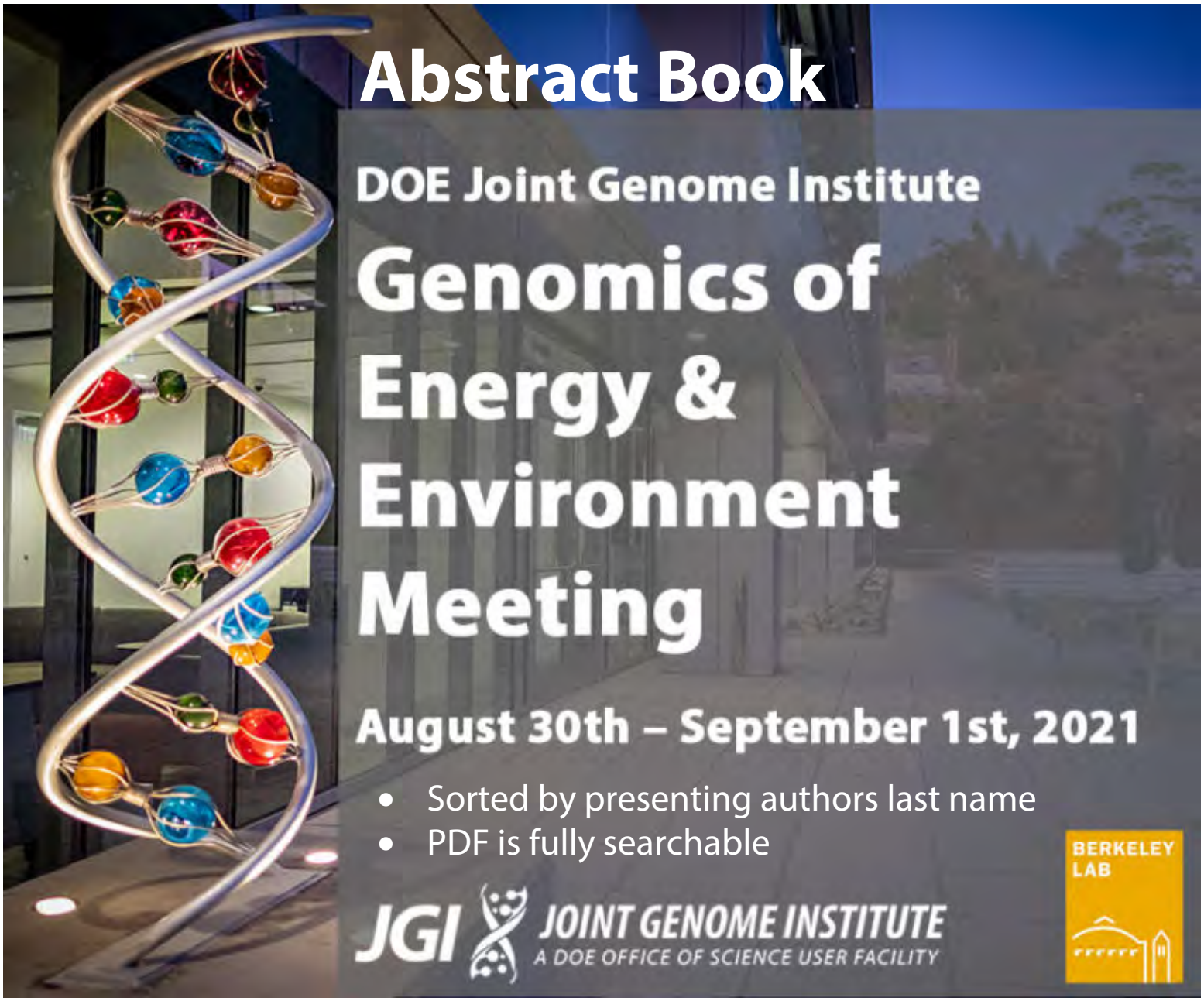


Abstract Book

DOE Joint Genome Institute

Genomics of Energy & Environment Meeting

August 30th – September 1st, 2021

- Sorted by presenting authors last name
- PDF is fully searchable

JGI JOINT GENOME INSTITUTE
A DOE OFFICE OF SCIENCE USER FACILITY

BERKELEY LAB

Title: Genome resolved metagenomic study of a membrane bioreactor treating produced water

Authors: Shwetha M Acharya[1], Brett van Houghton[2], James Rosenblum[2], Tzahi Y. Cath[2], Romy Chakraborty[1] and Susannah Tringe[1]

[1] Lawrence Berkeley National Laboratory, Berkeley, CA
[2] Colorado School of Mines, Golden, CO

Wastewater generated during production of oil and gas (O&G), commonly termed produced water (PW), contains a wide variety of contaminants such as hydrocarbons, heavy metals, and salts. Organic compounds can be effectively removed from PW through biological treatment, however little is known about the microbial community composition of a well-performing bioreactor or population dynamics during reactor operation.

To address this knowledge gap, we have operated and monitored a PW-fed membrane bioreactor (MBR) for over a year. This aerated, 70L reactor was inoculated with municipal wastewater sludge and incrementally fed with PW of increasing salinity ranging from 30 - 100 g/L. Microbial communities of the sludge and the feed were profiled via 16S rRNA gene amplicon sequencing and shotgun metagenomic sequencing. 16S rRNA amplicon reads were analyzed using DADA2/QIIME2 pipeline; shotgun metagenome reads were filtered, assembled and binned through JGI automated workflow.

The MBR operation remained stable, removing >80% of the influent dissolved organic carbon at all ranges of salinity tested. Analysis of 16S rRNA amplicon data revealed low microbial diversity in the sludge compared to the PW feed, primarily consisting of a core sludge microbiome encompassing 30 amplicon sequence variants (ASVs). Among these, 4 dominant ASVs belonging to the genera *Roseovarius, Iodidimonas, Methylophaga* and *Rehaibacterium* constituted greater than 50% of the sequence reads at all timepoints. In order to further understand how these ASVs evolve over time, MAGs generated from the shotgun sequencing will be analyzed to identify in-strain variants that may develop over time. In parallel, functional annotation of the MAGs will be performed to gain deeper insight into major pathways involved in hydrocarbon degradation, osmoprotection, and heavy metal resistance in the reactor. The results generated will help inform strategies to improve stability of biological treatment systems and thus increase their reliability in treating PW.

Fungal mitochondrial genome annotation - insights and challenges

Steven R. Ahrendt[1], Sajeet Haridas[1], Asaf Salamov[1], Stephen J. Mondo[1], Igor V. Grigoriev[1,2]

[1]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[2]Department of Plant and Microbial Biology, University of California-Berkeley, Berkeley, CA, USA

Mitochondria are specialized organelles found within cells of nearly all Eukaryotes and are responsible for ATP generation through oxidative phosphorylation. It is generally accepted that mitochondria are derived from an ancient endosymbiotic event, and therefore possess their own genomes (mitogenomes). Fungal mitogenomes retain a set of genes related to electron transport and oxidative phosphorylation, translation, and tRNA processing. However, this gene set is substantially reduced relative to many other Eukaryotes, with many functions having been transferred to the nucleus over the course of evolution. Fungal mitogenome annotation is challenging due in part to the large number of introns, which are primarily either group I or II and are self-splicing, often containing homing endonuclease genes. While there are increasing numbers of complete and annotated fungal mitogenomes, the relatively few existing sequences are largely from the Ascomycota and Basidiomycota. We have developed a new annotation pipeline for automatically annotating fungal mitogenomes, utilizing *ab initio* predictions as well as HMM-based predictions from a core set of conserved fungal mitochondrially-encoded genes. Here, we present a broad overview of around 1000 mitogenome annotations from across the Fungal Tree of Life. Results include a broad-scale overview of gene conservation across all fungi as well as topology comparison between mito- and nuclear genome trees. Single-scaffold fungal mitogenome assemblies range from 12 kb to 1.3 Mb. Going forward, annotated mitogenomes at the JGI will be presented with a new analysis tab on individual MycoCosm genome portals. This effort will expand the set of fungal mitogenome annotations, and work toward a better understanding of fungal mitochondrial genome structure and evolution.

Building Models of Carbohydrolytic-Active Enzymes With Machine Learning

David Bui, Sushant Mane, William B Andreopoulos

Department of Computer Science, San Jose State University

Carbohydrolytic-active enzymes or (CAZymes) are protein enzymes that facilitate the breakdown of complex carbohydrates, usually from plants. The study in Hess et al. (2011) took 90 candidate enzymes with similarities to known CAZyme proteins in order to discover enzymes that could efficiently break down plant matter for large scale conversion into biomass. They tested these enzymes against 10 different plant substrates. Around 57% (51/90) of these were enzymatically active against 6 of out the 10 different plant substrates. The plant substrates that showed enzyme reactions were Avicel, Carboxymethyl Cellulose (CMC), Xylan, Miscanthus, Switchgrass and Lichenan.

In this study we propose a machine learning-based approach to find more CAZymes that are enzymatically active against any of these substrates. Our goal is to build machine learning models of the 90 proteins with a focus on the 57% of CAZYmes that were enzymatically active. In the search for CAZymes with desired substrate reactions we created machine learning models on the dataset of 90 CAZyme candidate proteins. After gathering the fasta sequence files we turned the sequences into overlapping k-mers and applied the bag of word technique of count vectorizing these sequences. These sequences were then input and trained on various machine learning techniques. From experiments and gridsearch hyper parameter searching we determined Multilayer Perceptrons achieved the greatest accuracy on the training portion of the dataset. We trained one model on classifying sequences that had no enzymatic substrate reactions versus sequences that showed any enzymatic substrate reactions. We then trained an additional 6 models focusing on whether a sequence had a specific substrate reaction versus any other having other enzymatic substrate reactions. Applying randomized shuffles to the dataset we were able to achieve accuracies 90% and above on each of the models.

We propose a classification pipeline for discovery of enzymatically active CAZYme sequences using the models we trained. The pipeline will first apply a no reaction model on a sequence to determine if the enzyme has the potential to have any enzymatic substrate reactions; if there is a potential, we will apply on the sequence each of the 6 trained substrate specific models and then assign the corresponding substrate classification if our models award the sequence with the label. Future work would include running through our pipeline a large dataset of carbohydrate-active genes and evaluating the results. Code for the dataset and for this pipeline can be found on GitHub.

1. Matthias Hess et al. “Metagenomic discovery of biomass-degrading genes and genomes from cow rumen”. In: Science 331.6016 (2011), pp. 463–467.

Exploring and Visualizing a *Panicum Hallii* Mutant Library

Beryl N. Arinda, Matthew Mingay, Sharon Greenblum

Abstract

Through large-scale mutation induction experiments and high-throughput genome sequencing approaches, gene function can be extensively investigated. However, such experiments generate large data sets and expert knowledge of bioinformatics tools is required to manage, explore, and translate this data. Unfortunately, many researchers lack the experience in programming languages, data preprocessing or reformatting tools required to parse these data sets for biological insight. An interactive, web-based platform that can be used to effectively navigate, visualize, and analyze massive genomics datasets would be extremely useful to the research community at large.

A robust, user-friendly web application featuring a point, click or type workflow was developed to navigate genomic variant data from large resequencing projects. Initial design and testing was conducted using data from sequencing a fast-neutron (FN) induced mutant population of *Panicum hallii,* a more tractable relative of the biofuel feedstock species *Panicum virgatum*. However, our application was designed to accommodate any variant dataset provided in the standardized variant call format (VCF) with accompanying phenotype information. The tool was implemented using Shiny, an R based toolkit. The application takes search queries input by the user (ex. gene names, phenotype or keyword descriptors) and generates an output containing informative metrics (ex. mutation type, mutation frequency, mutation position, putative impact, and phenotypes) from an index of variants. Thus, our application provides users with the ability to easily explore, parse, and summarize information contained within large VCF datasets through a user-friendly, interactive web interface, requiring no programming experience whatsoever.

This resource will be available to researchers through a web browser and can be used to uncover new genotypic and phenotypic information that could facilitate hypothesis formation, experimental designs, and ultimately guide the genomic engineering of desired traits.

“Corals’ microbiome nitrogen metabolism impacts thermal tolerance”

Authors: Viridiana Avila-Magana, Bishoy Kamel, Monica Medina

Elevated temperatures as a consequence of global change poses a threat for coral species. Coral-microbiota interactions plays a key role in the survival of corals. Using a phylogenetic framework, we have performed a controlled bleaching experiment in three different coral species displaying distinct thermal tolerances. Using metatranscriptome analysis we show evidence that coexistent coral holobiont microbial associates display different responses and metabolic capabilities under high temperature stress. Specifically, the holobiont nitrogen metabolism and amino acid biosynthetic pathways exhibit distinct expression profiles across taxa during thermal stress. We find that each member has a unique response that can influence the holobiont’s final ability to withstand thermal stress.

Thermotolerant coral holobionts rely on the associated bacterial community for the maintenance of nitrogen metabolic pathways (i.e., assimilatory and dissimilatory nitrate reduction, denitrification and fixation) during stress. These microbial contributions to the coral holobiont can have conspicuous evolutionary and ecological outcomes under climate change.

By integrating our results in a comparative framework, we suggest that bleaching is a complex response influenced by the holobiont members’ interactions. These complex interactions during bleaching include the supply of nitrogen by microbial communities, highlighting the crucial role of bacteria the maintenance of the symbiotic state of the coral holobiont.

Are we there yet? Benchmarking low-coverage nanopore long-read sequencing for the assembling of mitochondrial genomes using the vulnerable silky shark *Carcharhinus falciformis*

J. Antonio Baeza[1, 2, 3,*] & F. J. García-De León[4,]

[1] *Department of Biological Sciences, 132 Long Hall, Clemson University, Clemson, SC 29634, USA.*

[2] *Smithsonian Marine Station at Fort Pierce , 701 Seaway Drive, Fort Pierce, Florida 34949, USA.*

[3] *Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, Larrondo 1281, Coquimbo, Chile.*

[4] *Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste, La Paz, Baja California Sur, México*

[*]correspondence: JA Baeza, baeza.antonio@gmail.com, phone:1 772 475 0053

Abstract

Whole mitochondrial genomes are quickly becoming markers of choice for the exploration of within-species genealogical and among-species phylogenetic relationships. Most often, 'primer walking' or 'long PCR' strategies plus Sanger sequencing or low-pass whole genome sequencing using Illumina short reads are used for the assembling of mitochondrial chromosomes. Herein, we tested whether mitochondrial genomes can be sequenced from long reads nanopore sequencing data exclusively. Secondly, we examined the accuracy of the long-reads assembled mitochondrial chromosomes when comparing them to a 'gold' standard reference mitochondrial chromosome assembled using Illumina short-reads sequencing. Low-pass whole genome

sequencing using a MinION ONT pocket-sized platform plus customized *de-novo* and *reference-based* workflows assembled and circularized a highly accurate mitochondrial chromosome in the silky shark *Carcharhinus falciformis*. Indels at the flanks of homopolymer regions explained most of the dissimilarities observed between the 'gold' standard reference mitochondrial chromosome assembled using Illumina short-reads and each of the long-reads mitochondrial genome assemblies. Although not totally accurate, mitophylogenomics and barcoding analyses (using entire mitogenomes and the D-Loop/Control Region, respectively) demonstrated that long-reads assembled mitochondrial genomes can and do reliably identify the sequenced individual as belonging to *C. falciformis* and distinguished the same individual from others belonging to closely related congeneric species. This study demonstrates that mitochondrial genomes can be sequenced from long-reads nanopore sequencing data exclusively. Nanopore technology can be used to quickly test *in situ* mislabeling in the shark fin fishing industry and thus, improve surveillance protocols, law enforcement, and the regulation of this fishery. This study will also assist with the transferring of high-throughput sequencing technology to moderate- and low-income countries so that international teams of scientists can explore population genomics in this shark using inclusive research strategies.

KEYWORDS: long-read sequencing, nanopore, elasmobranch

Carbon and nitrogen cycling niche differentiation of bacteria and fungi in a temperate forest topsoil: A metaproteomic approach

Petr Baldrian, Robert Starke, Rubén López Mondéjar, Zander Rainer Human, Martina Štursová, Tomáš Větrovský, Stephen J. Callister[2], Mary S. Lipton[3], Adina Howe[4], Lee Ann McCue[3], Christa Pennacchio[5] Igor Grigoriev[5]

[1]Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic

[2]Pacific Northwest National Laboratory, Richland, WA, United States of America

[3]Environmental Molecular Sciences Laboratory, Richland, WA, United States of America

[4]Iowa State University, Ames, IA, United States of America

[5]DOE Joint Genome Institute, Berkeley, CA, United States of America

Temperate coniferous forests sustain the highest levels of biomass of all terrestrial ecosystems and belong to the major carbon sinks on Earth. However, the composition of microbial communities and their functional diversity across various habitats of the structurally complex soil environment have yet to be fully understood. Here, we analyzed the metaproteomes from litter, plant roots, rhizosphere, and bulk soil in a temperate coniferous forest in summer and winter to improve the understanding of the interplay between bacterial and fungal communities in these distinct habitats. Our metaproteomic approach yielded a total of 139,127 proteins that allowed to differentiate the contribution of microbial taxa to protein expression as well as the general functionality based on KEGG Orthology in each habitat. The pool of expressed carbohydrate-active enzymes (CAZymes) was dominated by fungal proteins. While CAZymes in roots and litter targeted mostly the structural biopolymers of plant origin such as lignin and cellulose, the majority of CAZymes in bulk soil and in the rhizosphere targeted oligosaccharides, starch, and glycogen. Proteins involved in nitrogen cycling were mainly of bacterial origin. Most nitrogen cycling proteins in litter and roots participated in ammonium assimilation while those performing nitrification were the most abundant in bulk soil and rhizosphere samples. Together, our results indicated niche differentiation of the microbial involvement in carbon and nitrogen cycling in a temperate coniferous forest topsoil (1).

(1) Starke, R. et al., 2021. Niche differentiation of bacteria and fungi in carbon and nitrogen cycling of different habitats in a temperate coniferous forest: A metaproteomic approach. Soil Biology & Biochemistry 155, 108170.

1. Introduction

A large share of the Earth's carbon (C) resides in living trees, forest vegetation and soil microbial biomass of temperate coniferous forests (Pan et al., 2011). Understanding their functional responses to global change is important for predictive climate modelling as they act as both C sources and sinks. The C cycling in forest topsoil is highly integrated within the biosphere. Its complex dynamics are influenced by both soil community functions and a heterogeneous trophic environment. Soil microbes contribute to a variety of ecosystem functions such as the decomposition of organic matter, tree root symbiosis and pathogenicity. Microbes inhabiting litter, on the other hand, depend on complex litter-bound nutrients recycled by microbial decomposition that is partly mineralized and partly transformed into organic matter, which accumulates in the forest floor (Baldrian, 2017). Plant root symbionts, especially mycorrhizal fungi, act as important regulators of plant productivity and are responsible for the acquisition of limiting nutrients such as nitrogen (N) or phosphorus (Van Der Heijden et al., 2008), especially in nutrient-poor ecosystems (Franklin et al., 2014). Mycorrhizal fungi dominate the fungal community in the coniferous forest soils below the litter layer (Lindahl et al., 2007) and represent potentially the largest pool of microbial biomass (Ekblad et al., 2013). By extending the tree roots by their hyphae, they mediate the flow of tree-derived C through its roots into the soil, which can account for more than 30% of total primary production in these ecosystems (Clemmensen et al., 2013; Franklin et al., 2014). The topsoil clearly represents the major compartment of C storage and microbial activity in coniferous forests but understanding the processes of C and nutrient cycling is, however, complicated due to its spatial heterogeneity resulting from the existence of four habitats: litter, plant roots, rhizosphere, and bulk soil that are distinct in many respects (Baldrian, 2017). In addition to the fact that soil and rhizosphere contains smaller amount of organic matter than litter and roots, roots represent a habitat where the bulk of the plant biomass is living and thus not readily degraded. Furthermore, spruce litter and roots also differ in the content of major biopolymers: brown spruce needles contain around 32% lignin, 29% cellulose and 21% hemicelluloses (Johansson, 1995) while fine spruce roots (<2 mm) are reported to contain more lignin (34%) and less cellulose (20%) and hemicelluloses (9%; (Hobbie et al., 2010)). Therefore, the application of the same analytical tools to all these habitats is the best option to understand and to relate their functions in ecosystem processes that reflect their properties.

Specific functions of the topsoil organisms can be statistically inferred based upon the homology of genes, transcripts or proteins that are present to experimentally characterized genes and proteins in known organisms. This so-called ortholog annotation is performed by the comparison to phylogenomic

databases, covering either a wide range of functional classes, e.g., KEGG Orthology (Kanehisa et al., 2016a, 2016b) or more specific subsets, such as the carbohydrate-active enzymes (CAZymes) (Lombard et al., 2014) involved in C cycling. Processes of C cycling are closely related to N cycling as their ratio can be used to identify the origin of a source (Müller, 1977; Ishiwatari and Uzaki, 1987; Prahl et al., 1994). The theory of ecological stoichiometry suggests that strictly homeostatic organisms have low nitrogen use efficiency (NUE) but high carbon use efficiency (CUE) at low substrate C/N-ratios (Sterner and Elser, 2002). By contrast, these organisms are expected to lower their CUE while increasing their NUE at high substrate C/N-ratios (Mooshammer et al., 2014). Logically, there exists a threshold elemental ratio (TER) that defines the elemental ratio at which the metabolic control of an ecological system switches from C limitation to N limitation (Urabe and Watanabe, 1992; Anderson and Hessen, 1995). The result is an inverse relationship between C and N cycling separated by the TER (Mooshammer et al., 2014) that may yield a different set of partaking microbes (Starke, 2017). In coniferous forest soils, a variety of C sources of different complexity are decomposed with a vast arsenal of glycoside hydrolases (GHs) that hydrolyze glycosidic bonds (Lombard et al., 2014). The functional classification comprises families of structurally-related catalytic and carbohydrate-binding modules or functional domains of enzymes that degrade, modify or create glycosidic bonds (Lombard et al., 2014), and are thus important for the breakdown of the different components of biomass: cellulose, hemicellulose and pectin from plant, chitin from fungal and peptidoglycan from bacterial cell walls. Importantly, however, members of the same CAZymes family can catalyze different reactions and their family membership may not sufficiently indicate the targets of their activity (López-Mondéjar et al., 2016) as seen in the family GH5. In addition to GHs, CAZymes include glycosyl transferases (GTs), carbohydrate esterases (CEs) and polysaccharide lyases (PLs) involved in the decomposition processes as well as auxiliary activities (AA) that cover redox enzymes acting in conjunction with CAZymes. The AA class is key for the decomposition of lignin by oxidative enzymes (Levasseur et al., 2013). On the other hand, N is available in the environment in different chemical forms: organic nitrogen, ammonium, nitrite, nitrate, nitrous oxide, nitric oxide, and inorganic nitrogen gas. Several enzymes predominantly produced by microorganisms are involved in their interconversions during assimilation, denitrification, nitrification or nitrogen fixation (Simon and Klotz, 2013; Sparacino-Watkins et al., 2014). For the global N budget, nitrogen fixation in natural ecosystems (Vitousek et al., 2013) as well as agricultural systems (Fowler et al., 2013), the emission of oxidized N forms from soil (Pilegaard, 2013) and of ammonium in terrestrial ecosystems (Sutton et al., 2013) were previously deemed important.

The use of metaproteomics became more and more popular to study how microbes contribute to soil ecosystem services (Von Bergen et al., 2013) as proteins not only provide both functional and

taxonomic information (Hettich et al., 2013; Von Bergen et al., 2013; Wilmes et al., 2015), but also represent the catalysts of important biochemical reactions (Wong, 2009). In favor of using proteins are: (i) the extracellular activity of soil proteins that can persist and remain active through the stabilization by humic substances and clay (Burns et al., 2013) whereas up to 40% of DNA (Carini et al., 2016) and 6% of RNA (Papp et al., 2018) were reported to derive from dead cells; (ii) the limitations of nucleic acids sequencing such as quantitative accuracy (Feinstein et al., 2009; Hungate et al., 2015), which may result in biased proportions of community members or functions; (iii) the difference between mRNA and protein levels depending on the temporal scale, on the complexity of the biological system and on the type of perturbation (Liu and Aebersold, 2016). Methodological advances, such as two-dimensional liquid chromatography (Callister et al., 2018) and site specific databases from sequencing approaches, now allow for a high protein identification rate that makes it possible to detect proteins with low relative abundance, such as CAZymes or proteins related to N cycling. Even though they are essential for the ecosystem functioning, they make up only a small part of the proteome that is dominated by structural proteins and proteins participating in the central metabolism.

In this study, we used a metaproteomic approach combining a site specific database comprising metagenomes, metatranscriptomes and fungal genomes together with two-dimensional liquid chromatography (Callister et al., 2018) to unveil the taxonomic and functional composition of the metaproteome in a temperate coniferous forest. In addition to describing the general functionality using KEGG Orthology, we have specifically addressed CAZymes that are involved in C cycling and proteins involved in N cycling. Four contrasting habitats – litter, plant roots, rhizosphere, and bulk soil, were compared to unveil habitat specific functions. We hypothesized that the utilization of plant biopolymers performed by fungi will be highest in litter, while components of microbial biomass targeted largely by bacteria (Lladó et al., 2017; López-Mondéjar et al., 2018) will be more important in soil and rhizosphere. Due to the high C/N-ratio and the high share of biomass-derived C in litter, we also expected that the incorporation of ammonia will be higher in litter compared to rhizosphere and soil where the share of inorganic N is higher and the C/N-ratio is lower (Baldrian et al., 2012).

2. Material and methods

2.1. Study area and sample collection

The study area was located at high altitudes (1170-1200 m) of the Bohemian Forest mountain range (Central Europe; 49°02 N, 13°37 E) and was covered by an unmanaged Norway spruce (*Picea abies* (L.) H. Karst.) forest. The mean annual temperature was 5°C and the mean annual precipitation was 1000 mm. The understory was either missing or composed of grasses (*Avenella*, *Calamagrostis*), bilberries (*Vaccinium*) and mosses. The same study area was explored previously to identify the total and active microbial communities and the analysis of microbiome activity (Baldrian et al., 2012; Žifčáková et al., 2016). To account for potential seasonal variation, samples were collected at two time points, on 09/22/2015 - peak of vegetation season represented by the highest belowground allocation of photosynthates (Hansson et al., 2013) and on 03/22/2016 (end of winter). At five sites, located approximately 250 m from each other, sixteen soil cores (4 cm diameter) were collected from a defined area of 3 x 3 m. Litter (2-4 cm thickness), soil (organic horizon, 3-6 cm of depth), rhizosphere soil (soil adhering to plant roots that had to be manually detached) and plant roots (< 2 mm in diameter) themselves were separated from each other. Organic horizon of soil was selected for sampling since it contains the bulk of the microbial biomass, roots, and ectomycorrhizal mycelia in this system (Žifčáková et al., 2016). Moreover, due to the stony nature of the site, mineral horizon is often present only as random pockets between stones and cannot be reasonably sampled. Litter was cut into 0.5 cm pieces and mixed, soil was passed through a 5-mm sterile mesh and mixed, and the same was later done for rhizosphere soil. Due to the need for immediate freezing in liquid nitrogen, roots were washed on site in sterile deionized water until all visible soil material was removed. All root material was pooled per site and sampling time resulting in a total of 40 samples (5 sites × four habitats × two seasons). All samples were immediately frozen in liquid nitrogen and stored on dry ice. Samples for nucleic acid extraction were stored at -80°C, samples for metaproteome analysis were freeze-dried and stored at -45°C.

2.2. Litter and soil chemistry

Soil physicochemical parameters were measured in an external laboratory. Litter, soil and root samples were powdered prior to all analyses. Powdered root samples were included in C and N analysis. pH was measured by mixing powdered samples with either 5 ml of deionized water or KCl shaken for 5 min on orbital shaker before measurement on a laboratory pH meter (WTW Multilab 540). C and N content was measured by combustion of powdered material in 100% O2 at 1000°C, and the amount of oxidized C and N recorded in a Flash 2000 elemental analyser (Thermo Scientific) and analyzed with Eager Xperience software (Thermo Scientific) to get its relative content. Organic carbon and carbonate content were

measured with a second, parallel analysis of the sample after decomposition with hydrochloric acid and fumigation of the released carbon dioxide. Nitrate and ammonium ions were extracted in 0.5 M potassium sulphate solution, filtered, and then quantified by injection into a Quikchem FIA 8000 (Lachat Instruments, Colorado, USA) autoanalyser instrument and processed using Omnion 3.0 software (Lachat Instruments, Colorado, USA). Available Ca, K, and Mg was measured after extraction in Mehlich II solution in a 1:5 sample:extractant ratio and then quantified by atomic absorption spectrophotometry in a Analytik Jena ConrAA 700 (Jena, Germany).

2.3. Protein extraction

Protein extractions were performed according to previously described protocols (Nakayasu et al., 2016; Callister et al., 2018) with minor modifications. Extractions were performed from 2 g of lyophilized rhizosphere soil and bulk soil samples. Briefly, proteins were extracted by the sequential addition of Millipore ultrapure water (Millipore, Billerica, MA), ice-cold methanol and chloroform to a final ratio of 8:4:2 chloroform:methanol:water (v/v). A mixture of 0.9–2.0 mm stainless steel beads, 0.1 mm garnet beads and 0.1 mm zirconia beads were added to these samples. Cells were lysed using a combination of horizontal vortexing and sonication with a probe sonicator at 20% amplitude for 30 seconds. Protein extractions from roots and litter were performed from 1 g of lyophilized samples reconstituted in Millipore, ultrapure water. Samples were homogenized with an Omni handheld homogenizer (Omni International, Marietta, GA) with disposable tips after adding ice-cold methanol. After homogenization, chloroform was added to obtain a final ratio of 8:4:2 chloroform:methanol:water (v/v). The aqueous and organic phases of the extracts were separated by centrifugation, and after discarding the upper methanol phase, the middle chloroform interphase and the bottom debris pellets were collected in separate clean 50 mL tubes. The protein interphase and debris pellets were precipitated in methanol, and the resulting pellets resuspended in a protein solubilization buffer (4% SDS and 100 mM DTT in 50 mM Tris buffer, pH 8.0, aided by sonication and shaking. A final concentration of 20% trichloroacetic acid was added to the clean supernatants, after which it was centrifuged, and the pellets collected. These protein pellets were washed three times in acetone, and in the final wash step, the protein-containing solutions were transferred to 2 mL tubes and the pellets collected by centrifugation. The Filter-Aided-Sample-Preparation (FASP, Expedeon LTD, UK) digestion method was used for digestion of protein pellets. Protein pellets were resuspended in UPX buffer (Expedeon LTD, UK) and transferred to the spin columns provided in the FASP kit along with 8 M urea in 100 mM ammonium bicarbonate. Spin columns were washed three times with

a urea buffer, followed by two washes with 25 mM ammonium bicarbonate. The spin columns were then transferred to clean collection tubes, and protein digestion performed in 100 mM NH_4HCO_3 and 4 µg trypsin on the spin columns. The digested peptides were collected by centrifugation, and any remaining peptides were collected with another addition of NH_4HCO_3. The eluent was concentrated to 30 µL using a SpeedVac at room temperature and a vacuum of 1 Torr. The concentrations were measured with a bicinchoninic acid (BCA) assay (Thermo Fisher Scientific, Waltham, MA). Samples were diluted to 0.1 µg per µL for LC-MS analysis.

2.4. Mass spectrometric measurement

Peptides from each sample were separated using two different orthogonal separation strategies directly coupled to the mass spectrometer generating 12 fractions and 12 MS/MS datasets per sample (Callister et al., 2018) with a run time of 26 h per sample for a total of 562 MS/MS datasets. For the first and additional dimension of liquid chromatography to traditional mass spectrometric measurements, peptides were diluted in 10 mM NH_4HCO_2 (pH 10.0) and separated into 12 fractions using a commercial Waters (Milford, MA) XBridge 5 µm particle size C18 column (4.6 mm i. d. x 250 mm length) with an attached 20 mm long x4.6 mm i. d. guard column at 0.5 mL min^{-1} using an Agilent 1100 series HPLC system (Agilent Technologies, SantaClara, CA) with two mobile phases: A) 10 mM NH_4HCO_2 (pH 10.0), and B) 10 mM NH_4HCO_2 (pH 10.0) with acetonitrile (v/v 10:90) over 120 min. All fractions were dried under vacuum and suspended in 15 µl 10 mM NH_4HCO_2 (pH 10.0). For the second dimension of liquid chromatography, dried peptides of each of the 12 fractions were diluted in 10 mM NH_4HCO_2 (pH 10.0), separated and eluted with the same setup as describe above but, instead of fractionating the eluted peptides, the peptides were ionized via electrospray emitters made in-house using 150 µm o.d. x 20 µm i.d. chemically etched fused silica (Kelly et al., 2006) and sprayed for 120 min into the tandem mass spectrometry (MS/MS), a Q-Exactive HF mass spectrometer (Thermo Scientific, San Jose, CA) that included a Pacific Northwest National Laboratory (PNNL)-designed nano-electrospray ionization interface (Callister et al., 2018). The ion transfer tube temperature and spray voltage were 300°C and 2.2 kV, respectively. Data were collected for 100 min. Mass spectra were acquired from 400-2000 m/z at a resolution of 60k and automatic gain control (AGC) target of $3x10^6$. The top 12 precursors were selected for higher-energy collisional dissociation (HCD) MS/MS spectra acquired in data-dependent mode with an isolation window of 2.0 Th and at a resolution of 17,500 (AGC target of $1x10^5$) using a normalized collision energy of 30 and a 30 s dynamic exclusion time.

2.5. Database generation and metaproteome annotation

MSConvert (Mayampurath et al., 2008) was used to convert proteome datasets into .dta search files in preparation for amino acid sequence assignment using the MS-GF+ search algorithm (Kim et al., 2008) in combination with the study site specific database comprising metagenomes and metatranscriptomes that derived from the same samples as the proteins. To obtain these, DNA and RNA were extracted and libraries were prepared as described previously (Žifčáková et al., 2016, 2017) except for the spruce roots where RNA was extracted and purified using the Nucleospin RNA Plant Kit (Macherey Nagel). All libraries were sequenced on an Illumina HiSeq2000 in the Joint Genome Institute, Walnut Creek, USA, to generate 150-base paired-end reads. Metagenome and metatranscriptome samples were assembled as described previously (Žifčáková et al., 2016, 2017). Briefly, Trimmomatic 0.36 (Bolger et al., 2014) and FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) were used to remove adaptor contamination, trim low-quality ends of reads and omit reads with overall low quality (<30), sequences shorter than 50 bp were omitted. Combined assembly of all samples was performed using MEGAHIT 1.1.3 (Li et al., 2015). Despite high sequencing depth of the metagenome, only a few tens of predicted metagenome-assembled genomes (MAGs) were obtained that covered a small share of the total diversity (Nayfach et al., 2020). It was thus impossible to utilize MAGs for annotation or analysis purposes. Gene calling and annotation was performed using JGI Gold as well as using BLAST against all published fungal genomes available in January 2018 (Grigoriev et al., 2014). Of these two, the taxonomic identification with a higher bitscore was used as the best hit. Genes encoding CAZymes were annotated using the dbCAN HMM database V6 (Huang et al., 2018). CAZymes were grouped based on their participation in the utilization of distinct C sources based on their classification to known CAZymes families (Lombard et al., 2014). For genes predicted in the metagenome and metatranscriptome assemblies, taxonomic annotations with a bitscore lower than 135 and functional annotations with e-values higher than $1e^{-30}$ were disregarded. The phyla *Chordata, Streptophyta* and *Coniferophyta* were not analysed. Functional analysis comprised metabolism, cellular processes, environmental information processing and genetic information processing. Metagenome and metatranscriptome data are available at IMG Integrated Microbial Genomes of the JGI DOE (Taxon object IDs 3300031708 and 3300032515; (Chen et al., 2019)).

Search parameters included a 20-ppm precursor mass tolerance dynamic modification of methionine (oxidation at +15.9949 Da), static modification of cysteine residues resulting from alkylation during sample preparation, and the inclusion of partially tryptic peptides. Protein abundances as the area

under the curve of proteins with a false discovery rate (FDR) below 0.05 based on the propensity score matching (PSM) score. In total, 139,127 proteins were identified by our metaproteomic approach. Proteins assigned to *Chordata* were disregarded as those are likely either human origin or wrongly annotated. Proteins assigned to the plants *Coniferophyta* and *Streptophyta* were discarded from the microbiome analysis. The significance of treatment effects on total protein content, the relative abundance of ribosomal proteins, CAZymes, N cycling and general functionality on level 2 and level 3 KEGG Orthology from the different habitats were calculated based on the Tukey's honestly significance difference (HSD) test (Tukey, 1949; Morrison et al., 2013). Numbers of proteins per time and habitat that were measured and annotated by taxonomy and function can be found in **Table 1**.

3. Results

3.1. Chemistry of litter, roots, and soil

Litter showed a significantly higher pH (P =0.0055 with H_2O and 0.0171 with KCl) than soil (**Table 1**). Total nitrogen was significantly higher in litter as compared to soil (P =0.0003) where both ammonium and nitrate were comparable between the two habitats. Both total carbon (P =0.0002) and organic carbon (P =0.0011) was significantly higher in litter while carbonate carbon was higher in soil (P =0.0726). Calcium (P =0.0074), magnesium (P =0.0020), and potassium (P =0.0054) were all significantly higher in litter as compared to soil. Nitrogen content in the roots was significantly lower (P <0.05) as compared to soil and litter. Otherwise, the carbon content in the roots was significantly higher than in soil but significantly lower than in litter (P <0.05). Together, this resulted in significantly higher C/N-ratios in roots as compared to litter and soil (P <0.05).

3.2. Protein content and the relative abundance of ribosomal proteins, CAZymes and proteins related to N cycling

Plant roots in September showed significantly (P <0.05) higher protein counts compared to litter in March and September, which in turn contained significantly more proteins than found in plant roots in March as well as bulk and rhizosphere soil regardless the sampling time (**Figure 1a**). The relative

abundance of ribosomal proteins from plants was significantly higher in spruce roots regardless the sampling time than in litter, bulk, and rhizosphere soil where the share of plant proteins was relatively low (**Figure 1b**). In total, the ribosomal proteins from non-plant proteins made up approximately 0.5% in spruce roots, bulk, and rhizosphere soil and up to 2% in litter. The relative abundance of plant proteins assigned to CAZymes was significantly higher in spruce roots sampled in March compared to September and higher than in the other habitats regardless the sampling time. The relative abundance of non-plant CAZymes ranged from 0.2% in roots to 2.0% in rhizosphere soil. Plant proteins related to N cycling were significantly higher in roots regardless the sampling time but showed no significant contribution in litter, bulk, and rhizosphere soil. Non-plant proteins related to N cycling ranged from 0.02% in roots 0.87% in rhizosphere soil.

3.3. Total and ribosomal microbial community composition, and its general functionality

The total protein pool after the exclusion of the plant proteins was dominated by bacteria regardless of habitat and the sampling time with higher proportions of eukaryotes in roots and to a lesser extent in litter (**Figure 2**). Particularly *Proteobacteria*, *Actinobacteria* and *Acidobacteria* were the dominant bacterial phyla. Archaeal proteins were most abundant in bulk and rhizosphere soil during March, representing up to 0.2% and 0.1% relative protein abundance, respectively, but only 0.01-0.03% in the other habitats and sampling times. The fungal *Basidiomycota* and *Ascomycota* dominated among the eukaryotes regardless of the habitat or the sampling time, with a higher *Basidiomycota*-to-*Ascomycota* ratio in roots and litter without differences with sampling time. The composition of ribosomal proteins showed a much larger contribution of eukaryotes in roots, reaching approximately 65% of the total and in litter (up to 50%). The ribosomal proteins were dominated by the fungi from the *Basidiomycota* and *Ascomycota* like the total proteins, but higher abundances of *Arthropoda* were found particularly in roots.

In total, 16 of the 19 most abundant level 2 KEGG categories and 22 of the 26 most abundant level 3 categories showed significant differences between habitats in their relative contribution to the total proteome (**Figure 3**). Particularly on level 2, proteins involved in amino acid metabolism, cell growth and death, energy metabolism, signal transduction, translation and transport and catabolism were highly abundant in roots. Rhizosphere soil showed high number of proteins involved in biosynthesis of other secondary metabolites, carbohydrate metabolism, cellular community - prokaryotes and xenobiotic biodegradation and metabolism. In bulk soil, proteins related to cellular community - prokaryotes,

xenobiotic biodegradation and metabolism, cell motility and membrane transport were highly abundant. Proteins involved in amino acid metabolism, cell growth, and death, cellular community - eukaryotes, folding, sorting, and degradation, nucleotide metabolism, transcription, transport and catabolism, and xenobiotic biodegradation and metabolism were abundant in litter. On level 3, proteins involved in C fixation in photosynthetic organisms, glycolysis/gluconeogenesis, necroptosis, oxidative phosphorylation, and photosynthesis were abundant in roots. Proteins involved in citrate cycle, fructose and mannose metabolism, glycolysis/gluconeogenesis and quorum sensing were abundant in rhizosphere soil. Bulk soil showed high numbers of proteins related to ABC transporters, bacterial chemotaxis, flagellar assembly, fructose and mannose metabolism, quorum sensing and two-component system. Lastly, proteins involved in apoptosis, glycolysis/gluconeogenesis, glyoxylate and dicarboxylate metabolism, hippo signaling pathway, phagosome, purine metabolism, and pyrimidine metabolism were abundant in litter.

