton-NMR analyses of PPL sediment pore fluids have detected high concentrations of labile carbon substrates for sulfate reducing bacteria, including alcohols (ethanol, isopropanol), and organic acids (acetate), while metagenomic analyses have uncovered abundant functional genes linked to sulfur cycling (both oxidative and reductive dsrA), supporting a microbial role in S oxidation at the sediment-water interface. Surprisingly, abundant methane has also been detected in active sulfate-reducing zones, suggesting that non-competitive substrates such as

methanol may be utilized by methanogens. Supporting this hypothesis, NMR characterizations of pore fluids identified ~ 300 µM methanol. Overall, these data indicate that rapid sulfur cycling within PPL sediments drives extensive carbon mineralization, does not thermodynamically inhibit methanogenesis, and results in the PPR likely playing an oversized role in regional carbon fluxes to the atmosphere.

Comparative Genomic Analyses of the *Chaetomiaceae* Provide Insights into Evolution and Inform the Development of Experimental Genetic Systems

Hutchinson, Miriam I.* (miramira@unm.edu)[1]; Powell, Amy J.[2]; Tsang, Adrian[3]; O'Toole, Nicholas[3]; Berka, Randy M.[4]; Grigoriev, Igor V.[5]; Barry, Kerrie[5]; Natvig, Donald O.[1]

[1]Department of Biology, University of New Mexico, Albuquerque, NM. [2]Sandia National Laboratories, Albuquerque, NM. [3]Centre for Structural and Functional Genomics, Concordia University, Montreal, QC. [4]Novozymes Inc., Davis, CA. [5]US Department of Energy Joint Genome Institute, Walnut Creek, CA.

The fungal family Chaetomiaceae possesses a diversity of mesophilic and thermophilic species. It is the hallmark fungal group for thermophilic species that are also adapted to microhabitats rich in cellulosic substrates. Thermophilic Chaetomiaceae are of interest in industry because they produce thermally stable enzymes and grow under high temperatures that prevent contamination from other microorganisms. Moreover, multiple shifts between mesophily and thermophily have occurred in the family. Members of the family therefore present broad opportunities for studying fungal cell biology, evolution and ecology; and they present opportunities to develop cell factories for biotechnology, including the decomposition of complex carbohydrates for bioenergy applications. The genome sequences of multiple thermophilic and mesophilic members of the Chaetomiaceae (Sordariales) have been acquired as part of an effort to identify evolutionary events that have shaped reproductive and life-style differences among species in the family, including the origins of thermophily. These genomes have enabled an investigation of the genetics underlying diverse mating systems of the group. We have characterized sexual reproduction in the thermophile Myceliophthora heterothallica towards the goal of establishing this organism as a model for the family. We have successfully crossed M. heterothallica strains from Indiana, New Mexico and Germany. Genome sequences from M. heterothallica and other members of the family show that protein-coding genes in the mating regions of heterothallic species are conserved relative to other outcrossing Sordariales such as Neurospora, with the important exception that mating-type a strains from heterothallic species of Chaetomiaceae possess a partial mat A-1 gene. The mat A-1 in mating-type a strains is truncated for the alpha box region required for fertility in other Sordariales, but it has a conserved intact downstream reading frame. Among species of Chaetomiaceae with teleomorphic states, heterothallism is rarely reported while homothallism is common. Analyses of genomes of homothallic species--Chaetomium globosum, C. thermophilum, Thielavia hyrcaniae and T. appendiculata--demonstrate that they possess a true mat a-1 gene, but it is not linked to the mat A-1 region. Among filamentous Ascomycota, this represents one recurring evolutionary pathway for derived homothallism, which requires mat A and mat a regions in a single haploid genome. This comparative genomic effort has also revealed likely heterothallism in a second thermophilic species, Thielavia terrestris, which has been of interest in industry and is the subject of several patents. Preliminary efforts to explore the evolution of thermophily among Chaetomiaceae have identified genes for several proteins previously associated with thermal tolerance that appear to be up-expressed in thermophiles relative to mesophiles.

Automated Analog-Digital Optofluidic Platform to Detect and Sort Microdroplets for High-Throughput Genetic Engineering

Iwai, Kosuke* (kosukeiwai@lbl.gov)[1, 2]; Kim, Peter W.[1, 2]; Raje, Manasi[1, 2]; Duncombe, Todd A. [1, 2]; Heinemann, Joshua V.[1, 3]; Deng, Kai[1, 2]; Adams, Paul D.[1, 4, 5]; Singh, Anup K.[1, 2]

[1]Technology Division, Joint BioEnergy Institute (JBEI), Emeryville, CA/USA. [2]Applied Biosciences and Engineering, Sandia National Laboratories, Livermore, CA/USA. [3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Lab, Berkeley, CA/USA. [4]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Lab, Berkeley, CA/USA. [5]Department of Bioengineering, University of California at Berkeley, Berkeley, CA/USA.

In the recent years, synthetic biology has dramatically grown and became significantly important for both of scientific researches and industrial applications such as biofuel and pharmaceutical applications. However, multiple genetic engineering steps required for synthetic biology are often time-consuming and labor-intensive with repetitive pipetting and plating. Therefore, automated and efficient processes to perform molecular biology assays have been long desired.

Microdroplet technology that uses aqueous droplets (pl-µl size in volume) suspended in an oil phase as compartmentalized reaction chambers holds a powerful potential to drastically improve the genetic engineering steps with various benefits such as high-throughput chemical processes with faster diffusion rate, small volume of reagent consumption required, and better control of experimental environment. To fully utilize the benefits of microdroplet technologies, it is essential to develop reliable platform to manipulate droplets. Researchers have developed numbers of microfluidic components and systems, and most of such platforms can be categorized in two groups; continuous-flow (analog) microfluidics that manipulate the droplets by controlling the hydrodynamic force, and digital microfluidics that utilize surface tension from electrowetting on dielectric with arrayed electrodes. While continuous-flow microfluidics is useful to handle large numbers of droplets at once, digital microfluidics can actively manipulate target droplets in a programmable manner.

To maximize the advantages of both techniques, we have recently developed a hybrid platform of analog and digital droplet. The system has achieved multiple functions, including i) formation of droplets and encapsulation of cells suspended in aqueous medium, ii) mix and split the droplets to add specific amount and concentrations of various reagents, and iii) incubation process with localized temperature control. We successfully demonstrated the transformation of plasmids and expression of fluorescent proteins in several bacteria. However, it still requires off-chip optical measurements, which has been a bottleneck for the overall efficiency of the experimental protocols. To enhance the throughputness, here we propose a new analog-digital optofluidic platform. We integrate optical fibers in the microchannels to add on-chip capabilities for fluorescence-based detection of encapsulated cells and enzymatic activities in discrete droplets, and for autonomous sorting of the target droplets. Unlike conventional microfluidic systems, the analog-digital optofluidic device allows completely automated genetic engineering steps, and we believe that our technology would significantly accelerate the synthetic biology researches and applications such as directed evolutions.

The Ecology and Evolution of Secondary Metabolism in a Model Group of Marine Bacteria

Jensen, Paul* (pjensen@ucsd.edu); Amos, Greg; Letzel, Anne-Catrin; Ziemert, Nadine

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA USA.

Bacterial secondary metabolites are the biosynthetic products of large and complex gene clusters. These clusters are well suited for horizontal gene transfer, with acquisition events providing immediate opportunities to test small and often biologically active molecules for their effects on fitness. Those that provide a selective advantage are maintained while those that do not are lost. These acquisition events are mediated by the availability of gene clusters in the "local gene pool" and can spread through globally distributed populations. The marine actinomycete genus Salinispora has become a useful model to study the ecology and evolution of secondary metabolism. Based on 120 genome sequences provided by the Community Sequencing Program, we have observed high rates of pathway acquisition and exchange. More than 50% of the gene clusters occur in only one or two strains, suggesting recent acquisition. These recently acquired clusters largely occur in two of 21 genomic islands, suggesting a yet to be defined mechanism that directs them to specific chromosomal sites. Gene clusters that become fixed within a species are observed to migrate into more conserved regions of the core genome suggesting they are no longer available for exchange. A separate CSP transcriptomics project revealed that more than half of the gene clusters are silent under normal laboratory growth conditions. Comparative analyses indicate subtle regulatory differences among related gene clusters that explain why some clusters will remain permanently silent. These observations suggest that regulatory element deletion represents an alternative or intermediate strategy to gene cluster loss. Comparative "omics" are providing unique opportunities to infer the evolutionary histories of biosynthetic gene clusters independent of the bacteria in which they reside and to understand the processes that generate secondary metabolite structural diversity.

Rhizosphere Metagenomics and Metatranscriptomics of Biofuel Crops

Jiarong Guo* (guojiaro@gmail.com)[1]; Aaron Garoutte[1]; Adina Howe[2]; Qiong Wang[1]; Titus Brown[3]; James Cole[1]; James Tiedje[1]

[1]Michigan State University. [2]Iowa Sate University. [3]University of California, Davis.

We applied deep sequencing (~1TB) to study rhizosphere microbiomes of 3 biofuel crops: corn (annual) and switchgrass and Miscanthus (perennials). Soil samples were taken at senescence from seven replicate plants for each crop and from very close or adjacent to small and fine roots. We compared the rhizosphere metagenomes at three levels: overall community structure (SSU rRNA gene), overall function (annotation from global assembly), and N cycle genes (from Xander, a targeted gene assembly tool). All three levels showed corn had a significantly different community from Miscanthus and switchgrass (except for AOA), and that the two perennials showed a trend of separation. In terms of life history strategy, the corn rhizosphere was enriched with more copiotrophs while the perennials were enriched with oligotrophs, which is further supported by higher abundance of genes in "Carbohydrates" subsystem category and higher fungi/bacteria ratios. In addition, corn also had a less rich and even community, and with a larger dispersion of replicate data in ordination plots. The corn rhizosphere was also enriched in Penicillium (non-beneficial fungi) which may indicate stress to the corn system in

this severe drought year. The nitrogen fixing community of corn was dominated by nifH genes most closely affiliated to Rhizobium (perhaps a legacy from prior legume crops) while the perennials had nifH sequences most related to Coraliomargarita, Novosphingobium and Azospirillum, indicating that the perennials selected independent beneficial members. Moreover, higher numbers of genes for nitrogen fixation and lower number of genes for nitrite reduction suggest better nitrogen sustainability of the perennials. These data indicate that perennial bioenergy crops have advantages over corn in higher microbial species richness and functional diversity and in selecting members with beneficial traits, consistent with a higher level of sustainability of perennial biofuel crops. We also conducted a 'core functional microbiome' analysis of switchgrass using a multi-omic approach on samples taken at maximum plant growth in the year following the above study. We used metagenomics and metatranscriptomics to identify shared transcribed functions. The minimum functional core accounted for 99% or 92% of functional annotations depending on annotation database (SEED Subsystems or RefSeq respectively). We then used metaproteomic data to further confirm our minimum functional core. The metaproteome data had 460 unique SEED Subsystem annotations with an abundance of 876,429. All but 12 of the SEED Subsystem annotations were found in the minimum functional core further validating the minimum functional core. While functions related to cellular maintenance and housekeeping were highly abundant within the minimum functional core, functions related to biogeochemical cycling and plant growth promotion were present within the minimum functional core. Specifically we found evidence that the most abundant nitrogen cycling processes were related to ammonia assimilation. Phosphate metabolism was a highly active component of the phosphorous cycle. Carbon cycling enzyme related to glycoside hydrolases and lignin break down were abundant especially in the metaproteome.

Genetic and Genomic Studies of Local Adaptation in *Panicum hallii*

Juenger, Thomas E.* (tjuenger@austin.utexas.edu)[1]; Lovell, John[1]; Gould, Billie[2]; Lowry, David[2]; Jenkins, Jerry[3]; Schmutz, Jeremy[3, 4]

[1]University of Texas at Austin. [2]Michigan State University. [3]HudsonAlpha Institute for Biotechnology. [4]US Department of Energy Joint Genome Institute, Walnut Creek, California.

Panicum hallii is a C4 perennial grass adapted to a diversity of habitats and environmental gradients across the southwestern US. Interest in developing genomic resources for P. hallii has been driven by its close relationship to the bioenergy crop switchgrass (Panicum virgatum). Much like upland and lowland ecotypes of switchgrass, there are two major ecotypes of P. hallii. P. hallii var. hallii is found in xeric habitats and var. filipes occurs primarily in coastal mesic habitats. Our collaborative research group has used a variety of genetic and genomic tools to explore the genetic basis of this divergence, leveraging resources generated in collaboration with the DOE JGI. In particular, the completion of the version 2 genome assembly opens to researchers a host of tools for exploring C4 perennial grass biology. Here, I briefly report on progress on the genome assembly, genetic mapping and population genomic analyses, and expression QTL studies to elucidate divergence and adaptation to drought in P. hallii.

Understanding Mechanisms of Microbe-Microbe Interactions Using Patterns of Genome Content Evolution

Kamneva, Olga K.* (okamneva@stanford.edu)

Department of Biology, Stanford University, 371 Serra Mall, Stanford, CA 94305-5020, USA.

Inter- and intraspecific ecological interactions are widespread in microbial communities and can be predicted from concerted changes in microbial abundance using various statistical approaches. However, understanding the interaction mechanisms still requires extensive experimental investigations. Several studies of candidate pathways and metabolic modules have reported genome content differentiation within gene sets that facilitate the interaction of two partner species or strains. This differentiation, arising on the species tree branches where interaction also emerges, could serve as a genomic signature of the involvement of a gene set in the interaction at the molecular level. Here I report a new computational approach designed to detect such gene sets in the genomes of a pair of known interacting partners. Stochastic character mapping is used to predict expected genome content of the partners given genome content of related taxa not known to engage in the ecological association. I compare expected and observed genome content of partner species to find gene sets characterized by unexpectedly high differentiation in their representation, suggesting that these gene sets potentially facilitate microbe-microbe interactions in the given species pair. Finally, I apply this new method to members of phototrophic consortia Chlorochromatium aggregatum: Comamonadaceae bacterium CR and Chlorobium chlorochromatii CaD3, to predict several gene sets potentially mediating interaction between these microbes.

Comparative Genomics of C4 Grasses

Kellogg, Elizabeth A.* (ekellogg@danforthcenter.org)[1]; Mayfield-Jones, Dustin[1]; Studer, Anthony J.[2]; Santana de Carvalho, Daniel[3]; Schnable, James C.[3]; Brutnell, Thomas P.[1]

[1]Donald Danforth Plant Science Center. [2]University of Illinois. [3]University of Nebraska-Lincoln.

C4 photosynthesis is a high efficiency pathway that is shared by most biofuels crops. It has originated over 60 times independently in the history of flowering plants, including ca. 12 origins in the grass subfamily Panicoideae alone, providing a naturally replicated experiment in parallel evolution. Changes in gene regulation or protein sequence that occurred independently in multiple origins are likely to be required for C4 function. Until recently, genomic-scale comparisons of C3 and C4 photosynthesis in the grasses have compared the C4 grasses in the Panicoideae with distantly related C3 species such as rice, brachypodium and wheat, confounding differences resulting from the transition from the C3 to the C4 photosynthetic pathway with other unrelated changes over an estimated fifty million years of independent evolution. Focusing on both C3 and C4 species within the Panicoideae mitigates this problem while maximizing the number of independent C4 origins that can be compared. We are generating “draft grade” genome assemblies from five diploid panicoid grass species: Andropogon virginicus (broomsedge bluestem, C4, NADP-ME), Paspalum vaginatum (seashore paspalum, C4, NADP-ME), Dichanthelium oligosanthes (Heller’s rosette grass, C3), Chasmanthium laxum (slender woodoats, C3) and Urochloa fusca (browntop signalgrass, C4, PCK). High quality draft genomes for Dichanthelium and Urochloa have been completed and are providing a basis for syntenic analyses that would not have been possible using gene trees alone. The availability of the Dichanthelium genome has already helped

interpret which changes in C4 panicoid species are linked to the evolution of C4 photosynthesis, and which are simply the result of shared phylogenetic history. The Paspalum genome is proving to be useful for a wider community interested in salt tolerance.

Leaf transcriptomes were produced for all species, taking advantage of a continuous developmental gradient characteristic of all grass leaves. In developing leaves, the leaf base is less mature than the leaf tip, so gene expression in different leaf parts captures discrete developmental windows in the synthesis and assembly of the photosynthetic apparatus. Four leaf segments were collected from multiple individuals of all species except Andropogon and RNA seq libraries were constructed. Transcriptome profiling is complete for Paspalum and Dichanthelium, and in progress for Urochloa; RNA isolation is in progress for Chasmanthium.

Simplified Bacterial Communities Conferring Productivity Enhancements to Biofuels-Producing Algae

Kimbrel, Jeffrey A.* (kimbrel1@llnl.gov)[1]; Samo, Ty J.[1]; Nilson, Daniel J.[1]; Nuccio, Erin[1]; Pete-Ridge, Jennifer[1]; Siccardi, Anthony[2]; Lane, Todd[3]; Mayali, Xavier[1]

[1]Lawrence Livermore National Laboratory, Livermore CA. [2]Texas A&M Agrilife, Corpus Christi, TX. [3]Sandia National Laboratories, Livermore, CA.

The LLNL Biofuels SFA seeks to support robust and sustainable microalgae fuel production through a systems biology understanding of algal-bacterial interactions. We hypothesize that by understanding the factors that control cellular physiology and biogeochemical fluxes in and out of algal cells, particularly through the phycosphere, we can advance the efficiency and reliability of algal biofuel production. We surveyed nearly 100 simplified bacterial community compositions from several natural and algal-pond sources enriched with axenic Phaeodactylum tricornutum (Pt) or Nannochloropsis salina (Ns), both model biofuel-producing microalgal strains. These enrichments were evaluated for algal productivity including chlorophyll fluorescence, growth rate and yield, under 12- and 24-hour light-dark cycles and compared to axenic algal cultures. These measures of algal productivity were correlated to microbial abundance, either as clusters of co-occurring bacteria, groups of bacterial taxa or individual microbes.

In all simplified communities, the majority of bacteria belong to the Rhodobacteraceae family which are major contributors to algal health in natural ecosystems. While there were some shared taxa between the Pt and Ns enrichments, these algal-hosts largely selected for different genera in both the free-living and phycosphere-associated bacteria, with certain taxa having opposing effects on algal productivity. For example, some Loktenella species were positively correlated to Ns productivity and negatively correlated to Pt, while Loktenella species had the opposite effect. The positive growth effects seen so far suggest that maintenance of algal-bacterial mutualisms can assist with establishing economically feasible algal biofuels.

Future work will further characterize these relationships using experimental and bioinformatic approaches and identify the chemical and ecological mechanisms underpinning these complex symbioses. Using genomic and transcriptomic information we will build complex metabolic models to classify the role of these probiotic microbes, e.g. as nutrient providers or pathogen suppressors.

Exploring the Underpinnings of Plant-Microbial Community Interactions: a Defined Microbiome for *Brachypodium*

Knelman, Joseph* (jknelman@lbl.gov)[1, 5]; Katz, Andrew[1, 5]; Manganiello, Gelsomina[2]; Cartwright, Amy[1, 5]; Mueller, Urlich[3]; Dangl, Jeff[4]; Matteo, Lorito[2]; Louie, Katherine[5]; Northen, Trent R.[5]; Vogel, John[1, 5]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, CA. [2]University of Naples Federico II [3]University of Texas, Austin, TX. [4]University of North Carolina, Chapel Hill, NC. [5]Environmental Genomics and Systems Biology, LBNL, Walnut Creek, CA.

While past research has focused on the interactions between plants of economic interest and individual strains of microbes, in the field, plant-microbe interactions are driven by feedbacks between the plant and a diverse community of microbes in its entirety. Because the complexity of soil microbial communities hinders our ability to study plant-microbe interactions, we leverage an experimentally reproducible, defined community of microbes associated with the roots of Brachypodium distachyon, a JGI flagship and model organism, to examine the biological underpinnings of plant-microbe interactions at this community level. We have selected a diverse set of putatively plant growth promoting microbial isolates from B. distachyon and Arabidopsis in order to assemble a semi-complex, reproducible microbiome. Further, we have designed an experimental setup, by which microbial communities are assembled and inoculated into pots with Brachypodium, which are then grown up under controlled conditions of a growth chamber. Employing our experimental setup, we have tested the effect of the defined microbial community across nutrient gradients and assessed the effect of monoassociations on Brachypodium growth. Our preliminary data demonstrate that monoassociations with both fungi and bacteria as well as associations with the full constructed bacterial community may yield phenotypic responses in Brachypodium, and promote plant growth as measured by aboveground biomass. The addition of microbial communities under non-nutrient limiting conditions resulted in phenotypic differences, including increased biomass, indicating that the mechanism for microbial community effects on plant growth may draw on interactions outside of nutrient dynamics. DNA has been extracted from these samples 16S sequencing has been employed to assess the reproducibility and stability of root associated bacterial communities and generate higher resolution insights into community membership and structure. Using 16S rRNA gene sequencing as well as metatranscriptomics at the JGI, we will employ our experimental system to screen an existing collection of thousands of B. distachyon mutant lines to examine the genomic basis of plant-microbe interactions involved in plant growth promotion. In total, we aim to continue to build out this defined microbiome system as a foundational resource to explore questions in biology and ecology related to plant-microbe interactions that will be essential to growing agricultural plants amidst the demands of population growth and environmental change.

Genomic Approaches to Characterize Fungus Garden Ecology and Evolution

Kopac, Sarah* (Sarah.Kopac@uconn.edu); Lee, Kevin[1]; Klassen, Jonathan

University of Connecticut.

Fungus-farming ants are a model system to study complex ecological interactions. In most fungus-farming ant species, the worker ants tend a fungus garden (cultivar) that in return provides the ants'

primary food source. Most of these ant species also carry protective bacteria (Pseudonocardia) to defend the cultivar. This symbiosis also includes specialized pathogens of the fungus garden (Escovopsis) and Pseudonocardia (black yeast). Other bacteria live in the fungus garden whose ecology remains poorly defined. Because these fungus garden bacteria include members of Serratia, and Lactococcus, which are known secondary metabolite producers, these bacteria may play a defensive role in this symbiosis. Despite the potential ecological importance of fungus garden bacteria, it remains difficult to distinguish the persistent taxa from transient species that are introduced into the fungus garden from the soil and ant-provided food. Here, we propose to address this problem by comparing the genomes of fungus garden microbes to their relatives isolated from different environments. In particular, frameshifts may indicate that the strain is undergoing genome reduction, a common phenomenon in other symbiont genomes. As an example of this approach, we sequenced the genome of a Lactococcus strain isolated from the fungus garden of the ant Trachymyrmex septentrionalis. Our preliminary analysis indicates that this genome has an increased number of frameshift mutations relative to related strains that were isolated from different environments. This highlights how genomics may be able to differentiate persistent symbionts that contribute to community function from transient immigrants that may interrupt community homeostasis.

Genomics of Anaerobic Cellulose-Degrading Fungal Symbionts of the Herbivore Gut

Kuo, Alan* (akuo@lbl.gov)[1]; Mondo, Stephen[1]; Salamov, Asaf[1]; LaButti, Kurt[1]; Henske, John K.[2]; Solomon, Kevin V.[2]; Gilmore, Sean P.[2]; O'Malley, Michelle A.[2]; Grigoriev, Igor V.[1]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, CA. [2]UC Santa Barbara Dept. of Chemical Engineering, Santa Barbara, CA.

As highly proficient degraders of cellulose, the Neocallimastigomycota fungi are key members of the microbiota of large mammalian and reptilian herbivores. Their remarkable cellulolytic capabilities have great potential for use in biofuel processing. They differ from other fungi 1) in possessing an extracellular cellulose-degrading complex called a fungal cellulosome, 2) in being obligate anaerobes, with hydrogenosomes instead of mitochondria, and 3) in developing flagella in certain life stages, placing them near the base of fungal phylogenies. All these distinctive characteristics touch on areas of intense interest to the JGI, including bioenergy, biotechnology, carbon cycling, fungal evolution, and symbiosis. The genomes of these interesting and important fungi are very AT-rich and highly repetitive, posing challenges to their genomic investigation. We have deployed PacBio long-read technology to overcome these challenges and to sequence, assemble, and annotate 3 Neocallimastigomycota genomes. We found modest numbers of genes (11-13k per genome) and gene families (< 1k clade-specific clusters), and confirmed low GC-content (< 25%), high repeat content (> 50%), absence of many aerobiosis-related genes (> 30 Core Eukaryotic Genes), and near-basal location in the fungal tree (branching with flagellated but aerobic Chytridiomycota). We are now using the genome to complement recent progress in elucidating the cellulosome using non-genome-based biochemical, proteomic, and transcriptomic methods.

Metagenomic Insights into the Response of Microbial Communities to Polycyclic Aromatic Hydrocarbons in Soil

Kwak, Yunyoung* (yun@knu.ac.kr); Shin, Jae-Ho

School of Applied Biosciences, Kyungpook National University, Daegu 41566, Republic of Korea.

Involved in the effective bioremediation of polycyclic aromatic hydrocarbons (PAHs) in situ environments, the influential role of soil microbial communities were studied by a metagenomic approach. Soil microcosms were established with the treatment of PAHs (anthracene, phenanthrene, fluorene, fluoranthene, and pyrene) and used for screening the soil sample that presenting degradation activities of these compounds. Over a 6 months period, compound degradation rates were assessed by HPLC analyses, and distinctive different degradation patterns were observed in soil microcosms treated with the low molecular weight PAH anthracene. For a better understanding of the structure and functional components of soil microbial communities, total soil DNAs were prepared from soil microcosms and sequenced on the Ion Torrent Personal Genome Machine (PGM) platform. Generated metagenomic sequences were annotated using Metagenome Rapid Annotation using Subsystem Technology (MG-RAST) pipelines. Consequently, in soil samples that demonstrated with significant degradation rates to anthracene, microbial communities were analyzed for an increase of functional components, as well as shifts of taxonomic compositions. Furthermore, metagenomic datasets derived from the soil group that harboring the fastest degradation pattern were more characterized for the potential of microbial responses to the compound degradation. These kinds of informative results can contribute to suggest a significant meaning of soil microbial communities as important players on the degradation of PAHs, as well as the basis for the establishment of effective bioremediation strategies against various chemical pollutants in situ soil environments.

Insights into Fiber Degradation by *Ruminococcus albus* 7 Revealed by Growth in Continuous-Culture Chemostat Using Transcriptomic and Physiological Analyses

La Reau, Alex J.* (alareau@wisc.edu)[1]; Christopherson, Melissa R.[1]; Dawson, John A.[3]; Stevenson, David M.[2]; Weimer, Paul J.[1, 2]; Kendziorski, Christina[3]; Suen, Garret[1]

[1]Department of Bacteriology, University of Wisconsin—Madison, Madison, WI 53706, USA. [2]U.S. Dairy Forage Research Center, U.S. Department of Agriculture-Agricultural Research Services, (USDA-ARS), Madison, WI 53706, USA. [3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI 53706, USA.

Ruminococcus is an important genus of bacteria present in many mammalian gastrointestinal tracts. Some species are among the most abundant cellulolytic members of the bovine rumen microbial community, where they actively convert plant biomass to nutritive short-chain fatty acids. For example, Ruminococcus albus 7 degrades cellulose and ferments its constituent sugars predominantly into acetate which the host can utilize. Conversely, in pure culture conditions, it switches to higher production of ethanol relative to acetate and is one of a few organisms to do so at mesophilic temperatures. However, despite decades of research, the mechanisms by which R. albus 7 degrades cellulose are not well understood. It is now known that R. albus 7 does not appear to use a canonical cellulosome, and instead seems to utilize a combination of unique carbohydrate-binding modules, a glycocalyx and

pilin-like proteins. Here we use a combination of physiological and transcriptomic analyses to gain insights into the cellulolytic and fermentative capabilities of R. albus 7. We show that R. albus 7 is able to degrade and ferment an array of fibrous substrates into ethanol and acetate in vitro. Additionally, we provide transcriptomic evidence that at identical dilution rates in steady-state chemostat conditions, R. albus 7 grown on cellulose versus cellobiose appears to upregulate production of cellulases, in particular those containing CBM37 domains. Moreover, it also seems to upregulate the expression of the tryptophan biosynthetic operon while Pil loci expression is unchanged. However, comparisons of gene expression at a range of dilution rates show very different trends, revealing that growth substrate and rate are important variables in the expression of cellulolytic machinery. Our work helps to revise our understanding of the fiber degradation strategy for this critically important member of the rumen and provides key targets for future exploration into enzyme characterization.

Assessing the Environmental Effect of Biofiltration on Greenhouse Gas Emissions from Wastewater

Lai, Ellen* (elai@ucdavis.edu); Zhao, Yongjing; Hess, Matthias; Mitloehner, Frank M.

Department of Animal Science, University of California, Davis.

Industrial, municipal, and agricultural wastewater contain elevated levels of nitrogen (N), which has the potential to pollute groundwater and the atmosphere. To mitigate nitrogen loading in wastewater, a biofiltration system has been developed and built at food processing plants, semi-rural communities, and most recently, on a commercial dairy in Central California, on which the present study focuses. The biofilter uses woodchips inoculated with earthworms and microbes to remove solids and contaminants from wastewater. Although chemical analysis shows that the biofilter decreases nitrogen loading in wastewater, it is unclear whether the biofilter is doing so at the expense of increasing nitrous oxide (N_2O) emission. N_2O, a byproduct of incomplete denitrification, is a potent greenhouse gas with a global warming potential (GWP) 298 times greater than the GWP of carbon dioxide (IPCC 2014). Thus, the present study examines the whole system, comparing the nitrogenous emissions and microbial communities in addition to wastewater chemical data before and after filtration. The resulting emission profiles suggest that the biofilter reduces ammonia (NH3) emissions from lagoon water by 90.2% without increasing N_2O emissions. Phylogenomic 16S rRNA analysis, which is currently underway, will provide a biological basis to explain the observed emission profiles of the filtration system. The biofilter's ability to reduce NH3¬ emissions might render biofiltration a potential cost-effective environmental solution for wastewater not only at dairies, but also other industries and communities. An in-depth understanding of the biofilter's biological and chemical processes will allow us to fine-tune this technology, enhancing our ability to reduce ammonia and greenhouse gas emissions from wastewater.

Citations:
IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Multiplexing Strategies for Microbial Whole Genome SMRT® Sequencing

Lambert, Christine* (clambert@pacb.com); Harting, John; Lee, Walter; Baybayan, Primo

PacBio, Menlo Park, CA

The increased throughput of the RS II and Sequel Systems enables multiple microbes to be sequenced on a single SMRT® cell. This multiplexing can be readily achieved by simply incorporating a unique barcode for each microbe into the SMRTbell™ adapters after shearing genomic DNA using a streamlined library construction process. Incorporating a barcode without the requirement for PCR amplification prevents the loss of epigenetic information (e.g., methylation signatures), and the generation of chimeric sequences, while the modified protocol eliminates the need to build several individual SMRTbell libraries.

We multiplexed up to 8 unique strains of H. pylori. Each strain was sheared, and processed through adapter ligation in a single, addition only reaction. The barcoded strains were then pooled in equimolar quantities, and processed through the remainder of the library preparation and purification steps.

We demonstrate successful *de novo* microbial assembly and epigenetic analysis from all multiplexes (2 through 8-plex) using standard tools within SMRT Link Analysis using data generated from a single SMRTbell library, run on a single SMRT cell. This process facilitates the sequencing of multiple microbial genomes in a single day, greatly increasing throughput and reducing costs per genome assembly.

