Pangenomes reveal genomic signatures of microbial adaptation to experimental 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

Below-ground carbon transformations represent a natural climate change mitigation solution, but newly-acquired traits adaptive to climate stress may alter microbial climate feedback mechanisms. To better define microbial evolutionary responses to long-term climate 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 ºC above ambient temperatures. We hypothesize that across generations of chronic warming, genomic signatures within diverse bacterial lineages reflect trait-based 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 observed differences in global codon usage bias between heated and control genomes, suggesting potential adaptive traits related to growth or growth efficiency. This effect was more varied for organisms with fewer 16S rrn operons, suggesting that these organisms experience different selective pressures on growth efficiency . Together, these data illustrate the emergence of lineage-specific traits as well as common ecological-evolutionary microbial responses to climate change.

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

1 Significance Statement

Anthropogenic climate change threatens soil health in part by altering microbial below-ground carbon cycling. Evolutionary responses are often overshadowed by community-level ecological responses, but adaptive responses represent irreversible changes in traits and functional potential. We predict that microbes are adapting to climate change stressors, like soil warming. To test this, we analyzed the genomes of bacteria from a soil warming experiment where soil plots have been experimentally heated 5 ºC above ambient for over 30 years. While genomic attributes were unchanged by long-term warming, warming altered functional gene content related to carbon usage and genetic indicators of growth efficiency. These responses may represent new parameters in how soil systems feed back to the climate system.

2 Introduction

Humanity faces an existential threat with accelerating soil loss due to habitat degradation and climate change Handelsman and Cohen (2021), 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 Cavicchioli et al (2019); Gillings and Paulsen (2014); Tiedje et al (2022). Soils are home to the most diverse microbial communities on the planet, and soil health is predicated on microbial diversity and activity Wagg et al (2014); Delgado-Baquerizo et al (2017). 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 complex ways Butler et al (2012); Jansson and Hofmockel (2020). 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 Griscom et al (2017). Soils release considerable amounts of atmospheric carbon dioxide (CO_2) through microbial decomposition of organic matter Schlesinger and Andrews (2000) but are also the largest terrestrial reservoir of global carbon Bahiri-Elitzur and Tuller (2021); Oelkers and Cole (2008). Microbes mediate key roles in carbon fluxes between terrestrial and atmospheric systems, evidenced by the fact that incorporating microbial processes into Earth system models improves global carbon pool projections Wieder et al (2013, 2014, 2015). Temperate forests actively sequester 0.7 Pg C year-1, but the potential for forests to mitigate climate change is uncertain Dixon et al (1994); Lal (2005); Pan et al (2011). The fate of terrestrial carbon pools largely depends on how microbes respond to global change factors.

Long-term field warming experiments allow us to make insights into what soils could look like in a warming world. 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 Peterjohn et al (1994). 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 Pold et al (2017); Melillo et al (2017). Though fungal biomass and community structure were drastically changed with warming Frey et al (2008); Pec et al (2021), bacterial biomass and community structure remained relatively unperturbed by warming, with significant changes observed only after 20 years in the organic horizon DeAngelis et al (2015). Finally, bacteria isolated from the heated Harvard Forest plots showed greater potential to degrade lignin and other plant polymers compared to those isolated from control plots Pold et al (2015, 2016), suggesting that adaptations related to substrate use may explain long-term dynamics of climate stress on microbial systems.

Both ecological and evolutionary mechanisms shape how environmental microbes respond to climate change. The metacommunity concept Leibold et al (2004) describes how regional pools of microbial diversity, high dispersal rates, and a propensity for dormancy all determine how microbial communities change, or not, in response to environmental disturbance Shade et al (2012); Allison and Martiny (2008). In general, ecological filtering strongly shapes microbial community assembly Nemergut et al (2013); Freedman and Zak (2015). 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 natural microbial populations across years or decades Koskella and Vos (2015). Evolution introduces heritable traits into microbial populations, and we don't know the cascading impacts

of local or regional adaptive traits on global soil ecosystem functions. Theoretical carbon models need to account for ecological and evolutionary feedbacks across spatiotemporal scales within these dynamic selective landscapes Abs et al (2022).

Based on our previous works, ecological filtering in heated plots likely selects for microbes able to degrade more recalcitrant carbon substrates Pold et al (2016) in response to warming-induced decreases in carbon quality and quantity Pold et al (2017). Long-term warming causes thermal acclimation of soil respiration Bradford et al (2008, 2019), and though this effect may be seasonal Domeignoz-Horta et al (2022), adaptive processes may also be important. Microbial thermal adaption likely involves physiological trade-offs altering growth rate and efficiency Bradford (2013). Life history trade-offs between growth efficiency, resource acquisition, and stress tolerance shape organismal fitness linking population and community-level trait evolution Malik et al (2020). We hypothesize that Harvard Forest microbial lineages have evolved adaptive traits related to carbon substrate metabolism and growth efficiency under generations of chronic warming conditions. To maximize our ability to identify genomic signatures of adaption, we used comparative pangenome approaches on collections of closely related bacterial taxa across multiple independent lineages spanning Actinobacteria, Alphaproteobacteria, and Betaproteobacteria. We explored 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.

3 Results and Discussion

3.1 Harvard Forest soil warming experiment bacterial genome collection

We conducted comparative pangenomic analysis on bacteria isolated from a long-term warming field experiment to test our hypothesis that microbes adapt to local environmental change stressors. All bacterial isolates originated from a multi-decade climate manipulation experiment Melillo et al (2017) 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. Heated cables buried 10 cm below the surface continuously maintain soil temperatures in heated plots at 5 ºC above the ambient temperatures of control plots (6 replicate plots per treatment of 6 x 6 m) Peterjohn et al (1994). 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 DeAngelis et al (2015).

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 representing multiple independent lineages of dominant soil taxa from heated and control plots. The final data set contained 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 *Rhizobium* spp. (n=14), and Betaproteobacteria *Paraburkholderia* spp. (n=17), and Betaproteobacteria *Ralstonia* spp. (n=18) (Figure 1, Table S1).

Actinobacteria, Alphaproteobacteria, and Betaproteobacteria are all globally dominant soil phyla Janssen (2006); Delgado-Baquerizo et al (2018) and represent broad functional guilds delineated by life history strategies Schimel and Schaeffer (2012); Fierer et al (2007). Though many models consider bacteria in general to be copiotrophic, and there exists a spectrum of growth efficiencies among phyla. Soil-dwelling Betaproteobacteria and Alphaproteobacteria are generally considered more copiotrophic with rapid growth and strong responses to labile carbon, while Actinobacteria typically embody more oligotrophic traits including slower growth and catabolic enzyme production Fierer et al (2007); Trivedi et al (2013). Actinobacterial genomes harbor abundant cellulases that breakdown plant-derived biomass in soils Lewin et al (2016); Berlemont and Martiny (2013). We previously observed enrichment of Actinobacteria and Alphaproteobacteria in Harvard Forest heated plots compared to controls DeAngelis et al (2015); Pold et al (2015). (We also included Betaproteobacteria here as a contrasting example of taxa less responsive to warming). This aligns with a temperature sensitively lab incubation study by Oliverio et al (2017) who found that across soil samples, nearly all Actinobacteria taxa and approximately half of all Alphaproteobacteria responded positively to warming.