3.4. Proteins involved in C and N cycling

Most proteins assigned to CAZymes, involved in the utilization of complex C compounds, were assigned to Eukaryotes, with high *Basidiomycota* abundances in bulk and rhizosphere soil as opposed to high *Ascomycota* abundances in roots and litter (**Figure 4a**). The share of bacterial CAZymes ranged from approximately 15% in roots to 35% in rhizosphere soil and were dominated by the *Actinobacteria*. Different sources of C were targeted in each habitat. While the CAZymes targeting plant cell wall polysaccharides (lignin, cellulose, glucuroxylans, pectin, and galactomannans) represented as much as 68% of all CAZymes in roots and 58% in litter, their share in rhizosphere and bulk soil was only 25% and 23% (**Figure 4b**). Compounds of fungal origin (chitin, glucans, and glucomannans) were most frequently targeted in rhizosphere and litter (10% and 9% of CAZymes), but due to high variation, the differences among habitats were not significant. Highest utilization of cello- and xylooligosaccharides was observed in soil where they were targeted by 46% of CAZymes while their importance was low in the roots (15%). Both rhizosphere and soil were typical by a high share of CAZymes utilizing reserve compounds (starch, glycogen, and trehalose) that represented 30% and 24% of all CAZymes compared to only 11% in roots and 9% in litter. In roots, proteins involved in lignin degradation were more abundant in March, suggesting the utilization of roots that died during winter while pectin-targeting proteins were more abundant in September when mycorrhizal fungi colonized roots likely by dissolving their pectic middle lamellae (**Figure 4b**).

The recovered proteins involved in N cycling were mostly of bacterial origin, with a share between 75% in roots to 95% in rhizosphere soil (**Figure 5a**). Particularly the proteins of *Proteobacteria* were highly abundant in all habitats. *Acidobacteria* showed higher relative abundance in litter and in roots (especially in September), while the proteins produced by *Actinobacteria* were abundant in roots, bulk soil, and litter. Eukaryotic proteins were mostly assigned to fungi, both *Basidiomycota* and *Ascomycota*. Functionally, spruce roots and litter were dominated by proteins involved in ammonia assimilation into organic molecules whereas proteins involved in nitrification were most common in the rhizosphere soil and bulk soil, the latter showing higher share of ammonia assimilation (**Figure 5b**).

4. Discussion

4.1. Remarks to the metaproteomic approach

Generally, metaproteomic approaches yield thousands of proteins per sample (Kleiner, 2019). However, with the relative CAZymes abundance of up to 5% (Žifčáková et al., 2017) only tens to hundreds of proteins identified as CAZymes will be found per sample. This is the reason why protein abundances for CAZymes have not been reported until now. The total of 139,127 proteins recovered from four forest habitats in our study make it possible to generate sufficient absolute numbers of CAZymes proteins. Indeed, the relative abundance of CAZymes proteins depended on the habitat with 1.4 ±0.4% in litter, 3.0 ±0.1% in roots, 0.1 ±0.0% in rhizosphere, and 0.2 ±0.0% in soil. Noteworthy, the extraction of proteins from the four habitats must be taken into careful consideration as not only the total amount of proteins will differ in each habitat due to the relative importance of organisms from different domains reflected in different C and N content but also the sorption will change between the habitats as their matrices comprise of different molecules (Oonk et al., 2012). In fact, no pattern between the number of measured proteins and the C/N-ratios were visible between litter, roots, and soil, which indicated that the protein extractability is based on qualitative such as different C and N sources or different sorption properties in the environment rather than quantitative measures such as the total content of C and N in the environment (**Table 1**). Proteins related to the N cycling showed even lower relative abundances that would make it difficult to detect with smaller protein recovery, consistent with a previous study that showed only a few proteins related to N cycling in soil (Starke et al., 2016). Admittedly, the high number of identified proteins may derive from less stringent search parameters (20 ppm precursor mass tolerance

and 5% FDR). Using a stricter search parameter of 1% FDR reduced the number of peptide spectral matches from 1,401,092 (5% FDR) to 1,044,506 and the measured proteins dropped to 74,841. The FDR concept for protein identification was originally established for pure culture proteomics (Elias and Gygi, 2007), allowing for the comparison between different mass spectrometers and database search algorithms using a defined threshold of 1% (Barnouin, 2011). However, searches against large databases, such as the database used in this study, not only require long computation times but also decrease the number of identified proteins due to the overestimation of the FDR (Heyer et al., 2017). In fact, the limitation of the target-decoy controlled FDR approach in combination with large sequence databases in metaproteomics was responsible for missing valuable protein identifications (Muth et al., 2015) which makes FDR's higher than 1% common in metaproteomic approaches (Keiblinger et al., 2012; Schneider et al., 2012; Bastida et al., 2014; Figarska et al., 2018). That is why we believe that 5% FDR was a reasonable comprise for our study, but we also acknowledge that more stringent filtering on the probability scoring value does not only change the quantitative results but may also affect the qualitative results. Only 75.1% of the proteins were annotated by a taxonomy with high certainty, presumably due to the low number of fully sequenced and annotated genomes as they are mostly limited to those fungal and bacterial species that have undergone isolation and extensive characterization. Problematically, the vast majority of organisms were not yet studied (Pham and Kim, 2012; Martiny, 2019) and the annotation is based on the similarity to the genomes of the very few studied model organisms. However, the combination of metagenomes, metatranscriptomes and genomes of local microorganisms provided a comprehensive database that included fungal proteins, which are normally almost absent in metaproteomic data. Lastly, due to sequence homology between proteins, the annotation of deeper taxonomic depths than order level is unreliable (Starke et al., 2018) which is why a trait-based functional classification of fungi was impossible as the guilds require at least species-level identification.

4.2. The microbial community composition of the total and the ribosomal proteins

Only the ribosomal proteins showed a much larger contribution of fungal proteins in spruce roots and abundances like the ones of bacteria in litter. The relative contribution of fungal proteins increased with increasing C/N-ratio of the surrounding environment, consistent with higher C/N-ratios reported for fungi as compared to bacteria (Paul and Clark, 1988; Harris et al., 1997). Ribosomal proteins are components of the translation apparatus of microbes and could therefore reflect their abundance or a combination of abundance and activity, considering that more active taxa likely contain more ribosomes

(Baldrian et al., 2012). Due to the fact that the ribosomal proteome and the proteome showed similar patterns in Mediterranean semi-arid soils (Bastida et al., 2016), we exclude a differential production of ribosomes as compared to all proteins between bacteria and eukaryotes. The ribosomal proteome related better to the fungal dominance of transcripts found in the very same (Žifčáková et al., 2016) or similar ecosystems of coniferous forest soils (Bailly et al., 2007; Damon et al., 2012; Takasaki et al., 2013), which is in line with the role of fungi as the key players in the plant-root-microbiome. They not only regulate plant productivity and acquire limiting nutrients (Van Der Heijden et al., 2008), but also decompose recalcitrant organic polymers in the litter (De Boer et al., 2005; Štursová et al., 2012). Noteworthy, we report similar activities between bacteria and fungi in litter as opposed to the dominance of fungi in the more decomposed litter of coniferous forest soils reported before (Ekblad et al., 2013). Among the eukaryotes, *Annelida* were also frequently found in the ribosomal proteins, highlighting the potential impact of larger eukaryotes that are generally disregarded in microbiome research. In fact, the role of microbes as predators or prey in the food-chain is generally neglected (Hall and Meyer, 1998; Haubert et al., 2009; Halvorson et al., 2016; Anderson et al., 2017), even though terrestrial omnivores and carnivores were reported to contain proteins and lipids with microbial fingerprints before (Larsen et al., 2009; Arthur et al., 2014; Dharampal and Findlay, 2016; Ohkouchi et al., 2017; Steffan et al., 2017). The large biomass combined with relatively low C/N-ratios (Swift and Boddy, 1984) of saprotrophic cord-forming basidiomycetes (Boddy and Watkinson, 2008) make this nutrient source attractive for these soil invertebrates (Pollierer et al., 2009). Here, we provide further evidence for the involvement of these animals in the plant-root-microbiome. However, the absolute protein numbers were not sufficient to allow for a specific functional assessment of topsoil invertebrates.

4.3. The general and the specific functionality of the microbial community

Most of the most abundant level 2 (84.2%) and level 3 KEGG functions (84.6%) had significant ($P < 0.05$) differences between the habitats, indicating a different functional profile of the microbial community depending on the location in the temperate coniferous forest. Not surprisingly, the functions involved in bacterial motility, namely flagellar assembly and chemotaxis were significantly more represented in the proteome of soil, the habitat with the highest level of heterogeneity and scarce nutrients (Baldrian, 2017). The limitations of total C sources in soil as compared to litter and roots (**Table 1**), and with that presumably simple C sources as well are observed in the lower share of proteins involved in glycolysis compared to the other habitats. The importance of microbial sensing in soil (and rhizosphere)

is supported by the elevated level of proteins involved in Quorum sensing (QS) [PATH:ko02024] as the ability to detect and to respond to population density by gene regulation (Bassler, 1999). QS bacteria release chemical signal molecules, so-called autoinducers (AI), which increase in concentration as a function of cell density. Indeed, autoinducer-2 kinase [EC:2.7.1.189] and AI-2 transport system substrate-binding protein were among the most abundant QS proteins (more information on the abundance of specific QS proteins can be found in the supplementary information, **Table S1**).

Fungi dominated the C cycling in each of the four habitats regardless the sampling time mainly targeting cello/xylobiose, cellulose, glucuroxylans, lignin, pectin and starch/glycogen, which corresponds with their estimated share on transcription in the same coniferous forests (Žifčáková et al., 2017) and to the proteomic analysis of decomposing beech litter (Schneider et al., 2012). All these lines of evidence support the expected major role of saprotrophic fungi in the decomposition of recalcitrant organic polymers (De Boer et al., 2005; Štursová et al., 2012). The dominance of fungi in the utilization of complex carbohydrates such as cellulose and lignin is in line with the better adaption of fungi to decompose recalcitrant plant biopolymers (Van der Wal et al., 2013). Similar to the metatranscriptomic approach, *Acidobacteria*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* were the most abundant bacterial producers of CAZymes (Žifčáková et al., 2017). These phylotypes are known to be common in acidic forest topsoil (Žifčáková et al., 2016) and especially the members of the *Acidobacteria* and *Bacteroidetes* seem to be potent degraders of biopolymers of plant and fungal origin in forest topsoil (Lladó et al., 2016, 2019).

Our results also indicate that there are differences in the sources of C between the habitats as the spectra of genes that encode for carbohydrate-active enzymes differ. Forest litter is rich in various recalcitrant plant biopolymers and supports the activity of those microorganisms that are able to attack them as seen from a large proportion of proteins targeting plant cell wall (Johansson, 1995; Rumpel et al., 2002; Lorenz et al., 2007; Štursová et al., 2020). These proteins were also found abundant in the roots, though the reason is less obvious. The probable explanation is that the proteins degrading plant cell walls in roots either attack dead roots or serve ectomycorrhizal symbionts to colonize root tips. The latter would be supported by the observed increase of proteins related to pectin decomposition in September (**Figure 4b**), the peak of ectomycorrhizal activity (Žifčáková et al., 2016). The soil and rhizosphere habitats are rather typical by the high share of proteins involved in utilization of C containing reserve compounds (e.g., starch, glycogen, trehalose), which may indicate fluctuations in C supply in these environments including the seasonal activity of roots. The ability of microbes to store C and utilize these resources may be an adaptation to these fluctuations (Žifčáková et al., 2017).

In line with the hypothesis that different microbes are involved in C and N cycling (Starke, 2017) due to their inverse CUE/NUE relationship (Mooshammer et al., 2014), bacteria dominated the N cycling in each habitat (more information on the abundance of specific proteins for each N cycling process can be found in the supplementary information, **Table S2**). *Acidobacteria* showed higher abundances in litter and roots, and seem to be associated with higher abundances of proteins related to ammonia assimilation, consistent with the previously reported positive correlation of the phylotype with mineral N contents (Ruppel et al., 2007; Hayatsu et al., 2008). Indeed, acidobacterial communities were more abundant and underwent significant structural changes in girdled plots (Rasche et al., 2011) in which the uptake of available nutrients such as ammonia or nitrate by trees is prevented (Högberg et al., 2001). In rhizosphere and bulk soil, nitrification was the most abundant process indicated by hydroxylamine dehydrogenase (*hao*, K10535) and three methane/ammonia monooxygenases (*pmoA*, K10944; *pmoB*, K10945 and *pmoC*, K10946). Ammonia oxidation, the first step in nitrification, is crucial for the global N cycle (Pratscher et al., 2011) and is predominantly performed by bacteria in soil (Kowalchuk and Stephen, 2001). Consistently, bacteria dominated proteins involved in N cycling regardless the habitat. The contrast between high nitrification in bulk and rhizosphere soil and the importance of the incorporation of ammonia into organic acids in litter and roots demonstrated a divergence between the habitats. While the relatively high abundance of NH_4^+ in bulk and rhizosphere soil allows its energetic use, different C/N-ratios in litter and roots allow for the preferential incorporation into the living biomass. Consistent with the lower C/N-ratios reported for soil samples as compared to litter in this temperate coniferous forest (Žifčáková et al., 2016), our data showed a similar trend of lower C/N-ratios in soil as compared to litter and roots (**Table 1**).

This metaproteomic analysis of major habitats in a temperate forest topsoil demonstrated the importance of both bacterial and fungal taxa. Consistent with our first hypothesis, fungal proteins dominated the decomposition of recalcitrant biopolymers in litter while bacterial proteins were more abundant in rhizosphere and soil, consistent with higher C/N-ratios in litter and roots than in soil. A niche differentiation between bacteria and fungi in N and C cycling, respectively, was unveiled. Consistent with our second hypothesis, a higher number of proteins related to ammonium assimilation was detected in litter as opposed to proteins related to nitrification in rhizosphere and soil. The metaproteome composition suggests that each of the studied habitats has specific features with the most distinct separation between soil and rhizosphere on one side, and litter and plant roots on the other side. While the latter habitats provide an essentially unlimited amount of complex C sources, their use may be limited by N availability. Our results indicate sufficient levels of N enabling nitrification but probably lower and fluctuating availability of C compounds. Despite limitations, metaproteomics seems to be suitable for the

analysis of microbiome composition and its functionality, allowing a better functional focus than metagenomics while reducing the potential limitations of metatranscriptomics due to changes in expression over time or quantitative accuracy of sequencing (Feinstein et al., 2009; Hungate et al., 2015). Admittedly, the relative importance of fungi in the eukaryotic proteins was likely driven by the fungal-based database and further studies need to confirm the results with respect to other eukaryotes such as annelids and arthropods.

Acknowledgements

This work was supported by the Czech Science Foundation (18-25706S and 20-02022Y), by the Ministry of Education, Youth and Sports of the Czech Republic (LTT17022). A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS; #49499) program and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory (grid.436923.9), which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). RS thanks Nico Jehmlich for mentoring.

References

Anderson, T.R., Hessen, D.O., 1995. Carbon or nitrogen limitation in marine copepods? Journal of Plankton Research. doi:10.1093/plankt/17.2.317

Anderson, T.R., Pond, D.W., Mayor, D.J., 2017. The role of microbes in the nutrition of detritivorous invertebrates: A stoichiometric analysis. Frontiers in Microbiology. doi:10.3389/fmicb.2016.02113

Arthur, K.E., Kelez, S., Larsen, T., Choy, C.A., Popp, B.N., 2014. Tracing the biosynthetic source of essential amino acids in marine turtles using δ13C fingerprints. Ecology. doi:10.1890/13-0263.1

Bailly, J., Fraissinet-Tachet, L., Verner, M.C., Debaud, J.C., Lemaire, M., Wésolowski-Louvel, M., Marmeisse, R., 2007. Soil eukaryotic functional diversity, a metatranscriptomic approach. ISME Journal. doi:10.1038/ismej.2007.68

Baldrian, P., 2017. Forest microbiome: Diversity, complexity and dynamics. FEMS Microbiology Reviews. doi:10.1093/femsre/fuw040

Baldrian, P., Kolaiřík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., Žifčáková, L., Šnajdr, J.,

Rídl, J., Vlček, Č., Voříšková, J., 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME Journal. doi:10.1038/ismej.2011.95

Barnouin, K., 2011. Guidelines for experimental design and data analysis of proteomic mass spectrometry-based experiments. Amino Acids. doi:10.1007/s00726-010-0750-9

Bassler, B.L., 1999. How bacteria talk to each other: Regulation of gene expression by quorum sensing. Current Opinion in Microbiology. doi:10.1016/S1369-5274(99)00025-9

Bastida, F., Hernández, T., García, C., 2014. Metaproteomics of soils from semiarid environment: Functional and phylogenetic information obtained with different protein extraction methods. Journal of Proteomics. doi:10.1016/j.jprot.2014.02.006

Bastida, F., Torres, I.F., Moreno, J.L., Baldrian, P., Ondoño, S., Ruiz-Navarro, A., Hernández, T., Richnow, H.H., Starke, R., García, C., Jehmlich, N., 2016. The active microbial diversity drives ecosystem multifunctionality and is physiologically related to carbon availability in Mediterranean semi-arid soils. Molecular Ecology 25. doi:10.1111/mec.13783

Boddy, L., Watkinson, S.C., 2008. Wood decomposition, higher fungi, and their role in nutrient redistribution. Canadian Journal of Botany. doi:10.1139/b95-400

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. doi:10.1093/bioinformatics/btu170

Burns, R.G., DeForest, J.L., Marxsen, J., Sinsabaugh, R.L., Stromberger, M.E., Wallenstein, M.D., Weintraub, M.N., Zoppini, A., 2013. Soil enzymes in a changing environment: Current knowledge and future directions. Soil Biology and Biochemistry. doi:10.1016/j.soilbio.2012.11.009

Callister, S.J., Fillmore, T.L., Nicora, C.D., Shaw, J.B., Purvine, S.O., Orton, D.J., White, R.A., Moore, R.J., Burnet, M.C., Nakayasu, E.S., Payne, S.H., Jansson, J.K., Paša-Tolić, L., 2018. Addressing the challenge of soil metaproteome complexity by improving metaproteome depth of coverage through two-dimensional liquid chromatography. Soil Biology and Biochemistry. doi:10.1016/j.soilbio.2018.07.018

Carini, P., Marsden, P.J., Leff, J.W., Morgan, E.E., Strickland, M.S., Fierer, N., 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nature Microbiology. doi:10.1038/nmicrobiol.2016.242

Chen, I.M.A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J.R., Seshadri, R., Smirnova, T., Kirton, E., Jungbluth, S.P., Woyke, T., Eloe-Fadrosh, E.A., Ivanova, N.N., Kyrpides, N.C., 2019. IMG/M v.5.0: An integrated data management and

comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Research. doi:10.1093/nar/gky901

Clemmensen, K.E., Bahr, A., Ovaskainen, O., Dahlberg, A., Ekblad, A., Wallander, H., Stenlid, J., Finlay, R.D., Wardle, D.A., Lindahl, B.D., 2013. Roots and associated fungi drive long-term carbon sequestration in boreal forest. Science. doi:10.1126/science.1231923

Damon, C., Lehembre, F., Oger-Desfeux, C., Luis, P., Ranger, J., Fraissinet-Tachet, L., Marmeisse, R., 2012. Metatranscriptomics reveals the diversity of genes expressed by eukaryotes in forest soils. PLoS ONE. doi:10.1371/journal.pone.0028967

De Boer, W., Folman, L.B., Summerbell, R.C., Boddy, L., 2005. Living in a fungal world: Impact of fungi on soil bacterial niche development. FEMS Microbiology Reviews. doi:10.1016/j.femsre.2004.11.005

Dharampal, P.S., Findlay, R.H., 2016. Mercury levels in largemouth bass (Micropterus salmoides) from regulated and unregulated rivers. Chemosphere. doi:10.1016/j.chemosphere.2016.11.162

Ekblad, A., Wallander, H., Godbold, D.L., Cruz, C., Johnson, D., Baldrian, P., Björk, R.G., Epron, D., Kieliszewska-Rokicka, B., Kjøller, R., Kraigher, H., Matzner, E., Neumann, J., Plassard, C., 2013. The production and turnover of extramatrical mycelium of ectomycorrhizal fungi in forest soils: Role in carbon cycling. Plant and Soil. doi:10.1007/s11104-013-1630-3

Elias, J.E., Gygi, S.P., 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods. doi:10.1038/nmeth1019

Feinstein, L.M., Woo, J.S., Blackwood, C.B., 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. Applied and Environmental Microbiology. doi:10.1128/AEM.00120-09

Figarska, S.M., Gustafsson, S., Sundström, J., Ärnlöv, J., Mälarstig, A., Elmståhl, S., Fall, T., Lind, L., Ingelsson, E., 2018. Associations of circulating protein levels with lipid fractions in the general population. Arteriosclerosis, Thrombosis, and Vascular Biology. doi:10.1161/ATVBAHA.118.311440

Fowler, D., Coyle, M., Skiba, U., Sutton, M.A., Cape, J.N., Reis, S., Sheppard, L.J., Jenkins, A., Grizzetti, B., Galloway, J.N., Vitousek, P., Leach, A., Bouwman, A.F., Butterbach-Bahl, K., Dentener, F., Stevenson, D., Amann, M., Voss, M., 2013. The global nitrogen cycle in the Twentyfirst century. Philosophical Transactions of the Royal Society B: Biological Sciences. doi:10.1098/rstb.2013.0164

Franklin, O., Näsholm, T., Högberg, P., Högberg, M.N., 2014. Forests trapped in nitrogen limitation - an ecological market perspective on ectomycorrhizal symbiosis. New Phytologist. doi:10.1111/nph.12840

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X.,

Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Research. doi:10.1093/nar/gkt1183

Hall, R.O., Meyer, J.L., 1998. The trophic significance of bacteria in a detritus-based stream food web. Ecology. doi:10.1890/0012-9658(1998)079[1995:TTSOBI]2.0.CO;2

Halvorson, H.M., White, G., Scott, J.T., Evans-White, M.A., 2016. Dietary and taxonomic controls on incorporation of microbial carbon and phosphorus by detritivorous caddisflies. Oecologia. doi:10.1007/s00442-015-3464-6

Hansson, K., Fröberg, M., Helmisaari, H.S., Kleja, D.B., Olsson, B.A., Olsson, M., Persson, T., 2013. Carbon and nitrogen pools and fluxes above and below ground in spruce, pine and birch stands in southern Sweden. Forest Ecology and Management. doi:10.1016/j.foreco.2013.05.029

Harris, D., Voroney, R.P., Paul, E.A., 1997. Measurement of microbial biomass N:C by chloroform fumigation-incubation. Canadian Journal of Soil Science. doi:10.4141/S96-064

Haubert, D., Birkhofer, K., Fließbach, A., Gehre, M., Scheu, S., Ruess, L., 2009. Trophic structure and major trophic links in conventional versus organic farming systems as indicated by carbon stable isotope ratios of fatty acids. Oikos. doi:10.1111/j.1600-0706.2009.17587.x

Hayatsu, M., Tago, K., Saito, M., 2008. Various players in the nitrogen cycle: Diversity and functions of the microorganisms involved in nitrification and denitrification. Soil Science and Plant Nutrition. doi:10.1111/j.1747-0765.2007.00195.x

Hettich, R.L., Pan, C., Chourey, K., Giannone, R.J., 2013. Metaproteomics: Harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. Analytical Chemistry. doi:10.1021/ac303053e

Heyer, R., Schallert, K., Zoun, R., Becher, B., Saake, G., Benndorf, D., 2017. Challenges and perspectives of metaproteomic data analysis. Journal of Biotechnology. doi:10.1016/j.jbiotec.2017.06.1201

Hobbie, S.E., Oleksyn, J., Eissenstat, D.M., Reich, P.B., 2010. Fine root decomposition rates do not mirror those of leaf litter among temperate tree species. Oecologia. doi:10.1007/s00442-009-1479-6

Högberg, P., Nordgren, A., Buchmann, N., Taylor, A.F.S., Ekblad, A., Högberg, M.N., Nyberg, G., Ottosson-Löfvenius, M., Read, D.J., 2001. Large-scale forest girdling shows that current photosynthesis drives soil respiration. Nature. doi:10.1038/35081058

Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T., Yi, H., Yang, Z., Yin, Y., 2018. DbCAN-seq: A database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Research. doi:10.1093/nar/gkx894

Hungate, B.A., Mau, R.L., Schwartz, E., Gregory Caporaso, J., Dijkstra, P., van Gestel, N., Koch, B.J., Liu,

C.M., McHugh, T.A., Marks, J.C., Morrissey, E.M., Price, L.B., 2015. Quantitative microbial ecology through stable isotope probing. Applied and Environmental Microbiology. doi:10.1128/AEM.02280-15

Ishiwatari, R., Uzaki, M., 1987. Diagenetic changes of lignin compounds in a more than 0.6 million-year-old lacustrine sediment (Lake Biwa, Japan). Geochimica et Cosmochimica Acta. doi:10.1016/0016-7037(87)90244-4

Johansson, M.B., 1995. The chemical composition of needle and leaf litter from scots pine, Norway spruce and white birch in scandinavian forests. Forestry. doi:10.1093/forestry/68.1.49

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016a. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research. doi:10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., Morishima, K., 2016b. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. Journal of Molecular Biology. doi:10.1016/j.jmb.2015.11.006

Keiblinger, K.M., Wilhartitz, I.C., Schneider, T., Roschitzki, B., Schmid, E., Eberl, L., Riedel, K., Zechmeister-Boltenstern, S., 2012. Soil metaproteomics - Comparative evaluation of protein extraction protocols. Soil Biology and Biochemistry. doi:10.1016/j.soilbio.2012.05.014

Kelly, R.T., Page, J.S., Luo, Q., Moore, R.J., Orton, D.J., Tang, K., Smith, R.D., 2006. Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. Analytical Chemistry 78, 7796–801. doi:10.1021/ac061133r

Kim, S., Gupta, N., Pevzner, P.A., 2008. Spectral probabilities and generating functions of tandem mass spectra: A strike against decoy databases. Journal of Proteome Research 7, 3354–63. doi:10.1021/pr8001244

Kleiner, M., 2019. Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. MSystems. doi:10.1128/msystems.00115-19

Kowalchuk, G.A., Stephen, J.R., 2001. Ammonia-Oxidizing Bacteria: A Model for Molecular Microbial Ecology. Annual Review of Microbiology. doi:10.1146/annurev.micro.55.1.485

Larsen, T., Lee Taylor, D., Leigh, M.B., O'Brien, D.M., 2009. Stable isotope fingerprinting: A novel method for identifying plant, fungal, or bacterial origins of amino acids. Ecology. doi:10.1890/08-1695.1

Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., Henrissat, B., 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnology for Biofuels. doi:10.1186/1754-6834-6-41

Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.W., 2015. MEGAHIT: An ultra-fast single-node solution for

large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. doi:10.1093/bioinformatics/btv033

Lindahl, B.D., Ihrmark, K., Boberg, J., Trumbore, S.E., Högberg, P., Stenlid, J., Finlay, R.D., 2007. Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. New Phytologist. doi:10.1111/j.1469-8137.2006.01936.x

Liu, Y., Aebersold, R., 2016. The interdependence of transcript and protein abundance: new data–new complexities. Molecular Systems Biology. doi:10.15252/msb.20156720

Lladó, S., López-Mondéjar, R., Baldrian, P., 2017. Forest Soil Bacteria: Diversity, Involvement in Ecosystem Processes, and Response to Global Change. Microbiology and Molecular Biology Reviews. doi:10.1128/mmbr.00063-16

Lladó, S., Větrovský, T., Baldrian, P., 2019. Tracking of the activity of individual bacteria in temperate forest soils shows guild-specific responses to seasonality. Soil Biology and Biochemistry. doi:10.1016/j.soilbio.2019.05.010

Lladó, S., Žifčáková, L., Větrovský, T., Eichlerová, I., Baldrian, P., 2016. Functional screening of abundant bacteria from acidic forest soil indicates the metabolic potential of Acidobacteria subdivision 1 for polysaccharide decomposition. Biology and Fertility of Soils. doi:10.1007/s00374-015-1072-6

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research. doi:10.1093/nar/gkt1178

López-Mondéjar, R., Brabcová, V., Štursová, M., Davidová, A., Jansa, J., Cajthaml, T., Baldrian, P., 2018. Decomposer food web in a deciduous forest shows high share of generalist microorganisms and importance of microbial biomass recycling. ISME Journal. doi:10.1038/s41396-018-0084-2

López-Mondéjar, R., Zühlke, D., Větrovský, T., Becher, D., Riedel, K., Baldrian, P., 2016. Decoding the complete arsenal for cellulose and hemicellulose deconstruction in the highly efficient cellulose decomposer Paenibacillus O199. Biotechnology for Biofuels. doi:10.1186/s13068-016-0518-x

Lorenz, K., Lal, R., Preston, C.M., Nierop, K.G.J., 2007. Strengthening the soil organic carbon pool by increasing contributions from recalcitrant aliphatic bio(macro)molecules. Geoderma. doi:10.1016/j.geoderma.2007.07.013

Martiny, A.C., 2019. High proportions of bacteria are culturable across major biomes. The ISME Journal.

Mayampurath, A.M., Jaitly, N., Purvine, S.O., Monroe, M.E., Auberry, K.J., Adkins, J.N., Smith, R.D., 2008. DeconMSn: A software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. Bioinformatics 24, 1021–3. doi:10.1093/bioinformatics/btn063

Mooshammer, M., Wanek, W., Hämmerle, I., Fuchslueger, L., Hofhansl, F., Knoltsch, A., Schnecker, J.,

Takriti, M., Watzka, M., Wild, B., Keiblinger, K.M., Zechmeister-Boltenstern, S., Richter, A., 2014. Adjustment of microbial nitrogen use efficiency to carbon:Nitrogen imbalances regulates soil nitrogen cycling. Nature Communications. doi:10.1038/ncomms4694

Morrison, S., Sosnoff, J.J., Heffernan, K.S., Jae, S.Y., Fernhall, B., 2013. Aging, hypertension and physiological tremor: The contribution of the cardioballistic impulse to tremorgenesis in older adults. Journal of the Neurological Sciences. doi:10.1016/j.jns.2013.01.016

Müller, P.J., 1977. C N ratios in Pacific deep-sea sediments: Effect of inorganic ammonium and organic nitrogen compounds sorbed by clays. Geochimica et Cosmochimica Acta. doi:10.1016/0016-7037(77)90047-3

Muth, T., Kolmeder, C.A., Salojärvi, J., Keskitalo, S., Varjosalo, M., Verdam, F.J., Rensen, S.S., Reichl, U., de Vos, W.M., Rapp, E., Martens, L., 2015. Navigating through metaproteomics data: A logbook of database searching. Proteomics. doi:10.1002/pmic.201400560

Nakayasu, E.S., Nicora, C.D., Sims, A.C., Burnum-Johnson, K.E., Kim, Y.-M., Kyle, J.E., Matzke, M.M., Shukla, A.K., Chu, R.K., Schepmoes, A.A., Jacobs, J.M., Baric, R.S., Webb-Robertson, B.-J., Smith, R.D., Metz, T.O., 2016. MPLEx: a Robust and Universal Protocol for Single-Sample Integrative Proteomic, Metabolomic, and Lipidomic Analyses. MSystems. doi:10.1128/msystems.00043-16

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T.B.K., Nielsen, T., Kirton, E., Faria, J.P., Edirisinghe, J.N., Henry, C.S., Jungbluth, S.P., Chivian, D., Dehal, P., Wood-Charlson, E.M., Arkin, A.P., Tringe, S.G., Visel, A., Abreu, H., Acinas, S.G., Allen, E., Allen, M.A., Andersen, G., Anesio, A.M., Attwood, G., Avila-Magaña, V., Badis, Y., Bailey, J., Baker, B., Baldrian, P., Barton, H.A., Beck, D.A.C., Becraft, E.D., Beller, H.R., Beman, J.M., Bernier-Latmani, R., Berry, T.D., Bertagnolli, A., Bertilsson, S., Bhatnagar, J.M., Bird, J.T., Blumer-Schuette, S.E., Bohannan, B., Borton, M.A., Brady, A., Brawley, S.H., Brodie, J., Brown, S., Brum, J.R., Brune, A., Bryant, D.A., Buchan, A., Buckley, D.H., Buongiorno, J., Cadillo-Quiroz, H., Caffrey, S.M., Campbell, A.N., Campbell, B., Carr, S., Carroll, J.L., Cary, S.C., Cates, A.M., Cattolico, R.A., Cavicchioli, R., Chistoserdova, L., Coleman, M.L., Constant, P., Conway, J.M., Mac Cormack, W.P., Crowe, S., Crump, B., Currie, C., Daly, R., Denef, V., Denman, S.E., Desta, A., Dionisi, H., Dodsworth, J., Dombrowski, N., Donohue, T., Dopson, M., Driscoll, T., Dunfield, P., Dupont, C.L., Dynarski, K.A., Edgcomb, V., Edwards, E.A., Elshahed, M.S., Figueroa, I., Flood, B., Fortney, N., Fortunato, C.S., Francis, C., Gachon, C.M.M., Garcia, S.L., Gazitua, M.C., Gentry, T., Gerwick, L., Gharechahi, J., Girguis, P., Gladden, J., Gradoville, M., Grasby, S.E., Gravuer, K., Grettenberger, C.L., Gruninger,

R.J., Guo, J., Habteselassie, M.Y., Hallam, S.J., Hatzenpichler, R., Hausmann, B., Hazen, T.C., Hedlund, B., Henny, C., Herfort, L., Hernandez, M., Hershey, O.S., Hess, M., Hollister, E.B., Hug, L.A., Hunt, D., Jansson, J., Jarett, J., Kadnikov, V. V., Kelly, C., Kelly, R., Kelly, W., Kerfeld, C.A., Kimbrel, J., Klassen, J.L., Konstantinidis, K.T., Lee, L.L., Li, W.J., Loder, A.J., Loy, A., Lozada, M., MacGregor, B., Magnabosco, C., Maria da Silva, A., McKay, R.M., McMahon, K., McSweeney, C.S., Medina, M., Meredith, L., Mizzi, J., Mock, T., Momper, L., Moran, M.A., Morgan-Lang, C., Moser, D., Muyzer, G., Myrold, D., Nash, M., Nesbø, C.L., Neumann, A.P., Neumann, R.B., Noguera, D., Northen, T., Norton, J., Nowinski, B., Nüsslein, K., O'Malley, M.A., Oliveira, R.S., Maia de Oliveira, V., Onstott, T., Osvatic, J., Ouyang, Y., Pachiadaki, M., Parnell, J., Partida-Martinez, L.P., Peay, K.G., Pelletier, D., Peng, X., Pester, M., Pett-Ridge, J., Peura, S., Pjevac, P., Plominsky, A.M., Poehlein, A., Pope, P.B., Ravin, N., Redmond, M.C., Reiss, R., Rich, V., Rinke, C., Rodrigues, J.L.M., Rossmassler, K., Sackett, J., Salekdeh, G.H., Saleska, S., Scarborough, M., Schachtman, D., Schadt, C.W., Schrenk, M., Sczyrba, A., Sengupta, A., Setubal, J.C., Shade, A., Sharp, C., Sherman, D.H., Shubenkova, O. V., Sierra-Garcia, I.N., Simister, R., Simon, H., Sjöling, S., Slonczewski, J., Correa de Souza, R.S., Spear, J.R., Stegen, J.C., Stepanauskas, R., Stewart, F., Suen, G., Sullivan, M., Sumner, D., Swan, B.K., Swingley, W., Tarn, J., Taylor, G.T., Teeling, H., Tekere, M., Teske, A., Thomas, T., Thrash, C., Tiedje, J., Ting, C.S., Tully, B., Tyson, G., Ulloa, O., Valentine, D.L., Van Goethem, M.W., VanderGheynst, J., Verbeke, T.J., Vollmers, J., Vuillemin, A., Waldo, N.B., Walsh, D.A., Weimer, B.C., Whitman, T., van der Wielen, P., Wilkins, M., Williams, T.J., Woodcroft, B., Woolet, J., Wrighton, K., Ye, J., Young, E.B., Youssef, N.H., Yu, F.B., Zemskaya, T.I., Ziels, R., Woyke, T., Mouncey, N.J., Ivanova, N.N., Kyrpides, N.C., Eloe-Fadrosh, E.A., 2020. A genomic catalog of Earth's microbiomes. Nature Biotechnology. doi:10.1038/s41587-020-0718-6

Ohkouchi, N., Chikaraishi, Y., Close, H.G., Fry, B., Larsen, T., Madigan, D.J., McCarthy, M.D., McMahon, K.W., Nagata, T., Naito, Y.I., Ogawa, N.O., Popp, B.N., Steffan, S., Takano, Y., Tayasu, I., Wyatt, A.S.J., Yamaguchi, Y.T., Yokoyama, Y., 2017. Advances in the application of amino acid nitrogen isotopic analysis in ecological and biogeochemical studies. Organic Geochemistry. doi:10.1016/j.orggeochem.2017.07.009

Oonk, S., Cappellini, E., Collins, M.J., 2012. Soil proteomics: An assessment of its potential for archaeological site interpretation. Organic Geochemistry. doi:10.1016/j.orggeochem.2012.06.012

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S., Hayes, D., 2011. A large and persistent carbon sink in the world's forests. Science.

doi:10.1126/science.1201609

Papp, K., Mau, R.L., Hayer, M., Koch, B.J., Hungate, B.A., Schwartz, E., 2018. Quantitative stable isotope probing with H218O reveals that most bacterial taxa in soil synthesize new ribosomal RNA. ISME Journal. doi:10.1038/s41396-018-0233-7

Paul, E.A., Clark, F.E., 1988. Soil microbiology and biochemistry. Soil Microbiology and Biochemistry. doi:10.2307/4003217

Pham, V.H.T., Kim, J., 2012. Cultivation of unculturable soil bacteria. Trends in Biotechnology. doi:10.1016/j.tibtech.2012.05.007

Pilegaard, K., 2013. Processes regulating nitric oxide emissions from soils. Philosophical Transactions of the Royal Society B: Biological Sciences. doi:10.1098/rstb.2013.0126

Pollierer, M.M., Langel, R., Scheu, S., Maraun, M., 2009. Compartmentalization of the soil animal food web as indicated by dual analysis of stable isotope ratios (15N/14N and 13C/12C). Soil Biology and Biochemistry. doi:10.1016/j.soilbio.2009.03.002

Prahl, F.G., Ertel, J.R., Goni, M.A., Sparrow, M.A., Eversmeyer, B., 1994. Terrestrial organic carbon contributions to sediments on the Washington margin. Geochimica et Cosmochimica Acta. doi:10.1016/0016-7037(94)90177-5

Pratscher, J., Dumont, M.G., Conrad, R., 2011. Ammonia oxidation coupled to CO2 fixation by archaea and bacteria in an agricultural soil. Proceedings of the National Academy of Sciences of the United States of America. doi:10.1073/pnas.1010981108

Rasche, F., Knapp, D., Kaiser, C., Koranda, M., Kitzler, B., Zechmeister-Boltenstern, S., Richter, A., Sessitsch, A., 2011. Seasonality and resource availability control bacterial and archaeal communities in soils of a temperate beech forest. ISME Journal. doi:10.1038/ismej.2010.138

Rumpel, C., Kögel-Knabner, I., Bruhn, F., 2002. Vertical distribution, age, and chemical composition of organic carbon in two forest soils of different pedogenesis. Organic Geochemistry. doi:10.1016/S0146-6380(02)00088-8

Ruppel, S., Torsvik, V., Daae, F.L., Øvreås, L., Rühlmann, J., 2007. Nitrogen availability decreases prokaryotic diversity in sandy soils. Biology and Fertility of Soils. doi:10.1007/s00374-006-0122-5

Schneider, T., Keiblinger, K.M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G., Roschitzki, B., Richter, A., Eberl, L., Zechmeister-Boltenstern, S., Riedel, K., 2012. Who is who in litter decomposition Metaproteomics reveals major microbial players and their biogeochemical functions. ISME Journal. doi:10.1038/ismej.2012.11

Simon, J., Klotz, M.G., 2013. Diversity and evolution of bioenergetic systems involved in microbial

nitrogen compound transformations. Biochimica et Biophysica Acta - Bioenergetics. doi:10.1016/j.bbabio.2012.07.005

Sparacino-Watkins, C., Stolz, J.F., Basu, P., 2014. Nitrate and periplasmic nitrate reductases. Chemical Society Reviews. doi:10.1039/c3cs60249d

Starke, R., 2017. Untersuchung der Energie-und Nährstoffflüsse mikrobieller Gemeinschaften. University of Hohenheim.

Starke, R., Jehmlich, N., Bastida, F., 2018. Using proteins to study how microbes contribute to soil ecosystem services: The current state and future perspectives of soil metaproteomics. Journal of Proteomics. doi:10.1016/j.jprot.2018.11.011

Starke, R., Kermer, R., Ullmann-Zeunert, L., Baldwin, I.T., Seifert, J., Bastida, F., von Bergen, M., Jehmlich, N., 2016. Bacteria dominate the short-term assimilation of plant-derived N in soil. Soil Biology and Biochemistry 96. doi:10.1016/j.soilbio.2016.01.009

Steffan, S.A., Chikaraishi, Y., Dharampal, P.S., Pauli, J.N., Guédot, C., Ohkouchi, N., 2017. Unpacking brown food-webs: Animal trophic identity reflects rampant microbivory. Ecology and Evolution. doi:10.1002/ece3.2951

Sterner, R.W., Elser, J.J., 2002. Ecological Stoichiometry: The Biology of Elements, in: Molecules to the Biosphere. Princeton University Press.