Community Science Project: Systems Biology Approach to Fracking for Environmental Monitoring

Lamendella, Regina* (lamendella@juniata.edu)[1, 2]; Wright, Justin R.[1, 2]; Campa, Maria[3]; Hazen, Terry C.[3, 4]; McLimans, Christopher[1]; Ulrich, Nikea[1]

[1]Juniata College, Huntingdon, PA. [2]Wright Labs LLC, Huntingdon, PA. [3]University of Tennessee, Knoxville, TN. [4]Oak Ridge National Laboratory, Oak Ridge, TN.

Similar to most industrial activities, unconventional natural gas extraction (fracking) can lead to contamination and impact environmental health. Our Community Science Project uses high-throughput sequencing to elucidate the microbial biodegradation pathways and adaptive strategies harnessed by microbial communities in response to potential fracking impacts. Characterizing the microbial consortia associated with fracturing fluids can lead to improvements in treatment and reuse of these fluids. Further, knowledge of the biodegradative capabilities of the microorganisms associated with hydrofrack operations and potential recipients (e.g., aquifers, surface waters, and streambed sediments) will enable predictions about the fate and longevity of compounds of environmental concern. Here we compare metagenomics and metatranscriptomics profiles of microbial communities associated with waste fluids, and aquatic ecosystems in central and western Pennsylvania that have been impacted by hydrofrack operations, including sites with documented spills. In addition, non-impacted, geographically-matched sites were chosen as baseline samples for comparison. Comparative metagenomics and metatranscriptomics of source samples, and impacted and non-impacted sites were performed in JGI's

Integrated Microbial Genomes pipeline. Preliminary results are indicating an enrichment in several KEGG pathways involved in xenobiotic degradation within source samples and watersheds with documented fluid releases. Altogether this Community Science Project will provide a clearer understanding of the natural microbial community's response to hydrofracking and will better inform in situ remediation strategies, as well as the persistence of microbial markers related to biodegradation.

Genome-Wide Sequence-Indexed Collection of *Brachypodium* Mutants

Laudencia-Chingcuanco, Debbie* (debbie.laudencia@ars.usda.gov)[1]; Sibout, Richard[2]; Granier, Fabienne[2]; Schackwitz, Wendy[3]; Martin, Joel[3]; Cartwright, Amy[3]; Konishi, Colin[3]; Vogel, John[3]

[1]USDA-ARS WRRC, Crop Improvement and Genetics Research Unit, Albany, CA, USA. [2]INRA-AgroParisTech, Versailles, France. [3]US Department of Energy Joint Genome Institute, Walnut Creek, CA.

Brachypodium distachyon has emerged as a powerful model system to address fundamental questions in grass biology. Gaining insights into the function of more than 30,000 genes identified from its sequenced genome will further expand its utility for basic research. To this end, we are establishing a sequence-indexed library of mutations generated by chemical and radiation mutagenesis. Three Brachypodium mutagenized populations were produced using ethyl methanesulfonate, sodium azide and fast-neutron radiation. With the decreasing cost of DNA sequencing, we utilized whole genome sequencing to identify the genomic variations introduced by these mutagens. To date, we have sequenced 368 mutants. From the first 250 mutants we identified >200,000 high confidence mutations. The highest mutation rate, 1,200 SNP mutations per line, was observed in mutants treated with 10mM sodium azide. We will present as summary of the mutations identified to date and a comparison of mutagens and doses. In addition, we will present a collaborative framework of over 40 labs that are utilizing this resource for both forward and reverse genetic applications. Please contact us if you would like to join this effort.

Towards Raman-Activated Microbial Cell Sorting (RACS) in an Automated Microfluidic Platform

Lee, Kang Soo* (ks_lee@mit.edu)[1, 2]; Palatinszky, Márton[3]; Nguyen, Jen[1, 2]; Fernandez, Vicente I.[1, 2]; Menolascina, Filippo[1, 4]; Wagner, Michael[3]; Stocker, Roman[1, 2]

[1]Ralph M. Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Department of Civil, Environmental and Geomatic Engineering, Institute for Environmental Engineering, ETH Zurich, 8093 Zurich, Switzerland. [3]Division of Microbial Ecology, Department of Microbiology and Ecosystem Science University of Vienna, Wien, Austria. [4]SynthSys – Synthetic and Systems Biology, Institute for Bioengineering, University of Edinburgh, Scotland.

Raman microspectroscopy is a non-destructive technique for chemical measurements that has long been used on solids, liquids, and gases. In recent years, Raman approaches have been increasingly used to measure the chemical composition of microbial cells, largely because there is no need for

pretreatment (e.g., fluorescence labeling), since water does not interfere with the acquisition of Raman spectra, and because Raman spectra show sharp fingerprints. Here we report on our work to couple Raman measurements with microfluidics, aimed at developing a Raman-activated microbial cell sorting platform. Our goal is the high-throughput sorting of microbial cells based on functional traits of interest, for subsequent single-cell genomic analysis. Heavy water (D2O), which is easily incorporated into the cellular lipids, is used as a marker of the active/functional microbes and this is detected in the Raman spectrum. A sample is injected into the microfluidic device and focused in three dimensions. When a single cell is captured by the optical tweezer in the flow stream, it is transported into the sample-free region and sorted depending on its D2O-labeling status. Different features of the spectrum of individual cells were used to (i) identify a single cell in the optical tweezer in real time and (ii) calculate whether the cell is D2O-labeled or not. In this presentation, we illustrate each of these steps, demonstrate how they were coupled into a fully automated microfluidic platform, and present results for three different strains that we tested. Our aim is to use the sorted cells for subsequent single-cell genomic analysis, ultimately to enable to directly connect phenotypic with genotypic characteristics of individual cells.

Genetic Determinants of Bacterial Adaptation to Plants

Levy Asaf* (alevy@lbl.gov)[1]; Yourstone, Scott[2]; Clingenpeel, Scott[1]; Rangel Alvarez, Bryan[1]; Herrera Paredes, Sur[2]; Lundberg, Derek[4]; Lebeis, Sarah[2]; Pelletier, Dale[3]; Tringe, Susannah G.[1]; Dangl, Jeff[2]; Woyke, Tanja[1]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, California, 94598, USA. [2]Department of Biology and Howard Hughes Medical School, Bioinformatics and Computational Biology Program, Curriculum in Genetics, and Carolina Center for Genome Sciences, CB-3280, [3]University of North Carolina, Chapel Hill, North Carolina 27599, USA. [4]Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [5]Max Planck Institute for Developmental Biology in Tübingen, Germany.

Plants tightly associate with an array of diverse bacteria, with which they share complex and fascinating interactions, including parasitism, mutualism, and commensalism. These plant-associated (PA) bacteria have evolved a large set of genes that enable them to adapt to the plant environment. The functions of these genes are in many cases poorly studied and lack sufficient mechanistic understanding to facilitate microbiome engineering. Here, we employed a large-scale comparative genomics study to characterize PA genes, in particular root-associated (RA) genes. This was done by comparing 17 million genes from 3500 high quality and phylogenetically diverse bacterial genomes that were classified based on their isolation site. The PA genome set includes 300 newly sequenced bacteria isolated from the root environments of Poplar and Arabidopsis. Additionally, we deeply sampled the Arabidopsis root microbiome using various methods, including shotgun metagenomics and single cell genomics. PA and RA genes were found to be enriched also in the culture-independent samples. Several genomic features are common to most PA bacteria, such as the expansion of carbohydrate metabolism regulons. Many of the PA genes are encoded within putative gene operons, some of which encode for novel complex functions. Novel PA genes include a rapidly evolving virulence factor in the Acidovorax genus, a phage-like secretion system, and bacterial proteins with plant-like domains. The large PA and RA gene repository and the novel sequenced RA bacterial genomes will not only deepen our basic understanding of plant-microbe interactions, but will enable engineering of beneficial interactions for improved crop productivity and agricultural sustainability.

Use of KitBase to Facilitate Forward and Reverse Genetics Research in Rice

Li, Guotian* (gtli@ucdavis.edu)[1, 2]; Jain, Rashmi[1, 2]; Chern, Mawsheng[1, 2]; Pham, Nhan T.[1, 2]; Jones, Kyle[1, 2]; Martin, Joel[3]; Schackwitz, Wendy[3]; Lipzen, Anna[3]; Jiang, Liangrong[1, 2]; Barry, Kerrie[3]; Schmutz, Jeremy[3, 4]; Ronald, Pamela C.[1, 2]

[1]Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]Department of Plant Pathology and the Genome Center, University of California, Davis, CA 95616, USA. [3]US Department of Energy Joint Genome Institute, Walnut Creek, CA, 94598, USA. [4]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA.

To achieve cost efficient conversion of lignocellulosic biomass into biofuels, basic knowledge on the genes that control cell wall biosynthesis and modification is needed in grasses. To facilitate cell wall research, we have generated a mutant population using fast-neutron irradiation in the model rice cultivar Kitaake, a tractable model grass species. Kitaake is an early flowering, short-statured, short life cycle rice that is easy to transform, compared with other Japonica and Indica rice varieties. In collaboration with the Joint Genome Institute, we are sequencing 4,000 mutants. Genomic analysis of more than 1,000 mutants has been done, revealing that 13,469 genes are affected. Mutation types include single base substitutions, deletions, insertions, inversions, translocations, tandem duplications, and complex events. Single base substitutions predominate, but deletions affect the greatest number of genes, accounting for 57.2% of all affected genes. To make the genetic resource publicly available, we established KitBase, a comprehensive repository for rice mutant information. KitBase integrates JBrowse and BLAST to facilitate identification of mutations and searching of the database. KitBase includes genomic data, phenotypic data, and seed information for each of the mutant lines. Dozens of glycosyltransferase (GT) and glycohydrolase genes have been mutated in this mutant collection, and the function of multiple GTs in rice cell wall biosynthesis is being studied.

Characterization of the *Eucalyptus grandis* Terpenome Using Engineered Yeast Strains

Li, Rachel A.* (rachelli@berkeley.edu)[1, 2]; Yoshikuni, Yasuo[3]; Keasling, Jay D.[1, 2]

[1]UC Berkeley, Berkeley, CA. [2]Joint BioEnergy Institute, Lawrence Berkeley National Lab, Emeryville, CA. [3]JUS Department of Energy Joint Genome Institute, Walnut Creek, CA.

Sequenced in 2014, Eucalyptus grandis has more terpene synthases than any other plant studied to date. The essential oil of E. grandis is highly complex, making it difficult to purify and study the terpene constituents that may be useful as fuels, chemicals, and pharmaceuticals. To simplify the search for novel terpenes and to connect them with their corresponding synthases in E. grandis, we have engineered S. cerevisiae to be used as a chassis for terpene synthase expression. Each strain contains an engineered mevalonate pathway, including the expression of heterologous genes to circumvent the native feedback regulation at HMG1 and overexpression of the precursor prenyl diphosphate synthase. The remainder of the ergosterol pathway is downregulated by the introduction of a copper repressible promoter, diverting flux towards terpene production. In collaboration with the Joint Genome Institute, the 2014 E. grandis genome was scanned with stringent filters for terpene synthases, finding 73 terpene synthases and 25

of the triterpene-producing oxidosqualene cyclases. In total 98 enzymes were synthesized and are being expressed and characterized in the yeast strains engineered for high terpene flux.

Bacterial and Eukaryal Community Structures in Permanently Ice-Covered Antarctic Lakes

Li, Wei* (liw9@miamioh.edu)[1]; Kiss, Andor J.[2]; Dolhi, Jenna M.[1]; Teufel, Amber G.[1]; Morgan-Kiss, Rachael M.[1]

[1]Department of Microbiology, Miami University, Oxford, Ohio. [2]Center for Bioinformatics and Functional Genomics, Miami University, Oxford, Ohio.

The McMurdo Dry Valleys (MDVs) of Antarctica harbor numerous permanently ice-covered lakes which provide unique aquatic oases to support the food webs that are dominated by the microbial loop. The west and east lobes of Lake Bonney (WLB, ELB and respectively) and Lake Frxell (FRX) are three of several heavily studied lakes which are part of the McMurdo Long Term Ecological Research Program (MCM-LTER; www.mcmlter.org). The water columns of all three of our study lakes exhibit permanent physical and chemical stratification and minimal allochthonous nutrient inputs. Lake Bonney is oligotrophic and is generally oxic, while Lake Fryxell is mesotrophic with a deep anoxic zone. To examine the influence of lake physicochemistry on microbial community structure, we sequenced the SSU rRNA genes for bacteria and eukaryotes at multiple sampling depths in the water columns of our three study lakes. Preliminary analyses revealed that prokaryotic and eukaryotic populations are stratified in the water columns, and that microbial communities structure is strongly influenced by environmental parameters (i.e., temperature, salinity, nutrient, light and oxygen levels), despite that samples are clustering closer in WLB and WLB to FRX. Despite physical connectivity during the austral summer between the upper water columns of WLB and ELB, community structures are distinct in terms of distribution and abundance of major protists (e.g., photosynthetic haptophytes and heterotrophic chrysophytes) as well as specific bacterial groups in deep anoxic water.

Allele-Specific Transcriptome Analysis of *Populus* Hybrids

Liang-Jiao, Xue* (Liangjiao Xue); Harding, Scott A.; Tsai, Chung-Jui

Warnell School of Forestry and Natural Resources, Department of Genetics,University of Georgia, Athens, GA 30602.

Naturally-occurring Populus fremontii x Populus angustifolia hybrids exhibit variation in growth and defense traits with important ecological implications. However, molecular bases underlying segregation of these traits are not well understood. Here we report the transcriptome analysis at gene and allele levels in one F1 clone and one backcross (BC) clone of naturally-occurring Populus fremontii (Pf) and P. angustifolia (Pa) hybrids. Genomic resequencing data were obtained to call variants, which were then used to construct allele-specific (Pf and Pa) transcriptomes based on the P. trichocarpa reference genomes. RNA-seq data from mature leaves of both genotypes grown under normal or warm temperatures were processed to obtain maximum likelihood estimates of allelic expression. We identified more than 2200 and 2700 genes with allele-biased (AB) expression in the F1 and BC clones,

respectively. The BC leaves exhibit high levels of condensed tannins (CTs) characteristic of the Pa parent. Accordingly, transcript levels for the majority of CT biosynthetic pathway genes and their transcriptional activator MYB134 are much higher in the BC than the F1. MYB134 exhibits a Pa-biased expression in the F1 hybrid. In contrast, many CT pathway gene transcripts are Pf-biased, suggesting that the low rates of CT biosynthesis in the F1 clone may be epistasis, caused by genetic conflict. Besides flavonoid biosynthesis, GO categories associated with monosaccharide and starch metabolism are enriched among Pf-biased genes in F1, whereas amino acid metabolism, defense, and protein-DNA complex organization related functions are overrepresented among Pa-biased genes. Genes involved in photosynthesis are up-regulated by warm temperature in F1 but down-regulated in BC, consistent with different optimum temperatures for the two clones and their parental lineages. Interestingly, many photosynthesis-related genes that exhibit AB expression in F1 are also differentially expressed between F1 and BC, but the same is not observed for BC. Our analysis provides clues for the understanding of allelic interactions which may contribute to phenotypic diversity in Populus hybrid zones.

Transcriptomic Analysis for a Gorgonian Coral and Its Associated Organisms

Lin, Chan-Shing* (shinlin@faculty.nsysu.edu.tw); Lin, Ying-Rong

Department of Marine Biotechnology and Resources, National Sun Yat-sen Unversity, Gushan Disctrict, Kaohsiung City, 80424 TAIWAN.

The chemical constituents in the gorgonian corals such as Families Briareidae, Gorgoniidae and Plexauridae, including genus Junceella, Briareum, Rumphella, Ellisella, Menella, Echinomuricea, have many biological activities, e.g., cytotoxic, antiinflammatory, antiviral, insecticidal, antifouling, and immunomodulatory effects. Since there are a lot of associated organisms found in corals, the active compounds would be produced by their associations rather by corals themselves alone. We first compared ribosome-like sequences of a purple gorgonian coral with the ribosomal sequences from bacteria, fungi, chloroplasts, and mitochondria, respectively, to evaluate the associated organisms with a gorgonian coral. The transcriptomic data then were compared with the identified organisms to group out the data for each genus; which showed that symbiotic alga dominated at about 60% in our samples. The issues from the associated organisms also caused genome sequencing for the corals. With an example of Illumina genome sequencing, the sizes of the longest assembly scaffolds can be bigger than 150k; however, the annotations against viruses/phages and known databases of gene, refseq_RNA, or Uniprot were very diverse. We will bring up the discussions and collect suggestions during the poster presentation.

Genome Sequence of the Model Perennial Grass *Brachypodium sylvaticum* and the Identification of Transgene Combinations Conferring Tolerance to Multiple Abiotic Stresses

Liu, Lifeng* (lifeng.liu@lbl.gov)[1, 2]; Sade, Nir[3]; Gordon, Sean[1]; Cartwright, Amy[1]; Rubio-Wilhelmi, Maria del Mar[3]; Schmutz, Jeremy[1, 4]; Jenkins, Jerry[1, 4]; Tobias, Christian[2]; Thilmony, Roger[2]; Blumwald, Eduardo[3]; Vogel, John[1, 5]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598. [2]USDA-ARS Western Regional Research Center, Albany, CA 94706. [3]Department of Plant Sciences, UC Davis, Davis, CA 95616. [4]HudsonAlpha Instutite for Biotechnology Huntsville, AL, 35806. [5]Department of Plant & Microbial Biology, UC Berkeley, Berkeley, CA 94720.

Abiotic stresses such as drought, heat and salinity can have devastating economic and social impacts. Recent studies have linked the increased frequency of severe abiotic stress events with global warming, emphasizing the urgent need to develop crops with enhanced tolerance to abiotic stresses. Abiotic stress tolerance is particularly important for biomass crops because they must be grown under low-input conditions to maximize the net carbon ratio and on marginal land to minimize competition with food crops. Since many transgenic technologies proposed to improve biomass crops may negatively impact their perenniality, it would be extremely useful to have model perennial grass to rapidly test these technologies. To this end, we have turned to the perennial grass Brachypodium sylvaticum. B. sylvaticum has all the biological traits necessary to serve as a model perennial grass (e.g. diploid, small size, rapid generation time, self-fertile, small genome size, simple growth requirements). We have transformed B. sylvaticum with 20 genes that have been shown or predicted to enhance abiotic stress tolerance in other species. We are currently characterizing the stress tolerance of these lines and have identified two genes that when combined confer tolerance to simultaneous drought and salt stress. In order to further develop B. sylvaticum as a model system and enable transcriptomic analysis of our transgenic plants, we sequenced the B. sylvaticum genome using PacBio the sequencing platform. The current assembly contains 358 Mb in 1,118 contigs with an N50 contig length of 874Kb. We used a genetic map to arrange the contigs into chromosome scale scaffolds. Our initial assembly and annotation will be presented.

Identification of a *Ruminococcaceae* Species that Could Mediate O-Demethylation of Methyl Tert-Butyl Ether (MTBE) by Metagenomics Analysis of a Methanogenic Consortium

Liu, Tong* (tongliu@berkeley.edu)[1]; Yu, Ke[2]; Häggblom, Max M.[3]

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA. [2]School of Environment and Engery, Peking University Shenzhen Graduate School. [3]Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ.

Methyl tert-butyl ether (MTBE) was widely used as a gasoline oxygenate in U.S. until 2000. It caused major contamination in groundwater due to its offensive taste and odor as well as its toxicity. Also, the stable molecular structure of MTBE makes it persistent in groundwater and, therefore, affects the quality of drinking water. Anaerobic bioremediation of MTBE contaminated aquifer is a promising solution to address this issue. However, the essential microorganisms that mediated anaerobic biodegradation of MTBE are still unclear. The aim of this research is to shed light on the microorganisms actively degrading MTBE. Our previous study indicated that phylotypes belonging to the Ruminococcaceae in the Firmicutes were active in the methanogenic MTBE-degrading community. Here, we recovered the genome of this Ruminococcaceae species via metagenomics analysis. The results showed that the recovered genome phylogenetically matched with the Ruminococcaceae identified from stable isotope probing and clone library analysis. We further elucidated its metabolic pathways by KEGG database annotation. The almost completed Wood-Ljungdahl pathway strongly supports for the acetogenesis by the Ruminococcaceae, which is hypothesized to have the potential of cleavage of ether bond and O-demethylation of MTBE.

Metagenomics of Western Arctic Ocean Microbial Communities

Lovejoy, Connie* (connie.lovejoy@bio.ulaval.ca)[2, 3, 4]; Walsh, David[1]

[1]Concordia University Department of Biology, Montreal, QC, Canada. [2]Département de Biologie Université Laval Québec, QC Canada. [3]Institut de biologie intégrative et des systèms (IBIS), Université Laval. Québec QC Canada. [4]Unite Mixte Internationale Takuvik (UMI3376), Université Laval Québec, QC, Canada.

The Arctic Ocean (AO) is rapidly changing. The Arctic atmosphere is warming two to three-fold faster than the global rate. Climate change monitoring programs, such as the Joint Ocean Ice Studies (JOIS) lead by Fisheries and Oceans Canada (DFO), have documented decadal-scale changes in the AO. Surface waters are warming due to sea ice retreat and freshening due to increased sea ice melt water and river runoff. Warming and freshening is having profound effects on arctic marine ecosystems, including shifts in nutrient availability, primary productivity, and microbial community composition. Further changes in microbial distributions and activities are anticipated. However, only a few studies have described microbial diversity patterns in the AO and studies demonstrating temporal variability are exceptionally rare. As change continues, knowledge on the structure and function of Arctic marine microbial communities will become critical for assessing and predicting the consequences of a warmer, fresher AO. Overall, this projects aims to provide an Arctic Ocean metagenomic resource that can be used by us (and others) in studies on the genomic and functional diversity of marine microbes. In such studies, it is common practice to use publically available metagenomic data to test hypotheses on the biogeographical distribution of particular taxa and metabolic pathway, or to combine these two by exploring population and pangenome structure across environments. Compared to lower latitudes, there is far less metagenomic representation from high latitude seas, particularly the open Arctic Ocean. Hence the availability of a metagenomic dataset representative of the Arctic Ocean will provide important insights into the genomic and metabolic characteristics of Arctic microbes and fill an important void in metagenomic coverage of the global oceans.

Universal Expression Tools to Improve Nutrient Acquisition of Energy Crops

Mansoori Zangir, Nasim* (NMansooriZangir@lbl.gov)[1]; Hernandez-Garcia, Carlos[1]; Masson, Fabrice[1]; Shih, Patrick[1]; Vuu, Charlee[1]; Goyal, Garima[1]; Nath, Sangeeta[2]; Hillson, Nathan J.[1, 2]; Deutsch, Samuel[2]; Loqué, Dominique[1]

[1]Joint BioEnergy Institute, Physical Biosciences division, Lawrence Berkeley National Laboratory, 1 Cyclotron road, Berkeley, CA 94720, USA. [2]US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, B310-346A, Walnut Creek, CA 94598, USA.

Plant growth and development relies on roots as a means to anchor the plant body as well as to absorb water and nutrients such as N, P and K. As deficiency in these nutrients negatively impacts photosynthesis, plant growth and yield, chemical fertilizers have been used to compensate and meet the growing demand for plant material. Unfortunately the excessive use of fertilizers has come at high environmental and economic costs; and its production utilizes a substantial proportion of worldwide energy consumption. The current development of growing crops on marginal land (low water content, low nutrient supply, vulnerability to erosion and heavy-metal pollution) for bioenergy will reduce

competition with food crops and the pressure on high-quality arable lands utilization. As sufficient nutrient uptake by the plant root is vital especially when soil conditions restrict its availability, root systems in energy crops can be engineered to acquire/ accumulate the required nutrients.

Our focus is to design “universal” expression tools for plant root engineering functional across diverse plant species. Through targeted-cell engineering and synthetic biology we aim to engineer plant root systems to generate metabolic pathways and improve nutrient acquisition/accumulation in energy crops. This aim will be carried out through development and validation of a large set of root-specific and nutrient-responsive promoters. In house transcriptome analysis will help identify root specific genes that are constitutively expressed, induced or repressed through conserved metabolic responses to N, P and Fe starvation in hydroponic conditions across a widely diverse selection of plant species. Promoters from the identified genes will be isolated and characterized for spatiotemporal expression patterns and expression levels across different taxonomic classes using composite plants and conventional transgenic approaches. Subsequently monocot and dicot “universal” promoter libraries of the detected genes alongside those previously reported will be generated using a semi-high-throughput promoter cloning approach and yeast homologous recombination assembly method and ultimately used for tissue specific metabolic pathway engineering to improve nitrogen use efficiency in bioenergy crops.

We expect that the results of this research will generate a diversity of building blocks for plant engineering and will directly contribute to advance the DOE’s mission for the sustainable production of bioenergy.

http://www.jbei.org/research/divisions/feedstock/cell-wall-engineering/

NanoSIMS Isotope Imaging to Investigate Algal-Bacterial Interactions in Biofuel-Producing Communities

Mayali, Xavier* (mayali1@llnl.gov)[1]; Samo, Ty J.[1]; Nilson, Daniel J.; Siccardi, Anthony[2]; Lane, Todd[3]; Kimbrel, Jeffrey[1]; Pett-Ridge, Jennifer[1]; Weber, Peter[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA. [2]Texas A&M AgriLife, Corpus Christi, TX, USA. [3]Sandia National Laboratory, Livermore, CA, USA.

To better understand how bacteria promote the growth of microalgae, we are investigating algal-bacterial interactions in batch laboratory cultures of two algal species (Nannochloropsis salina CCMP 1776 and Phaeodactylum tricornutum CCMP 2561). Outdoor algal pond samples from Corpus Christi, Texas, were obtained and the free-living bacterial fraction (< 1 micron) were added to previously bacteria-free (axenic) cultures to create primary enrichments. These samples were further enriched for algal cell surface (phycosphere) associated bacteria by collecting the algal cells, washing them gently, and creating dilution cultures into secondary enrichments. A number of primary and secondary enrichments were investigated for the ability of attached bacteria to influence the cell-specific carbon fixation rates of the microalgal cells using NanoSIP. Cultures were incubated for 18 hours with 13C bicarbonate to track inorganic C fixation and 15N-leucine to track bacterial production. Samples were analyzed with LLNL’s NanoSIMS 50 to quantify isotope incorporation by individual cells. Most primary bacterial enrichments did not exhibit any significant effects on microalgal cell specific C fixation, and attached bacteria generally exhibited faster growth than free-living bacteria. Most secondary enrichments exhibited the

opposite patterns: algal cells with attached bacteria showed increased cell specific C fixation rates, and free-living bacteria showed increased growth rates compared to phycosphere-attached bacteria. We further showed that in some cases, attached bacteria incorporated more algal-derived organic C than free-living cells, suggesting that cell attachment can be critical for C cycling in these ecosystems. Current and future efforts include the isolation of single phycosphere-associated bacterial species with probiotic effects, and whole genome sequencing to eventually carry out metabolic modeling experiments in order to understand the genetic basis for the probiotic effects.

This research was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039, and a JGI CSP (#1939). Work at LLNL was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344

Expanding the Wisconsin Diversity Panel to Improve GWAS of Biofuel Traits in Maize

Mazaheri, Mona* (mmazaheri@wisc.edu)[1, 2]; Vaillancourt, Brieanne[3, 4]; Gage, Joe[1]; de Leon, Natalia[1, 2]; Mayer, Manfred[1, 5]; Barry, Kerrie[6]; Buell, Robin C.[3, 4]; Kaeppler, Shawn M.[1, 2]

[1]Department of Agronomy, University of Wisconsin, Madison, Wisconsin 53706. [2]Department of Energy, Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, Wisconsin 53706. [3]Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824. [4]Department of Energy, Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan 48824. [5]Plant Breeding, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany, 85354. [6]US Department of Energy, Joint Genome Institute, Walnut Creek, California 94598.

Maize stover is the largest source of crop residue in the U.S. and, therefore, a valuable source of biomass in biofuel production. The high level of natural genetic variation available among maize inbreds can be exploited to detect genomic regions associated with enhanced biomass traits such as late flowering and increased plant height and stalk diameter. The objective of this study was to expand an existing panel of 596 diverse maize inbreds to enhance our ability to detect novel genes associated with biomass traits in genome wide association studies (GWAS). A total of 334 inbreds including elite, exotic, and breeding lines were selected and added to the Wisconsin Diversity population. The updated set (WiDiv 2.0) consists of 837 inbred lines adapted to the upper Midwest region of the United States. The WiDiv 2.0 population was genotyped with 430,947 RNA-Seq based single nucleotide polymorphism (SNP) markers. The utility of the diversity panel was tested by GWAS of a highly heritable trait (maize cob color). A significant association (P = 10-38) was detected within a 13 kb region of the cob color candidate gene (p1 locus). The WiDiv population combined with the genomic data provided in this study is a foundation for detecting genomic regions underlying biomass and biofuel related traits in maize.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

Studying Coordination of Microbial Community Behavior via Context-Based Mutual Information Networks

McClure, Ryan S.* (Ryan.Mcclure@pnnl.gov); Overall, Christopher C.; Hill, Eric A.; Bernstein, Hans; McDermott, Jason E.; Beliaev, Alex S.

Pacific Northwest National Laboratory, Richland, WA.

Association networks built from parallel transcriptomic datasets can answer fundamental questions regarding gene control and species interactions in microbial systems. Here, we combine several RNA-seq datasets examining a model consortium comprised of a cyanobacterial phototroph (Thermosynechococcus elongatus BP1) and a heterotrophic partner (Meiothermus ruber strain A). During co-cultivation, M. ruber is completely dependent upon T. elongatus for sources of organic carbon and nitrogen, which leads to development of specific metabolic coupling and interactions between the two organisms. To understand and identify these interactions at the systems-level we utilized RNA-seq in conjunction with the Context Likelihood of Relatedness (CLR) program to build a network of genes (nodes) from both species and identify instances of co-expression (edges) both within and between species of this consortium. Resulting networks were then parsed to identify possible points of interaction and the genes and processes that show the greatest number of links between species. Networks comprised of each species individually showed structures similar to what has been previously observed with genes involved in central functions including photosynthesis, oxidative phosphorylation, cell motility, nitrogen and amino acid metabolism, carbohydrate metabolism and translation grouped separately into tight modules of highly co-expressed genes. When examining both species networks became much larger (~10x more edges with only ~2x more genes). When focusing on edges between species, M. ruber genes with the highest number of interspecies links include those involved in nitrogen metabolism, amino acid metabolism and cell adhesion. T. elongatus genes with multiple interspecies links include those involved with photosynthesis, carbon concentration mechanisms, nitrogen metabolism and terpenoid metabolism. When examining links between processes as a whole energy metabolism, transport and ribosomal proteins were strongly linked. Through a view of species interactions at the network level these studies identify putative points of interaction in phototroph-heterotroph communities. These include both central and expected processes (carbon, nitrogen and energy metabolism) as well as additional areas of interaction (adhesion, terpenoid metabolism). Both the developed methodology and conclusions derived from this work, are widely applicable to microbial communities for identification of interactions between species and characterization of community functioning as a whole. Ongoing efforts are focused on characterizing interaction in multi-organism consortia to determine how network structure changes in response to environmental perturbations and changes in community complexity.

Genomic Resources and 3D Phenotyping to Accelerate Sorghum Improvement

McCormick, Ryan F.* (ryanabashbash@tamu.edu)[1, 2]; Truong, Sandra K.[1, 2]; Jenkins, Jerry[3, 4]; Morishige, Daryl T.[2]; Sreedasyam, Avinash[4]; Shu, Shengqiang[3]; Weers, Brock D.[2]; McKinley, Brian A.[2]; Mattison, Ashley J.[1, 2]; Kennedy, Megan[3]; Amirebrahimi, Mojgan[3]; Grimwood, Jane[3, 4]; Schmutz, Jeremy[3, 4]; Mullet, John E.[1, 2]

[1]Interdisciplinary Program in Genetics, Texas A&M University, College Station, TX. [2]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX. [3]US Department of Energy Joint Genome Institute, Walnut Creek, CA. [4]HudsonAlpha Institute for Biotechnology, Huntsville, AL.