The objective of this study was to identify genomic signatures that may link functional potential to trait-based adaptation. We generated high quality draft genomes according to community standards Bowers et al (2017) using a combination of sequencing platforms (Table S2 and Table S3). A 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 at the University of Massachusetts Amherst (Amherst MA, USA) (Table S3) with ONT MinION (Oxford Nanopore Technologies, Oxford, UK), 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 in assembled genomes varied considerably between sequencing platforms, so we assessed genome assembly quality by estimating completion based on the presence of bacterial single-copy genes. All genomes have high estimated completion, greater than 94% (Table S2-S3). Since gene synteny was less relevant for our downstream analyses, we concluded that these draft genomes were 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 assembly 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 also referenced in Table S1.

3.2 Phylogeny and genomic attributes

Phylogenetic reconstruction of concatenated single-copy ribosomal protein genes showed six monophyletic clades with varying levels of intra-group relatedness (Figure 1). Within each clade, genome wide average nucleotide identity (ANI) was always greater than 76% (Table 1) indicating these groups represent genus-level comparisons Barco et al (2020). Genomes within each clade varied from inter-species to intra-species comparisons (ANI $\geq$ 95%) Konstantinidis and Tiedje (2005), with some isolates sharing nearly identical genomes (ANI $>$ 99%). However, based on ANI alone the *Ralstonia* spp. clade is the only group that was exclusively the same species (Figure 1, Table 1). 16S rRNA gene phylogeny showed concordant relationships to the ribosomal protein tree (Figure S1), although the relationship between *Kitasatospora* spp. and *Streptomyces* spp. clades was less distinct at this locus. All intra-clade genomes shared 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 warming treatment. For the most part, these genome traits did not vary by treatment in any of the clades, including genome size, total number of genes, genes per kb, average gene length, G+C 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 (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) (Figure 2, Table S2). When controlling for phylogeny, these attributes also did not vary by treatment across all clades (phylogenetic ANOVA, P-value $>$ 0.1). There were two genomic traits that did vary by warming treatment within specific clades. The first was 16S rRNA gene copy number, which differed between treatments for *Kitasatospora* spp. genomes (P-value = 0.033) (Figure 2). SSU rRNA copy number reflects microbial ecological growth traits and ecological strategies Klappenbach et al (2000).

While most *Kitasatospora* spp. genomes harbored 9-10 16S rRNA gene copies, heated genomes are more likely to have a single 16S rRNA gene copy (Figure 2). The second was that average gene length varied between treatments for *Streptomyces* spp. genomes (P-value = 0.067) (Figure 2). On average, heated *Streptomyces* spp. genomes had longer genes compared to control genomes, 976 versus 909 bp, respectively.

3.3 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 compared to control treatments, and determined 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, the *Ralstonia* pangenome contained relatively few gene clusters (GCs) (n=5,164 GCs) with nearly all of these belonging to the conserved core genome (n=5,164 GCs) and very few strain-specific or singleton genes. Conversely, the other Betaproteobacteria clade *Paraburkholderia*, which has a similar number of genomes, showed considerably higher gene richness with approximately 5X more gene clusters (n=25,952), only 10.4% GCs (n=2,696) belonging to the core gene pool, and many singletons. We also observed a contrasting pattern between the more open pangenome of Alphaproteobacteria *Bradyrhizobium* and the more closed pangenome of Alphaproteobacteria *Rhizobium*. Both Actinobacteria clades showed high levels of gene cluster diversity, which is unsurprising as Actinobacteria are well known for their genomic and ecological diversity Lewin et al (2016).

We did not observe gene clusters strongly associating with heated or control treatments for any of the clades (Figure 2). In other words, a discrete heated or control core does not exist in our genome collection. The total number of individual gene clusters that were differently abundant between heated and control genomes varied across clades, but this pattern did not reflect more open or closed pangenome structure. Gene clusters in *Kitasatospora* and *Rhizobium* clades showed the strongest relationships with control (n=409 or 1.3% GCs with enrichment scores with an an unadjusted P-value < 0.01) and heated (n=474 or 5.9% GCs with enrichment scores with an an unadjusted P-value < 0.01) treatment plots, respectively. However, it's possible that this reflects uneven contributions of control and heated genomes in these clades.

Patterns of gene variation across populations evolve through various mechanisms that depend on both population parameters and ecological interactions McInerney et al (2017); Domingo-Sananes and McInerney (2021), and often these mechanisms are impossible to disentangle Baumdicker and Kupczok (2023). Pangenomes are constantly in flux and are themselves shaped by environmental dynamics and ecological-evolutionary feedbacks Cummins et al (2022). For example, lineages with large accessory gene pools have a deeper reserve of genes that may be shared within a population during times of stress

or environmental change, facilitating ecological adaptation Conrad et al (2022). However, a recent pangenome study found that environmental stressors including acidity, heat, drought, and salt stress all led to reductions in gene richness of *Bradyrhizobium diazoefficiens* Simonsen (2022), with accessory genes more prone to habitat-specific loss. These dynamics may reflect a transition between niche generalists possessing more open pangenomes and niche specialists more closed pangenomes Brockhurst et al (2019).

Because we didn't see a strong treatment-specific signal for individual gene clusters, we also investigated differences in functional gene content between treatment to discover annotations that were over- or underrepresented in genomes derived from chronically heated soils. Genomes were annotated hierarchically: gene cluster annotations varied by functional orthologs (n=452 KOfams, Table S4) or by functional units of genes within metabolic pathways (n=14 KEGG pathway modules, Table S5). Finally, the estimated metabolic capabilities of strains also varied between treatments (n=29 metabolic pathways, Table S6).

Genomes from heated plots were relatively depleted for some pathways associated with carbohydrate metabolism. This included 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 *Paraburkholderia* (Table S5). Conversely, genomes from heated plots were relatively enriched in nitrogen metabolism for *Kitasatospora* (Table S5). Genomes from heated plots were relatively depleted in trans-cinnimate degradation for *Kitasatopora* and in purine degradation, nitrate assimilation, and amino acid degradation for *Rhizobium* (Table S6). Genomes from heated plots were relatively enriched for modules involved in central carbon metabolism including the gamma-aminobutyrate (GABA) shunt, vitamin biosynthesis, and amino acid biosynthesis for *Kitasatapora*; GABA biosynthesis for *Streptomyces*; and methionine degradation for *Rhizobium* (Table S6). Since it's unlikely that the presence or absence of entire metabolic pathways resulted from single 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 Proteobacteria clades. *Rhizobium* and *Paraburkholderia* genomes from heated plots were relatively enriched in KEGG pathways for biosynthesis of beta-lactams (Table S5) and in beta-lactam resistance KEGG modules for *Bradyrhizobium* (Table S6). This could indicate 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 Melillo et al (2017), accompanied by relatively greater depletion of carbohydrates and other relatively more accessible forms of heterotrophic substrates Pold et al (2017).