Štursová, M., Šnajdr, J., Koukol, O., Tláskal, V., Cajthaml, T., Baldrian, P., 2020. Long-term decomposition of litter in the montane forest and the definition of fungal traits in the successional space. Fungal Ecology. doi:10.1016/j.funeco.2020.100913

Štursová, M., Žifčáková, L., Leigh, M.B., Burgess, R., Baldrian, P., 2012. Cellulose utilization in forest litter and soil: Identification of bacterial and fungal decomposers. FEMS Microbiology Ecology. doi:10.1111/j.1574-6941.2012.01343.x

Sutton, M.A., Reis, S., Riddick, S.N., Dragosits, U., Nemitz, E., Theobald, M.R., Tang, Y.S., Braban, C.F., Vieno, M., Dore, A.J., Mitchell, R.F., Wanless, S., Daunt, F., Fowler, D., Blackall, T.D., Milford, C., Flechard, C.R., Loubet, B., Massad, R., Cellier, P., Personne, E., Coheur, P.F., Clarisse, L., Van Damme, M., Ngadi, Y., Clerbaux, C., Skjøth, C.A., Geels, C., Hertel, O., Kruit, R.J.W., Pinder, R.W., Bash, J.O., Walker, J.T., Simpson, D., Horváth, L., Misselbrook, T.H., Bleeker, A., Dentener, F., de Vries, W., 2013. Towards a climate-dependent paradigm of ammonia emission and deposition. Philosophical Transactions of the Royal Society B: Biological Sciences. doi:10.1098/rstb.2013.0166

Swift, M.J., Boddy, L., 1984. Animal–microbial interactions in wood decomposition, in: Anderson, J.M., Rayner, A.D.M., Walton, D.W.H. (Eds.), Invertebrate–Microbial Interactions. Cambridge University

Press, Cambridge, pp. 89–131.

Takasaki, K., Miura, T., Kanno, M., Tamaki, H., Hanada, S., Kamagata, Y., Kimura, N., 2013. Discovery of Glycoside Hydrolase Enzymes in an Avicel-Adapted Forest Soil Fungal Community by a Metatranscriptomic Approach. PLoS ONE. doi:10.1371/journal.pone.0055485

Tukey, J.W., 1949. Comparing Individual Means in the Analysis of Variance. Biometrics. doi:10.2307/3001913

Urabe, J., Watanabe, Y., 1992. Possibility of N or P limitation for planktonic cladocerans: An experimental test. Limnology and Oceanography. doi:10.4319/lo.1992.37.2.0244

Van Der Heijden, M.G.A., Bardgett, R.D., Van Straalen, N.M., 2008. The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology Letters. doi:10.1111/j.1461-0248.2007.01139.x

Van der Wal, A., Geydan, T.D., Kuyper, T.W., De Boer, W., 2013. A thready affair: Linking Fungal diversity and community dynamics to terrestrial decomposition processes. FEMS Microbiology Reviews. doi:10.1111/1574-6976.12001

Vitousek, P.M., Menge, D.N.L., Reed, S.C., Cleveland, C.C., 2013. Biological nitrogen fixation: Rates, patterns and ecological controls in terrestrial ecosystems. Philosophical Transactions of the Royal Society B: Biological Sciences. doi:10.1098/rstb.2013.0119

Von Bergen, M., Jehmlich, N., Taubert, M., Vogt, C., Bastida, F., Herbst, F.A., Schmidt, F., Richnow, H.H., Seifert, J., 2013. Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology. ISME Journal. doi:10.1038/ismej.2013.78

Wilmes, P., Heintz-Buschart, A., Bond, P.L., 2015. A decade of metaproteomics: Where we stand and what the future holds. Proteomics. doi:10.1002/pmic.201500183

Wong, D.W.S., 2009. Structure and action mechanism of ligninolytic enzymes. Applied Biochemistry and Biotechnology. doi:10.1007/s12010-008-8279-z

Žifčáková, L., Větrovský, T., Howe, A., Baldrian, P., 2016. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. Environmental Microbiology. doi:10.1111/1462-2920.13026

Žifčáková, L., Větrovský, T., Lombard, V., Henrissat, B., Howe, A., Baldrian, P., 2017. Feed in summer, rest in winter: microbial carbon utilization in forest topsoil. Microbiome. doi:10.1186/s40168-017-0340-0

Figure legends

Figure 1: Count of protein groups identified by at least one unique peptide in logarithmic scale (a). The relative abundance of ribosomal proteins and of proteins affiliated to CAZymes or nitrogen cycling (b). Data with the same letter across all habitats for the protein count and within one habitat for the relative abundance is not significantly different according to the HSD test ($P < 0.05$).

Figure 2: The relative abundance of proteins of the most abundant phyla (excluding plants) in the forest topsoil in the total and the ribosomal protein pool. Eukaryotic phyla are shown in shades of red and bacteria in shades of blue.

Figure 3: Dominant functional groups of proteins in the forest topsoil metaproteome grouped by habitats. The 19 most abundant KEGG categories for level 2 and the 24 most abundant for level 3 are shown. The categories accounted for on average 98.5% and 60.7% of all level 2 and level 3 functions, respectively. Sampling times and replicates were combined for each habitat. Data with the same letter within one row is not significantly different according to the HSD test ($P < 0.05$).

Figure 4: The relative abundance of CAZymes of the most abundant phyla (excluding plants) in the forest topsoil (a), and the relative share of CAZymes assigned to target molecules (b). Eukaryotic phyla are shown in shades of red and bacteria in shades of blue. Chitin, glucomannans and glucans represent compounds of fungal origin, while galactomannans, glucuroxylans, pectin, cellulose lignin, cellobiose and xylobiose are components of plant cell walls or their degradation products and peptidoglycan is a component of bacterial cell wall.

Figure 5: The relative abundance of proteins involved in nitrogen cycling of the most abundant phyla (excluding plants) in the forest topsoil (a), and their functional distribution (b). Eukaryotic phyla are shown in shades of red and bacteria in shades of blue.

Tables and table legends

Table 1: Chemistry of the litter, roots, and soil (each n=5).

	Litter	**Roots**	**Soil**
pH (H_2O)	3.51±0.05	n.d.	3.36±0.07
pH (KCl)	2.92±0.02	n.d.	2.82±0.06
N (%)	1.82±0.16	1.04±0.04	1.22±0.15

$N\text{-}NH_4^+$ (mg kg^{-1})	10.8±4.12	n.d.	9.73±2.68
$N\text{-}NO_3^-$ (mg kg^{-1})	5.83±1.17	n.d.	6.70±1.20
C (%)	50.3±3.19	41.23±0.94	29.2±5.42
C_{Org} (%)	47.4±2.94	n.d.	23.6±7.55
$C_{Carbonate}$ (%)	2.91±1.29	n.d.	5.60±2.45
C/N-ratio	27.7±3.18	39.64±0.96	23.9±5.44
Ca (mg kg^{-1})	504±114	n.d.	235±124
Mg (mg kg^{-1})	220±39.7	n.d.	116±30.5
K (mg kg^{-1})	490±138	n.d.	171±32.3

n.d. – not determined due to insufficient mass

Table 2: Number of measured proteins and proteins annotated by both a non-plant taxonomy with a bitscore lower than 135 and a function with an e-value below e^{-30} (each n=3). In total, 1 g of litter and roots as well as 2 g of rhizosphere and bulk soil were used for protein extraction, but a final concentration of each 1 µg $µL^{-1}$ extracted proteins was measured.

	Measured proteins		**Annotated proteins**	
	March	September	March	September
Litter	89,229±7,095	74,228±7,616	30,673±611	28,960±2,275
Roots	39,375±23,681	157,293±14,633	4,657±2,876	20,138±10,491
Rhizosphere	6,953±1,725	8,340±310	1,509±324	1,629±271
Soil	28,002±23,388	16,280±2,683	8,339±8,979	4,014±1,079

"I'll be there for you: Microbe-mediated resilience to abiotic stress in *Sorghum bicolor*"

Authors: Elle M. Barnes, Dawn Chiniquy, Kyle Hartman, Susannah G. Tringe

Background/Methods

Sorghum bicolor is a genetically diverse crop cultivated for a variety of agronomic uses, including grain, sugar, and energy production. However, cultivation of energy sorghum for biofuel production will require the use of marginal lands with potentially low nutrient availability and/or periods of water stress. All plants growing in soil harbor diverse communities of microbes that inhabit the areas in, on, and around their roots. Selected members of these microbial communities can provide benefits to their plant hosts, including direct growth promotion and conferring tolerances to abiotic and biotic stress. To explore a possible microbial solution to increase the nutrient use efficiency and resilience to water stress in sorghum, we used 16S rRNA sequencing, metagenomics, and SPIEC-EASI network analysis to survey the diversity, structure, and functional potential of sorghum bacterial communities. We report how drought, nitrogen deficiency, and plant genotype alter the sorghum microbiome throughout the growing season and correlate these changes with sorghum biomass.

Results/Conclusions

We found a significant effect of growing condition (N and water availability) on the alpha and beta diversity of the sorghum rhizosphere. However, the strength of the effect was dependent on sorghum genotype and location within the field. In addition to affecting microbial diversity and composition, network analysis suggested that growth condition also influenced the structure and co-associations of specific beneficial bacterial taxa. We subset the rhizosphere to focus on bacterial taxa identified as putatively plant growth promoting (PGPB). We found that sorghum dry weight and height were positively correlated with the relative abundance of PGPB, regardless of genotype or growth condition. However, the composition of PGPB varied by treatment. Additionally, abiotic stressed plants showed increases over time in the abundance of bacterial orthologous groups associated with root elongation, nitrogen fixation, siderophore biosynthesis, and plant hormone biosynthesis. Overall, our results suggest that under abiotic stress, sorghum individuals are able to recruit beneficial bacteria to their rhizosphere and that microbe-mediated interactions play a critical role in plant resilience.

2021 DOE JGI Genomics of Energy & Environment Meeting Abstract Submission

Jonelle T. Basso[1], Maria Dzunkova[1], Maureen Berg[1], Cristian Espinosa[2], Zhiying Zhao[1], Simon Roux[1], Benjamin Cole1 and Axel Visel[1],

(1) DOE JGI Berkeley Lab, Berkeley, CA, (2) UC Merced, Merced, CA

Abstract title: Assessment of viral influence on plant root colonization by plant growth promoting rhizobacteria

Abstract text:

Beneficial members of the plant microbiome can increase nutrient availability for their hosts, protect their hosts against pathogens, and enhance host resilience against abiotic stress. While previous and ongoing studies of the rhizosphere microbiome have been critical for assessing the impact of specific plant-microbe interactions, their focus has overwhelmingly targeted bacterial and fungal members of the microbiome. Viruses are ubiquitous, outnumbering all other biological entities on the planet, yet they are remarkably understudied in the rhizosphere. Prior work from our group identified functional roles for hundreds of genes in a plant growth promoting rhizobacterium *Pseudomonas simiae* that are important for its colonization of the rhizosphere. Two of these genes that cause reduced fitness in the rhizosphere when mutated are components of a latent bacteriophage, and are present among three phage loci ranging in size from 15-65kbp. We observed significant differences in bacterial cell lysis and fluorescently stained populations upon mitomycin c induction and flow cytometry respectively, between induced and control cultures over a 24-hour period. Taken together, these findings suggest the possibility that bacteriophages are involved in modulating the ability of bacteria to colonize plants. The quantitative impact of these phage genes on root colonization and the molecular underpinnings of this presumptive plant-bacterial-phage interaction are currently being investigated.

Evaluation of binning tools for microbial eukaryotes

Lucia Bazan-Williamson, Asaf Salamov, Igor Grigoriev

Abstract

Microorganisms are found almost everywhere on Earth; they play an essential role in biological processes and are present in nearly all environments. These microorganisms exist in communities, interact with plants, animals, and other microorganisms. Studying these communities is vital for understanding the ecological role they represent and their participation in the different biological processes. However, within these microbial communities it is often hard to separate one microorganism from another or cultivate them in a laboratory. *Metagenomics* studies microbial communities containing uncultivated microbes from diverse environments. To analyze these communities and find which organisms are present in a sample, it is necessary to bin metagenome sequences, or cluster metagenome assembly contigs into bins representing individual genomes. Most metagenomics studies primarily focus on prokaryotes due to complexity of eukaryotic genomes. We compared the performance of binning tools on eukaryote-containing metagenomes from three diverse environments (Agave, Populus, Lake Erie) to recover the full genomes of microbial eukaryotes. Five binning tools; MetaBAT2 (Kang, 2015), MaxBin2 (Wu, 2014), Binsanity (Graham, 2017), CONCOCT (Alneberg, 2014) & VAMB (Nissen,2021), have been tested on sequenced metagenomes for their ability to extract clean and complete eukaryotic bins based on BUSCO completeness tests, percentage of eukaryotic content, and genome/bin sizes.

Comparison of short-read and long-read metagenome assembly for viral genome recovery

Authors: Maureen Berg, Taylor Reiter, Russell Neches, Hannah Houts, christian Santos Medellin, Tessa Pierce, Jackson Sorensen, Preston Tasoff, Peter Turnbaugh, C. Titus Brown, Joanne Emerson, Simon Roux

Comparisons of long-read and short-read metagenome assemblies typically show that short-read sequence assemblies are less error prone, but struggle to assemble complicated genome regions (e.g. repeats) compared to long-read sequence assemblies. Viral genome assembly, particularly from environmental data, can be challenging as viral genomes tend to be significantly less abundant than their microbial counterparts, they can have significant proportions of repeat regions and hypervariable regions, and the higher mutation rates of some viruses can lead to increased strain variation within the community. Long-read sequencing can be challenging to perform for many environmental samples, since this method requires large amounts of input DNA, but preliminary data has shown that long-reads metagenome assembly could improve viral genome recovery in some cases. Here, we use metagenomic data with paired long-read and short-read sequences to identify specific factors that impact viral genome assembly, and assess their relative importance in a natural phage community. Our preliminary data suggests that coverage, coverage variation along the genome, and sequence diversity (i.e. “hypervariable” regions) are the three main factors leading to failures in short-read assemblies. In our soil viral community, coverage variation along the genome seemed to impact short-read genome assembly more than sequence diversity. Once refined and expanded to other sample types, such direct comparisons between short-read and long-read assemblies should help us collectively understand the “blind spots” of short-read metagenomics, and how large and/or critical these blindspots are.

The supragenic organization of glycoside hydrolase encoding genes reveals the distribution of polysaccharide utilization loci across sequenced bacterial genomes

R. Berlemont. PhD.

Department of Biological Sciences, California State University - Long Beach.

The identification of “polysaccharide utilization loci” (PULs) in *Bacteroides* highlighted the clustering of GH-genes targeting polysaccharide and transporters. However, GH-genes account 3.52% of the genes in *Bacteroides* (n=30) and only 0.95% in bacterial genomes (n=15,374). Thus, PULs in *Bacteroides* might result from stochastic and/or evolutionary processes. Indeed, across Bacteroidetes, the number of genes, GH-genes and the frequency PULs varied extensively. In this context, a major question is to understand the aggregation of GH-genes in microbial genomes, beyond the Bacteroides genus and the Bacteroidetes phylum.
Here, I first investigated the clustering of GH-genes in randomized bacterial chromosome of varying size and GH-gene content. These simulations provide a baseline to estimate the contribution of stochastic processes to the clustering of genes, in real genomes.
Next, I identified the 406,337 GH-genes in 15,822 “complete genomes” using GeneHunt. The clustering of GH-genes in sequenced bacterial genomes is phylogenetically conserved and reflect the overall conserved genome organization and the conserved distribution of GH-genes across sequenced bacterial genomes. Finally, focusing on genomes with many GH-genes, I investigated the association between GH-genes and transporter-genes (ABC, PTS, and SUS family). Using this approach, I first re-analyzed the Bacteroides genomes with many GH-genes being associated with genes encoding susCD transporters in polysaccharide utilization loci (PUL). Then, I identified phylogenetically conserved PUL-like structures across bacterial lineages. These clusters of genes might represent an evolutionary intermediate between scattered genes and operons and highlight the extreme specialization towards carbohydrates deconstruction in multiple lineages, beyond the Bacteroidetes phylum.

Mapping plant-mycorrhizal interactions with spatial transcriptomics and single-nuclei sequencing

Margot Bezrutczyk[1], Karen Serrano[2], Danielle Goudeau[1], Rex Malmstrom[1], Ronan O'Malley[1], Henrik V. Scheller[2], Axel Visel[1], Benjamin J Cole[1]

[1] Joint Genome Institute, Lawrence Berkeley National Lab, Berkeley, California, USA
[2] Joint BioEnergy Institute, Emeryville, California, USA

The interaction of plants with arbuscular mycorrhizal fungi (AMF) is both ancient and widespread. In this symbiotic relationship, plants (including the legume model, *Medicago truncatul*a) provide AMF with carbon in exchange for phosphorous, nitrogen and water, making this interaction a prime target for crop improvement. Despite decades of research, we still lack a comprehensive understanding of all of the molecular features involved in establishing and maintaining AMF symbiosis. This is in part due to the difficulty of working with a biological system that only occurs between specific cell types of plant roots, precluding many powerful functional genomics strategies. To overcome these hurdles, we are piloting the use of two recently-developed and highly complementary transcriptomics methods: single-nuclei sequencing and spatial transcriptomics, to create a two-dimensional integrated map of the transcriptomes of both plant and fungal cells during colonization. Furthermore, we are applying this transformative technology to *M. truncatula* mutants to determine critical genetic pathways that can be exploited to engineer crops, with the goal to convert more atmospheric carbon to useable plant biomass while reducing the need for chemical fertilizer applications.

Extreme solutions to extreme problems: studies of life at the edge using Cyanidiophyceae red algae - Debashish Bhattacharya, Timothy McDermott, Julia Van Etten, Christa Pennacchio, Katherine Louie, Benjamin Bowen, Trent Northen,

Harsh environments with extremes of temperature, pH, solar radiation, salt concentration, and pressure produce exquisitely tailored solutions by resident biota. Bacteria and Archaea are the usual masters of life at the extreme, and until recently, algae that have evolved under similar conditions, have been largely overlooked. We are studying the Cyanidiophyceae, a group of unicellular aquatic and terrestrial red algae (e.g., Galdieria, Cyanidium, and Cyanidioschyzon) that occupy a variety of hot springs and acid mining sites characterized by variable light levels, high temperature, low pH, with high salt and toxic heavy metal concentrations. Work by our group has shown that the Cyanidiophyceae follow the “1% rule”; i.e., on average, approximately 1% of their gene inventory comprises prokaryote genes acquired via horizontal gene transfer (HGT). We have also demonstrated that a majority of HGT candidates encode proteins with functions related to “polyextremophily”, including metal and xenobiotic resistance/ detoxification, cellular oxidant reduction, carbon metabolism, amino acid metabolism, osmotic resistance, and salt tolerance. These traits make Cyanidiophyceae ideal models for achieving DOE objectives, for example, by engineering resistance genes and pathways into commercially important algae and plants to protect them from environmental stresses such as drought and heavy metal contamination. Other applications include alga-based production of compounds such as phycocyanin, floridosides, and glycogen, recovery of rare earth elements, and detoxification of heavy metals. In this presentation, I will discuss our JGI-CSP project goals and preliminary data derived from metabolomic analysis of Cyanidiophyceae from Yellowstone National Park (YNP). Specifically, we are generating Illumina short-read metagenomic data for a time series at 3 distinct hot springs habitats at Lemonade Creek, YNP, Illumina metatranscriptomic, and untargeted and targeted polar metabolomic data. The metabolomic data are critical to gain a functional perspective on the impact of HGTs and will be included in multipartite networks, with other omics data, to understand how novel proteins are integrated into existing pathways or give rise to novel pathways. --

USE OF TELL-SEQ TECHNOLOGY FOR DE NOVO ASSEMBLY OF MICROBIAL GENOMES

BACKGROUND: Microbial genomes exhibit varying sequence composition, often containing repetitive or low-complexity regions that can complicate *de novo* genome assembly. Short reads characterize the genomes efficiently when reference genomes are available, but when sequencing genomes for the first time some regions can be difficult, leading to gaps in otherwise contiguous assemblies. Using clonally barcoded beads, TELL-Seq™ technology allows reads originating from the same long parent template molecule to be associated informatically, thereby bridging intervening gaps in sequence.

METHODS: Eight microbial species with genomic GC content ranging from 28% to 69% were chosen for TELL-Seq™ library preparation. Genomic DNA was extracted from lyophilized cell pellets (ATCC) using the MagAttract kit (Qiagen) according to the manufacturer's instructions, yielding gDNA fragments with an average length of ~30kb. 0.5ng of gDNA input was used for each prep and bead input into PCR was proportional to genome size, which ranged from 1.6 to 5.4 Mb. All eight libraries were sequenced in a single MiSeq run and reads were downsampled to achieve a mean coverage of 100X for each genome. Sequence data was processed and assembled with the TELL-Read and TELL-Link software (Universal Sequencing Technology), respectively.

RESULTS: All library products showed the expected fragment size distribution regardless of GC content. Yields decreased at >65% GC but were still sufficient for multiple sequencing runs. A total of 1-3M read pairs was needed to achieve 100X coverage of each microbial genome. TELL-Read analysis indicated input parent template molecule lengths of 15-28kb. >96% of each genome was captured and TELL-Link assembly gave very large contigs and NG50 values that approached the size of the largest chromosome present in each genome, and misassembly was limited in most organisms.

CONCLUSIONS: While Illumina standard WGS sequencing solutions are more than adequate for isolate resequencing and draft *de novo* assembly, TELL-Seq™ technology enables near complete *de novo* assembly of microbial genomes from short read sequences using sub-nanogram inputs of genomic DNA. Optimizing gDNA extractions to increase input fragment lengths and removal of fragments <10kb is expected to further improve performance. Inaccurate input quantitation can contribute to decreased performance, as was observed for *B. cereus*, and highly repetitive genomes such as that of *B. pertussis* may be more prone to misassembly.

Non-coding RNA genes from Fungi

Chris Brown 1,6, Chun Shen Lim 1, Emir Padilla 1,2 ,Teyuan Chyou 1, Brooke Weinstein 3,4, Andy Nilsen 5, Tina Summerfield 5, David Orlovich 5,6 , Scott Roy 3,4

1. Biochemistry, University of Otago, Dunedin, New Zealand
2. Centro de Ciencias Genómicas , Mexico City, Mexico
3. Quantitative & Systems Biology, School of Natural Sciences, University of California-Merced, Merced, CA, USA
4. Department of Biology, San Francisco State University, San Francisco, CA, USA
5. Botany, University of Otago, Dunedin, New Zealand
6. JGI project 'Fungal evolution of truffle like bodies'

Fungi are a diverse group of eukaryotes, ranging from multicellular mushrooms to unicellular yeasts. These Fungi are divided into two major phyla: ascomycetes, which includes *S. cerevisiae,* and basidiomycetes, which includes mushrooms, that diverged about a billion years ago. Fungi play important roles in the biosphere and in biotechnology, and are a focus of JGI research. We have been analysing their genomes, as part of studies aimed at understanding their evolution and development.

Surprisingly, non-coding RNAs have received little attention in multicellular fungi (e.g. mushrooms). Very few are annotated in Mycocosm or Ensembl Fungi. In contrast, in other eukaryotes intensive study has identified key ncRNA roles, particularly in development and regulating protein coding gene expression.

Important classes of noncoding RNAs are: small RNAs (e.g. miRNAs, tRNAs); intermediate sized RNAs (e.g. small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs)); and long noncoding RNAs (e.g. rRNA, or long ncRNAs involved in development).

We are using comparative genomic, transcriptomic, and ribosome profiling data to discover ncRNAs. We aim firstly to identify the common ncRNAs involved in the core processes of transcription and translation. We have found that fungal kingdom or clade specific covariation models are required for this.

ncRNA genes are commonly found in intergenic regions or in introns. In our recently published study, we analysed introns across 263 diverse fungal species (Lim et al 2021). Overall, the data indicate there has been massive intron loss in multiple intron-poor fungal clades. For ncRNAs, the two most conserved introns, whose positions are retained for about a billion years, encode snoRNAs that are involved in gene expression.

This work provides avenues for identifying and annotating core ncRNAs and new functional RNAs, particularly in multicellular fungi.

Lim, C. S., Weinstein, B. N., Roy, S. W., & Brown, C. M. (2021). Analysis of fungal genomes reveals commonalities of intron gain or loss and functions in intron-poor species. *Molecular Biology and Evolution*. doi:10.1093/molbev/msab094

Multi-omics investigation of abiotic stress responses in algal biofuel strains

S. Calhoun[1], T.A.S. Bell[2], C. Kruse[4], B. Kamel[1], K. Louie[1], D. Treen[1], B. Bowen[1], T.R. Northen[1], M. Huesemann[5], S.R. Starkenburg[4], I.V. Grigoriev[1,6]
[1]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley CA
[2]University of Montana, Missoula, MT
[3]Los Alamos National Laboratory, Los Alamos, NM
[4]Pacific Northwest National Laboratory, Richland, WA
[5]Department of Plant and Microbial Biology, University of California-Berkeley, Berkeley, CA

Microalgae are of interest for applications in biofuel production due to their ability to convert carbon dioxide into lipids. During outdoor cultivation, algae are exposed to environmental changes, but the underlying mechanisms that control and regulate physiological responses and adaptation to environmental pressures are largely unknown. Systems-level characterization enabled by functional genomics can help identify biochemical pathways that promote stability and productivity of algae. Screening of candidate production strains has identified top performers in lab-based cultivation simulations and outdoor testbed facilities, including *Scenedesmus* sp. NREL 46B-D3 and *Monoraphidium minutum* 26B-AM, which showed robust growth in the summer and winter screens, respectively. We applied our multi-omics pipeline to profile these high potential strains under temperature and salt perturbations. Metabolomics and transcriptomics measurements were taken as a time course on samples grown under environmental perturbations. We observed that osmolytes, such as trehalose, proline and betaine, increase in abundance under salt stress, coinciding with upregulation of genes involved in proline biosynthesis and antioxidant processes. Through co-expression analysis, transcription factors with correlated expression to differentially expressed metabolism genes were identified as potential target genes for engineering. This multi-omics analysis provides a foundation for strain improvement in biofuel applications and expands our general understanding of metabolic and regulatory mechanisms in algae.

Title: **Enriched root bacterial microbiome in invaded vs native ranges of the model grass allotetraploid *Brachypodium hybridum***

Brooke Pickett[1,2], Chelsea J. Carey[2,3], Keshav Arogyaswamy[2], Jon Botthoff[4], Mia Maltz[2,4], Pilar Catalán[5,6,7], and Emma Aronson[2,4]

1) Division of Biological Sciences, University of California San Diego, San Diego, USA.
2) Department of Microbiology and Plant Pathology, University of California Riverside, Riverside, USA.
3) Point Blue Conservation Science, Petaluma, USA.
4) Center for Conservation Biology, University of California Riverside, Riverside, USA.
5) Departamento de Ciencias Agrarias y del Medio Natural, Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Spain.
6) Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Spain.
7) Tomsk State University, Tomsk, Russia.

e-mail:

Brooke Pickett: brookepic22@gmail.com

Chelsea J. Carey: ccarey@pointblue.org

Keshav Arogyaswamy: arogyaswamy@gmail.com

Jon Botthoff: botthoff@ucr.edu

Mia Maltz: maltz@ucr.edu

Pilar Catalán: pcatalan@unizar.es

Emma Aronson: emmaa@ucr.edu

Invasive plant species are disrupting natural areas around the world. Invasive species can shift the composition of key soil microbial groups, thus creating novel soil microbial communities. To better understand the biological drivers of invasion, we studied plant-microbial interactions in species of the *Brachypodium distachyon* complex, a model system for functional genomic studies of temperate grasses and bioenergy crops which contain plants native to the Mediterranean region including the allotetraploid invader *B. hybridum*. While *Brachypodium hybridum* invasion in California is currently in an incipient stage, threatening natural and agricultural systems, its diploid progenitor species *B. distachyon* is not invasive in California. We investigated the root, soil, and rhizosphere bacterial composition of *Brachypodium hybridum* in both its native and invaded range, and of *B. distachyon* in the native range. We

used high-throughput, amplicon sequencing to evaluate if the bacteria associated with these plants differ, and whether biotic controls may be driving *B. hybridum* invasion. Bacterial community composition of *B. hybridum* differed based on provenance (native or invaded range) for root, rhizosphere and bulk soils, as did the abundance of dominant bacterial taxa. *Brachypodium hybridum* roots were more diverse in the invaded range, as compared to roots from the native range. Root bacterial composition differed between *B. hybridum* and *B. distachyon* in the native range, while rhizosphere soil and bulk soil composition was equivalent. Bacteroidetes, Cyanobacteria and *Bacillus* spp. were significantly more abundant in *B. hybridum* roots from the invaded range, whereas Proteobacteria, Firmicutes, *Erwinia* and *Pseudomonas* were more abundant in the native range roots. The Cyanobacterial symbiosis may occur in drier invaded habitats, whereas root-associated pathogen bacteria may be lower within the invaded range, facilitating the invasiveness of *B. hybridum*. *Brachypodium hybridum* forms novel biotic interactions with a diverse suite of rhizosphere microbes from the invaded range, which may not exert a similar influence within its native range, ostensibly contributing to *B. hybridum's* invasiveness. These associated plant microbiomes could inform future management approaches for *B. hybridum* in its invaded range. These dynamics could be key to understanding, predicting, and preventing future plant invasions.

Keywords: Endophytic bacteria, *Brachypodium*, Invasive species, Rhizosphere, Root, Soil
Conference: Botany 2021 (http://2021.botanyconference.org/)
Presentation type: Poster
Topic: Symbioses: Plant, Animal, and Microbe Interaction
http://2021.botanyconference.org/engine/search/656.html

Fungi are what they secrete: evolution of zygomycete secretomes and the origins of terrestrial fungal ecologies

Ying Chang[1,2], Yan Wang [3,4], Sabrina Blankenship[1], Jeff Beard[1], Kyle A. Gervers[1], Alessandro Desiro[5], Gregory Bonito[5], Nicole Reynolds[6], Matthew E. Smith[6], Gerald L. Benny[6], Igor V. Grigoriev[7,8], Kerrie Barry[7], Kerry O'Donnell[9], Jason E. Stajich[10,11], Joseph W. Spatafora[1]

[1] Department of Botany and Plant Pathology, College of Agricultural Sciences, Oregon State University, Corvallis, Oregon, USA. [2] Life Sciences, Division of Science, Yale-NUS college, Singapore. [3] Department of Biological Sciences, University of Toronto Scarborough, Toronto, Canada. [4] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada. [5] Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, Michigan, USA. [6]Department of Plant Pathology, University of Florida, Gainesville, Florida, USA. [7] Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [8]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA. [9]USDA, Agricultural Research Service, National Center for Agricultural Utilization Research, Mycotoxin Prevention and Applied Microbiology Research Unit. 1815 N. University, Peoria, IL, USA.[10] Department of Microbiology and Plant Pathology, University of California Riverside, Riverside, CA, USA. [11]Institute for Integrative Genome Biology, University of California, Riverside, CA, USA.

Fungi survive in diverse ecological niches by secreting various proteins and other molecules to the extracellular environment to acquire food, to compete with other microorganisms and defend against various biotic and abiotic stresses. Fungal secretome content is therefore believed to be tightly linked to fungal ecologies. Representing the earliest branches of terrestrial fungi, the zygomycete fungi (Mucoromycota and Zoopagomycota) took two different trajectories, with the former adapted to plant hosts and substrates and the latter to non-plant hosts and substrates. Characterizing the secretome content of zygomycete fungi and reconstructing the evolution of these secretomes can improve our understanding of the evolution of fungal ecologies associated with the origins of terrestrial fungi and help us explore the ecological potentials of the extant zygomycete fungi.

In this study, we sampled 132 zygomycete fungal genomes and performed analyses on the secreted proteins, with a focus on digestive enzymes, putative small secreted proteins (SSPs), and other proteins hypothesized to function in the environment. Our analyses revealed that phylogeny played an important role in shaping the secretome composition of zygomycete fungi while trophic mode contributed for a smaller part. Reconstruction the evolution of secreted digestive enzymes revealed lineage-specific expansions, including the diversification of plant-cell-wall-degrading-enzymes (PCWDEs) in Mucoromycotina, chitinases and peptidases in Mortierellomycotina, and independent expansions of chitinases and proteases within Zoopagomycota. We identified the presence of multiple pathogenicity-related SSPs and other proteins in lineages known to have saprotrophic ecologies, suggesting that either the ecologies of these fungi are incompletely known, or that these pathogenicity-related SSPs are involved in important functions associated with saprotrophic ecologies, or both.

In previous work by Hess and colleagues, thousands of putative cellulase genes were identified from cow rumen microbiome. In this project, we explored the cellulase activity of two confirmed cellulase enzymes encoded by genes CJD5-106 and CJD5-110. We are interested in exploring different variables that can affect their cellulase activity. Previous research on cellulases found that increased temperature and presence of cobalt ion can lead to increased activity in some cases. In two different experiments, we tested the effects of cobalt and higher incubation temperatures on these two enzymes (CJD5-106 and CJD5-110) respectively. We quantified the cellulase activity using a DNS (3,5-Dinitrosalicylic acid) assay. With our results, we hope to gain insight into the ways CJD5-106 and CJD5-110 can be optimized for maximum cellulase activity in the application of biofuels production.

Molecular strategies for adaptation by legumes to toxic environments (CSP NI 506073)

Michael Clear[1]*, Aditi Bhat[1], M. Mercedes Lucas[2], Kerrie Barry[3], Anna Lipzen[3], Matt Blow[3], Jeremy Schmutz[3], José J. Pueyo[2] ,Tim Paape[1]

*presenting author

[1]Brookhaven National Laboratory, Upton, NY, USA

[2]Department of Soil, Plant and Environmental Quality, Institute of Agricultural Sciences, ICA-CSIC, Madrid, Spain

[3]Joint Genome Institute/Lawrence Berkeley Laboratory, Berkeley, CA USA

Legumes form symbioses with nitrogen-fixing bacteria which allows them to replenish nitrogen in marginal soils. Some marginal soils have high concentrations of heavy metals due to contamination and soil degradation. Cadmium (Cd) and mercury (Hg) are two of the most toxic heavy metals to animals, plants, and bacteria. We previously tested a large collection of the legume species *Medicago truncatula* (230 plant genotypes) for tolerance to Cd and Hg to identify high and low tolerant genotypes. To better understand the molecular mechanisms of heavy metal tolerance in legumes, we analyzed transcriptional changes in leaf and root tissues following exposure to Cd or Hg contrasting plant genotypes that had low or high tolerance to either heavy metal. We observed higher numbers of differentially expressed genes (DEGs) in response to Cd compared to the Hg treatment with more DEGs in root versus leaf tissues. GO enrichment analysis showed that glutathione transferase, heme binding, and oxioreductase activity were consistently overrepresented in response to either metal treatment. We identified a core set of 131 genes that were induced in highly metal tolerant strains and were lowly expressed in less tolerant strains. Genes associated with glutathione and sulfur assimilation were among the most highly differentially expressed genes within this core set. Because heavy metal tolerance, transport, and accumulation are complex processes involving many genes (i.e., polygenic), we are conducting co-expression network analysis to identify modules that are enriched for stress and transport related functions. This study will provide insights into detoxification mechanisms that are necessary for successful symbiosis. Our findings will also demonstrate macroevolutionary evidence about whether metal tolerance is largely conserved across plant families by their use of common orthologous metal homeostasis genes and detoxification pathways.

First Author Email
LECote@lbl.gov

First Author
Laleh E. Coté

First Author's Affiliations
Lawrence Berkeley National Laboratory, University of California-Berkeley

Second Author
Anne M. Baranger

Second Author's Affiliations
University of California-Berkeley

Presentation Title
Career impacts of the COVID-19 crisis on life science, health, environmental science, and geoscience students and professionals

Abstract (320 words)
The science, technology, engineering, and mathematics (STEM) workforce is critical to solving the large-scale multi-faceted problems faced by society today. Considering the emergence and spread of the new respiratory Coronavirus Disease 2019 (COVID-19) across the U.S. (and globally), it is inevitable that COVID-19 will impact the academic and professional trajectories of our STEM workforce at every stage. This study explored the ways in which the COVID-19 crisis impacted students and professionals studying and/or working in life sciences, health, environmental science, and geoscience. Survey data was collected in 2020 and 2021, and includes 1,284 survey responses from across the U.S. The respondents come from different institution types, career stages, and personal backgrounds. Many experienced challenges associated with being separated from family and friends. Others explicitly connected their decline in mental health to COVID-19-related circumstances, commonly citing isolation while living alone, being separated from family and friends, increased work load, and balancing work with childcare at home. 74% of respondents reported that their academic or professional plans had been changed and/or altered, and 87% of respondents reported that their current or future projects had been changed and/or altered by the COVID-19 crisis. Many described the cascading effects of laboratory and/or field site shut-downs, which led to decreased productivity, delayed publication, hindered collaborations, reduction in staff, slowed training, and delayed graduation for students who were unable to complete their projects as planned. Some described pivoting their research to include projects related to SARS-CoV-2 and/or COVID-19, while others shifted to the use of theoretical or computational methods and/or previously collected data. Across all career levels, people are reflecting on their choice of profession or area of focus, some of whom are considering shifting to a different pathway altogether. This study will

include some discussion of the needs and desires of those who have described feeling insufficiently supported by their institution, department, program, group, peers, etc.

A Constitutive Stress Response is an Adaptation to Low Temperature in the Antarctic green alga *Chlamydomonas* sp. UWO241

Marina Cvetkovska, Xi Zhang, Beth Szyszka-Mroz, David R. Smith, Norman P.A. Hüner

The Antarctic green alga *Chlamydomonas* sp. UWO241 is an obligate psychrophile that thrives in the depths of the Antarctic Lake Bonney at permanently low temperature (4-6°C) but is unable to survive at temperatures ≥18°C. Little is known how exposure to heat affects its physiology or whether it mounts a heat stress response in a manner comparable to mesophiles. Here, we dissect the responses of UWO241 to temperature stress by examining its growth, primary metabolome and transcriptome under steady-state low temperature and heat stress conditions. UWO241 exhibits slow growth at 4°C, a temperature closest to its natural habitat, and faster growth at higher temperatures of 10-15°C. In comparison with *Chlamydomonas reinhardtii*, UWO241 constitutively accumulates metabolites and proteins commonly considered as stress markers under a range of growth temperatures, including soluble sugars, antioxidants, polyamines, and heat shock proteins to ensure efficient protein folding at low temperatures. We propose that this permanent stress metabolism is an adaptive advantage to life at extreme conditions. Despite low growth rates, 4°C-grown UWO241 cultures retained the capacity to respond to heat stress and accumulated increased amounts of stress-related compounds, when compared to cultures grown at 10-15°C. Thus, growth of UWO241 at higher permissive temperatures (10°C and 15°C) does not provide enhanced heat protection. UWO241, however, fails to induce the accumulation of HSPs when exposed to heat, unlike the the mesophile *C. reinhardtii* exposed to comparable conditions. We suggest that the inability of UWO241 to fine-tune its response to heat its one of the reasons behind its sensitivity to increased temperatures. Our work adds to the growing body of research on temperature stress in psychrophiles, many of which are threatened by climate change.

“Free-living Soil Diazotrophs Provide Nitrogen to Pastures after Land Use Conversion from Amazon Rainforest”

Author: Rachel Danielson, Kyle Meyer, Jordan Sayre, Klaus Nuesslein, Brendan Bohannan, Jorge Rodrigues

The Amazon Rainforest is a global diversity hotspot and crucial carbon sink, although around 20% of its total extent has been deforested and replaced with monocultures of exotic bunchgrass for cattle pasture. Understanding the impact of this large-scale disturbance on soil microbial community composition and activity is crucial to understand consequential shifts in nutrient or greenhouse gas cycling, as well as adding to the body of knowledge concerning how these complex communities respond to human disturbance. In this study, we sought to determine the impact of rainforest conversion on microbial communities involved in asymbiotic nitrogen fixation. Surface soils (0-10cm) from three rainforests and three adjacent pastures were collected in Rondônia, Brazil, the Amazon state with the highest rate of deforestation. Soil chemical and physical parameters were paired with measurements of microbial activity and genetic profiles to determine how community composition and process rates relate to environmental conditions. Measuring both the natural abundance of 15N in total soil N, as well as incorporation of enriched 15N2 under incubation has revealed that conversion of primary forest to cattle pasture results in a significant 47x increase in the rate of nitrogen fixation by free-living diazotrophs. Quantitative PCR of a gene encoding the nitrogenase reductase enzyme correspondingly reveals a 23x increase of genes in pasture compared to forest soils, and shows significant correlation with asymbiotic fixation rates (ρ = 0.66). Additionally, genetic sequencing of both nifH genes and transcripts shows a significant increase in the diversity of present (3.31x) and transcriptionally-active (2.12x) diazotrophs within pasture soil communities. Additionally, principal components of DNA-based community ordination constrained by land use type (ρ = 0.79). With respect to soil chemical measurements, levels of both organic and inorganic nitrogen tended to be lower in pastures compared to forests, with ammonium rather than nitrate as the dominant inorganic form. However no significant or consistent differences in total, extractable, permanganate-oxidizable, or loss-on-ignition carbon are present between the two land use types. Forest conversion was associated with a 0.5-1.0 unit pH increase, but concentrations of many biologically-relevant nutrients such as phosphorus do not increase consistently. Crucially, molybdenum, a key micronutrient for nitrogen fixers, had lower availability in pasture soils. Considered together, this study indicates a significant change in nitrogen cycling and diazotroph community composition with the conversion of Amazon Rainforest. This may have important implications for the sustainability of cattle pastures once established, since nitrogen is a crucial nutrient for forage grass productivity.

