Pangenomes reveal genomic signatures of microbial adaptation to chronic soil warming

Mallory J. Choudoir[1,2], Achala Narayanan[1,3], Damayanti Rodriguez-Ramos[1,4], Rachel Simoes[1], Alon Efroni[1], Abigail Sondrini[1] and Kristen M. DeAngelis[1*]

[1]Department of Microbiology, University of Massachusetts Amherst, Amherst, 01003, MA, USA.
[2]Present address: Department of Plant and Microbial Biology, North Carolina State University, Raleigh, 27606, NC, USA.
[3]Present address: Department of Plant and Microbial Biology, University of Minnesota, St. Paul, 55108, MN, USA.
[4]Present address: Department of Bacteriology, University of Wisconsin Madison, Madison, 53706, WI, USA.

*Corresponding author(s). E-mail(s): deangelis@microbio.umass.edu;
Contributing authors: mchoudo@ncsu.edu; naray262@umn.edu; drodriguezr2@wisc.edu; rsimoes@umass.edu; aefroni@umass.edu; asondrini@umass.edu;

Abstract

Evolutionary responses to anthropogenic climate change are irreversible and largely uncaptured by climate models. Below-ground carbon transformations represent an important natural mitigation solution, but novel adaptive traits may alter microbial climate feedback mechanisms. To better define microbial evolutionary responses to warming, we study microorganisms from an ongoing *in situ* soil warming experiment at the Harvard Forest Long-term Ecological Research (LTER) site where, for over three decades, soils are continuously heated 5 ^{o}C above ambient temperatures. We hypothesize that across generations of chronic warming, genomic signatures within diverse bacterial lineages reflect adaptations related to growth and carbon utilization. From our culture collection of soil bacteria isolated from experimental heated and control plots, we sequenced genomes representing taxa dominant

in soil communities and sensitive to warming, including independent lineages of Alphaproteobacteria, Actinobacteria, and Betaproteobacteria. We investigated differences in genomic attributes and patterns of functional gene content to identify genetic signatures of adaptation. Comparative pangenomics revealed differently abundant gene clusters with functional annotations related to carbon and nitrogen metabolism. We also observe differences in genome-wide codon usage bias between heated and control genomes, suggesting potential adaptive traits related to growth or growth efficiency. Together, these data illustrate the emergence of diverse lineage-specific adaptive traits as well as common ecological-evolutionary microbial responses to climate change.

Keywords: climate change, soil, adaptation, pangenome, codon usage

1 Introduction

Humanity faces an existential threat with accelerating soil loss due to habitat degradation and climate change [1], but the irreversible consequences of microbial adaptation to climate change remain insufficiently unexplored. Microbes are both impacted by climate change and are also themselves agents of global change [2–4]. Soil habitats are home to the most diverse microbial communities on the planet, with soil health predicated on microbial diversity and activity [5, 6]. Ultimately, how soil microbes respond and evolve to changing soil habitats will determine terrestrial ecosystem function under future climate change scenarios.

Rising soil temperatures due to climate change are altering microbial global nutrient cycles in cryptic ways [7, 8]. Worse, chronic warming may disrupt the microbial mechanisms governing carbon fluxes between the biosphere and atmosphere, mechanisms which may provide a natural attenuation of the self-reinforcing feedbacks to climate [9]. Soils are a considerable source of atmospheric carbon dioxide (CO_2) released through microbial decomposition of organic matter [10] and are also the largest terrestrial carbon reservoir. Microbes have pivotal roles in carbon fluxes between terrestrial and atmospheric systems, and incorporating microbial processes improves global carbon pool projections [11–13]. Temperate forests actively sequester 0.7 Pg C year-1, but the potential for forests to mitigate climate change is uncertain [14–16]. The fate of terrestrial carbon pools largely depends on how microbes respond to global change factors.

For over three decades, temperate forest soils at the Harvard Forest Long-term Ecological Research (LTER) site are continuously heated *in situ* 5 ^{o}C above ambient temperatures [17]. After 26 years, warming degraded the soil organic carbon quantity and quality, increased net CO_2 emissions, and altered microbial activity in heated versus control plots [18, 19]. Though fungal

biomass and community structure were drastically altered with these treatments [20, 21], bacterial biomass and community structure remained relatively unperturbed by warming, with significant changes observed only after 20 years in the organic horizon [22]. Finally, bacteria isolated from the heated Harvard Forest plots show greater potential to degrade lignin and other plant polymers compared to those isolated from control plots [23, 24], suggesting that adaptations related to substrate use may explain long-term dynamics of climate stress on microbial systems.

The metacommunity concept [25] describes how regional pools of microbial diversity, high dispersal rates, and a propensity for dormancy determine how microbial communities change, or not, in response to environmental disturbance [26, 27]. In general, ecological filtering is a strong force in shaping microbial community assembly [28]. Canonically we expect to observe similar ecological forces shaping microbial community rapid responses to soil warming, but due to large population sizes and short generation times, co-occurring ecological and evolutionary forces can act on more similar time scales for microbes across years or decades. Evolution introduces heritable adaptive traits into the soil ecosystem, and we don't know the cascading impacts of new adaptive traits for global carbon models. Theoretical models need to account for ecological and evolutionary feedbacks across spatiotemporal scales within these dynamic selective landscapes [29].

Our previous works indicate ecological filtering in heated plots in response to decreases in carbon quality and quantity [18] that likely selects for microbes able to degrade more recalcitrant carbon substrates [24]. Long-term warming causes thermal acclimation of soil respiration [30, 31], and though this effect may be seasonal [32], adaptive processes may also be important. Microbial thermal adaption likely involves physiological trade-offs altering growth rate and efficiency [33]. Life history trade-offs between growth efficiency, resource acquisition, and stress tolerance shape organismal fitness landscapes that link population and community-level trait evolution [34]. We hypothesize that Harvard Forest microbial lineages have evolved adaptive traits related to carbon substrate metabolism and growth efficiency through generations under chronic warming conditions. To maximize our ability to identify genomic signatures of adaption, we use comparative pangenome approaches on collections of closely related bacterial taxa across multiple independent lineages spanning Actinobacteria, Alphaproteobacteria, and Betaproteobacteria. We explore genomic attributes, patterns of functional gene content presence and absence, and codon usage between heated and control genomes to define adaptive traits which have emerged in this system since the onset of the long-term warming experiment.

2 Results and Discussion

2.1 Harvard Forest soil warming experiment bacterial genome collection

We conducted comparative pangenomics on bacteria from a long-term warming field experiment to test our hypothesis that microbes adapt to local environmental change stressors. All bacterial isolates originate from a multi-decade climate manipulation experiment [19] at the Harvard Forest Long-Term Ecological Research (LTER) site (Petersham, MA, USA). The Prospect Hill warming experiment was established in 1991 in a mixed deciduous forest stand. It operates through heated cables buried 10 cm below the surface which continuously maintain soil temperatures in heated plots at 5 ^{o}C above the ambient temperatures of control plots (6 replicate plots per treatment of 6 x 6 m) [17]. Between 2013-2020, we built a diverse culture collection of bacteria isolated from heated and control plots (see Methods and Table S1 for isolation details) targeting taxa dominant in soil communities and sensitive to long-term warming [22].