3.4 Carbohydrate-active enzymes

Given that differences in available carbon is the major environmental determinate driving ecological differentiation between heated and control soils, next we identified genes functionally annotated as carbohydrate-active enzymes Drula et al (2022), or CAZymes. These enzymes catalyze the breakdown, degradation, or modification of carbohydrates and glycoconjugates, and CAZyme annotations can include both catalytic and carbohydrate-binding module enzyme domains.

The total number of CAZyme annotations did not differ across treatment for all clades (Table S2), suggesting that potential clade-specific differences associated with long-term warming would derive from CAZyme composition and not abundance. Clade membership was a strong predictor for both CAZyme family and subfamily composition (PERMANOVA, P-value=0.001, R2=0.82). Across the entire data set, the interaction between treatment and clade did not drive variation in CAZyme composition at the enzyme family classification (PERMANOVA, P-value=0.109, R2=0.012), but the interaction of treatment and clade explained variation in CAZyme composition at the subfamily enzyme classification (P-value=0.055, R2=0.013).

Within clades, treatment is a significant predictor of CAZyme family composition for *Rhizobium* (PERMANOVA, P-value=0.016, R2=0.21) but not for the other clades (Figure S2). However, treatment predicts CAZyme subfamily composition for *Kitasatospora* (P-value=0.05, R2=0.069), *Rhizobium* (P-value=0.02, R2=0.21), and *Ralstonia* (P-value=0.018, R2=0.16) (Figure 3). CAZyme subfamily annotations support enzyme substrate prediction Zheng et al (2023a), suggesting that strains from heated versus control plots differ in their genomic potential to utilize specific carbon substrates as opposed to differing in the presence or absence of broad enzyme classes.

We also observed that patterns of CAZyme families across pangenome GCs were differently abundant in genomes from heated and control plots (Table S7). For example, glycosyl-transferase (GT) enzymes were depleted in heated *Bradyrhizobium* and *Paraburkholderia* clades. Heated genomes of *Streptomyces*, and *Rhizobia* were relatively enriched in Auxiliary Activity enzymes involved in the breakdown of ligin (Table S7). Chronic soil warming depleted total soil organic matter and altered carbon composition, resulting in lower quantity as well as quality of carbon substrates in the heated versus control plots Pold et al (2017). Further, a greater portion of bacterial strains isolated from heated plots utilized complex carbon substrates (including xylan and cellulose) compared to strains isolated from control plots Pold et al (2016). Together, these data suggest that microbial substrate availability and carbon utilization shape microbial adaptation to soil warming.

3.5 Codon usage bias

Codon usage bias (CUB) is an estimate of the degree to which an organism demonstrates non-random, or preferential, use of synonymous codons in genes.

Codon bias varies with microbial physiology through mechanisms involving protein translation accuracy and/or efficiency Sharp et al (2005, 2010); Bennetzen and Hall (1982). Different measurements of codon usage bias offer various insights into codon usage across genes and gene expression levels across genomes Bahiri-Elitzur and Tuller (2021). We measured CUB as the effective number of codons (ENC) Wright (1990) and as the Measurement Independent of Length and Composition (MILC) Supek and Vlahoviček (2005). CUB varied between heated and control genomes for both metrics.

ENC 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 calculated CUB distributions between heated and control strains and across core and accessory genes, which may reflect differing selection pressures across gene pools or differences in horizontal gene exchange dynamics Moszer et al (1999). Genome wide gene-level ENC distributions differed between heated and control genomes for all clades (Wilcox rank sum test, P-value < 0.0001) except *Ralstonia* spp. (P-value $= 0.73$) (Figure S3A). Core gene ENC distributions differed between heated and control genomes for *Streptomyces*, *Bradyrhizobium*, and *Rhizobium* (P-value < 0.0001) but not for *Kitasatospora*, *Paraburkholderia*, and *Ralstonia* (P-value > 0.34) (Figure S4A). Accessory gene ENC distributions differed between heated and control genomes for all clades (P-value < 0.0001) except *Paraburkholderia* spp. and *Ralstonia* spp. (P-value > 0.14) (Figure S5A).

MILC calculates codon usage of a sequence as the distance against the expected distribution of a reference gene set. When calculated against a reference set of highly expressed genes, like ribosomal proteins, higher values indicate similar patterns of codon usage. Genome wide gene-level MILC distributions differed between heated and control genomes for all clades (Wilcox rank sum test, P-value < 0.001) except *Rhizobium* spp. (P-value $= 0.18$) (Figure S3B). Core gene MILC distributions differed between heated and control genomes for all clades (P-value < 0.022) besides *Bradyrhizobium* (P-value $= 0.48$) (Figure S4B). Accessory gene MILC distributions differed between heated and control genomes for all clades (P-value < 0.02) except *Bradyrhizobium* spp. and *Rhizobium* (P-value > 0.13) (Figure S5B).

While gene-level CUB metrics show genome wide distribution patterns, codon usage summarized for each genome provides a single metric to compare between treatment and gene pools. Strain-level median ENC and MILC codon usage did not vary between heated and control groups for all clades and for all gene pools (Wilcox rank sum test, P-value > 0.1), but median ENC and MILC did vary across gene pools for all clades (ANOVA, P-value < 0.002) (Figure 4). In general, core genes tend to have stronger codon bias patterns than accessory genes, and these differences contribute to global codon usage patterns.

Long-term warming was associated with less ENC codon bias (higher values) for Actinobacteria and Alphaprotebacteria clades, but Betaproteobacteria clades did not share this relationship (Figure 4). We observed a similar pattern in MILC codon bias with higher values, or codon usage more similar to highly

expressed ribosomal protein coding genes, in control versus heated genome for *Kitasatopora* and *Streptomyces* (Figure 4). Conversely, MILC values were greater in *Paraburkholderia* heated genomes compared to control genomes. Genomes with different MILC distributions likely differ in their protein expression profiles, which may relate to growth rate and efficiency. For example, this may indicate that genomes from long-term heated soils have lower global expression levels (i.e. related to maximum growth rate), or possess a smaller pool of genes with relatively less efficient or slower expression patterns. Subsequent work investigating tRNA variation between heated and control genomes will provide additional insights into how variation in amino acid usage may lead to fitness effects through enhanced transnational efficiency.