“

“The Plasminogen-Apple-Nematode (PAN) domain suppresses plant immunity”

Authors: Kuntal De, Debjani Pal, Carly Shanks, Tim Yates, Kai Feng, Sara Jawdy, Wellington Muchero

Abstract:

Immunosuppression is a phenomenon that can block the ability of host natural defense. It is involved in precise gamete fertilization, cell growth, and proliferation during development. Additionally, the fine-tuning of immunosuppression in host-pathogen interaction in a different divergent organism is incredibly essential. Here, we show for the first time that the Plasminogen-Apple-Nematode (PAN) domain of the G-type lectin receptor-like kinases is essential for immunosuppression in plants. Hormonal pathways involving jasmonic acid and ethylene are critical for plant defense against microbes, necrotrophic pathogens, parasites and insects. We identified two Salix purpurea G-type lectin receptor kinases that were able to trigger defense signaling via jasmonic and ethylene signaling pathways upon expression in Arabidopsis and tobacco. These receptors accumulated mutations in conserved amino acid residues of the PAN domain. Restoration of these mutations to their conserved state impaired defense signaling as evidenced by receptor degradation and diminished MAPK phosphorylation. Further, we demonstrated that the domain is prerequisite for oligomerization, ubiquitination and proteolytic degradation of these receptors. Collectively, our results suggest that ubiquitination and proteolytic degradation mediated by the PAN domain plays a role in receptor turn-over to impair jasmonic acid and ethylene defense signaling in plants.

The physiological and transcriptional control of growth rate variation in *Brachypodium*

David L. Des Marais, Caio Guilherme-Pereira, and Marjorie R. Lundgren

Plant growth rate is a complex trait that reflects a balance between resource acquisition and allocation. We exploited growth rate diversity within and between species of the model grass genus *Brachypodium* to identify the gene regulatory architecture of carbon assimilation, partitioning, and developmental allocation. First, we use diversity panels in annual *B. distachyon* and perennial *B. sylvaticum* to show that growth rate variation is primarily controlled by specific leaf area and the balance between rates of carbon assimilation and respiration. Genetic correlations between these complex traits and major axes of gene expression variance suggest candidate regulatory pathways involved in determining growth rate. We next validate our regulatory hypotheses by directly manipulating the carbon available for growth in a series of growth chamber manipulations, with the aim of perturbing carbon source and sink dynamics in these two species. We find that annual *B. distachyon* is carbon sink-limited, as an increase in the rate of carbon assimilation does not translate into additional biomass. Conversely, perennial *B. sylvaticum* appears to be source-limited, as it readily increased tillering and thus whole plant growth rate in response to the extra source of carbon. These manipulations reinforce the role played by assimilation rate and carbon partitioning in controlling growth rate and point towards core metabolic processes driving growth rate variation, particularly with respect to key bioenergy traits including tillering and root biomass production. Through the course of these analyses, we generated new genomic resources for *Brachypodium*, providing a powerful system to study the genetic basis of functional variation within and between annual and perennial grass species.

“Genomics-enabled exploration of the algal HGT landscape through time and space”

Authors:
Richard Dorrell, Alan Kuo, Benoit Perez-Lamarque, Adrien Villain, Zoltan Fussy, Federico Ibarbalz, Elisabeth Richardson, Nikola Zarevski, Helene Morlon, Guillaume Blanc, Connie Lovejoy, Chris Bowler, Igor Grigoriev

Abstract:
Horizontal gene transfer (HGT), or the non-sexual movement of genetic material between cells, has played important roles in both the early and recent evolution of microbial eukaryotes; however, its overall dynamics, taxonomic biases, and associated functional contributions to eukaryotic evolution remain poorly understood, in part due to a paucity of sequenced genomic information from key eukaryotic lineages. In this short talk, I will demonstrate how comparative genomic, and environmentally-enabled phylogenomic techniques, taking advantage of the rapid expansion in particular of eukaryotic algal genomes and transcriptomes available, can be used to transform our understanding of who, when, where, and why HGT occurs, focussing particularly on the evolution of eukaryotic algal groups with chloroplasts of “secondary” or eukaryotic endosymbiotic origin.

In one project, my group has used the manual and automated resolution of thousands of single-gene trees, alongside experimental characterisation, to trace the dynamic patterns of HGT within the ochrophytes, a major group of eukaryotic algae encompassing diatoms, diverse small-sized flagellates such as pelagophytes and chrysophytes, and kelps. We have demonstrated that the ochrophyte HGT landscape is defined by two contrasting signals: eukaryotic HGTs into and out from the ochrophytes with eukaryotic algal partners, which reveal a complex origin of ochrophyte and other algal chloroplasts via secondary and higher endosymbiosis; and a continuous wave of bacterial HGT into the ochrophytes that has predominantly served to remodel the ochrophyte secretome. Our data illustrate the complexity, timing, and functions of the algal HGT landscape, alongside identifying new candidate genes for phenotypic characterisation.

In a second project, we have studied the distinct patterns of HGT in distantly related small flagellated algae native to the Arctic Ocean. Profiting from JGI genome assemblies of an Arctic cryptomonad, haptophyte, chrysophyte, and pelagophyte, alongside environmental sequence assemblies from the Tara Oceans expedition, we show that distantly related algae native to the Arctic have converged on similar genome coding contents, independent of their underlying phylogenetic distance, and this convergence is mediated in part by within-oceanic HGT of at least 35 gene families with Arctic-adaptive functions, typified by ice-binding proteins. Our data serve to remodel our understanding of HGT from a phylogenetically- to biogeographically-structured process, and provide the first glimpse into the genomic diversity of an ecologically important and environmentally fragile marine biome.

From trophic cascade to mutualism: ecological genomics rectifies the concept of *Trichoderma* – plant interactions

Feng Cai[1], Andrei Steindorff[2], and Irina S. Druzhinina[1]

[1] Fungal Genomics laboratory (FungiG), the Key Laboratory of Plant Immunity, Nanjing Agricultural University, Nanjing, China
[2] US Department of Energy Joint Genome Institute (JGI), Berkeley, CA, USA

What happens to plants if there would be no *Trichoderma*? We addressed this question by integrating the evolutionary genomics, genus-wide taxonomic profile, and metadata analysis on *Trichoderma* interactions in nature.

The species-rich mycoparasitic fungal genus *Trichoderma* (Hypocreales, Ascomycota) contains outstandingly plant-beneficial species used as commercial bioeffectors. Consequently, *Trichoderma* is frequently referred to as a *mutualistic symbiont of plants*, an *ectomycorrhizal partner*, a *commensal,* or a *hyperparasite* (parasite of plant parasites). However, the precise ecological specification of *Trichoderma* – plant interactions in nature is different compared to agricultural conditions and remains cloudy.

Ecophysiological profiling and the genus-wide metadata analysis documented the generally poor growth in soil, wood phytosaprotrophy, and multifaced direct interactions with taxonomically diverse fungi (*mycoparasitism, ammensalism,* and *competition*). *In vitro* studies of direct interactions with plants pointed to *commensalism* as *Trichoderma* frequently remains neutral. However, some *Trichoderma* species get isolated as facultative endophytes and thus form putative *mutualistic* interactions with plants. In line with this, the addition of *Trichoderma* inoculants (11 species) resulted in a significant (p <0.05) increase of the *Solanum lycopersicum* (tomato) biomass and root growth by 9-44% and 36-118%, respectively. The evolutionary genomics study revealed that the origin of the genus coincided with the Cretaceous-Paleogene extinction event 66.5 MYA when it diverged from the ancestor shared *Hypomyces* and *Escovopsis* and showed that all these mycophagous genera were closely related to entomopathogenic Cordycipitaceae. Thus, this analysis explained mycoparasitism as the innate property of *Trichoderma*, while it did not uncover the nature of interactions with plants. The subsequent evolutionary analysis of plant cell wall degrading CAZymes (pcwdCAZymes) revealed that the formation of the genus was accompanied by the substantial enrichment of the initially limited pool of these enzymes through the massive lateral gene transfer from closely related Pezizomycotina fungi. However, numerous pcwdCAZymes families remained underrepresented in genomes making *Trichoderma per se* a still poor colonizer of plants but a phytosaprotroph. We then looked at the diversity of the effector-like proteins (cerato-platanins, hyphosphere proteins, and hydrophobins) that play a role in the plant pathogenicity of many fungi and found that despite some of them were slightly phytotoxic, their evolution in Hypocreales was explained by a function other than the interactions with plants.

Thus, being the top *mycoparasite* (~top predator) of mainly plant-pathogenic fungi, *Trichoderma* has a positive *indirect* impact on plants (*trophic cascade model*). These powerful indirect interactions occur when a predator (*Trichoderma*) limits the density of its prey (plant-pathogenic fungi) and thereby enhances the fitness of the next lower trophic level (plants). The results also suggest that the still rare *direct* interactions with plants can gain importance in the evolutionary future of some *Trichoderma* spp. However, whether they develop towards *mutualism* or *parasitism* is still challenging to predict.

ISME Abstract Submission | 3.3.20

Title: Bacteria that metabolize phosphite for energy likely have ancient origins and are prevalent on Earth today

Authors: Sophia D. Ewens[a,b], Alexa F. S. Gomberg[a], Tyler P. Barnum[a], Mikayla A. Borton[c], Hans K. Carlson[d], Kelly C. Wrighton[c], John D. Coates[a,b,d]

[a]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; [b]Energy & Biosciences Institute, University of California, Berkeley, CA 94720;
[c]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523; and [d]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Body (500 words max): Phosphite (PO_3^{3-}) is the most energetically favorable chemotrophic electron donor known, with a half-cell potential (E_o') of −650 mV for the oxidation of phosphite to phosphate (PO_4^{3-}). Dissimilatory phosphite oxidizing (DPO) microbes are capable of capturing this energy for growth. Until recently, only two bacterial species were known to perform DPO, lending to the belief that DPO is an esoteric metabolism. Here we describe the identification of 21 novel DPO microorganisms (DPOM) via selective enrichment and genome-resolved metagenomics, which expands the known taxonomic diversity of DPO to six classes of Gram-positive and Gram-negative bacteria. Analyses of DPO-genes allowed for the construction of HMM models that were used to mine publicly available genomes in the Joint Genome Institute, Integrated Microbial Genomes and Metagenomes-database. We show that the taxonomic diversity of DPO enrichments is representative of global DPOM diversity, and metadata reveals that DPOM are concentrated in diverse anoxic environments including wastewaters, sediments, and subsurface aquifers. Evolutionary analyses of DPO genes further indicates that DPO metabolism is predominantly vertically transferred, with ancient origins that may date back to ~3.2 Gya, only ~0.8 Gya after the proposed evolution of methanogenesis. We discuss the implications of the unexpected prevalence, diversity, and antiquity of DPOM.

Metagenomic analysis of algae colonizing the near surface aquifer of a temperate glacier

Quincy Faber, Christina Davis, Brent C. Christner

Glacier algae have been recently studied for their role as primary producers in the cryosphere as well as in ice darkening, lowering albedo, and contributing to the acceleration of ice melting. Although visibly dense algal blooms can be observed on glacier surfaces and snowpack during the melt season, sunlight penetrates ~2m into ice, forming a near surface aquifer, termed the weathering crust aquifer (WCA), that can support photosynthesis beneath the ice surface. Previous work using amplicon sequencing of WCA communities in the Matanuska Glacier (Alaska) during 2014 and 2015 has shown the presence of unique phototrophic algal and cyanobacterial taxa in the near surface aquifer waters. Here we present a functional and taxonomic analysis of the microbial communities based on DNA sequences extracted from the WCA samples in comparison with data from nearby supraglacial streams and deeper horizons within the englacial ice. Parallel Meta and QIIME were used to extract small subunit rRNA gene sequences and assign OTUs, respectively, and revealed that algal taxa are the most abundant phototrophs in near surface ice. Based on phylogenetic analyses using small subunit rRNA and *rbcL* genes, the most abundant taxa are closely related to *Ochromonas* CCMP 1899, a mixotrophic species of golden algae isolated from Antarctic sea ice, and *Ancylonema nordenskioeldii*, a species of green algae commonly associated with glacial ice algal blooms. Functional analysis showed similar abundances of photosynthetic genes in near surface ice and the weathering crust aquifer, and photosynthetic genes were rare in the taxa from the englacial ice (4 to 30m below the surface). These results show that specific algae inhabit sympagic environments in the near surface ice, providing as yet unquantified biogeochemical contributions to the glaicier and meltwater. Future studies using metatranscriptomics can be used to characterize photosynthesizing algae within the WCA and their adaptations to low light and nutrient availability (e.g. pigment expression, mixotrophy). Melt season duration is increasing with global temperatures, increasing the length and extent of melt on ice sheets and glaciers. Therefore, understanding the role of microorganisms inhabiting WCA ecosystems is critical for assessing their contributions to nutrient export and impacts of downstream systems that receive meltwaters.

Gas fermentation has emerged as a sustainable biomanufacturing platform capable of producing valuable chemicals from non-food and waste feedstocks that would normally be considered pollutants. LanzaTech is a worldwide leader in gas fermentation having commercialized and scaled up the production of ethanol from steel mill off gas using a proprietary biocatalyst, acetogen *Clostridium autoethanogenum*, that has demonstrated the ability to metabolize waste gases with various combinations of CO/CO2/H2 into useful products like acetone, isopropanol, and precursors for plastics (https://doi.org/10.1146/annurev-chembioeng-120120-021122)

Only a decade ago, acetogens were considered genetically inaccessible and poorly characterized. DOE-sponsored efforts from several groups have advanced the tools and knowledgebase of C1 recycling acetogens (doi: 10.1038/s41598-017-12712-w, 10.1038/s41467-018-06993-6 10.1371/journal.pcbi.1006848, 10.2172/1543199). These efforts have generated transcriptomic, proteomic, and metabolomic profiles for acetogens when grown under different conditions that have demonstrated differential expression of key metabolic pathways depending on feedstock. Yet, there remains a plethora of uncharacterized genes, and a limited knowledge of endogenous regulatory systems and gene essentiality.

For industrial fermentation, control of metabolic flux can be the difference between a profitable and failed enterprise. Tools for controlling metabolism of acetogens exist including optimizing heterologous genes, swapping promoter sequences, and removing gene sequences via knockout (KO). By far the most employed and effective tool is gene KO (e.g. improving product yield via KO of stress tolerance systems (doi: 10.1038/srep17874)); however, KO can lead to instability or lethal phenotypes in less characterized strains.

Although the pace of genome engineering has been sped up by advances in tool development and foundational research, it remains relatively costly, unpredictable, and inefficient to generate a KO strain. To perturb the system with minimal time and cost LanzaTech has established an effective CRISPRi knockdown (KD) system that can prevent or diminish transcriptional activity of a gene (doi: 10.1093/synbio/ysab008). Building on this tool, we developed first genome-wide CRISPRi libraries in collaboration with JGI.

Through the JGI synthesis program, we designed and synthesized a sgRNA library of ~13,000 plasmids targeting >99% of putative protein coding regions for *C. autoethanogenum*. Using LanzaTech's biofoundry, CRISPRi clones were generated, harvested, pooled, and cultured at various temperatures. Outgrowth cultures were analyzed by NGS and gene fitness scores were calculated by relating normalized sgRNA reads between a sample and its starting population. Fitness scores are leveraged to demonstrate the library is subject to a strong biological influence upon transformation, performs in a reproducible manner between replicates, and leads to both shared and differential fitness aspects between experimental conditions.

The ability to interrogate fitness performance after a single transformation and outgrowth experiment enables scientists to verify rationally derived KO candidates do not have fitness detriment while identifying novel beneficial KO targets in an irrational manner. This work speeds up the strain development timeline, decreases risks associated with KO generation, and elucidates new targets for genome optimization.

Metagenomic Profiles of Bacterial Communities in the Northwest Arabian Gulf with Respect to Depth, Location, and Time

Saja A. Fakhraldeen, Nazima Habibi, Sakinah Al-Haddad, Sabeena Farvin K. Habeebullah, and Surendraraj Alagarsamy

Kuwait Institute for Scientific Research, Salmiya, Kuwait

Located in the Northwest Arabian Gulf, Kuwait Bay is a unique aquatic environment that serves as an important regional nursery ground for myriad marine biota and is a water body of vast cultural, economic, and environmental significance. Due to extreme temperatures, hypersalinity, and exceptional shallowness, the Bay serves as an ideal environment for development of novel genetic adaptations that are unique and necessary for survival of resident organisms. To begin the process of systematically identifying putative unique functional genetic adaptations, one can probe genomic profiles of microorganisms such as bacteria. Marine bacteria are known to play essential functional roles in maintaining biogeochemical equilibrium. Disturbances in marine biogeochemical equilibrium can have detrimental effects on the overall health and stability of the marine ecosystem and it is thus important to understand how marine bacteria individually function to collectively regulate the surrounding ecosystem and prevent detrimental disequilibria. This can be achieved by firstly metagenomically profiling the resident bacterial communities within a given water body. Thus, for this study, seawater samples were collected from various sites strategically distributed throughout Kuwait Bay from both the surface and bottom of the water column on a monthly basis over a period of six months to construct community profiles and investigate their variability with respect to depth, location, and time. Bacterial nucleic acid was extracted from the collected samples and subjected to shotgun metagenomic sequencing at a depth of 20 million paired-end reads per sample. The vast amount of data generated are being subjected to *in silico* interrogations and bioinformatic analyses to profile resident bacterial communities with respect to depth, location, and time, and various physicochemical parameters of the seawater (salinity, temperature, and dissolved oxygen levels) at the time of sampling. Preliminary results reveal that the most prominent bacterial phyla within the collected samples are *Proteobacteria* (55.4%), *Cyanobacteria* (18.7%), and *Bacteroidetes* (11.8%). Within the phylum Proteobacteria, the classes *Alphaproteobacteria* (74.8%) and *Gammaproteobacteria* (25.2%) are the most abundant. The five most abundant bacterial families are *Synechococcaceae* (18.7%), *Rhodobacteraceae* (14.0%), *Alteromonadaceae* (7.3%), *Flavobacteriaceae* (5.7%), and *Geminicoccaceae* (5.6%) while the five most abundant bacterial species are *Synechococcus_sp_WH_8109* (18.6%), *Rhodobacteraceae_bacterium_HIMB11* (13.0%), *Alteromonas_macleodii* (7.0%), *Geminicoccus_sp* (5.6%), and *Formosa_sp_Hel3_A1_48* (3.7%). Members of the class *Gammaproteobacteria* displayed significant differential abundance with respect to depth, favoring the surface versus bottom of the water column. Bacteria that displayed spatiotemporal preferences were more diverse and belonged to various classes and genera. Comprehensive metagenomic community profiles and their spatiotemporal variability as well as their variability with respect to depth down the water column and changes in physicochemical parameters of the seawater at the time of sampling will be shown. In the future, outcomes from investigations of functional adaptations (e.g. adaptations in stress

response genes, salt tolerance genes, etc.) developed by resident marine bacteria to survive within these uniquely extreme conditions will be performed.

Diverse subsurface Thaumarchaeota populations in hydrologically-variable floodplain sediments revealed through genome-resolved metagenomics

Christopher Francis[1]*, Bradley Tolar[1], Linta Reji[1], Emily Cardarelli[1], Kristin Boye[2], and John Bargar[2]

[1] Stanford University, Stanford, CA
[2] SLAC National Accelerator Laboratory, Menlo Park, CA

Formatted: Space After: 0 pt

The terrestrial subsurface microbiome contains vastly underexplored phylogenetic diversity and metabolic novelty, with critical implications for global biogeochemical cycling. Among key microbial inhabitants of subsurface soils and sediments are Thaumarchaeota, an archaeal phylum encompassing the ammonia-oxidizing archaea (AOA). Although AOA have been studied extensively in topsoils, our understanding of the diversity, ecophysiology, and activity of these critically important nitrogen-cycling organisms in deeper soils/sediments is limited. Our recent analysis of subsurface microbial communities (based on both 16S rRNA and *amoA* ammonia monooxygenase genes) within hydrologically-variable floodplain sediments in the Wind River Basin near Riverton, WY revealed that AOA were the predominant ammonia-oxidizers and that their community structure shifted dramatically with depth. In order to understand the ecophysiological adaptations of these subsurface AOA, we used genome-resolved metagenomics to examine Thaumarchaeota populations spanning 11 distinct depths along a 234 cm depth profile at Riverton site KB1. Phylogenomic analysis of metagenome-assembled genomes (MAGs) indicated a pronounced shift in AOA populations with depth. This vertical zonation in thaumarchaeal population structure was similarly evident in the relative abundances of lineages, estimated based on read mapping against MAGs. Our results suggest that hydrological variables, particularly proximity to the water table, impart a strong control on the ecophysiology of Thaumarchaeota in alluvial sediments. To complement our detailed spatial (vertical) characterization of AOA communities at site KB1, we also investigated subsurface microbial community composition over both time and space (depth) at a nearby Riverton site (Pit2) through a full seasonal hydrologic cycle of water table rise, flooding, and summer drought. In particular, have obtained metagenomes (and 1000s of MAGs) from samples collected monthly (April to September 2017) from 7 distinct subsurface layers [topsoil, evaporite, sand, evaporite/clay, transiently-reduced zone (TRZ), clay, and aquifer], allowing spatiotemporal AOA dynamics to be examined within the context of the overall microbial communities and a suite of hydrogeochemical measurements. Overall, this project is yielding unprecedented genomic and ecophysiological insights into the microbial communities responsible for nitrification in subsurface floodplain sediments directly influenced by hydrological fluctuations.

Untargeted metabolomics identifies the potential impact of secondary metabolites on species interactions among xylarialean fungi

Mario E.E. Franco[1], Malak Tfaily[2], Lillian P. Moore[1], Megan N. Nickerson[1], Roya AminiTabrizi[2], Christian O. Ayala-Ortiz[2], Katherine Louie[3], Trent Northen[3], Jana M. U'Ren[1]

[1]BIO5 Institute and Department of Biosystems Engineering, The University of Arizona, Tucson, AZ 85721 USA; [2]Department of Environmental Science, University of Arizona, Tucson, Arizona 85721 USA; [3]Department of Energy, The Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA

The Xylariaceae (Sordariomycetes) comprise one of the largest and most diverse families of Ascomycota. Xylariaceous fungi commonly occur as saprotrophs, pathogens, as well as endophytes of phylogenetically diverse plants and lichens. Previous genomic investigation revealed that Xylariaceae s.l. taxa contain highly abundant and novel secondary metabolite gene clusters (SMGCs), often exceeding the abundance found in bioactive fungi such as *Penicillium* and *Aspergillus*. Here, we investigated the ecological roles of secondary metabolites in species interactions among xylarialean fungi using a co-culture experiment with seven phylogenetically diverse isolates. Isolates were grown alone and in all pairwise combinations to identify the diversity and identity of metabolites produced during fungal-fungal interactions. To assess the impact of phylogenetic relationships on interaction outcomes, we screened three closely related isolates of *Xylaria flabelliformis*, as well as four isolates that represent increasingly distant taxa (i.e., *Xylaria arbuscula, Nemania diffusa, Poronia punctata*, and *Biscogniauxia mediterranea*). Fungi were grown on defined media with glucose at room temperature and ambient light/dark conditions. After one month of growth, we removed the fungal mycelium and agar in the interaction zone between isolates from each plate. Tissues were freeze dried and homogenized with a bead beater prior to untargeted metabolomic characterization at the Joint Genome Institute. Overall, we observed that the majority of fungal interactions resulted in deadlock between isolates (24/28), indicative of the production of inhibitory

secondary metabolites. Analysis of untargeted metabolomics data revealed that 1,370 features were significantly up-regulated in all co-culture treatments compared to isolates grown alone or media controls. For example, the antifungal compound Griseofulvin was up-regulated in all isolates that contained the Griseofulvin SMGC, yet not in all interactions. Our ongoing analyses will shed additional light on the role of secondary metabolism to govern species interactions in one of the most chemically diverse clades of filamentous fungi.

Unraveling the regulation of sugar beet pulp utilization in the industrially relevant fungus *Aspergillus niger*

Sandra Garrigues[1], Roland S. Kun[1], Mao Peng[1], Keykhosrow Keymanesh[2], Dianne Bauer[2], Anna Lipzen[2], Igor V. Grigoriev[2], Vivian Ng[2] and Ronald P. de Vries[1]

[1] Fungal Physiology, Westerdijk Fungal Biodiversity Institute & Fungal Molecular Physiology, Utrecht University, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

[2] USA Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720, United States

Abstract

Sugar beet pulp is the main by-product of industrial sugar beet (*Beta vulgaris)* processing, and is currently sold as low-value animal feed. It is especially rich in pectin and hemi-cellulose, mainly xyloglucan. Efficient degradation and utilization of agro-industrial waste, such as sugar beet pulp, is crucial for the bio-based economy. The filamentous fungus *Aspergillus niger* possesses a wide array of hydrolytic and oxidative enzymes that degrade complex plant biomass substrates, and several regulators have been reported to be involved in their production. The role of the pectinolytic regulators GaaR, AraR and RhaR in sugar beet pectin degradation has previously been reported. However, genetic regulation of the degradation of sugar beet pulp has not been assessed in detail.

In this study, we generated a set of *A. niger* single and combinatorial deletion mutants targeting the pectinolytic regulators GaaR, AraR, RhaR and GalX as well as the (hemi-)cellulolytic regulators XlnR and ClrB to address their relative contribution to the utilization of sugar beet pulp. We show that abolished growth on sugar beet pulp and all sugar beet pulp components can only be accomplished with the combinatorial deletion of all the regulators under study. In addition, transcriptomic analysis showed that *A. niger* has a flexible regulatory network, adapting to the utilization of (hemi-)cellulose at early timepoints when pectin degradation is impaired. Finally, our study demonstrated that AraR, XlnR and ClrB are responsible for the regulation of

the major enzyme activities involved in efficient degradation and utilization of sugar beet pulp components by *A. niger*.

Despite the ecological importance and industrial promise of *Emiliani huxleyi* there currently exists no transfection system for this species. The combination of *E. huxleyi's* calcium carbonate synthesis and photosynthesis, coupled with its propensity to form massive blooms in the Earth's oceans give the organism monumental importance in global carbon cycling (Read et al., 2013). Industrially, *E. huxleyi* has many proposed uses including biofuels, cosmetics, and superfine calcium carbonate. The use of alkenones, long hydrocarbons produced by *Emiliani huxleyi, Gephyrocapsa oceanica,* and *Isochrysis galbana*, as a source of biofuel is of primary interest. Despite its global distribution, industrial usefulness, and ecological importance, over half of the predicted genes in the *E. huxleyi* genome are of unknown function and fundamental processes such as calcification and alkenone biosynthesis are not fully understood at the genetic level (Read et al., 2013; Marsh, 2003; Shi, 2019). Considering the lack of knowledge regarding *E. huxleyi* and alkenone biosynthesis it is urgent that a transfection system be developed to unravel the genetic mysteries of these abundant microalgae and tap into their full industrial potential.

Cell penetrating peptides (CPPs) are a possible agent for successful transfection of *E. huxleyi* and other haptophytes where more common methodologies have failed. CPPs are short polypeptides that are capable of translocating across the plasma membrane. CPPs are easily synthesized, economical, simple to use, and have low cytotoxicity (Magzoub & Grasland, 2004). The primary goal for this project was to complete studies of translocation of three CPPs, antennapedia homeodomain (pAntp), trans-activator of transcription (TAT), and murine vascular endothelial cadherin (pVEC) in three related haptophyte species, *Emiliani huxleyi, Gephyrocapsa oceanica,* and *Isochrysis galbana*. The belief is that these peptides will prove useful for transfection experiments that will allow genetic manipulation of haptophytes.

Our studies show that the CPPs examined are capable of complexation with and intracellular delivery of a 27 kilodalton green fluorescent protein (GFP) cargo. pVEC-fluorescein isothiocyanate (FITC) conjugate outperformed TAT-FITC and pAntp-FITC in translocation ability when administered without cargo, both in the DOE and titer. It was further demonstrated that *I. galbana* and *G. oceanica* were able to proliferate even after exposure to all concentrations of peptide tested. The DOE was able to identify the optimal peptide, incubation time, and peptide concentration for translocation.

Future directions with pVEC include the possibility of plasmid DNA (pDNA) and ribonucleoprotein (RNP) complex transfection (for subsequent CRISPR editing) in haptophytes.

Abstract Title:
Cytotype Variation and Adaptation in Switchgrass

Authors:
Paul P. Grabowski (1), Joseph D. Napier (2), Sujan Mamidi (1), John T. Lovell, Jason Bonette (2), Maria Jose Gomez-Hughes (2), Thomas E. Juenger (2), Jeremy Schmutz (1)

Author Affiliations:
(1) Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL
(2) Department of Integrative Biology, University of Texas at Austin, Austin, TX.

Abstract Text:
Polyploidy is pervasive in plants. The success of higher ploidy taxa is commonly attributed to fitness and adaptive advantages resulting from genome duplication. Testing this hypothesis between taxa is difficult because ploidy difference is generally confounded with genetic divergence. Species with natural ploidy variation, however, can have less correlation between ploidy and genetic divergence and fewer genetic differences between cytotypes, making them good systems for evaluating the fitness and adaptive consequences of ploidy differences. The wide ranging bioenergy grass switchgrass (*Panicum virgatum* L.) has two main naturally occurring cytotypes, 4X and 8X. Here, we use a combination of whole genome resequencing and common garden phenotypes to show that the transition from 4X to 8X has occurred multiple independent times in switchgrass and 8X switchgrass tends to tolerate greater climate variation compared to 4X switchgrass. Furthermore, we demonstrate that 8X switchgrass has, on average, higher signals of admixture between distant lineages than 4X switchgrass, and that variation from distant lineages in 8X is associated with climate adaptation. Together, these data suggest that wide intraspecific crosses promote genome duplication in switchgrass, and the presence of variation from multiple diverged lineages helped 8X switchgrass succeed in niches less suited to 4X switchgrass populations

A New Structural Paradigm In Heme Binding – A Novel Family Of Plant Heme Oxidases.

Nicolas Grosjean*†[1] (ngrosjean@bnl.gov), Desigan Kumaran†[1], Meng Xie[1], Ian Blaby[1,2] and Crysten E. Blaby-Haas[1]

[1]Biology Department, Brookhaven National Laboratory, Upton, NY
[2]Current affiliation: US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA.
†Contributed equally

Up to 90% of the iron found in leaves is located in the chloroplasts, where every membrane-spanning photosynthetic complex has an absolute requirement for iron cofactors, such as heme. Specialized biogenesis pathways involved in heme trafficking and insertion have been described specifically for cytochrome *b*- and *c*-type hemoproteins. These pathways exist to ensure fidelity of cofactor synthesis and limit potential oxidative stress caused by free heme. However, the existence of a generalized heme chaperone that can interact with the labile heme pool in photosynthetic organisms to protect and deliver heme has yet to be identified. Using a phylogenomic approach, we identified a large protein family consisting of uncharacterized or putative heme-binding proteins. Through this analysis we identified three distinct subfamilies of phototroph-specific homologs that might have evolved discrete functions such as i) heme-biosynthesis regulators, ii) heme oxidases, and iii) heme chaperones.

To test our computationally derived hypotheses, we purified protein homologs from the green alga, *Chlamydomonas reinhardtii*, the bioenergy feedstock *Populus trichocarpa*, and the cyanobacterium, *Synechocystis* sp. PCC 6803. We demonstrated that the algal and land plant proteins can bind and degrade heme *in vitro*, suggesting that these chloroplast-localized proteins represent a new family of plant heme oxidases. However, by combining *in vitro* and *in vivo* experiments in *Synechocystis* sp. PCC 6803, we hypothesize that the cyanobacteria-specific subfamily may function as either heme storage or heme chaperones. Additionally, determination of crystal structures of the cyanobacterial homolog in the presence of heme revealed unprecedented features. The protein forms a dimer where the heme is saddled by two zinc ions. By analogy with the protective axial histidines found in a bacterial heme transporter, we propose that the zinc saddle found here protects heme from oxidation. By homology modelling, this structure also helps to understand the structure-function relationship of plant homologs where the heme binding residues are not conserved. The discovery of this family of novel heme oxidases and putative heme chaperones provides new molecular and genomic insights into the evolution of heme regulation in photosynthetic organisms.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER) as part of the Quantitative Plant Science Initiative (QPSI) SFA.

"Conducting STEM industry internships remotely during the pandemic: The Biotech Partners experience"

Author:
Alex Gurn

Abstract:

Biotech Partners (BP), in collaboration with Rockman et al (REA), is exploring the local context of educational partnership with business and industry. This 3-year ethnographic case study uses biotech industry internships as a lens to ask: In what roles and in what ways do business and industry workforce members motivate students from diverse underrepresented populations to become aware of, interested in, and prepared for careers in the STEM workforce? In this Dr. Jennifer Hugenberger (BP) and Alex Gurn (REA) will discuss the role and impact of industry internships with high school students, focusing on student learning experiences and outcomes in data science/bioinformatics internships, and strategies that industry partners and mentors can use to increase youth learning opportunities and develop successful internship experiences.

Mechanisms of infection and response of the fungal wheat pathogen *Zymoseptoria tritici* during compatible, R-gene resistant and non-host interactions

Wheat is the second-most important cereal crop, grown by 30 million farmers worldwide. It is affected by *Zymoseptoria tritici*, the cause of Septoria tritici blotch. This disease has an asymptomatic stage for 1–13 days, followed by a rapid transition to necrotrophy. The lifestyle of *Z. tritici* makes this pathogen an attractive model to investigate infection phase-specific gene expression, and host-pathogen specificity mechanisms in susceptible, R-gene and non-host interactions. Differential gene expression and KEGG pathway enrichment were determined in *Z. tritici* during a susceptible interaction with the cultivar Taichung 29, two R-gene interactions with the resistant cultivars Veranopolis (*Stb2*) and Israel 493 (*Stb3*), and one non-host interaction with barley. Differential gene expression was calculated at 1, 3, 6, 10, 17 and 23 days after inoculation (DAI). The highest number of differentially expressed genes (DEGs) was observed at 10 DAI in the susceptible interaction compared to the R-gene interactions with Veranopolis and Israel 493. This corresponds with the transition to the necrotrophic stage, which suggests that *Z. tritici* activated genes to initiate and maintain the necrotrophic lifestyle around 10 DAI. The enriched pathways from 10 to 23 DAI were biosynthesis of secondary metabolites, amino acids and antibiotics, and protein processing in endoplasmic reticulum. Only six DEGs in *Z.tritici* were observed at 23 DAI when comparing expression during the R-gene interaction between *Z. tritici* and Veranopolis (*Stb2* resistance gene) with that between *Z. tritici* and Israel 493 (*Stb3* resistance gene). No DEGs were identified when comparing these two resistant interactions from one to 17 DAI. This suggests that *Z. tritici* activated the same sets of genes during its interaction with each of the two resistant wheat cultivars. More DEGs were obtained at 3 DAI in the susceptible and R-gene interactions compared to the non-host. This suggests that 3 DAI is the point at which the fungus recognizes the plant as a host and activates specific genes to trigger subsequent responses. The enriched pathways at 3 DAI in the three comparisons were metabolic pathways and peroxisome, while autophagy – yeast and glycerolipid metabolism were enriched in the susceptible interaction compared to the non-host. The biosynthesis of secondary metabolites, amino acids and antibiotics pathways were enriched in the susceptible interaction compared to the non-host from 17 to 23 DAI. There was a dramatic decrease in the number of upregulated DEGs from 6 until 23 DAI when comparing the R-gene interactions to the non-host interaction

between *Z. tritici* and barley. In contrast, from 6 to 23 DAI, there was a substantial increase in the number of upregulated DEGs during the susceptible interaction compared to the non-host. This suggests that from 6 to 23 DAI, *Z. tritici* activates comparable mechanisms to interact with both R-gene resistant and non-host species. In contrast, from 6 to 23 DAI *Z. tritici* activates a large number of DEGs during susceptible interaction other than those that are activated during a non-host interaction, presumably to overcome plant defenses, establish a successful invasion of the host and obtain nutrients.

A highly diverse and novel microbial community mediating reductive sulfur-transformations in an early earth analogue

C. Ryan Hahn, Ibrahim F. Farag, Chelsea L. Murphy, Mircea Podar, Mostafa S. Elshahed, and Noha H. Youssef

Abstract

Life evolved and diversified in the absence of molecular oxygen. The prevailing anoxia and unique sulfur chemistry in the Paleo-, Meso- and Neoarchean, and early Proterozoic eons would support microbial communities that are drastically different than those currently thriving on the earth's surface. However, given preservation limits of biological material; the nature, identity, and fate of microorganisms that thrived under such conditions are currently unknown. Zodletone spring in southwestern Oklahoma represents a unique habitat where spatial sampling could substitute for geological eons: from the anoxic, surficial light exposed sediments simulating a preoxygenated earth, to overlaid water column where air exposure simulates the relentless oxygen intrusion during the Neo Proterozoic. We document a remarkably diverse microbial community in the anoxic spring sediments, with 340/516 (65.89%) of genomes recovered in a metagenomic survey belonging to 210 bacterial and archaeal families that are either novel or that exhibit an extremely rare distribution on the current earth. Such diversity is underpinned by the widespread occurrence of sulfite-, thiosulfate-, tetrathionate-, and sulfur-reduction, and paucity of sulfate-reduction machineries in these novel and rare taxa, hence greatly expanding lineages mediating reductive sulfur-cycling processes in the tree of life. Analysis of the overlaying water community demonstrated that oxygen intrusion led to a drastic decrease in the diversity and novelty of the overall as well as the reductive sulfur cycling microbial communities. Such transition from ancient novelty to modern commonality underscores the catastrophic impact of the great oxygenation event on the earth's surficial anoxic community. It also suggests that novel and rare lineages encountered in current anaerobic habitats could represent taxa once thriving in an anoxic earth but have failed to adapt to earth's progressive oxygenation.

“Microbial Niche of Novel Black Yeast Fungi from Biological Soil Crusts Based on their Genomes, Phenotypic Plasticity, and Metabolite Potential”

Authors: Steven Harris, Erin Carr, Sarah Grambo, Stephen Allen, Mitchell Sullivan, Igor Grigoriev, Alan Kuo, Kerrie Barry, Wayne Riekhof

Black yeast fungi are polyextremotolerant fungi that contain high amounts of melanin in their cell wall and primarily maintain a yeast form. These fungi grow in xeric, nutrient deplete environments which implies that they require highly flexible metabolisms and the ability to form lichen-like symbioses with nearby algae and bacteria. However, the exact ecological niche and interactions between these fungi and their surrounding community is not well understood. We have isolated two novel black yeast fungi of the genus Exophiala: JF 03-3F “Goopy” E. viscosium and JF 03-4F “Slimy” E. limosus, which are from biological soil crusts located in Canada. These crusts are a community of microbes that live in xeric subarctic soils. Whole genome sequencing through JGI, and various phenotyping experiments have been performed on these isolates to investigate their potential niche within the biological soil crust microbial consortium. From our results we have determined that these Exophiala spp. are capable of utilizing a wide variety of carbon and nitrogen sources potentially from symbiotic microbes, they can withstand many abiotic stresses, and potentially provide UV resistance to the crust community in the form of secreted melanin. Further experiments such as their secreted metabolome and co-culturing these isolates with neighboring algae and bacteria from the same samples in the future will provide us with even more information about the niches of these understudied fungi.

“Methanogenic archaea and aerobic methane-synthesizing bacteria from diverse Yellowstone habitats”

Authors:
Roland Hatzenpichler, Brian Bothner, Emiley Eloe-Fadrosh, Zackary Jay, Anthony Kohtz, Viola Krukenberg, Mackenzie Lynes, Timothy McDermott, Angela Tarver, Neha Varghese, Qian Wang

Abstract:
Methane is a potent greenhouse gas. Understanding its sources and sinks is a crucial endeavor in earth system sciences, biogeochemistry, and microbial ecology. It was previously assumed that methanogenesis is restricted to a metabolically highly specialized group of archaea within the phylum Euryarchaeota. However, studies within the last decade have proposed the involvement of several other archaeal lineages in anaerobic methane cycling. These predictions were based on finding the marker gene of anaerobic methane cycling, methyl-coenzyme M reductase subunit A, mcrA, in metagenome-assembled genomes (MAGs) from diverse habitats. While these discoveries are exciting, none of these hypotheses have been experimentally tested and none of the predicted organisms have been brought into culture. Via a combination of targeted mesocosm and cultivation experiments, in situ observations, and cutting-edge single cell targeted physiology measurements we are testing hypotheses on novel methane-producing microorganisms in diverse habitats of Yellowstone National Park (YNP).