The development of genomic and phenomic resources will accelerate the genetic improvement of the food, forage, and bioenergy crop plant Sorghum bicolor. In collaboration with the DOE Joint Genome Institute and the HudsonAlpha Institute for Biotechnology, a suite of improved genomic resources for sorghum were generated, including a more accurate genome assembly (Sbi3), transcriptome annotation (v3.1), and atlas of transcript expression across multiple sorghum tissues and developmental stages. Genomic resources combined with high throughput, image-based phenotyping platforms are able to rapidly identify genetic bases of complex traits useful for plant performance. We developed a portable phenotyping platform based on a depth camera to semi-automatically generate and automatically measure 3D reconstructions of sorghum plants. This platform identified quantitative trait loci (QTL) underlying phenotypic variation in shoot architecture across multiple developmental timepoints, including QTL for plant height and leaf length. These results show that sorghum Dwarf-3 (Dw3) is associated with a pleiotropic effect on leaf angle at least 7 days prior to an effect on plant height. Continued development of these and additional genomic and phenomic resources for sorghum will facilitate more rapid advances in plant productivity.

Microbiome of Anaerobic Digesters at 51 Municipal Wastewater Reclamation Plants

Mei Ran* (ranmei2@illinois.edu)[1]; Narihiro, Takashi[1, 2]; Nobu, Masaru K.[1]; Kuroda, Kyohei[1, 3]; Muñoz, Julian S.[4]; Lee Patrick K.H.[5]; Lee, Po-Heng[6]; van Lier, Jules[4]; Kamagata, Yoichi[2]; McInerney, Michael J.[7]; Liu, Wen-Tso[1]

[1]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [2]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibarak, Japan. [3]Department of Environmental Systems Engineering, Nagaoka University of Technology, Kami-tomioka, Niigata, Japan. [4]Faculty of Civil Engineering and Geosciences, Department of Water Management, Section Sanitary Engineering, Delft University of Technology, Delft, the Netherlands. [5]School of Energy and Environment, City University of Hong Kong, Kowloon, Hong Kong. [6]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. [7]Department of Botany and Microbiology, University of Oklahoma, Norman, OK, USA.

Anaerobic digestion (AD) is an essential waste(water) treatment technology that not only biodegrades waste but also produces methane as a sustainable energy source. Microbes involved in this process include fermenters, syntrophs and methanogens as three key major guilds, as well as many uncharacterized prokaryotes (Bacteria and Archaea) and eukaryotes (protozoa). While this process has been widely implemented, the metabolic capacity and diversity of the major guilds and other organisms remain to be fully uncovered. In this study, 148 anaerobic digester sludge samples were taken from 51 municipal wastewater treatment plants in North America, Japan, Hong Kong, and Netherland. The genomic DNA extracted from individual sludge samples were used for analyses of 16S rRNA gene through Illumina Miseq sequencing. A total of 7.04 million effective sequences were obtained, which generated 263 thousand operational taxonomic units (OUTs at 97% sequence similarity) including 190 thousand singletons and 30 thousand doubletons. Clustering analyses based on weighted UniFrac distance matrix and UPGMA

algorithm demonstrated that operation temperature (mesophilic and thermophilic) was the top contributor to divide AD microbiome into major clusters, and within each cluster treatment plant was the primary grouping unit for digester microbial community rather than geological location (e.g., country). Analyses of core communities of mesophilic and thermophilic digesters identified signature microorganisms that were unique for different temperature conditions and shared populations that were mainly associated with feed sludge. In addition, we identified more than 20 uncharacterized phyla and classes (e.g., EM3, Hyd24-12, OP8), and will present the detail information in the poster. In conclusion, this global community profiling enabled us to advance the current understanding of microbial phylogeny, physiology, and ecology in AD microbiome, and allowed us to select AD samples to better understand the key metabolic functions of uncharted populations through further metagenomic and metatranscriptomic analyses.

Microbial, Physical and Chemical Drivers of COS and ^{18}O-CO_2 Exchange in Soils

Meredith, Laura K.* (laurameredith@email.arizona.edu)[1, 2]; Singer, Esther[3]; Boye, Kristin[2]; Whelan, Mary[4, 5]; von Sperber, Christian[2]; Pang, Erin[2]; Wingate, Lisa[6]; Ogée, Lisa[6]; Berry, Joseph A.[5]; Welander, Paula V.[2]

[1]University of Arizona, Tucson, Arizona, USA. [2]Stanford University, Stanford, California, USA. [3]UUS Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [4]University of California, Merced, California, USA. [5]Carnegie Institution for Science, Stanford, California, USA. [6]INRA, Villenave d'Ornon, France.

Carbonyl sulfide (COS) and the oxygen isotope composition ($\delta^{18}O$) of CO_2 are potential tools for differentiating the contributions of photosynthesis and respiration to the balance of global carbon cycling. These processes are coupled at the leaf level via the enzyme carbonic anhydrase (CA), which hydrolyzes CO_2 in the first biochemical step of the photosynthetic pathway ($CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$) and its structural analogue COS ($COS + H_2O \rightarrow CO_2 + H_2S$). CA also accelerates the exchange of oxygen isotopes between CO_2 and H_2O leading to a distinct isotopic imprint [1]. The biogeochemical cycles of these tracers include significant, yet poorly characterized soil processes that challenge their utility for probing the carbon cycle. In soils, microbial CA also hydrolyze COS and accelerate O isotope exchange between CO_2 and soil water. To account for these soil processes, measurements are needed to identify the key microbial, chemical, and physical factors.

In this study, we survey COS and $\delta^{18}O$ exchange in twenty different soils spanning a variety of biomes and soil properties. By comparing COS fluxes and $\delta^{18}O$-CO_2 values emitted from moist soils we investigate whether the same types of CA catalyze these two processes. These data are compared with soil physical (bulk density, volumetric water content, texture), chemical (pH, elemental analysis, sulfate, sulfur K-edge XANES), and microbial measurements (biomass and phylogeny). Furthermore, we will determine the abundance and diversity of CA-encoding genes to directly link CA with measured soil function. As samples are being prepared for submission, we analyzed existing JGI integrated microbial genomes data sets to reveal the distribution of the 6 known CA classes in different environments. Ultimately, this work will define the best predictors for COS fluxes and $\delta^{18}O$-CO_2 values from our suite of biogeochemical measurements. The suitability of identified predictor variables can be tested in follow-up studies and applied for modeling purposes.

Construction and Optimization of Lignin Degradation in Escherichia Coli

Michener, Joshua K.* (michenerjk@ornl.gov)[1, 2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN. [2]Bioenergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN.

Lignin is produced as a byproduct of cellulosic biofuel production and would improve the economics of biofuel production if it could be converted into high-value chemicals. However, the heterogeneity of lignin prevents many potential applications. Microbes are capable of metabolizing many of the diverse compounds produced during lignin degradation, channeling these compounds into a limited set of intermediates. However, these lignin-degrading enzymes are poorly understood. We are reconstructing these pathways in Escherichia coli to aid in characterization of potential lignin-degrading enzymes. Leveraging the capabilities of the Joint Genome Institute, we are pursuing an iterative process of gene synthesis to introduce new pathways, experimental evolution to optimize the heterologous pathways, and genome resequencing to identify mutations that improve pathway function. In our first example, we have constructed strains that are capable of growing with a model lignin compound, protocatechuate, at rates approaching the growth rate with glucose. These strains will be a useful foundation to construct more complicated lignin degradation pathways. In addition, by tackling a complex pathway with challenging biochemistry, we are learning important lessons for future metabolic engineering projects.

How Does a Mat Function? Organismal, Metabolic, and Geochemical Microgradients in an Epsomitic Microbial Mat Ecosystem

Mobberley, Jennifer* (jennifer.mobberley@pnnl.gov)[1]; Lindemann, Stephen[1]; Bernstein, Hans[1]; Moran, James[1]; Renslow, Ryan[1]; Babauta, Jerome[2]; Beyenal, Haluk[2]; Nelson, William C.[1]; Jansson, Janet K.[1]

[1]Pacific Northwest National Laboratory, Richland, WA, USA. [2]Washington State University, Pullman, WA.

Functional and species-level diversification with microbial communities is driven by environmental gradients. A model system to study this relationship, across compact sub-millimeter spatial scales, is the phototrophic microbial mats from the epsomitic Hot Lake (Oroville, WA) that develop at the sediment-water interface. Initial microbial diversity studies suggested that the Hot Lake mat hosts hundreds to thousands of taxa, however little is known about the spatial distribution and metabolic diversity relative to in situ environmental gradients. In this study, we linked metabolic potential to environmental and metabolic gradients by measuring depth-resolved metagenomic data, microsensor measurements of dissolved O_2, net rates of oxygenic photosynthesis, scalar irradiance, and primary productivity as determined by 13C-enrichment from labeled bicarbonate uptake. Additionally, fine scale depth-resolved 16S rRNA gene amplicon libraries spanning the diel cycle were paired with the above analyses to assess shifts in microbial diversity. In order to examine genes involved in biogeochemical cycling we generated a *de novo* metagenomic coassembly of five horizontally-sectioned microbial mat layers to obtain a spatial distribution of organisms and metabolic potential. We obtained 41 bacterial genomes with greater than 50 percent estimated genome completeness as well as gene-resolved unbinned contigs. These data showed differential distribution of the three dominant cyanobacterial populations through the mat depth profile suggesting phototrophic niche differentiation. The microsensor profiles and fixed carbon accumulation patterns suggested that light is a major driver of biogeochemical cycling within the top 3 mm of the mat

through energy and carbon acquisition from oxygenic photosynthesis. While cyanobacterial photosynthesis was the main conduit for energy and nutrient acquisition in the upper photic layers of the mat, the molecular data indicated that the deep mat contained a diverse phototrophic and mixotrophic community. Carbon acquisition pathways below 3 mm revealed an increase in heterotrophic bacteria, as well as a shift in carbon fixation ability from Cyanobacteria to members of Proteobacteria and Bacteroidetes. As Hot Lake water contained low concentrations of nitrate and ammonia and high concentrations of sulfate, there was a steeper distribution gradient of assimilatory nitrate reduction cycle genes compared to a more even distribution of genes in assimilatory sulfate reduction cycle. Detected anaerobic respiratory pathways included dissimilatory nitrate reduction to ammonia, denitrification, and dissimilatory sulfate reduction. The metabolic diversity of energy and carbon, nitrogen, and sulfur cycles in the deeper portions of the mat, corresponding to decreased oxygen concentration and light penetration, suggests that the organisms are exploiting available chemical and energy gradients through metabolic interactions. Genome reconstruction from environmental metagenomic samples provides increased resolution of the genetic content and specific metabolic pathways of community members, enabling more detailed predictions of interactions and metabolic exchange within an ecosystem. In this study, correlating species-resolved metagenomics with geochemical and metabolic gradient measurements helped us understand the interactions that are important to community structure and nutrient and energy flow within an epsomitic microbial mat community.

N^6-Methyladenine Marks Active Gene Expression in Early Diverged Fungi

Mondo, Stephen, J.* (sjmondo@lbl.gov)[1]; Dannebaum, Richard O.[1]; Ramamoorthy, Govindarajan K.[1]; Gryganskyi, Andrii[2]; James, Timothy[3]; O'Malley, Michelle A.[4]; Stajich, Jason[5]; Grigoriev, Igor V.[1]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, CA. [2]Duke University, Durham, NC. [3]Michigan State, East Lansing, MI. [4]University of California, Santa Barbara, Santa Barbara, CA. [5]University of California, Riverside, Riverside, CA.

DNA base pair modifications play essential roles in gene regulation and genome maintenance across all domains of life. DNA methylation, in its most commonly studied form 5-methylcytosine (5mC), has generally been understood as a repressive epigenetic mark. Interestingly, a new form of DNA base modification, N^6-Methyldeoxyadenosine (6mA), was recently discovered in eukaryotes and thought to be correlated with active gene expression. Here, we surveyed 15 lineages distributed across kingdom Fungi for presence of 6mA. We found that 6mA was very abundant in almost all early-diverged lineages surveyed, while in the Dikarya levels were fairly low. On average, ≈1.68% of adenines were methylated in early-diverged fungi, with the highest amount being found in *Hesseltinella vesiculosa* (Mucoromycotina), at 2.8%. This is 7x larger than the highest level reported previously (*Chlamydomonas reinhardii* - 0.4%), making it the highest level of 6mA found in any eukaryote to date. We discovered a conserved VATB (IUPAC nomenclature) motif associated with 6mA across all lower fungi, while the Dikarya showed no consistent pattern. Furthermore, >90% of 6mA marks in early-diverged fungi group together into dense clusters, or 'islands', the majority of which overlapped promoter sequence. Specifically, these islands frequently originated near the transcriptional start site (TSS) and tapered off around 500 bp into the gene body. Like *Chlamydomonas*, 6mA island presence appeared to have a positive impact on gene expression, but there was no correlation with the actual level of expression per se. Lastly, we identified that 6mA islands are more frequently found at promoters of highly conserved genes and we have

identified several pfam domains which may contribute to regulation of 6mA in lower fungi. We speculate that changes in 6mA utilization represents a major transition in the evolution of fungi, and that in early-diverged fungi 6mA islands mark genes for active transcription.

Genomes OnLine Database (GOLD): a Curated Catalogue of Genome and Metagenome Sequencing and Analysis Projects

Mukherjee, Supratim* (supratimmukherjee@lbl.gov); Reddy, T.B.K.; Stamatis, Dimitrios; Bertsch, Jonathon; Ovchinnikova, Galina; Isbandi, Michelle; Verezemska, Olena; Kyrpides, Nikos

US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA

The Genomes Online Database (GOLD) is a manually curated online resource that catalogs sequencing projects from around the world. The current version of GOLD implements a four level classification system in the form of a Study, Biosample (Organism), Sequencing Project and Analysis Project. GOLD provides an up-to-date status of complete and ongoing sequencing projects along with associated metadata. As of March 2016, GOLD contains information for nearly 24000 studies, 80000 sequencing projects and biosamples and 65000 analysis projects. The GOLD web interface facilitates submission of a diverse range of sequencing projects (such as isolate genome, single-cell genome, metagenome, metatranscriptome) and complex analysis projects (such as genome from metagenome, combined assembly from multiple sequencing projects). GOLD provides a seamless interface with the Integrated Microbial Genomes (IMG) family of analysis tools and also supports and promotes the Genomic Standards Consortium (GSC) Minimum Information standards. GOLD is available as an intuitive web interface for users to access genome reports and launch customized search tools.

URL: https://gold.jgi.doe.gov

Decoding DOC Degradation with Metatranscriptomics: How Is Microbial Metabolism Altered When Carbon Source Has Been Exposed to Sunlight?

Nalven, Sarah G.* (sarahnalven@gmail.com)[1]; Crump, Byron C.[1]; Cory, Rose M.[2]; Ward, Collin P.[2]; Kling, George W.[2]

[1]Oregon State University, Corvallis, OR. [2]University of Michigan, Ann Arbor, MI.

Arctic soils are warming, making vast stores of organic carbon available for conversion to CO_2. However, the processes that convert this carbon into CO_2 are not well understood. Once soil carbon enters surface water as dissolved organic carbon (DOC), it is transformed and mineralized by the combined activities of microbes and sunlight; but little is known about the biological or photochemical reactions that degrade DOC in surface waters, or how these reactions interact. Therefore we conducted an experiment to investigate the chemical and biological mechanisms of DOC degradation by comparing microbial metabolism of DOC that has been exposed to natural

sunlight versus kept in the dark. We leached DOC from the organic soil layer of moist acidic tundra in triplicate, exposed DOC to sunlight or kept it in the dark, added a common inoculum to both treatments, and compared metatranscriptomes of microbial communities after 4 h incubations. We found that the taxonomic composition of active microbial communities (those expressing genes) was similar between treatments, but expression patterns were significantly different (MANOVA, p = 0.001). This difference was driven primarily by the higher relative expression of genes coding for transcriptional and translational machinery in the light treatments and genes involved in metabolism in the dark treatments. Elevated expression of transcriptional and translational machinery in light treatments likely indicates that cells were preparing to synthesize new proteins in order to adapt to a new, light-exposed carbon source. In contrast, microbes in dark treatments did not have to adapt to a new carbon source, and were therefore able to maintain higher levels of metabolism than microbes in the light treatment. Expression of genes involved in cell motility and signal transduction was also higher in the dark, while expression of genes involved in structure and protein secretion was higher in the light. Several individual genes within these and other pathways also differed between light and dark treatments, including genes for transporters and the degradation of aromatic compounds, demonstrating a shift in the metabolic function of microbial communities following photo-degradation of soil DOC in arctic surface waters.

Manifestations of Microdiversity in a Model Microbial Community

Nelson, William C.* (william.nelson@pnnl.gov); Wiley, Steven H.; Fachko, Devin; Maezato, Yukari; Mobberley, Jennifer M.; Romine, Margaret F.; Jansson, Janet K.

Pacific Northwest National Laboratory, Richland, WA, USA.

Culture-independent high-throughput sequencing technology has enabled the discovery of unexpected phylogenetic diversity among bacteria and archaea in the environment. Although current evidence suggests that 16S rRNA-based amplicon analysis might overestimate phylogenetic diversity, metagenomic data indicate functional diversity within species is likely to be underestimated. The co-existence of organisms that are closely related (and thus may be conflated within an OTU definition) but exhibit functional differences caused by varying gene content, allele content, and/or regulatory response is referred to as microdiversity.

To investigate the prevalence of microdiversity within microbial communities, we took advantage of two recently described 20-member phototrophic model microbial communities with nearly identical community composition (UCC-A and UCC-O). We have generated complete or near-complete genomic information for all member organisms using a combination of genome reconstruction from metagenomic sequence and sequencing of cultivated isolates. We found that, even within these relatively simple communities, microdiversity is present. Evidence for this includes: 1) assembly and segregation of two Halomonas species where only one was predicted to be present from amplicon analysis, 2) anomalous assembly efficiency for high-coverage genomes in the metagenome, 3) identification of conserved patterns of single nucleotide polymorphisms (SNPs) present at consistent frequency within the metagenome, and 4) aberrant coverage profiles for certain genomic regions.

Analysis of the gene content of the two Halomonas species, HL-48 and HL-93, demonstrates that large-scale functional difference can be hidden within an identified OTU. We found 445 genes unique to HL-48 and 895 genes unique to HL-93, and demonstrated that HL-93 can, as predicted, grow on

glucuronate, while HL-48 cannot. Intraspecies microdiversity, detected by conserved SNPs, was observed for Halomonas HL-48, Rhizobiales HL-109, and Rhodobacteraceae bacterium Bin18. Evaluation of SNPs positions showed distribution within intergenic regions consistent with a random distribution, suggesting that functional differentiation between subpopulations is not primarily driven by changes in regulatory regions. Functional analysis of genes containing SNPs indicated a high fraction of affected genes are involved in transport or have no identified function. Metatranscriptome data identified SNPs in expressed genes and differences in the pattern of gene expression between subpopulations. Data from the Rhizobiales HL-109 genome suggests a third mechanism for generating microdiversity is gain/loss of genomic regions, possibly via the activity of mobile elements.

Our data indicate that even within relatively simple communities, the level of functional diversity is greater than suggested by taxonomic surveys. Some of the differences in functional potential that we identified could help closely related organisms avoid competition. Microdiversity might benefit a community through providing redundancy for key metabolic functions across organisms with varying optimal growth conditions. Our system provides an excellent opportunity to explore these hypotheses and examine functional redundancy, division of labor, and cooperative metabolic activity in microbial communities.

Isolation of Novel Mammal-Associated *Fibrobacter* spp.

Neumann, Anthony P.* (apneumann2@wisc.edu); McCormick, Caroline A.; Suen, Garret

University of Wisconsin-Madison, Madison, WI.

Bacteria in the genus *Fibrobacter* have attracted interest due to their superior ability to degrade crystalline cellulose and ferment the resulting sugars to succinate, one of the Department of Energy's top value added chemicals from biomass. Since their original isolation from the bovine rumen by Robert Hungate in the mid-twentieth century, molecular evidence for the presence of *Fibrobacter* in the gastrointestinal tracts of diverse herbivorous mammals has been observed. However, despite being conserved and widely distributed among mammalian herbivores, this genus remains poorly understood due in part to a lack of representative axenic cultures. The purpose of our research is to gain insights into the phylogenetic diversity and ecological niche of *Fibrobacter* spp. in diverse mammalian herbivores. We hypothesize that distinct lineages within the genus *Fibrobacter* have evolved to efficiently colonize their respective gastrointestinal environments, whether that is the foregut of ruminants or the hindgut of odd-toed ungulates, but that they have retained their ecological role as cellulose-degrading specialists in these ecosystems. To test this hypothesis we have developed a novel approach for recovering axenic cultures of *Fibrobacter* from gastrointestinal contents or feces. Using this approach we have successfully isolated 35 distinct strains of *Fibrobacter* representative of populations in nine mammalian hosts as well as previously described and novel diversity within the genus *Fibrobacter*. One notable observation is the recovery of previously uncharacterized strains representative of populations in the hindgut of odd-toed ungulates, i.e. horses, rhinos, and tapirs. With these isolates we now have a foundation to begin addressing our goal through the next phase of this project, which will involve whole-genome sequencing and phenotypic characterization of these novel *Fibrobacter* cultures.

Improved High-Quality Draft Genome Sequences for All *Carnobacterium* spp. Type Strains

Nicholson, Wayne L.* (WLN@ufl.edu)[1]; Davis, Christina L.[1]; Shapiro, Nicole[2]; Huntemann, Marcel[2]; Clum, Alicia[2]; Reddy, T.B.K.[2]; Pillay, Manoj[3]; Markowitz, Victor[3]; Varghese, Neha[2]; Pati, Amrita[2]; Ivanova, Natalia[2]; Kyrpides, Nikos[2, 4]; Woyke, Tanja[2]

[1]Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Merritt Island, FL 32953 USA. [2]US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. [3]Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [4]Dept. of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia.

The genus Carnobacterium was proposed in 1987 to encompass a group of closely related bacteria originally classified as unusual species of *Lactobacillus*. The genus *Carnobacterium* includes heterofermentative, facultatively anaerobic, psychrotolerant, either motile or non-motile, Gram-positive rod-shaped lactic acid bacteria that produce mostly L-lactic acid by fermentation from glucose. At present the genus contains 11 species with validly published names, which can be roughly divided into two groups. As the genus name implies, most *Carnobacterium* species (*C. divergens*, *C. gallinarum, C. inhibens, C. jeotgali, C. maltaromaticum, C. mobile, C. viridans*) belong to a group that were originally isolated from biological sources such as living fish or foods derived from animal sources. A second group of Carnobacterium spp. has been isolated from cold, low-nutrient environments such as Antarctic ice lakes (C. funditum, C. alterfunditum, C. iners) or Arctic permafrost (C. pleistocenium). We present here the improved high-quality draft sequences of the type strains for all 11 Carnobacterium species, as well as two new isolates obtained from Siberian permafrost, C. inhibens subsp. gilichinskyi strain WN1359 and C. viridans strain WN1374. This work was sponsored by DOE-JGI project number CSP1165.

Diverse Syntrophic Catabolism for Escaping Thermodynamic Limitations

Nobu, Masaru K.* (masarunobu@gmail.com)[1]; Narihiro, Takashi[2]; Liu, Miaomiao[1]; Kuroda, Kyohei[1]; Mei, Ran[1]; Liu, Wen-Tso[1]

[1]University of Illinois, Urbana-Champaign, Urbana, IL, USA. [2]The National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan.

Microbial respiration is a ubiquitous and diverse phenomenon that drives organic compound degradation, even in the most thermodynamically extreme environments where many bacteria resort to using H+ as an electron acceptor, producing H_2. Hydrogenogenesis rapidly becomes endergonic even with the slightest accumulation of H_2, so many H2-generating bacteria intimately cooperate with H2-oxidizing partners – a fascinating mutualism termed syntrophy. Specialized heterotrophic bacteria (syntrophs) and H2-oxidizing Archaea that reduce CO_2 to CH_4 (methanogens) form this mutualistic interaction to degrade compounds ranging from fatty acids to aromatic compounds. Syntrophs typically degrade such compounds to acetate, H_2, and formate, which methanogens mineralize to CH_4 and CO_2. However, a few studies discover unprecedented deviant pathways, such as production of butyrate from propionate degradation and cyclohexanecarboxylate from benzoate. Only a handful of syntrophs have had genome sequencing, let alone thorough biochemical characterization, leaving the diversity, ecology,

and biochemical rationale of such unique metabolic pathways largely unexplored. We may be drastically underestimating the biochemical capabilities and strategies of heterotrophic bacteria thriving at the thermodynamic limit.

To investigate the diverse catabolism of syntrophic aromatic acid degraders, we select inocula from methanogenic communities (U1 and B) hosting various syntrophs related to those known to degrade BA (Syntrophus) and additional aromatic acids including TA (Pelotomaculum and Syntrophorhabdus) along with syntrophic partners – hydrogenotrophic and aceticlastic methanogens. In addition, we identify a novel Syntrophorhabdus-related clade "Candidatus Syntropharomatica." Amidst the close physical association of syntrophs and methanogens, the spatial heterogeneity in the U1 and B communities may create various microenvironments supporting diverse syntrophic niches. We characterize the catabolic potential, interactions, and pathways of such syntrophs by metagenomically recovering draft genomes and tracking the syntrophs' gene expression in response to amending individual aromatic substrates – BA, TA, and TM. Based on total gene expression, BA, TA and TM degradation is respectively predominated by Syntrophus, Syntrophorhabdus, and "Ca. Syntropharomatica" in community U1 and Syntrophus, Pelotomaculum, and Syntrophorhabdus in community B; however, the degradation of all substrates peculiarly involves multiple syntrophs. Besides competition, perhaps cooperation between syntrophs and niche differentiation using specialized metabolic pathways may also lead to parallel and seemingly redundant activity of syntrophic populations. Here, we implement omics approaches to show evidence for novel catabolic pathways that may enable syntrophs to circumvent thermodynamic restrictions and diversify into discrete ecological niches, transforming our current understanding of metabolism and ecology in methanogenic ecosystems.

Agricultural Nitrogen Management Affects Microbial Communities, Enzyme Activities and Functional Genes for Nitrogen Cycle Processes

Norton, Jeanette M.* (Jeanette.norton@usu.edu)[1]; Ouyang, Yang[1]; Stark, John M.[1]; Shapiro, Nicole[2]

[1]Utah State University, Logan, UT USA. [2]DOE Joint Genome Institute, Walnut Creek, CA USA.

Improved understanding of nitrogen (N) cycling in agroecosystems is essential for increasing N use efficiency and sustainable food production. Availability of N from organic sources and fertilizers is the result of the enzymatic processes that comprise N mineralization, immobilization and nitrification. These transformations between organic N and inorganic N form a central part of the internal soil nitrogen cycle. A multi-year experiment was conducted in Utah to examine N-source effects on nitrification and mineralization in agricultural systems. N-sources include low and high levels of ammonium sulfate fertilizer (100 and 200 kg N/ha) and steer manure composts. We used a combination of enzyme activity, real-time quantitative PCR for target genes, amplicon pyrosequencing and metagenomic sequencing to examine functional changes in responsible organisms. Changes in bacterial communities, ammonia oxidizer communities, nitrite oxidizers, and ureolytic organisms in response to N treatments were found. Differential inhibition and quantitative PCR revealed that while ammonia-oxidizing archaea gene counts were higher, ammonia-oxidizing bacterial populations were more dynamic and responsible for an equal or greater fraction of the ammonium oxidized. The combination of enzyme kinetic and metagenomic approaches has brought us closer to the goal of linking the capable organisms to the process rate and extent in the soil environment.

Support from DOE-JGI Community Sequencing/Science Program 2010, 2014, Utah Agric. Expt. Sta. #UTA00371, 01260, NSF:0541797 and USDA Agriculture and Food Research Initiative Competitive Grants Program Grant no. 2011-67019-30178

The Interconnected Rhizosphere: Roots Stimulate High Microbial Network Complexity and Alter the Expression of Enzymes Related to Organic Matter Decomposition in Soil

Nuccio, Erin* (nuccio1@llnl.gov)[1]; Shi, Shengjing[2, 3]; Shi, Jason[3]; Karaoz, Ulas[4]; Boaro, Amy[5]; Nicora, Carrie[5]; Stuart, Rhona K.[1]; Weber, Peter[1]; He, Zhili[3]; Brodie, Eoin[4]; Lipton, Mary[5]; Zhou, Jizhong[3]; Firestone, Mary K.[2]; Pett-Ridge, Jennifer[1]

[1]Lawrence Livermore National Laboratory. [2]University of California, Berkeley. [3]University of Oklahoma. [4]Lawrence Berkeley National Laboratory. [5]Pacific Northwest National Laboratory.

The soil surrounding plant roots, the rhizosphere, has long been recognized as a zone of great functional importance to plants and the terrestrial ecosystems they inhabit. Plants play the central role in transferring atmospheric CO_2 to belowground soil C pools, while microbes are primary mediators of C transformation and mineralization in the soil. However, the molecular mechanisms underlying soil-plant-microbial interactions and their role in C cycling are poorly understood.

To identify potential microbe-microbe interactions within the rhizosphere over two growing seasons, we examined the co-occurrence patterns of bacteria in the rhizosphere soil of wild oat (Avena fatua) using random matrix theory (RMT). Our results revealed that bacterial networks in rhizosphere were substantially more complex than those in the surrounding bulk soils, and the complexity of the bacterial networks increased as the plants grew, even as univariate diversity decreased. Increased network complexity coupled to decreased diversity highlights that interactions are a crucial dimension of community organization overlooked by univariate diversity metrics. Within the rhizosphere networks, groups of highly connected modules formed over time, which most likely represent both interactions as well as microbial niches developing in response to root-induced changes of the soil environment. Consistent with the hypothesis that extensive mutualistic interactions occur among rhizosphere bacteria, covariations were predominantly positive (>80%); we identified quorum-based signaling as one potential interaction strategy. Highly connected taxa identified as putative keystone species (module hubs and connectors) often had low relative abundance in the rhizosphere (< 0.1%), which suggests that focusing on abundant taxa may overlook organisms that could play important roles in maintaining rhizosphere community structure and function. Network complexity appears to be a defining characteristic of the rhizosphere microbiome, and is a previously undescribed property of this habitat.

Rhizosphere microorganisms can also alter the breakdown of plant tissues and root litter, and in many cases have been shown to accelerate the decomposition of detrital plant biomass. However, the soil microbial communities and carbohydrate and lignolytic gene transcripts mediating the decomposition of root litter in soil are largely unidentified. We hypothesized that root exudates stimulate the expression of enzymes that are involved in decomposition of macromolecular C compounds. To assess how the abundance and diversity of decomposition enzymes differs in the rhizosphere relative to the surrounding bulk soil, we sequenced the metatranscriptomes of rhizosphere and bulk soil over time. Our results showed that a large

number of enzymes related to C decomposition more highly expressed in the rhizosphere compared to bulk soil. In addition, we are using genomic and proteomic approaches (including stable-isotope techniques) to identify key metabolic pathways responsible for C transformation and mineralization during root in-growth and root death/decay. By integrating stable isotopes as tracers of natural resource utilization (i.e. root litter), and analysis of the functional properties of the communities that respond to those resources, we can identify the microbial communities that are stimulated in the soil microbiome in response to root litter, living roots, and their intersection.