Genomes with higher 16S rRNA gene copy numbers have stronger codon usage bias (ENC), but this relationship is weakened in genomes of organisms exposed to long-term warming. We observed a significant negative relationship between median ENC codon bias and 16S rRNA gene copy number for both heated (Spearman's rho = -0.31, P-value < 0.0001) and control genomes (rho = -0.46, P-value < 0.0001), but the strength of this relationship differed between treatment (slope of linear model differs between groups, P-value = 0.026) (Figure 5). Conversely, we did not observe a predictive relationship between median MILC codon bias and 16S rRNA gene copy number for either treatment group. While SSU copy number does not vary between heated and control genomes (Figure 2), there may be genomic traits related to a relationship between codon usage and ribosomal operons that result in growth efficiency adaptations related resource availability Roller et al (2016) in warming soils.

Growing evidence suggests environmental drivers of microbial community-level codon usage. For instance, a recent cross-biome metagenomic study demonstrated direct impact of microbial habitat on codon bias and amino acid usage after controlling for other non-environmental variables Panda and Tuller (2023). Ultimately, a number of factors influence codon usage patterns in microbial genomes including genome wide GC content Chen et al (2004) as well as tRNA abundances Ikemura (1985); Dong et al (1996). Global CUB correlates with microbial life history strategies, growth rate, and ecological adaptations Botzman and Margalit (2011); Roller et al (2013); Chen et al (2021), traits which appear to emerge in soil microbiomes with long-term warming DeAngelis et al (2015); Liu et al (2021).

4 Conclusion

We examined genomes of bacteria exposed to three decades of climate warming to understand whether warmer soils select for common ecological traits, or if microbes also evolve and adapt to warmer soils. Warming caused a reduction in fungal but not bacterial biomass Frey et al (2008); DeAngelis et al (2015), suggesting a seasonal shift towards a more bacterial-dominated ecosystem and local adaptation to enable this change. We identified genomic signatures related to growth efficiency (Figures 4-5) and resource acquisition (Figure 3,

Tables S4-S7), which supports our previous observations that depleted soil organic matter in warming plots leads to more oligotrophic life history strategies DeAngelis et al (2015) and utilization of more complex carbon substrates Pold et al (2016).

Microbial populations tend to adapt to local environmental conditions Kraemer and Boynton (2017). However, evidence for ecological adaptation (or lack thereof) varies across environmental, spatial, and phylogenetic distances Connor et al (2010); Prosser et al (2007); Leducq et al (2014); Kraemer and Kassen (2015). The question here is, does local adaptation act along along the spatial and temporal scales of the Harvard Forest warming experiment? Based on an average generation time of two weeks for soil microbial communities estimated using 18O-labeled water Domeignoz-Horta et al (2022), 30 years corresponds to about 500 generations in our Harvard Forest system.

Under laboratory setting, microbes adapt to increasing temperatures rapidly Bennett et al (1992); Tenaillon et al (2012). After 1500 generations of selective temperature stress, the fungus *Neurospora* adapted to increased temperature through life history trait trade-offs, including increased resource allocation to spores but decreased growth rate and higher respiration Romero-Olivares et al (2015). 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. Even more rare are observations of local adaptation *in situ*. A small, but growing, body of research demonstrates ecological and evolutionary responses resulting in adaptation to environmental change on similar timescales Martiny et al (2023). For example, in a recent reciprocal transplant study across a natural temperature gradient, ecological and evolutionary feedbacks were detected on a timescale of 1.5 years Chase et al (2021). In this analysis, consistent mutations were detected in genes related to nutrient acquisition and stress responses.

The pangenomes of six clades (Figure 2) reflect organisms with different nutrient acquisition strategies, based on carbohydrate-activated enzymes, functional gene content, and codon usage bias. Our comparative pangenome data supports our previous works suggesting microbial trait trade-offs between growth efficiency and carbon substrate utilization Domeignoz-Horta et al (2020); Pold et al (2016). Functional genetic potential of genomes from heated plots suggests increased competition under a regime of more limited resources due to chronic warming. Ultimately, population-level adaptations to climate change may aggregate at the community-level resulting in changes in biogeochemical functions Wallenstein and Hall (2012). Future research on the physiology and growth traits of these organisms will further strengthen our ability to predict microbial function from genome adaptation, and to better predict ecosystem function in a changing world from microbial eco-evolutionary responses to climate change.

5 Methods

5.1 Experimental site and bacterial culture collection

We isolated bacteria from the Harvard Forest long-term soil warming experiment plots between 2013–2020 using a range of enrichment conditions (Pold et al (2016); Domeignoz-Horta et al (2020) and Table S1) targeting taxa of interest. Briefly, we sampled soils (0–10 cm) using a steel soil corer, manually separated mineral and organic horizons, and sieved (2 mm). We plated soil dilutions onto solid 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) or Reasoner's 2A (R2A), genotyped with full length 16S rRNA gene sequences (27F-AGAGTTTGATCMTGGCTCAG and 1492R-TACGGYTACCTTGTTACGACTT Weisburg et al (1991)), and cryopreserved in 20% glycerol at -80 ºC.

5.2 Genomic DNA extraction and whole genome sequencing

We inoculated colonies into 10 mL 10% tryptic soy broth (TSB), grew them to stationary phase at 28 ºC with shaking, and pelleted the cells by centrifugation. We extracted high molecular weight genomic DNA (gDNA) from cell pellets using either the DNeasy Blood & Tissue Kit (Qiagen, Germantown, MD, USA) or a modified CTAB-lysozyme extraction protocol Larsen et al (2007) (with phenol-chloroform instead of chloroform and extended centrifugation times).

We confirmed average fragment size of 30-50 kb with gel electrophoresis and quantified samples using a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) to verify at least 1 µg of gDNA. We evaluated DNA quality using a NanoDrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) with absorbance ratios 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 PacBio RS (Pacific Biosciences, Menlo Park, CA, USA) or NovaSeq S4 (Illumina, San Diego, CA, USA) (Table S1). A subset of betaproteobacterial genomes were also sequenced at SeqCenter (Pittsburgh, PA, USA) using NextSeq 2000 (Illumina, San Diego, CA, USA). All remaining genomes were sequenced in house using ONT MinION (Oxford Nanopore Technologies, Oxford, UK).

We prepared ONT MinION sequencing libraries using the Ligation Sequencing Kit SQK-LSK-109 (ONT, Oxford, UK). We multiplexed 6-8 isolates 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, we prepped samples with the NEBNext® FFPE DNA Repair Mix and NEBNext® Ultra™ II End Repair/dA-Tailing kits (New England Biolabs, Ipswich,

MA, USA) and performed DNA cleanup Ampure XP beads (Beckman Coulter, Pasadena, CA, USA). We pooled approximately 150 ng of each barcoded sample 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. We primed the flow cell using the Flow Cell Priming Kit (ONT, Oxford, UK), mixed, and loaded approximately 15 fmols of the pooled library, sequencing buffer, and loading beads. Sequencing runs completed after 72 hours, and we ran high accuracy base calling on the raw fast5 data with the ONT software Guppy v4.5.4 or v6.1.1 (Oxford Nanopore Technologies, 2021).