In order to gain a better understanding of methane cycling in extreme ecosystems, we screened 100 geothermal features (pH 1.6-9.4, 17-94 ºC) in YNP for the presence of potential methanogens via 16S rRNA gene and mcrA gene amplicon sequencing, in combination with elemental and gas analyses of the geothermal features. From three of these hot springs, we obtained MAGs of both canonical and proposed methanogenic archaea. In addition to interpretating the methanogenic potential encoded by 12 MAGs, we conducted mesocosm experiments to test the activity response of the hot spring communities to the amendment of eight (combinations of) methanogenic substrates over the course of up to 50 days. We observed lineage-, substrate, and feature-specific responses of the methanogenic communities and demonstrated that classical methanogens within the Euryarchaeota as well as methanogenic members of the Verstraetearchaeota and Archaeoglobi are implicated in methane production in these mesocosms. Cultivation efforts are ongoing.

In addition to anaerobic methane cycling, we studied the sources of methane in oxic habitats. Because methanogens are sensitive to O2, methane supersaturation in O2-rich ocean and freshwater environments is a conundrum for both microbiologists and biogeochemists, and has been termed the “methane paradox”. We isolated an aerobic freshwater bacterium, a member of the genus Acidovorax, from Yellowstone Lake. This Acidovorax sp. encodes a single enzyme capable of releasing methane from glycine betaine or methylamine under oxic conditions. When heterologously expressed, this single gene transforms E. coli into a methane producer. This finding provides a yet unaccounted-for source for methane in oxic systems and a direct link between the carbon and nitrogen cycles. Our discovery radically breaks from previous observations that attribute biologically derived methane mainly to anaerobic archaea. After screening all publicly available isolate genomes as well as >15,000 public metagenomes on IMG, we found that homologs of the gene responsible

for methane synthesis in Acidovorax are present in the genomes of virtually all cellular life forms. We are currently screening different environmental samples for this new metabolic potential and use an expression system to screen diverse sequences from environmental metagenomes for their methane-generating potential.

The polyploid genome assembly of the modern sugarcane cultivar R570

Adam Healey, John Lovell, Jerry Jenkins, Olivier Garsmeur, Nicolas Pompidor, Simon Rio, Chris Plott, Karen Aiken, Jane Grimwood, Angélique D'hont, Robert Henry, Jeremy Schmutz

Sugarcane is one of the most important economic crops in the world, both for sugar production and bioenergy usage. Modern, high performance cultivars are a result of hybridization and repeated introgression ('nobelization') between two progenitor species, the high sugar yielding, domesticated Saccharum officinarum and wild S. spontaneum, selected by breeders to provide hardiness and disease resistance and ratooning capability. Each progenitor varies in both chromosome number and ploidy (S.officinarum: 2n = 80, x = 10; S.spontaneum: (2n = 40–124, x = 8), resulting in modern cultivars with incredibly complex genomes that are large (~10Gb), polyploid (10-14X), interspecific, recombinant, aneuploid, heterozygous. Cultivar R570 is the most well-researched genotype to date with many sequencing resources dedicated to untangling its complex genome. Using a combination of sequencing technologies including single dose genetic map markers, Bionano optical map scaffolds, flow-sorted single chromosome libraries, Illumina short reads, Pacbio long read "HiFi" data, and chromatin conformation capture (HiC) libraries, we present the 5 Gb primary genome assembly of R570, constructed into 72 chromosomes. The assembly has been validated with both unique transposable element repeats and fluorescent genomic in situ hybridization (GISH) probes that distinguish between progenitor sequences. This homeolog resolved genome assembly for polyploid R570 represents perhaps the most complex genome ever attempted and will be a critical resource for improving one of the world's most important and valuable crops.

Using an archaeal metabolite to improve growth of bacteria and plants

Connor Hines, Alicia M. Salvi, Aline Rodrigues de Queiroz, Jeremy Brown, Jennie Catlett, Katarzyna Glowacka, Rebecca Roston, Nicole R. Buan

Department of Biochemistry, University of Nebraska-Lincoln

A goal of agricultural research is to find sustainable ways to grow more with less, and an overarching hypothesis is that synthetic biology can be used to meet this pressing social need by introducing genetic enhancements to crops that may never occur without human intervention. We have used a first-principles design strategy to identify a metabolite from extremophilic archaea that thrive in harsh and energy-limited conditions which improves bacterial and plant growth. The metabolite, which has an unknown natural biosynthetic pathway, protects strictly anaerobic methanogenic archaea from oxidative damage and increases growth rate under unstressed conditions. This effect was also seen in *E. coli* bacteria during aerobic growth, both when the metabolite is included in culture medium or when synthesized endogenously from synthetic operons. When applied exogenously to plants, the metabolite improves biomass accumulation and photosynthetic efficiency in *Arabidopsis thaliana* and soybean under normal conditions. Our study reinforces the need to address gaps in biochemistry and physiology knowledge of non-model organisms in order to successfully translate findings from genomic studies to new sustainable agriculture technologies.

Metabat-Hi-C: An Efficient Method for Improving Metagenomic Binning by Integrating Hi-C Sequencing

Metagenomic binning is responsible for computationally grouping scaffolds, assembled from whole genome shotgun (WGS) reads, based on the species of origin. Currently, binning still struggles to generate metagenomic bins with high genome completeness from a single environmental sample. We hypothesized that by using machine learning to integrate long range DNA interaction information provided through Hi-C sequencing into our binning process, we will be able to improve the genome completeness of metagenomic bins. Our software, Metabat-Hi-C, recruits previously unbinned scaffolds into the correct bins and merges bins that are predicted to come from the same genome. We confirmed that the addition of Hi-C sequencing and not simply an increase in sequencing data, leads to the increase in binning completeness. Applying the algorithm to a cat fecal dataset, we found that the integration of Hi-C data led to an average improvement in genome completeness of 19% among all merged bins, while not significantly increasing contamination. This method was able to improve the completeness of 62% of all incomplete bins, bins with less than 80% completeness, through recruiting unbinned scaffolds to existing bins. The remaining bins were held at equal completeness. In the future, we will evaluate our method's ability to retrieve metagenomic bins against other popular binning algorithms such as ProxiMeta and bin3C

Effect of drought on soil microbial metabolisms driving carbon allocation and volatile organic compound cycling in the tropical rainforest at Biosphere 2

Linnea K Honeker, Giovanni Pugliese, Johannes Ingrisch, Jane Fudyma, Juliana Gil-Loaiza, Elizabeth Carpenter, Esther Singer, Gina Hildebrand, Lingling Shi, Erik Daber, Michaela Dippold, Jürgen Kreuzwieser, S. Nemiah Ladd, Christiane Werner, Malak Tfaily, Laura M. Meredith,

Droughts are occurring with increased frequency and duration in tropical rainforests (TRFs) due to climate change, having a significant impact on soil C dynamics. The role of microbes as drivers of changing C flow, particularly in relation to metabolic pathways and volatile organic compound (VOC) cycling, remains largely unknown. Here, we aimed to characterize microbial responses to drought using an integrative, multiple 'omics approach, and hypothesized that microbial communities will adapt by altering their C allocation strategies. Specifically, during pre-drought, primary metabolic pathways will be more active with microbes using C towards growth, whereas during drought, microbes will divert C to secondary metabolite (including VOC) production in response to stress. To test this, we conducted an ecosystem-wide 66-day drought experiment in the TRF biome at Biosphere 2, a glass- and steel-enclosed facility near Tucson, AZ. To track carbon allocation by microbes, we injected C1 or C2 position-specific 13C-pyruvate (PYR) solution into a 25 cm2 region of soil within a soil flux chamber collar (n=6 locations) and measured C isotope ratios of VOC and CO2 emissions. Soil was collected at 0, 6, and 48 hours after PYR addition to examine responses in soil metatranscriptomics, metagenomics, and metabolomics (1H NMR [nuclear magnetic resonance] and Fourier-transform ion cyclotron resonance [FTICR]). Our results indicated that 13CO2 (primarily emitted from locations receiving C1-13C-PYR) fluxes decreased during drought, indicating diminished microbial activity. 13C-VOCs (primarily emitted from C2-13C-PYR) fluxes also differed between pre-drought and drought. Furthermore, drought-induced increases in activity of VOC-producing metabolic pathways, including diterpenoid and monoterpenoid biosynthesis, were evident, as inferred from volatilome, metabolome, and metatranscriptome data. Overall, these results indicate that integration of multiple 'omics datasets reveal specific impacts of drought on microbial activity affecting carbon flow in the TRF soil.

Searching for novel antibiotic resources in eukaryotic algae with a computational approach

Blake T. Hovde and Taehyung Kwon
Los Alamos National Lab

As bacterial adaptations to existing antibiotics continue to threaten human health, the scientific community has been actively seeking novel antibiotic resources. However, these efforts have not been fully extended to eukaryotic algal species. The recent production of high-quality algal genome sequences accelerates resource searches within these untapped genome materials. While the huge genetic diversity in algae has worked as a hindrance for algal genome studies, it could also grant us a broad spectrum of undiscovered antibiotics. Thus, we analyze 218 eukaryotic algal genomes with or without genome annotation and identify a total of 2,902 biosynthetic gene clusters (BGCs) including polyketide synthases (PKS), non-ribosomal peptide sythetases (NRPS), and ribosomally synthesized and post-translationally modified peptides (RiPPs). We investigate the core domain architectures and functional domain frequencies of various eukaryotic algal BGCs, which can be utilized in target selection prior to experimental identification. and phylogenetic reconstruction using the core synthase domain reveals their taxonomic representation within algal clades. In addition, we suggest a down-selection approach by comparing the domain architectures between the candidates and a set of known and experimentally characterized BGCs. Lastly, we investigate the frequencies of the adjacent transporter and antibiotic resistance genes in the eukaryotic algal BGCs. Our finding outlines the efficacy of algal resources in antibiotic discovery through various layers of computational methods, which also provides an *in silico* strategy for the initial step of commercialization of high-value products.

Introduction: Accurately identifying and quantifying metabolites in an untargeted fashion is an important goal for the medical and life sciences. Being able to do so will improve the ability to characterize biochemistry in disease and healthy states, improve establishment of better clinical biomarkers as well as enable better understanding of biochemistry in organisms across the plant and animal kingdoms. Dramatic improvement in mass spectrometry hardware, such as the introduction of the Orbitrap and other high resolution instruments, which offer mass resolution at the single digit parts per million (ppm) error range now produce large and complex datasets. However, detection of meaningful signals remains particularly challenging. Novel informatics are needed to accurately identify metabolites as mass-features in large scale LC-MS data.

Methods: An outstanding challenge in LC-MS driven metabolomics is calling metabolite peaks rapidly and accurately in large datasets (dozens to thousands of samples). While existing algorithms are useful, they have limitations that become pronounced at scale and lead to false positives as well as signal dropouts. We have developed a suite of computational tools to overcome the challenges posed by unreliable algorithms: Isolock, Autocredential and anovAlign. Isolock uses isopairs, or metabolite-istopologue pairs, to calculate and correct for mass drift across LC-MS runs. Autocredential leverages statistical features of LC-MS data to amplify naturally present 13C isotopologues and validate metabolites through isopairs. anovAlign, an anova-derived algorithm, is used to align retention time windows across samples to accurately delineate retention time windows for mass-features.

Preliminary data: Using a large previously published dataset of hundreds of samples from the Broad Intsitute, we demonstrate that this suite of tools is more sensitive and reproducible than the commercial software Progenesis QI. Over 14,000 high priority, stringently filtered metabolite masses are detected with our pipeline, encompassing nearly all of the metabolite masses predicted by Progenesis that pass the same filtering thresholds, and providing thousands of additional signals. A comparison to the open source pipeline, XCMS, will also be provided. Additionally, an analysis of preliminary metabolomics data from the plant species *Setaria viridis* and *Sorghum bicolor* associated with a DOE funded study into genetics of drought tolerance in the C4 grasses will be provided

Novel Aspects: Novel methods for mass-drift correction, improved metabolite peak-picking, and retention time drift correction.

Seasonality and habitat specificity of microbial transcription in a temperate coniferous forest soil

Zander Human, Lucas Auer, Alonso Serrano, Simon Law, Martina Štursová, Tomáš Větrovský, Adina Howe, Christa Pennacchio, Igor Grigoriev, Vaughan Hurry, Nathaniel Street, Francis Martin, Håvard Kauserud and Petr Baldrian

Coniferous forests are globally important due to the large land area that they cover and their role as global carbon sink. Coniferous forest soils play an important role in carbon turnover and nutrient supply to trees and can broadly be divided into different habitats, based on physical and chemical properties and the accompanying microbiome. As a result, these forest soil habitats, such as litter, bulk soil, plant roots and rhizosphere are also distinguished based on the prevailing C and N-cycling processes. The influence of tree photosynthates released by roots may be a further driver of differences between forest soil habitats and should be evident when comparing transcription in summer and winter seasons. In this study, we analysed metatranscriptomes from *P. abies* roots, rhizosphere, bulk soil and litter in March, June, September and December. Our aims were to determine microbiome function in different forest soil habitats, such as N and C cycling. Furthermore, we test the hypothesis that seasonality in forest soil habitats are driven by spruce root activity and ECM symbiosis and thus the extent of seasonality depends on the degree of influence of root activity. Roots and litter had the highest proportion of unique transcripts while rhizosphere and bulk soil shared the majority of transcripts, although bulk soil had more unique functions. Carbon cycling in roots and litter was mainly directed towards plant polymers such as lignin and cellulose, while in rhizosphere and soil, CAZymes targeted mostly α-glucans and bacterial and fungal biomass. Ammonia assimilation was the most transcribed N-cycling process overall and peaked in early summer. Expression was most seasonal in roots, followed by rhizosphere and bulk soil, while litter was the least seasonal. Early summer (June) was the season most differentiated from other seasons, and had the highest expression of root-growth associated transcripts. ECM symbiosis-related gene expression was higher in June and September compared to December and March. Together, our results show that microbial activity in forest topsoil largely reflects substrate quality and content, and that root activity and root-derived compounds are likely the largest source of this seasonality.

Local adaptation of nitrogen-fixing bacteria, Microvirga, to high-stress serpentine soils

Allie Igwe[1] and Michelle Afkhami[1]

1. Department of Biology, University of Miami, Miami, FL

Plant-microbe associations impact plant phenotypes, distributions and biodiversity and range in their effects on a continuum from costly parasitic to beneficial mutualistic interactions. Research into these associations have traditionally focused on obligate or endosymbiotic relationships within the nodules of legumes because of their role in biological nitrogen fixation.

Nitrogen-fixation is an integral part of the nitrogen cycle and the ability of bacteria to perform this chemical reaction makes it important for plants because associations with nitrogen-fixing bacteria provide plant-growth promoting properties and can extend the environmental range of plants. There is much to be gained from exploring how mutualistic relationships between plants and free-living nitrogen-fixers are established and maintained in non-legumes in environments that exert strong selective pressures.

Serpentine soils are an extreme and less well-understood terrestrial environment that can be used as a model system to understand biogeochemical and environmental microbial processes. Previous research revealed that microbial communities enhance plant survival in high stress serpentine soils and implicated one group of bacteria, *Microvirga*, in this effect. Leveraging the strong selection pressure exerted by the extreme environment of serpentine soils on *Microvirga* will allow us to identify organisms and genes that will contribute to the DOE mission of Environmental Microbial Processes.

We propose to isolate *Microvirga* from serpentine and nonserpentine soils in order to characterize the taxonomic and functional diversity observed in free-living nitrogen-fixers. To do this we will assemble draft genomes of the isolated *Microvirga*, annotate assembled genes by comparing them to a reference *Microvirga* genome and identify core and accessory genes, then analyze the genomic variation of Microvirga between serpentine and nonserpentine soils and across California.

We predict that *Microvirga* isolated from serpentine soils will contain accessory genes that confer heavy-metal tolerance to the microbe. Particularly, I expect to see genes related to heavy metal and magnesium transport as nickel concentrations and the Mg/Ca ratio in serpentine soil are the major challenges that living organisms must overcome to survive in the environment.

Title: Elucidation of biochemical pathways underlying multiple stress tolerance in the marine alga *Picochlorum celeri*

Authors: Bishoy Kamel[1], Colin Kruse[2], Sara Calhoun[1], Katherine Louie[1], Ben Bowen[1], Trent Northen[1], Scott J Edmundson[3], Michael Huesemann[3], Shawn Starkenburg[2], Igor Grigoriev[1,4]

[1]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley CA

[2] Los Alamos National Laboratory, Los Alamos, NM

[3] Pacific Northwest National Laboratory, Richland, WA

[4] Department of Plant and Microbial Biology, University of California-Berkeley, Berkeley, CA

Algae are a promising and practical alternative to fossil fuels. Additionally, algae are now being utilized as an important source of pharmaceuticals and nutraceuticals. Understanding algal metabolism and response to stress is a cornerstone to developing industrially relevant strains. *Picochlorum celeri* is an industrially relevant marine alga that is characterized by a high biomass productivity combined with a rapid growth rate. Algae from the genus *Picochlorum* are unique in many aspects having an exceptionally small cell size, the smallest known algal genomes. In nature this genus occupies harsh environments characterized by high salinity and temperature fluctuation exhibiting multi-stress tolerance.

In this experiment we exposed *P. celeri* to outdoor-relevant cultivation conditions in two salinities: 15 PSU and 50 PSU under simulated summer growth conditions and under a simulated Fall cold front temperature event. Earlier observations have shown increased growth under cold stress, earmarking this strain for both winter and summer production. Using a combination of transcriptomics, metabolomics, network analysis and metabolic modeling we identified transcription regulatory modules involved in diurnal regulation of gene expression in addition to response to salt and temperature stress. Pathways enriched in the analysis involved amino acid, fatty acid and osmolyte metabolism with overlap between the different stressors, salt, and temperature pinpointing *P. celeri* unique adaptation to harsh environmental conditions. Furthermore, due to the unique genome organization of *P. celeri* we were able to identify chromosome level modules related to stress and diurnal gene expression.

Additional experiments are underway to confirm the predicted regulatory modules, and metabolic pathways using DAP-Seq technology and proteomics. By combining metabolic modeling with multiomics data generated, candidate genes and genomic loci can be targeted for genetic manipulation paving the way towards rapid strain improvement and prototyping of novel biosynthetic pathways.

“Leveraging the potential of Fabricated Ecosystems (EcoFAB) and model microbial synthetic community for the mechanistic dissection of plant-microbiome interactions”

Authors: Eoghan King, Bradie Lee, Connor Fitzpatrick, Kathrin Wippel, Jacob Calabria, Josefine Kant, Peter Andeer, Kateryna Zhalnina, Karsten Zengler, Borjana Orsava, Michelle Watt, Ute Roessner, Paul Schulze-Lefert, Jeff Dangl, John Vogel, Trent Northen

Abstract:

The major role root-associated microbes play in ecosystems and plant physiology is increasingly apparent, yet the lack of standardized and reproducible experimental systems represents a major challenge for the plant-microbiome field. Fabricated ecosystems (EcoFABs) are sterile devices providing unique capabilities in the control of simplified microbial communities and measurement of their effect on plants, with the potential to advance a mechanistic understanding of soil and plant microbiomes. The Trial Ecosystems for the Advancement of Microbiome Science (TEAMS) project is creating, validating, and disseminating EcoFAB technologies. As part of an international ring-trial experiment, we are distributing to 4 labs a model community of 17 bacteria isolated from the switchgrass rhizosphere and EcoFAB devices. We will assess the reproducibility of host plant selection by inoculating this community on Brachypodium distachyon and Lotus japonicus roots. Owing to the use of EcoFABs, a unique combination of analyses can be performed using the same biological samples: 16S community profiling of root-colonizing bacteria, root system architecture analysis, and metabolomic analysis of root exudates. We will present results obtained from pilot studies to highlight the progress of this project. The purpose of this multi-institution project is to produce common protocols and datasets standards for the plant-microbiome community. Ultimately leading its researchers to build on each other’s work and design predictive models.

Relative Accuracy of Fungal Secondary Metabolite Prediction Pipelines

Talia Korobkin[1-3], Igor Grigoriev[1-3], Asaf Salamov[1-2]

[1]US DOE Joint Genome Institute
[2]Lawrence Berkeley National Laboratory
[3]University of California, Berkeley

Fungal secondary metabolites (SMs) are becoming increasingly important in the study of fungal ecological interactions and the application of fungi to medicine and green technology. Thus, prediction pipelines that identify biosynthetic gene clusters (BGCs) from fungal genome sequences have become crucial in modern research. The most widely used pipelines for fungal BGC prediction are antiSMASH (Blin, 2019, https://doi.org/10.1093/nar/gkz310) and SMURF (Khaldi, 2010, https://doi.org/10.1016/j.fgb.2010.06.003). Directly comparing these tools, I intended to determine the most accurate pipeline for predicting various classes of BGCs, including PKS, NRPs, alkaloids, RiPP, terpenes, and hybrids. Here, I performed a relative accuracy comparison of antiSMASH and SMURF, as well as an augmented version of SMURF (new SMURF), in which new classes of terpene core biosynthetic genes were added for prediction. The platform MiBIG (Kautsar, 2019, https://doi.org/10.1093/nar/gkz882) provided experimental BGC data as a baseline for the comparison of 145 BGC ID's across four different fungal genera. Analysis of BGC prediction accuracy relative to MiBIG gene clusters revealed the new SMURF pipeline as the most comprehensively accurate across BGC classes. For some BGC families, including thiopeptide, antiSMASH has greater or equal accuracy of predicting gene clusters, making the pipeline complementary in these areas. Further analysis of pipeline accuracy on full genomes is necessary to confirm these results.

Stories in Science: Stimulating Interest in STEM Through Student Made Science Videos

Mike Kostolansky, De La Salle High School

Abstract

The purpose of this project, *Stories in Science*, is about high school students exploring STEM by having them create simple, hand drawn videos explaining scientific concepts. There is a rich history making scientific topics more approachable through visual means, with the amount of such content greatly increasing via internet distribution. One example of recent research is Felice Frankel's NSF-funded *Learning by Picturing* program where undergraduate science and engineering students develop visual representations of scientific concepts. *Stories in Science* leverages storytelling, the creative process, and peer learning to engage students in a journey from passive to more engaging learning. In particular, *Stories in Science* seeks to target students in the Bay Area who have had minimal exposure to STEM topics and activities. The "alpha test" phase entails recruiting several small cohorts of students, a teacher moderator, and science and storyteller mentors. Students are organized into groups of three, assigned a scientific advisor and a visual/content storytelling guide. Technical assistance is provided to the cohorts by a project lead. When finished, the videos are presented to their classmates for feedback. It is the hope that *Stories in Science* can find a larger audience beyond the classroom, and ultimately serve to increase the confidence of student-creators to deepen interest in science classes and science-related activities outside of school.

Dennis Suazo, Michaeline B.N. Albright, Thomas Yoshida, John Dunbar, **Marie E. Kroeger**.

Products of plant litter decomposition are major inputs to the soil organic carbon pool in terrestrial ecosystems. Historically, the rate of decomposition and amount of carbon sequestered in soil has been attributed to abiotic factors such as climate and litter chemistry. However, we have found significant variation in carbon flow (i.e., dissolved organic carbon (DOC) and carbon dioxide (CO_2)) during litter decomposition that was driven exclusively by microorganisms. Moreover, we have discovered that high DOC communities have significantly lower species richness compared to low DOC communities, which suggests potential empty niche space or functional limitations. In this study, we aimed to determine whether increased DOC in high DOC communities is caused by species absence (null hypothesis) or suppression (alternative hypothesis). We used a 'common garden' approach with microcosms containing sterile sand and non-sterile blue grama grass litter that was inoculated with 11 mixed dosages of previously characterized microbiomes and incubated for 56 days. These communities were selected based on their consistently high or low DOC phenotypes. Respiration was measured throughout while DOC and microbial community composition was measured on day 56. As expected, we found that the un-mixed high and low DOC controls had significantly different DOC abundance and microbial community composition. We observed an array of intermediate DOC abundances in the mixtures which supports our alternative hypothesis of suppression. Further supporting evidence of our alternative hypothesis is that both fungal and bacterial richness in the mixtures was lower than the low DOC control community and *in silico* mixtures. Overall, both bacterial ($p=0.001$) and fungal ($p=0.001$) communities were significantly altered by mixing ratios. Bacterial ASVs were more commonly associated with the low DOC inoculum while fungal ASVs were more commonly associated with high DOC inoculum. When all mixing ratios were considered, there were more unique bacterial and fungal ASVs in the mixed communities than to either low or high DOC communities. This study provides preliminary evidence that biotic interactions are controlling carbon flow during litter decomposition through functional suppression leading to increased DOC accumulation.

Labyrinthulomycota: not fungi, not algae, what are they?

Alan Kuo, Asaf Salamov, Joshua Rest, Jackie Collier, Igor Grigoriev
DOE Joint Genome Institute, Berkeley, CA. Stony Brook University, Stony Brook, NY.

The eukaryotic Labyrinthulomycota ("labys") play major roles as decomposers in marine ecosystems, and as such are the aquatic counterparts to the mostly terrestrial fungi. Besides their parallel lifestyles, there are morphological and developmental similarities that pre-molecular taxonomies used to classify Labyrinthulomycota as Fungi (hence the -mycota suffix). Molecular taxonomies show that labys are not fungi at all, but are a clade within the Heterokonta, a diverse kingdom of mostly photosynthetic algae, including diatoms, kelps, biofuel candidates such as *Nannochloropsis*, and plant pathogenic oomycetes. Labys themselves, however, are not photosynthetic and lack plastids. To address this apparent case of ecological divergence from algae and convergence with fungi, we sequenced the genomes of the 3 diverse labys *Aurantiochytrium*, *Schizochytrium*, and *Aplanochytrium*, and uploaded them into the comparative genomics platform PhycoCosm (https://phycocosm.jgi.doe.gov). We also imported the genomes of 2 other Labys, a *Labyrinthula* and a *Hondaea*, into PhycoCosm in collaboration with their authors. We interrogated the laby and other PhycoCosm genomes to provide evidence that labys likely evolved from an aplastic ancestor and are not former algae that lost their plastids. We also compared the laby genomes with other eukaryotic genomes in PhycoCosm and MycoCosm (https://mycocosm.jgi.doe.gov) to search for molecular evidence for evolutionary convergence between labys and fungi.

“High throughput mutation maker platform”

Authors:
Emre Kusakci, Lisa Simirenko, Ian Blaby

Abstract:
De novo protein engineering requires powerful tools to generate and screen mutation libraries, which can be used for research, or industrial purposes. Here we present our work on augmenting a mutation maker platform that is automated to perform mutations on a batch of any input DNA or protein sequence allowing high throughput screening of these and generation of optimal oligos by back-tracking algorithms. The platform also screens for possible sites of nicking mutagenesis on the input DNA strain. Therefore, the platform can be used to create single- or multi-site saturation mutagenesis.

Exploring the role of DNA methylation in early-diverging fungi

Carlos Lax[a], Macario Osorio-Concepción[a], Francisco E. Nicolás[a], Junhuan Yang[b], Yuanda Song[b], Victoriano Garre[a]

[a]*Department of Genetics and Microbiology, Faculty of Biology, University of Murcia, Murcia, Spain*
[b]Colin Ratledge Center for Microbial Lipids, School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo, 255000, Shandong, People's Republic of China.

N6-methyladenine (6mA) is an epigenetic DNA modification whose implications have been widely characterized in bacteria. In prokaryotic organisms, this mark is involved processes such as the discrimination between self and foreign DNA, cell cycle regulation, gene expression, and DNA repair and replication. In contrast, although it was a well-known regulator of RNA stability in eukaryotic organisms, the implications of 6mA in those organisms are poorly understood because of the relatively low levels found in eukaryotic genomes. In this project, we will delve into the implications of this mark in the genome of early diverging fungi. In this group, the presence of adenine methylation is higher than in other eukaryotic organisms which suggests that this epigenetic modification could play a relevant role in the biology of those fungi. To better understand the impact of 6mA, we have developed knockout mutant strains in putative adenine methyltransferases using *M. lusitanicus* as a model organism. Knockout mutant strains in the *mta1* gene showed a reduction in growth. When complementing this mutant strain with the wild-type allele of the gene, it recovered the wild-type phenotype. However, complementation with an allele that expressed a protein with a single substitution in the catalytic center failed to restore wild-type growth, indicating that catalytic activity of the protein is required for normal growth. For its part, the disruption of a second gene (*mt2*) encoding a hypothetical methyltransferase caused a severe reduction in genomic 6mA levels, determined analyzing the dA/6mA ratio by HPLC-MS, and an increase in lipid content. Also, we have characterized the implications of 6mA in the yeast to mycelia transition with the study of a putative 6mA demethylase (Dmt1). Deletion of the *dmt1* gene provoked a delay in the dimorphic transition that was associated to an elevated 6mA levels in the mycelium. Transcriptomic, 6mA and 5mC profiling is being running by the Joint Genome Institute (JGI) to correlate phenotypic changes observed in mutants for 6mA modifying enzymes with genome epigenetic modifications and gene expression. The combination of phenotypic and molecular data is expected to reveal the implications of 6mA in the regulation of gene expression in early-diverging fungi and potentially other eukaryotes.

Title: Genome-wide survey of artificial mutations in *Brachypodium distachyon*

Li Lei[1], Lesly L. Fang [1,2], Joel Martin[1], Amy Cartwright[1], Jacob Espinosa[12], Mingqin Shao[1], Richard Sibout[5#,] Debbie Laudencia-Chingcuanco[3#], and John P. Vogel[1,4#]

1 U.S. Dept. of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, USA
2 The University of California, Merced, CA
3 USDA-ARS Western Regional Research Center, Albany, CA, USA
4 UC Berkeley Department of Plant and Microbial Biology, Berkeley, CA, USA
5. Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, 78000, Versailles, France

Abstract: Much work needs to be done to domesticate the grasses being developed as feedstocks for the nascent biofuel industry. Knowledge of the genes controlling traits relevant to biomass crops (e.g. yield, abiotic stress tolerance, disease resistance, cell wall composition) could be used to accelerate domestication and help overcome the inherent difficulties (e.g. polyploidy, self-incompatibility, difficulty evaluating yield traits) in breeding these new crops. The ability to identify and order plants with mutations in virtually any gene in the genome is a powerful research tool that can be used to accelerate research. This approach has been applied to great effect by Arabidopsis researchers thanks to an enormous number of freely available sequence-indexed T-DNA mutants. Unfortunately, as a dicot, Arabidopsis is not suitable to study the many biological questions unique to the grasses. Due in large part to the much greater effort required to create transgenic grasses, the insertional mutant collections available for grasses like *Brachypodium distachyon* and rice have not reached saturation. With the rapid decrease in the cost of DNA sequencing and a large number of mutations, a single chemically mutagenized line can harbor, it is now more cost-effective to identify mutations by whole-genome sequencing than by T-DNA mutagenesis. We created a powerful reverse and forward genetic tool by sequencing a collection of 2,000 chemical and radiation mutants in the model grass *B. distachyon*. We identified ~2 million mutations and annotated their functional impact, particularly, predicting the functional impact of the huge number of nonsynonymous SNPs. This will increase the utility of the mutations collection for studying all aspects of grass biology.

Glycoside Hydrolase family 30 harbors fungal subfamilies with distinct polysaccharide specificities

Xinxin Li[a], Dimitrios Kouzounis[b], Mirjam A. Kabel[b], Adiphol Dilokpimol[a] , Ronald P. de Vries[a]

[a]Fungal Physiology, Westerdijk Fungal Biodiversity Institute & Fungal Molecular Physiology, Utrecht University, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

[b]Laboratory of Food Chemistry, Wageningen University & Research, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands

Abstract

Glycoside hydrolases are critical for the enzymatic hydrolysis of plant biomass, and belong to different families in Carbohydrate-Active enZyme (CAZy) database[1]. Glycoside Hydrolase family 30 (GH30) is a diverse but understudied family, containing enzymes from bacteria, fungi and animals. To deepen our understanding of the fungal members in GH30, we used more than 150 amino acid sequences to construct a GH30 phylogenetic tree, of which eleven candidates covering different fungal SFs were selected for biochemical characterization using polysaccharides and crude plant biomass.

Our phylogenetic analysis of GH30 updated the fungal SFs (GH30_SF3, SF5, and SF7) included in the CAZy database. Fungal members were clustered in SF3, SF10, SF5, and SF7. Functional characterization revealed that enzymes from different SFs exhibited distinct substrate specificities. However, while the different subfamilies act on different polysaccharides, they all mainly release 'short' non-digestible di- and oligosaccharides, which could be of interest in food and feed industries. One new xylobiohydrolase shows high potential for commercial xylobiose production. These findings contribute to understanding the fungal GH30 subfamily and facilitate industrial applications of fungal GH30 enzymes.

References

1. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013.* Nucleic acids research, 2014. **42**(D1): p. D490-D495.

Title: Ectomycorrhizal fungi regulate rhizosphere metal ion processes and plant ion uptake under high heavy metals

Authors: Hui-Ling Liao[1,2], Haihua Wang[1,2], Ryan Tappero[3], Kaile Zhang[1,2], Steven Wu[4], Khalid Hameed[5], Sara Branco[6], Joske Ruytinx[7], Jenny Bhatnagar[8], Alejandro Rojas[9], Eudora Miao[5], Abigail Maciejewski[5], Alan Kuo[10], Kerrie Barry[10], Igor V. Grigoriev[10,11], Rytas Vilgalys[5]

1. North Florida Research and Education Center, University of Florida, Quincy, FL, USA
2. Soil and Water Science Department, University of Florida, Gainesville, FL, USA
3. Brookhaven National Laboratory, NSLS-II, Upton, NY, USA
4. Independent Researcher, Davis, CA, USA
5. Department of Biology, Duke University, Durham, NC, USA
6. Department of Integrative Biology, University of Colorado Denver, Denver, Colorado, USA
7. Research groups Microbiology and Plant Genetics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
8. Department of Biology, Boston University, Boston, MA, USA
9. Department of Entomology and Plant Pathology, University of Arkansas, Fayetteville, AR, USA
10. U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
11. Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA

Metal ion-mediated rhizosphere processes are largely controlled by local biological activities, which in turn, affect soil nutrient fluxes and plant fitness. Specific soil metal ions, such as Fe2+/3+, Cu2+, Mn ions, and Zn2+, are essential for the growth and metabolism of plants and soil biota. Metals are needed at appropriate concentrations, as excessive levels can be toxic. In addition, many metal ions regulate soil organic matter (SOM) decomposition through electron-transferring as well as via promoting the structures of mineral-associated SOM. Ectomycorrhizal fungi (EMF), the keystone symbionts of many tree species, are known to regulate C/N fluxes across host plants and rhizosphere, but it is unclear how much of a role these fungi play in regulating metal-associated chemistry and changing metal bioavailability across the ideal vs. stress levels of external metals. Using *Suillus-Pinus* as the model system, we employed three EMF-plant bioassay setups followed by metatranscriptomic and chemical image approaches to study the molecular function of EMF in response to different external metal ion conditions. The three substrates used for three bioassay studies respectively include sterile sand treated with gradients of ferrihydrite coated sand, Zn^{2+}, and heavy metal contaminated soil. Our preliminary data showed that *Suillus* was able to regulate the rhizosphere metal ion flux and plant ion uptake. For example, in the absence of Fe, *Suillus* broke down the Fe-coded sand into smaller particles in the rhizosphere and enhanced Fe absorption of *Pinus*. The fungal sheath performed dual regulation of plant Zn uptake, which depended upon the concentration of external Zn. In addition, 491 *Suillus* genes were significantly up-regulated, and 203 genes were down-regulated in mycorrhizal roots in response to high Zn/Cd contaminated soil compared to non-contaminated forest soil or sterile sand. Of these 694 *Suillus* genes, 183 genes (27% of total counted genes) were identified as transporters. We found reduced expression in genes coding

for Na+/H+, K+, and NH_3 transport and increased expression of Cu, Fe2+/Zn2+, and P transporter genes, heavy metal exporter HMT1, monocarboxylate and amino acids. In addition to transporters, *Suillus* genes associated with rhizosphere-*Pinus* interaction were also up-regulated in response to heavy metal condition, including genes for plant growth/defense (chitinase), P cycling (dual phosphatase), intracellular iron regulation, program cell death (Metacaspase), GTP-signaling (tyrosine kinase; G-protein subunits) and lipid metabolism (Fatty acid desaturase, Enoyl-CoA hydratase). The changes of *Pinus* genes identified from EMF vs. non-EMF roots in response to heavy metal environment will also be discussed. Overall, our study revealed the ability of EMF symbionts on regulating metal ion allocation between plant and rhizosphere. A core set of EMF genes involved in detoxification and metal ion uptake may be responsible for this process.

Metagenomics reveals stratification of microbial taxa and functional genes in soil aggregates of varying sizes

Jonathan Y. Lin[1], Christian B. Erikson[1], Daoyuan Wang[2], Kate M. Scow[1], and Jorge L.M. Rodrigues[1,3]

[1]Department of Land, Air, and Water Resources, University of California, Davis, CA, USA
[2]Department of Environmental Science and Engineering, Shanghai University, Shanghai, China
[3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Most soil microorganisms live in communities within and on the surface of soil aggregates – interconnected structures of varying size made up of organic matter and mineral particles that comprise the soil. However, these structures often occur at spatial scales much smaller than that traditionally captured in homogenized soil cores. Here, we investigated microbial community taxa and functional potential in soil aggregates ranging from micron to millimeter diameter sizes (< 53 μm, silt & clay; 53-250 μm, microaggregates; 250 μm – 2 mm, small macroaggregates; > 2 mm, large macroaggregates). Soils were collected from fields after 6 years of organic or conventional fertility management at the Russell Ranch Agricultural Experiment Station (Davis, CA) and sieved to obtain each aggregate size. Total DNA was extracted from each aggregate size fraction and whole-genome shotgun sequencing was performed using an Illumina Novaseq platform (PE 150). The resulting reads were assembled using SPAdes and mapped with BBMap, while taxonomy was assigned with Kraken2 and functional genes were annotated using the JGI IMG metagenome annotation pipeline v. 5.0. Both microbial taxonomic and functional gene composition were more strongly shaped by aggregate size (PERMANOVA $R^2 = 0.052$, $P = 0.004$; $R^2 = 0.097$, $P = 0.001$) than by fertility management, respectively ($R^2 = 0.016$, $P = 0.283$; $R^2 = 0.025$, $P = 0.024$). Further analysis showed that the silt & clay particles and microaggregates supported a significantly ($P < 0.05$) higher abundance of microbial genes for denitrification (NarG, NarK, NarY, NirB), whereas small and large macroaggregates supported a higher abundance of sulfur oxidation and superoxide dismutase genes, reflecting differences in oxygen availability based on aggregate pore size. In addition, the small and large macroaggregates contained a higher abundance of genes for glycogen debranching enzyme (GDB1) while the silt & clay and microaggregate fractions had a higher abundance of genes associated with degradation of aromatic compounds (LigB), corroborating differences in the chemical makeup of different sized aggregates. Our results show that microbial taxa and functions are shaped by the spatial structure of the soil, with differences likely reflecting the variation of abiotic conditions in different sized aggregates. These findings are important to more accurately predict biogeochemical cycling than by studying bulk soils alone.

JGI User meeting abstract:

Title:

Pan-genome annotations: an evolutionary framework to assess functional gene variation among de novo genome assemblies.

Authors:

John T. Lovell

Jeremy Schmutz

Abstract:

Recent advances in long-read sequencing technology have permitted assembly of very complex genomes and multiple reference genomes within many species. This leap — from a single small-genome of a related model species to many genomes across and within related taxa — is currently driving a paradigm shift in how biologists view and use molecular variation. We can now test for functional variants in situ within the large and often polyploid genomes of our species of interest, without relying on tenuous comparisons to distantly related genetic models. However, tools to leverage genomic diversity among multiple de novo assemblies are in their infancy. Here, we describe a computational framework to integrate multiple genome annotations into a single database, a 'pan-genome annotation'. To accomplish this, we present GENESPACE, an R package that leverages gene order collinearity ('synteny') and gene sequences that share a common ancestor ('orthologs') across genomes. Finally, we demonstrate that queries of pan-genome annotations result in a nuanced perspective of gene presence-absence and putatively functional sequence variation, even across broad taxonomic scales.

Title

Exploring genomics, transcriptomics, and endophytic fungi of Agave, a promising energy crop for semi-arid and arid regions

Authors

Marina Marone[1], Maria Fernanda Campanari[1], Fabio Raya[1], Gonçalo Pereira[1,2], Marcelo Carazzolle[1,2]

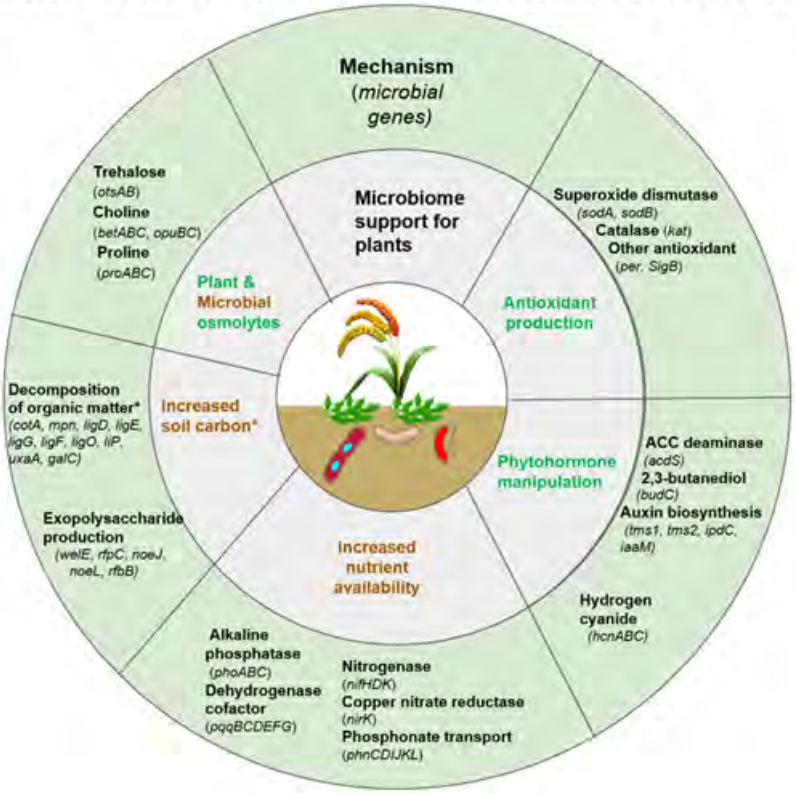
1: Laboratory of Genomics and bioenergy - University of Campinas - Campinas - São Paulo/Brazil.
2: Center for Computing and Engineering Sciences - University of Campinas – Campinas - São Paulo/Brazil.