Isolates for whole genome sequencing were selected from our culture collection based on their full-length 16S ribosomal RNA gene phylogeny and taxonomy. Our goal was to build collections of closely related genomes with replicates from heated and control plots that represent multiple independent lineages of dominant soil taxa. The final collection contains 108 genomes belonging to 6 clades spanning three phyla, including Actinobacteria *Streptomyces* spp. (n=15), Actinobacteria *Kitasatospora* spp. (n=29), Alphaproteobacteria *Bradyrhizobium* spp. (n=15), Alphaproteobacteria *Rhizobia* spp. (n=14), and Betaproteobacteria *Paraburkholderia* spp. (n=17), and Betaproteobacteria *Ralstonia* spp. (n=18) (Table S1).

We generated high quality draft genomes according to community standards [35] using a combination of sequencing platforms (Table S2 and Table S3). A small subset of genomes (n=20) were sequenced previously at the DOE Joint Genome Institute (JGI) (Berkeley, CA, USA) with PacBio RS (Pacific Biosciences, Menlo Park, CA, USA). All other genomes were sequenced either in house (University of Massachusetts Amherst, Amherst MA, USA) with Oxford Nanopore Technologies (ONT, Oxford, UK) minION, at SeqCenter (Pittsburgh, PA, USA) with NextSeq 2000 (Illumina, San Diego, CA, USA), or at DOE JGI with NovaSeq S4 (Illumina, San Diego, CA, USA) or PacBio Sequel IIe (Pacific Biosciences, Menlo Park, CA, USA) (Table S1). Total contigs varied considerably between sequencing platforms, so we assessed genome quality by estimating completion based on the presence of bacterial single-copy genes. Nearly all genomes have high estimated completion, greater than 94% (Table S2). Since gene synteny is irrelevant for our downstream analyses, we concluded that these draft genomes are adequate for our study. All genome assemblies are archived on the DOE JGI's Integrated Microbial Genomes and Microbiomes (IMG/M) under study names "Sequencing of bacterial isolates from soil warming experiment in Harvard Forest, Massachusetts,

USA" and "Using genomics to understand microbial adaptation to soil warming" with GOLD Study IDs Gs0121248 and Gs0156716. IMG Taxon/Genome ID for each genome is referenced in Table S1. Sequence data is archived on the SRA under BioProjects PRJNA944970, PRJNA944972, PRJNA944974, PRJNA944975, PRJNA944977, and PRJNA944978 with BioSample accession numbers referenced in Table S1.

2.2 Phylogeny and genomic attributes

Phylogenetic reconstruction of concatenated single-copy ribosomal protein genes shows six monophyletic clades with varying levels of intra-group phylogenetic relatedness (Figure 1). Within each clade, genome-wide average nucleotide identity (ANI) is always greater than 76% (Table 1) indicating these groups represent genus-level comparisons [36]. Genomes within each clade vary from inter-species to intra-species comparisons (ANI $\geq$ 95%) [37], with some isolates sharing nearly identical genomes (ANI $>$ 99%). However, based on ANI the *Ralstonia* spp. clade is the only group that is exclusively the same species (Figure 1, Table 1). 16S rRNA gene phylogeny shows concordant relationships to the ribosomal protein tree (Figure S1), although the relationship between *Kitasatospora* spp. and *Streptomyces* spp. clades are less distinct at this locus. All intra-clade genomes share greater than 97% similarity at 16S rRNA gene sequences (with the exception of *Streptomyces* spp. with greater than 95.8% 16S rRNA gene similarity)(Figure S1, Table 1) suggesting intra-species comparisons. Our goal here is not to demarcate taxonomic boundaries, only to frame the phylogenetic and genomic relatedness of our genomes within the ongoing discussion of bacterial taxonomic diversity.

To identify adaptive genomic traits associated with long-term soil warming that are shared among the clades, we first determined if broad genome attributes varied by treatment. A number of traits did not vary with treatment groups in any of the clades, including genome size, total number of genes, genes per kb, average gene length, GC content, 16S rRNA gene copy number, total number of pangenome gene clusters, number of unique or singleton genes, and the number of carbohydrate active enzymes or CAZymes (Kruskal-Wallis rank sum test on the interaction between clade and treatment followed by post-hoc pairwise Dunn test across intra-clade treatment groups, P-value $>$ 0.1). The two exceptions were that 16S rRNA gene copy number varied between treatments for *Kitasatospora* spp. genomes (P-value = 0.033), and average gene length varied between treatments for *Streptomyces* spp. genomes (P-value = 0.067). On average, heated *Kitasatospora* spp. genomes had fewer 16S rRNA gene copies than control genomes, 6.2 versus 9.5, respectively. On average, heated *Streptomyces* spp. genomes had longer genes comapred to control genomes, 976 bp versus 909 bp, respectively.

We examined genomic signatures of bacteria exposed to three decades of climate warming to understand whether warmer soils select for certain taxa, or if evolution also leads to local adaptation. Based on an average generation time of two weeks for soil microbial communities estimated using 18O-labeled

water, 30 years corresponds to about 500 generations in this system [32]. In a reciprocal transplant study across a natural temperature gradient, ecological and evolutionary feedbacks were detected on a timescale of 1.5 years [38]. In this analysis, consistent mutations were detected in genes related to nutrient acquisition and stress responses. Our insights into microbial responses to climate change across longer time scales that impact ecosystem-level functions are narrow due to limited access to long-term warming experiments. Further, the few climate change studies that extend beyond a decade largely focus on ecological community responses using methods that would obscure detection of evolutionary responses. For example, a 2012 study investigating 18 years of warming in an Arctic soil system found that warming altered alpha diversity and microbial community composition, driven by increased dominance of Actinobacteria [39]. Across a 7 year climate manipulation experiment in a grassland ecosystem, a recent study found that warming decreased microbial richness, phylogenetic diversity, and overall biomass [40]. Authors also noted an increased in abundance of spore forming taxa, including Actinobacteria and Firmicutes, as well as functional genes related to sporulation. These results could indicate that climate warming affects regional seed banks of microbial diversity.

2.3 Pangenome structure, gene content, and functional enrichment between treatments

We used comparative pangenomics to determine if genomes from heated treatments harbored different patterns of functional gene content, and if these differences were consistent between clades. Pangenome structure and composition varied across phyla and between clades, ranging from more open to more closed pangenomes (Figure 2, Table 1). For instance, *Ralstonia* pangenome contains relatively few gene clusters (GC) (n=5,164 GCs) with nearly all of these belonging to the conserved core genome (n=3,735 GCs) and very few unique genes. Conversely, the other Betaproteobacteria clade *Paraburkholderia*, with similar number of genomes, shows considerably more diversity with approximately 5X more gene clusters (n=25,952) with only 10.4% (n=2,696) belonging to the core gene pool and many unique genes. This contrasting pattern is similar between the more open pangenome of Alphaproteobacteria *Bradyrhizobium* and more closed pangenome of Alphaproteobacteria *Rhizobium*. Actinobacteria clades show high levels of gene cluster diversity and more open structure, which is unsurprising as Actinobacteria are well known for their genomic and ecological diversity [41].

The number of individual gene clusters that are differently abundant between heated and control genomes varies across clades, with a trend for clades with smaller core pangenomes and larger accessory gene cluster pools to have greater associations of gene clusters with treatment groups. Lineages with more open pangenomes and larger accessory pools have a deeper reserve of genes that may be shared within a population during times of stress, which can lead to more rare genes to sweeping to fixation due to strong selection

and fitness effects [42]. Differences in gene flow dynamics and the presence of locally fit functional genes likely contributes to how microbes evolve to climate change.