5.3 ONT MinION genome assembly

For all genomes sequenced using ONT MinION, 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 Kolmogorov et al (2019) for *de novo* assembly. We used Racon v1.4.2 Vaser et al (2017) 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 both short- and long-read sequence data, we used unicycler Wick et al (2017) to perform hybrid assembly. Quast Gurevich et al (2013) and CheckM Parks et al (2015) assessed the genome assembly quality, and we generally considered a genome assembly as high-quality if it had less than 5% contamination and more than 90% estimated completeness (Table S3) Bowers et al (2017).

5.4 Genome annotation

We conducted genomic analyses within the anvi'o v7.1 software ecosystem Eren et al (2021). 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 Hyatt et al (2010). The program anvi-run-kegg-kofams assigned KEGG Ortholog (KO) functions to protein coding genes Aramaki et al (2020), and the program anvi-estimate-metabolism reconstructed the metabolic capabilities of each isolate genome Kanehisa et al (2012). Finally, we predicted catalytic and carbohydrate-binding module domains of carbohydrate-active enzymes (CAZymes) using the dbCAN3 web server Zhang et al (2018); Zheng et al (2023a). Hidden Markov Model (HMM) searches against the dbCAN HMM database (HMMdb release 11.0) provided CAZyme family annotations, and HMM searches against the dbCAN-sub subfamily HMMdb predicted glycan substrates Zheng et al (2023b,a). We restricted downstream analyses to genes with hits (E-values $<$ 1e-15 and coverage $>$ 0.35) to both family and subfamily HMM databases. Finally, we imported CAZyme annotations into anvi'o contig databases with anvi-import-functions.

5.5 Pangenome construction

We constructed pangenomes for each clade with the program anvi-pan-genome. Protein comparisons were computed with DIAMOND Buchfink et al (2015) on the sensitive mode and with minbit heuristic = 0.5 (default value) to eliminate weak matches. Finally, the MCL algorithm Dongen and Abreu-Goodger (2012), with a MCL inflation value of 6, 8, or 10, depending on the phylogenetic relationships of the clade (Table 1), identified shared gene clusters (GCs) or collections of similar genes. Core GCs are defined as GCs present in all genomes. Accessory GCs are those present in at least one but not all genomes. Finally, singletons are GCs unique to a single genome. Within each clade, the program anvi-compute-genome-similarity computed pairwise average nucleotide identity using the software program pyANI and ANIb method Pritchard et al (2016).

5.6 Phylogenetic analysis

Phylogenetic relationships were reconstructed from conserved genes. 16S ribosomal RNA (rRNA) genes and single-copy housekeeping genes were identified with HMM searches using the command anvi-run-hmms with anvi'o databases 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 Edgar (2004). 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 Stamatakis et al (2005), and bootstrap support was determined from 100 iterations using the RAxML rapid bootstrapping algorithm Stamatakis et al (2008). Two Acidobacteria genomes, *Acidiphila* sp. EB88 and *Edaphobacter dinghuensis* EB95 (IMG Genome IDs 2767802315 and 2747843220), were included as the outgroup. We used iTol Letunic and Bork (2019) to visualize trees. Finally, we used the R package phytools to perform phylogenetic ANOVA with 1000 simulations Revell (2012).

5.7 Genome trait analyses

We used program anvi-compute-functional-enrichment-in-pan to identify functional annotations, metabolic pathway modules, and individual gene clusters differently enriched in heated and control treatment groups Shaiber et al (2020). All discussions of functional annotations were based on unadjusted P-value cutoffs, but we acknowledge that this may generate false positives. We used principal coordinate analysis (PCoA) of Bray–Curtis dissimilarities to visualize differences in CAZyme composition between treatment groups with the R package phyloseq McMurdie and Holmes (2013) and permutational ANOVA to asses statistical differences following 999 permutations with vegan Oksanen et al (2022).

We used the R package coRdon Elek et al (2019) to calculate codon bias. We filtered genes less than 80 codons before calculating codon usage metrics including the effective number of codons (ENC) Wright (1990) and the Measure Independent of Length and Composition (MILC) Supek and Vlahoviček (2005). For MILC values we calculated genome wide codon bias with respect to a subset of ribosomal proteins, which we speculated to be highly expressed genes. All other statistical analyses were performed using R R Core Team (2022) statistical software, including FSA Ogle et al (2022) for Dunn's post-hoc test of multiple comparisons.

6 Figure and Table Legends

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, with heated in red and control in black. Phylogenetic clades are also 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.

Figure 2. Pangenomes Panels illustrate pangenome composition and structure across six clades spanning three phyla. Concentric circles show gene cluster presence/absence for each isolate genome, and internal dendrograms order gene clusters by Euclidean distance. Note, for visualization singletons were removed from pangenomes before plotting. Genomes are ordered by ANI similarity. Finally, core gene clusters are labeled for each pangenome. Pangenomes are colored by clade, and genomes are colored by treatment as shown in the legend. Bar graphs show the relative comparisons of genome attributes between strains including genome size (Length), G+C content (GC), total number of genes (Genes), average gene length (Ave len), genes per kb, number of singleton gene clusters, total number of gene clusters, 16S rRNA gene copies (SSU copies), and total number of carbohydrate-activated enzymes (CAZymes). Actual values for all genome attributes are reported in Table S2.

Figure 3. Ordination plot of CAZyme sub-family composition across clades Plots show PCoA ordinations of CAZyme sub-family abundance and composition (Bray-curtis distances) for each clade. Each point represents the collection of carbohydrate-active enzymes classified at the sub-family level Zheng et al (2023a) for each genome. Heated genomes are colored according to legend. R2 and P-values report PERMANOVA results asking if treatment explains significant variation in CAZyme sub-family composition.

Figure 4. Strain-level median CUB values across gene pools Plots illustrate the strain-level (i.e. summarized per genome) median codon usage between treatment groups across clades and across gene pools. Columns are codon usage summary for all genes (All), core genes (Core), or accessory genes

(Accessory). Core genes belong to gene clusters that are present in all genomes within a pangenome clade. Accessory genes belong to gene clusters that are shared by at least two, but not all genomes within a pangenome clade. Note, singletons genes were not included in the accessory gene pool for this analysis. Boxplots show the interquartile range of codon usage values for control and heated genomes, and the line is the median, whiskers are 1.5 interquartile range, and black points are outliers. Circles show median codon usage values for each strain and are colored by clade. Figure 4A shows ENC Wright (1990) values, and Figure 4B shows MILC Supek and Vlahoviček (2005) values. We calculated MILC values against a set of conserved ribosomal protein genes.

Figure 5. CUB versus 16S rRNA gene copy number Plot illustrates the relationships between global codon usage bias (y-axis) and 16S rRNA gene copy number (x-axis) (Table S2). Circles show median values of strain-level ENC codon usage for each genome and are colored according to pangenome clade. Linear regression is plotted for heated (dashed line) and control (solid line) treatment. Shaded areas depict 95% confidence intervals. Spearman's rho for heated and control treatments are included in the legend.