Abstract

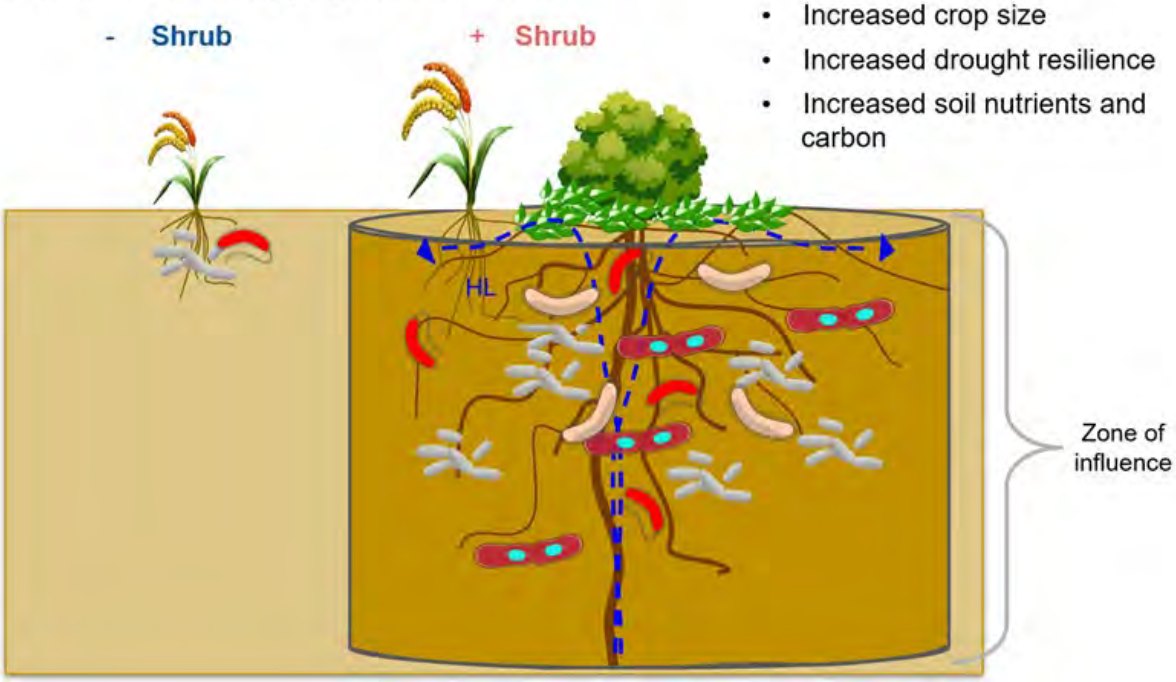
Plants from the genus *Agave* are studied because of their drought resistance mechanisms and for their bioenergy production potential. Currently Agaves are only used to produce sisal fibers and alcoholic beverages (such as tequila and mezcal), in Brazil and in Mexico, respectively. However, Agaves are also interesting to be used as a biomass feedstock to bioenergy production in marginal areas because they present high productivities, low lignin content and high shoot to root ratio. Brazil is the main producer and exporter of sisal fiber in the world and the production is done at the Brazilian Northeast semi-arid region. There are several published works about the morphological, physiological, and genetics of mainly *A. tequilana*, but not many ones about the fiber-producing cultivars. To expand the knowledge about these plants and generate information for breeding programs, we are working with genomics, transcriptomics, and genotyping data of three fiber-producing cultivars (*A. sisalana*, hybrid 11648, and *A. fourcroydes*). We assembled draft genomes of these three cultivars to carry out functional annotation and phylogenomic analysis. In addition, we are using these genomes as reference to guide a genotyping experiment in the Brazilian semiarid regions using ddRAD-Seq and SSRs (single sequence repeats), aiming at building molecular markers for increased yield and disease resistance selection. Also, we have assembled the transcriptomes from leaf, stem, and root from the same three cultivars collected from a germplasm bank located in the state of Paraíba/Brazil, a region that faced a 7-year drought. We explored most expressed transcripts and tissue-specific transcripts in the comparison between the three cultivars. We observed that the cultivars activated a highly overlapping set of stress-response genes, which were the most expressed transcripts. Also, raffinose was detected at high concentrations, possibly acting as an osmolyte, though differences have been found at its biosynthesis routes. Surprisingly, we found within these three transcriptomes around 10% of transcripts belonging to endophytic fungi, and the majority (58%) of root-specific transcripts are fungal. Among the most expressed transcripts, many were annotated as heat shock proteins, which are highly related to heat stress resistance. Other works have reported endophytic fungi in Agave's roots, but the presence of so many heat shock proteins is novel. Our results provide advances about the molecular genetics of three fiber-producing Agave species with great potential to be used as energy crops.

The Sahel of W Africa is a climate change hotspot, where soil degradation and recurring drought significantly reduce agricultural productivity. Erratic rainfall, exacerbated by climate change, t contributes to ongoing and future food insecurity in this region. The staple crop, pearl millet (*Pennisetum glaucum*), is grown by subsistence farmers during the rainy season without fertilizer or irrigation (conditions also desired for US domestic biofuel cultivation). **Indigenous farming practices demonstrate one solution: where farmers intercrop with an native woody shrub** (*Guiera senegalensis*)**, millet drought resilience - and crop yields - are dramatically increased.** One mechanism of this beneficial interaction: shrubs perform hydraulic redistribution of water to surface soil, assisting crops through growing-season drought periods. Additionally, *the moister, carbon-rich soils under the shrub canopy harbor a distinct and active microbial community; this work examines the role of these microbes in aiding the millet (Figure 1)*. We are characterizing this potential in partnership with the Joint Genome Institute (JGI) to metagenomically sequence the millet rhizo- and endo-sphere communities, with and without shrub intercropping, in long-term experimental plots in Senegal. Using the framework of known plant-microbe beneficial interactions (Figure 1) we are analyzing metagenomes for (i) gene ecology, of microbial genes known or hypothesized to impact plant drought; (ii) lineage ecology, of known plant growth-promoting microbial lineages; and (iii) genome ecology, of recovered metagenome-assembled genomes that exhibit differential abundance in the presence of shrubs and/or encode the target drought-relevant genes. We are examining these interactions across nested spatial and experimental scales (Figure 2) to assess the applicability of this system across landscapes.

1 A. Mechanisms of microbial influence on millet drought resilience



1 B. Shrub intercropping confers benefits



- Increased crop size
- Increased drought resilience
- Increased soil nutrients and carbon

2. Beyond JGI support: spatial and experimental scale

Project

Latitudinal Gradient

Key Questions

How does the microbial community influence millet drought resilience with and without shrubs along a rainfall and soil type gradient? How does season affect the structure and function of this community?

Samples & Procedures

	Rainy						Dry						Total Samples	Totals by procedure
	Shrub			No Shrub			Shrub			No Shrub				
	N	C	S	N	C	S	N	C	S	N	C	S		
Millet Rhizosphere	4	4	4	4	4	4	4	4	4	4	4	4	48	48 metaG, 48 PLFA
Millet Endosphere	4	4	4	4	4	4	0	0	0	0	0	0	24	24 metaG
Bulk Soil	4	4	4	4	4	4	4	4	4	4	4	4	48	48 metaG, 48 PLFA
Totals samples	**Rainy Season 2019: 72**						**Dry Season 2020: 48**						**120**	120 metaG, 96 PLFA

1.3 TB sequencing supported by JGI

How does the microbial community influence millet drought resilience in an Optimized Shrub-Intercropping System? How does season affect the structure and function of this community?

	Rainy Season				Dry Season				Total Samples	Totals by procedure
	Shrub		No Shrub		Shrub		No Shrub			
	0NPK	1NPK	0NPK	1NPK	0NPK	1NPK	0NPK	1NPK		
Millet Rhizosphere	4	4	4	4	0	0	0	0	16	16 metaG, 16 PLFA
Millet Endosphere	4	4	4	4	0	0	0	0	16	16 metaG
Bulk Soil	4	4	4	4	4	4	4	4	32	32 metaG, 32 PLFA
Total samples	**Rainy Season: 48**				**Dry Season: 16**				**64**	64 metaG, 48 PLFA

Greenhouse

How does the microbial community from +/- OSS soils influence millet drought resilience under an imposed drought? What is the plant response?

Sampling Location	Phase				Total samples
	Planting	Growth	Drought	Recovery	
Millet rhizosphere microbiome	36 16S iTags; 36 ITS iTags; 36 PLFA	36 16S iTags; 36 ITS iTags; 36 PLFA; 36 metaG; 36 metaT; 36 microbial extracellular enzyme activity (EA)	36 16S iTags; 36 ITS iTags; 36 PLFA, 36 metaG, 36 metaT,	36 16S iTags; 36 ITS iTags; 36 PLFA; 36 EA	144 16S iTags; 144 ITS iTags; 72 metaG; 72 metaT; 144 PLFA; 72 EA
Millet above ground biomass	0	36 chlorophyll content; 36 soluble sugars; 36 Glycine Betaine	36 chlorophyll content; 36 soluble sugars; 36 Glycine Betaine	36 chlorophyll content; 36 soluble sugars; 36 Glycine Betaine; 36 biomass C:N	Chlorophyll content: 108; Soluble Sugars: 108; Glycine Betaine: 108; Biomass C:N: 36
Millet below ground biomass	0	0	0	36 root structure scans; 36 biomass C:N; 36 endophyte metaGs	Root scans: 36; Biomass C:N: 36; Endophyte metaGs: 36
Rhizosphere soil nutrient	12 C:N	36 Anions; 36 plant available nitrogen	0	36 Anions; 36 plant-avail. N	Anions: 72; Plant-avail. N: 72; C:N: 12

Title: Alternating Oxic-anoxic Conditions in Engineered Denitrification Systems Enhance Denitrification Rates via Fungi-mediated Carbon Release

Authors: Philip McGuire, James Shapleigh, Matthew Reid

Denitrifying woodchip bioreactors (WBRs) are an increasingly popular tool for removing nitrate from agricultural tile drainage in a cost-effective and sustainable manner. However, denitrification – particularly in aging WBRs – is frequently carbon-limited due to the recalcitrance of carbon found in lignin-rich woodchip biomass. Managed hydraulic regimes, such as drying and rewetting of the woodchip bed, have been shown to enhance nitrate removal rates as well as mitigate the additional production of nitrous oxide. These benefits introduced from oxic-anoxic cycling of the lignocellulosic media are observed in conjunction with increased levels of dissolved organic and inorganic carbon – indicative of higher concentrations of bioavailable and metabolized carbon within the reactors. We theorize that faster denitrification rates are therefore due to oxygen-dependent ligninolytic enzymes produced by fungal community members that break down recalcitrant lignin sheaths and liberate more-readily bioavailable cellulosic and hemicellulosic carbon for denitrifying microbes. The goal of this research is to elucidate the driving biogeochemical factors that improve WBR performance to aid in the development of novel operating practices, such as bioaugmentation or variable hydraulic regimes.

We present experimental chemical profiles from laboratory flow-through WBR systems supplemented by metagenomic sequencing of DNA isolated from woodchip-attached biofilms. Nitrate removal rates were shown to be enhanced under drying-rewetting conditions without concurrent increases in nitrous oxide. Lignin modifying peroxidases were identified from raw reads via a protein-protein similarity-based approach using the Carbohydrate-Active enZYmes Database and subsequent taxonomic assignment classified the majority of the reads as derived from fungi with *Ascomycota* as the primary phylum. This suggests that fungi play a pivotal role in the WBR in deconstructing lignin to release bioavailable carbon for downstream denitrifying metabolisms, and to our knowledge is the first documentation of fungi as a member of the WBR microbial community. Further examination of the denitrifying community via partial metagenome assembled genomes revealed that many denitrifiers lacked the genetic capability for hemicellulose sugar degradation. This alludes to the necessity of an intermediate niche that promotes the breakdown of primary lignocellulose materials into denitrifier-accessible carbon. These preliminary findings will be further examined via forthcoming metagenomes and metatranscriptomes from JGI complemented by ITS2 amplicon sequencing that will aid in assessing the importance of fungi within denitrifying WBRs.

A sustainable perspective of microalgal cell factory for co-production of biomass, biofuel and by-products (B3) via CO_2 supplementation

Anju Mehra, Pannaga P. Jutur*

Abstract

Microalgal cell factories are fast growing system that can sequester high quantities of CO_2 and convert the same in to biomass, biofuel and byproducts (B3). The effectiveness of microalgae is closely linked to the concentration of CO_2 in the environment for growth and the greater the CO_2 levels, the better the growth and therefore the production of (B3). Our study examined the effect of different CO_2 concentrations (i.e., 0.03%, 1%, 3%, and 6%), in a freshwater *Chlorella* sp. on its macromolecular composition, lipids and biomass productivities. The results revealed that 3% CO_2 was optimal concentration for B3 generation showing biomass, lipid, and pigments productivities up to 344 mg L^{-1} D^{-1} , 53 mg L^{-1} D^{-1} and 4.5 mg L^{-1} D^{-1} respectively. In addition, higher photosynthetic efficiency (F_V/F_m) in 3% CO_2 denotes an improvement in photosystem II maximum quantum yield which in turn reflects in biomass production. Further metabolomics profile was analyzed resulting in identification of 33 metabolites, showing a significant change in metabolites associated with glycolysis and TCA cycle. Up-regulation of TCA cycle metabolites and individual amino acids was observed under 3% CO_2 supplementation; this can be linked with the fact that carbon flux is channelized towards protein biosynthesis and lipid biosynthesis hence enhancing cellular biomass and lipid productivity. In conclusion, the findings of the study show that CO_2 supplementation in microalgal cells is a sustainable approach for the enhanced production of biomass, biofuel, and bioproducts from microalgae.

“The root of the problem: transferring knowledge from model to crop”

Authors:
Maria Mendoza, Lorenzo Scaturchio, Margot Bezrutczyk, Axel Visel, Benjamin Cole

Abstract:
In order to improve drought tolerance, nutrient uptake, and carbon sequestration in crop plants, we need a better understanding of the molecular underpinnings of root morphology and physiology. Single cell RNA sequencing (scRNA-seq) brings us closer to this goal by allowing us to map the transcriptional activity of the root at cellular resolution. Most single-cell studies in plants to date have been performed on Arabidopsis roots, which are morphologically and transcriptionally distinct from distantly-related monocot crops. Cross species integration of datasets is critical for transferring existing knowledge from well-annotated scRNA-seq studies from Arabidopsis to crop plants, but evolutionary distance makes this challenging. For this research project, we established parameters for evaluating the quality of integration of scRNA-seq datasets from rice and Arabidopsis roots. We plan to compare different inter-species integration techniques and improve upon current gene lists to optimize dataset mixing. Our goal is to produce a well-annotated dataset that can be used as an atlas for monocot crops.

Diversity of Dissimilatory Phosphite Oxidizing Microorganisms from Enrichment Cultures of Estuarine Sediments and Groundwater

Kyle S. Metcalfe (kyle.metcalfe@berkeley.edu)*,[1] Sophia D. Ewens,[1,2] Yi Liu,[1,2], Sepideh Sadeghi,[3] W. Andrew Jackson,[3] Alfonso F. Davila,[4] and **John D. Coates**[1,2]

[1] Department of Plant and Microbial Biology, University of California, Berkeley
[2] Energy & Bioscience Institute, University of California, Berkeley
[3] Department of Civil, Environmental, and Construction Engineering, Texas Tech University
[4] Space Science and Astrobiology Division, NASA Ames Research Center

Project Goals: This project aims to investigate the role of microbial dissimilatory phosphite oxidation (DPO) in the global phosphorus and carbon biogeochemical cycles. We are examining the prevalence of DPO and phosphite ($P^{(3+)}O_3^{3-}$) in a broad range of natural environments to examine fundamental physiological and biochemical aspects of DPO. Phosphorus (P) is essential for biology and typically has been assumed to exist at Earth's surface in its 5+ oxidation state, as phosphate ($P^{(5+)}O_4^{3-}$). However, recent work has shown that the reduced P species phosphite can constitute >10% of total dissolved P in diverse environments[1]. In enrichment cultures[2,3] and isolates[4] from the built environment, dissimilatory phosphite-oxidizing microorganisms (DPOM) can conserve energy by phosphite oxidation using the *ptx-ptd* gene cluster[5], but DPOM in natural environments have thus far only been predicted computationally[2]. Here, we present new geochemical and biological constraints on the prevalence of DPOM in natural environments through a synthesis of insights from enrichment cultures of DPOM from estuarine sediment and groundwater. We established 240 enrichments for DPOM from 10 samples of groundwater and 11 samples of estuarine sediment. Enrichments were supplied with phosphite and an array of electron acceptors to enrich for DPOM and were maintained in the laboratory for 10+ months. IC and IC-MS-MS measurements of these enrichments documented active DPO and further revealed that DPO rates were not correlated with endogenous (t_0) phosphite concentrations. These results suggested the metabolic potential for DPO exists in many environments, even those with low (<100 nM) phosphite concentrations. Additionally, DPO rates did not vary by electron acceptor, indicating DPO was likely coupled to reduction of CO_2 as in previous DPO enrichments from wastewater[2,3]. 16S rRNA amplicon sequencing revealed that the most abundant 16S amplicons in several enrichments shared >95% sequence identity with 16S sequences of DPOM MAGs from wastewater enrichments[2]. We thus inferred that DPOM can be successfully enriched from groundwater and estuarine sediments and therefore inhabit these natural environments. Future whole genome sequencing will reconstruct and characterize DPO MAGs to examine the metabolic and phylogenetic diversity of DPOM enriched from groundwater and estuarine sediments.

References

1. Figueroa, I. A. & Coates, J. D. Microbial Phosphite Oxidation and Its Potential Role in the Global Phosphorus and Carbon Cycles. *Adv. Appl. Microbiol.* **98**, 93–117 (2017).

2. Ewens, S. D. *et al.* The diversity and evolution of microbial dissimilatory phosphite oxidation. *PNAS* **118**, (2021).
3. Figueroa, I. A. *et al.* Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E92–E101 (2018).
4. Schink, B. & Friedrich, M. Phosphite oxidation by sulphate reduction. *Nature* **406**, 37–37 (2000).
5. Simeonova, D. D., Wilson, M. M., Metcalf, W. W. & Schink, B. Identification and Heterologous Expression of Genes Involved in Anaerobic Dissimilatory Phosphite Oxidation by Desulfotignum phosphitoxidans. *Journal of Bacteriology* **192**, 5237 (2010).

Examining *Carnobacterium* Species Response to Pressure Extremes Via Growth and Global Gene Expression

Kathleen Miller[1], Flora Tang[2], Sixuan Li[2], Kelli Mullane[2], Bronte Shelton[2], Lam Bui[2], Douglas Bartlett[2], and Wayne Nicholson[1]

[1] Departments of Microbiology and Cell Science, University of Florida
[2] Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego

The complete pressure range for prokaryotic life has yet to be fully characterized. Multiple studies have been undertaken to determine the limits of life in high pressure, but fewer studies have investigated the growth and genetic responses of organisms to low pressure environments. Most prokaryotes grow in a relatively limited range of pressure, but the psychrotolerant, facultatively anaerobic, Gram-positive genus *Carnobacterium* encompasses species capable of surviving a pressure range of over 5 orders of magnitude, from 10^3 Pa (Mars surface) to $\geq 10^7$ Pa (Earth's ocean floor). *Carnobacterium* spp. have been isolated from a wide variety of cold niches including the Aleutian trench and Siberian permafrost.

In order to understand the ability of microbes to adapt in a range of pressure, including low pressure found in the near subsurface of Mars, we cultivated 14 *Carnobacterium* spp. in a cold (2°C), CO_2 atmosphere environment, in ten-fold increments of pressure spanning 10^3 – 10^7 Pa, then harvested samples for RNA-seq and Methyl-seq. Growth patterns of the 14 species indicate that most species, including permafrost isolates, displayed the fastest growth at the lowest pressure tested (10^3 Pa). However, some species showed a divergence in their pressure optimum and grew fastest at higher pressure, suggesting that growth optimums at a particular pressure are species dependent. Previous research on *C. inhibens* subsp. *gilichinskyi* indicated that this species alters its DNA methylation pattern when cultured at low pressure. We propose that this epigenetic response, in conjunction with the global transcriptomic response, allows the species to adapt to different pressures. Bioinformatic analyses of RNA-seq data indicates that genes involved in transmembrane transport as well as motility are differentially expressed in high and low pressure among the different species. Methyl-Seq and RNA-Seq data generated from the 14 *Carnobacterium* spp. cultures grown at different pressures will facilitate our understanding of genetic responses of prokaryotes living in diverse pressure environments.
Supported by DOE-JGI (CSP 502927) and NASA-Exobiology (NNX17AK84G).

Transcriptomics of temporal- versus substrate-specific wood decay in the brown-rot fungus *Fibroporia radiculosa*

Byoungnam Min[a,b] Steven Ahrendt[a,b] Anna Lipzen[a] Cristina E. Toapanta[c] Robert A Blanchette[c] Dan Cullen[d] David S Hibbett[e] and Igor V Grigoriev[a,b]#

[a]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[b]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

[c]Department of Plant Pathology, University of Minnesota, St. Paul, Minnesota, USA

[d]USDA Forest Products Laboratory, Madison, Wisconsin, USA

[e]Biology Department, Clark University, Worcester, Massachusetts, USA

Brown-rot fungi lack many enzymes associated with complete wood degradation, such as lignin-attacking peroxidases, and have developed alternative rapid wood breakdown mechanisms. To see the transcriptomic effect of culture conditions and wood species, we grew *Fibroporia radiculosa* in submerged cultures containing Wiley milled wood wood (5 days) and in solid wood wafers (30 days) with aspen, pine, and spruce as substrate. The analysis revealed wood species had a limited effect on the transcriptome: <3% of genes were differentially expressed between wood species. In the comparison of milled wood and wafer conditions, however, the transcriptome analyses indicated that the genes encoding Fenton chemistry enzymes (such as hydroquinone biosynthesis enzymes and oxidoreductases) were activated during submerged

growth on ground wood, while the genes encoding plant cell wall-degrading enzymes (such as glycoside hydrolases and peptidases) were activated during growth on wood wafers. The activation of plant cell wall-degrading enzymes in the wood wafer samples was expected based on previous reports, but it is surprising that Fenton chemistry was activated in the submerged ground substrate, which lacks intact wood and contains ample moisture and soluble sugars.

Adaptation of Chromatin Condensate Proteins to Temperature

Evan J. Mizerak[1], Serafin Colmenares[2], Gary Karpen[2]

[1]Brown University, Providence, RI, USA
[2]Lawrence Berkeley National Laboratory and UC Berkeley, Berkeley, CA, USA

Abstract

Liquid-liquid phase separation (LLPS) has emerged as having a prominent role in the formation of membraneless organelles — also called biological condensates — such as P granules, nucleoli, and Cajal bodies. LLPS also potentially affects DNA by influencing chromatin organization, transcription, replication, and repair. Heterochromatin Protein 1 (HP1) is a highly conserved family of nuclear proteins whose members exert crucial functions including the repression of gene expression via the assembly and structural maintenance of heterochromatin. Previous work has reported that LLPS mediates the formation of heterochromatin domains, and that *Drosophila melanogaster* HP1a undergoes liquid-liquid demixing *in vivo*. Further, preliminary data suggest that condensates formed by HP1 orthologs from organisms which grow at different optimal temperatures display significant variability in temperature sensitivity.

Here, we leverage multiple sequence alignment, tertiary structure prediction, molecular visualization, and live-cell 3-D imaging to isolate amino acid differences that influence condensate stability following temperature shifts. Focusing primarily on Drosophilids and fungal species that grow optimally at different temperatures, we identify amino acid changes in the conserved HP1 hinge domain which alter biophysical properties such as hydrodynamic radii and isoelectric points. These structural and electrostatic differences correlate with optimal organismal growth temperatures, with larger hydrodynamic radii and greater isoelectric points associated with thermophilic species. Ongoing experiments utilize hinge domain swaps to elucidate potential connections between hinge domain amino acid content and variations in condensate stability *in vivo*. This research has potential applications in understanding and mediating the biological effects of climate change.

Nucleosomes are highly conserved protein-DNA complexes found across Eukaryota. These repeating units, comprised of ≈150bp DNA sequence wrapped around a histone octamer, link together to form larger structures such as chromosomes. In addition to helping structure DNA in the nucleus, they are critical for regulating gene activity. Consequently, a substantial number of epigenetic modifications serve to move these complexes, thereby making genes easier (or more difficult) to access by transcriptional machinery. However, previous studies in model eukaryotes have revealed the potential for DNA sequence itself to play a role in nucleosome organization, although its broader significance *in vivo* is currently poorly understood. Here, we report the presence of periodic DNA patterns at key genomic positions, with peaks separated by ≈150bp that correlate with nucleosome positions. Comparing these periodic patterns across multiple genomic features, we find that they are most prominent surrounding transcriptional start sites and are largely independent of intron/exon structure. While uncommon, this pattern was found across diverse eukaryotes, with a strong phylogenetic component to its distribution. The strongest patterns were observed in Microbotryomycetes (Fungi, Basidiomycota) and Chlorophyta (Viridiplantae, aka, green algae). Through *in vitro* and *in vivo* assays, we confirmed that presence of these patterns is correlated with increased nucleosome occupancy. Interestingly, while dinucleotide frequencies spanning nucleosome-bound DNA sequences varied, occupied DNAs converged on a similar GC pattern across all lineages. Using machine learning approaches, we 1) found that GC content and DNA shape are generally the most important feature in predicting nucleosome occupancy, and 2) developed a model to predict nucleosome occupancy using only DNA features. We applied our method to predict occupancy in other eukaryotes, which reveals a correlation with *in vivo* nucleosome locations. Combined, our data identify DNA sequence as a potentially important contributor to *in vivo* nucleosome positions and should be considered in conjunction with epigenetic modifications when attempting to disentangle nucleosome organization in eukaryotes.

“Developing a novel microbial host and synthetic biology tools for valorizing waste polyethylene terephthalate and lignin”

Authors:
Tae Seok Moon, JinJin Diao, Yifeng Hu

Abstract:
Polyethylene terephthalate (PET) represents 8% (by weight) of global solid waste. PET chemical recycling has been an option to solve this global problem, but it suffers from its relatively high process cost and the extremely low price of virgin PET. One solution is to upcycle waste PET rather than recycle it to generate the same PET typically with low quality. PET upcycling can be achieved by depolymerizing PET into terephthalic acid (TPA) and ethylene glycol (EG) and biologically converting these monomers into value-added products. However, there are only a handful of reports demonstrating microbial strains capable of growing on both TPA and EG generated from PET as sole carbon sources. To overcome this critical challenge, we have performed strain screening to discover a Rhodococcus strain (named RPET) that can grow well on the alkaline hydrolysis products of PET as the sole carbon source without any purification step. Notably, this strain was able to tolerate and grow on a mixture of TPA and EG at extremely high concentrations (up to 0.3M each, total 0.6M) and high osmolarity resulting from alkaline hydrolysis and pH neutralization. The resultant pH neutralized media supported RPET’s growth (up to 0.4 g dry cell weight per g PET) without any purification and sterilization step except for their dilution to make up to 0.6M of monomer concentrations.

Adipic acid, a monomer for nylon production, is currently produced from petroleum derivatives, requiring an alternative process for its sustainable “green” production. Muconate can be converted into various chemicals, including adipic acid. Using non-model organisms, multiple labs have demonstrated muconate production from lignin-derived aromatic compounds, with glucose used as a growth substrate. Rhodococcus opacus is well suited for valorizing lignin, but developing this promising chassis had been challenging due to limited genetic engineering tools. To address this issue, we have developed various synthetic biology tools, including a gene repression system based on CRISPR interference (CRISPRi) and a knockout method. Notably, many synthetic biology tools, developed for Rhodococcus opacus, were functional in different related species and strains such as RPET.

In this presentation, we will discuss our CRISPRi tool’s utility for waste valorization. We have developed and optimized the CRISPRi system, which uses a T7 RNA polymerase system to express a small guide RNA, demonstrating the maximum repression efficiency up to 85%. We also provide a cloning strategy that enables constructing multiple CRISPRi plasmids without any PCR step, facilitating this GC-rich organism’s engineering. Using the optimized CRISPRi system, we confirmed the annotated roles of four metabolic pathway genes related to the consumption of benzoate, vanillate, catechol, and acetate. Additionally, we showed our tool’s utility by demonstrating the inducible accumulation of muconate. While the maximum muconate yield obtained using our tool was 30% of the yield obtained using gene knockout, our tool showed its inducibility and partial repressibility. Finally, we will discuss our effort for PET conversion into carotenoids and muconate as

two demonstration products. This work represent the promise for valorizing waste PET and lignin-derived compounds using Rhodococcus.

Spacer Coverage and Genome Conservation in the *Myoviridae* Virus Family

Isabella Morimoto, Russell Neches, and Nikos Kyrpides
Joint Genome Institute

Many bacteria and archaea use the CRISPR-Cas system to defend against viral infection. In order for this system to function, these populations incorporate spacer genes into their genome that match to specific viral genomic sequences. Based on this, it may be evolutionarily beneficial for the CRISPR-Cas systems to target more conserved genes in the viral genome. The purpose of this study is to examine metagenomic data to identify possible patterns in which genes of the family *Myoviridae* match to CRISPR spacers and specifically, to determine if the protospacers found more frequently in *Myoviridae* host genomes are found in more conserved regions in the *Myoviridae* viral genome. This study used a bipartite network assembled from data from IMG [1,2] and IMG/VR [3]. Habitat metadata was attached to spacer nodes, and host prediction data was attached to viral nodes. The data was then filtered to include only information about *Myoviridae* and sorted to determine the genes present in the largest number of viral genomes. These genes were separated from the viral contigs to form a multiple sequence alignment and build a phylogenetic tree.

[1] Mukherjee, Supratim, Dimitri Stamatis, Jon Bertsch, Galina Ovchinnikova, Jagadish Chandrabose Sundaramurthi, Janey Lee, Mahathi Kandimalla, I-Min A. Chen, Nikos C. Kyrpides, and T. B. K. Reddy. "Genomes OnLine Database (GOLD) v. 8: overview and updates." *Nucleic Acids Research* 49, no. D1 (2021): D723-D733.

[2] Chen, I-Min A., Ken Chu, Krishnaveni Palaniappan, Anna Ratner, Jinghua Huang, Marcel Huntemann, Patrick Hajek et al. "The IMG/M data management and analysis system v. 6.0: new tools and advanced capabilities." *Nucleic Acids Research* 49, no. D1 (2021): D751-D763.

[3] Roux, Simon, David Páez-Espino, I-Min A. Chen, Krishna Palaniappan, Anna Ratner, Ken Chu, T. B. K. Reddy et al. "IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses." *Nucleic Acids Research* 49, no. D1 (2021): D764-D775.

Green Yeast Revisited: Gene targeting in Auxenochlorella & Chlamydomonas

Jeffrey L. Moseley[1,2] , Daniela Strenkert[1], Sean D. Gallaher[1] & Sabeeha S. Merchant[1]
[1]Quantitative Biosciences, University of California, Berkeley, Berkeley, California 94720
[2]Phycoil Biotechnology International, Inc., Fremont, California 94580

The unicellular, oleaginous Trebouxiophyte, *Auxenochlorella protothecoides*, is capable of robust photoautotrophic, mixotrophic or heterotrophic growth in defined media. Heterotrophic fermentation of Auxenochlorella has been deployed for manufacture of algae flours and triacylglycerol oil at industrial scale. Genome sequences have been published for *A. protothecoides* strains Cp0710, UTEX 25 and UTEX 2341, but none of the nuclear assemblies was resolved to the level of phased, diploid genomes. We have leveraged sequence information from the public Auxenochlorella genomes to resolve polymorphic alleles at multiple loci for *A. protothecoides* UTEX 250 and UTEX 2341. Public transcriptome data was used to improve models for genes involved in central carbon metabolism, fatty acid and lipid biosynthesis, chlorophyll biosynthesis, light-harvesting and photosynthetic electron transfer. Furthermore, we have developed methods for Auxenochlorella transformation, transgene expression and nuclear gene targeting by homologous recombination, increasing the utility of this species as both a reference organism for basic research and a biotechnology platform.

We undertook reverse genetic investigation of chlorophyll biosynthesis in order to demonstrate the use of Auxenochlorella and Chlamydomonas as complementary experimental systems. Chlamydomonas *CTH1* and *CRD1* encode paralogs of the diiron enzyme component of the Mg-protoporphyrin IX monomethylester (Mg-PIXMME) cyclase, involved in formation of the chlorophyll isocyclic E ring. *CTH1* and *CRD1* are reciprocally expressed in response to Cu- and O_2-availability under the control of the CRR1 transcription factor, and we demonstrated previously that chlorophyll accumulation is reduced by 60-70% in Cu-deficient or hypoxic *crd1* mutants. Novel *cth1* mutants generated by ribonucleoprotein-mediated editing had the opposite phenotype, with Cu-replete cells displaying reduced growth and chlorophyll. Auxenochlorella *CHL27* is the homolog of *CTH1*/*CRD1*, and sequential ablation of both *CHL27* alleles resulted in non-photosynthetic mutants that were completely chlorophyll-deficient and accumulated a red intermediate pigment with an absorption spectrum that was consistent with Mg-PIXMME. Chlorophyll biosynthesis and photosynthetic growth were restored to *chl27* double knockouts by knock-in of *CHL27* or *CTH1*, but not by *CRD1*, suggesting that additional factors might be required for CRD1 to function in the heterologous system.

The Chlamydomonas ferredoxins FDX5 and FDX6 are also reciprocally regulated by Cu. FDX5 expression is activated in Cu-deficient cells, but FDX6 accumulates only under Cu-replete conditions. *FDX5* homologs are only found within the Chlamydomonadales, and their function is unclear. FDX6 is a homolog of plant FdC2, a broadly conserved ferredoxin implicated as an electron donor in the cyclase reaction. We found that Chlamydomonas *fdx6* mutants recapitulate the Cu-conditional chlorophyll-accumulation defect of *cth1* mutants, and homozygous Auxenochlorella *fdx6* knockouts were similarly chlorophyll-deficient. Using the Auxenochlorella mutants as hosts for heterologous expression, we will test the hypothesis that Chlamydomonas FDX6 provides electrons for the cyclase reaction to CTH1, and FDX5 performs the same function for CRD1.

Exploring Genomes OnLine Database (GOLD) - A Summer Intern Perspective

Paul Nicolopoulos* (PaulNicolopoulos@lbl.gov), Alessandro Favognano* (Afavognano@lbl.gov), Supratim Mukherjee, Dimitrios Stamatis, Cindy Li, Jagadish C. Sundaramurthi, Galina Ovchinnikova, Mahathi Kandimalla, Nikos C. Kyrpides, I-Min Chen, T.B.K. Reddy

DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, CA, USA

The Genomes Online Database (GOLD) is a manually curated and freely accessible database (https://gold.jgi.doe.gov) that allows users to enter and access finished and ongoing sequencing projects from around the world. GOLD lets users search, sort, and categorize genome reports in a simplified manner using advanced search tools. GOLD is organized in a four level system that follows the form of Study, Biosample/Organism, Sequencing Project, and Analysis Project. Projects that are in GOLD come from different sources including internal projects from the Department of Energy Joint Genome Institute (DOE-JGI), external projects entered by users, and projects sourced from public databases such as NCBI. As of August 2021, GOLD contains approximately 49,780 Studies, 146,270 Biosamples, 415,580 Organisms, 431,380 Sequencing Projects, and 324,760 Analysis Projects.

As interns at GOLD, we have been reviewing the data of different organism and metagenome entries and curating them. Specifically, we have been reviewing geographic location information by editing latitude/longitude coordinates and by checking the name and formatting of the entries. This is important because if a user were to look up an organism or metagenome entry on GOLD and the coordinates did not match the geographic location, this could cause confusion and could lead to analysis errors. Currently, there are nearly 120,000 organisms and 150,000 metagenome entries that we are reviewing.

One of the most common errors with these entries are coordinates which lead to inaccurate geographic locations. An example of this is from a biosample that was identified as coming from Knoxville, Tennessee with the coordinates 35.9606, 83.9207. This set of coordinates leads to somewhere in China rather than the US. To find the root of this problem it is good to see if either the longitude or latitude is roughly similar to the desired geographic location in order to see which

coordinate is off. In this example, the latitude seemed to be accurate but the longitude seemed off. The next step would be to compare the faulty coordinate (longitude in this case) with other biosamples entries from Knoxville in order to see the typical longitude range. It seems that Knoxville longitudes are usually around -84, which is a clear indicator that the longitude should have been -83.9207 rather than +83.9207).

We can use advanced searches in GOLD to find interesting information. For example, we can use advanced search tools to search for and select sand metagenomes as well as see their distribution from around the world. Another example is to search for engineered biosamples based on their habitats. This allows us to get a general idea of where these biosamples are coming from. We will discuss such search cases in more detail.

Genomes of novel Myxococcota reveal severely curtailed machineries for predation and cellular differentiation

Chelsea L. Murphy1, R. Yang[1], T. Decker[1], C. Cavalliere[1], V. Andreev[2], N. Bircher[1], J. Cornell[1], J. Bowen[1], R. Dohmen[1], C. J. Pratt[3], A. Grinnell[1], J. Higgs[1], C. Jett[1], E. Gillett[1], R. Khadka[1], S. Mares[1], C. Meili[1], J. Liu[4], H. Mukhtar[1], Mostafa S. Elshahed[1], and Noha H. Youssef[1]*

1Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA
2Department of Plant Biology, Ecology, and Evolution, Oklahoma State University, Stillwater, OK, USA
3Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK, USA
4Department of Animal Sciences, Oklahoma State University, Stillwater, OK, USA

Cultured Myxococcota are predominantly aerobic soil inhabitants, characterized by their highly coordinated predation and cellular differentiation capacities. Little is currently known regarding yet-uncultured Myxococcota from anaerobic, non-soil habitats. We analyzed genomes representing one novel order (o__JAFGXQ01) and one novel family (f__JAFGIB01) in the Myxococcota from an anoxic freshwater spring in Oklahoma, USA. Compared to their soil counterparts, anaerobic Myxococcota possess smaller genomes, and a smaller number of genes encoding biosynthetic gene clusters (BGCs), peptidases, one- and two-component signal transduction systems, and transcriptional regulators. Detailed analysis of thirteen distinct pathways/processes crucial to predation and cellular differentiation revealed severely curtailed machineries, with the notable absence of homologs for key transcription factors (e.g. FruA and MrpC), outer membrane exchange receptor (TraA), and the majority of sporulation-specific and A-motility-specific genes. Further, machine-learning approaches based on a set of 634 genes informative of social lifestyle predicted a non-social behavior for Zodletone Myxococcota. Metabolically, Zodletone Myxococcota genomes lacked aerobic respiratory capacities, but encoded genes suggestive of fermentation, dissimilatory nitrite reduction, and dissimilatory sulfate-reduction (in f_JAFGIB01) for energy acquisition. We propose that predation and cellular differentiation represent a niche adaptation strategy that evolved circa 500 Mya in response to the rise of soil as a distinct habitat on earth.

“SIP-OMICS: A SEMI-AUTOMATED PIPELINE FOR ISOTOPICALLY-TARGETED COMMUNITY ANALYSIS AND METAGENOMICS”

Authors: Erin Nuccio, Steve Blazewicz, Marissa Lafler, Ashley Campbell, Rex Malmstrom, Jennifer Pett-Ridge

Abstract:

Linking the identification of uncultivated microbes with their environmental function has long been a ‘Holy Grail’ for microbial ecologists. While many techniques attempt to meet this goal, stable isotope probing—SIP—remains the most comprehensive for studying whole microbial communities in situ. Since it was introduced in 2003, over 500 publications report using SIP to link metabolic activity to microbial identity. In SIP, microbes who take up an isotopically heavy substrate end up with heavier DNA, which can be sorted by density and divided into multiple fractions. Compounds labeled with 13C or 15N are frequently used to study the ecophysiology of organisms that consume a substrate of interest, while 18O water is used as a universal tracer to measure the taxon-specific growth of all active taxa. However, SIP is not as broadly used as it could be because it requires specialized equipment and is relatively low throughput and time-consuming. Via the JGI ETOP program, we designed a high throughput semi-automated SIP pipeline that can be combined with either amplicon or metagenomic sequencing. Our pipeline decreases operator time, reduces operator error, and improves reproducibility by targeting the most labor-intensive steps of traditional SIP—fraction collection, cleanup, and DNA processing. In addition, we have developed a method for pre-screening nucleic acids for isotopic enrichment, to ensure samples are adequately enriched prior to density gradient separation. Since establishing our pipeline, we have run over 1000 SIP samples, including well-replicated studies of annual grassland soil taxa active during key points in the water-year (fall wet up, spring growing season), plus analyses of drought, redox, soil habitat and mycorrhizal effects. Overall, the pipeline reduces the per sample processing time from 9 hours each to 1.7 hours, with ca. 30 minutes of manual work per sample. Using in silico analysis, we determined the least number of fractions needed—without increasing error rate. Through the development of this pipeline and implementation at JGI, our goal is to make isotope-enabled techniques both high-throughput and accessible to the greater scientific community.