We also investigated differences in function gene content between treatment groups. Each pangenome was analyzed independently to assess which genes were over- or underrepresented in genomes of organisms derived from chronically heated soils. Functional annotations relating to carbon and nitrogen metabolisms are significantly associated with heated and control genomes, and these annotations vary across clades (Table S4-S6). Genomes are annotated hierarchically, where gene cluster functional differences vary by functional orthologs or KOfams (x=452 KOfams, Table S4) or functional units of genes within in metabolic pathways or KEGG pathway modules (x=14 KEGG modules, Table S5). Finally, the estimated metabolic capabilities of strains also varied (x=29 metabolic pathways, Table S6). All analyses of functional annotations were based on unadjusted p-value cutoffs, which may generate some false positives. However, P-value correction would eliminate any possibility of inference, an issue that is commonly encountered with high-throughput genetic analyses.

Genomes from heated plots were relatively depleted for some pathways associated with carbohydrate metabolism including lipid and methane metabolism for *Kitasatospora*; co-factor and vitamin metabolism for *Rhizobium*; amino acid metabolism (polyamine biosynthesis) for *Bradyrhizobium*; and central carbohydrate metabolism for the *Paraburkholderia*. Genomes from heated plots were relatively enriched in nitrogen metabolism for *Kitasatospora*. Since it's unlikely that the presence or absence of entire metabolic pathways resulted in simple gene gain and loss events, pathway differences between heated and control groups reflect either longer timescales or ecological and evolutionary feedbacks mechanisms shaping genome evolution.

Pathways associated with competition were enriched in two Proteobacteria clades from heated plots. *Rhizobium* and *Paraburkholderia* genomes from heated plots were relatively enriched in KO pathways for biosynthesis of beta-lactams (Table S5) and in beta-lactam resistance for *Bradyrhizobium* (Table S6). This could be an indication of increased competition due to warming stress, or an indirect effect of the severe soil organic carbon depletion in the chronically heated soils. Over three decades, heated soils have lost 34% of soil organic matter [19], accompanied by relatively greater depletion of carbohydrates and other relatively more accessible forms of heterotrophic substrates [18].

2.4 Codon usage bias

Differences in global codon usage bias (CUB) correlate to microbial life history strategies, growth rate, and ecological adaptations [43–45], traits which appear to have been altered in soil microbiomes with long-term warming [22, 46]. CUB is an estimate of the degree to which an organism demonstrates non-random, or preferential, use of of synonymous codons in genes and organisms. Different

measures of codon usage bias offer different insights into codon usage across genes and genomes (see [47]. We measured the effective number of codons (ENC) [48], which calculates the frequencies of different codons in a gene sequences and ranges from 20 (extreme bias, only a single codon is used for each amino acid) to 61 (no bias, all synonymous codons are equally used). We also measured the independent of length and composition (MILC) [49], which calculates codon usage distance between a gene against the expected distribution based on background sequence composition or a set of reference genes, with higher values indicated stronger bias. Both CUB indices differed with long-term warming treatment.

Genome-wide (i.e. calculated for each gene) ENC distributions differ between heated and control genomes for all clades (Wilcox rank sum test, P-value < 0.0001) except *Ralstonia* spp. (P-value $= 0.7$) (Figure 3A). Genome-wide gene-level MILC distributions differ between heated and control genomes for all clades (P-value < 0.001) except *Paraburkholderia* spp. and *Ralstonia* spp. (Wilcox rank sum test, P-value > 0.4) (Figure 3B). We also determined CUB distributions between heated and control genes across core and accessory gene pools (Table S1). Core gene ENC distributions differ between heated and control genomes for *Streptomyces*, *Bradyrhizobium*, and *Rhizobium* (P-value < 0.0001) but not for *Kitasatospora*, *Paraburkholderia*, and *Ralstonia* (P-value > 0.3). Core gene MILC distributions differ between heated and control genomes for *Kitasatospora*, *Streptomyces*, and *Rhizobium* (P-value < 0.05) and are nearly significant for *Bradyrhizobium* (P-value $= 0.086$), but not for *Paraburkholderia* and *Ralstonia* (P-value > 0.1). Accessory gene ENC distributions differ between heated and control genomes for all clades (P-value < 0.0001) except *Paraburkholderia* spp. and *Ralstonia* spp. (P-value > 0.1). Accessory gene MILC distributions differ between heated and control genomes for all clades (P-value < 0.04) except *Ralstonia* spp. (P-value > 0.65). In summary, differences in genome-wide codon usage bias is, in some instances, driven by differences in codon usage across core and accessory genes, which may reflect differing selection pressures across gene pools or gene exchange dynamics.

While gene-level CUB metrics show the genome-wide distribution patterns, CUB summarized for each genome provides a single metric to compare between strains. When summarized per genome (i.e. calculated for each genome) (Table S7), ENC mean, median (Figure 3), and standard deviation did not vary between heated and control genomes for all clades. When summarized per genome, MILC mean and median (Figure S3) also did not vary between heated and control genomes for all clades. However, MILC standard deviation varied between *Kitasatospora* (Wilcox rank sum test, P-value $= 0.026$), *Streptomyces* (P-value $= 0.012$), and *Rhizobium* (P-value $= 0.029$) heated and control genomes. MILC standard deviation was greater in control Actinobacteria genomes but higher in heated *Rhizobium* genomes. When summarized per genome for core and accessory genes (Table S7), CUB mean, median (Figure 3), and standard deviation largely did not vary between heated and control

core gene pools for all clades (P-value > 0.1). Exceptions include differences in median MILC (P-value = 0.088) values between heated and control *Streptomyces* core genes; average ENC (P-value = 0.062), average MILC (P-value = 0.023), and median MILC (P-value = 0.016) values between heated and control *Kitasatospora* accessory genes; and standard deviation ENC (P-value = 0.06) and average MILC (P-value = 0.08) values between heated and control *Rhizobium* accessory genes. In summary, the relative differences in codon usage indices (both directional and magnitude) vary between gene pools and between phyla and clades (Figure 3, Figure S3, Table S7).

There is a significant negative relationship between ENC codon usage values and 16S rRNA gene copy number for both heated (Spearman's rho = -0.32, P-value < 0.0001) and control genomes (rho = -0.46, P-value < 0.0001), and the slope of these relationships differs P-value = 0.026) (Figure 4). There also is a weaker but still significant negative relationship between MILC codon usage values and 16S rRNA gene copy number for both heated (Spearman's rho = -0.24, P-value = 0.0011) and control genomes (rho = -0.41, P-value < 0.0001), and the slope of these relationships is nearly significantly different (P-value = 0.058) (Figure S4). In summary, this indicates that genomes with higher 16S rRNA gene copy numbers also have stronger codon usage bias, and this relationship is stronger in control genomes.

3 Conclusion

Soils in the heated plots at the Harvard Forest long-term warming experiment sustained chronic warming for decades. Warming caused a reduction in fungal but not bacterial biomass [20, 22], suggesting a seasonal shift towards a more bacterial-dominated ecosystem and local adaptation to enable this change. Pangenome analysis of 6 clades reflected organisms that have different nutrient acquisition strategies, based on functional gene content and codon usage bias. The genomes also suggest increased competition for these more limited resources due to chronic warming. Further physiological assays will help to connect changes in genome sequence to microbial function and ultimately inform parameters of ecosystem models of climate change.