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

Figure S1. 16S rRNA gene phylogeny Tree shows the phylogenetic relationships between all isolates (Table S1) constructed from 16S rRNA gene nucleotide sequences. Genome names are colored according to treatment group, with 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. Note, we failed to extract best-hit 16S rRNA gene sequences from *Paraburkholderia* genomes GAS33, GAS82, GAS205, and GAS206C, so these strains are omitted from the tree.

Figure S2. Ordination plot of CAZyme family composition across clades Plots show PCoA ordinations of CAZyme family abundance and composition (Bray-curtis distances) for each clade. Each point represents the collection of carbohydrate-active enzymes classified at the family level against the dbCAN HMM database (HMMdb release 11.0) Zhang et al (2018); Zheng et al (2023b) for each genome. Heated genomes are colored according to legend. R2 and P-values report PERMANOVA results asking if treatment explains significant variation in CAZyme family composition.

Figure S3. Gene-level genome-wide codon usage distributions Panels illustrate global gene-level codon usage distributions for all strains across clades. Each boxplot shows the interquartile range of codon usage values for all genes, and the line is the median, whiskers are 1.5 interquartile range, and black points are outliers. Boxplots are colored by clade and treatment. Plots to the right show the codon usage density distributions by treatment. Figure S2A shows ENC Wright (1990) distributions, and Figure S2B shows MILC Supek and Vlahoviček (2005) distributions. We calculated MILC values against a set of conserved ribosomal protein genes. Wilcox rank sum test P-values are reported for each clade and codon usage metric.

Figure S4. Gene-level core genome codon usage distributions Panels illustrate gene-level core genome codon usage distributions for all strains across all clades. Each boxplot shows the interquartile range of codon usage values for core genes, and the line is the median, whiskers are 1.5 interquartile range, and black points are outliers. Boxplots are colored by clade and treatment. Plots to the right show the codon usage density distributions by treatment. Figure S3A shows ENC Wright (1990) distributions, and Figure S3B shows MILC Supek and Vlahoviček (2005) distributions. We calculated MILC values against a set of conserved ribosomal protein genes. Wilcox rank sum test P-values are reported for each clade and codon usage metric.

Figure S5. Gene-level accessory genome codon usage distributions Panels illustrate the gene-level accessory genome codon usage distributions for all strains across all clades. Note, singletons genes were not included in the accessory gene pool for this analysis. Each boxplot shows the interquartile range of codon usage values for accessory genes, and the line is the median, whiskers are 1.5 interquartile range, and black points are outliers. Boxplots are colored by clade and treatment. Plots to the right show the codon usage density distributions by treatment. Figure S4A shows ENC Wright (1990) distributions, and Figure S4B shows MILC Supek and Vlahoviček (2005) distributions. We calculated MILC values against a set of conserved ribosomal protein genes. Wilcox rank sum test P-values are reported for each clade and codon usage metric.

Table S1. Table S1 describes the isolation metadata for bacterial isolates included in this study from *Kitasatospora* spp., *Streptomyces* spp., *Bradyrhizobium* spp., *Paraburkholderia* spp., and *Rhizobia* spp. clades. Columns report isolate id, phylum, organism name, warming experiment treatment, soil horizon, original isolation media, year isolated, genome sequencing platform, sequencing facility, NCBI BioSample accession number, and IMG taxon id (na values are unknown). Media abbreviations include: VL55 amended with xylan, plant ploymers (PP), or readily oxidized carbon (ROC); humic acid vitamin agar (HVA); lignin soybean flour vitamin agar (LSFA); modified carboxymethyl cellulose (CMC); International Streptomyces Project-2 Medium (ISP2); and 1% glucose/0.4% potato infusion (1% glu). See references within Pold et al (2016); Domeignoz-Horta et al (2020) for media formulations and

additional isolation conditions. Note the 20 genomes in the sequencing facility column that were previously sequenced, indicated with asterisks. All other genomes were generated for this study.

Table S2. Table S2 describes metadata for all draft genomes. Columns report genome name, phylum and clade, warming experiment treatment, number of contigs in the assembly, total assembly length, percent G+C, estimated genome completeness and redundancy (plus estimated confidence), total number of genes, average gene length, number of genes per kb, number of unique or singleton gene clusters, total number of gene clusters, 16S rRNA gene copy number, and the total number of carbohydrate-active enzymes (i.e. CAZymes). We used the program anvi-estimate-genome-completeness to estimate genome completion and redundancy based on presence of single-copy genes. Of note, the estimated completeness of *Streptomyces* GP11 is 63.4% but with a low confidence score of 0.6. However GP11's CheckM Parks et al (2015) genome completeness estimate is 99% (Table S3). We used the program anvi-run-hmms to determine 16S rRNA gene copy numbers with Hidden Markov Model (HMM) searches to anvi'o HMM sources Ribosomal_RNA_16S. We used the dbCAN3 metaserver to annotate CAZymes Zhang et al (2018); Zheng et al (2023b).

Table S3. Table S3 describes assembly metadata and quality control metrics for all draft genomes sequenced and assembled in house using ONT minION. Columns report isolate name, phylum, draft genome assembly IMG taxon id, *de novo* or hybrid assembly (see Methods), assembly coverage (X), taxonomic classification Parks et al (2015), assembly size, percent G+C, number of contigs, N50, and completeness and contamination as estimated by CheckM Parks et al (2015).

Table S4. Table S4 reports the functional enrichment metrics for KOfam annotated gene clusters. Columns are KOfam descriptions, enrichment score (see Methods and Shaiber et al (2020)), unadjusted P-values, Q-values adjusting for multiple tests, associated treatment group, KO identifier, corresponding gene clusters, the portion of heated (p_heated) and control (p_control) genomes that harbor each functional annotation, and clade. Note, only functional annotations that are significantly (unadjusted P-value < 0.05) enriched between treatments are included.

Table S5. Table S5 reports the functional enrichment metrics for KEGG annotated gene clusters. Columns are KEGG class descriptions, enrichment score (see Methods and Shaiber et al (2020)), unadjusted P-values, Q-values adjusting for multiple tests, associated treatment group, KEGG class identifier, corresponding gene clusters, the portion of heated (p_heated) and control (p_control) genomes that harbor each functional annotation, and clade. Note, only functional annotations that are significantly (unadjusted P-value < 0.05) enriched between treatments are included.

Table S6. Table S6 reports the functional enrichment metrics for KEGG metabolic modules. Columns are KEGG metabolic module description Kanehisa et al (2012), enrichment score (see Methods and Shaiber et al (2020)),

unadjusted P-values, Q-values adjusting for multiple tests, associated treatment group, KEGG metabolic module identifier, corresponding genomes, the portion of heated (p_heated) and control (p_control) genomes that harbor each functional annotation, and clade. Note, only functional annotations that are significantly (unadjusted P-value < 0.05) enriched between treatments are included.