Genomes to Dynamic Decay Communities: Understanding Fungal Interactions During Early Decomposition Events in Natural Lignocellulosic Substrate

O'Leary, Jade M.* (olearyJM@Cardiff.ac.uk)[2]; Eastwood, Dan[1]; Boddy, Lynne[2]; Hiscox, Jen[2]; Müller, Carsten T.[2]; Orr, Galya[3]

[1]Swansea University, Swansea, United Kingdom. [2]Cardiff University, Cardiff, United Kingdom. [3]EMSL, Richland, WA/USA.

Understanding the microbial decomposition of lignocellulose and its contribution to global carbon cycling requires realistic model systems that go beyond in silico and single culture analyses. Fungal decay of lignocellulosic substrates evolved in competitive microbial environments. To truly understand the evolution and function of complex enzymatic decay systems, genetic and biochemical data are needed from controlled competitive interactions. Natural decomposition involves a succession of interacting microbes exhibiting varying degrees of competitive and decay ability. In this proposal we will target agaicomycete and ascomycete primary decay species which initiate natural decomposition and, by altering the woody substrate, influence the microbial decay communities that will follow. This multidisciplinary approach will involve genome sequencing underrepresented ecologically important species, and the detailed analysis and charactrization of interacting mycelia in 3 dimensional wood based microcosms using transcriptomic, proteomic, mass spectroscopy and fluorescence microscopy. The data generated by this study will give unparalleled insights into how microbial decay communities develop, interact and bring about substrate decomposition. Key processes and changes to substrate will be identified and models will be developed to consider the evolutionary pressures that produced the diverse chemistry and enzyme mechanisms associated with fungal wood decay.

Automating the Design for Synthesis and Assembly Process Using the Sequence Polishing Library (SPL)

Oberortner, Ernst* (eoberortner@lbl.gov)[1]; Cheng, Jan-Fang[1]; Hillson, Nathan J.[2]; Deutsch, Samuel[3]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, California. [2]DOE Joint BioEnergy Institute.

Over the last years the community has acquired knowledge regarding the feasibility of "writing" DNA. Nowadays, DNA synthesis vendors manufacture DNA in a fairly cost and time efficient manner. However,

ordering synthetic DNA from a DNA synthesis vendors can involve severe inefficiences. Specifically:
(1) Large and complex sequences cannot be synthesized with state-of-the-art technologies and, hence, are rejected by the DNA synthesis vendor.
(2) If not rejected, then the sequences are synthesized, but at higher costs and longer cycle times.

The DNA synthesis community has identified common sequence features that impact the success/failure rates of synthetic DNA. Such features include, for example, repeating sequences, sequences with low or high GC content, or sequences that contain restriction sites. In the scope of the DNA Synthesis Program, we developed the Sequence Polishing Library (SPL). SPL is a software library that consists of various software tools to higher the success rate of synthesizing DNA, to deliver the DNA to our users in shorter cycle times and to keep the synthesis costs as minimal as possible.

Here, we demonstrate the Sequence Polishing Library (SPL) and its integration into the DOE JGI's semi-automated "Design for Synthesis and Assembly" workflow. SPL enables the verification of DNA sequences against specified synthesis constraints, such as repeats or GC content. In case of violations, careful modifications can be suggested that alter the DNA sequence without changing its functionality, such as swapping codons in the DNA's coding regions. SPL provides expressive, comprehensive feedback about violations and the performed modifications. The SPL partitioning tool decomposes DNA sequences that exceed the maximum length of synthesis into synthesizable building blocks. The challenge here is to find appropriate overlapping sequences between properly sized building blocks in order to assemble the synthesized building blocks effectively. Thus, the partitioning tool enables the user to configure the characteristics of the overlaps, which are mostly determined by the utilized assembly protocol, such as length or melting temperature.

SPL and its tools can be invoked programmatically via an Application Programming Interface (API) and utilized manually via a web-based User Interface (UI). Both interfaces support commonly used data exchange formates, such as FASTA, GenBank, and Synthetic Biology Open Language (SBOL). Therefore, we believe that synthetic biology researchers as well as DNA synthesis vendors can benefit from utilizing SPL, contributing to the automation of internal and cross-organizational workflows.

Effects of Rumen Microbiome Composition on Methane Emission from Enteric Fermentation

Okatsu, Yuki* (yokatsu@ucdavis.edu); Dykier, Casey; Sainz, Robert D.; Hess, Matthias

Department of Animal Science, University of California, Davis, Davis, CA 95616 USA.

Methane (CH_4) accounts for 10% of total greenhouse gas emissions in the United States and increasing concentrations of atmospheric methane have been of particular concern since it has a global warming potential that is 28 times greater than that of than that of carbon dioxide(U.S. Environmental Protection Agency, 2016)

Enteric fermentation of ruminant livestock is the second largest source of anthropogenic methane accounting for ~28% of CH_4 that is generated by human related activities. Ruminants have a unique digestive system that contains a complex mixture of bacteria, archaea, protozoa and fungi that enable animals to digest and absorb energy from plant material. The ruminants' microbes play a crucial role in the overall digestion process and the conversion of plant material into metabolic intermediates or final

fermentation products that are further metabolized or released into the environment. Methane, which is one of the end products of rumen fermentation, is generated by the ruminants' archaea and is being released into the atmosphere by eradication (Carberry et al., 2014). In a recent study it was shown that cattle with higher feed efficiency produce less methane than cattle with lower feed efficiency (Fitzsimons et al., 2013). The molecular mechanisms and microbial interactions that cause these metabolic differences in the host animals are still poorly understood. An enhanced understanding of the rumen microbiome and how it affects the amount of methane and subsequently the animals' feed efficiency; will facilitate the identification and development of sustainable strategies for CH_4 migration from cattle, a significant part of terrestrial ecosystems.

Here I will outline my research project and provide initial results of my master's thesis through which I am studying the microbiome and metabolic rates of cattle that have low and high feed efficiency. I anticipate the results of my thesis will represent the first step toward on understanding of how microbe-host interactions change feed efficiency and consequently the amount of methane that is being generated by an animal.

References

Carberry CA, Kenny DA, Kelly AK, Waters SM, 2014. Quantitative analysis of ruminal methanogenic microbial populations in beef cattle divergent in phenotypic residual feed intake (RFI) offered contrasting diets. J Anim Sci Biotechnol 5, 41.
Fitzsimons C, Kenny DA, Deighton MH, Fahey AG, Mcgee M, 2013. Methane emissions, body composition, and rumen fermentation traits of beef heifers differing in residual feed intake1. Journal of Animal Science 91.
U.S. Environmental Protection Agency, 2016. Draft U.S. Greenhouse Gas Inventory Report: 1990-2014. In. (2016.)

Return to the Sea: the Genome of the Seagrass *Zostera marina*

Olsen Jeanine* (j.l.olsen@rug.nl.)[1]; van de Peer, Yves[3]; and the 35 members of the *Zostera Marina* Consortium including JGI

[1]University of Groningen, NL. [2]Ghent University, BE.

The genome of Z. marina, the first marine angiosperm to be fully sequenced (online Nature 27 Jan 16, hard copy 18 Feb 16) reveals unique insights into the genomic losses and gains involved in achieving the structural and physiological adaptations required for its fully marine lifestyle, arguably the most extreme habitat shift ever accomplished by flowering plants. Seagrasses returned to the sea some 100 mya and Zostera underwent a whole genome duplication near the Cretaceous-Tertiary boundary thus opening new ecological opportunities.

In returning to the sea, seagrasses found a vast new habitat free of terrestrial competitors, insect pests and drought stress; but they also had to adapt to high salinity, develop flexible leaves able to take up nutrients, and perform photosynthesis in the spectrally reduced light of the submarine environment.

Here we present a number of highlights including modified cell walls that are more algal-like than plant-like, loss of all stomata and all of the genes, reduced plant defense genes and volatiles, and modifications for underwater pollination involving exineless pollen.

The genome resource will advance a wide range of functional ecological studies from adaptation of marine ecosystems under climate warming, to unraveling the mechanisms of osmoregulation under high salinities that may further inform our understanding of the evolution of salt tolerance.

Uncovering Earth's Virome

Paez-Espino, David* (adpaezespino@lbl.gov)[1]; Eloe-Fadrosh, Emiley[1]; Thomas, Alexander D.[3]; Huntemann, Marcel[1]; Pati, Amrita[1]; Rubin, Edward[1, 2]; Ivanova, Natalia[1]; Kyrpides, Nikos[1]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, CA. [2]Lawrence Berkeley National Laboratory, Berkeley CA. [3]UC Berkeley, Berkeley CA.

Viruses are the most abundant biological entities on Earth, but challenges in detecting, isolating, and classifying unknown viruses have prevented exhaustive surveys of the global virome. Here we analyzed over 5 Tb of metagenomic sequence data from 4,100 geographically diverse samples to assess the global distribution, phylogenetic diversity, and host specificity of viruses. We discovered over 125,000 partial viral genomes, including the 7 largest phages yet identified, and increased the number of known viral genes by 16-fold. Half of the predicted partial viral genomes were clustered into genetically distinct groups, most of which included genes unrelated to those in known viruses. Using CRISPR spacers and tRNA matches to link viral groups to microbial host(s), we doubled the number of microbial phyla known to be infected by viruses, and identified viruses that can infect organisms from different phyla. Distribution analysis of viral groups across diverse ecosystems revealed strong habitat specificity for the vast majority of viral groups, but also identified groups of cosmopolitan viruses. Our results highlight an extraordinary global diversity of viruses and provide detailed insight into viral habitat distribution and host-virus interactions.

Understanding the Genomic Basis of Syntrophic Relationships Between Rumen Anaerobes

Peng, Xuefeng* (xpeng@engineering.ucsb.edu)[1, 2]; Gilmore, Sean P.[1]; Henske, John K.[1]; Theodorou, Michael K.[3]; O'Malley, Michelle A.[1]

[1]Department of Chemical Engineering, University of California, Santa Barbara, CA, USA. [2]Department of Earth Science, University of California, Santa Barbara, CA, USA. [3]Animal Production, Welfare and Veterinary Sciences, Harper Adams University, Newport, Shropshire, UK.

Anaerobic microbial consortia in rumens of herbivore are highly effective in converting lignocellulose-rich biomass into sugars and fermentation products such as carbon dioxide and methane. They consist of bacteria, protozoans, fungi, and methanogens, whose interdependency remains elusive largely due to difficulties in cultivation. With the goal of understanding competitive and mutualistic interactions among different groups of microorganisms, we developed a culturing technique that allows us to isolate and

characterize co-cultures of fungi and methanogens. In particular, we seek to determine the specificity of the syntrophy between fungi and methanogens, and whether the composition of biomass substrates exerts selective pressure on anaerobic consortia.

Here we present preliminary results from a pilot investigation, and lay out the plan of our subsequent sampling campaigns and experimental designs. Fresh fecal pellets from ruminant herbivores and fresh rumen material from fistulated cows will serve as source material of anaerobic consortia. Enrichments of anaerobic consortia will be maintained via continuous batch culture. At each transfer, the headspace of culturing tubes will be analyzed with gas chromatography, with a focus on hydrogen, methane, and carbon dioxide. Dissolved fermentation products will be analyzed on high performance liquid chromatography. Genomic DNA will be extracted for iTag sequencing and metagenomics analysis to examine the evolution of microbial communities. RNA will be extracted to assess transcriptional response within aerobic consortia during enrichment.

In the long run, understanding the genomic basis of anaerobic consortia from rumens will provide a framework to build synthetic co-cultures that convert biomass into value-added products.

Using a Single Cell Type Model to Advance Our Understanding of the Impact of the Plant Epigenome on Gene Transcription in Response to Environmental Stresses

Pingault, Lise* (lise.pingault@ou.edu); Qiao, Zhenzhen; Ren, Yuqi; Libault, Marc

Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, 73019, USA.

The structural organization of eukaryotic genomes is dependent on various epigenomic marks including, among others, the methylation pattern of the genomic DNA (gDNA) and the pool of small non coding RNAs. These epigenomic marks also regulate transcription, gDNA replication, recombination and DNA repair. However, mostly due to the cellular complexity of the tissues collected, modifications of the epigenome in response to stresses and their impact on gene expression remain unclear. To better characterize the role of the epigenome in controlling gene expression, we are focusing our effort on one single plant cell type: the root hair cell. The main function of these single tubular root cells is to improve water and nutrient uptake by the roots. As a consequence, root hair cells are highly responsive to environmental stresses, and they are considered as a model to investigate in detail the molecular response of a single cell type to various biotic and abiotic stresses including those related with climate change.

Using soybean as a model and taking advantage of the recent updates of the annotation of its genome, we initiated the project by comparing the modifications of the root hair methylome and transcriptome in response to 11 environmental stresses including extreme temperatures and pHs, nutriments depravations, drought, atmospheric CO_2 and salinity elevation.

These transcriptomic and epigenomic datasets are currently analyzed to (1) draw the expression profiles of each genes in response to the 11 stress conditions, (2) identify new non coding transcripts, (3) study in details the sub- or neo-functionalization of soybean duplicated genes, (4) and identify differentially methylated regions and their impact on gene expression. The detailed results of our analysis will be

presented as well as the implications in term of relationship between gene space organization and genome methylation in the context plant response to different environmental stresses.

Identification of Novel Methane Mitigation Strategies Using an In-Vitro Rumen System (RUSITEC)

Polley, Tamsen* (tmpolley@ucdavis.edu); Hess, Matthias

Department of Animal Science at the University of California, Davis, Davis, California.

Methane (CH_4) is a greenhouse gas with a 28-fold greater global warming potential than carbon dioxide over a 100-year time horizon. Enteric fermentation, mainly from ruminant livestock, contributes to 17% of global methane and 3.3% of total greenhouse gas emissions. On average, a cow produces 485 L of methane per day. Different mitigation strategies have been proposed to reduce eructed CH_4 emitted from livestock. Several natural and synthetic compounds have been found to reduce microbial CH_4 production during the degradation of lignocellulosic plant material within in-vitro and in-vivo systems.

In the project presented here we are focusing on one synthetic and one naturally derived feed additive compounds, respectively: 3-nitrooxypropanol (NOP) and walnut hulls (*Juglans regia*). A recent study found that NOP, added to either in-vitro and in-vivo rumen systems, resulted in a significant decrease (~60% and 70% respectively) of CH_4 production. Walnut hulls are a novel feed additive with similar properties to almond hulls, which are commonly used as a dairy cow feed supplement and are a industry byproduct. Importantly, walnut hulls have high tannin content, which has been shown to inhibit methanogenesis during enteric fermentation. To broaden the repertoire of chemical and biological feed additives for methane mitigation, we are focusing on elucidating the methodology of NOP in the rumen system and its effect on the microbial population, and test the hypothesis that walnut hulls as feed additive will reduce the microbial rumen production of methane.

The tightly controlled rumen simulation technique (RUSITEC) will be utilized to study the multi-scalar response of the microbial community to CH_4 mitigation techniques. In a preliminary study we will generate 16S rRNA profiles and measure environmental data such as greenhouse gas and volatile fatty acid (VFA) emission, which will allow a whole-system assessment of methane production during enteric fermentation.

The understanding of microbial carbon cycling in the cow's rumen, which is a significant terrestrial source of CH_4, and the development of sustainable methods for greenhouse gas mitigation strategies will be advanced by this pilot study through the use of the highly controlled RUSITEC system.

How to Bin Thousands of High Quality Genomes from an Ecosystem with Moderate Complexity

Probst, Alexander J.* (alexander.j.probst@gmail.com)[1]; Sieber, Christian M.K.[1, 2]; Thomas, Brian C.[1]; Sharon, Itai[1]; Diamond, Spencer[1]; Jarett, Jessica[2]; del Rio, Tijana G.[2]; Tringe, Susannah G.[2]; Woyke, Tanja [2]; Banfield, Jillian F.[1]

[1]UC Berkeley, Department of Earth and Planetary Sciences, Berkeley, CA USA. [2]US Department of Energy Joint Genome Institute, Walnut Creek, CA USA.

Linking metabolic functions to organisms is one of the most important but also more challenging aspects of environmental genomics. Genome-resolved metagenomics has proven to be a valuable approach to recover hundreds of near-complete genomes from a biotope, enabling resolution of metabolic function at organism level. Many different algorithms exist to reconstruct bacterial and archaeal genomes from assembled genomic fragments. Here, we evaluated the power of six different binning algorithms to recover genomes from a deep-subsurface environment with moderate complexity and substantial genomic novelty. Additionally, we show that time-consuming, semi-automated binning tools that rely upon emergent self-organizing maps based on differential coverage or nucleotide composition improve results only marginally compared to those achieved using automated binning tools. We evaluated the quality of recovered genomic bins from 24 different samples using bacterial and archaeal single-copy gene sets. Our results provide evidence that the combination of six different binning methods substantially outperforms each single binning method. Overall, we reconstructed 2490 archaeal and bacterial genomes with a completeness of greater than 70%. These genomes represent a substantial fraction of organisms in this ecosystem and enable comprehensive metabolic prediction of microbial function. This research was supported by a JGI ETOP project.

Designer Small Molecule Biosensors for High-Throughput Phenotyping of Genomes and Genes

Raman, Srivatsan M.* (sraman4@wisc.edu)[1, 2, 3]

[1]Great Lakes Bioenergy Center, University of Wisconsin-Madison, Madison, WI 53706. [2]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706. [3]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706.

Genetically-encoded biosensors are emerging as a powerful tool for high-throughput, single cell measurement of intracellular small molecules. Changes in the concentration of specific small molecules can serve as a readout of enzymatic activity, function of operons, and effects of regulatory mutants on protein stoichiometry. In energy and environment applications, biosensors enable high-throughput screening of variants of cellulases for specific glycosidic cleavages, transporters and degradation enzymes of valorized lignin-based phenolics to reduce toxicity, biosynthesis of value-added products from lignocellulosic breakdown products, and for the mining of bioremediation pathways from environmental microbiomes. These biosensors are usually allosteric transcription factors (aTF), which activate transcription of a reporter gene when bound to a specific small molecule. The aTFs bind to a wide variety of small molecules such as sugars, phenolics, polycyclic aromatics, alkanes and other industrially useful molecules. However, only small molecules with a known naturally-occurring aTF partner can be sensed, thereby significantly impeding the broad applicability of biosensors for any arbitrary small molecule. Here, we describe a general approach to make designer biosensors for new small molecules by redesigning aTF specificity. We evaluate tens of thousands of computational protein design candidates in a high-throughput screen to engineer new biosensors. We redesigned the lac repressor to respond to four new molecules – gentiobiose, fucose, lactitol and sucralose – with activity and specificity comparable to wild-type lac repressor with its inducer IPTG. We are currently working on building a suite of biosensors for natural products like flavonoids and terpenoids, next-generation biofuels like isobutanol, environmental pollutants like chlorinated phenols and redox sensors (NADH/NAD+) to

dynamically regulate redox pathways. As an enabling technology, we expect our designer biosensor platform to be of broad interest to the DoE bioenergy community by facilitate strain engineering, bioconversion and environmental bioprospecting.

The Distribution, Diversity, and Function of Microbial and Viral Communities Inhabiting Anoxic Crustal Fluids Within Deep Subseafloor Basalt

Rappé, Michael S.* (rappe@hawaii.edu)[1]; Jungbluth, Sean P.[2]; Lin, Huei-Ting[3]; Nigro, Olivia D.[4]; Steward, Grieg F.[4]

[1]Hawaii Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Kaneohe, HI, USA. [2]University of Southern California, Center for Dark Energy Biosphere Investigations, Los Angeles, CA, USA. [3]Institute of Oceanography, National Taiwan University, Taipei, Taiwan. [4]Department of Oceanography, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, HI, USA.

Hydrothermally heated fluids circulate everywhere within the permeable basement rock of the upper ocean crust, creating an environment that is conducive to supporting a deep subsurface marine biosphere. While little is known regarding microbial life within our planet's rock-hosted deep subseafloor, boreholes drilled through deep ocean sediment and into the underlying basalt provide invaluable windows of access that have been used previously to document the presence of microorganisms within fluids percolating through the subseafloor crust. During expeditions from 2008 to 2014, we repeatedly sampled basalt-hosted, deep subseafloor crustal fluids from five different boreholes drilled along the Juan de Fuca Ridge flank in the Northeast Pacific Ocean using pumps and samplers capable of collecting whole water and filtered particulates in situ. The instrumented boreholes, sitting at a seafloor depth of 2600 m, penetrate ~100 to 260 m of bottom sediments and another ~48 to 300 m of igneous basement where they tap into hot (up to 65°C), anoxic fluid within Earth's largest aquifer. Recently, we analyzed small subunit ribosomal RNA genes amplified and sequenced from marine sediment, bottom seawater, and basalt-hosted deep subseafloor fluids that span multiple years and locations on the Juan de Fuca Ridge flank in order to quantitatively delineate a subseafloor microbiome comprised of distinct Bacteria and Archaea, including phylogenetically unique lineages of bacteria related to Nitrospirae, Aminicenantes, Calescamantes, and Chloroflexi and unique, uncultivated lineages of Archaea including marine benthic group E, the Terrestrial Hot Spring Crenarchaeotic Group, Bathyarchaeota, and distant relatives of cultivated, sulfate-reducing Archaeoglobi. We have also begun to characterize the metabolic potential and genomic characteristics of microbes and viruses residing in the subseafloor crust of the Juan de Fuca Ridge flank by sequencing both environmental DNA and the genomes of single sorted cells originating from crustal fluids of two recently drilled boreholes, U1362A and U1362B. Here we describe these current DNA sequencing efforts, and outline plans for future metagenome, metatranscriptome, and single cell genome sequencing through the Joint Genome Institute's 2016 Community Sequencing Program.

Comparative Genomics of Biotechnologically Important Yeasts

Riley, Robert* (rwriley@lbl.gov)[1]; Haridas, Sajeet[1]; Wolfe, Kenneth H.[2]; Lopes, Mariana[3, 4]; Hittinger, Chris, Todd[3, 5]; Göker, Markus[6]; Salamov, Asaf[1]; Wisecaver, Jen[7]; Long, Tanya M.[8]; Aerts, Andrea L.[1]; Barry, Kerrie[1]; Choi, Cindy[1]; Clum, Alicia[1]; Coughlan, Aisling Y.[2]; Deshpande, Shweta[1]; Douglass, Alexander P.[2]; Hanson, Sara J.[2]; Klenk, Hans-Peter[6, 9]; LaButti, Kurt[1]; Lapidus, Alla[1]; Lindquist, Erika[1]; Lipzen, Anna[1]; Meier-Kolthoff, Jan P.[6]; Ohm, Robin A.[1]; Otillar, Robert P.[1]; Pangilinan, Jasmyn[1]; Peng, Yi[1]; Rokas, Antonis[7]; Rosa, Carlos A.[4]; Scheuner, Carmen[6]; Sibirny, Andriy A.[10]; Slot, Jason C.[11]; Benjamin, Stielow J.[6, 12]; Sun, Hui[1]; Kurtzman, Cletus P.[13]; Blackwell, Meredith[1, 4, 15]; Grigoriev, Igor, V.[1]; Jeffries, Thomas W.[8]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598. [2]UCD Conway Institute, School of Medicine, University College Dublin, Dublin 4, Ireland. [3]Laboratory of Genetics, Genome Center of Wisconsin, Wisconsin Energy Institute. [4]J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI. [5]Departamento de Microbiologia, ICB, C.P. 486, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil. [6]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI. [7]Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, 38124 Braunschweig, Germany. [8]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA. [8]University of Wisconsin-Madison, Department of Bacteriology and USDA Forest Products Laboratory, Madison, WI, USA. [9]Newcastle University, School of Biology, Ridley Building, Newcastle upon Tyne, UK. [10]Department of Molecular Genetics and Biotechnology, Institute of Cell Biology, NAS of Ukraine, Lviv 79005 Ukraine and Department of Biotechnology and Microbiology, University of Rzeszow, 35-601 Poland. [11]Department of Plant Pathology, Ohio State University, Columbus, OH 43210. [12]CBS-KNAW Fungal Biodiversity Centre, Utrecht, the Netherlands. [13]US Department of Agriculture ARS NCAUR 1815 N University St. Peoria, IL 61604. [14]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803 USA. [15]Department of Biological Sciences, University of South Carolina, Columbia, SC, 29208 USA.

Ascomycete yeasts are metabolically diverse with great potential for biotechnology. Here we report the comparative genome analysis of 29 taxonomically and biotechnologically important yeasts, including 16 newly sequenced. We identify a genetic code change, CUG-Ala, in Pachysolen tannophilus which is sister to the known CUG-Ser clade. Our well-resolved yeast phylogeny shows that some traits such as methylotrophy are restricted to single clades, whereas others such as L-rhamnose utilization have patchy phylogenetic distributions. Many pathways of interest are encoded by gene clusters, with variable organization and distribution. Genomics can predict some biochemical traits precisely, but the genomic basis of others such as xylose utilization remains unresolved. Our data also provide insight into early evolution of ascomycetes. We document the loss of H3K9me2/3 heterochromatin, the origin of ascomycete mating-type switching, and pan-ascomycete synteny at the MAT locus. These data and analyses provide strategies for engineering efficient biosynthetic and degradative pathways, and gateways for genomic manipulation.

Metabolic and Regulatory Networks for Carbohydrate Utilization in Human Gut *Clostridia*

Rodionov, Dmitry* (rodionov@burnham.org)[1, 2]; Khoroshkin, Matvei[1]; Leyn, Semen[1]

[1]Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia.
[2]Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, CA, USA.

Human microbiota is the complex and dynamic community of commensal, symbiotic and pathogenic microorganisms that are present on and within the human body and has an enormous impact on humans. Commensal Clostridia consist of gram-positive bacteria in the phylum Firmicutes and make up a substantial part of the total bacteria in the gut microbiota. Carbohydrates comprise a key natural source of carbon and energy for a variety of heterotrophic microbes in gut microbiome. The chemical diversity of dietary and host-derived polysaccharides in gut ecosystem is matched by a multitude of species-to-species variations in sugar utilization networks – extracytoplasmic degradation of polysaccharides, uptake and biochemical transformations of oligo- and monosaccharides, and regulatory mechanisms involved in feeding of carbohydrates into central carbon metabolism. We use a comparative genomics-based approach to assess in silico the metabolic potential of key Clostridia members of the human gut microbiome reflecting their abilities to utilize various sources of carbon/energy and produce fermentation products such as short-chain fatty acids. This approach is based on functional gene annotation and prediction using three comparative genomics techniques: (i) homology-based methods; (ii) genome context analysis using the IMG and SEED resources; (iii) co-regulation by the same transcription factor analyzed using the RegPredict/RegPrecise resources.

We performed detailed reconstructions of metabolic pathways and regulons for utilization of >20 carbohydrates and polysaccharides, as well as the central glycolytic pathways and fermentation pathways for production of acetate, propionate and butyrate in the complete and draft genomes of Ruminococcus gnavus, Clostridium symbiosum, Faecalibacterium prausnitzii and several other Clostridia that define a program of normal development of the microbiota in infants. A particular emphasis of the reconstruction effort was on highly divergent and complex carbohydrate catabolic machinery including extracellular polymer-hydrolyzing, uptake and intracellular conversion/fermentation components. The reconstructed pathways include numerous novel carbohydrate-specific transporters, regulators and enzymes that are non-orthologous to the previously characterized proteins and were previously defined only at the level of general class. In R. gnavus, we predicted 28 novel regulators, most of which are from the AraC family, identified their DNA binding motifs and reconstructed the respective sugar catabolic regulons. In contrast, genes encoding key glycolytic enzymes, as well as pathways for production of short-chain fatty acids were predicted to be co-regulated by the global transcriptional regulator Rex responding to redox state of the cell.

In summary, the obtained in silico reconstructions of carbohydrate metabolism contribute to our understanding of metabolic crossfeeding processes in human gut microbiome. The key practical deliverable of these reconstruction efforts will be a phenotype matrix, aggregating numerous predicted phenotypes. These deliverables will yield numerous testable hypotheses that can be tested with focused analyses of additional –omics data. The refined reconstructions and phenotypes will be used to build a model of *in vivo* modulatory effects of gut microbiota in response to diets and novel prebiotics.

Combined –Omics Approaches Highlight Functional Diversity for Lignocellulose Breakdown Within a Single Genus of the Order Polyporales, Basidiomycetes

Rosso, Marie-Noelle* (marie-noelle.rosso@univ-amu.fr)[1, 2]; Miyauchi, Shingo[1, 2]; Drula, Elodie[1, 2]; Forquin-Gomez, Marie-Pierre[1, 3]; Piumi, Francois[1, 2]; Navarro, David[1, 2, 4]; Favel, Anne[1, 2, 4]; Ruiz-Duenas, Javier F.[5]; Grigoriev, Igor V.[6]; Ritley, Robert[6]; Lipzen, Anna[6]; Henrissat, Bernard[7, 8]; Berrin, Jean-Guy[1, 2]; Record, Eric[1, 2]

[1]INRA, UMR1163 Biodiversité et Biotechnologie Fongiques, F-13288 Marseille, France. [2]Aix Marseille Université, UMR1163 Biodiversité et Biotechnologie Fongiques, F-13288 Marseille, France. [3]Polytech'Marseille, UMR1163 Biodiversité et Biotechnologie Fongiques, F-13288 Marseille, France. [4]CIRM, UMR1163 Biodiversité et Biotechnologie Fongiques, F-13288 Marseille, France. [5]Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, E-28040 Madrid. [6]US Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [7]Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257 CNRS, Université Aix-Marseille, F-13288 Marseille, France. [8]INRA, USC 1408 AFMB, F-13288 Marseille, France.

Saprotrophic fungi play a critical role in the decomposition of dead organic matters and their conversion into fungal biomass, CO_2, and small molecules, such as organic acids. Among saprotrophs growing on lignocellulosic material, wood decay fungi produce and secrete a plethora of enzymes able to degrade the three main polymers in plant biomass : cellulose, hemicellulose and lignin by diverse hydrolytic and oxidative mechanisms. Those fungi are thereby a source of enzymes of interest to the modern human industries. The Basidiomycetes fungi from the genus *Pycnoporus*, order *Polyporales*, have been studied for their efficiency in numerous biotechnological applications related to their capacity to degrade lignin and to secrete laccases. Recently, the genome of *P. cinnabarinus strain* CIRM-BRFM137 was sequenced and 10,442 predicted genes were functionally annotated using a phylogenenomic procedure (Levasseur et al., BMC Genomics, 2014). The genome revealed a consistent repertoire of genes shared with wood-decaying basidiomycetes with the classical families involved in cellulose and hemicellulose degradation, while its pectinolytic repertoire was relatively limited. Considering lignin breakdown, a complete versatile enzymatic arsenal was identified including a set of laccases and class II peroxidases.

To deepen the analysis of the functional diversity within the genus *Pycnoporus*, we compared by a combined genomic, transcriptomic and secretomic approach three *Pycnoporus* strains that belong to three *Pycnoporus* species and originate from different geoclimatic regions. *Pycnoporus cinnabarinus* is widely distributed, especially in the Northern hemisphere. P. coccineus is found in countries bordering the Indian and Pacific Oceans and *P. sanguineus* is found in the tropics and subtropics of both hemispheres. All three species are white-rotters mainly found on hardwoods.

Our analyses show that the three genomes share highly conserved gene repertoires coding for lignin- and cellulose-active enzymes. However, their responses to lignocellulosic substrates of various composition are different and involve different enzymatic machineries for lignocellulose degradation. This study shows the importance of exploring functional diversity of wood decay fungi beyond gene repertoires.

Microbial Endophytes for Plant Biomass, Biostimulant for Feedstock Biofuel Production

Ruiz-Font, Angelica* (convenios22@hotmail.com)[1, 2]; Ramos-Robles, Isabel[2]; Villa-Ramirez, Sugeyrol[1]; Sanchez-Tozapantzi,Gabriel[2]

[1]Applied Biotechnology Reserach Center, National Polythecnic Institute, MEXICO. [2]Productive Biotechnology Graduate Program, National Polythecnic Institute, MEXICO.