4 Methods

4.1 Experimental site and bacterial culture collection

Bacteria were isolated from the long-term warming experiment plots between 2013–2020 using a range of enrichment conditions ([24] and Table S1) targeting taxa of interest. Briefly, soils were sampled (0–10 cm) using a steel soil corer, mineral and organic horizons were separated manually, and sieved (2 mm). Soil dilutions were plated onto solid enrichment media (Table S1) and incubated at room temperature for 2-8 weeks. Following the initial enrichment, colonies were streaked to isolation on 10% tryptic soy agar (TSA) and Reasoner's 2A (R2A), genotyped with full length 16S

rRNA gene sequences (27F-AGAGTTTGATCMTGGCTCAG and 1492R-TACGGYTACCTTGTTACGACTT [50]), and cryopreserved in 20% glycerol at -80 ºC.

4.2 Genomic DNA extraction and whole genome sequencing

We inoculated colonies into 5 mL 10% tryptic soy broth (TSB), grew them to stationary phase at 28 ºC with shaking, and pelleted them. High molecular weight gDNA was extracted from actinobacterial and alphaproteobacterial cell pellets using an adapted CTAB-lysozyme extraction protocol [51], with phenol-chloroform instead of chloroform and extended centrifugation times. Betaproteobacterial cell pellets were extracted using the DNeasy Blood & Tissue Kit (Qiagen, Germantown, MD, USA).

Average fragment size of 30-50 kb was confirmed with gel electrophoresis, and samples were quantified using a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) to verify at least 1 µg of gDNA. DNA quality was evaluated using a NanoDrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) with quality control requirements of absorbance ratio 260/280 within 1.8-2.0 and 260/230 within 2.0-2.3.

A subset of genomes were sequenced and assembled at the DOE Joint Genome Institute (JGI) (Berkeley, CA, USA) using either the PacBio RS (Pacific Biosciences, Menlo Park, CA, USA) or NovaSeq S4 (Illumina, San Diego, CA, USA) platforms (Table S1). All remaining genomes were sequenced in house using the MinION (Oxford Nanopore Technologies (ONT), Oxford, UK) platform. A subset of betaproteobacterial genomes were also sequenced at SeqCenter (Pittsburgh, PA, USA) using the NextSeq 2000 (Illumina, San Diego, CA, USA) platform.

MinION sequencing libraries were prepared using the Ligation Sequencing Kit SQK-LSK-109 (ONT, Oxford, UK). Between 6-8 isolates were multiplexed per sequencing run using the Native Barcoding Expansion Kit EXP-NBD104 (ONT, Oxford, UK). We skipped the Covaris g-TUBE shearing step to target long fragment DNA. Starting with 1 µg of DNA per isolate, samples were prepped with the NEBNext® FFPE DNA Repair Mix and NEBNext® Ultra™ II End Repair/dA-Tailing kits (New England Biolabs, Ipswich, MA, USA), and DNA cleanup was performed using Ampure XP beads (Beckman Coulter, Pasadena, CA, USA). 150 ng of each barcoded sample was pooled for a final library of 700-1000 ng, and adapters were ligated to the sample with Blunt/TA ligase (New England Biolabs, Ipswich, MA, USA). The long fragment buffer was used in an extended 10 minute incubation at 37 ºC to enrich for high molecular weight DNA. The flow cell was primed using the Flow Cell Priming Kit (ONT, Oxford, UK). Approximately 15 fmols of the pooled library, sequencing buffer, and loading beads were mixed and then loaded through Spot-On port of the flow cell. Sequencing runs completed after 72 hours, and high accuracy base calling was performed on the raw fast5 data with

the ONT software Guppy v4.5.4 or v6.1.1 (Oxford Nanopore Technologies, 2021).

4.3 minION genome assembly pipeline

For all genomes sequenced in house, we performed either *de novo* long-read assembly or hybrid short and long-read assembly. We used Filtlong (https://github.com/rrwick/Filtlong) to sub-sample high quality reads (–min_length 1000 –min_mean_q 85) to approximately 40X estimated genome coverage and Flye v2.8.1 [52] for *de novo* assembly. We used Racon v1.4.2 [53] to generate a consensus assembly and performed final polishing with medaka v1.2.1 (Oxford Nanopore Technologies, 2018). For the subset of betaproteobacterial genomes with short and long-read sequences, we used unicycler [54] to perform hybrid assembly. Quast [55] and CheckM [56] assessed the quality of our assemblies, and we generally considered a good assembly to have less than 5% contamination and more than 90% estimated completeness.

4.4 Genome annotation and pangenome construction

All comparative genomic analyses were either facilitated by or conducted with the software ecosystem anvi'o v7.1 [57]. Genome assemblies were imported into anvi'o with the program anvi-gen-contigs-database, and open reading frames were identified using Prodigal v2.6.3 [58]. The program anvi-run-kegg-kofams assigned KEGG Orthologs (KOs) functions to protein coding genes [59], and the program anvi-estimate-metabolism reconstructed the metabolic capabilities of each isolate genome [60]. Finally, genomes were annotated for carbohydrate active enzymes (CAZymes) and their associated substrates using the dbCAN3 meta server [61, 62]. Within each clade, the program anvi-compute-genome-similarity computed pairwise average nucleotide identity using the software program pyANI and ANIb method [63].

Pangenomes were constructed for each clade with the program anvi-pan-genome. Protein comparisons were computed with DIAMOND [64] on the sensitive mode and with minbit heuristic = 0.5 (default value) to eliminate weak matches. Finally, the MCL algorithm [65] with a MCL inflation value of 6, 8, or 10, depending on the phylogenetic relationships of the clade (Table S3), identified shared gene clusters (GCs). Opperationally, GCs can be thought of as a single gene but are actually collections of similar genes. Core gene GCs are defined as GCs present in all genomes, and within the core GCs are single-copy core genes (SCGs). Accessory GCs are those present in at least two but not all genomes. Finally, singletons are GCs unique to a single genome.

4.5 Phylogenetic analysis

Phylogenetic relationships were reconstructed from conserved housekeeping genes. 16S ribosomal RNA (rRNA) genes (best hit) and single-copy genes were identified with Hidden Markov Model (HMM) searches using the command anvi-run-hmms with anvi'o HMM sources Ribosomal_RNA_16S and

Bacteria_71. Nucloetide sequences for 16S rRNA genes and concatenated sequences for 39 ribosomal protein genes were extracted with the program anvi-get-sequences-for-hmm-hits and aligned with MUSCLE [66]. Maximum likelihood (ML) trees were built using the generalized time reversible nucleotide substitution model with gamma distributed rate heterogeneity among sites (GTRGAMMA) in RAxML v8.2.11, and bootstrap support was determined from 100 iterations using the RAxML rapid bootstrapping algorithm [67, 68]. Two Acidobacteria genomes, *Acidiphila* sp. EB88 and *Edaphobacter dinghuensis* EB95 (IMG Genome IDs 2767802315 and 2747843220), were included as the outgroup. We used iTol [69] to visualize trees.

4.6 Genome and pangenome trait analyses

The program anvi-compute-functional-enrichment-in-pan was used to identify functional annotations, metabolic pathway modules, and individual gene clusters enriched in heated and control genomes (see [70]). The R package coRdon [71] was used to calculate codon usage bias metrics. First, we removed sequences less than 80 codons. The we calculated two different codon usage metrics, the effective number of codons (ENC) [48] and Measure Independent of Length and Composition (MILC) [49]. All other statistical analyses were performed using R [72] statistical software including FSA [73] for Dunn's post-hoc test of multiple comparisons.