Table S7. Table S7 reports the functional enrichment metrics for carbohydrate-active enzymes (CAZymes) annotated using the dcbCAN3 meta server Zhang et al (2018); Zheng et al (2023a). Columns are CAZyme class description, CAZYme family, enrichment score (see Methods and Shaiber et al (2020)), unadjusted P-values, Q-values adjusting for multiple tests, associated treatment group, enzyme EC number (if known), predicted substrate (if known), corresponding gene clusters, the portion of control (p_control) and heated (p_heated) genomes that harbor each functional annotation, and clade. Note, only CAZyme families that are significantly (unadjusted P-value < 0.05) enriched between treatments are included.

Acknowledgments. This project was conducted with support by a grant from the National Science Foundation (No. DEB-1749206). The soil warming experiments at Harvard Forest are maintained with support from the National Science Foundation Long Term Ecological Research Program (DEB-1832110) and a Long Term Research in Environmental Biology grant (DEB-1456610). We are grateful to Drs. Serita Frey and Mel Knorr for maintaining this infrastructure for our community. 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, 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

Abs E, Chase AB, Allison SD (2022) Burning questions: How do soil microbes shape ecosystem biogeochemistry in the context of global change? Environmental Microbiology

Allison SD, Martiny JB (2008) Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences 105(supplement_1):11,512–11,519

Aramaki T, Blanc-Mathieu R, Endo H, et al (2020) Kofamkoala: Kegg ortholog assignment based on profile hmm and adaptive score threshold. Bioinformatics 36(7):2251–2252

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

Barco R, Garrity G, Scott J, et al (2020) A genus definition for bacteria and archaea based on a standard genome relatedness index. MBio 11(1):e02,475–19

Baumdicker F, Kupczok A (2023) Tackling the pangenome dilemma requires the concerted analysis of multiple population genetic processes. Genome Biology and Evolution p evad067

Bennett AF, Lenski RE, Mittler JE (1992) Evolutionary adaptation to temperature. i. fitness responses of escherichia coli to changes in its thermal environment. Evolution 46(1):16–30

Bennetzen JL, Hall BD (1982) Codon selection in yeast. Journal of Biological Chemistry 257(6):3026–3031

Berlemont R, Martiny AC (2013) Phylogenetic distribution of potential cellulases in bacteria. Applied and environmental microbiology 79(5):1545–1554

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

Bowers RM, Kyrpides NC, Stepanauskas R, et al (2017) Minimum information about a single amplified genome (misag) and a metagenome-assembled genome (mimag) of bacteria and archaea. Nature biotechnology 35(8):725–731

Bradford MA (2013) Thermal adaptation of decomposer communities in warming soils. Frontiers in microbiology 4:333

Bradford MA, Davies CA, Frey SD, et al (2008) Thermal adaptation of soil microbial respiration to elevated temperature. Ecology letters 11(12):1316–1327

Bradford MA, McCulley RL, Crowther T, et al (2019) Cross-biome patterns in soil microbial respiration predictable from evolutionary theory on thermal adaptation. Nature Ecology & Evolution 3(2):223–231

Brockhurst MA, Harrison E, Hall JP, et al (2019) The ecology and evolution of pangenomes. Current Biology 29(20):R1094–R1103

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using diamond. Nature methods 12(1):59–60

Butler SM, Melillo JM, Johnson J, et al (2012) Soil warming alters nitrogen cycling in a new england forest: implications for ecosystem function and structure. Oecologia 168(3):819–828

Cavicchioli R, Ripple WJ, Timmis KN, et al (2019) Scientists' warning to humanity: microorganisms and climate change. Nature Reviews Microbiology 17(9):569–586

Chase AB, Weihe C, Martiny JB (2021) Adaptive differentiation and rapid evolution of a soil bacterium along a climate gradient. Proceedings of the National Academy of Sciences 118(18):e2101254,118

Chen SL, Lee W, Hottes AK, et al (2004) Codon usage between genomes is constrained by genome-wide mutational processes. Proceedings of the National Academy of Sciences 101(10):3480–3485

Chen Y, Neilson JW, Kushwaha P, et al (2021) Life-history strategies of soil microbial communities in an arid ecosystem. The ISME Journal 15(3):649–657

Connor N, Sikorski J, Rooney AP, et al (2010) Ecology of speciation in the genus bacillus. Applied and environmental microbiology 76(5):1349–1358

Conrad RE, Viver T, Gago JF, et al (2022) Toward quantifying the adaptive role of bacterial pangenomes during environmental perturbations. The ISME Journal 16(5):1222–1234

Cummins EA, Hall RJ, McInerney JO, et al (2022) Prokaryote pangenomes are dynamic entities. Current Opinion in Microbiology 66:73–78

DeAngelis KM, Pold G, Topçuoğlu BD, et al (2015) Long-term forest soil warming alters microbial communities in temperate forest soils. Frontiers in microbiology 6:104

Delgado-Baquerizo M, Trivedi P, Trivedi C, et al (2017) Microbial richness and composition independently drive soil multifunctionality. Functional Ecology 31(12):2330–2343

Delgado-Baquerizo M, Oliverio AM, Brewer TE, et al (2018) A global atlas of the dominant bacteria found in soil. Science 359(6373):320–325

Dixon RK, Solomon A, Brown S, et al (1994) Carbon pools and flux of global forest ecosystems. Science 263(5144):185–190

Domeignoz-Horta LA, Pold G, Liu XJA, et al (2020) Microbial diversity drives carbon use efficiency in a model soil. Nature communications 11(1):3684

Domeignoz-Horta LA, Pold G, Erb H, et al (2022) Substrate availability and not thermal-acclimation controls microbial temperature sensitivity response to long term warming. Global Change Biology

Domingo-Sananes MR, McInerney JO (2021) Mechanisms that shape microbial pangenomes. Trends in Microbiology 29(6):493–503

Dong H, Nilsson L, Kurland CG (1996) Co-variation of trna abundance and codon usage inescherichia coliat different growth rates. Journal of molecular biology 260(5):649–663

Dongen Sv, Abreu-Goodger C (2012) Using mcl to extract clusters from networks. In: Bacterial Molecular Networks. Springer, p 281–295

Drula E, Garron ML, Dogan S, et al (2022) The carbohydrate-active enzyme database: functions and literature. Nucleic acids research 50(D1):D571–D577

Edgar RC (2004) Muscle: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32(5):1792–1797

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

Eren AM, Kiefl E, Shaiber A, et al (2021) Community-led, integrated, reproducible multi-omics with anvi'o. Nature microbiology 6(1):3–6

Fierer N, Bradford MA, Jackson RB (2007) Toward an ecological classification of soil bacteria. Ecology 88(6):1354–1364

Freedman Z, Zak DR (2015) Soil bacterial communities are shaped by temporal and environmental filtering: evidence from a long-term chronosequence. Environmental microbiology 17(9):3208–3218

Frey S, Drijber R, Smith H, et al (2008) Microbial biomass, functional capacity, and community structure after 12 years of soil warming. Soil Biology and

Biochemistry 40(11):2904–2907

Gillings MR, Paulsen IT (2014) Microbiology of the anthropocene. Anthropocene 5:1–8