Paight, C. University of California Santa Barbara. cpaight@ucsb.edu
Johnson, M.D. Woods Hole Oceanographic Institution. mattjohnson@whoi.edu
Lasek-Nesselquist, E. Wadsworth Center. elasekness@gmail.com
Moeller, H.V. University of California Santa Barbara. hvmoeller@ucsb.edu

From Eating to Enslavement: Comparative Transcrptomics of the Chloroplast-stealing Ciliate Genus *Mesodinium*

A number of eukaryotic protists obtain their ability to photosynthesize through kleptoplasty, the retention of photosynthetically active chloroplasts from their prey. Organisms capable of kleptoplasty use the photosystems of ingested prey to produce sugars that make up a significant amount of the carbon captured and energy demands of the host. Kleptoplasty has independently evolved multiple times in diverse lineages, but the mechanisms involved in kleptoplasty evolution, acquired plastid maintenance, and regulation are not well understood. The ciliate genus *Mesodinium* contains karyokleptic (steals prey nuclei as well as plastids), kleptoplastidic, and heterotrophic members, making it an excellent system to study the evolution of kleptoplasty. Comparing the transcriptomes (and eventually genomes) of *M. pulex* (heterotroph), M. *chamaeleon* (kleptoplastidic), and *M. rubrum* (kleptoplastidic and karyokleptic) we have gained insights into the relationships between these ciliate species and their prey. Our findings suggest that, as *Mesodinium* lineages become more adept at retaining prey plastids, their metabolisms also become more reliant on other prey metabolic products in addition to photosynthate. Here we focus on the algal prey metabolic contributions in *M. rubrum* and *M. chameleon.* Both ciliate species rely on photosynthate from stolen plastids, but while *M. chamaeleon* adjusts its own metabolic machinery according to prey type, *M. rubrum* appears to rely on its prey for the production of numerous pre-formed amino acids and metabolites. Thus, at least within the *Mesodinium* genus, increased photosynthetic activity correlates with decreased metabolic independence.

“Machine learning prediction of novel pectinolytic enzymes in Aspergillus niger through integrating heterogeneous (post-) genomics data “

Authors: Mao Peng, Ronald de Vries

Pectinolytic enzymes are a variety of enzymes involved in breaking down pectin, a complex and abundant plant cell wall polysaccharide. In nature, pectinolytic enzymes play an essential role in facilitating bacteria and fungi to depolymerize and utilize pectin. In addition, pectinases have been widely applied in various industries, such as food, wine, textile, paper and pulp industry. Therefore, discovery of novel pectinolytic enzymes has received global interest. However, traditional enzyme discovery relies heavily on biochemical experiments, which are time consuming, laborious and expensive. To accelerate identification of novel pectinolytic enzymes, we developed a machine learning (ML) approach for predicting pectinases in the industrial workhorse fungus Aspergillus niger.

“Temporal transcriptomic analysis to understand the host and non-host aphid feeding on sorghum”

Authors: Lise Pingault, Joe Louis

Sorghum is the fifth most important cereal crop in the world and a model for C4 grass bioenergy crops. However, sorghum is also susceptible to insect pests that can dramatically decrease its yields. Phloem-feeding insects, such as aphids, severely damages the plant by sucking sap from leaves, thereby reducing its photosynthetic ability. In addition, aphids vector plant viruses that result in considerable yield loss. The goal of this project is to decipher the transcriptome of a JGI flagship plant: sorghum, in response to host [i.e. pest cultivar or accession specific] and non-host [i.e. can occur in all the cultivars of the host plant species] pests [i.e. aphids]. In this project, we utilized the most updated version of the sorghum reference sequence of the cultivar BTX623 and a RNA-seq libraries to characterize the transcriptomes of sorghum leaves in response to insect-pest infestation. The experiments were conducted on sorghum plants grown under controlled environment that were infested with host: sugarcane aphid (Melanaphis sacchari) and corn leaf aphid (Rhopalosiphum maidis) and non-host: green peach (Myzus persicae) and greenbug (Schizaphis graminum) aphid species.
In addition to the identification of differentially expressed genes, we will utilize transcriptomic information to identify the functional elements such as lncRNA, cisNAT and cis-regulatory element controlling sorghum response to host and non-host insect pests. Working with host and non-host pests will help to characterize new transcriptomic pathways and regulatory elements, which provides an opportunity to elucidate sorghum defenses against phloem-feeding aphids. We will also take advantage of the time course transcriptomic study to highlight the temporal activation or inactivation of the sorghum defense pathways in response to the different aphids. The information generated from this project will provide needed scientific foundation for rationally designing more resistant crop varieties/cultivars through modern breeding techniques and/or through transgenic approaches.

“Genomes to ecosystems - harnassing genomes to study public good exploitation in natural bacterioplankton communities”

Authors: Shaul Pollak

Abstract:

Bacteria often interact with their environment through extracellular molecules that increase access to limiting resources. These secretions can act as public goods, creating incentives for exploiters to invade and “steal” public goods away from producers. This phenomenon has been studied extensively in vitro, but little is known about the occurrence and impact of public good exploiters in the environment. Here, we develop a genomic approach to systematically identify bacteria that can exploit public goods produced during the degradation of polysaccharides. Focusing on chitin, a highly abundant marine biopolymer, we show that public good exploiters are active in natural chitin degrading microbial communities, invading early during colonization, and potentially hindering degradation. In contrast to in vitro studies, we find that exploiters and degraders belong to distant lineages, facilitating their coexistence. Our approach opens novel avenues to use the wealth of genomic data available to infer ecological roles and interactions among microbes.

Molecular Signatures of *Janthinobacterium lividum* in Contaminated Soil from Trinidad Support High Potential for Crude Oil Metabolism

Amanda C. Ramdass [1] and Sephra N. Rampersad [2*]

[1] Biochemistry Research Laboratory (Rm216), The University of the West Indies, Faculty of Science and Technology, Department of Life Sciences, St. Augustine, Trinidad and Tobago - West Indies ac_ramdass@hotmail.com
[2] Biochemistry Research Laboratory (Rm216), The University of the West Indies, Faculty of Science and Technology, Department of Life Sciences, St. Augustine, Trinidad and Tobago - West Indies sephra.rampersad@sta.uwi.edu
***** Correspondence: sephra.rampersad@sta.uwi.edu

Trinidad and Tobago (10° 32' 11.1156" N and 61° 18' 43.0236" W) is the southernmost twin-island country of the Caribbean, and is positioned ~11 kilometres off the north eastern coast of Venezuela in mainland South America. Trinidad has a century-old history of petroleum exploration, and production which began in 1857; this makes Trinidad one of the world's oldest oil producers. The first oil wells were drilled close to the Pitch Lake in La Brea which is the largest, most significant, natural deposit of asphalt in the world. Leaking oil conduits, naturally-occurring oil seeps, and asphaltic mud volcanoes are all part of the terrestrial landscape in South Trinidad. Recent exploration of these polluted sites uncovered strains of *Janthinobacterium lividum*, which for the first time in the literature, was isolated from oil-contaminated soils. These Trinidad strains were identified as highly efficient crude oil degraders presenting with atypical features. Our data suggest that chemotaxis, elective inactivation of non-complementary mechanisms of co-metabolism and secreted lipase activity of these strains support the ability of Trinidad strains to outcompete and survive in these contaminated environments as both a self-regulating microorganism and as a key consortia member. The success of natural attenuation and bioremediation approaches of organic contaminants depends largely on understanding of microbial regulatory mechanisms and cellular responses to various environmental factors/stressors. As such, we propose that chemotaxis enables the Trinidad *J. lividum* strains to navigate through the anoxic micro-environment of crude oil-contaminated soil to encounter compounds that are metabolizable for optimum growth and survival. The action of secreted lipase enzyme contributed to the demonstrated lipolytic activity of the Trinidad *J. lividum* strains in culture. All identified mutations in the deduced *EstA*/*B* (α/β hydrolase) amino acid sequence of the Trinidad strains were predicted to be neutral, and were located outside of the main catalytic sites. Lipases are one of the largest groups of commercially important enzymes, and bacterial α/β hydrolases are highly efficient in the hydrolysis of polyaromatic esters. Additionally, the Trinidad *J. lividum* strains were only recovered in co-culture with different species of yeast, and were resistant to tetracycline and streptomycin antibiotics. They were also non-violacein producers in culture, and colonies were non-pigmented, despite several attempts at induction of gene expression of the *VioABCDE* operon. Analyses indicated no functional disruption of the *VioA* gene, and its deduced amino acid sequence. The absence of violacein may be the result of elective inactivation of the operon which would serve the co-metabolic synergy of Trinidad *J. lividum* strains with different yeast species for survival crude oil polluted sites. The *VioA* and *EstA*/*B* mutations may have accumulated over evolutionary time without compromising the ability of these strains to survive in crude oil contaminated environments, and without affecting consortia-based interactions with different yeast species. These synergistic relationships are

complex reactions that can lead to critical behavioural shifts for the best survival outcome for members of a particular partnership.

Genome-wide cutting scores enable sgRNA activity predictions and definition of essential genes in the yeast *Yarrowia lipolytica*

Adithya Ramesh[1], Dipankar Baisya[2], Varun Trivedi[1], Cory Schwartz[1, †], Stefano Lonardi[2], Ian Wheeldon[1]*

1. Department of Chemical and Environmental Engineering, University of California, Riverside, CA 92521
2. Department of Computer Science, University of California, Riverside, CA 92521
3. Center for Industrial Biotechnology, University of California, Riverside, CA 92521

[†] Current address: Pfenex Inc., 10790 Roselle St., San Diego, CA 92121

Abstract

Genome wide functional genomics screens present an opportunity to determine the genetic underpinnings of a biological process and investigate the contribution of genes to desired phenotypes. Non-conventional organisms are good targets for such screens as they can present a range of desirable traits that help avoid complex and intensive metabolic engineering of less suitable model hosts. As a drawback, their genome and metabolic networks are often less well understood, and they lack the range of genetic engineering tools available in model organisms. A significant bottleneck in engineering the non-conventional oleaginous yeast *Yarrowia lipolytica* has been the lack of synthetic biology tools for multiplex genome editing, functional genomic screening, and rapid strain development. We have sought to overcome these limitations by developing CRISPR-Cas9 and -Cas12a based tools for gene knockout, integration, regulation, and genome-wide screening. However, prediction of highly active sgRNA which are crucial in effective genome editing and improving confidence in hit calling, remains a challenge. In addition, *Y. lipolytica* lacks a well-defined consensus set of essential genes that would help further our understanding of this organism and ease metabolic engineering efforts. Thus, we constructed two genome-wide libraries, one using SpCas9 and the other using LbCas12a, to target all protein coding sequences. In the absence of DNA repair by non-homologous end joining, screens provided a cutting score (CS) for each guide, while screens in the wild type background provided a fitness score (FS) for each gene. We used the genome-wide CS values to develop a new machine learning based guide-activity prediction algorithm called DeepGuide, and also used these values to provide a guide-activity correction to more accurately determine FS for each gene in the genome. Combined with results from a previously published essential gene set identified using a transposon screen, the outcomes of our CRISPR screens define a consensus set of essential genes for *Y. lipolytica*.

Metagenome-assembled genomes from San Francisco Bay water column yield insights into blooming ammonia-oxidizing archaea and abundant bacterioplankton

Anna N. Rasmussen and Christopher A. Francis

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

Metagenome assembled genomes (MAGs) have greatly expanded our ability to examine the metabolic capacity of microorganisms from a wide variety of environments and to gain insight into biogeochemically relevant microbes that have yet to be enriched in the laboratory or were missed by previous primer-based studies. We assembled and binned metagenomes from eight pelagic samples collected from along the salinity gradient in San Francisco Bay (SFB). We subset the metagenomes at different percentages and co-assembled samples to optimize genome coverage and assemble MAGs for 'highly abundant' and 'less abundant' organisms, respectively. After dereplicating (95% Average Nucleotide Identity) all bins > 50% complete and < 10 % contaminated, we have a total of 449 MAGs from 17 phyla. Through this process, we recovered two high-quality ammonia-oxidizing archaea (AOA) MAGs representing a *Nitrosomarinus*-like genome from the South SFB and a *Nitrosopumilus* "SCM1-like" genome originating from North SFB. The South SFB AOA MAG represents a highly abundant bloom organism, and analysis of it along with other *Nitrosomarinus*-like MAGs reveals that this group has streamlined genomes with low GC content, high coding density, and is enriched in urease genes compared to other AOA. We also recovered six *Pelagibacteraceae* MAGs with distinct distributions along the salinity gradient. Our dataset allows for the comparison of closely related organisms across a fresh-to-marine salinity gradient from several different phyla and will yield insight into marine-freshwater transitions and niche separation.

PHYLOGENOMICS AND ENRICHMENT ANALYSIS OF LONG CHAIN-POLYUNSATURATED FATTY ACID (LC-PUFA) METABOLIC PATHWAY IN MICROALGAE

Mohammed Rehmanji [1, 2], Ashish Kumar, Asha Arumugam Nesamma, Pannaga Pavan Jutur *

[1]Omics of Algae Group, Industrial Biotechnology and DBT-ICGEB Centre for Advanced Bioenergy Research, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110067 INDIA

Tasneem Fatma, Nida Jamil Khan

[2]Department of Biosciences, Jamia Millia Islamia, Jamia Nagar, Okhla, New Delhi 110025 INDIA

* Corresponding author

Pannaga Pavan Jutur, Omics of Algae Group, Industrial Biotechnology, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110067 INDIA; E-mail: jppavan@icgeb.res.in (P. P. Jutur); Phone - +91 11 26742357; Fax - +91 11 17 26742316

Co-authors Email addresses

mohammedrehmanji@gmail.com (MR); ashish.ashu15@gmail.com (AK); asha22 @gmail.com (AAN); tfatima@jmi.ac.in (TF); njkhan@jmi.ac.in (NJK)

Abstract

Enhancing the production of long chain polyunsaturated fatty acids (LC-PUFAs) in the microalgae need a comprehensive understanding on their arrangement and functional relationship among genes involved in their biosynthesis. The characterization and elucidation of such protein-encoding genes and metabolic pathways will provide insights in finding key targets in the synthesis of LC-PUFAs. In the present study, bioinformatics tools have been employed to identify potential gene targets in LC-PUFA biosynthesis. *In silico* prediction of 125 functional genes involved in the production of LC-PUFAs within microalgal lineages were depicted. Phylogenomics demonstrate that LC-PUFAs synthesis genes (FAD; ELO, ACS), regulatory elements (WIN1, MYB family members, LEC, ABI4 etc) and transporter proteins (ABCG11, ABCG2, ABCD1 etc.) are highly conserved among these microalgal and plant lineages. Furthermore, integrating transcriptome datasets employing pathway enrichment analysis show a crosstalk between LC-PUFA biosynthesis proteins (ELO, LACS, FAD), regulatory elements (LEC2, MYB, WIN) and transporters (ABCD1), a new perspective to enhance the precursors for nutraceutical industries.

“Phylogeny and distribution of dissimilatory iodate-reducing bacteria in oxygen minimum zones”

Authors: Victor Reyes-Umana, Zachary Henning, Tyler Barnum, John Coates

Abstract:
Iodine is oxidized and reduced as part of a biogeochemical cycle that is especially pronounced in the oceans, where the element naturally concentrates. The use of oxidized iodine in the form of iodate (IO3−) as an electron acceptor by microorganisms is poorly understood. In our recent work, we outline genetic, physiological, and ecological models for dissimilatory IO3− reduction to iodide (I−) by a novel estuarine bacterium, Denitromonas sp. IR-12. Our results show that dissimilatory iodate reduction (DIR) by strain IR-12 is molybdenum-dependent and requires an IO3− reductase (idrA). Based on genetic and physiological data, we identify a respiratory pathway that provides an energy yield equivalent to that of nitrate or perchlorate respiration. Using a variety of public databases, we identify at least 145 organisms that possess the requiste genes for DIR in a mobile cluster with a conserved association across known and predicted DIR microorganisms (DIRM). Consistent with the ecological niche expected of such a metabolism, idrA is enriched in the metagenome sequence databases of marine sites with a specific biogeochemical signature (high concentrations of nitrate and phosphate) and diminished oxygen. Taken together, these data suggest that DIRM help explain the disequilibrium of the IO3−:I− concentration ratio above oxygen-minimum zones and support a widespread iodine redox cycle mediated by microbiology.

Intraspecific variation in time-based transcriptional networks underlies differential stress responses in *Brassica rapa*

Brassica rapa is a physiologically and genetically diverse species containing several economically important specialty crops (Chinese cabbage, pak choi, oilseed, turnip and leafy vegetable varieties). Following divergence from Arabidopsis approximately 24 million years ago, *B. rapa* underwent a whole genome triplication. We are interested in how expansion of the genome has affected diel regulation of plant stress responses. Morphotypes within *B. rapa* have varying cold tolerance, with some that are cold-hardy (Chinese cabbage) while others are relatively sensitive (oilseed varieties). To uncover what transcriptional networks underlie these differences we performed a five-day cold-stress time-series experiment in eight morphotypes (CSP 504418). Incorporating time into co-expression network analyses allowed us to identify differential diel transcriptional patterns. We found that a large proportion of retained paralogs between morphotypes have altered expression patterns. Within these paralogs, we find significant enrichment for one copy being cold responsive over both, suggesting functional divergence. Surprisingly, we also find intraspecific variation for the paralog that is cold responsive, indicating morphotype-specific regulation of cold response. Identifying the regulatory genes driving these dynamic network responses will facilitate future efforts to manipulate transcriptional patterns to improve crop tolerance.

“MetaHipMer: a distributed assembler for terabases of metagenomic data”

Authors: Robert Riley, Rob Egan, Steven Hofmeyr, Alicia Clum, Evangelos Georganas, Alex Copeland, Emiley Eloe-Fadrosh, Simon Roux, Eugene Goltsman, Aydin Buluc, Jeff Kimbrel, Jennifer Pett-Ridge, Daniel Rokhsar, Leonid Oliker, Katherine Yelick

Terabases of environmental metagenomic datasets present an opportunity to characterize complex environments, study biogeochemical processes, and discover novel lineages of life. Metagenome coassembly, in which many samples are assembled together, has distinct advantages over multiassembly, in which samples are assembled one at a time. Coassembly yields more contiguous assemblies, more high quality bins, greater taxonomic diversity, and allows the detection of rare, low-abundance microbes. However, coassembly of terabases of metagenome data is difficult, if not impossible, due to the memory limitations of assembly software, most of which can only run on single compute nodes. The ExaBiome Project is applying MetaHipMer, an assembler that runs distributed across hundreds or thousands of compute nodes, to coassemble terabases of metagenomic data from JGI user science projects covering a broad range of environments, with dataset sizes ranging from 3 TB to 25 TB, and beyond.

The molecular response of the rumen microbiome to the methane-reducing red seaweed *Asparagopsis armata*

Enteric fermentation represents the largest single source of anthropogenic methane (CH_4) emission in the United States, accounting for ~30% of the total CH_4 emitted nationwide. Due to the significant impact of CH_4 on climate change and the negative correlation of animal productivity and enteric CH_4 production, there is great interest in identifying feed additives that might mitigate CH_4 synthesis in the rumen ecosystem. Macroalgae belonging to the genus *Asparagopsis* have shown to reduce methane production during rumen fermentation up to ~90% when added to the feed. In this project we used a combination of metagenomes and metatranscriptomics to determine the molecular response of the rumen microbiome in response to *Asparagopsis armata* added to commercial cattle feed. Analysis of metatranscriptome data generated as part of a New Investigator CSP suggests that bioactives that are present in *A. armata* result in the downregulation of several methanogenesis genes. These genes could potentially serve as targets for advanced methane mitigation strategies using synthetic biology and allow a highly regulated manipulation of the hydrogen flow and therefore of CH_4 production in the rumen ecosystem.

Stable isotope enabled metabolomics of methanogenic soils

Abstract:

Wetland restoration efforts are critical to enhance carbon (C) sequestration in biomass and reestablish ecosystem vitality and functions. The success of such restoration effortsare frequently assessed in terms of microbial processes like slow decomposition of soil organic matter, and development of chemically reduced soil environment resulting in iron reduction and methane production. The mechanistic understanding of how the complex dynamics of microbial community development, metabolic expression and activity co-occur or compete in their local redox environment can greatly impact our predictive understanding of these processes, generally in submerged environments, and specifically in ecosystems undergoing rapid changes in historic hydrologic conditions.

Wetlands are among the largest natural contributors to the global emission of methane. Methane producing microorganisms or methanogens have a very limited substrate range and their *in situ* activities are often linked to intermediary ecosystem metabolism, i.e., a complex food web of interconnected microorganisms that catalyze essential intermediary processes that ultimately drive methanogenesis. For example, fermentation products like short-chain fatty acids and alcohols can be utilized by both iron reducers and methanogens. Thus, methane production may be competing for intermediary substrates formed as a result of microbial metabolism located higher up in the redox ladder, a concept rarely tested in natural soils. We approach this scarcely studied paradox in the context of microbial community stability (insensitivity to disturbance, i.e. altered redox state) and resilience (a community's return to a pre-disturbance condition) and test its relevance to wetland restoration goals.

The energetic favorability of processes associated with alternative terminal electron acceptors (nitrate, sulfate, iron) govern wetland carbon flux and methane biogeochemistry. This study seeks to establish mechanistic links between microbial metabolism to trace gas fluxes to landscape-scale changes in physical (water-level, temperature) and geochemical (redox potential, pH, electron acceptor profiles) properties. Since, organic substrates or metabolites form the primary currency of exchange for microbial growth and activity, understanding how redox potential impacts the net accumulation/consumption of these compounds might be key to quantitatively link process rates. By taking advantage of the high-resolution quantification of soil metabolites, we hope to provide a proof-of-principle for how redox chemistry affects the soil metabolome in natural soils. We collected intact soil cores from freshwater wetlands and subjected them to a time-course stable isotope probing using ^{13}C-acetate as tracer in a controlled sub-oxic environment in the laboratory. Samples were analyzed for metabolomics using a novel ^{13}C-NMR (Nuclear Magnetic Resonance) method. Metabolites were identified and quantified using 2D-NMR and ^{13}C-enrichments confirmed for a number of secondary metabolites including fermentation products. Preliminary results indicate an active pathway for methane oxidation under oxygen-limited conditions. This presentation will expand upon these findings and propose future applications of stable-enabled metabolomics in soils.

Improving the Performance of Apache Spark-based Metagenomic Read Clustering

Sean Saki[1,2], Brain Bushnell[2], and Zhong Wang[2,3].

1. Department of Computer Science, University of California at Merced, Merced, CA; 2. DOE Joint Genome Institute, Berkeley, CA; 3. School of Natural Sciences, University of California at Merced.

Metagenomics is one of the most data intensive biological research areas as thousands of different genomes are being analyzed at the same time. Despite that task-parallel programming paradigm enabled efficient metagenomics solutions on DOE's supercomputers, these solutions are challenging to develop and maintain. In contrast, Apache Spark uses a data parallel paradigm to achieve scalability on big metagenomic datasets. Additionally, Spark supports Python programming language, which makes the platform attractive for data scientists. However, pure Python code has suboptimal performance in Pyspark, Apache Spark's Python API. In this study, we explored various strategies to refactor a traditional Python-based metagenomics solution, metagenome read clustering, to take full advantage of the Spark platform. In our preliminary trials, we replaced functions that are not currently supported by Pyspark with Java and increased the speed by three-folds. We expect further optimization and integration with Pyspark will further increase performance, at least ten times faster than the baseline. In conclusion, we demonstrated that the Apache Spark platform is attractive route to achieve both scalability and performance on large metagenomics datasets.

Predicting plastid targeting in diverse algae using representations from protein language models

Asaf Salamov

DOE Joint Genome Institute

JGI's Algal Genomics Program aims to sequence diverse algal species spanning nearly all of eukaryotic tree of life and there is an interest in algal community to reliably identify and annotate all plastid genes in corresponding genomes.
Most of plastid proteins are encoded in nuclear genome in precursor form, with N-terminal presequences called 'transit peptides' (TPs), mediating the protein through a post-translational targeting pathway.
Many of computational programs developed for predicting plastid targeting were trained and tested on plant sequences and extreme heterogeneity of TPs both in sequence and length led to the development of few algal-clade specific programs, such as Predalgo for green algae and ASAFind for diatoms.
Recently protein language models (LM) have been trained on very large protein databases in analogy with LMs initially developed for natural language processing, which led to significant improvements in protein structure prediction. We developed a general machine learning method for plastid targeting, based on representations (embeddings) from such LMs, that could be used both for plant and algal sequences and which significantly outperforms all other available programs.

Title: **Abundance, distribution and evolution of Transposable elements (TE) families in the model grass genus *Brachypodium***

Rubén Sancho[1,2], Virginia Scarlett[3,4], John Vogel[3,4], David Des Marais[5] and Pilar Catalán[1,2,6]

1 Department of Agricultural and Environmental Sciences, High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain

2 Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Spain

3 DOE Joint Genome Institute, Berkeley, CA, USA

4 The University of California at Berkeley, Berkeley, CA, USA

5 Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA. USA

6 Tomsk State University, Tomsk, Russia

e-mail:

Rubén Sancho: rsancho@unizar.es

Virginia Scarlett: virginia.t.scarlett@gmail.com

John Vogel: brachypodium@gmail.com

David Des Marais: dldesmar@mit.edu

Pilar Catalán: pcatalan@unizar.es

Transposable elements (TEs) are mobile DNA sequences that represent a large genomic fraction and a source of genetic variability in the eukaryotic genomes. We investigate the abundance, distribution and evolution of TE families in five species of the model grass genus *Brachypodium*: the three annual species of the *B. distachyon* polyploid complex (*B. hybridum*-4x and its progenitors *B. distachyon*-2x and *B. stacei*-2x), and the ancestral *B. mexicanum*-4x and recently evolved *B. sylvaticum*-2x perennials. Our preliminary results show that the *B. distachyon* genome is more enriched in TEs (36.4% of masked genome) than that of *B. stacei* (31.1%), for a similar genome size, although their descendant allopolyploid *B. hybridum* shows more TE enrichment (37.3%) than any of its progenitors for a duplicated genome size. The two *B. hybridum* homeologous subgenomes D (distachyon-type) and S (stacei-type) show TE enrichments with respect to the chromosomes of their respective progenitor genomes (1.5-2% more than *B. distachyon* and 4-5% more than and *B. stacei* masked genomes per chromosome), which would support the hypothesis of increased enrichment of the repeteome after the polyploid shock. *B. hybridum* has an intermediate proportion of Gypsy TEs (10.9% of total masked genome: 13% B. hybridum-D and 8.5% B. hybridum-S) between those of its progenitor genomes (*B. distachyon*: 12.7%; *B. stacei*: 6.5%) but a larger proportion of TIRs and non-TIR/helitrons (9.8%: 9.6% B. hybridum-D and 10% B. hybridum-S, and 12.4%: 11.8% B. hybridum-D and 13% B. hybridum-S, respectively) than them (*B. distachyon,* TIR: 8.8% and non-TIR: 9.6%; *B. stacei*, TIR: 8.5% and non-TIR: 10.7%). *B. mexicanum* accumulates the highest repetitive genomic proportion within the genus (71.2% of its genome corresponds to TEs), being 40.4% of it Gypsy LTRs. This value duplicates the TE proportion of any other species

studied in *Brachypodium*, and largely explains the paramount differences found among its large genome size (1.5 Gbp) and the small genomes of the other polyploid species of the genus (e.g., *B. hybridum*: 509 Mbp). The abundance, distribution and dynamics of the TE families in the studied *Brachypodium* species are interpreted in the light of their phylogeny.

"In the allotetraploid Brachypodium hybridum, diploidization begins with transposable elements" -

Authors:
Virginia Scarlett, John Lovell, Jeremy Phillips, Shengqiang Shu, Jacob Espinosa, Jerry Jenkins, Jane Greenwood, Buolos Chaloub, Pilar Catalan, John P Vogel

Abstract:
The ‘genomic shock’ hypothesis posits that unusual challenges to genome integrity such as polyploidization may induce genome restructuring and activation of transposable elements (TEs). Decades of polyploidy research have revealed that this is often, but not always the case. While some species show massive genome restructuring and a ‘burst’ of TE activity in the first few generations following polyploidization, other species show little to no response. Here, we present a high-quality reference genome for an allopolyploid species, Brachypodium hybridum, that has previously revealed no signs of ‘genomic shock’. Two natural, independent lineages of B. hybridum, one that formed 1.4 million years ago, and another that formed 140 thousand years ago, represent two snapshots in polyploid evolution. In the newly sequenced genome of the older lineage, we see some evidence of minor gene loss and genome rearrangement, consistent with the relaxed selection that is characteristic of polyploids. We also find that transposable elements are gradually being exchanged between the two subgenomes, indicating that in allopolyploids with strict disomic inheritance, gradual transposition, rather than massive and sudden chromosomal rearrangements, may be a critical mechanism of evolutionary novelty. This system provides an opportunity to understand genome evolution in genetically stable allopolyploids, and counters the paradigm that ‘genomic shock’ is the norm for polyploid plants.

“From genome-scale to ecosystem-level microbial metabolism using KBase”

Authors: Daniel Segre

Abstract:

Are microbial communities and their ecological functions predictable through mechanistic computational models? In addressing this question, we hypothesize that metabolism operates as a multiscale process that ultimately dictates dominant interactions between microbial species and leads to division of labor and cross-feeding in complex microbial communities. One of the main tools we use to study ecosystem-level metabolism is constraint-based stoichiometric modeling. In particular, we combine dynamic flux balance modeling with convection-diffusion equations to model the spatio-temporal dynamics of communities in complex environments, starting from the genomes of individual species. By enabling systematic analysis of genomes to construct and analyze models, KBase plays an important role in this endeavor. We envisage that hybrid approaches combining data-driven analyses and mechanistic models will be increasingly valuable for understanding the role of microbes and their interactions in large-scale ecosystem processes. In addition to being relevant for the study of complex natural communities (e.g. plant-associated consortia), these approaches can guide targeted synthetic ecology experiments with artificial communities that can be used both as benchmark for our models and as valuable avenues for multi-species metabolic engineering.

Title: Developing an *in-silico* pipeline for the mining of anaerobic *o*-demethylase enzyme systems from lignocellulose-rich metagenomes

Anupama A. Sharan[a*], Camilla L. Nesbø[a,b], Emma R. Master[a,c], Elizabeth A. Edwards[a,d]

[a]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Ontario, Canada. *E-mail: anupama.sharan@mail.utoronto.ca
[b]Department of Biological Sciences, Faculty of Science, University of Alberta, Edmonton, Canada. [c]Department of Bioproducts and Biosystems, Aalto University, Espoo, Finland.
[d]Department of Cell and Systems Biology, University of Toronto, Ontario, Canada.

Corrinoid dependent *o*-demethylase enzyme systems are found in anaerobic acetogens and organohalide-respiring organisms and typically consist of four component gene products, two to three of which occur co-localised[1]. There are two methyltransferases (substrate interacting MT1 and metabolite interacting MT2), a cobalamin dependent corrinoid protein (CP) which is the methyl group shuttle and an activating enzyme (AE) that regenerates the CP in the catalytic cycle. The first three components occur together, co-localised either in triplets or in pairs of one CP and MT1 or MT2. The AE occurs elsewhere in the genome. These gene products work together to catalyse reversible demethylation of aromatic methyl ether compounds[2]. Aromatic methyl ethers are oligomeric degradation products of lignin, the most abundant heteroaromatic renewable biopolymer in nature. Demethylation of lignin and related oligomers is important for increasing the reactivity of their inert structure. Therefore, *o*-demethylation becomes very important for development of development of high-value bioproducts from lignin.

Despite the potency of these enzyme systems in transforming aromatic biopolymers, publicly available sequence and biochemical data about them is not particularly organized. To enable the discovery of more promising candidates of these enzyme systems from big biological datasets such as metagenomes, a streamlined bioinformatics pipeline was created[3]. This pipeline uses the unique co-localization feature of the enzyme system to selectively search for robust candidates for downstream heterologous production and biochemical assaying, searching against HMM models generated from enzymes previously published with biochemical data[1,2].

Using this pipeline, fifteen metagenomic datasets, including nine datasets from JGI-IMG, were analysed and a total of thirty novel *o*-demethylase co-localised system genes were recovered. Sequence similarity networks[4] and phylogenetic trees were also created with sequences from metagenomic datasets and sequences from NCBI RefSeq database mined using the same pipeline. This visualization also helped uncover trends in the phylogenetic and ecological distribution of the different components of the enzyme system. Within the scope of the metagenomic environments referenced in this study, the *o*-demethylase function seems to be carried out by a relatively conserved set of microorganisms across different environments. The most prominent taxa belong to *Firmicutes* phylum and include *Sporomusaceae*, *Oscillospiraceae, Anaerovirgula*; sulfate reducers such as *Dethiobacteraceae* and *Desulfosprosinus* and organohalide respiring *Dichloromethanomonas*. Other taxa include *Acidobacteria* and

Desulfobacteraceae. All these taxa represent novel sources of *o*-demethylase enzymes that have not been described in literature previously. Furthermore, these visualizations show that the MT1 gene is the least conserved and MT2 is most conserved across taxonomies in terms of sequence similarity. This is also supported by the biological function of these gene products. MT1 typically interacts with the phenolic substrates that vary across environments while MT2 interacts with metabolic intermediates that are similar in the organisms catalysing *o*-demethylation.

The developed pipeline has therefore not only led to the discovery of novel *o*-demethylase targets for downstream production, but also added important insights about *o*-demethylase system gene occurrences and organization across important biological spaces.

1. Schilhabel, A. et al. The ether-cleaving methyltransferase system of the strict anaerobe *Acetobacterium dehalogenans* : analysis and expression of the encoding genes. Journal of Bacteriology 191, 588–599 (2009).
2. Farnberger, J. E. et al. Biocatalytic methylation and demethylation via a shuttle catalysis concept involving corrinoid proteins. Communications Chemistry 1, (2018).
3. Sharan, A. A. *O*-demethylase_metagenome_search: Instructional tool to help retrieve co-localised *o*-demethylase operons from metagenomic data. GitHub (2021). Available at: https://github.com/anupamasharan/o-demethylase_metagenome_search.
4. Viborg, A. H. et al. A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). Journal of Biological Chemistry 294, 15973–15986 (2019).

Diel metabolism in microbial mats uncovered by metatranscriptomics

Amanda N. Shelton*, Zeqian Li*, Arthur Grossman, Seppe Kuehn, Devaki Bhaya
*contributed equally

The alkaline siliceous hot spring microbial mats of Yellowstone National Park are moderately complex ecosystems that represent excellent photoautotrophic model ecosystems for studying interactions in microbial communities. Cyanobacteria, the primary producers, and the filamentous anoxygenic phototrophs of the phylum Chloroflexi dominate the upper 1 mm of the mat, along with less abundant heterotrophs. During the day-night cycle, members of the mat experience large environmental fluctuations in light intensity, oxygen levels, pH, and other parameters. Performing in situ metatranscriptomics over the day-night cycle has allowed us to examine how members of the microbial mat community alter their expression in response to diel changes in the environment, both individually and as a community. Supported by a 2018 JGI CSP, we generated metagenomic sequences from 34 samples derived from two springs, over a period of 5 years and at 4 temperatures (from 50◦C-65◦C). We additionally sampled three metatranscriptomic time series. We mapped the metatranscriptome time series to the metagenome from the same sample, followed by normalization and co-expression analysis. We focused additional analyses on expression of genes associated with carbon and nitrogen metabolism.

Many genes (with KEGG Orthology (KO) annotation), exhibited diel-dependent expression dynamics. For carbon and nitrogen metabolism genes, co-expression analyses placed the majority into two groups, one highly expressed in the afternoon/evening (TCA cycle, photosynthesis, certain fermentation genes, certain nitrogen metabolism genes) and one peaking in the early morning (glycolysis, nitrogen fixation, certain fermentation genes, certain other nitrogen metabolism genes). While some patterns are easy to interpret (for example, photosynthesis and nitrogen fixation), metabolic implications of all of the correlated expression patterns are currently being explored.

To further interrogate the metabolism of the major mat genera (*Synechococcus*, *Chloroflexus*, *Roseiflexus*, and *Chloracidobacterium*), the time series of genes and pathways involved in organic acid transformations were examined. Organic acids produced by the cyanobacterium *Synechococcus* are hypothesized to be a major source of carbon and reducing power for other mat organisms. *Synechococcus* had large diel changes in expression of many genes involved in pyruvate and acetate metabolism, while other species showed constitutive expression of these genes.

In a parallel approach, isolates of the major hot spring mat genera such as *Synechococcus* OS-B', are being analyzed in culture at the transcriptome and metabolite levels for their responses to specific environmental perturbations. Our ongoing in situ metatranscriptomic analyses will improve our understanding of the ways in which the temporal metabolic networks of the microbial mat community are controlled and how different organisms contribute to community survival.

Orchestration daytime metabolism of sugarcane and energy-cane: a metabolomic, transcriptomic and non-structural carbohydrate analysis

Jovanderson J. B. da Silva, Luís Guilherme F. de Abreu, Allan Jhonathan R. Ferrari, Lucas M. de Carvalho, Adriana Grandis, Marcos Silveira Buckeridge, Taícia Pacheco Fill, Gonçalo A. G. Pereira, Marcelo Falsarella Carazzolle

Energy-cane is a commercial hybrid originated from the crossing of commercial sugarcane (high sugar content) and *Saccharum spontaneum* (high fiber content) plants. The result of this cross resulted in plants with high fiber content, low sucrose content and higher productivity (2,5 times more) compared to commercial sugarcane hybrids, which made energy-cane an ideal plant to produce ethanol from second generation and bioelectricity. In Brazil, some companies have developed commercial varieties of energy-cane, such as GranBio S/A. Although sugarcane is widely studied from a genomic point of view and through molecular biology, very little is known about energy-cane. Here, we analyze carbohydrates and use transcriptomics and metabolomics approaches to understand how the profile of carbohydrates, metabolites and transcripts from the leaf (source) and stalk (sink) tissues of energy-cane and sugarcane behave during diurnal cycle and how these tissues interact metabolically with each other to adapt to the environment and sustain their growth. For this, we carried out an experiment in which the leaf and stalk tissues of sugarcane and energy-cane were collected during a 24-hour diurnal cycle. Then, carbohydrates were extracted and analyzed by HPAEC-PAD. To explore metabolic differences, samples were subjected to LC and GC-MS approaches. The spectra obtained were quantified by MZMine2 and the compounds identified through the GNPS database (Global Natural Product Social Molecular Networking) and as complementary identification we used CANOPUS (SIRIUS v. 4.9.3). To analyze the transcriptome, the RNA samples were sequenced by Illumina/HiSeq 4000. The transcriptome was assembled by Trinity (2.12.0), gene expression was obtained through Kallisto (v. 0.46.2) and annotation of transcripts was performed with PANNZER (Protein ANNotation with Z-scoRE). As a result, we identified that sugars (sucrose, glucose, fructose) and some metabolites underwent a phase shift during the diurnal variation in leaf and stem tissues and many biological pathways were more active in energy-cane than in sugarcane, such as the phenylpropanoid pathway. Although energy-cane has less sucrose than sugarcane, a higher concentration of starch was observed in the internodes of the energy-cane, suggesting that energy-cane uses this polymer to supply its bioenergetic metabolism. Most of the metabolites identified during the diurnal cycle were classified as organoheterocyclic compounds, oxygenated organic compounds, organic acids and derivatives, benzeneides, lipids and related molecules and phenylpropanoids. We also show that the greatest metabolic differences between sugarcane and energy-cane occur during the night period. The transcriptome is under analysis, it is expected that the expression of genes related to biomass production are more active in energy-cane, as seen in the metabolome. As next steps, we will integrate transcriptomics and metabolomics data through co-expression network analysis. Thus, the knowledge generated about the metabolism of energy-cane compared to sugarcane will enable the development of more productive varieties to supply the bioenergy sector.

