

Title: Adaptive strategies in a cosmopolitan and abundant soil bacterium: genetic microdiversity or core metabolic flexibility?

Description:

We request de novo genome sequencing for 36 bacterial isolates obtained from three soil depths at three distinct locations in the Catalina-Jemez Critical Zone Observatory (CJCZO) (4 isolates \times 3 depths \times 3 sites = 36 genomes). These isolates are Bradyrhizobium species that are very likely to be non-symbiotic. Based on existing genomes for Bradyrhizobia, we expect the genomes to be on the order of 6-10 Mbp each. Assuming a maximum of 10 Mbp per genome, and 100X coverage of each genome, the total isolate genome sequencing request will not exceed 36 Gbp. For these genomes, short read genome sequencing (Illumina) should be sufficient as there are numerous genomes available in IMG to which the reads can be mapped to assist with assembly (currently 130 Bradyrhizobium spp, the vast majority of which are symbiotic with leguminous plants). The Carini lab will scale up cultures and isolate high quality DNA.

Although symbiotic Bradyrhizobia are extremely metabolically flexible, we know little about the metabolic flexibility and associated gene expression in free-living Bradyrhizobia. Therefore, we request transcriptomes from one representative isolate from each depth, at each site, grown under one of three distinct conditions: i) heterotrophically, ii) autotrophically with CO₂ + thiosulfate and iii) autotrophically with carbon monoxide as a sole source of C and energy. The total request, with replication, is 81 transcriptomes (3 depths \times 3 sites \times 1 isolate \times 3 conditions \times 3 replicates = 81 isolate transcriptomes). Assuming ~40 transcriptomes per Nextseq run (120 Gbp), we request 2 Nextseq runs for sufficient transcriptome coverage for downstream analyses. Thus our total transcriptome sequencing request is 240 Gbp. The Carini lab will perform the proposed experiments and isolate high quality RNA.

To contextualize the genome and transcriptome sequencing results to environmental dynamics, we request 81 stable isotope probing (SIP) enabled metatranscriptomes from the same depths at each of three locations that the bacterial isolates were obtained (3 depths \times 3 sites \times 3 labelled substrates \times 3 replicates = 81 SIP-enabled metatranscriptomes). The goal with these SIP experiments is to determine what the relative contribution of soil Bradyrhizobia is to heterotrophy, thiosulfate-fueled or carbon monoxide-fueled autotrophy, as a function of soil depth and across soil physicochemical gradients. The Carini lab will perform bottle incubations of soil with labelled substrates and deliver extracted RNA to JGI. We request that the JGI conducts the fractionation using their high throughput SIP fractionation capabilities, heavy-labeled RNA extraction, cDNA synthesis and cDNA sequencing. For SIP-assisted metatranscriptome sequencing, we request RNA sequencing via paired-end Illumina sequencing of ~3 Gbp per sample. This translates to ~40 samples per Nextseq run (2 Nextseq runs) and translates to approximately 240 Gbp.

Our total sequencing request for the entire project is 240 Gbp (SIP-enabled metatranscriptomes) + 240 Gbp (isolate transcriptomes) + 36 Gbp (isolate genomes) = 516 Gbp.

Justification:

Alphaproteobacteria are among the most abundant members of soil microbial communities where they play essential roles in plant-microbe dynamics in the rhizosphere and as free-living microbes in bulk soils. Of these alphaproteobacteria, Bradyrhizobia species are among the most extensively characterized genera, almost exclusively because of their symbiotic role in N-fixation with leguminous plants. Symbiotic Bradyrhizobia species enter into the roots of leguminous plants, where they induce the formation of root nodules that act as a focal point to fix atmospheric N₂ gas into ammonium (Poole, et al., 2018 Nat. Rev. Micro.). The bacteria exchange this fixed nitrogen with the plant in return for fixed carbon from the plant. The genetic details of nitrogen fixation and this symbiosis are well understood as the N-fixation process is economically and agriculturally important.