This project seeks to develop a microbial consortium of plant growth promoting bacteria in order to ensure the survival of these plant species in arid lands and soils poor in nutrients.

The first source of bacterial diversity was restoration process site. We established two experimental units with 4 different nutrient treatments, We used 6 seedling pine species (*Pinus ayacahuite, P. cembroides*, *P. rudis, P. montezumae, P. maximartinensis, P. hartwegii*), one specie of agave (*Agave atrovirens*). We sampled bacterial diversity in treatments showed the best survival rate.

The second source of bacterial diversity was three collections of microorganism previously isolated from different pristine sources: pines forest, Chihuahua desert and agave's land.

Each strain was screened for a) solubilization of phosphates in NBRIP media b) chitinase activity c) pectinase and d) cellulases were evaluated. The production of indoleacetic acid (AIA) in the presence of tryptophan was performed by the colorimetric technique using Salkowski reagent.

At first source of microbial diversity 56 morphotypes were isolated, among which 26 were obtained with the capacity to solubilize phosphates, 5 strains with the ability to produce chitinase, two strains with the ability to degrade pectin (pectinase enzyme) and six strains with the ability to degrade cellulose (activity cellulase). Finally we could only select three strains had the ability to produce indole compounds. At second source we isolated 69 morphotypes, 39 from desert, 25 from agave land and 6 from pine forest.

Establishing a Method for High Precision Cell Dispensing with the Echo® 500 Series of Liquid Handlers

Ruppert, Noël* (nruppert@labcyte.com); Jarman, Carl; Lejeune, Aurore; Orren, Linda; Lesnick, John

Labcyte Inc., Sunnyvale, CA, USA.

Cell-based assays have become indispensable for both academia and industry-based research toward drug discovery. Cell-based assays are found in numerous applications from compound screening, target identification and validation, to monitoring cellular events, efficacy and biosafety. Basic research relies on numerous cell-based assays to elucidate basic mechanisms of cell proliferation, differentiation and function. Since live cells are used, more biologically relevant data are acquired, revealing complex cellular information associated with the compound and target in the cells.

Cell-based assays are clearly an important part of drug discovery, and successfully dispensing viable cells is a key step in the process. The cell dispense must be quick, accurate and gentle enough

to ensure cell viability. Challenges include establishing cell stability and uniformity to enable a high-quality start point for a cell-based assay. This requirement makes the cell seeding step a key component for all cell-based screening.

In this study, we demonstrate the utility of the Labcyte Echo liquid handler to accurately dispense PC-3 cells, an adherent line, from the Labcyte Echo Qualified Reservoir to 1536 well assay plates. The results are measured with the CellTiter-Glo® luminescent cell viability assay and read on the PHERAstar FS HTS microplate reader. The addition of HistodenzTM to the cell culture suspension prior to the dispense step provides cell stability and dispense uniformity. Our data indicate that the Echo liquid handler is capable of high-quality cell seeding providing uniform concentrations of viable cells.

www.labcyte.com

Reconstruction of Genome-Scale Metabolic Networks and Models from Arabidopsis Thaliana Root-Associated Bacterial Isolates as Probes to Examine Host-Microbe Interactions

González, Isai Salas* (isai@email.unc.edu)[1]; Finkel, Omri M.[1]; Castrillo, Gabriel[1]; Gao, Tianxiang[1]; Dangl, Jeff[1, 2]

[1]University of North Carolina, Chapel Hill. [2]HHMI-GBMF Plant Science.

Plants assemble a taxonomically limited root-associated microbiome. These plant associated bacterial communities promote a diverse number of beneficial phenotypes in the host, ranging from plant growth promotion to pathogen inhibition. Because plant roots exudate a vast array of organic and inorganic compounds into the rhizosphere, it has been suggested that the host may have a broad influence over the metabolic activity of their associated microbes, nevertheless the specific metabolic influence that the plant has over specific isolates has not been elucidated yet.

In order to explore the potential metabolic requirements of the plant root-associated bacteria and to determine the correlation of these requirements with specific plant exudation patterns, the metabolic networks for around ~200 *A. thaliana* root-associated bacterial isolates were reconstructed. Based on the exploration of the topological features of each bacterial network a set of compounds were identified as presumably taken from the environment. The relevance of these compounds is that they may be correlated with plant exudation profiles under different growth conditions and may have a major role explaining bacterial colonization and host phenotypic patterns previously observed.

Together, our results suggest that, for the ~200 metabolic networks reconstructed, there is a general trend for closely related bacteria to have similar metabolic requirements. However, there is still a significant degree of metabolic diversity among the same closely related isolates that may be correlated with differences in bacterial colonization and host phenotypic state previously reported in mono association experiments.

Finally it is important to address that unraveling the complex plant-microbe interactions is a multidisciplinary task , because of this it is necessary to cross the results of this study with more complex and sophisticated modeling techniques , such as Genome Scale Models (GEMs) and Flux Balance

Analysis. As well, it is fundamental to integrate into the model an array of different experimental data , such as bacterial colonization surveys, host transcriptomics and metabolomic profiling of root exudates.

Correlating Taxonomic Composition of Algal-Associated Bacterial Communities with Productivity in Biofuel-Producing Algal Species

Samo, Ty J.* (samo1@llnl.gov)[1]; Kimbrel, Jeffrey A.[1]; Nilson, Daniel J.[1]; Nuccio, Erin[1, 3]; Siccardi, Anthony; Lane, Todd[2]; Pett-Ridge, Jennifer[1]; Mayali, Xavier[1]

[1]Lawrence Livermore National Laboratory. [2]Sandia National Laboratory. [3]Texas A & M Agrilife.

Mutualistic algal-bacterial interactions arise when bacteria provide metabolically beneficial substances to the algae in exchange for energy-dense organic compounds. We seek to ultimately ecologically engineer these positive exchanges by establishing closely interacting algal-bacterial cultures, further simplifying the algal-attached bacterial communities through enrichments, and monitoring for elevated growth and biomass characteristics via chlorophyll fluorescence measurements in two model biofuel-producing microalgal strains, *Phaeodactylum tricornutum* (Pt) and *Nannochloropsis salina* (Ns). The DNA of bacterial communities attached to the algae versus the total community was differentially collected and the 16S rRNA gene partially sequenced and utilized for comparisons and correlations with algal productivity measurements. We confirmed bacterial attachment to algae (via phycospheres) using fluorescence microscopy, scanning electron microscopy, and NanoSIMS imaging. All of the Ns enrichments displayed elevated growth and/or yield. Meanwhile, ~75% of the Pt enrichments showed enhanced growth and/or yield. In all enrichments, the majority of bacteria belonged to the *Rhodobacteraceae* family, a group known for probiotic interactions with algae. At the genus level, there was an abundance of *Phaeobacter* regardless of algal host, while the Pt enrichments had an abundance of *Labrenzia, Loktanella and Hyphomonas*. The Ns enrichments were dominated by *Marivita*, *Marinobacter* and *Algoriphagus*. These abundant taxa, however, were not necessarily correlated with a probiotic effect. Interestingly, two *Loktenella* species appeared to have opposing effects depending on the host, with L. vestfoldensis being beneficial to Ns and detrimental to Pt, while *L. rosea* had the opposite effect. Taken together, the growth effects and bacterial consortia in the co-cultures suggest that maintenance of algal-bacterial mutualisms can assist with establishing economically feasible and sustainable algal biofuels.

Effects of Bacterial Inoculants on the Development and Molecular Response to Changes in Microbial Communities in *Nicotiana benthamiana* Plants

Sanchez Barrios, Andrea* (asanchez-barrios@uky.edu)[1]; Shearin, Tom[1]; Lundberg, Derek[3];. De Lorenzo, Laura[2]; Nair, Meera[1]; Debolt, Seth[1]

[1]Department of Horticulture and Plant Physiology/Biochemestry/Molecular Biology Program, University of Kentucky Lexington, Kentucky 40546, USA. [2]Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY, 40546-0312, USA. [3]Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen 72076, Germany.

Efforts to optimize crop productions through traditional breeding are slow, and the use of transgenic approaches is expensive and often blocked by negative public perception and fertilizers present environmental consequences. An additional plant-improvement possibility may be the application of bacterial treatments –used as amendments- before germination to improve plant development. This approach has the potential to sustainably improve productivity and modulate metabolism via non-transgenic, non-toxic approaches. In order to achieve this we treated *Nicotiana benthamiana* seeds with three different bacterial strains, which we called S41 (isolated from switchgrass), S413 and S343 (isolated from tomato plants), as well as mock treatments. An evaluation of bacterial communities present inside roots of treated and non-treated plant, as well as a comparison between physiological and molecular responses to these possible community changes was performed to assess the bacterial inoculants effects on the development and improvement of N. benthamiana. Thus, we examined 16S rRNA amplicons in inoculated-plants vs. controls, showing a change in the composition of the microbial communities in roots of the different treatments. In addition, we have compared the differential plant height (H), length (L), leaf width (LW), number of leaves (NL), and number of flowers (NF) in response to the application of these bacterial inoculants on the seed coat. S413 and S343 inoculants had beneficial effects on seedling development and vegetative growth of plants; in contrast, S41 had a retarded growth effect. Analysis of gene expression using a genome-scale approach (PAT-seq, Poly(A) Tag coupled to high-through-sequencing) revealed that genes involved in stress response were downregulated for S343 and S413 treatments and upregulated for s41, which correlates with some of the patterns of growth that they present and reporter lines were used to confirm the expression of some of them.

JBrowse and GrainGenes

Herin, Sayer Y.* (herin@ucdavis.edu)[1, 2]

[1]UC Davis, Albany, CA. [2]USDA-ARS, Albany, CA.

The growth of genomics data in the grasses links historic genetic maps and characterizations of traits to the genome of wheat, barley and oat. Comparative analyses of the genomes of these organisms is more informative with the computational comparison of the sequence of the genes. Researchers can take seriously goals of moving a trait from one species to the other. Underlying this interface between researchers and the genome is the genome browser. JBrowse takes the place of GBrowse with a clearer presentation of data and an enhanced distribution of labor and responsiveness. Graingenes' integration with the genome browser is improving the researchers ability to navigate an ever more complete genome view. As the genomes of these plants are finalized, the genomic tools on GrainGenes, especially JBrowse, will become the trusty companions of researchers dedicated to improving wheat, barley and oat.

https://t3sandbox.org/sayer/
https://t3sandbox.org/sayer/jbrowse1/?data=wheat
https://t3sandbox.org/sayer/JBrowse-1.11.6

Identification of Glycosyltransferases Involved in Biosynthesis of Arabino-Galactans Via RNA-Seq and Virus-Induced Gene Silencing

Scheller, Henrik V.* (hscheller@lbl.gov)[1]; Stonebloom, Solomon H.[1]; Ebert, Berit[1]; Barry, Kerrie[2]

[1]Joint Bioenergy Institute, Lawrence Berkeley National Laboratory. [2]US Department of Energy Joint Genome Institute, Walnut Creek, California.

Arabinogalactan proteins (AGPs) are a group of plant cell wall glycoproteins consisting of heavily arabino-galactosylated hydroxyproline-rich peptides, which are often anchored to the plasma membrane with GPI-anchors. The arabinogalactan polysaccharides attached to AGPs are structurally complex and diverse. AGPs have numerous biological functions in plant growth and development, e.g. as cell non-autonomous signaling molecules. Functional studies of AGPs have been hindered by a lack of understanding of AGP glycan biosynthesis. To identify AGP glycan biosynthetic enzymes we performed a transcriptome study of AGP gum secretion in Meryta sinclairii. Upon wounding M. sinclairii secretes copious amounts of gum consisting largely of AGPs, and hence we can expect that AGP biosynthetic genes are up-regulated in response to wounding. We identified 34 up-regulated glycosyltransferase-like genes as good candidates for glycan biosynthetic activities. The role of candidate AGP glycan biosynthetic genes was tested by transiently silencing the Nicotiana benthamiana orthologs of each gene using Virus Induced Gene Silencing and analyzing the cell wall composition of silenced tissues.

The Geneious Sequence Classifier Tool: Identifications of Quarantine Significant Plant Pathogenic Fungi as an Illustrated Use Case

Shearman, Helen* (helen@biomatters.com)[1]; Kennedy, Aaron[2]; Olsen, Christian[1]; Kearse, Matt[3]

[1]Biomatters, Inc. Newark, NJ. [2]USDA-APHIS-PPQ-National Identification Services, Beltsville, MD. [3]Biomatters, Ltd. Auckland, New Zealand.

The Sequence Classifier plugin for Geneious automates the taxonomic classification (i.e., identification) of an organism from DNA sequence data. In this poster, we describe the Sequence Classifier plugin and present a use case from the USDA-APHIS Plant Protection and Quarantine (PPQ).

The Sequence Classifier plugin taxonomically classifies an organism based on similarity between its DNA sequence(s) and orthologous sequences in a researcher's database of reference sequences using a BLAST-like algorithm. The user creates a folder within Geneious and populates it with their own reference sequences. The Classifier plugin creates global pairwise alignments between the query sequence and each sequence in the researcher's reference database. The Classifier plugin then evaluates each pairwise alignment and selects an appropriate taxon for identification based on user-specified similarity levels. Multiple genes may be used simultaneously for classification and multiple classifications may be run at once. Phylogenetic trees may also be produced to optimize identification results based on automated multiple sequence alignments and phylogenetic analyses that are integrated into the plugin and run according to user selected parameters.

Potential pests to US agriculture and the environment are intercepted from around the world at US ports of entry on imported plants and plant products and identified by the USDA-APHIS-PPQ. Identifications

are routinely based on morphology; however, in 2014, PPQ piloted DNA sequence-based identifications of fungal plant pathogens in the genus Colletotrichum. DNA sequences from multiple loci derived from type specimens and other well-identified individuals representing all described Colletotrichum species were gathered as a reference set and imported into Geneious as a user defined database. The Classifier plugin was used to automate the manual bioinformatics tasks necessary to conduct DNA sequence-based identifications, and perform the analyses necessary to identify unknowns against the user defined database. This study found that the Classifier plugin enabled a standardized identification procedure, lowered the level of expertise required by the end user, and dramatically reduced the time necessary to identify an organism. The Classifier plugin facilitated resolution of many morphological based identifications, the appreciation that many intercepted Colletotrichum fungi are undescribed species, and that many Colletotrichum species have much wider host and geographic ranges than previously known.

The ability to create an in-house reference database and control a wide range of analysis parameters and options makes the Geneious Sequence Classifier plugin a versatile and powerful tool that may be applied to a wide variety of taxa and tasks, including improving and expediting standard DNA Barcoding analysis or other identification efforts, facilitating biodiversity discovery, and creating diagnostic identification protocols that may be validated for use in a variety of environments.

A User Friendly CRISPR Design Tool

Shearman, Helen* (helen@biomatters.com)[1]; Olsen, Christian[1]; Moir, Richard[2]; Cheung, Matthew[2]; Kearse, Matt[2]; Dunn, Sebastian[2]; Kuhn, Jonas[2]; Markowitz, Sidney[2]; Cooper, Alex[2]; Smith, Jo[2]

[1]Biomatters, Inc. Newark, NJ. [2]Biomatters, Ltd. Auckland, New Zealand.

The CRISPR/Cas9 system, a pioneering genome-editing technique, allows for the manipulation of genes and enables researchers to examine the consequence of sequence modifications in a precise and repeatable manner. The CRISPR tool in Geneious R9 will identify potential CRISPR target sites, including paired nickase sites, in a selected sequence and assess the target sites based on off-target interactions. Identified CRISPR sites are annotated on the target sequence using innovative heatmap-style annotations.

Each CRISPR site in the selected sequence is scored according to how many off-target sites it potentially will bind to and how similar the off-target sites are to the original sequence. Scores are calculated according to the method developed by the Zhang lab at the Massachusetts Institute of Technology1. CRISPR sites are displayed as an annotation track on your original sequence and the annotations are colored according to the corresponding score to enable you to see the best sites at a glance.

CRISPR sites need to be targeted to the gene of interest and not cut elsewhere in the genome; however, many other design tools provide a small number of choices for the target genomes. In Geneious, the researcher creates their own off-target database by simply creating a folder within Geneious and placing sequences within this folder. Any sequence may be placed inside this folder, including reference genome sequences downloaded from external databases and/or proprietary sequences created in-house. This allows researchers to use genomes from their organism of interest or from closely related organisms. The researcher then specifies the folder in the CRISPR options.

As this tool is contained within Geneious, researchers can leverage all of the other functionality in Geneious, including creating and sharing workflows for CRISPR design standard operating procedures. This poster aims to demonstrate the functionality of the CRISPR tool, including the benefits of providing the researcher the control to create their own off-target database.

References

[1] Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013 Nov;8(11):2281-308. doi: 10.1038/nprot.2013.143. Epub 2013 Oct 24.

Gene-Centric and Genome-Centric Approaches to Understanding Freshwater Microbial Communities

Shrader, Donna Grace* (d.grace.shrader@gmail.com); Hamilton, Joshua; Linz, Alexandra; McMahon, Katherine D.

University of Wisconsin-Madison, Madison, WI.

Microorganisms play essential roles in nutrient cycling in freshwater ecosystems, and differential approaches to analyzing these communities can yield complementary insights. Our lab has used time series of shotgun metagenomes sampled from freshwater lakes to examine the activity of uncultured microorganisms with both gene-centric and genome-centric approaches. Gene-centric approaches include assessing the functional potential for nutrient cycling in lakes by analyzing changes in lake gene content over time. Genome-centric approaches include assessing the functional potential within single-cell genomes and genomes assembled from metagenomes by enabling the reconstruction of an organism's functional pathways and assessing pathway completeness. Whereas gene-centric approaches represent more inclusively the sequencing read data captured by shotgun metagenomes, genome-centric approaches enable deeper insights about the functional potential of individual organisms and their specific interactions. Both approaches can shine light on the response of microorganisms to ecological disturbances, such lake mixing or algal blooms. Alternative insights from both approaches combine to provide a more complete picture of the ecology of freshwater lakes.

Discovery and Transfer of Novel Pathways for Phosphate Solubilization

Shulse, Christine N.* (cnshulse@lbl.gov)[1]; Chovatia, Mansi[1]; Lei, Yan[1]; Deutschbauer, Adam[2]; Blow, Matthew J.[1]

[1]JUS Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.

High yield agricultural plant growth is currently dependent on costly and environmentally damaging phosphate fertilizers. One approach to alleviating this dependency is to develop bacterial strains that can convert existing phosphorus sources in the soil to soluble forms available for plant uptake. Past attempts at developing such strains have been hindered by incomplete knowledge of the genes required for phosphorus solubilization, and failure of bacterial strains to survive in the plant root environment. To address these challenges, we are using genome wide mutagenesis of phosphate solubilizing bacteria

to discover novel genes and pathways underlying solubilization of phosphorus sources, as well as bioinformatics approaches to identify homologs of known phosphate-solubilization genes from microbial genomes and environmental metagenomes in the Integrated Microbial Genomes database. We have used synthetic biology approaches to synthesize these novel genes and transfer them to the plant-associated bacterium Pseudomonas fluorescens WCS417r. This research has the potential to result in significantly improved understanding of the genetics of phosphorus solubilization; novel biological tools for studying the interactions between plants, microbes and nutrients in the environment; and a first step in the development of alternative approaches to sustainable phosphorus use in agriculture.

Improving Genome-Resolved Metagenomics by Combining Binning Predictions

Sieber, Christian M.K.* (csieber@lbl.gov)[1, 2]; Probst, Alexander J.[2]; Sharon, Itai[2]; Thomas, Brian C.[2]; Tringe, Susannah G.[1]; Banfield, Jillian F.[2, 3]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. [2]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, USA. [3]Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA, USA.

Genome resolved metagenomics provides insight into the role of single organisms in context of the whole microbial community. Binning algorithms cluster assembled contigs from shotgun sequencing data into putative genomes and constitute an important step in the reconstruction of organism specific metabolic pathways and community interaction networks. However, existing binning methods often fail to predict a reasonable number of genomes and report many bins of low quality and completeness. Furthermore, their performance varies between samples and biotopes. Here, we present a new computational method, which considerably increases the amount of near complete bins by combining the strengths of different binning algorithms. Using a single copy gene based scoring system our method selects the highest quality genomes from an arbitrary number of predictions and determines a non-redundant and improved set of genome bins. With our method we combined predictions of six binning algorithms that were applied to samples from environments of different complexity including soil, natural oil seeps and the human gut. Results provide evidence that our method reports substantially more near complete genomes than any of the single binning methods alone. The ability to reconstruct a higher number of near-complete genomes from metagenomics data will contribute to comprehensive genome-centered analyses of ecosystems.

Linking Microbial Genomic Capacity to Geochemical Process in the Deep Terrestrial Biosphere

Simister, Rachel L.* (rlsimister@gmail.com)[1]; Vuillemin, Aurèle[2]; Kallmeyer, Jens[2]; Hallam, Steven J.[1]; Crowe, Sean A.[1]

[1]University of British Columbia. [2]GeoForschungsZentrum.

Lake Towuti is a, 560 km2, 200-m deep tectonic lake at the downstream end of the Malili lake system, a set of five, ancient (1-2 Ma) tectonic lakes located in central Sulawesi, Indonesia. Lake Towuti has high

rates of floral and faunal endemism and is surrounded by one of the most diverse tropical forests on Earth making it a hotspot of Southeast Asian biodiversity. Lake Towuti's geochemistry is controlled by the ultramafic (ophiolitic) rocks that make up its catchment basin. Weathering of these rocks supports a strong flux of iron to the lake, leading to highly oligotrophic lake waters poor in the nutrient phosphorus and sediments extremely rich in iron (ferruginous). Physical stratification in Lake Towuti's deep basin also leads to persistently anoxic ferruginous waters below 130m depth. Though rare on Earth's surface today, such ferruginous conditions were a prominent feature of the deep ocean throughout much Earth's early history and may also be characteristic of much of the deep terrestrial subsurface. In July 2015, the Towuti Drilling Project recovered ~100m of sediment core, with sedimentation rates indicating a bottom age of ~560,000 years. Sediment was recovered in 10cm whole round core sections at the following depths; 50cm resolution for the first 5m, 1m resolution for depths 5-20m, 3m resolution for depths 20-50m and 5m resolution for the rest of the core up to 120m, for a total of 50 sediment samples. To explore the biogeochemistry and microbiology of sediment from Lake Towuti, we are sequencing small sub-unit rDNA and rRNA to recover extant and fossil biological information. Based on community fingerprinting results and in combination with geochemical profiles and process rate measurements (e.g. metabolites like CH_4, NH_4+, DIC, Fe(II), and rate measurements for methanogenesis, methane oxidation, sulphate reduction, bacterial production, dark carbon fixation, protein biosynthesis, ATP turnover, hydrogenase activity) a subset of samples will be targeted for whole community genomic DNA and RNA sequencing (metagenomic and metatranscriptomic samples), to determine the potential and expressed metabolic capacity of the deep biosphere microbial community. Our preliminary results indicate that reactive ferric Fe appears to persist throughout the sediment core at concentrations up to nearly 2 wt. %,its persistence throughout the sediment core suggests abundant reactive Fe to fuel microbial Fe reduction We have also successfully extracted DNA from samples corresponding to sediments 2000 to ~560,000 years old, showing that DNA does persist in Lake Towuti's sediments for 10s or 100s of thousands of years, possibly due to shielding against degradation through Fe-mineral association. These ferruginous sediments can support a diversity of microbial taxa (Beta and gamma-proteobacteria, Bacteroidetes, Actinobacteria and uncultivated Thaumarchaotea and MCG group I) and recovery of phototrophic and eukaryotic sequences throughout the core, may indicate the presence and persistence of ancient DNA.

Characterization of Microbial "Hotspots" of Organic Matter Degradation in a Fast-Flowing Estuary

Simon, Holly M.* (simonh@ohsu.edu)[1]; Smith, Maria W.[1]; Herfort, Lydie[1]; Rivers, Adam R.[2]

[1]Center for Coastal Margin Observation and Prediction and Institute of Environmental Health, Oregon Health & Science University, Portland, OR, USA. [2]US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA.

In fast-flowing, river-dominated estuaries, powerful physical forcing may constrain the biology in the system. In these environments, regions of extended water retention are likely to be "hotspots" of microbial biogeochemical cycling. Several lateral bays off of the main channels of the Columbia River estuary are proposed to be such hotspots. Recent evidence suggests that these areas serve as both sources and sinks of particles; in particular we seek to understand the origin and fate of biogenic particles in relation to organic matter fluxes in the estuary. To better understand the importance of the bays for processing of organic matter, we performed metagenome analysis on sediment samples from three disparate bays. These are, in order of increasing salinity and ocean influence: the freshwater Cathlamet Bay, and the brackish Youngs and Baker Bays located nearer the mouth to the south and

north of the main channel, respectively. Sediments from 11 sites across the lateral bays were collected under similar estuarine regimes in August of two different years (2011 and 2013). Total DNA was extracted, purified and subjected to Illumina paired-end sequencing, generating >10 Gbp of assembled sequence, 20 million identified coding sequences, a median ORF size of ~ 500 bp, and approximately 45 to 75% of CDS functional and phylogenetic annotations, respectively. All metagenomes were dominated by bacterial sequences, with an abundance of diatom sequences also present in the samples that were most influenced by ocean waters. Unsupervised 2D hierarchical clustering analysis divided the metagenomes into three distinct clusters, two of which were characterized by an increased relative abundance of various *Bacteroidetes taxa*. These abundance patterns were positively and significantly correlated with both diatom and bacteriophage sequences. Analysis of functional gene categories indicated corresponding enrichment of genes associated with phytoplankton utilization. Detailed analysis of the data also suggested that the ratio of diatom RuBisCO subunits provided an indication of bloom degradation stage (early vs. late), which appeared to influence the bacterial community composition in the sediments. In contrast with the other two, the third cluster contained sediment metagenomes with very low relative abundance of diatom and *Bacteroidetes* sequences, and a metabolic potential biased towards microbial growth under nutrient limitation. A working model is presented to explain the data within the context of the microbial loop and greater estuarine food web. The results of this work highlight the central role of *Bacteriodetes taxa* in phytoplankton bloom degradation and organic matter recycling in a fast-flowing estuary.

Towards Understanding the *Panicum virgatum* Microbiome — Switchgrass Geno-/Ecotype and Treatment Practices Influence the Microbial Community

Singer, Esther* (esinger@lbl.gov)[1]; Juenger, Thomas E.[2]; Woyke, Tanja[1]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, California.. [2]UT Austin.

Growing energy demands and concerns for climate change have urgently pushed forward the timeline for the implementation of biofuel energies. There are still surprisingly large gaps in our understanding of switchgrass, *Panicum virgatum*, which has been championed as one promising biofuel species. This study looks at critical plant-microbe-soil traits that may be manipulated, through breeding or agronomic management, to improve the sustainability switchgrass.

Using V4 16S rRNA as well as metagenome sequencing, we studied the bacterial and fungal microbiome in four different genotypes of switchgrass plants, including upland and lowland ecotypes, which were subjected to various fertilization and harvest treatments. Microbial community composition was studied in leaf episphere, leaf endosphere, root episphere, root endosphere, and surrounding soil. We observed 1000s of bacterial and archaeal, as well as 100s of fungal OTUs. Ordination analysis shows that both, microbial community structure and richness, are distinct according to niche origin at the family and genus level. In all compartments, differences in community composition were mainly explained by harvest practice (~33% in the bacteria and archaea, 21-32% in the fungi) and geno-/ecotype differences (39 - 59% in the bacteria and archaea, 25-43% in the fungi). Indicator species and differential abundance analysis by factor grouping are complemented with functional gene analysis associated with switchgrass geno- and ecotype, as well as harvest practice. This dataset is the first to discuss the switchgrass

microbiome and allows insight into important microbial players associated with various switchgrass growth scenarios.

Cellulose Synthase (CesA) Genes of Bread Wheat (*Triticum aestivum*, L.)

Singh, Jaswinder* (jaswinder.singh@mcgill.ca)[1]; Kaur, Simerjeet [1]; Dhugga, Kanwarpal S.[2]; Gill, Kulvinder[3]

[1]Department of Plant Science, McGill University, Sainte Anne de Bellevue, QC, Canada. [2]International Maize and Wheat Improvement Center (CIMMYT), Carretera Mexico-Veracruz, Km. 45, El Batán, Texcoco, Estado de México, C.P. 56237. [3]Department of Crop and Soil Science, Washington State University, Pullman, WA, USA.

Wheat is the most widely grown of all the crops globally, yet information on its CesA gene family is limited. We have identified 22 CesA genes from bread wheat, which include homoeologs from each of the three genomes, and named them as TaCesAXA, TaCesAXB or TaCesAXD, where X denotes the gene number and the last suffix stands for the respective genome. Sequence analyses of the CESA proteins from wheat and their orthologs from barley, maize, rice, and several dicot species (*Arabidopsis*, beet, cotton, poplar, potato, rose gum and soybean) revealed motifs unique to monocots (Poales) or dicots. Novel structural motifs CQIC and SVICEIWFA were identified, which distinguished the CESAs involved in the formation of primary and secondary cell wall (PCW and SCW) in all the species. We also identified several new motifs specific to monocots or dicots. The conserved motifs identified in this study possibly play functional roles specific to PCW or SCW formation. The new insights from this study advance our knowledge about the structure, function and evolution of the CesA family in plants in general and wheat in particular. This information will be useful in improving culm strength to reduce lodging or alter wall composition to improve biofuel production.

Antarctic Lake Metagenomes from Cyanobacterial Mats and Planktonic Microbes

Slonczewski, Joan L.* (slonczewski@kenyon.edu)[1]; Li, Wei[2]; Bush, Sean P.[1]; Froula, Jeffrey[3]; Sedlacek, Christopher[2]; Blankenhorn, Darcy[1]; Morgan-Kiss, Rachael M.[2]

[1]Kenyon College, Gambier, OH, USA. [2]Miami University of Ohio, Oxford, OH, USA. [3]DOE, JGI, CA, USA.

Many Antarctic organisms appear morphologically similar to their mesophilic close relatives, yet the Antarctic organisms can survive for years in ice including six months of annual darkness. In culture, the cyanobacteria recover within minutes of hydration, while protists and microscopic invertebrates appear within another week. Metagenomic analysis and genome assembly will enable us to compare the cold-adapted organisms with mesophilic taxa, and new taxa may be discovered. In addition, reference-independent analysis using MetaBat and other binning pipelines may reveal novel taxa.

Antarctic glacier-fed lakes of the Taylor Valley contain distinctive cyanobacterial mats and planktonic phototrophs. The cyanobacterial “liftoff” mat communities, dominated by Microcoleus, grow slowly at the benthos at low light intensity. The benthic mat reaches upward, buoyed by oxygen bubbles. Pieces of mat

break off and float to the undersurface of the ice cover. During winter, water freezes, adding to the ice layer, but above the valley's dry winds ablate the ice. After 5-10 years, the dessicated pieces of mat emerge and are blown to new habitats. In the summer, the mat fragments may enter thawed "moats" around the permanent ice, or else meltwater outside glaciers such as the Canada Glacier that dams Lake Fryxell. The cyanobacteria support a diverse ecosystem, including ciliates and flagellates, rotifers, tardigrades, and nematodes. Planktonic microbes were collected via Niskin bottle from beneath the ice of Lake Fryxell and Lake Bonney and concentrated by filtration. Cyanobacterial mats were collected from emerging clumps at the surface of Lake Fryxell. Following two months storage at -80 °C, mat samples were cultured at 10 °C with 5% PAR. Within two weeks the samples bubbled oxygen and showed diverse morphotypes of cyanobacteria as well as stalked ciliates and nematodes. DNA from original stored samples was isolated by the PowerBiofilm kit (MoBio). In all, DNA was isolated for six mat samples from Lake Fryxell, three planktonic samples from Lake Fryxell, and three planktonic samples from Lake Bonney. The DNA was sent to JGI for Illumina sequencing and metagenome analysis through the Community Science Program.