Supplementary information. If your article has accompanying supplementary file/s please state so here. Please refer to Journal-level guidance for any specific requirements.

Acknowledgments. This project was supported by grants from the National Science Foundation (No. DEB-1749206), the NSF Long-Term Research in Environmental Biology (DEB-1456528), and the NSF Long-Term Ecological Research (DEB-1832210) programs. This work was also conducted with support (proposal 507977: 10.46936/10.25585/60008103 to M.J.C. and K.M.D.) from the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. We are grateful to all the people who contributed to the isolation, genotyping, sequencing, and annotation of isolates in this study, including (but probably not limited to) Erin Bergeron, Andrew Billings, Isabella Bushko, Gina Chaput, Emily Clark, Spencer Moore, Samantha Murphy, Grace Pold, Bianca Surjawan, and Wing Yin Tam.

Author Contributions Statement. **Choudoir, MJ**: Conceptualization, Formal analysis, Writing - Original Draft, Visualization; **Narayanan, A**: Methodology, Resources; **Rodriquez-Ramos, D**: Methodology, Resources; **Simoes, R**: Resources; **Efroni, A**: Resources; *Sondrini, A*: Resources; **DeAngelis, KM**: Conceptualization, Writing - Review & Editing, Supervision, Funding acquisition

References

[1] Handelsman, J. & Cohen, K. *A world without soil: The past, present, and precarious future of the earth beneath our feet* (Yale University Press, 2021).

[2] Cavicchioli, R. *et al.* Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology* **17** (9), 569–586 (2019) .

[3] Gillings, M. R. & Paulsen, I. T. Microbiology of the anthropocene. *Anthropocene* **5**, 1–8 (2014) .

[4] Tiedje, J. M. *et al.* Microbes and climate change: a research prospectus for the future. *Mbio* e00800–22 (2022) .

[5] Wagg, C., Bender, S. F., Widmer, F. & Van Der Heijden, M. G. Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proceedings of the National Academy of Sciences* **111** (14), 5266–5270 (2014) .

[6] Delgado-Baquerizo, M. *et al.* Microbial richness and composition independently drive soil multifunctionality. *Functional Ecology* **31** (12), 2330–2343 (2017) .

[7] Butler, S. M. *et al.* Soil warming alters nitrogen cycling in a new england forest: implications for ecosystem function and structure. *Oecologia* **168** (3), 819–828 (2012) .

[8] Jansson, J. K. & Hofmockel, K. S. Soil microbiomes and climate change. *Nature Reviews Microbiology* **18** (1), 35–46 (2020) .

[9] Griscom, B. W. *et al.* Natural climate solutions. *Proceedings of the National Academy of Sciences* **114** (44), 11645–11650 (2017) .

[10] Schlesinger, W. H. & Andrews, J. A. Soil respiration and the global carbon cycle. *Biogeochemistry* **48** (1), 7–20 (2000) .

[11] Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nature climate change* **3** (10), 909–912 (2013) .

[12] Wieder, W., Grandy, A., Kallenbach, C. & Bonan, G. Integrating microbial physiology and physio-chemical principles in soils with the

microbial-mineral carbon stabilization (mimics) model. *Biogeosciences* **11** (14), 3899–3917 (2014) .

[13] Wieder, W., Grandy, A., Kallenbach, C., Taylor, P. & Bonan, G. Representing life in the earth system with soil microbial functional traits in the mimics model. *Geoscientific Model Development* **8** (6), 1789–1808 (2015) .

[14] Dixon, R. K. *et al.* Carbon pools and flux of global forest ecosystems. *Science* **263** (5144), 185–190 (1994) .

[15] Lal, R. Forest soils and carbon sequestration. *Forest ecology and management* **220** (1-3), 242–258 (2005) .

[16] Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *science* **333** (6045), 988–993 (2011) .

[17] Peterjohn, W. T. *et al.* Responses of trace gas fluxes and n availability to experimentally elevated soil temperatures. *Ecological applications* **4** (3), 617–625 (1994) .

[18] Pold, G., Grandy, A. S., Melillo, J. M. & DeAngelis, K. M. Changes in substrate availability drive carbon cycle response to chronic warming. *Soil Biology and Biochemistry* **110**, 68–78 (2017) .

[19] Melillo, J. M. *et al.* Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science* **358** (6359), 101–105 (2017) .

[20] Frey, S., Drijber, R., Smith, H. & Melillo, J. Microbial biomass, functional capacity, and community structure after 12 years of soil warming. *Soil Biology and Biochemistry* **40** (11), 2904–2907 (2008) .

[21] Pec, G. J. *et al.* Fungal community response to long-term soil warming with potential implications for soil carbon dynamics. *Ecosphere* **12** (5), e03460 (2021) .

[22] DeAngelis, K. M. *et al.* Long-term forest soil warming alters microbial communities in temperate forest soils. *Frontiers in microbiology* **6**, 104 (2015) .

[23] Pold, G., Melillo, J. M. & DeAngelis, K. M. Two decades of warming increases diversity of a potentially lignolytic bacterial community. *Frontiers in microbiology* **6**, 480 (2015) .

[24] Pold, G. *et al.* Long-term warming alters carbohydrate degradation potential in temperate forest soils. *Applied and environmental microbiology*

82 (22), 6518–6530 (2016) .

[25] Leibold, M. A. *et al.* The metacommunity concept: a framework for multi-scale community ecology. *Ecology letters* **7** (7), 601–613 (2004) .

[26] Shade, A. *et al.* Fundamentals of microbial community resistance and resilience. *Frontiers in microbiology* **3**, 417 (2012) .

[27] Allison, S. D. & Martiny, J. B. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences* **105** (supplement_1), 11512–11519 (2008) .

[28] Nemergut, D. R. *et al.* Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews* **77** (3), 342–356 (2013) .

[29] Abs, E., Chase, A. B. & Allison, S. D. Burning questions: How do soil microbes shape ecosystem biogeochemistry in the context of global change? *Environmental Microbiology* (2022) .

[30] Bradford, M. A. *et al.* Thermal adaptation of soil microbial respiration to elevated temperature. *Ecology letters* **11** (12), 1316–1327 (2008) .

[31] Bradford, M. A. *et al.* Cross-biome patterns in soil microbial respiration predictable from evolutionary theory on thermal adaptation. *Nature Ecology & Evolution* **3** (2), 223–231 (2019) .

[32] Domeignoz-Horta, L. A. *et al.* Substrate availability and not thermal-acclimation controls microbial temperature sensitivity response to long term warming. *Global Change Biology* (2022) .

[33] Bradford, M. A. Thermal adaptation of decomposer communities in warming soils. *Frontiers in microbiology* **4**, 333 (2013) .

[34] Malik, A. A. *et al.* Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *The ISME journal* **14** (1), 1–9 (2020) .

[35] Bowers, R. M. *et al.* Minimum information about a single amplified genome (misag) and a metagenome-assembled genome (mimag) of bacteria and archaea. *Nature biotechnology* **35** (8), 725–731 (2017) .

[36] Barco, R. *et al.* A genus definition for bacteria and archaea based on a standard genome relatedness index. *MBio* **11** (1), e02475–19 (2020) .

[37] Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences* **102** (7), 2567–2572 (2005) .