New evidence indicates the majority of soil Bradyrhizobia are free-living, non-symbiotic microbes with cosmopolitan distribution and diverse metabolisms. For example, a recent global survey of soil microbial communities showed that Bradyrhizobia were the most widely distributed and abundant taxa worldwide (Delgado-Baquerizo, et al., 2018 Science). Other studies show that Bradyrhizobium isolates from forest soils lacked *nif* genes - required for N-fixation - and *nod* genes - which are involved, but not required, for symbiosis. Metagenomes from these forest soils were consistent with the interpretation that symbiotic Bradyrhizobia are rare members of the forest soil community, as *nif* and *nod* genes were several orders of magnitude less abundant than other core Bradyrhizobium genes (VanInsberghe, et al 2015 ISME J). Similar findings were observed in Bradyrhizobia isolated from grasslands and fallow agricultural fields. Thus, there is a renewed interest in studying the physiology and genomics of soil Bradyrhizobia in the context of understanding their roles in soil biogeochemistry and carbon cycling.

We surveyed the microbial communities across three depth profiles in the CJCZO, outside of Tucson, Arizona. These depth profiles span distinct soil and climate types ranging from mid and high elevation conifer forests to a low elevation desert rangeland. We sampled these depth profiles every 10 cm from the surface up to 1 m. DNA was extracted and the V4-V5 region of the bacterial 16S rRNA gene was amplified and sequenced using standard protocols. The resulting sequences were processed into operational taxonomic units (OTU) based on 97% sequence identity and plotted as a function of depth. We observed a single Bradyrhizobial OTU was either the most abundant, or among the top 5 most abundant taxa at all depths, despite huge differences in the soil characteristics across sites and along each the depth profile.

The observation that Bradyrhizobia are present and abundant up to 1 m depth in our samples and across different soil types is consistent with several scenarios: 1) there is substantial genomic variation within this abundant OTU that can help explain spatial distributions that are not apparent at the OTU level; 2) environmental sensing and differential regulation of genes enables different modes of metabolism under different conditions; or 3) the environmental factors that enable the success of Bradyrhizobium, are consistent across depths and sites.

In this CSP, we will identify the genomic, transcriptional and physiological factors that determine the success of Bradyrhizobia in these soil depth profiles by using a combination of genome sequencing, transcriptomes, and stable isotope probing of soils from these sites to determine what substrates Bradyrhizobia are incorporating. This work will: 1) transform our

knowledge of the function of non-symbiotic Bradyrhizobia in varied ecosystems; 2) identify what substrates in situ Bradyrhizobia are capable of assimilating and 3) provide contextualization of in situ transcription patterns with cultured isolates.

Utilization:

We will analyze the genomes, culture transcriptomes and SIP-derived transcriptomes to identify the life strategies non-symbiotic Bradyrhizobia employ to be successful across physicochemical gradients in terrestrial ecosystems. The experimental design, which includes appropriate replication, will enable statistically robust comparisons of genome content, gene expression and SIP-enabled transcriptomics, across isolates and experimental conditions. The genomes sequenced as part of this CSP will add to JGI's genome portfolio and broaden the phylogenetic representation of for the Bradyrhizobial phyla, especially as it pertains to non-symbiotic members. We will analyze the genome sequences using mauve and other software packages to identify core and pan-genomic features that may have relevance in the context of the environmental distribution of each isolate. Of particular interest will be studying genomic islands that may harbor mobile gene suites that confer additional capabilities that might be useful across soil physicochemical gradients.

The isolate transcriptomes obtained in this study will be mapped to their respective genomes (using Bowtie2, or similar software) and differentially abundant genes across treatments will be detected using DEseq transcriptome analysis software. This analysis will identify "core responses" that all isolates from all soil samples share to common growth conditions. We will also identify responses that are not shared across isolates, but may be specific to isolates from a single site or isolated from a particular depth. Also of interest will be the differential expression of genes that are not shared across isolates, as these genes and their transcription may help explain the distribution of Bradyrhizobial ecotypes.