A preliminary sample of mat DNA was analyzed by the MOCAT pipeline. The taxonomic profile of the metagenome was assessed by mapping sequence reads to a set of marker genes in the mOTU database. Preliminary results show 69% Microcoleus, 11% Nostoc, 2.5% Acidovorax, 2.5% Phenylobacterium, and 4% various actinomycetes. The Microcoleus and Nostoc will be of interest to obtain whole genomes for comparison with closely related mesophiles, with the aim of revealing mechanisms of low-temperature adaptation. The unexpected prominence of actinomycetes suggests a potential source of novel antibiotics.

Soil Microbial Community Responses to a Subterranean Coal Mine Fire Revealed by Metagenome Analysis

Sorensen, Jackson W.* (sorens75@msu.edu); Shade, Ashley

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824

The Centralia, Pennsylvania coal seam fire has been burning near-surface since 1962. It has created an extreme environment that supports coal fire-adapted microbial life. Heat, steam and combustion products vent upward from the fire through the overlying soils, increasing soil surface temperatures to over 80°C. Soil chemistry is altered by both spontaneous and microbial-mediated chemical reactions. As the fire expands into new areas, it also retreats from some affected sites, which can then recover. This unusual habitat provides an opportunity to investigate the selective pressures and community processes that promote microbial community stability in the face of extreme, ongoing disturbance. Supported by a JGI small-scale community sequencing project, we generated and analyzed 13 metagenomes from surface soils along a chronosequence of historical fire activity. For our preliminary analyses, we first compared changes in the metagenome content to changes in bacterial and archaeal community structure by 16S rRNA bacterial amplicon analysis, using the chronosequence to interpret changes along the fire gradient. We also analyzed the KEGG ortholog content of currently fire-affected in comparison to recovered and reference soils, and identified pathways overrepresented within hot soils, including trehalose metabolism (which is involved in both dormancy and heat and desiccation stress responses) and the hydroxypropionate-hydroxybutyrate cycle, an autotrophic carbon fixation pathway. Finally, we used the targeted gene assembler Xander to analyze the housekeeping gene rplB in the Centralia metagenomes and make additional taxaonomic assignments. Our results indicate that the Centralia soils

harbor yet-unidentified diversity, and can provide insights into the eco-evolutionary processes that drive microbial community changes in response to an extreme environmental filter.

Accessing the Potential of the Core Microbiome of a Bioenergy Crop in Plant Development and Biomass Accumulation

Souza, Rafael S.C.* (scs.rafael@gmail.com)[1, 5]; Okura, Vagner K.[1]; Armani, Jaderson S. L.[1, 5]; Damasceno, Natália B.[1, 5]; Jorrín, Beatriz[2]; Lozano, Núria[2]; Silva, Márcio J.[1]; González-Guerreiro, Manuel[2]; Araújo, Laura M.[1, 5]; Verza, Natália F.[1]; Bagheri, Homayoun C.[3]; Imperial, Juan[2, 4]; Arruda, Paulo[5]

[1]Center of Molecular Biology and Genetic Engineering (UNICAMP), Campinas, São Paulo/Brazil. [2]Centro de Biotecnología y Genômica de Plantas (CBGP), Campus de Montegancedo, Universidad Politécnica de Madrid, Madrid/Spain. [3]Repsol Technology Center, Madrid/Spain. [4]Consejo Superior de Investigaciones Científicas, Madrid/Spain. [5]Department of Genetics and Evolution (UNICAMP), Campinas, São Paulo/Brazil.

The plant microbiomes have emerged as a potential biotechnological resource for sustainable crop production. However, studies evaluating the microbiomes of economically important crops are scarce. Here, we present a comprehensive inventory of the structure and assemblages of the bacterial and fungal communities associated with sugarcane and evaluate its potential benefit for plant growth. Naturally occurring bacterial and fungal communities were accessed in plants grown without the addition of fertilizers. Roots, stalks, top leaves and young shoots were sampled from plants 4, 6, 8 and 10 months after budding. Analysis of 312 libraries at a deepness of 1 million reads per library revealed 23,811 bacterial OTUs and 11,727 fungal OTUs inhabiting the endophytic and exophytic compartments of roots, shoots and leaves. These bacterial and fungal communities originate primarily from the native soil around plants and colonize plant organs in distinct patterns. The sample type is the main driver of fungal community assemblage, and the organ compartment plays a major role in bacterial community assemblage. Core bacterial and fungal communities composed of less than 20% of the total microbial richness but accounting for over 90% of the total microbial relative abundance persists along plant development. The core microbiome composition varies among organs and compartments and is formed by highly abundant groups. A microbial culture collection representing a significant part of the sugarcane roots and stalks microbiome was constructed. Cross-referencing the microbiome assemblages with the microbes in the culture collection allowed the development of a consortium inoculum representing the most abundant endophytic and exophytic bacteria. Inoculation experiments using maize as a model plant revealed that the potential of microbial consortium in helping plant growth and development.

Project website: canein.cbmeg.unicamp.br/saccharome

JGI Plant Gene Atlas: Improving Functional Annotations of JGI Plants Using Transcriptional Data

Sreedasyam, Avinash* (asreedasyam@hudsonalpha.org)[1]; Plott, Christopher[1]; Qualls, Morgan[1]; Jenkins, Jerry[1]; Grimwood, Jane[1]; Carlson, Joseph[2]; Goodstein, David[2]; Juenger,

Thomas E.[4]; Yuhong Tang[5]; Tuskan, Gerald A.[6]; Brutnell, Tom[7]; Merchant, Sabeeha S.[8]; Rensing, Stefan A.[9]; Mullet, John E.[10]; Mockler, Todd[7]; Stacey, Gary[3]; Schmutz, Jeremy[1, 2]

[1]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA. [2]US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. [3]Divisions of Plant Science and Biochemistry, National Center for Soybean Biotechnology, University of Missouri, Columbia, MO, USA. [4]University of Texas at Austin, Department of Integrative Biology, Austin, TX, USA. [5]Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, OK, USA. [6]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [7]Donald Danforth Plant Science Center, St. Louis, MO, USA. [8]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA. [9]University of Marburg, Karl-von-Frisch-Str. 8, 35043 Marburg, Germany. [10]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, USA.

The JGI Plant Gene Atlas is a large, coordinated effort to add functional information to JGI Plant Flagship Genomes. The primary goal of this project is to develop dense RNA-seq data sets for Plant Flagships to form reference transcriptomes across common tissues and conditions and to provide the ability to compare expression across conditions within a plant and from orthologous genes across the JGI Plant genomes. This substantial amount of updateable transcriptomic resource is directly available to JGI users through the JGI Plant Portal at phytozome.jgi.doe.gov. The secondary goal is to provide a technology test bed to further develop genomic techniques that illuminate the function of plant genes and plant regulatory and pathway elements. Gene Atlas project currently comprises 1,078 samples from twelve JGI Plant Flagship genomes: *Chlamydomonas* (algal model), *Physcomitrella* (moss model), *Brachypodium* (a C3 grass model), switchgrass (a woody perennial crop plant), Hall's panicgrass (grass model), *Setaria italica* (grain and forage crop), *Setaria viridis* (model C4 grass), Sorghum (a C4 grass bioenergy crop and model) [6 monocots], *Arabidopsis* (model for plant genetics and biology), soybean (legume model and crop plant), Medicago (legume model) and poplar (biomass tree crop).

Results from our genome version controlled data analysis pipelines which facilitates identifying gene clusters showing tissue/condition specific expression patterns, visualize expression profiles of gene sets across Gene Atlas plants and adding experimental and orthologous based functional annotations will be presented.

In addition to standard tissues and conditions, the Gene Atlas includes a comparative condition across the plants focused on nitrogen metabolism providing a single condition that can be compared broadly from the minimal Chlamydomonas, into the early plant model Physcomitrella, and through the dicots and monocots included in the study. We will present our results on transcriptional modulation in response to changes in available nitrogen source and conserved gene expression networks across the phylogeny.

Tracking Distinct Freshwater Populations Using Single Cell Genomes and Metagenomics

Stevens, Sarah L.R.* (sstevens2@wisc.edu)[1]; Garcia, Sarahi L.[1, 2]; Crary, Benjamin[3]; Martinez-Garcia, Manuel[4]; Stepanauskas, Ramunas[5]; Woyke, Tanja[6]; Tringe, Susannah G.[6]; Andersson, Siv[7]; Bertilsson, Stefan[2]; Malmstrom, Rex[6]; McMahon, Katherine D.[1, 3]

[1]Department of Bacteriology, University of Wisconsin-Madison Madison, WI, USA . [2]Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [3]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA. [4]University of Alicante, Alicante, Spain. [5]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA. [6]US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. [7]Department of Molecular Evolution, Uppsala University, S-752 36 Uppsala, Sweden.

Prior to this work, species have been defined using the average nucleotide identity (ANI) of sequenced isolate genomes and populations defined based on genetic discontinuities found by mapping metagenomic reads to references. To unite both of these definitions, we compared 33 single-cell genomes (SAGs) collected from three freshwater lakes and the metagenomic reads from Lake Mendota (WI, USA) which map to these references. Subsets of the 33 freshwater SAGs included 14 and 10 SAGs from acI Actinobacteria and alfV Betaproteobacteria (freshwater sister clade to marine SAR11 clade) lineages, respectively. We found 16S diversity, genome wide ANI, and average amino acid identity (AAI) generally followed the same trend but these summary statistics obfuscated some fine-scale genomic differences. Breaking down the ANI calculations, by cutting the genomes into 300bp shreds, we found that the number of shreds that recruit as well as their range of nucleotide identities separated them into genetically discrete populations (sequence-discrete). We observed the natural sequence-discrete populations to which these references belong by mapping the metagenomic reads from Lake Mentoda back to the SAG reference. Using SAGs collected from several geographically isolated lakes, we saw that these lineages are globally distributed but adapted for their local environment. We also tracked the diversity and relative abundance of these populations through the Lake Mendota metagenomic time series. We found that even closely related (>99% 16S rRNA gene identity over 412 bp partial alignment) genomes from the same lake recruited different reads and showed differential abundance across the time series if they belong to different sequence-discrete populations, suggesting that these populations are not only genetically discrete but ecologically distinct.

A Day in the Life of *Chlamydomonas*

Strenkert, Daniela* (DanielaS@chem.ucla.edu)[1]; Gallaher, Sean D.[1]; Schmollinger, Stefan[1]; Pellegrini, Matteo[2]; Umen, James G.[3]; Merchant, Sabeeha S.[1]

[1]Institute for Genomics and Proteomics, University of California, Los Angeles, California, USA. [2]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California, USA. [3]Donald Danforth Plant Science Center, St. Louis, Missouri, USA.

The green alga *Chlamydomonas reinhardtii* is a reference organism for many questions in biology, especially photosynthesis. We generated a multilayer dataset, combining transcriptomic data with physiological measurements over the course of a *Chlamydomonas* cell cycle. Cultures were grown in a flat panel bioreactor system for reproducible synchronization and sampledin triplicate at regular intervals over the course of a 12-h light/12-h dark cycle to measure the transcriptome (by RNA-Seq). Nearly the entire genome (81% of all transcribed genes) were differentially expressed at one or more time points, and these were grouped into 16 major expression patterns describing the sequence of events during a day in the alga's life. Among the most interesting responses were the orchestration of cell division and fermentation during the night, and the balance between photoprotection and photosynthesis during the day.

We will exploit this experimental setup to answer fundamental questions about chromatin dynamics in photosynthetic organisms and beyond, which are difficult to address in any other species. Correlation of

our transcriptome data with changes of specific histone modifications under different environmental conditions will reveal combinatorial patterns of different histone modifications that define levels of transcriptional activity, so-called chromatin states. These patterns can vary significantly between different organisms and it will be interesting to see how the green alga´s chromatin states relate to those species outside the Chlorophyte lineage. Our diurnal experiments where thousands of genes cycle periodically might also reveal epigenetic memory marks that might contrast with those on facultative genes or constitutively-expressed genes. In parallel, we will also probe genes that are transcribed only under specific nutrient deficiency to study epigenetic memory. The copper regulon consisting of dozens of genes is one target of interest. We will establish whether stable epigenetic memory prevails at promoters that are activated by copper limitation after the cells have returned to copper-replete media.

Linking Microbial Identity and Function in Phototrophic Mats and Biofilms

Stuart, Rhona K.* (stuart25@llnl.gov)[1]; Mayali, Xavier[1]; Boaro, Amy[2]; Everroad, Craig[3]; Nilson, Daniel J.[1]; Weber, Peter[1]; Lipton, Mary[2]; Bebout, Brad M.[3]; Pett-Ridge, Jennifer[1]; Thelen, Michael P.[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, USA. [2]Pacific Northwest National Laboratory, Richland, WA, USA. [3]Exobiology Branch, NASA Ames Research Center, Moffet Field, CA, USA.

Cyanobacteria are globally distributed primary producers and the fate of their fixed C influences microbial biogeochemical cycling. This fate is complicated by cyanobacterial degradation and assimilation of extracellular organic matter (EOM), but because cyanobacteria are often assumed poor competitors for EOM consumption, the regulation of this activity has not been well-tested. In mats and biofilms, this is especially relevant because cyanobacteria produce an extensive organic extracellular matrix, providing the community with a rich source of nutrients. Light is a well known regulator of cyanobacterial metabolism, so here we characterized the effects of light availability on the incorporation of EOM by a filamentous cyanobacterium. We used stable isotope tracing at the single cell level to quantify photoautotroph assimilation in a mixed biofilm community under different metabolic conditions, and we integrated these results with metaproteo-genomics of the biofilm community to elucidate metabolic status. We find cyanobacteria in our biofilms not only make EOM, but use it directly both in the light and in the dark, and are important competitors for EOM consumption. Additionally, response of associated microbes to light availability suggests both cooperation and competition may dictate resource partitioning in these systems. The distinct light-dependent cyanobacterial organic matter incorporation patterns imply storage of C in the extracellular domain, and light-driven incorporation for micronutrient acquisition.

This research was supported by the U.S. Department of Energy Office of Biological and Environmental Research, Genomic Science Program under FWP SCW1039, a JGI CSP (#701), and at Pacific Northwest National Lab was supported by the OBER/GSP under the Pan-Omics project. Work at LLNL was performed under the auspices of DOE Contract DE-AC52-07NA27344.

Synthetic Biology Pipeline for Constructing and Testing Combinatorics of Cellulose- and Lignin-Degrading Enzymes

Suzuki, Yo* (ysuzuki@jcvi.org); Kostylev, Maxim; Otwell, Anne E.; Hanly, Timothy J.

J. Craig Venter Institute, La Jolla, CA.

Synergism among enzymes is commonly observed in nature for degradation of biopolymers such as cellulose and lignin. Although aspects of synergy have been incorporated in artificial systems such as commercial cellulase cocktails for hydrolyzing pretreated biomass for biofuel applications, new paradigms of enzyme synergy continue to emerge through studies of mixed samples of enzymes and populations of microbes (reviewed in Kostylev & Wilson, Biofuels. 2012). We have an ongoing BETO project to create mixtures of biomass-active enzymes on the yeast cell surface to capture some of the natural and artificial complexity in a single organism. With DNA synthesis support from JGI, we would like to expand on our current success with the initial set of cellulose-acting enzymes to add more cellulases and also ligninases to the arsenal of enzymes to be tested in our system. Our motivations are to identify sets of enzymes that facilitate the degradation of dilute acid-pretreated corn stover, where the solid fraction is ~27% lignin. With genes from JGI, we would also aim to find sets of enzymes that act effectively on lignin model compounds. The inclusion of ligninases in our work would promote the understanding of these important enzymes and could lead to the discovery of new synergistic interactions between different enzyme classes. DNA synthesis is required for incorporating enzymes that have been only detected in metagenomics studies where genomic DNA samples are unavailable. Moreover, DNA synthesis enables the rapid conversion of the digital information into expression constructs with proper elements for expression, secretion, and cell surface display. We have an established pipeline for mining databases, designing sequences, incorporating synthetic genes into expression constructs, introducing these enzyme constructs into yeast, testing for enzyme activity, constructing multi-enzyme strains, and selecting top-performing strains. Our data will be useful in the future for designing enzyme cocktails for saccharification or engineering fermentative organisms that express a subset of enzymes that synergize with currently available cocktails. The synthetic approach for discovery of multi-enzyme synergy has the potential to produce a step change in addressing the cost-effective conversion of biomass to fuels and chemicals.

Generation of Octaploid Switchgrass by Seedling Treatment

Tobias, Christian* (christian.tobias@ars.usda.gov)[1]; Yoon, Sangwoong[3]; Hernlem, Bradley[1]; Edme, Serge[2]; Sarath, Gautam[2]; Palmer, Nathan[2]

[1]USDA-ARS, Albany, California. [2]USDA-ARS, Lincoln, NE. [3]University of California, Davis.

Octaploid and tetraploid switchgrass populations are reproductively isolated. To create new methods of population improvement and to study the effects of whole genome duplication on switchgrass a method to efficiently induce fertile octaploid switchgrass by seedling treatment of tetraploid individuals with microtubule inhibitors was developed. This improved method was adapted from methods established on barley and was applied on the switchgrass bioenergy cultivar ‘Liberty’.

Tetraploid seedlings at the 2-3 leaf stage were treated with colchicine or amiprophos methyl to induce genome duplication. Resulting chimeric lines were fixed by intercrossing octaploid sectors to produce a population of 20 octaploid families. Fertility of octoploid sectors was significantly reduced relative to

tetraploid sectors however caryopsis (seed) size was significantly increased. The estimated genome sizes of the initial chimeric lines did not always agree with expectations as a result of higher order, incomplete, or partial genome duplications. Overall rates of octaploid identification were 12%-28%. This method is an improvement over existing methods because it does not involve tissue culture and can be used to generate many different unique polyploid genotypes in a single experiment. The new population will be incorporated into breeding efforts and be used to study questions relating to the effects of polyploidy on genome structure and gene expression. Increases in seed size and seed weight are correlated with high rates of germination, seedling vigor, and increased establishment in switchgrass, and this trait is a selected criterion in most forage breeding programs. Large seed size can confer an advantage under sub optimal establishment conditions. Induced polyploidy may benefit breeding programs in the northernmost switchgrass growing regions where octaploid cultivars outperform by promoting genetic exchange between populations.

Monitoring the Transcriptional Response of a Representative Low-Salinity Ammonia-Oxidizing Thaumarchaeote to Shifts in Environmental Conditions

Tolar, Bradley B.* (btolar1@stanford.edu); Francis, Christopher A.

Department of Earth System Science, Stanford University, Stanford, CA.

Ammonia oxidation plays a key role in aquatic and terrestrial ecosystems, with the organisms capable of this process as major players in the nitrogen and carbon cycles. Ammonia-oxidizing Archaea (AOA) are members of the phylum Thaumarchaeota, and are generally more abundant than their bacterial counterparts in marine systems; however, in terrestrial and estuarine environments their contribution varies. Though physiological studies on available thaumarchaeal pure and enrichment cultures have provided further insight into their metabolism beyond ammonia oxidation, we still know relatively little about what factors affect their growth. Our lab has maintained two closely-related strains of Candidatus Nitrosoarchaeum limnia (SFB1 and BG20; 96.3% identity across the entire genome) that were isolated from the San Francisco Bay-Delta. These stable enrichment cultures (90-97% purity) represent an abundant clade of Thaumarchaeota found in a variety of estuarine and freshwater samples across the globe; it is also one of the few thaumarchaeotes with genes for chemotaxis and motility, indicating its ability to respond to changes in its environment. Our CSP project seeks to understand how such changes in environmental conditions impact the growth and metabolism of Thaumarchaeota, using transcriptomics with strain SFB1 and BG20. Relevant environmental changes targeted in this study include temperature, salinity, pH, and concentrations of oxygen, ammonia, and nitrite. Strains will be grown at optimal conditions as well as at viable high and low extremes for each environmental factor, to determine which genes show increased (or decreased) transcription under stress. The resulting data will be compared between the two strains in order to define general characteristics of Ca. N. limnia and determine if genomic differences translate into physiological differences. For example, previous work has shown that BG20 is more sensitive to elevated nitrite concentrations and has a more narrow salinity tolerance. By performing these transcriptomic studies across a range of environmental conditions, we hope to gain novel physiological insights into this widespread yet scarcely understood species of low-salinity AOA.

Functional genomics and plant Modeling for Sorghum Improvement

Truong, Sandra K.* (thkhavi@tamu.edu)[1, 2]; McCormick, Ryan F.[1, 2]; Sreedasyam, Avinash[3]; Shu, Shengqiang[4]; Weers, Brock D.[2]; McKinley, Brian A.[2]; Mattison, Ashley J.[1, 2]; Jenkins, Jerry[3, 4]; Kennedy, Megan[4]; Amirebrahimi, Mojgan[4]; Grimwood, Jane[3, 4]; Schmutz, Jeremy[3, 4]; Mullet, John E.[1, 2]

[1]Interdisciplinary Program in Genetics, Texas A&M University, College Station, TX. [2]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX. [3]HudsonAlpha Institute for Biotechnology, Huntsville, AL. [4]United States Department of Energy Joint Genome Institute, Walnut Creek, CA.

Sorghum bicolor is a prominent food, forage, and bioenergy crop, and sorghum productivity will benefit from continued development of genomic resources and models to predict sorghum performance. In collaboration with the Joint Genome Institute and the HudsonAlpha Institute for Biotechnology, a suite of improved genomic resources for sorghum were generated, including a more accurate genome assembly, transcriptome annotation, and atlas of transcript expression across multiple tissues and developmental stages. These genomic resources combined with modeling approaches to predict plant performance will improve the rate at which performance gains can be achieved in sorghum. As an example, genomics and plant modeling were used to identify a sorghum gene regulating leaf angle and to predict the effect of leaf angle on light interception and performance. These procedures are being extended to additional phenotypes using 3D reconstructions of sorghum plants obtained from depth imaging. Continued development and application of these resources will accelerate sorghum improvement.

De Novo Reconstruction of the Spinach Coding Genome Using Full-Length Transcriptome

Tseng, Elizabeth* (etseng@pacb.com)[1]; Ashrafi, Hamid[2, 3]; Hulse-Kemp, Amanda M.[3]; Van Deynze, Allen[3]

[1]Pacific Biosciences, Menlo Park, CA. [2]North Carolina State University, Raleigh, NC. [3]University of California, Davis, Davis, CA.

For highly complex and large genomes, a well-annotated genome may be computationally challenging and costly, yet the study of alternative splicing events and gene annotations usually rely on the existence of a genome. Long-read sequencing technology provides new opportunities to sequence full-length cDNAs, avoiding computational challenges that short read transcript assembly brings. The use of single molecule, real-time sequencing from Pacific Biosciences to sequence transcriptomes (the Iso-SeqTM method), which produces *de novo*, high-quality, full-length transcripts, has revealed an astonishing amount of alternative splicing in eukaryotic species. With the Iso-Seq method, it is now possible to reconstruct the transcribed regions of the genome using just the transcripts themselves. We present Cogent, a tool for finding gene families and reconstructing the coding genome in the absence of a high-quality reference genome. Cogent uses k-mer similarities to first partition the transcripts into different gene families. Then, for each gene family, the transcripts are used to build a splice graph. Cogent identifies bubbles resulting from sequencing errors, minor variants, and exon skipping events, and attempts to resolve each splice graph down to the minimal set of reconstructed contigs. We apply Cogent to the Iso-Seq data for spinach, Spinacia oleracea, for which there is also a

PacBio-based draft genome to validate the reconstruction. The Iso-Seq dataset consists of 68,263 full-length, Quiver-polished transcript sequences ranging from 528 bp to 6 kbp long (mean: 2.1 kbp). Using the genome mapping as ground truth, we found that 95% (8045/8446) of the Cogent gene families found corresponded to a single genomic loci. For families that contained multiple loci, they were often homologous genes that would be categorized as belonging to the same gene family. Coding genome reconstruction was then performed individually for each gene family. A total of 86% (7283/8446) of the gene families were resolved to a single contig by Cogent, and was validated to be also a single contig in the genome. In 59 cases, Cogent reconstructed a single contig, however the contig corresponded to 2 or more loci in the genome, suggesting possible scaffolding opportunities. In 24 cases, the transcripts had no hits to the genome, though Pfam and BLAST searches of the transcripts show that they were indeed coding, suggesting that the genome is missing certain coding portions. Given the high quality of the spinach genome, we were not surprised to find that Cogent only minorly improved the genome space. However the ability of Cogent to accurately identify gene families and reconstruct the coding genome in a *de novo* fashion shows that it will be extremely powerful when applied to datasets for which there is no or low-quality reference genome.

Diversity of Transcriptome Regulatory Networks in Maize and Association with Biomass and Biofuel Traits

Vaillancourt, Brieanne* (vaillan6@msu.edu)[1, 2]; Beddows, Ian[1, 2, 3]; Brohammer, Alex B.[4]; Hirsch, Cory D.[5]; Wang, Mei[6]; Barry, Kerrie[6]; de Leon, Natalia[7, 8]; Kaeppler, Shawn M.[7,8]; Hirsch, Candice N.[4]; Buell, Robin C.[1, 2]

[1]Department of Plant Biology, Michigan State University, East Lansing, MI 48824. [2]DOE Great Lakes Bioenergy Research Center, East Lansing, MI 48824. [3]Department of Biology, Heinrich-Heine University Düsseldorf, D-40225 Düsseldorf, Germany. [4]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108. [5]Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108. [6]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598. [7]Department of Agronomy, University of Wisconsin-Madison, Madison, WI 53706. [8]DOE Great Lakes Bioenergy Research Center, Madison, WI 53706.

Interrogation of the maize pan-transcriptome in a diversity panel of 503 inbred maize inbred lines revealed that not only genomic but also transcriptomic variation contributed to phenotypic variation of biofuel feedstock traits. Construction of transcriptional networks across diverse maize inbred lines will permit the exploration of the gene regulatory mechanisms controlling biomass and biofuel related plant architecture and biochemical traits. Regulatory networks involving both mRNA and miRNAs allow the identification of 1) co-regulated transcripts, 2) regulatory elements controlling expression (i.e., transcription factors, DNA regulatory regions such as promoters), and 3) miRNAs that negatively regulate mRNA transcript abundance. Phenotypic data, when coupled with mRNA and miRNA transcriptome data, can be used to perform comparative analyses of transcriptional networks from divergent genotypes to identify genotype specific regulatory networks. We are using a dataset of RNA-seq (mRNA-Seq and small RNA-Seq) and whole genome resequencing across 35 diverse maize inbred lines and their hybrid progeny to evaluate transcriptional networks in five diverse tissues (root, shoot, endosperm, internode, and leaf). The dataset will also be used to associate genetic variation including single nucleotide polymorphisms, copy number variation, and presence/absence variation with genetic control of key plant architecture and biochemical traits relevant to biomass and biofuel traits.

This work was funded by the Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494). The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231. CDH was supported by a National Science Foundation National Plant Genome Initiative Postdoctoral Fellowship in Biology (Grant No. 1202724). ABB was supported by the DuPont Pioneer Bill Kuhn Honorary Fellowship. IB was supported with funds from iGRAD-Plant Graduate School GRK1525.

Functional Genomics of Moss-Cyanobacteria Interactions in Boreal Forest Ecosystems

Warshan, Denis* (denis.warshan@su.se)[1]; Espinoza, Josh L.[2, 3]; Richter, Alexander R.[2, 3]; Dupont, Christopher L.[2]; Weyman, Philip D.[3]; Rasmussen, Ulla[1]

[1]Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden. [2]Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA. [3]Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, La Jolla, CA.

Boreal forests are important ecosystems involved in global carbon cycling. The nitrogen cycling in these ecosystems is largely determined by a symbiotic association between feather mosses and diazotrophic cyanobacteria that fix majority of nitrogen flowing into boreal ecosystems. Recent secondary ion mass spectrometry (SIMS) data shows accumulation of cyanobacterial fixed nitrogen in mosses colonized by Nostoc cyanobacteria, establishing the flow of nitrogen from symbiont to host. Because nitrogen is often limiting in boreal forests, the interaction between the cyanobacteria and the mosses greatly affects the productivity of this ecosystem that makes up almost 30% of Earth's forested land. We seek to understand the genetic diversity of the cyanobacteria associated with the mosses, the molecular steps leading to the moss-cyanobacterial association, and the molecular currencies of the symbioses.

We present current progress in a Joint JGI/EMSL Community Science Project combining genomics, transcriptomics, and proteomics to characterize the feather moss-cyanobacteria association that covers the understory in boreal forests worldwide. Recent data from phylogenetic analysis of moss-associated cyanobacteria and from moss-cyanobacteria reconstitution experiments has informed our two central hypotheses: I) the moss-associated cyanobacteria are very diverse and different moss species host unique cyanobacterial communities, and II) the moss and cyanobacteria exchange and sense diffusible signals to establish and maintain a successful nitrogen-fixing association. Our approach is to test these hypotheses targets all levels of the central dogma and will culminate in genetic testing of hypotheses using targeted gene knockout in the symbiont of key genes derived from the results.

We sequenced the genomes of five different Nostoc spp. that are able to form symbiotic associations with two feather moss species (Pleurozium schreberi and Hylocomium splendens). As a control, we also sequenced the genome of one Nostoc sp. that is unable to form symbioses with the mosses. Comparative genome analysis of these cyanobacterial species allowed us to probe the genomic diversity of feather moss-associated Nostoc strains and identify a set of 32 genes differentially retained in genomes of symbiotic competent cyanobacteria compared to the non-symbiotic competent strain. We assembled gene knockout cassettes for five of these differentially retained genes

to create knock-out mutants related to functions like cellular differentiation, signaling and sulfur metabolism.. The five knock-out mutants will thus be assayed for their capacity to form nitrogen-fixing associations.

We also obtained transcriptomic and proteomic data for both the feather moss and Nostoc partners grown in isolation, together, or with chemical contact only through filter separation. The generated genome sequences provided the sequence database necessary for these studies: comparing the transcriptomic and proteomic responses of both partners during symbiotic competent and non-competent interactions will allow us to identify potential mechanisms used to establish and maintain the symbiosis and design additional targets for gene knockout and testing of symbiotic association.

Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: The Genomes of Soil and Plant-Associated and Newly Described Type Strains

Whitman, William B.* (whitman@uga.edu)[1]; Woyke, Tanja[2]; Kyrpides, Nikos[2]; Klenk, Hans-Peter[3]; Zhou, Yuguang[4]

[1]Department of Microbiology, University of Georgia, Athens, GA. [2]US Department of Energy Joint Genome Institute, Walnut Creek, California. [3]School of Biology, Newcastle University, Newcastle upon Tyne, UK. [4]China General Microbiological Culture Collection Center (CGMCC), Institute of Microbiology, Chinese Academy of Sciences, Beijing, China.