“Protein-dipeptide interaction network regulating Arabidopsis central carbon metabolism”

Authors:
Aleksandra Skirycz, Juan Moreno, Boris Zuhlke, Zoran Nikoloski

Abstract:
Plants constantly rewire their metabolism to meet ever-changing cellular demands. Elucidation of the underlying regulatory mechanisms constitutes one of the grand challenges in metabolism research with a direct translational impact on plant growth. Protein-metabolite interactions (PMIs) play a central role in controlling key enzymatic activities and thus metabolic fluxes. To uncover novel regulatory PMIs, we developed PROMIS, an original biochemical approach that relies on the co-fractionation of proteins and associated small-molecule ligands, combining size exclusion chromatography of native protein-metabolite complexes with untargeted metabolomic and proteomic analysis. Combination of the multiple PROMIS datasets for the model plant Arabidopsis with a supervised machine learning method uncovered a high-confidence PMI network comprising multiple proteinogenic dipeptides and 18 (iso)enzymes of central carbon metabolism. Dipeptides are protein degradation products, and though they have long been known to accumulate in response to environmental perturbations, their biological roles remain elusive. Mechanistic characterization of a selected enzyme (GAPDH) - dipeptide (Tyr-Asp) pair revealed a direct, inhibitory interaction associated with a significant change in carbon flux and steady-state metabolite levels directly impacting plant growth. Against this background, we hypothesize that the newly discovered enzyme-dipeptide interaction network contributes to the regulation of central carbon metabolism. Furthermore, we speculate that the different dipeptides will act at distinct points of flux control, leading to diverse metabolic and growth phenotypes. Future research will focus on the characterization of a hidden world of dipeptide-enzyme interactions that remains largely un-probed to date, and its role in the critical, yet still poorly understood, regulatory nexus of proteostasis and metabolism.

Comparative genomics of pyrophilous macrofungi

Andrei S. Steindorff[1*], Kyungyong Seong[1,2], Akiko Carver[1,2], Sara Calhoun[1], Monika Fischer[2], Kyra Stillman[2], Haowen Liu[2], Anna Lipzen[1], Guifen He[1], Mi Yan[1], Bill Andreopoulos[1], Jasmyn Pangilinan[1], Kurt LaButti[1], Vivian Ng[1], Matthew Traxler[2], Thomas D. Bruns[2], Igor V. Grigoriev[1,2]

[1]US DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA
[2]Plant and Microbial Biology, UC Berkeley, Berkeley, CA

Understanding post-fire soil systems are essential because they have significant direct and indirect effects on global carbon storage. Fires result in a large amount of carbon that remains resident on the site as dead and partially "pyrolyzed" (i.e., burnt under low oxygen) material with long residency times and constitutes a significant pool in fire-prone ecosystems. Besides, fire-induced hydrophobic soil layers, caused by condensation of pyrolyzed waxes and lipids, increase post-fire erosion and lead to long-term productivity losses. Soil microbes are likely involved in the degradation of all these compounds, yet little is currently known about the organisms or metabolic processes involved. So far, we have sequenced and annotated four pyrophilous Basidiomycetes and seven Ascomycetes genomes. In our previous work on Basidiomycetes fungi, we found expansion of genes potentially involved in the degradation of the hydrophobic layer, pyrolyzed organic matter, and mushroom formation. In this work, we focused on the seven ascomycetes genomes and compared them with the other 12 non-pyrophilous in the same order and 124 genomes at a larger scale, including pyrophilous Basidiomycetes and other organisms with heterogenous lifestyles. Additionally, we explored enriched Pfam domains and CAZymes to identify patterns associated with these organisms' 'charcoal-loving' lifestyle. Our analyses uncovered gene families related to the degradation of pyrolyzed organic matter, but these gene families were distinct from those expanded in the pyrophilous fungi in Basidiomycota. The enrichment analysis revealed families like peritrophin-A, arthropod defensin, aminopeptidases, beta-glucosidase, heat shock proteins, and fungal fucose-specific lectin. These families might be involved with the pyrophilous fungi' capacity to survive in a toxic environment like post-fire soil. We found a CAZyme CBM14 expanded exclusively in the Pyronemataceae family. This family is mainly found in metazoans and in fungi; it is only found in some Eurotiomycetes. Since it is a chitin-binding domain, this suggests that secreted CBM14 domain proteins might protect the fungus from microbial attacks in its soil habitat. Another interesting finding is that pyrophilous fungi have larger proteins and higher GC3 content than non-pyrophilous, being in an intermediate state to thermophiles. Pyrophilous fungi are commonly found fruiting after fire events, passing through their sexual stages in this process. To make an in-depth comparison of these conditions, we analyzed the available transcriptomic data of *Pyronema domesticum* grown in charcoal and during sexual development. We performed a co-expression network analysis and found two modules with the most differentially expressed genes in charcoal and sexual development. Gene Ontology categories like chitin/carbohydrate/lipid/superoxide metabolism and transport were found in both modules, showing that such processes are likely required to grow in the presence of

charcoal and sexual development. Also, the transcription factors STE12, LreA, LreB, VosA, and EsdC involved in mating response and environmental sensors in yeasts and filamentous ascomycetes were up-regulated in charcoal. This study will improve our understanding of this unique lifestyle of pyrophilous fungi and their role in post-fire carbon cycling.

Characterizing Novel Carbohydrate-Active Enzymes from the Cow Rumen Microbiome: A Course-based Undergraduate Research Experience (CURE)

Madison Stewart, Sheyenne Black, Matthew Escobar, Matthias Hess

Cellulose is a vital structural component of plant cell walls that is composed of β1-4 linked glucose molecules. When cellulose is broken down, its glucose molecules can be fermented into ethanol. Cellulose-based bioethanol offers a promising approach to mitigating carbon emissions through a cleaner, more renewable form of energy. However, cellulose's rigid structure is resistant to chemical breakdown through conventional methods, especially on an industrial scale. Still, many microorganisms can rapidly degrade cellulose using natural enzymatic mechanisms. A previous study identified thousands of genes encoding potential carbohydrate-active enzymes (CAZymes) in the cow rumen microbiome. As a part of an undergraduate Molecular Biotechnology course laboratory at California State University San Marcos, we (M.S and S.B.) identified and characterized a novel cellulase enzyme from the cow rumen microbiome. The CJD5-110 gene encodes a protein with a glycoside hydrolase 5 (GH5) domain. This gene was PCR amplified from cow rumen microbial DNA and inserted into a pET-based *Escherichia coli* expression vector. Production of the resulting recombinant protein in *E. coli* demonstrated that the enzyme was soluble and could efficiently hydrolyze carboxymethylcellulose, a structural analog of cellulose. The His-tagged recombinant protein was also partially purified using nickel ion chromatography. Overall, this work highlights metagenome data as a potential starting point for impactful undergraduate course research experiences without the need for sophisticated infrastructure and laboratory equipment.

Temperate forest versus grassland ecosystems: soil pH predictability and stochastic versus deterministic assembly of soil microbiomes

Laura Super

The relative importance of stochastic versus deterministic community assembly of microbes in ecosystems is still not well understood. This research tests the prediction that soil microbiome communities in temperate forests are more deterministic than in temperate grasslands across the seasons and years, and this effect is related to pH and soil layer (organic versus mineral). Data from the National Ecological Observatory Network (NEON), a continental-scale, open data project representing North American ecosystems, were explored for July (summer) and October (fall) in the years 2016 and 2017 to compare the soil pH predictability (lower variation means higher predictability) and also the relative importance of stochastic versus deterministic assembly (with the normalized stochasticity ratio [NST]: deterministic being less than 50% and more stochastic being greater than 50%) of soil microbiomes in temperate forests (four sites with up to 10 plots and three subsamples) and grasslands (three sites with up to 10 plots and three subsamples). These preliminary data are still being fully analyzed, but the current results will be presented.

A robust mutation filtering-followed QTL mapping assay to uncover the causative gene for ethanol resistance in an industrial bioethanol yeast strain

Larissa Escalfi Tristão[a,1], Marcelo Falsarella Carazzolle[a], Juliana José[a], Gonçalo Amarante Guimarães Pereira[a], Fellipe da Silveira Bezerra de Mello[a]

[a]Departamento de Genética, Evolução e Bioagentes, UNICAMP, Campinas, SP, Brazil

Ethanol tolerance is a desirable trait in yeast strains used in the alcohol industry, especially for corn ethanol, since high concentrations of this hydrocarbon is considered to be a growth inhibitor of *Saccharomyces cerevisiae*, posing burdening productivity loss. Like other stress-resistance phenotypes, yeast vigor to ethanol is a complex and polygenic trait - described as quantitative. Therefore, in this work we performed a Quantitative Trait Loci (QTL) mapping followed by a robust variant-calling filtering to identify alleles related to this phenotype. First, we have identified a superior strain that can endure strong selective pressure, regarding ethanol concentration in the culture medium. Using a library of 78 industrial and laboratory *Saccharomyces spp* , we performed a high-throughput screening on different ethanol concentrations (from 8 to 14%), which revealed that Brazilian bioethanol strain SA-1 (MATa/MATα) ranked amongst the best performing - while laboratory CEN.PK122 (MATa/MATα) failed to keep up with the increasing ethanol-induced stress. Segregant haploid cells of SA-1 were collected and phenotyped in 14% ethanol, which allowed selection of segregant SA49 (MATa) as the top-performing strain, that was further crossed with susceptible CEN.PK113-1A (CEN.PK122, MATα), resulting in hybrid LTY003. A F_2 population of 400 individual haploids was screened and a pool of 60 segregants was selected as ethanol-resistant for QTL mapping. A bulk segregant analysis assay was performed with an ethanol-susceptible pool of the same size. After DNA sequencing and QTL analysis, four peaks of Δ(SNP-index) - a variable that correlates mutations between phenotypically distinct pools - above a confidence interval of 99% were observed in chromosomes III, VII, XIII and XV, the last representing the most prominent. Next, we annotated all high quality single nucleotide polymorphisms at the statistically significant *loci* at chromosome XV surrounding its peak (177 kbp), and non-coding and synonymous mutations were discarded. Finally, non-synonymous mutations were filtered out based on allelic frequency using the 1,011 publicly available yeast genomes, in which only genes containing at least one rare mutation (<= 10% in population) were retained in the final set. We identified only one gene containing three rare non-synonymous mutations with a prevalence lower than 7% in the *S. cerevisiae* population, suggesting that this allele might be the responsible for ethanol resistance in strain SA-1. Reciprocal hemizygosity analysis is now being performed for confirmation.

The contribution of bacteriophages to genetic and phenotypic diversity of *Burkholderia pseudomallei*.

Iris Irby, Pacharapong Khrongsee, and Apichai Tuanyok

Department of Infectious Diseases and Immunology, College of Veterinary Medicine, University of Florida, Gainesville, FL 32610.

Burkholderia pseudomallei causes melioidosis, a severe tropical disease in humans and animals with great biodefense implications. It has been estimated that the global fatality of human melioidosis is comparable to death from measles and substantially more significant than those from dengue and leptospirosis. *B. pseudomallei* can be found in soil, and fresh water in most tropical regions but is highly endemic in Southeast Asia and northern Australia. The life cycle of *B. pseudomallei* is unknown, even though it has been postulated that environmental factors could contribute to its virulence and increase disease incidence. Bacteriophages are key biological factors influencing genetic recombination and predators of *B. pseudomallei*. Phage transduction is one of the most common genetic recombination events found in most *B. pseudomallei* strains. Here, we have shown that multiple functional prophages are excised spontaneously from the *B. pseudomallei* genomes during normal bacterial growth, replicate, and then lyse the host cell to disperse as lysogenic or temperate phages. Subsequent infection of another bacterial cell is facilitated by mobile genetic elements that recognize the 3' end sequence of tRNA genes. This mechanism is known as the tRNA-mediated site-specific recombination or tRNA-SSR. At least 1,800 genomes of *B. pseudomallei* were used in the analysis. Most *B. pseudomallei* strains are in lysogenic forms that possess at least one functional prophage in their genomes. Analysis of prophages and phage genome sequences has identified ten hotspots for phage recombination in *B. pseudomallei*, eight of which are associated with various tRNA genes including tRNA- Phe-GAA, Met-CAT, Pro-TGG, Arg-TCT, Cys-GCA, Arg-CCG, Ser-GGA, and Sec-TCA. These prophages are also common in other species within the *B. pseudomallei* complex (BPC). *By morphology and genomic composition, B. pseudomallei* - temperate phages are classified into two prominent phage families, Myoviridae and Siphoviridae. Further investigation on the interactions between *B. pseudomallei*, bacteriophages, and other environmental and biological factors would provide a bigger picture of *B. pseudomallei* genomic diversity that potentially influences its survival in the environments and pathogenic specialization in hosts.

Lytic archaeal viruses of the continental deep biosphere challenge a long-standing paradigm

Victoria Turzynski[1], Janina Rahlff[2], Sarah P. Esser[1], Indra Monsees[1], Till L.V. Bornemann[1], Lea Griesdorn[1], Perla Abigail Figueroa-Gonzalez[1], Frederik Schulz[3], Tanja Woyke[3], Andreas Klingl[4], Cristina Moraru[5], Alexander J. Probst[1]

[1]Group for Aquatic Microbial Ecology, Biofilm Center, University of Duisburg-Essen, Universitätsstraße 5, 45141 Essen, Germany

[2] Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Department of Biology and Environmental Science, Linnaeus University, SE-39182, Kalmar, Sweden

[3] DOE Joint Genome Institute, 1 Cyclotron Rd, Berkeley, CA, 94720, USA

[4] Plant Development & Electron Microscopy, Biocenter LMU Munich, Großhaderner Str. 2-4, 82152 Planegg-Martinsried, Germany

[5] Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl-von-Ossietzky-University Oldenburg, Carl-von-Ossietzky-Straße 9-11, 26111 Oldenburg, Germany

Introduction: *Candidatus* Altiarchaea can dominate deep subsurface ecosystems by fixing carbon dioxide and thus act as the main primary producer under oligotrophic conditions. Microbial interactions of *Ca.* Altiarchaea with bacteria and a symbiosis with other uncultivated archaea are known, yet archaeal viruses infecting these primary producers and their subsequent impact on the ecosystems' food webs remain uncertain.

Objectives: This study was set out with the aim of identifying uncultivated viruses of the *Ca.* Altiarchaea in different deep biosphere ecosystems.

Material & Methods: We linked metagenomics with virus-targeted direct-geneFISH, a modified fluorescence *in situ* hybridization method of direct-geneFISH and viral FISH, to detect putative viral infections in *Ca.* Altiarchaeum biofilms. FISH probes were specifically designed for targeting a putative 8.9 kb circular viral genome, which was previously *in silico* identified as a putative virus. Key findings were obtained by an enumeration of viral infections, classified into three infection categories.

Results: Viral infections of *Ca.* Altiarchaeum biofilms showed lytic features that could have implications for the integrity of the archaeal host and thus on the main carbon cycle in the ecosystem**.** 502 infections of 18,411 counted archaeal cells revealed a low percentage of virus adsorption (8.5%), many advanced infections with intracellular virus signals (76.5%), and states of cell lysis (15%). Moreover, imaging of several biofilm flocks and the analysis of the CRISPR development suggest a potential resistance of *Ca.* Altiarchaeum against their virus.

Conclusion: Overall, we conclude that infections of viruses with a putative lytic lifestyle on, e.g., primary producers, may stimulate heterotrophic carbon cycling in the deep biosphere, which shed new light on the standing paradigm that lysogeny represents the predominant viral lifestyle in the deep biosphere.

Galyna Vakulenko, Xi Zhang, Marina Cvetkovska.

Towards characterizing the abnormal heat stress response of the Antarctic alga *Chlamydomonas* sp. UWO241

Chlamydomonas sp. UWO241 is a psychrophilic alga found 17 m below the permanently ice-covered surface of the Antarctic Lake Bonney, where it experiences a myriad of harsh environmental conditions such as low temperature, low light, and high salinity. While this habitat is extreme, it is also very stable, and this alga rarely experiences changes in its environment. Heat shock proteins (HSPs) are a ubiquitous family of chaperone proteins that perform important housekeeping roles. In general, HSP expression is induced during abiotic stress to regain protein homeostasis – a process regulated by heat shock transcription factors (HSFs). Previous work has shown that UWO241 constitutively accumulates high protein levels of HSPs in steady-state conditions but fails to induce additional HSP accumulation during heat stress. It is hypothesized that UWO241 has lost the ability to regulate HSPs in its extreme but unchanging environment. In this study, a single HSF was identified in UWO241 genome. Comparative sequence analysis with related species revealed conservation in all the domains characteristic of a functional HSF. Next, we performed targeted analysis of the UWO241 transcriptome in heat-stressed cultures. We show that ~26% of UWO241 HSPs were differentially expressed during heat stress; however, the HSF1 transcript was not. Additionally, we have not been able to experimentally detect a full-length HSF1 transcript. More work needs to be done to fully elucidate the regulation of HSPs in UWO241.

Title: Legume invasions: did rhizobia escape phage control?

Authors: Jannick Van Cauwenberghe and Ellen Simms.

Abstract: Global trade has transported many species outside their natural range; some proliferate and negatively impact their new ecosystems. Almost 10% of the invasive plants recorded for North America are legumes (Fabaceae). Legumes benefits from access to atmospheric nitrogen (N) via a specialized mutualistic interaction with soil-dwelling bacteria called rhizobia. This dependence hampers the invasion ability of some legumes, yet rhizobia sometimes co-invade. Thus, understanding legume invasions requires elucidating factors that limit and facilitate rhizobium invasions. Bacteriophages play an important in role in controlling and shaping bacterial communities. We hypothesize that bacteriophage-rhizobium interactions are a key factor in successful legume invasions. Either (1) introduced rhizobia have escaped predation by their home range phages, or (2) native and introduced bacteriophages cause greater mortality in native rhizobium communities than in the introduced communities. To test this hypothesis, we assess the phage host range and the genomic diversity of rhizobia and associated phages from three Californian populations of yellow bush lupine (Lupinus arboreus, native Californian legume), and three Californian and three European populations of French broom (Genista monspessulana, invasive European legume). Furthermore, we will assess the effect native and introduced bacteriophages have on the competition between native and introduced rhizobia using microcosm experiments. Preliminary host range assays indicate that the phage-rhizobium interactions are shaped by the associated legume and geographic origin.

Reactive oxygen species mediate thylakoid membrane remodeling and triacylglycerol synthesis under nitrogen starvation in the alga *Chlorella sorokiniana*

Jithesh Vijayan[2,3], Nishikant Wase[4], Liu Kan[3], Chi Zhang[1,3] and Wayne R. Riekhof[1*]

School of Biological Sciences, University of Nebraska-Lincoln[1], Department of Biochemistry, University of Nebraska-Lincoln[2]; Center for Plant Science Innovation, University of Nebraska-Lincoln[3] School of Medicine, University of Virginia System Charlottesville, VA[4];

*author for correspondence: wriekhof2@unl.edu

ABSTRACT

Many microbes accumulate energy storage molecules such as triglycerides and starch during nutrient limitation. In eukaryotic green algae grown under nitrogen limiting conditions, triglyceride accumulation is coupled with chlorosis and growth arrest. In this study we show that accumulation of reactive oxygen species (ROS) under nitrogen limitation in the microalga *Chlorella sorokiniana* is involved in thylakoid membrane remodeling, leading to chlorosis. We show that ROS accumulation under nitrogen limitation is an active process involving downregulation of expression of ROS-quenching enzymes, such as superoxide dismutases, catalase, peroxiredoxin, and glutathione peroxidase-like, and upregulation of enzymes involved in generating ROS, such as NADPH oxidase, xanthine oxidase and amine oxidases. Expression of enzymes involved in ascorbate and glutathione metabolism are also affected under these conditions. We also show that calcium influx plays a putative role in activation of NADPH oxidases, leading to ROS generation and membrane remodeling. Quenching ROS under nitrogen limitation reduces TAG accumulation, adding additional evidence for the role of ROS signaling in the process.

New flow-through cultivation chambers for the first chemosynthetic symbiosis model system *Zoothamnium niveum*

Volland JM, Emboulé E, Raju Narayanasamy S, Holman HY, T, Andeer P, Northen T, Tyml T, Woyke T, Gros O, Date S.

Chemosynthetic symbioses are partnerships between invertebrate animals and chemosynthetic bacteria. They fuel some of the most productive ecosystems on Earth, such as deep-sea hydrothermal vent communities and can also be found in shallow-water environments such as mangroves. Even while being ubiquitous and evolutionarily relevant, they have received far less attention than heterotrophic (e.g. rumen) and photosynthetic (e.g. coral) symbioses and their research has been hampered by the lack of model systems. The only fast reproducing chemosynthetic symbiosis which has been cultivated in the laboratory is *Zoothamnium niveum*. This giant colonial ciliate thrives in sulfidic marine shallow waters and is entirely covered with a monolayer of sulfur oxidizing bacteria, *Ca*. Thiobios zoothamnicoli. In this mutualistic symbiosis, the ciliate host benefits from carbon transfer, while the bacteria benefit from a competition-free environment and enhanced access to an electron donor (hydrogen sulfide) and acceptor (oxygen). The current state-of-art requires cultivation using custom-made flow-through chambers which are needed to provide the symbiosis with a continuous supply of sulfide and oxygen. Unfortunately, these high-maintenance chambers are expensive and not readily available to the community, limiting the potential of *Z. niveum* to become a true model system. We aim to overcome these obstacles and help establish *Z. niveum* as a first fast reproducing, easy to culture chemosynthetic symbiosis model system in the laboratory that is reproducible and experimentally tractable. As a first step towards this goal, we have developed a new low-cost cultivation device that includes a modified tesla valve ensuring the mixing of sulfide and oxic seawater. High-fidelity complex computational fluid dynamics simulations were performed to determine the ideal fluidic and structural parameters. These devices can easily be produced in any laboratory using 3D models of printable molds for casting slide-bound polydimethylsiloxane (PDMS) chambers. Using this approach, we managed to reduce the production costs by more than 10 fold, while eliminating the need to contract external manufacturing companies. We also reduced maintenance from daily to once every three days allowing the culture to survive over the weekends. We inoculated this improved cultivation system with symbiotic single-cells of *Z. niveum* and successfully cultured centimeter-long colonies which then produced a second generation *in vitro*. Using light and scanning electron microscopy, we confirmed that the lab-grown colonies remained covered with a monolayer of ectosymbiotic bacteria. We believe establishing this cultivation method and chemosynthetic symbiosis system is crucial to research efforts that focus on aquatic symbioses, while allowing more detailed and comprehensive studies on chemosynthetic systems.

Linking specific microorganisms to their functions is one of the fundamental goals of microbial ecology. However, the inability to isolate and cultivate most environmental microbes remains a significant challenge for the field. The recent development of stable isotope probing (SIP) methods provides an alternative approach for culture-independent identification of microbial populations performing a specific metabolic activity in complex ecosystems. In a SIP experiment, a substrate of interest labeled with a heavy isotope (e.g., 13C, 18O) is incubated with a microbiome sample. Members of the microbial community responsible for utilizing the substrate will incorporate heavy isotopes in their nucleic acids. DNA can then be extracted to identify community members whose nucleic acids are enriched with heavy isotopes. This approach allows pinpointing active microbial populations even at a low abundance.

While SIP has traditionally relied on 16S PCR amplicons [1], recent advances in sequencing technologies and bioinformatics tools allow combining DNA-SIP and metagenomics, the study of the bulk genomic DNA recovered directly from environmental samples. This enabled the linkage of functions and metabolic reconstructions to (near-)complete microbial genomes from complex communities instead of single marker genes, as was done previously [2-5].

Community interest in DNA-SIP metagenomics is quickly rising. Unfortunately, the lack of standardized protocols and analysis pipelines makes it impractical to apply this technique in high-throughput. To tackle this challenge, we developed a standardized DNA-SIP metagenomic experiment and analysis framework that can scale the adoption of the method. We demonstrate that the addition of isotopically-labeled synthetic spike-ins for pre- and post-fractionation facilitates quality control, detection of poorly handled samples, normalization, and quantification of the data. With a fully integrated pipeline, researchers will be able to automatically quality control SIP metagenome data, extract a combined set of de novo reconstructed genomes from a single experiment, robustly quantify abundances of these genomes across fractions and samples, and automatically derive enrichment statistics reflecting activity levels of individual microbial populations.

1. Saidi-Mehrabad A, He Z, Tamas I, Sharp CE, Brady AL, Rochman FF, Bodrossy L, Abell GC, Penner T, Dong X et al: Methanotrophic bacteria in oilsands tailings ponds of northern Alberta. The ISME journal 2013, 7(5):908-921.
2. Ziels RM, Sousa DZ, Stensel HD, Beck DAC: DNA-SIP based genome-centric metagenomics identifies key long-chain fatty acid-degrading populations in anaerobic digesters with different feeding frequencies. The ISME journal 2018, 12(1):112-123.
3. Eyice O, Namura M, Chen Y, Mead A, Samavedam S, Schafer H: SIP metagenomics identifies uncultivated Methylophilaceae as dimethylsulphide degrading bacteria in soil and lake sediment. The ISME journal 2015, 9(11):2336-2348.
4. Verastegui Y, Cheng J, Engel K, Kolczynski D, Mortimer S, Lavigne J, Montalibet J, Romantsov T, Hall M, McConkey BJ et al: Multisubstrate isotope labeling and metagenomic analysis of active soil bacterial communities. mBio 2014, 5(4):e01157-01114.
5. Chemerys A, Pelletier E, Cruaud C, Martin F, Violet F, Jouanneau Y: Characterization of novel polycyclic aromatic hydrocarbon dioxygenases from the bacterial metagenomic DNA of a contaminated soil. Applied and environmental microbiology 2014, 80(21):6591-6600.

“Optimizing the thermotolerance of photorespiration in higher plants through robust parameterization of pathway kinetics and in planta validation”

Authors: Berkley Walker, Roze Ludmila, Aaron Lipeman

Abstract:

Although photorespiratory losses of CO2 and energy reduce production in C3 crops by ~30% under field conditions, its network structure and efficiency has predominately been resolved only under standard laboratory temperatures. Furthermore, there is evidence that photorespiration is not currently optimized, for example, photorespiration releases excess CO2 under elevated temperatures and can be re-engineered to increase biomass production in the field. Metabolic reaction-kinetic models of photorespiration have great potential to guide synthetic biology-guided improvement of bioenergy crops, but these models lack functional parameterizations for photorespiratory enzymes with appropriate temperature responses for biofuel crops. Our long-term goal is to identify targets for improving photorespiratory efficiency in bioenergy crops using a systems approach. The goals of this project are to functionally characterize the temperature responses of photorespiratory enzymes to parameterize a reaction-kinetic model of photorespiratory flux identify specific photorespiratory reactions to target for crop improvement. By measuring kinetics of photorespiratory enzymes from C3 bioenergy crops and other species with photorespiration adapted to high temperatures, we will be able to create a more accurate model of photorespiration in current crop systems and have a reservoir of thermotolerant enzymes to test in silico for improved efficiency under elevated temperatures. Ultimately, promising enzyme combinations predicted from the model will be transformed into Arabidopsis thaliana and the efficiency of photorespiration probed using a unique suite of in vivo gas exchange approaches.

Walker, E. J. L., Department of Biochemistry at Western University, Canada, ewalke72@uwo.ca; Pampuch, M., Department of Biochemistry at Western University, Canada, mpampuch@uwo.ca; Cochrane, R. R., Department of Biochemistry at Western University, Canada, rcochra3@uwo.ca; Davis, J. G., Department of Biochemistry at Western University, Canada, jdavi33@uwo.ca; Gloor, G. B., Department of Biochemistry at Western University, Canada, ggloor@uwo.ca; Karas, B. J., Department of Biochemistry at Western University, Canada, bkaras@uwo.ca

Chloroplasts represent a unique opportunity for the metabolic engineering of microalgae. Characteristics of their prokaryotic-derived genomes and compartmentalized micro-environments render the organelle amenable to high levels of transgene expression and foreign protein accumulation. However, the chloroplast's development as a synthetic biology tool has been limited by a lack of robust multigenic engineering techniques and methods for targeted organellar genome delivery. The first objective of this project is to explore two cloning approaches for creating synthetic chloroplast genomes in the diatom *Phaeodactylum tricornutum*: (i) a 'direct' PCR-based approach and (ii) an 'indirect' modular plasmid-based approach. The most efficient approach will then be used to introduce genetic manipulations at multiple loci throughout the exogenous chloroplast genome. Following this, we will explore different methods for the targeted delivery of whole synthetic genomes to the *P. tricornutum* chloroplast.

2021 DOE JGI Annual Meeting

Formation, characterization, and modeling of emergent synthetic microbial communities

Jia Wang[a], Dana L. Carper[a], Leah H. Burdick[a], Him K. Shrestha[a,b], Manasa R. Appidi[a,b], Paul E. Abraham[a], Collin M. Timm[a,c], Robert L. Hettich[a], Dale A. Pelletier[a], Mitchel J. Doktycz[a]

[a]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[b]Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN, USA
[c]Current address: Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA.

Abstract

The microorganisms present in soil are recruited by the plant to form microbial communities in the rhizosphere. Although several important factors such as soil properties, plant genotype and agricultural practices have been studied for their effects in the construction of the rhizosphere microbiome, how these microorganisms contribute to shaping their own community is still poorly understood. Synthetic microbial communities composed of well-studied individual isolates can be valuable model systems for elucidating the organizational principles of communities. Using genome-defined strains, systematic analysis by computational modeling can lead to mechanistic insights and metabolic interactions among species. In this study, 10 bacterial strains isolated from the *Populus deltoides* rhizosphere were co-cultured and passaged in two different media environments to form stable microbial communities. The membership and relative abundances of the strains in community stabilized after around 5 passages and resulted in only a few dominant strains that depended on the medium. To unravel the underlying metabolic interactions within the community, flux balance analysis was used to model microbial growth and predict metabolic interaction involved in organizing the microbial communities. These analyses were complemented by individual growth curves of bacteria, metaproteomics of the community and pairwise interaction between species. A fast growth rate can be advantageous for maintaining survival in the microbial community, and the final presence of a member also depends on selective antagonistic relationships and metabolic exchanges within the community. Revealing the mechanisms of interaction among plant-associated microorganisms provides insights into strategies for engineering microbial communities that can potentially increase plant growth and disease resistance. Deciphering the membership and metabolic potentials of

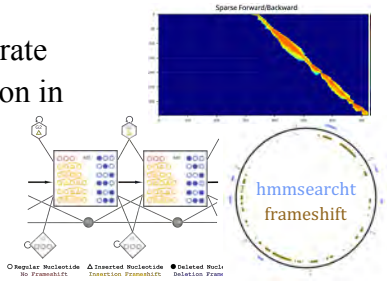
microbial community will also enable the design of synthetic consortia with desired biological functions.

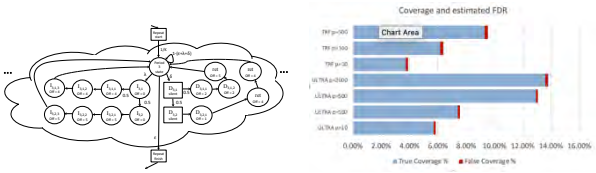
Methods for fast and sensitive (meta-)genome annotation with confidence assessment, and a new lightweight and flexible javascript visualization library

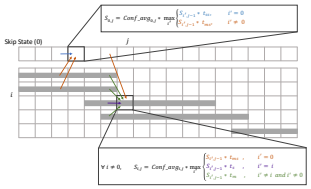
Travis Wheeler, Kaitlin Carey, Genevieve Krause, George Lesica, Daniel Olson, David Rich, Jack Roddy, Audrey Shingleton
Department of Computer Science, University of Montana

Annotation of genomic sequence depends on fast, accurate and reliable methods for classification. Here, we describe advances in methods for annotation with hidden Markov models (HMMs):

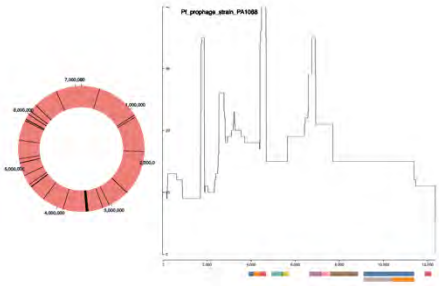
- a sparse dynamic programming algorithm to produce highly accurate Forward/Backward profile HMM alignments with 20-100x reduction in memory and runtime requirements (MMOREseqs);
- an extended profile HMM for labeling protein-coding DNA with frameshifts due to pseudogenization or sequencing error (FATHMM);
- a model for labeling tandemly-repetitive sequence with better sensitivity in the face of degenerate (gappy) repeats, good speed on large repeat regions (e.g. centromeres), and consistent statistics that enable principled soft-masking (ULTRA);
- a method for adjudicating between competing sequence annotations, including uncertainty quantification, overlap arbitration, and identification of nested insertions and instances of recombinations (PolyA).

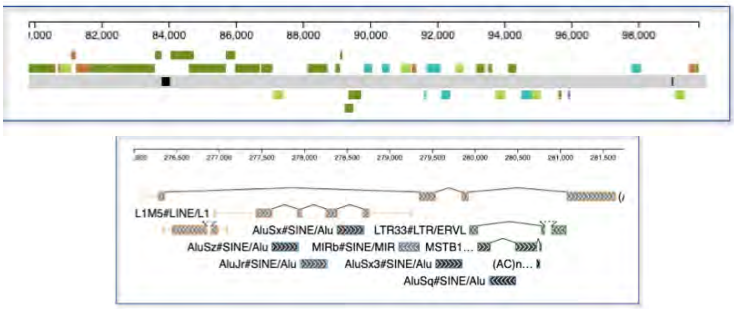






We also present an open-source javascript library that supports efficient development of custom, dynamic, and interactive visualizations of annotations of linear and circular genomic sequence (SODA).





The transcriptome atlas of the extremophyte model *Schrenkiella parvula* provides tissue specific insight into gene expression associated with abiotic stress responses

Chathura Wijesinghege, Guannan Wang, Pramod Pantha, Kieu-Nga Tran, Dong-Ha Oh, Maheshi Dassanayake

Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA

The extremophyte model, *Schrenkiella parvula* thrives in saline soils high in sodium and other salts including potassium and lithium. Its genomic and transcriptomic resources provide an ideal opportunity to investigate how plants cope with environmental stresses and will be instrumental in designing how we may use crops for biofuels grown in marginal lands that will not compete with prime agricultural fields. A spatio-temporal gene expression atlas for *S. parvula* provides a foundational resource for us to identify key genes that show tissue or developmental stage specific expression in response to salt stress. We have analyzed 16 distinct developmental stages and tissues between control and salt treated *S. parvula* transcriptomes. We have identified gene expression modules that are preferentially expressed in limited tissue types or developmental stages compared to core modules that have systemic expression across developmental stages and tissues. Additionally, we have identified genes that are differentially expressed in response to salt stress in each tissue type and have calculated a tissue specificity index for expressed protein-coding genes in the *S. parvula* genome. We also plan to discuss how some of the stress responsive gene networks we have identified in *S. parvula* compare with salt responsive pathways identified for Arabidopsis and highlight how tissue specific expression is achieved by potential gene subfunctionalization following gene duplication. Together these results provide insight into how gene expression programs contribute to environmental adaptations in *S. parvula.*

Genomic analysis of family UBA6911 (Group 18 Acidobacteria) expands the metabolic capacities of the phylum and highlights adaptations to terrestrial habitats.

A. Yadav, J. Borrelli, N. Youssef, M. S. Elshahed; Oklahoma State Univ., Stillwater, OK

Abstract:
Acidobacteria is a widespread bacterial phylum, with diverse metabolism and ecological roles. However, our understanding of Acidobacteria's physiological characteristics and metabolic capabilities is largely derived from the characterization of few Acidobacterial soil isolates. In this study we analyzed 27 genomes belonging to the yet-uncultured Acidobacteria Group 18; previously identified in a wide range of environments beside soil, to understand the physiology, evolution, metabolic and biosynthetic capabilities and ecological abundances of this group. The 27 metagenome assembled genomes (MAGs) analyzed clustered into 2 families and 5 genera. Distinct differences between soil-dwelling versus non-soil dwelling genomes were observed. Soil-dwelling members of group 18 exhibited larger genome sizes, and more regulatory genes, and pathways. Soil-derived group 18 genomes also possessed the capacity to degrade one-carbon compounds (e.g., methanol, methylamines), many of which are abundant as biproducts in soil ecosystems. The non-soil group 18 in contrast, appears to possess more streamlined genome and a capacity for anaerobic nitrate and sulfate reduction. Further, all genomes analyzed also possessed the capability to synthesize a wide range of secondary metabolites. Our results shed light on an enigmatic, yet-uncultured group within the phylum Acidobacteria, and highlight the importance of niche differentiation in shaping genomic architecture and metabolic capacities within the microbial world.

Title: Comparative Analysis on Enzyme Annotation Tools

Authors: Eddie Yang[1], Bishoy Kamel[1], Igor Grigoriev[1]
[1]US Department of Energy, Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley CA

Enzymes contribute to a variety of mechanisms in various organisms, and therefore assigning enzymes with EC numbers becomes a crucial step to understand these mechanisms. While there exists multiple types of tools to annotate enzymes, it is not yet determined which of them will work the best. In this project, we performed comparative analysis on the performance of enzyme annotation tools for a green alga *Chlamydomonas reinhardtii* using eggNOG, PRIAM, Kofam, and Swissprot with MMseqs2. *C. reinhardtii* is a well-studied model organism with its complete genome sequence being published, which will be ideal as the first target for this project. We extracted useful information by comparing annotations from each tool in order to understand the differences between these tools and contribute to a possible annotation pipeline in the future. First, we chose 30 genes from the TCA cycle of *C. reinhardtii* as a small, well-curated test set. Then, we expanded our comparative analysis to 16000+ predicted genes of this organism as the general set.

On average, around 27 of the 30 genes in the TCA cycle were annotated as enzymes, while only around 10% - 15% of all genes in *C. reinhardtii* were annotated as enzymes, with PRIAM and Kofam tending to annotate fewer genes than other tools. The E-value distribution of the TCA-cycle gene set also displays an entirely different trend from the large set. For instance, E-values from eggNOG tend to cluster at a lower value in the small set, which indicates the tool to be more accurate in certain pathways. This finding reminds us to also focus on per-pathway accuracy, as the general-set accuracy may not tell the whole story here. Additionally, two of the tools, eggNOG and PRIAM, provide coordinates of the predicted enzymatic domains, which were used in our investigation of consistency in their annotations. By comparing these coordinates, we discovered that the sequences corresponding to domains predicted by these tools primarily overlap with each other at a level of over 99% on average. Overall, by comparing the annotations with the reference set of enzymes, we find out that eggNOG has the highest accuracy among all 4 tools for the small TCA-cycle gene set, while PRIAM has the highest accuracy for the large whole-genome set.

In the next steps, we will explore more challenging annotations across different genomes, such as multi-enzyme proteins and protein complexes. We believe that this preliminary work is fueling the development of annotation tools to improve our understanding of algal and fungal metabolism using genomes sequenced at the JGI.

Frontiers in Genetic Design with Nonconventional Fungi

Nilesh K. Sharma, Emma Tobin, Jaclyn Thompson, Andres Carrillo, Joseph H. Collins, Kevin Keating, Natalie Farny, **Eric M. Young**

Nonconventional fungi have vital ecological roles and vast potential as cell factories. Yet, these organisms are challenging to engineer as the genetic tools are limited. Here, we describe the application of modern genomics, DNA synthesis, and modular molecular cloning to unlock nonconventional fungi for secondary metabolite biosynthesis and communication in soil. First, we show that DNA synthesis can enable the rapid testing of genetic parts in the CTG clade yeast *Debaryomyces hansenii*. By adhering to a modular cloning standard, we characterize antibiotic selection markers, fluorescent reporters, and transcriptional parts, resulting in a toolkit for manipulating gene expression over three orders of magnitude in this halotolerant ascomycete. Second, we show that high quality genomics and transcriptomics, in addition to DNA synthesis, permits selection of both constitutive and regulated parts. We demonstrate this in the basidiomycete yeast *Xanthophyllomyces dendrorhous*. From differential gene expression analysis of cultures grown in the dark and under ultraviolet light, we derive strong constitutive promoters and light-regulated promoters with over a ten fold dynamic range in gene expression. This –omics derived parts set holds promise for dynamic pathway regulation and photobiology in basidiomycetes. Finally, we show that the unique properties of filamentous fungi enable long distance activation of genetic circuits in soil. We use the mycorrhizal fungi *Lyophyllum atratum* as a fungal highway to deliver engineered soil bacteria *Pseudomonas putida* deep in soil, detect a chemical signal, and send that signal back to the surface. These three applications demonstrate the utility of modern systems and synthetic biology for realizing the potential of nonconventional fungi for ecological and economical purposes.

Increased planting density in bioenergy sorghum enhances GA mediated stem elongation by inducing GA3-oxidase expression in leaves

Authors: Ka Man Jasmine Yu, William L. Rooney & John E. Mullet

Sorghum bicolor is a C4 grass and prominent bioenergy feedstock, that produces ~40 Mg/ha of dry biomass yield under ideal growing conditions. At harvest, ~84% of total bioenergy sorghum biomass accumulates in the stem, which can be >4m long and contain >40 internodes. In this study, we investigated the impact shading has on stem growth and biomass accumulation by growing bioenergy sorghum hybrids at four planting densities ranging from 20,000 to 132,000 plants/ha, under field conditions for 60 days. We found that typical shade avoidance responses were induced by increasing planting density, including an over 2-fold increase in sorghum internode length and a ~22% decrease in stem diameter. This shade-induced internode elongation was due to an increase in cell length and number of cells spanning the length of internodes. We also found that *SbGA3ox2* (Sobic.003G045900), a gene encoding the last step of the GA biosynthesis pathway, was expressed ~20-fold higher in leaf collar tissue of developing phytomers in plants grown at high vs. low planting density. External application of GA3ox to exposed bioenergy sorghum stems increased plant height, internode length, cell length and the number of cells throughout the internodes. We propose that increasing planting density induces the expression of GA3-oxidase in leaf collar tissue, which then increases the synthesis of GA at the stem, thus stimulating internode elongation.