The metatranscriptomes from SIP experiments (from both the "heavy" and "light" fractions) will be mapped to the genomes of the isolates and a differential transcript abundance across the heavy and light fractions will be calculated only for those genes that map to the Bradyrhizobial genomes using DEseq. This analysis will tell us 1) what genes were expressed in in situ Bradyrhizobial populations that assimilated the labeled substrate; 2) the proportion of the Bradyrhizobial community that is assimilating the different labeled substrates; and 3) how the assimilation of different substrates by Bradyrhizobia changes with depth. These patterns will be compared to nutrient profiles from each of the soils to identify explanatory variables that can be directly measured.

Collectively, the products of this CSP will be used to produce 2-4 peer reviewed publications and generate ideas for future funding calls through the DOE or other agencies. One such publication will describe the core and pan genomes of non-symbiotic Bradyrhizobia and how they relate to previously sequenced genomes. A second manuscript will highlight the "core" Bradyrhizobial transcriptional response to different growth conditions that all isolates share and identify responses that differ by the site or depth from which the isolate was obtained. A third publication will be focused on the stable isotope tracer experiments that describes 1) which genes are active in situ; 2) what proportion of the bradyrhizobial population assimilates each substrate; and 3)

how (1) and (2) change with depth and across sites. Other publications that may arise from this work will depend on the particular insight gained from the genomes or transcriptomes of organisms. All products of this work will be deposited into public databases promptly as follows: Genomes will be deposited and made public in Joint Genome Institute's IMG-ER; transcriptomes will be deposited into the NCBI BioSample Database.

Community interest:

This work will be broadly used by microbiologists, biogeochemists, and microbial ecologists. Particularly, those researchers studying i) the microbiology and biogeochemistry of extreme environments; ii) microbe-microbe interactions; and iii) microbial ecology of surface and subsurface soils as it pertains to abundant microbial groups.

Life in subsurface soils is dominated by microbes that typically are restricted to the 'rare biosphere' (at low abundance) in surface soils. A striking exception to this generality are those taxa related to Bradyrhizobia. In our study and several other studies, Bradyrhizobia are found in high relative abundance throughout the shallow subsurface environments, despite sharp gradients of soil physicochemical parameters. For example, many of these subsurface soils are oligotrophic, acidic, have relatively high CO₂ concentrations, and have temperatures lower than surface soils. CSP-supported sequencing of the Bradyrhizobia isolates from both surface and deeper environments will allow researchers to compare results to more extensively studied surface systems and identify important genetic features that may explain the success of Bradyrhizobia in hostile environments.

Though the number of genome sequences available to researchers continues to grow, we still have a poor understanding of how genomic information relates to environmental distributions and how to predict functional roles based on this information. Thus, the Bradyrhizobial genomes, transcriptomes and SIP-enabled metatranscriptomes we will characterize will provide fruitful information to biogeochemists studying the sources and sinks of carbon in the subsurface. Although our isolates are derived from just a few sites, given that most subterranean environments are oligotrophic to a certain degree, understanding the impact of nutrient deprivation on carbon cycling by abundant members of the microbial community is likely important to a number of systems. Related to this, by obtaining culture transcriptomes alongside the SIP enabled metatranscriptomes, we will be able to directly relate changes in the transcription of genes to the dynamics of individual sub-populations in situ. Thus, researchers interested in understanding the integration of different types of meta-omic data will find our experiments, and the public data they will yield, beneficial.

Moreover, a current theme of DOE supported research is to use multi-omics to gain deeper insight into the function of microbial communities. One obstacle to gaining this understanding is a lack of appropriate model systems in which to study the relationships between multi-omic tools. Thus, researchers seeking to understand the linkages between 'omics technologies will benefit from the development of environmentally relevant and phylogenetically diverse model microbial systems.