The Genomic Encyclopedia of Bacteria and Archaea (GEBA) project was launched by the JGI in 2007 as a pilot project to sequence about 250 bacterial and archaeal genomes of elevated phylogenetic diversity. Herein, this approach was extended to type strains of prokaryotes associated with soil or plants and their close relatives as well as type strains from newly described species. Because cultivation and preparation of DNA is one of the major limitations in genomic sequencing of type strains, this project utilizes the expertise of individual investigators to submit DNA to JGI for sequencing and annotation. Understanding the microbiology of soil and plants is critical to many DOE mission areas, such as biofuel production from biomass, biogeochemistry, and carbon cycling. Sequencing of type strains of novel species while they are being described is also targeted. Since 2006, about 630 new species have been described per year, many of which are closely aligned to DOE areas of interest in soil, agriculture, degradation of pollutants, biofuel production, biogeochemical transformation, and biodiversity. Including these into the expanding database of genome sequences will provide further understanding of prokaryotic diversity. Since the project began in the Fall of 2013, individual investigators proposed 777 type strains for genome sequencing, and 534 of these strains were approved. Sequences for 130 genomes have been completed or are in progress. Projects approved were largely for type strains from soils, plant associated and saline soils and were contributed by investigators from 14 nations, chiefly India, Spain, United Kingdom, South Africa, and China. In addition, approval was obtained for 380 type strains provided by the China General Microbiological Culture Collection Center (CGMCC), which possesses a large collection of type strains isolated in China. Sequences for 266 of these have been completed or are in progress. Therefore, this project has significantly increased the number of genome sequences for type strains, especially among plant and soil associated species

Large Scale Discovery and Deorphanization of Natural Products Using Fungal Artificial Chromosomes and Untargeted Metabolomics (FAC-MS)

Wu, Chengcang* (cwu@intactgenomics.com)[4]; Clevenger, Kenneth D.[1, 2]; Bok, Jin Woo[3]; Ye, Rosa[4]; Miley, Galen P.[1]; Velk, Thomas[3]; Chen, Cynthia[4]; Yang, KaHoua[3]; Gao, Peng[2]; Lamprecht, Matthew[4]; Thomas, Paul M.[2]; Islam, M.N.[4]; Keller, Nancy P.[3]; Kelleher, Neil L.[1, 2]

[1]Department of Chemistry, Northwestern University, 633 Clark St, Evanston, IL 60208, USA. [2]Proteomics Center of Excellence, Northwestern University, 633 Clark St, Evanston, IL 60208, USA. [3]Departments of Medical Microbiology and Immunology and Bacteriology, University of Wisconsin at Madison, 500 Lincoln Dr, Madison, Wisconsin, 53706, USA. [4]Intact Genomics, Inc. 1100 Corporate Square Drive, Suite 257, St Louis, Missouri, 63132, USA.

The vast majority of fungal-encoded chemical space is uncharted due to difficulties culturing and genetically manipulating many fungi. We report a scalable technology using fungal artificial chromosomes (FACs) to capture full-length secondary metabolite biosynthetic gene clusters (BGCs) derived from unbiased large-insert libraries of genomic DNA1 of Aspergillus terreus, A. aculeatus, and A. wentii (fig. A). Host A. nidulans strains transformed with FACs (fig. B) are screened by untargeted liquid chromatography-mass spectrometry with ultrahigh mass accuracy (fig. C), and FAC-encoded products are identified by a robust scoring system to identify spectral features most likely associated with each FAC2 (fig. D). Deletions of specific backbone and tailoring genes within FACs (achieved through facile E. coli genetics) then confirm assignment of products to BGCs and facilitate analyses of biosynthesis (fig. E). We use this “FAC-MS” platform to screen 56 FACs and report detection of SM products from 17/56 (30% hit rate), including an A. terreus FAC encoding three distinct products identified as a novel lipopeptide, a terpenoid from the ophiobolin family of compounds, and the orphan benzodiazepine benzomalvin A, which is 300-fold overexpressed relative to endogenous levels in A. terreus. Deletants of benzomalvin A backbone and tailoring genes allow us to propose and test the first biosynthetic model for this molecule. Wide application of the FAC-MS pipeline should have a major impact on fungal natural products research in the mid-term future. For example, the application of FAC-MS to a collection of over 200 Wisconsin Aspergillus strains collected Dr. Martha Christensen, with an average of 50 BGCs per strain and an estimated hit rate of >20%, would lead to validated assignments of at least 2,000 compounds to their BGCs.

The Development of a Bioinformatic Pipeline for ssu-rRNA Novelty Measurement

Wu, Dongying* (dygwu@ucdavis.edu)[1, 2]; Eisen, Jonathan, A.[2]

[1]US Department of Energy Joint Genome Institute, Walnut Creek, California. [2]University of California, Davis, Davis, California.

Small subunit rRNA (ssu-rRNA) has been widely adopted as a “phylogenetic marker” for studies of diverse living organisms. Ssu-rRNA sequences are usually the only information available for the deep evolutionary branches (phyla) of bacteria and archaea that contain no cultivated representatives. In an effort to help us prioritize the selection of such novel organisms for sequencing by the single cell genomics technology, we’ve developed a automatic pipeline to measure the phylogenetic novelty of ssu-

rRNA. By building phylogenetic trees of environmental ssu-rRNA with reference sequences, we first score rRNA sequences by how distant they are, in a phylogenetic tree, from organisms that have had their genome sequenced already. This we refer to as “Genomic Novelty” or GN. We also developed a scoring scheme for how distant an rRNA sequence is from any rRNA sequence in SILVA database. This we refer to as “rRNA novelty” or RN. By analyzing ssu-rRNA of different lengths, we conclude from this that the automated tree-base novelty scoring pipeline performs well only for ssu-rRNA sequences that are larger than 1000 bases.

Genome of Blue Mold Causing Fungus *Penicillium solitum* Compared to Related Species to Reveal Genetic Features Likely Involved in Virulence

Wu, Guangxi* (guangxiwu@gmail.com)[1]; Yin, Guohua[2]; Yu, Jiujiang[3]; Jurick II, Wayne M.[3]; Bennett, Joan W.[2]; Shelton, Daniel R.[3]

[1]Oak Ridge Institute for Science and Education. [2]Department of Plant Biology and Pathology, Rutgers University, New Brunswick, NJ 08901. [3]USDA/ARS, Beltsville Agricultural Research Center, Beltsville MD 20705.

Postharvest blue mold decay of apples and pears is caused by Penicillium species in the United States and in many countries worldwide. This genus is responsible for severe economic losses and produces an array of mycotoxins that contaminate processed apple products. Among the species that cause blue mold, P. expansum is the most prevalent, causing more than 50% of the postharvest decay during storage. P. solitum is less virulent compared with P. expansum but also causes decay. To analyze the genetic mechanisms contributing to pathogen virulence, spore germination, and mycotoxin production, the whole genome of P. solitum (RS1 isolate) was sequenced using PacBio. The raw sequences were assembled into seven scaffolds with an N50 of 7.7Mbp. The genome size is estimated to be 35.8 Mbp and contains 10,672 coding genes, which is similar to what is previously reported for P. expansum and P. chrysogenum. Further, we compared the P. solitum genome to a panel of six related Penicillium species, including P. expansum, another pome fruit pathogen, and two other plant pathogens. We revealed their phylogeny using a multi-gene approach, and then profiled gene sets unique to either of the two or both of the pome fruit pathogens. Aided with functional analysis, we identify genes that are putatively involved in virulence. These genes could be further studied experimentally and could be keys to devising novel strategies to control blue mold decay on apples and pears during storage.

RNA Profiling of *Neurospora crassa*: Insights Into Plant Cell Wall Degradation by Filamentous Fungi

Wu, Vincent* (vwu104@berkeley.edu)[1]; Kowbel, David[1]; Xiong, Yi[1, 2]; Lipzen, Anna[2]; Singan, Vasanth[2]; Grigoriev, Igor V.[1, 2]; Glass, Louise N.[1]

[1]UC Berkeley, Berkeley, CA. [2]JGI, Walnut Creek, CA.

Conversion of cellulosic plant biomass to biofuels holds great potential for alleviating our reliance to fossil fuels. A major goal of this research is to develop fungi to be suitable for secreting large volumes of plant cell wall degrading enzymes (CWDEs). In order to productively direct our efforts, a solid understanding of how these fungi sense and respond to their nutritional environment must first be established. Using *Neurospora crassa* as a model, we have started to characterize how filamentous fungi respond to different carbon, nitrogen, sulfur and phosphate sources. To do this we are comparing the transcriptional profiles of *N. crassa* induced by a variety of these nutritional conditions. This data has not only shed light on the role of previously identified genes, it has uncovered novel candidates that we hypothesize are important *N. crassa* metabolism. Here we focus on the novel annotation transcription factors orchestrating the huge transcripitional response to presences or absences of nutrients, as well as transporters, which are the front line of nutrient detection and response.

Integrating Metagenomics and Metatranscriptomics to Analyze the Impact of Ionic Liquids on the Bioenergy Feedstock Degrading Microbiome and Transcription of Enzymes Relevant to Polysaccharide Hydrolysis

Wu, Yu-Wei* (ywwei@lbl.gov)[1, 4]; Higgins, Brendan[2]; Yu, Chaowei[2]; Reddy, Amitha P.[2]; Ceballos, Shannon[2;] Joh, Larry D.[2]; Simmons, Blake A.[1, 3]; Singer, Steven W.[1, 4]; VanderGheynst, Jean S.[1, 2]

[1]Joint BioEnergy Institute, Emeryville, CA 94702, USA. [2]Biological and Agricultural Engineering, University of California, Davis, CA 95616. [3]Biological and Materials Science Center, Sandia National Laboratories, Livermore, CA, 94551. [4]Biological Systems and Engineering, Lawrence Berkeley National Laboratory, CA, USA.

Pretreatment using ionic liquids (IL) is a promising approach for the conversion of lignocellulose to biofuels. One challenge facing this approach is that residual IL can be inhibitory to industrial enzymes and microorganisms involved in hydrolysis and fermentation. In this study a thermophilic microbial community was cultured on switchgrass amended with varying levels of the ionic liquid 1-ethyl-3-methylimidazolium acetate ([C2mim][OAc]). Changes in the microbial community and transcription of genes relevant to IL tolerance and lignocellulose hydrolysis were measured through metagenomics and metatranscriptomics. Increasing the level of [C2mim][OAc] up to 1% resulted in an increase on the relative abundances and transcription activities of organisms from the phylum Firmicutes. Interestingly, moderate IL (up to 1%) resulted in increased transcription of xylanase, endoglucanase, beta-glucosidase, and IL tolerance genes compared to communities without IL. [C2mim][OAc] levels above 1% resulted in decreased microbial activity, enzyme activity, and transcription of genes involved in lignocellulose hydrolysis. The results indicate that moderate levels of [C2mim][OAc] select for thermophilic microorganisms that not only tolerate IL but effectively hydrolyze lignocellulose in switchgrass. Discovery of IL tolerant organisms and enzymes is critical for the development of biological processes that convert IL-pretreated biomass to biofuels and chemicals. Our approach of employing metagenomics and metatranscriptomics in the analysis of IL-enriched cultures can facilitate discovery of microorganisms and enzymes that are active in the presence of IL.

Towards a Mobile Ecogenomic Sensor: Development and Applications of the 3rd Generation Environmental Sample Processor

Yamahara, Kevan M.* (kyamahara@mbari.org)[1]; Preston, Christina M.[1]; Pargett, Doug[1]; Jensen, Scott[1]; Roman, Brent[1]; Romano, Anna[2]; Cardwell, John[2]; Birch, James[1]; Delong, Edward[2]; Scholin, Christopher[1]

[1]Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA. [2]University of Hawaii, Honolulu, HI, USA.

The development of new high-throughput sequencing methods over the last decade has revolutionized the study of microorganisms. These advancements have profound implications for use in environmental research, resource management, agriculture, product stream quality assurance, and public health. However, one of the long-standing challenges common to all of these applications is acquiring and handling samples prior to the application of these new analytical methods. We demonstrate how the new Environmental Sample Processor (ESP) can address the challenge of sample collection and processing over an extended period under non-traditional laboratory conditions, and preview its capability to support a variety of downstream molecular detection systems. The ESP has been redesigned to be smaller and less complex relative to its predecessor in order to enable deployment on mobile platforms such as a long-range autonomous underwater vehicle (LRAUVs), profiling floats, or serve as a hand-carried instrument. Here, we present an overview of the redesigned ESP that utilizes a novel cartridge-based system for sample collection and processing, including preservation for subsequent analysis in the laboratory, or on-board nucleic acid/protein extractions for in situ, real-time, molecular analyses. Integration of the new ESP with the LRAUV has created a novel capability to find and sample particular features of interest, allowing for “smart” in-situ sampling and analytical operations over large geographical areas absent human intervention. Real-time data access, and Internet connectivity also allows the instrument to be redirected according to changing environmental conditions or sampling priorities. Preliminary results from the new ESP prototype reveal that samples collected and preserved in situ can be utilized for a variety of downstream molecular analytical methods, such as qPCR, tag-sequencing, and metatranscriptomics. We show that the abundance and diversity of microbial genes and transcripts from samples collected and preserved using the new ESP are statistically similar to those processed using accepted laboratory procedures.

Genome-Scale Metabolic Modeling with the PSAMM Software

Zhang, Ying* (yingzhang@uri.edu); Steffensen, Jon Lund; Dufault-Thompson, Keith

Department of Cell and Molecular Biology, College of the Environment and Life Sciences, University of Rhode Island.

The broad application of genome-scale metabolic modeling has made it a useful technique for tackling fundamental questions in biological research and engineering. Today over 100 models have been constructed for organisms that carry out a diverse array of metabolic activities spanning all three kingdoms of life. These models, however, have been curated independently following different conventions. The maintenance of model consistency has been challenging due to the lack of consensus in model representation and the absence of integrated modeling software for associating mathematical simulations with the annotation and biological interpretation of metabolic models. To solve this

problem, we developed a new software package, PSAMM, and a new model format that incorporates heterogeneous, model-specific annotation information into modular representations of model definitions and simulation settings. PSAMM provides significant advances in standardizing the workflow of model annotation and consistency checking. Compared to existing tools, PSAMM supports more flexible configurations and is more efficient in running constraint-based simulations. All functions of PSAMM are freely available for academic users and can be downloaded from a public Git repository at https://zhanglab.github.io/psamm/ under the GNU General Public License.

In this presentation, we will introduce a new metabolic modeling software named PSAMM. The software can be accessed at https://zhanglab.github.io/psamm/.

Functional Selection of Microbial Genes in the *Arabidopsis Thaliana* Rhizosphere

Yourstone, Scott* (105 Lone Tree Ct. Durham, NC 27713)[1]; Herrera Paredes, Sur[1]; Cameron, Hunter[1]; Finkel, Omri M.[1]; Gonzalez, Isai Salas[1]; Lundberg, Derek[2]; Lebeis, Sarah[3]; Clingenpeel, Scott[4]; Levy, Asaf[5]; Woyke, Tanja[5]; Tringe, Susannah G.[5]; Jones, Corbin D.[1]; Dangl, Jeff[1,6]

[1]University of North Carolina, Chapel Hill, NC. [2]Max Planck Institute for Developmental Biology, Tubingen, Germany. [3]University of Tennessee, Knoxville, TN. [4]University of Utah, Salt Lake City, UT. [5]US Department of Energy Joint Genome Institute, Walnut Creek, CA. [6]HHMI-GBMF Plant Science Investigator.

Bacterial communities living in close association with eukaryotes are important for maintaining the overall well being of their host by providing resistance to biotic and abiotic stress along with a variety of other beneficial phenotypes. However, observing the function characteristics of metagenomes associated with these beneficial phenotypes is a challenge that requires massive sequencing efforts and advanced bioinformatics. *De novo* assembly of DNA from complex microbial communities is marginally useful as only a small fraction of reads assemble into contigs. An alternative to *de novo* assembly involves mapping reads to a reference sequence. In metagenomics, mapping reads to a reference is also a challenge for two reasons: 1) enumerating reference genomes for each unique community member of a complex microbiota is beyond our current capabilities and 2) it is unlikely that a substantial number of the metagenome reads will map to any of the reference genomes at a species level of specificity. Despite these challenges, it is possible to sequence a variety of isolates sampled from a given community, and by relaxing read mapping parameters each isolate can be used to capture reads in a taxa (e.g. family) specific manner. Here we show that mapping metagenomic reads to a database of hundreds of isolate genomes using low specificity mapping parameters allows for inference about family level selection of functional characteristics. We demonstrate this analysis on data generated from wild soil and rhizosphere samples, however this method can be applied to any comparative metagenomic dataset. We find many functional groups (i.e. COGs) are differentially abundant between rhizosphere and bulk soil metagenomes at a taxa specific level. Interestingly, taxa with a large number of enriched functional groups in rhizosphere samples are also more likely to have many depleted functional groups. Along with delineating general functional patterns, this method can be used to investigate specific functional elements (e.g. COGs), genes, or alleles associated with differential abundance. To illustrate, we show differential abundance patterns for rimK, a gene known to be important for mono-association of Pseudomonads in *Arabidopsis* roots. In conclusion, this type of metagenome analysis can be applied to any comparative metagenomics project to uncover high or low level functional patterns associated with colonizing a microbiome.

Identification of Missing Enzymes and Transporters Involved in the Synthesis and Salvage of Queuosine by Comparative Genomics

Zallot, Rémi* (remizallot@ufl.edu)[1]; de Crécy-Lagard, Valérie (vcrecy@ufl.edu)[1]; Ross, Robert[2]; Chen Wei-Hung[3]; Bruner, Steven D.[3]; Limbach, Patrick A.[2]

[1]Institute of Food and Agricultural Sciences, Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32601, USA. [2]Rieveschl Laboratories for Mass Spectrometry, Department of Chemistry, University of Cincinnati, Cincinnati, OH 45221. [3]Department of Chemistry, University of Florida, Gainesville, FL 32601, USA.

Queuosine (Q) is a modification of the wobble position of tRNAs with GUN anticodons. Q is involved in translational accuracy, modulating misreading by tRNAs. Q is present in most eukaryotes and bacteria but only some bacteria can synthesize Q *de novo*. The Q synthesis pathway is well characterized in Escherichia coli. Phylogenetic profiling and examination of physical gene clustering encoding Q biosynthesis enzymes allowed the identification of an alternate final step and found in nearly 25% of the bacteria predicted to harbor Q. This prediction is being validated *in vivo* by complementation of an Escherichia coli queG mutant strain with genes from this alternate family.

Several bacteria only have genes encoding for the final steps in the Q synthesis pathway implying Q precursor salvage. For salvage to occur, specific transporters are needed and these have yet to be identified. Candidate genes were identified based on a combination of bioinformatics evidence. These were then tested experimentally by following the rate of Q incorporation into tRNA in a double mutant of E. coli impaired in Q synthesis and for the candidate transporters.

This work demonstrates the power of using comparative genomics approaches for the identification of missing transporters and enzymes.

Recovery and Expression of Intact Secondary Metabolite Biosynthetic Pathways from a Large-Insert Soil Metagenomic Library

Zhou, Jinglie* (jzz0026@auburn.edu)[2]; Monsma, Scott[1]; Pereira, Alinne[2]; Pfeifer, Blaine[3]; Santos, Scott R.[2]; Niebauer, Megan[1]; Ferguson, Erin[1]; Godiska, Ron[1]; Wu, Chengcang[1]; Mead, David[1]; Liles, Mark[2]

[1]Lucigen Corporation, 2905 Parmenter St., Middleton WI, USA. [2]Department of Biological Sciences, Room 101 Rouse Life Sciences Building, 120 West Samford Avenue, Auburn University, Auburn AL, 36849 USA. [3]Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260-4200, USA.

Soil microorganisms contain vast reservoirs of bioactive natural products; however, the majority of them are recalcitrant to cultivation in the lab. In this study a large-insert soil metagenomic clone library (~110kb and 19,200 clones) was constructed from an agricultural soil (Cullars Rotation, Auburn, AL) using a broad host range shuttle BAC vector, pSmartBAC-S. This insert size is capable of harboring many intact secondary metabolite biosynthetic pathways such as Type I PKS pathways (usually > 40 kb). Identification of secondary metabolite biosynthesis clusters was conducted using multiple methods, including DNA hybridization, PCR and next-generation sequencing. In the first two methods we targeted a conserved domain of Type

I polyketide synthases (PKS) and identified clones by macroarray hybridization or PCR, resulting in 6 and 110 pathway-containing clones, respectively. In addition, we used a strategy in which plates, rows and columns were separately pooled, and bar-coded DNA sequences from each pool were subjected to Illumina HiSeq sequencing. Contigs were assembled from each pool and screened for secondary metabolite gene clusters using antiSMASH3.0, resulting in identification of 884 clones that contained a PKS and/or non-ribosomal peptide synthetase pathway, among 1,910 total biosynthetic pathways identified. The cloned pathways are very divergent from known pathways, with the %G+C content varying from 34 to 79% and the nearest BLAST hit of keto-synthase domains ranging from 19 to 95% amino acid identity. Biosynthetic clusters identified via PCR were a subset of the clones identified via next-gen sequencing, which were both numerically more abundant and represent novel pathways highly divergent from known pathways. 96 identified pathway-containing BAC clones with limited homology to known PKS pathways were transformed into an E. coli BTRA strain engineered for PKS expression. BAC clones expressed in BTRA were screened for the synthesis of antibacterial compounds by various bioassay against the pathogens Methicillin-resistant Staphylococcus aureus (MRSA), Pseudomonas aeruginosa, Acinetobacter baumannii, Klebsiella pneumonia and Cryptococcus neoformans. Clones expressing antimicrobial activity were further characterized by LC/MS analysis. These results indicate that highly novel biosynthetic clusters can be cloned intact from complex metagenomes and heterologously expressed to produce secondary metabolites, thereby expanding our available resources for natural product discovery.

Improved MicroRNA Library Preparation Workflow for Next-Generation Sequencing Allows Ultra-Low Inputs and Eliminates Gel Purification

Zon, Gerald* (gzon@trilinkbiotech.com); Shore, Sabrina; Henderson, Jordana M.; McCaffrey, Anton P.; Hogrefe, Richard I.

TriLink BioTechnologies, San Diego, CA/USA.

MicroRNAs (miRNAs) are small non-coding RNA molecules (~22 nt) that play important roles in cellular proliferation, differentiation, apoptosis, and development by simultaneously controlling the expression level of hundreds of genes. Numerous recent studies have shown that miRNA expression profiles differ between normal tissues and cancerous cells derived from the same organ, between cancer types, and can act as oncogenes or tumor suppressors, contributing to different pathways in tumorigenesis. They may be used for diagnostic and prognostic purposes and they also constitute novel targets for cancer treatment. The overwhelming majority of microRNAs deposited in miRBase, which are ~3,000 in number that grows exponentially, have been predicted from small RNA deep-sequencing using various “next-generation sequencing” (NGS) platforms and sample preparation methods. However, commercially available sample preparation kits typically require relatively large inputs of RNA (~1 ug) and a laborious gel purification step, which is not amenable to automation. Performance of these kits are limited by side reactions forming “adapter dimers” by ligation of 5’ and 3’ adapters without an intervening RNA insert, which is exacerbated at low RNA inputs. Herein we describe use of an improved workflow enabled by novel, chemically modified 5’ and 3’ adapters that greatly suppress adapter dimer formation. This new technology preferentially blocks adapter dimer formation during the ligation steps for miRNA, and other types of endogenous small RNA species, during NGS library preparation. Heretofore unrealized advantages provided by this unique approach include (1.) increase in mapped sequence reads, (2.) ability to handle ultra-low RNA inputs (≤ 1 ng), which is (3.) especially relevant to clinical samples, such as plasma, urine or FFPE, (4.) allows bead-based isolation, thus eliminating the need for gel purification (5.) making it possible to achieve full automation.

Index

C

D

E

F

G

H

K

L

M

N

S

T

U

V

W

X

Y

Z

Attendees

Catherine Adam
DOE Joint Genome Institute
Cadam@lbl.gov

Steven Ahrendt
DOE Joint Genome
Institutesahrendt0@gmail.com

Cathie Aime
Purdue University
maime@purdue.edu

Ed Allen
DOE Joint Genome Institute
eaallen@lbl.gov

Benjamin Allen
Oak Ridge National Laboratory
allenbh@ornl.gov

Karthik Anantharaman
University of California, Berkeley
karthik.a@berkeley.edu

Peter Andeer
Lawrence Berkeley National
Laboratorypfandeer@lbl.gov

Lisa Anderson
MIT
laanders@mit.edu

Bill Andreopoulos
DOE Joint Genome Institute
wandreopoulos@lbl.gov

Adam Arkin
Lawrence Berkeley National Laboratory
aparkin@lbl.gov

Jaderson Armanhi
State University of Campinas
jader.armanhi@gmail.com

Aude Aznar
Joint BioEnergy Institute
aaznar@lbl.gov

Orli Bahcall
Nature
o.bahcall@us.nature.com

Massie Ballon
DOE Joint Genome Institute
mlballon@lbl.gov

Kerrie Barry
DOE Joint Genome Institute
kwbarry@lbl.gov

Michael Barton
DOE Joint Genome Institute
mbarton@lbl.gov

Diane Bauer
DOE Joint Genome Institute
dmbauer@lbl.gov

Alexander Beliaev
Pacific Northwest National Laboratory
alex.beliaev@pnnl.gov

Michael Beman
University of California, Merced
mbeman@ucmerced.edu

Paul Blainey
Broad Institute of MIT and Harvard
pblainey@broadinstitute.org

Jeffrey Blanchard
University of Massachusetts
jeffb@bio.umass.edu

Matthew Blow
DOE Joint Genome Institute
mjblow@lbl.gov

Katie Bockrath
US Fish and Wildlife Service
katherine_bockrath@fws.gov

Harvey Bolton
Pacific Northwest National Laboratory
harvey.bolton@pnnl.gov

Ben Bowen
Lawrence Berkeley National Laboratory
bpbowen@lbl.gov

Robert Bowers
DOE Joint Genome Instituter
mbowers@lbl.gov

Chris Bowler
Ecole Normale Superieure
cbowler@biologie.ens.fr

Jim Bristow
Lawrence Berkeley National Laboratory
jbristow@lbl.gov

Luciano Brocchieri
University of Florida
lucianob@ufl.edu

Eoin Brodie
Lawrence Berkeley National Laboratory
elbrodie@lbl.gov

Natasha Brown
DOE Joint Genome Institute
nzvenigorodsky@lbl.gov

Steven Brown
Oak Ridge National Laboratory
brownsd@ornl.gov

Dan Browne
Texas A&M University
dbrowne.up@gmail.com

David Bruce
LANL
dbruce@lanl.gov

Jennifer Brum
Ohio State University
brum.11@osu.edu

James Burrows
Pacific Biosciences
jburrows@pacificbiosciences.com

Cristina Butterfield
University of California, Berkeley
butterfc@berkeley.edu

Barbara Campbell
Clemson University
bcampb7@clemson.edu

Manu Capoor
Rockefeller University
mcapoor@rockefeller.edu

Emily Cardarelli
Stanford University
ecardare@stanford.edu

Zoe Cardon
Marine Biological Laboratory
cardon@mbl.edu

Joe Carlson
DOE Joint Genome Institute
JWCarlson@lbl.gov

Amy Cartwright
DOE Joint Genome Institute
acartwright@lbl.gov

Madeli Castruita
UCLA
madeli.c@gmail.com

David Catcheside
Flinders University
David.Catcheside@flinmders.edu.au

David Cavalier
GLBTC
cavalie8@msu.edu

Meghan Chafee
Max Planck Institute for Marine Microbiology
mchafee@mpi-bremen.de

Romy Chakraborty
Lawrence Berkeley National Laboratory
rchakraborty@lbl.gov

Cindy Chen
DOE Joint Genome Institute
cindychen@lbl.gov

Jay Chen
Oak Ridge National Laboratory
chenj@ornl.gov

Jianwei Chen
University of Calgary
jchen@ucalgary.ca

Jan-Fang Cheng
DOE Joint Genome Institute
jfcheng@lbl.gov

Dawn Chiniquy
DOE Joint Genome Institute
dmchiniquy@lbl.gov

Jennifer Chiniquy
DOE Joint Genome Institute
JLChiniquy@lbl.gov

Mila Chistoserdova
University of Washington
milachis@uw.edu

Charles Chiu
UCSF
charles.chiu@ucsf.edu

Dylan Chivian
Lawrence Berkeley National Laboratory
DCChivian@lbl.gov

Hee jung Cho
University of California, Berkeley
beeologai@berkeley.edu

Mansi Chovatia
DOE Joint Genome Institute
mrchovatia@lbl.gov

Doina Ciobanu
DOE Joint Genome Institute
dgciobanu@lbl.gov

Cameron Coates
DOE Joint Genome Institute
rccoates@lbl.gov

Benjamin Cole
DOE Joint Genome Institute
bjcole@lbl.gov

Maureen Coleman
University of Chicago
mlcoleman@uchicago.edu

David Collins
San Diego State University
collins.david.andrew@gmail.com

Jason Corwin
University of North Carolina at Chapel Hill
jacorwin@email.unc.edu

Kyle Costa
California Institute of Technology
costak@caltech.edu

Robert Cottingham
Oak Ridge National Laboratory
cottinghamrw@ornl.gov

Ivana Crnovcic
The Scripps Research Institute (TSRI)
icrnovci@scripps.edu

Byron Crump
Oregon State University
bcrump@coas.oregonstate.edu

Chris Daum
DOE Joint Genome Institute
cgdaum@lbl.gov

Patrick Davidson
Phytozome Group, Joint Genome Institute
padavidson@lbl.gov

Ronald De Vries
CBS-KNAW Fungal Biodiversity Centre
r.devries@cbs.knaw.nl

Shweta Deshpande
DOE Joint Genome Institute
Sdeshpande@lbl.gov

Sam Deutsch
DOE Joint Genome Institute
sdeutsch@lbl.gov

Manpreet Dhami
Stanford University
mdhami@stanford.edu

Spencer Diamond
University of California, Berkeley
sdiamond@berkeley.edu

Stephen DiFazio
West Virginia University
spdifazio@mail.wvu.edu

David Dilworth
DOE Joint Genome Institute
djdilworth@jgi.gov

Jose Dinneny
Carnegie Institution for Science
jdinneny@carnegiescience.edu

Rhyan Dockter
DOE Joint Genome Institute
rbdockter@lbl.gov

Mitchel Doktycz
Oak Ridge National Laboratory
doktyczmj@ornl.gov

Veronika Dollhofer
Bavarian State Research Center for Agriculture
veronika.dollhofer@lfl.bayern.de

Xiaoli Dong
University of Calgary
xdong@ucalgary.ca

Timothy Donohue
UW Madison- Great Lakes Bioenergy Research Center
tdonohue@bact.wisc.edu