[38] Chase, A. B., Weihe, C. & Martiny, J. B. Adaptive differentiation and rapid evolution of a soil bacterium along a climate gradient. *Proceedings of the National Academy of Sciences* **118** (18), e2101254118 (2021) .

[39] Deslippe, J. R., Hartmann, M., Simard, S. W. & Mohn, W. W. Long-term warming alters the composition of arctic soil microbial communities. *FEMS microbiology ecology* **82** (2), 303–315 (2012) .

[40] Wu, L. *et al.* Reduction of microbial diversity in grassland soil is driven by long-term climate warming. *Nature Microbiology* **7** (7), 1054–1062 (2022) .

[41] Lewin, G. R. *et al.* Evolution and ecology of actinobacteria and their bioenergy applications. *Annual review of microbiology* **70**, 235–254 (2016) .

[42] Vernikos, G., Medini, D., Riley, D. R. & Tettelin, H. Ten years of pan-genome analyses. *Current opinion in microbiology* **23**, 148–154 (2015) .

[43] Botzman, M. & Margalit, H. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome biology* **12** (10), 1–11 (2011) .

[44] Roller, M., Lucić, V., Nagy, I., Perica, T. & Vlahoviček, K. Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic acids research* **41** (19), 8842–8852 (2013) .

[45] Chen, Y., Neilson, J. W., Kushwaha, P., Maier, R. M. & Barberán, A. Life-history strategies of soil microbial communities in an arid ecosystem. *The ISME Journal* **15** (3), 649–657 (2021) .

[46] Liu, X. J. A., Frey, S. D., Melillo, J. M. & DeAngelis, K. M. Physical protection regulates microbial thermal responses to chronic soil warming. *Soil Biology and Biochemistry* **159**, 108298 (2021) .

[47] Bahiri-Elitzur, S. & Tuller, T. Codon-based indices for modeling gene expression and transcript evolution. *Computational and Structural Biotechnology Journal* **19**, 2646–2663 (2021) .

[48] Wright, F. The 'effective number of codons' used in a gene. *Gene* **87** (1), 23–29 (1990) .

[49] Supek, F. & Vlahoviček, K. Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. *BMC bioinformatics* **6**, 1–15 (2005) .

[50] Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16s ribosomal dna amplification for phylogenetic study. *Journal of bacteriology* **173** (2), 697–703 (1991) .

[51] Ioerger, T. R. *et al.* Variation among genome sequences of h37rv strains of mycobacterium tuberculosis from multiple laboratories. *Journal of bacteriology* **192** (14), 3645–3653 (2010) .

[52] Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology* **37** (5), 540–546 (2019) .

[53] Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome research* **27** (5), 737–746 (2017) .

[54] Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS computational biology* **13** (6), e1005595 (2017) .

[55] Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. Quast: quality assessment tool for genome assemblies. *Bioinformatics* **29** (8), 1072–1075 (2013) .

[56] Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **25** (7), 1043–1055 (2015) .

[57] Eren, A. M. *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nature microbiology* **6** (1), 3–6 (2021) .

[58] Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11** (1), 1–11 (2010) .

[59] Aramaki, T. *et al.* Kofamkoala: Kegg ortholog assignment based on profile hmm and adaptive score threshold. *Bioinformatics* **36** (7), 2251–2252 (2020) .

[60] Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. Kegg for integration and interpretation of large-scale molecular data sets. *Nucleic acids research* **40** (D1), D109–D114 (2012) .

[61] Zhang, H. *et al.* dbcan2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **46** (W1), W95–W101 (2018) .

[62] Zheng, J. *et al.* dbcan-seq update: Cazyme gene clusters and substrates in microbiomes. *Nucleic Acids Research* **51** (D1), D557–D563 (2023) .

[63] Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G. & Toth, I. K. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* **8** (1), 12–24 (2016) .

[64] Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using diamond. *Nature methods* **12** (1), 59–60 (2015) .

[65] Dongen, S. v. & Abreu-Goodger, C. in *Using mcl to extract clusters from networks* 281–295 (Springer, 2012).

[66] Edgar, R. C. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32** (5), 1792–1797 (2004) .

[67] Stamatakis, A., Ludwig, T. & Meier, H. Raxml-iii: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21** (4), 456–463 (2005) .

[68] Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the raxml web servers. *Systematic biology* **57** (5), 758–771 (2008) .

[69] Letunic, I. & Bork, P. Interactive tree of life (itol) v4: recent updates and new developments. *Nucleic acids research* **47** (W1), W256–W259 (2019) .

[70] Shaiber, A. *et al.* Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome biology* **21** (1), 1–35 (2020) .

[71] Elek, A., Kuzman, M. & Vlahovicek, K. cordon: Codon usage analysis and prediction of gene expressivity. *R package version* **1** (3) (2019) .

[72] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria (2022). URL https://www.R-project.org/.

[73] Ogle, D. H., Doll, J. C., Wheeler, P. & Dinno, A. *FSA: Fisheries Stock Analysis* (2022). URL https://github.com/fishR-Core-Team/FSA. R package version 0.9.3.

Main Figures and Tables

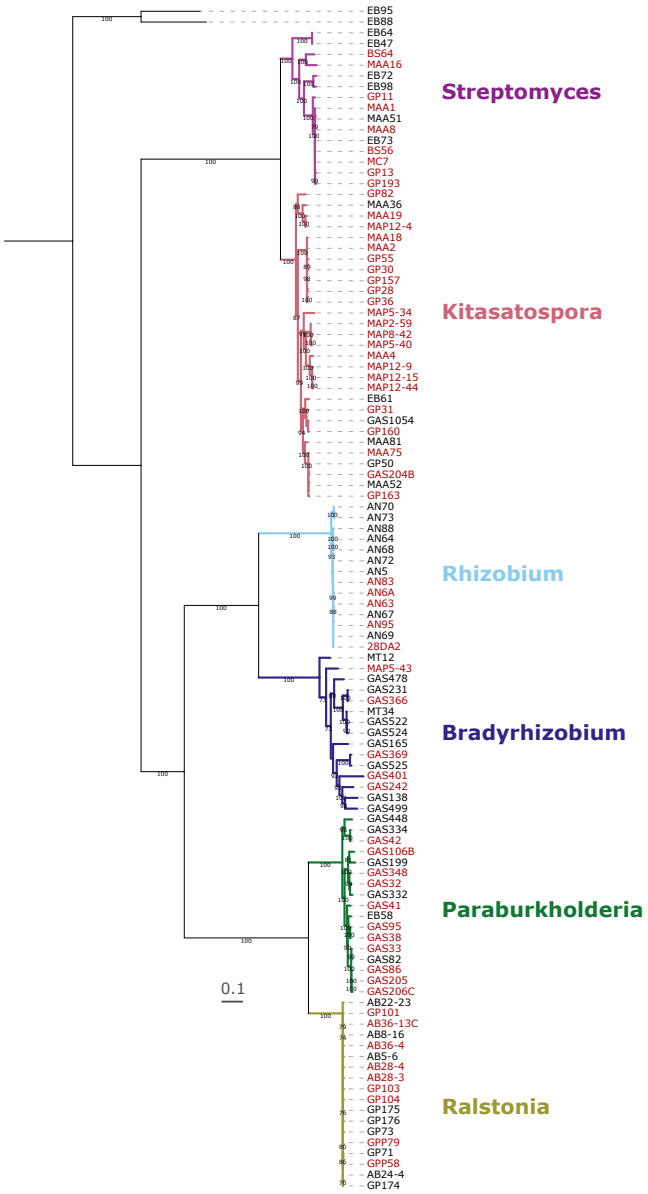
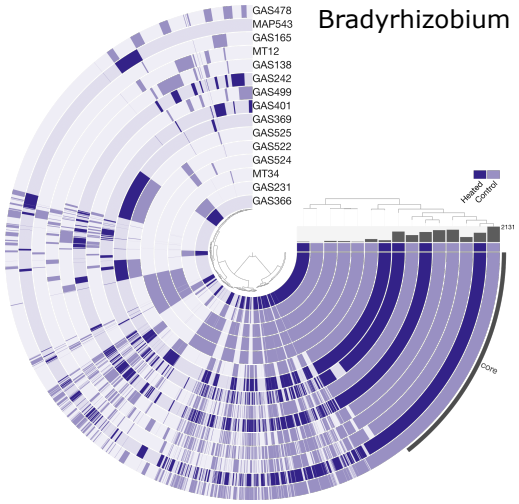
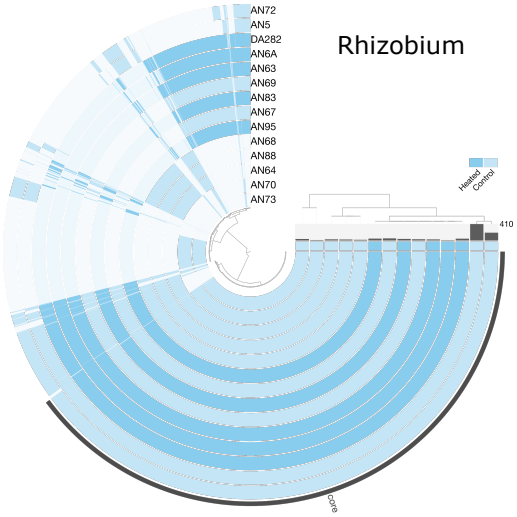