DOE mission:

The microbial communities in subsurface terrestrial soils have a huge impact on the spatial and temporal dynamics of carbon sequestration and net ecosystem carbon fluxes. However, it is unclear how specific microbial taxa and their relative physiologies under different nutrient stresses affect these processes. Thus, the findings from this study are particularly relevant to DOE efforts to characterize carbon cycling dynamics in less well understood and extreme ecosystems. We have an extremely limited knowledge of how carbon dioxide fixed in surface environments propagates through the soil profile and what the roles of oligotrophic microbes play in subsurface carbon cycling processes. Moreover, we do not understand how metabolically flexible bacteria such as Bradyrhizobia contribute to carbon fixation in subsurface environments, relative to surface environments. Thus, the data collected from this study can directly be linked to the impact of microorganisms on carbon cycling and further developed to improve capabilities for carbon cycle predictions. This work will combine multiple lines of evidence derived from i) culture-based experiments; ii) the transcriptional dynamics under various stressors; and iii) SIP enabled metatranscriptomes under the same stressors to form a holistic picture of the potential roles of Bradyrhizobia in both surface and subsurface carbon cycling.

Terrestrial environments are ideal systems to study the roles of microbes in oligotrophic environments and the interactions between physical, chemical, and geochemical processes and biological processes. While these interactions are highly complex, controlled experimentation with important microbial taxa from these environments is an essential component of reducing that complexity to identify targets for future research. For example, by studying a dominant microbe in the laboratory, we previously identified a new component to vitamin biogeochemistry in the ocean (Carini, et al., 2014 ISME J). We know soil microbes are key players in the biogeochemical cycles in all habitats, but the details from subsurface soils are just beginning to emerge. This project will help improve the understanding of microbial mechanisms controlling carbon cycling in nutrient poor subsurface environments.

Sample preparation:

Dr. Paul Carini has more than 10 years of experience cultivating and conducting experiments with extremely fastidious oligotrophic microbes for transcriptome and metabolomics experiments (see attached CV). The Bradyrhizobia proposed in this experiment are readily cultivatable from a variety of environments. Several Bradyrhizobia isolates from these soils have been cultured already and are part of the Arizona Culture Collection, run by the Carini lab at the University of Arizona. The Carini Lab is housed within the University of Arizona's BIO5 Institute (<http://www.bio5.org/>) and is part of the new College of Agriculture's EcoSystem Genomics Cluster. This new Cluster also has an affiliation with the Arizona Genomics Institute (AGI, <http://genome.arizona.edu/>), an existing partner with JGI, that routinely isolates high yield, high quality, and high molecular weight nucleic acids for long-read DNA sequencing. The Carini lab is capable of performing high molecular weight DNA extractions and RNA extractions for each proposed experiment in a timely fashion.

Several of the strains we will sequence as part of this CSP are either actively growing or cryo-preserved and are expected to be readily revivable in the same chemically defined growth medium they were initially isolated and subcultured on. Other isolates will be isolated on YMA medium, typically used to propagate Bradyrhizobia. These isolates will be subcultured on to a

chemically defined mineral medium that will facilitate the experiments investigating the transcriptome dynamics under different growth conditions. Because Bradyrhizobia are fairly straightforward to grow in the laboratory, we do not foresee substantial bottlenecks in supplying genomic DNA. We estimate that genomic DNA should be available for sequencing within 6 months of the award date. Because of the experimental design of the isolate transcriptome experiments, we anticipate all RNA work to be completed 1 - 1.5 years after the award date.

The SIP metatranscriptome experiments may need extended incubations in order to obtain sufficient RNA label, especially in subsurface soils, as the microbial biomass is lower. The incubations that will be performed on each sample are ¹³C-labeled Mannitol (a proxy for heterotrophy), ¹³C-labeled CO₂ as a proxy for autotrophy and ¹³C-labeled CO as a proxy for CO-derived autotrophy. We will likely need to empirically determine the appropriate length of incubation before performing the actual experiments. Because of the uncertainty of incubation times, and the need for empirical testing, we estimate SIP metatranscriptome RNA will not be available until mid-late year 2 of the award.