Kevin Dorn
Kansas State University
dorn@ksu.edu

Sharon Doty
University of Washington
sldoty@uw.edu

Devin Doud
DOE Joint Genome Institute
dfdoud@lbl.gov

Meghan Drake
Oak Ridge National Laboratory
drakemm@ornl.gov

Daniel Drell
U.S. Department of Energy
daniel.drell@science.doe.gov

Tommy Duong
Clontech Lab
tommy_duong@clontech.com

Dan Eastwood
Swansea University
d.c.eastwood@swansea.ac.uk

Daniel Ebbole
Texas A&M University
d-ebbole@tamu.edu

Rob Egan
DOE Joint Genome Institute
RSEgan@lbl.gov

Louise Egerton-Warburton
Chicago Botanic Garden
lwarburton@chicagobotanic.org

Robert Evans
DOE Joint Genome Institute
rse@lbl.gov

Kjiersten Fagnan
NERSC
kmfagnan@lbl.gov

Jesus Fernandez Bayo
University of California Davis
jdfbayo@ucdavis.edu

Omri Finkel
UNC Chapel Hill
ofinkel@live.unc.edu

Daniel Fishman
U.S. Department of Energy, EE-3B
Daniel.Fishman@ee.doe.gov

Mark Forrer
Joint BioEnergy Institute
mark.forrer@lbl.gov

Bryce Foster
DOE Joint Genome Institute
brycefoster@lbl.gov

Christopher Francis
Stanford University
caf@stanford.edu

Philip Gach
Joint BioEnergy Institute
pcgach@lbl.gov

Difeng Gao
Clemson University
difengg@clemson.edu

Haifeng Geng
Sandia National Laboratory
hngeng@sandia.gov

Jaimie Gibbons
South Dakota State University
jaimie.gibbons@jacks.sdstate.edu

David Gilbert
DOE Joint Genome Institute
degilbert@lbl.gov

Sean Gilmore
University of California, Santa Barbara
SeanGilmore@umail.ucsb.edu

Randy Girsang
DOE Joint Genome Institute
rgirsang@lbl.gov

Tijana Glavina del Rio
DOE Joint Genome Institute
tglavina@lbl.gov

Barry Goldman
Monsanto
barry.s.goldman@monsanto.com

Eugene Goltsman
DOE Joint Genome Institute
egoltsman@lbl.gov

David Goodstein
DOE Joint Genome Institute
dmgoodstein@lbl.gov

Sean Gordon
DOE Joint Genome Institute
sgordon@lbl.gov

Danielle Goudeau
DOE Joint Genome Institute
dgoudeau@lbl.gov

Danielle Graham
DOE Joint Genome Institute
dgraham@lbl.gov

Elizabeth Green
Lawrence Livermore National Laboratory
green74@llnl.gov

Richard Green
Univ. of California, Santa Cruz
ed@soe.ucsc.edu

Marc Griesemer
Lawrence Livermore National Laboratory
marcgri@gmail.com

Igor Grigoriev
DOE Joint Genome Institute
ivgrigoriev@lbl.gov

Jenny Gu
Pacific Biosciences
jgu@pacb.com

Liping Gu
South Dakota State University
liping.gu@sdstate.edu

Yong Gu
USDA-ARS-WRRC
yong.gu@ars.usda.gov

Lee Gunter
Oak Ridge National Laboratory
gunterle@ornl.gov

Aria Hahn
University of British Columbia
hahnaria@gmail.com

Charles Haitjema
University of California, Santa Barbara
haitjema@engineering.ucsb.edu

Richard Hall
Pacific Biosciences
rhall@pacificbiosciences.com

Trinity Hamilton
University of Cincinnati
trinity.hamilton@uc.edu

Nancy Hammon
DOE Joint Genome Institute
nmhammon@lbl.gov

Sajeet Haridas
DOE Joint Genome Institute
sharidas@lbl.gov

Miranda Harmon-Smith
DOE Joint Genome Institute
mlharmon-smith@lbl.gov

Mohamed FauziHaroon
KAUST
fauziharoon@gmail.com

Nomi Harris
Lawrence Berkeley National Laboratory
nlharris@lbl.gov

Howard Harris
Lawrence Livermore National Laboratory
harris123@llnl.gov

Susanna Harris
The University of North Carolina at Chapel Hill
Susanna_Harris@unc.edu

Steven Harris
University of Nebraska
sharris2@unl.edu

Wyatt Hartman
DOE Joint Genome Institute
whhartman@lbl.gov

Alyse Hawley
University of British Columbia
alysekh@mail.ubc.ca

Richard Hayes
DOE Joint Genome Institute
rdhayes@lbl.gov

Guifen He
DOE Joint Genome Institute
ghe@lbl.gov

Monica Heger
GenomeWeb
mheger@genomeweb.com

Cheryl Heiner
Pacific Biosciences
cheiner@pacificbiosciences.com

John Henske
University of California, Santa Barbara
johnhenske@engineering.ucsb.edu

Sayer Herin
UC Davis, Dept. of Plant Sciences
herin@ucdavis.edu

Sur Herrera Paredes
The University of North Carolina at Chapel Hill
sur@email.unc.edu

Matthias Hess
University of California Davis
mhess@ucdavis.edu

Steven Higgins
University of Tennessee
shiggin8@vols.utk.edu

Ryan Hill
DOE Joint Genome Institute
rhill@lbl.gov

Maureen Hillenmeyer
Stanford Genome Technology Center
maureenh@stanford.edu

Nathan Hillson
DOE Joint Genome Institute
njhillson@lbl.gov

Chris Hittinger
GLBRC & Laboratory of Genetics, UW-Madison
cthittinger@wisc.edu

Joe Ho
University of British Columbia
joeho776@gmail.com

Kirsten Hofmockel
Pacific Northwest National Laboratory
khof@iastate.edu

Cody Holik
Bolt Threads
cholik@boltthreads.com

Cindi Hoover
DOE Joint Genome Institute
cahoover@lbl.gov

Chiaki Hori
Hokkaido University
chori@for.agr.hokudai.ac.jp

David Hoyt
EMSL- Pacific NW National Lab
david.hoyt@pnnl.gov

Szu Yi Hsu
University of Minnesota
hsuxx166@umn.edu

Ping Hu
Lawrence Berkeley National Laboratory
phu@lbl.gov

Sui-Sheng Hua
USDA-ARS-WRRC
sylvia.hua@ars.usda.gov

Martin Huh
USC
huh@usc.edu

Hope Hundley
DOE Joint Genome Institute
HNTice@lbl.gov

Marcel Huntemann
DOE Joint Genome Institute
mhuntemann@lbl.gov

Miriam Hutchinson
University of New Mexico
miramira@unm.edu

Natalia Ivanova
DOE Joint Genome Institute
nnivanova@lbl.gov

Kosuke Iwai
Joint BioEnergy Institute / Sandia National Laboratories
KosukeIwai@lbl.gov

Daniel Jacobson
Oak Ridge National Laboratory
jacobsonda@ornl.gov

Tim James
University of Michigan
tyjames@umich.edu
Jessica Jarett
DOE Joint Genome Institute
jkjarett@lbl.gov

Jaswinder Singh
McGill University
jaswinder.singh@mcgill.ca

Jerry Jenkins
HudsonAlpha Institute
jjenkins@hudsonalpha.org

Chris Jennings
Labcyte Inc
cjennings@labcyte.com

Paul Jensen
Scripps Institution of Oceanography, UCSD
pjensen@ucsd.edu

Tianying Ji
DOE Joint Genome Institute
tianying@lbl.gov

Beibei Jing
Joint Bioenergy Institute
Bjing@lbl.gov

Jenifer Johnson
DOE Joint Genome Institute
jkaplan@lbl.gov

Chijioke Joshua
Joint BioEnergy Institute
cjjoshua@lbl.gov

Thomas Juenger
University of Texas at Austin
tjuenger@austin.utexas.edu

Sean Jungbluth
University of Southern California
jungbluth.sean@gmail.com

Shigeyuki Kakizawa
National Institute of Advanced Industrial Science and Technology (AIST)
s.kakizawa@aist.go.jp

Olga Kamneva
Stanford University
okamneva@stanford.edu

Don Kang
DOE Joint Genome Institute
ddkang@lbl.gov

Ulas Karaoz
Lawrence Berkeley National Laboratory
ukaraoz@lbl.gov

Achim Karger
Thermo Fisher Scientific
achim.karger@thermofisher.com

Andrew Katz
DOE Joint Genome Institute
andrewkatz@lbl.gov
Lisa Kegg

DOE Joint Genome Institute
lrkegg@lbl.gov

Elizabeth Kellogg
Danforth Plant Science Center
ekellogg@danforthcenter.org

Megan Kennedy
DOE Joint Genome Institute
mckennedy@lbl.gov

Hae Jin Kim
Chung-Ang University, Department of Systems Biotechnology
hjinkim327@gmail.com

Joonhoon Kim
Joint BioEnergy Institute
joonhoonkim@lbl.gov

Kristi Kim
Pacific Biosciences
kspittle@pacificbiosciences.com

Jeff Kim
Radiant Genomics
jeff@radiantgenomics.com

Jeff Kimbrel
Lawrence Livermore National Laboratory
kimbrel1@llnl.gov

Linda Kinkel
University of Minnestoa
kinkel@umn.edu

Edward Kirton
DOE Joint Genome Institute
eskirton@lbl.gov

Corby Kistler
USDA ARS
hckist@umn.edu

Manuel Kleiner
University of Calgary
manuel.kleiner@ucalgary.ca

Joey Knelman
DOE Joint Genome Institute
jknelman@lbl.gov

Annegret Kohler
INRA
kohler@nancy.inra.fr

Sebastian Kolinko
Lawrence Berkeley National Laboratory
sdkolinko@lbl.gov

Angela Kollmer
DOE Joint Genome Institute
akollmer@lbl.gov

Sarah Kopac
University of Connecticut
Sarah.kopac@uconn.edu

Suzanne Kosina
Lawrence Berkeley National Laboratory
smkosina@lbl.gov

Faisal Koua
University of Virginia
fk3p@virginia.edu

Stephen Kresovich
Clemson University
skresov@clemson.edu

Dave Kudrna
Ariizona Genomics Inst., University of Arizona
dkudrna@email.arizona.edu

Sunita Kumari
Cold Spring Harbor Labs
kumari@cshl.edu

Govindarajan Kunde Ramamoorthy
DOE Joint Genome Institute
gkramamoorthy@lbl.gov

Victor Kunin
DuPont Pioneer
kunivi@gmail.com

Alan Kuo
DOE Joint Genome Institute
akuo@lbl.gov

Rita Kuo
DOE Joint Genome Institute
rckuo@lbl.gov

Yunyoung Kwak
Kyungpook National University
yun@knu.ac.kr

Nikos Kyrpides
DOE Joint Genome Institute
nckyrpides@lbl.gov

Alex La Reau
UW-Madison
alareau@wisc.edu

Kurt LaButti
DOE Joint Genome Institute
klabutti@lbl.gov

Bridget Ladell
USGS
bladell@usgs.gov

Ellen Lai
University of California, Davis
elai@ucdavis.edu

Kathleen Lail
DOE Joint Genome Institute
klail@lbl.gov

Eric Lam
Rutgers University
ericL89@hotmail.com

Christine Lambert
Pacific Biosciences
clambert@pacb.com

Regina Lamendella
Juniata College
reginalamendella@gmail.com

Miriam Land
Oak Ridge National Laboratory
landml@ornl.gov

Robert Landick
University of Wisconsin-Madison
landick@bact.wisc.edu

Rebecca Lau
Lawrence Berkeley National Laboratory
rlau@lbl.gov

Debbie Laudencia-Chingcuanc
USDA-ARS
debbie.laudencia@ars.usda.gov

Carol Lauzon
California State University East Bay
carol.lauzon@csueastbay.edu

Sarah Lebeis
University of Tennessee
slebeis@utk.edu

Kang Soo Lee
MIT / ETH Zurich
ks_lee@mit.edu

Hayan Lee
DOE Joint Genome Institute
hayan.lee@lbl.gov

Janey Lee
DOE Joint Genome Institute
jlee2@lbl.gov

Ji Lee
DOE Joint Genome Institute
jel@lbl.gov

Jaeheon Lee
FMC Corporation
jaeheon.lee@fmc.com

KyuChan Lee
Systems Microbial Ecology Lab,
Chung-ang University
lkcmsd@gmail.com

Ming Lei
Bolt Threads
mlei@boltthreads.com

Peggy Lemaux
University of California, Berkeley
lemauxpg@berkeley.edu

Asaf Levy
DOE Joint Genome Institute
alevy@lbl.gov

Guotian Li
Joint BioEnergy Institute
gtli@ucdavis.edu

Rachel Li
Joint BioEnergy Institute
rachelli@berkeley.edu

Wei Li
Miami University
liw9@miamioh.edu

Junyan Lin
DOE Joint Genome Institute
junyanlin@lbl.gov

Chan-Shing Lin
National Sun Yat-sen University
shinlin@faculty.nsysu.edu.tw

Anna Lipzen
DOE Joint Genome Institute
alipzen@lbl.gov

Tie Liu
Carnegie Institution for Science/Stanford U
tliu@ciw.edu

Lifeng Liu
DOE Joint Genome Institute
lifeng.liu@lbl.gov

Tong Liu
University of California, Berkeley
tongliu@berkeley.edu

Wen-Tso Liu
University of Illinois at Urbana-Champaign
wentso.liu@gmail.com

Alfredo Lopez De Leon
Novozymes, Inc.
allo@novozymes.com

Dominique Loque
Lawrence Berkeley National Laboratory
dloque@lbl.gov

Katherine Louie
DOE Joint Genome Institute
kblouie@lbl.gov

Connie Lovejoy
Laval University
connie.lovejoy@bio.ulaval.ca

Andrea Lubbe
Lawrence Berkeley National Laboratory
alubbe@lbl.gov

Lauren Lui
Lawrence Berkeley National Laboratory
lmlui@lbl.gov

Qin Ma
South Dakota State University
qin.ma@sdstate.edu

David Macaya-Sanz
West Virginia University
david.macayasanz@mail.wvu.edu

Spencer Macdonald
UBC
spence.macdonald@alumni.ubc.ca

Sally Madden
DOE Joint Genome Institute
smadden@lbl.gov

Ramana Madupu
U.S. Department of Energy
ramana.madupu@science.doe.gov

Jon Magnuson
Pacific Northwest National Laboratory
Jon.Magnuson@pnnl.gov

Rex Malmstrom
DOE Joint Genome Institute
rrmalmstrom@lbl.gov

Nasim Mansoori Zangir
Joint BioEnergy Institute
nmansoorizangir@lbl.gov

Joel Martin
DOE Joint Genome Institute
j_martin@lbl.gov

Gracie Martin
USDA, ARS
gracie.martin@ars.usda.gov

Christopher Mason
Weill Cornell Medicine
chm2042@med.cornell.edu

Xavier Mayali
Lawrence Livermore National Laboratory
mayali1@llnl.gov

Mona Mazaheri
University of Wisconsin
mmazaheri@wisc.edu

Grace McCalla
U.S. Geological Survey
smccalla@usgs.gov

Ryan McClure
Pacific Northwest National Laboratory
ryan.mcclure@pnnl.gov

Ryan McCormick
Texas A&M University
ryanabashbash@tamu.edu

Margaret McFall-Ngai
University of Hawii
mcfallng@hawaii.edu

Ran Mei
University of Illinois at Urbana-Champaign
ranmei2@illinois.edu

John Meng
DOE Joint Genome Institute
xianweimeng@lbl.gov

Xiandong Meng
DOE Joint Genome Institute
xiandongmeng@lbl.gov

Laura Meredith
University of Arizona
laurameredith@email.arizona.edu

Gene Myers
Max-Planck Institute for Molecular Cell Biology
& Genetics
myers@mpi-cbg.de

Joshua Michener
Oak Ridge National Laboratory
michenerjk@ornl.gov

Sirma Mihaltcheva
DOE Joint Genome Institute
smihaltcheva@lbl.gov

Chris Miller
University of Colorado Denver
chris.miller@ucdenver.edu

Jennifer Mobberley
Pacific Northwest National Laboratory
jennifer.mobberley@pnnl.gov

Todd Mockler
Donald Danforth Plant Science Center
tmockler@danforthcenter.org

Stephen Mondo
DOE Joint Genome Institute
sjmondo@lbl.gov

William Moore
Joint BioEnergy Institute
wmoore@lbl.gov

Jenny Mortimer
Joint BioEnergy Institute
jcmortimer@lbl.gov

Supratim Mukherjee
DOE Joint Genome Institute
supratimmukherjee@lbl.gov

Nu-Ri Myeong
Chung-Ang University
myungnuri@gmail.com

Christine Naca
DOE Joint Genome Institute
clnaca@lbl.gov

Sarah Nalven
Oregon State University
sarahnalven@gmail.com

Kishore Nannapaneni
NewLeaf Symbiotics
knannapaneni@newleafsym.com

Nandita Nath
DOE Joint Genome Institute
nnath@lbl.gov

Sangeeta Nath
DOE Joint Genome Institute
snath@lbl.gov

Don Natvig
University of New Mexico
dnatvig@gmail.com

Ali Navid
Lawrence Berkeley National Laboratory
navid1@llnl.gov

Bill Nelson
Pacific Northwest National Laboratory
william.nelson@pnnl.gov

Anthony Neumann
University of Wisconsin-Madison
apneumann2@wisc.edu

Dianne Newman
Caltech/HHMI
dkn@caltech.edu

Vivian Ng
DOE Joint Genome Institute
vng@lbl.gov

Wayne Nicholson
University of Florida
WLN@ufl.edu

Torben Nielsen
DOE Joint Genome Institute
torben@lbl.gov

Tina Nielsen
Great Lakes Bioenergy Research Center
tnielsen@glbrc.wisc.edu

Masaru Nobu
University of Illinois, Urbana-Champaign
masarunobu@gmail.com

Intawat Nookaew
Oak Ridge National Laboratory
nookaewi@ornl.gov

Angela Norbeck
EMSL at the Pacific Northwest National Laboratory
angela.norbeck@pnnl.gov

Trent Northen
DOE Joint Genome Institute
trnorthen@lbl.gov

Daniel Northrup
Booz Allen
daniel.northrup@hq.doe.gov

Jeanette Norton
Utah State University
jeanette.norton@usu.edu

Erin Nuccio
Lawrence Livermore National Laboratory
nuccio1@llnl.gov

Khouanchy Oakgrove
DOE Joint Genome Institute
ksoakgrove@lbl.gov

Ernst Oberortner
DOE Joint Genome Institute
eoberortner@lbl.gov

Yuki Okatsu
University of California, Davis
yokatsu@ucdavis.edu

Jade O'Leary
Cardiff University
olearyJM@cardiff.ac.uk

Jeanine Olsen
Groningen Institute of Evolutionary Life Sciences
j.l.olsen@rug.nl

Michelle O'Malley
UC-Santa Barbara
momalley@engineering.ucsb.edu

Hiroshi Otani
DOE Joint Genome Institute
hotani@lbl.gov

David Paez
DOE Joint Genome Institute
adpaezespino@lbl.gov

Jasmyn Pangilinan
DOE Joint Genome Institute
jlpangilinan@lbl.gov

Kabir Peay
Stanford University
kpeay@stanford.edu

Dale Pelletier
Oak Ridge National Laboratory
pelletierda@ornl.gov

Paul Peluso
Pacific Biosciences
ppeluso@pacificbiosciences.com

Ze Peng
DOE Joint Genome Institute
zpeng@lbl.gov

Xuefeng Peng
University of California, Santa Barbara
xpeng@engineering.ucsb.edu

Christa Pennacchio
DOE Joint Genome Institute
cppennacchio@gmail.com

Len Pennacchio
DOE Joint Genome Institute
lapennacchio@lbl.gov

Pamela Peralta-Yahya
Georgia Tech
pperalta-yahya@chemistry.gatech.edu

Fernando Perez
Lawrence Berkeley National Laboratory
fperez@lbl.gov

Rolando Perez
Stanford University
rcperez@stanford.edu

Rene Perrier
DOE Joint Genome Institute
raperrier@lbl.gov

Jennifer Pett-Ridge
Lawrence Livermore National Laboratory
pettridge2@llnl.gov

Jeremy Phillips
DOE Joint Genome Institute
jlphillips@lbl.gov

Yvette Piceno
Lawrence Berkeley National Laboratory
ympiceno@lbl.gov

Lise Pingault
University of Oklahoma
lise.pingault@ou.edu

Joseph Chris Pires
University of Missouri
piresjc@missouri.edu

Jonathan Plett
Western Sydney University
j.plett@westernsydney.edu.au

Tamsen Polley
University of California, Davis
tmpolley@ucdavis.edu

Joshua Pritchett
DOE Joint Genome Institute
jpritch@lbl.gov

Alexander Probst
University of California, Berkeley
alexander.j.probst@gmail.com

Simon Prochnik
DOE Joint Genome Institute
prochnik@gmail.com

Christine Queitsch
University of Washington
queitsch@u.washington.edu

Betania Quirino
Embrapa-Agroenergy
bfq@uwalumni.com

Srivatsan Raman
GLBRC & UW-Madison
sraman4@wisc.edu

Priya Ranjan
Oak Ridge National Laboratory
ranjanp@ornl.gov

David Rank
Pacific Biosciences
drank@pacificbiosciences.com

Michael Rappe
University of Hawaii
rappe@hawaii.edu

Tatiparthi Reddy
DOE Joint Genome Institute
tbreddy@lbl.gov

Rebecca Reiss
DOE Joint Genome Institute
rreiss@lbl.gov

Sarah Richardson
Joint BioEnergy Institute
SMRichardson@lbl.gov

Robert Riley
DOE Joint Genome Institute
rwriley@lbl.gov

Adam Rivers
DOE Joint Genome Institute/ LBNL
arrivers@lbl.gov

Dan Rokhsar
DOE Joint Genome Institute
dsrokhsar@lbl.gov

Dmitry Rodionov
Sanford Burnham Prebys Medical Discovery Institute
rodionov@sbpdiscovery.org

David Rodriguez
Thermo Fisher Scientific
david.rodriguez2@thermofisher.com

Juan Romero
Thermo Fisher Scientific
juan.romero@thermofisher.com

Jeffrey Ross-Ibarra
University of California Davis
rossibarra@ucdavis.edu

Marie-Noëlle Rosso
INRA- Aix Marseille Université
marie-noelle.rosso@univ-amu.fr

Priya Roy Chowdhury
UMass Amherst
priyanka.roy_chowdhury@okstate.edu

Eddy Rubin
DOE Joint Genome Institute
emrubin@lbl.gov

Angelica Ruiz-Font
National Polythecnic Institute
convenios22@hotmail.com

Noel Ruppert
Labcyte Inc.
nruppert@labcyte.com

Douglas Rusch
Indiana University
doug.rusch@gmail.com

Elizabeth Ryan
DOE Joint Genome Institute
eryan@lbl.gov

Isai Salas Gonzalez
UNC-CH
isai@email.unc.edu

Howard Salis
Pennsylvania State University
salis@psu.edu

Ty Samo
Lawrence Livermore National Laboratory
samo1@llnl.gov

Andrea Sanchez Barrios
University of Kentucky
asanchez-barrios@uky.edu

Laura Sandor
DOE Joint Genome Institute
LCSandor@lbl.gov

Siov Sarreal
USDA
siov.sarreal@ars.usda.gov

Wendy Schackwitz
DOE Joint Genome Institute
wsschackwitz@lbl.gov

Christopher Schadt
Oak Ridge National Laboratory
schadtcw@ornl.gov

Katherine Scheibel
University of California, Berkeley
kscheibel@berkeley.edu

Henrik Scheller
Lawrence Berkeley National Laboratory
hscheller@lbl.gov

Jeremy Schmutz
DOE Joint Genome Institute
jschmutz@hudsonalpha.org

Timo Schuerg
Joint BioEnergy Institute
tschuerg@lbl.gov

Frederik Schulz
DOE Joint Genome Institute
fschulz@lbl.gov

Patrick Schwientek
Bayer
patrick.schwientek@bayer.com

Aditi Sengupta
University of Arizona
asengupta@email.arizona.edu

HoonJe Seong
Chung-Ang University
s4645@naver.com

Rekha Seshadri
DOE Joint Genome Institute
rseshadri@lbl.gov

Volkan Sevim
DOE Joint Genome Institute
vsevim@lbl.gov

Rina Shaikh-Lesko
Freelancer
rina.shaikh.lesko@gmail.com

Aditi Sharma
DOE Joint Genome Institute
aditisharma@lbl.gov

Allison Sharrar
University of California, Berkeley
asharrar@berkeley.edu

Helen Shearman
Biomatters
helen@biomatters.com

Reshma Shetty
Ginkgo Bioworks, Inc.
reshma@ginkgobioworks.com

Donna Shrader
University of Wisconsin-Madison
d.grace.shrader@gmail.com

Shengqiang Shu
DOE Joint Genome Institute
sqshu@lbl.gov

Christine Shulse
DOE Joint Genome Institute
cnshulse@lbl.gov

Christian Sieber
DOE Joint Genome Institute
CSieber@lbl.gov

Orzenil Silva
Brazilian Agricultural Research Corporation
obonfimdasilvajr@lbl.gov

Lisa Simirenko
DOE Joint Genome Institute
lsimirenko@lbl.gov

Rachel Simister
University of British Columbia
rlsimister@gmail.com

Holly Simon
Oregon Health & Science University
simonh@ohsu.edu

Vasanth Singan
DOE Joint Genome Institute
VRSingan@lbl.gov

Esther Singer
DOE Joint Genome Institute
esinger@lbl.gov

Steven Singer
Lawrence Berkeley National Laboratory
SWSinger@lbl.gov

Kanwar Singh
DOE Joint Genome Institute
ksingh@lbl.gov

Joan Slonczewski
Kenyon College
slonczewski@kenyon.edu

Tatyana Smirnova
DOE Joint Genome Institute
tsmirnova@lbl.gov

Hamilton Smith
J. Craig Venter Institute
hsmith@jcvi.org

Kim Smuga-Otto
Freelance journalist
kimsmuga@gmail.com

Zewei Song
University of Minnesota
songx208@umn.edu

Jackson Sorensen
Michigan State University
sorens75@msu.edu

Rafael Souza
State University of Campinas
scs.rafael@gmail.com

Alexander Spunde
DOE Joint Genome Institute
ayspunde@lbl.gov

Avinash Sreedasyam
HudsonAlpha Institute
asreedasyam@hudsonalpha.org

Shawn Starkenburg
Los Alamos National Laboratory
shawns@lanl.gov

Evan Starr
University of California, Berkeley
starr.evan@berkeley.edu

Sarah Stevens
University of Wisconsin – Madison
sstevens2@wisc.edu

Kyra Stillman
DOE Joint Genome Institute
kstillman@lbl.gov

Daniela Strenkert
UCLA
DanielaS@chem.ucla.edu

Rhona Stuart
Lawrence Livermore National Laboratory
stuart25@llnl.gov

Woo-Jun Sul
Chung-Ang University
sulwj@cau.ac.kr

Tarah Sullivan
Washington State University
t.sullivan@wsu.edu

Diane Joyce Sutton
State of California
djsutton@mac.com

Yo Suzuki
J. Craig Venter Institute
ysuzuki@jcvi.org

Tami Swenson
Lawrence Berkeley National Laboratory
tlswenson@lbl.gov

Taichi Takasuka
Hokkaido University
tendo.nd@gmail.com

Jyoti Taneja
University of California, Berkeley
jyotitaneja@berkeley.edu

Angela Tarver
DOE Joint Genome Institute
amtarver@lbl.gov

Neslihan Tas
Lawrence Berkeley National Laboratory
ntas@lbl.gov

Owatha Tatum
DOE Joint Genome Institute
oltatum@lbl.gov

Kate Telma
Bolt Threads
ktelma@boltthreads.com

Michaela TerAvest
Michigan State University
teraves2@msu.edu

Michael Thelen
Lawrence Livermore National Laboratory
mthelen@llnl.gov

James Tiedje
Michigan State University
tiedjej@msu.edu

Christian Tobias
USDA-ARS
christian.tobias@ars.usda.gov

Bradley Tolar
Stanford University
btolar1@stanford.edu

Denis Tolkunov
DOE Joint Genome Institute
dtolkunov@lbl.gov

Andy Tomatsu
DOE Joint Genome Institute
atomatsu@lbl.gov

Kathleen Treseder
University of California Irvine
treseder@uci.edu

Susannah Tringe
DOE Joint Genome Institute
sgtringe@lbl.gov

Andrew Tritt
DOE Joint Genome Institute
ajtritt@lbl.gov

Stephan Trong
DOE Joint Genome Institute
strong@lbl.gov

Sandra Truong
Texas A&M University
thkhavi@tamu.edu

Adrian Tsang
Concordia University
adrian.tsang@concordia.ca

Elizabeth Tseng
Pacific Biosciences
etseng@pacb.com

Ray Turner
DOE Joint Genome Institute
crturner@lbl.gov

Maren Tuttle-Lau
USFWS
maren_tuttle-lau@fws.gov

Daniel Udwary
NERSC
dwudwary@lbl.gov

Sarah Unruh
University of Missouri-Columbia
sauvd3@mail.missouri.edu

Brieanne Vaillancourt
Michigan State University
vaillan6@msu.edu

Jacob Valenzuela
Institute for Systems Biology
jvalenzu@systemsbiology.org

Daniel Van der Lelie
FMC Corporation
daniel.vanderlelie@fmc.com

Neha Varghese
DOE Joint Genome Institute
njvarghese@lbl.gov

Rytas Vilgalys
Duke University
fungi@duke.edu

Axel Visel
DOE Joint Genome Institute
avisel@lbl.gov

John Vogel
DOE Joint Genome Institute
jpvogel@lbl.gov

Jana Voriskova
Lawrence Berkeley National Laboratory
jvoriskova@lbl.gov

Mary Voytek
NASA
mary.voytek-1@nasa.gov

Gaoyan (Natalie) Wang
DOE joint Genome Institute
gaoyanwang@lbl.gov

Mei Wang
DOE Joint Genome Institute
mwang@lbl.gov

Zhong Wang
DOE Joint Genome Institute
zhongwang@lbl.gov

Denis Warshan
Stockholm University DEEP
denis.warshan@su.se

Eric Webb
USC
eawebb@usc.edu

Peter Weber
Lawrence Livermore National Laboratory
weber21@llnl.gov

Eric Wei
Stanford University
ewei2@stanford.edu

William Whitman
University of Georgia
whitman@uga.edu

Heather Wilkinson
Texas A&M University
h-wilkinson@tamu.edu

Ryan Williams
Monsanto
ryan.j.williams@monsanto.com

Tanja Woyke
DOE Joint Genome Institute
twoyke@lbl.gov

Justin Wright
Juniata College
justinwright523@gmail.com

Kelly Wrighton
The Ohio State University
kwrighton@gmail.com

Chengcang Charles Wu
Intact Genomics, Inc.
cwu@intactgenomics.com

Yu-Wei Wu
Joint BioEnergy Institute / Lawrence Berkeley National Laboratory
ywwei@lbl.gov

Guangxi Wu
ORISE
guangxiwu@gmail.com

Dongying Wu
University of California Davis
dygwu@ucdavis.edu

Vincent Wu
University of California, Berkeley
vwu104@berkeley.edu

Yi Xiong
DOE Joint Genome Institute
yxiong@lbl.gov

Ling Xu
University of California, Berkeley
xuling@berkeley.edu

Liangjiao Xue
University of Georgia
lxue@uga.edu

Kevan Yamahara
Monterey Bay Aquarium Reaearch Institute
kyamahara@mbari.org

Juying Yan
DOE Joint Genome Institute
juyingyan@lbl.gov

Mi Yan
DOE Joint Genome Institute
miyan@lbl.gov

Shijie Yao
DOE Joint Genome Institute
syao@lbl.gov

Yasuo Yoshikuni
DOE Joint Genome Institute
yyoshikuni@lbl.gov

Yuko Yoshinaga
DOE Joint Genome Institute
yyoshinaga@lbl.gov

Scott Yourstone
UNC -Chapel Hill
scott.yourstone81@gmail.com

Eric Yu
Bolt Threads
eyu@boltthreads.com

Remi Zallot
University of Florida
remizallot@ufl.edu

Matthew Zane
DOE Joint Genome Institute
mczane@lbl.gov

Jinxiang Zhang
Joint BioEnergy Institute
jinxiangzhang@lbl.gov

Ying Zhang
University of Rhode Island
yingzhang@uri.edu

Jean Zhao
DOE Joint Genome Institute
zyzhao@lbl.gov

Jinglie Zhou
Auburn University
jzz0026@auburn.edu

Ruanbao Zhou
South Dakota State University
ruanbao.zhou@sdstate.edu

Naupaka Zimmerman
University of Arizona
naupaka@gmail.com

Gearld Zon
Trilink Biotechnologies
gzon@trilinkbiotech.com

16-JGI-2307

JGI JOINT GENOME INSTITUTE
A DOE OFFICE OF SCIENCE USER FACILITY

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY
Office of Science