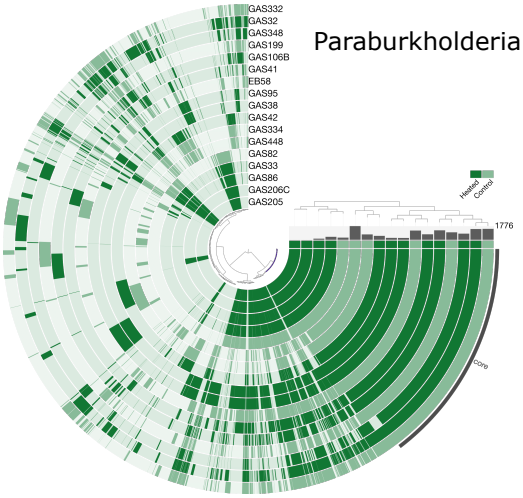
Figure 1. Ribosomal protein gene phylogeny. Tree shows the phylogenetic relationships between all isolates (Table S1) constructed from 39 concatenated ribosomal protein gene nucleotide sequences. Genome names are colored according to treatment group (heated in red and control in black). Phylogenetic clades are labeled and colored according to their pangenome group. Scale bar indicates nucleotide substitutions per site. Bootstrap values from 100 iterations are noted on all branches. Tree was rooted with two Acidobacteria genomes *Acidiphila* sp. EB88 and *Edaphobacter dinghuensis* EB95 as the outgroup.

Kitasatospora
MAP12-44
MAP1215
MAP12-9
MAA4
MAP5-34
GP82
MAP12-4
MAA19
MAA36
GP160
GAS1054
GP31
EB61
MAA81
MAP8-42
MAP5-40
MAP2-59
GP50
GAS204B
MAA52
GP163
MAA75
MAA2
MAA18
GP55
GP30
GP157
GP28
GP36
Heated
Control
1864
core

Streptomyces
MAA16
BS64
EB98
EB72
EB64
EB47
MAA1
BS56
EB73
MC7
GP193
GP13
MAA8
MAA51
GP11
core
Heated
Control
3662


Bradyrhizobium
GAS478
MAP543
GAS165
MT12
GAS138
GAS242
GAS499
GAS401
GAS369
GAS525
GAS522
GAS524
MT34
GAS231
GAS366
Heated
Control
2131
core


Rhizobium
AN72
AN5
DA282
AN6A
AN63
AN69
AN83
AN67
AN95
AN68
AN88
AN64
AN70
AN73
Heated
Control
410
core


Paraburkholderia
GAS332
GAS32
GAS348
GAS199
GAS106B
GAS41
EB58
GAS95
GAS38
GAS42
GAS334
GAS448
GAS82
GAS33
GAS86
GAS206C
GAS205
Heated
Control
1776
core

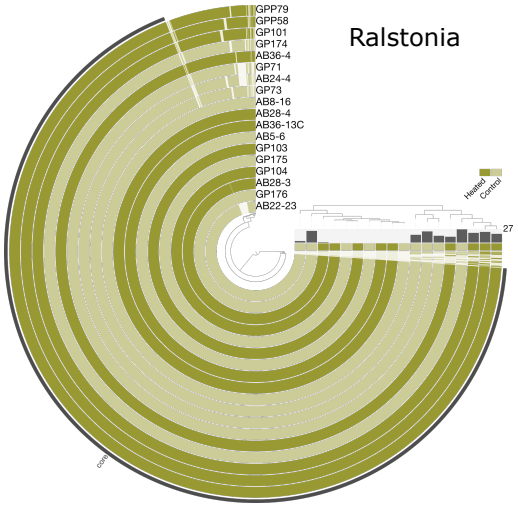

Ralstonia
GPP79
GPP58
GP101
GP174
AB36-4
GP71
AB24-4
GP73
AB8-16
AB28-4
AB36-13C
AB5-6
GP103
GP175
GP104
AB28-3
GP176
AB22-23
Heated
Control
27
core

Figure 2. Pangenome structure. Panels illustrate pangenome structure across six clades spanning three phyla. Concentric circles show gene cluster presence/absence for each isolate genome within the pangenome. Note, for visualization singleton genes were removed from before plotting, and bar graphs show the number of singleton genes for each genome. Internal dendrograms order the presence/absence of gene clusters by Euclidean distance. Pangenomes are colored by clade, and genomes are colored by treatment with heated genomes always being a darker shade. Finally, core gene clusters are highlighted for each pangenome.

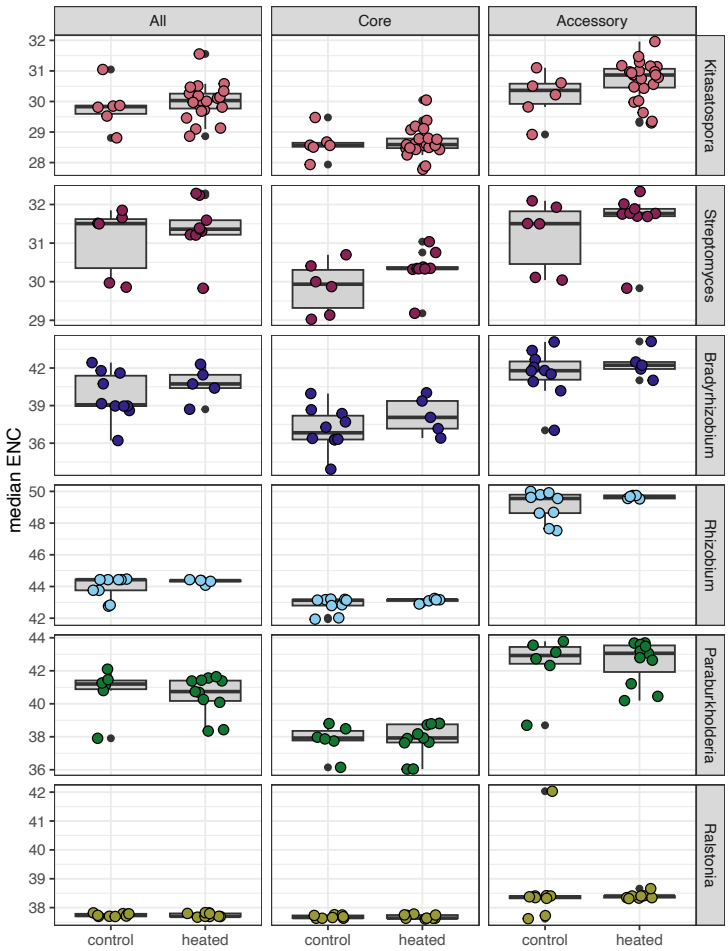


Figure 3. Strain-level median genome-wide codon usage. Plots illustrate the strain-level (i.e summarized per genome) median ENC codon usage across pangenome clades. Columns are median codon usage for all genes, core genes, or accessory genes. Core genes belong to gene clusters that are present in all genomes within a single pangenome group. Accessory genes belong to gene clusters that are shared by at least two, but not all genomes within a single pangenome group. Note, singeltons genes were not included in this analysis. Boxplots show the interquartile range of codon usage values for control and heated genomes from each clade. The line within each boxplot is the median, whiskers are 1.5 interquartile range, and black points are outliers. Circles show values of median genome-wide codon usage for each genome and colored by clade.

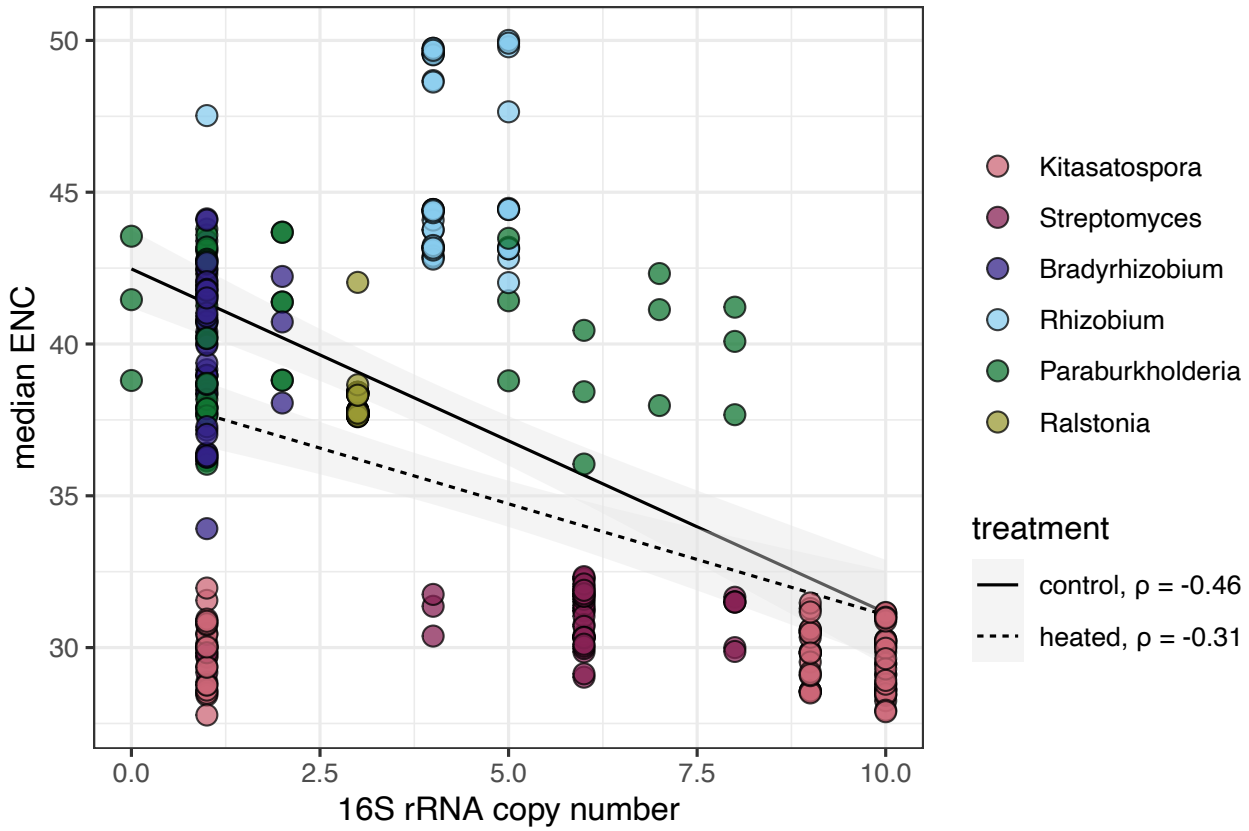


Figure 4. CUB versus 16S rRNA gene copy number. Plots illustrate the relationships between codon usage bias (y-axis) and 16S rRNA gene copy number (x-axis) (see Table S2) for each genome. Circles show values for median genome-wide (i.e. summarized across all genes within a genome) ENC codon usage for each genome and are colored according to pangenome group. Linear regression is plotted for heated (dashed line) and control (solid line) treatments. Shaded areas depict 95% confidence intervals. Spearman's rho for each treatment group is included in the legend.

Pangenome	Phylum	n	ANIb Identity	16S rRNA Similarity	MCL Inflation	Total GCs	Core GCs	SCGs	% Core	Accessory	% Accessory	Singletons	% Singletons
Kitasatospora	Actinobacteria	29	0.795–0.999	0.978–1.00	6	32,646	2,051	799	6.3%	19,050	58.4%	11,544	35.4%
Streptomyces	Actinobacteria	15	0.769–0.999	0.956–1.00	6	29,515	2,107	1,118	7.1%	12,817	43.4%	14,590	49.4%
Bradyrhizobium	Alphaproteobacteria	15	0.788–0.999	0.976–1.00	6	25,935	2,583	1,866	10.0%	10,858	41.9%	12,486	48.1%
Rhizobium	Alphaproteobacteria	14	0.903–0.999	0.995–1.00	10	7,984	3,714	2,558	46.5%	3,268	40.9%	1,001	12.5%
Paraburkholderia	Betaproteobacteria	17	0.799–0.999	0.977–1.00	8	25,952	2,696	1,966	10.4%	11,150	43.0%	12,106	46.6%
Ralstonia	Betaproteobacteria	18	0.999	0.995–1.00	10	5,164	4,517	3,735	87.5%	467	9.0%	180	3.5%

Table 1. Pangenome summary. Table describes pangenome characteristics for each clade. Columns list pangenome, phylum, number of genomes (n), average nucleotide identity (ANI) range, 16S rRNA gene similarity range, MCL inflation values used for generating gene clusters, total number of gene clusters (GCs), number of core GCs, number of single-copy core GCs, portion of GCs that belong to the core pool (% Core), number of accessory GCs, portion of GCs that belong to the accessory pool (% Accessory), number of singletons, and portion of GCs are singletons (% Singletons).