Griscom BW, Adams J, Ellis PW, et al (2017) Natural climate solutions. Proceedings of the National Academy of Sciences 114(44):11,645–11,650

Gurevich A, Saveliev V, Vyahhi N, et al (2013) Quast: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075

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

Hyatt D, Chen GL, LoCascio PF, et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics 11(1):1–11

Ikemura T (1985) Codon usage and trna content in unicellular and multicellular organisms. Molecular biology and evolution 2(1):13–34

Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16s rrna and 16s rrna genes. Applied and environmental microbiology 72(3):1719–1728

Jansson JK, Hofmockel KS (2020) Soil microbiomes and climate change. Nature Reviews Microbiology 18(1):35–46

Kanehisa M, Goto S, Sato Y, et al (2012) Kegg for integration and interpretation of large-scale molecular data sets. Nucleic acids research 40(D1):D109–D114

Klappenbach JA, Dunbar JM, Schmidt TM (2000) rrna operon copy number reflects ecological strategies of bacteria. Applied and environmental microbiology 66(4):1328–1333

Kolmogorov M, Yuan J, Lin Y, et al (2019) Assembly of long, error-prone reads using repeat graphs. Nature biotechnology 37(5):540–546

Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences 102(7):2567–2572

Koskella B, Vos M (2015) Adaptation in natural microbial populations. Annual Review of Ecology, Evolution, and Systematics 46:503–522

Kraemer SA, Boynton PJ (2017) Evidence for microbial local adaptation in nature. Molecular Ecology 26(7):1860–1876

Kraemer SA, Kassen R (2015) Patterns of local adaptation in space and time among soil bacteria. The American Naturalist 185(3):317–331

Lal R (2005) Forest soils and carbon sequestration. Forest ecology and management 220(1-3):242–258

Larsen MH, Biermann K, Tandberg S, et al (2007) Genetic manipulation of mycobacterium tuberculosis. Current protocols in microbiology 6(1):10A–2

Leducq JB, Charron G, Samani P, et al (2014) Local climatic adaptation in a widespread microorganism. Proceedings of the Royal Society B: Biological Sciences 281(1777):20132,472

Leibold MA, Holyoak M, Mouquet N, et al (2004) The metacommunity concept: a framework for multi-scale community ecology. Ecology letters 7(7):601–613

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

Lewin GR, Carlos C, Chevrette MG, et al (2016) Evolution and ecology of actinobacteria and their bioenergy applications. Annual review of microbiology 70:235–254

Liu XJA, Frey SD, Melillo JM, et al (2021) Physical protection regulates microbial thermal responses to chronic soil warming. Soil Biology and Biochemistry 159:108,298

Malik AA, Martiny JB, Brodie EL, et al (2020) Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. The ISME journal 14(1):1–9

Martiny JB, Martiny AC, Brodie E, et al (2023) Investigating the eco-evolutionary response of microbiomes to environmental change. Ecology Letters

McInerney JO, McNally A, O'connell MJ (2017) Why prokaryotes have pangenomes. Nature microbiology 2(4):1–5

McMurdie PJ, Holmes S (2013) phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data. PloS one 8(4):e61,217

Melillo JM, Frey SD, DeAngelis KM, et al (2017) Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. Science 358(6359):101–105

Moszer I, Rocha EP, Danchin A (1999) Codon usage and lateral gene transfer in bacillus subtilis. Current opinion in microbiology 2(5):524–528

Nemergut DR, Schmidt SK, Fukami T, et al (2013) Patterns and processes of microbial community assembly. Microbiology and Molecular Biology Reviews 77(3):342–356

Oelkers EH, Cole DR (2008) Carbon dioxide sequestration a solution to a global problem. Elements 4(5):305–310

Ogle DH, Doll JC, Wheeler P, et al (2022) FSA: Fisheries Stock Analysis. URL https://github.com/fishR-Core-Team/FSA, r package version 0.9.3

Oksanen J, Blanchet FG, Friendly M, et al (2022) vegan: Community ecology package. r package version 2.5-7. 2020

Oliverio AM, Bradford MA, Fierer N (2017) Identifying the microbial taxa that consistently respond to soil warming across time and space. Global Change Biology 23(5):2117–2129

Pan Y, Birdsey RA, Fang J, et al (2011) A large and persistent carbon sink in the world's forests. science 333(6045):988–993

Panda A, Tuller T (2023) Determinants of associations between codon and amino acid usage patterns of microbial communities and the environment inferred based on a cross-biome metagenomic analysis. npj Biofilms and Microbiomes 9(1):5

Parks DH, Imelfort M, Skennerton CT, et al (2015) Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome research 25(7):1043–1055

Pec GJ, van Diepen LT, Knorr M, et al (2021) Fungal community response to long-term soil warming with potential implications for soil carbon dynamics. Ecosphere 12(5):e03,460

Peterjohn WT, Melillo JM, Steudler PA, et al (1994) Responses of trace gas fluxes and n availability to experimentally elevated soil temperatures. Ecological applications 4(3):617–625

Pold G, Melillo JM, DeAngelis KM (2015) Two decades of warming increases diversity of a potentially lignolytic bacterial community. Frontiers in microbiology 6:480

Pold G, Billings AF, Blanchard JL, et al (2016) Long-term warming alters carbohydrate degradation potential in temperate forest soils. Applied and environmental microbiology 82(22):6518–6530

Pold G, Grandy AS, Melillo JM, et al (2017) Changes in substrate availability drive carbon cycle response to chronic warming. Soil Biology and Biochemistry 110:68–78

Pritchard L, Glover RH, Humphris S, et al (2016) Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods 8(1):12–24

Prosser JI, Bohannan BJ, Curtis TP, et al (2007) The role of ecological theory in microbial ecology. Nature Reviews Microbiology 5(5):384–392

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

Revell LJ (2012) phytools: An r package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223

Roller BR, Stoddard SF, Schmidt TM (2016) Exploiting rrna operon copy number to investigate bacterial reproductive strategies. Nature microbiology 1(11):1–7

Roller M, Lucić V, Nagy I, et al (2013) Environmental shaping of codon usage and functional adaptation across microbial communities. Nucleic acids research 41(19):8842–8852

Romero-Olivares AL, Taylor JW, Treseder KK (2015) Neurospora discreta as a model to assess adaptation of soil fungi to warming. BMC Evolutionary Biology 15(1):1–8

Schimel J, Schaeffer S (2012) Microbial control over carbon cycling in soil. front microbiol 3: 348

Schlesinger WH, Andrews JA (2000) Soil respiration and the global carbon cycle. Biogeochemistry 48(1):7–20

Shade A, Peter H, Allison SD, et al (2012) Fundamentals of microbial community resistance and resilience. Frontiers in microbiology 3:417

Shaiber A, Willis AD, Delmont TO, et al (2020) Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. Genome biology 21(1):1–35

Sharp PM, Bailes E, Grocock RJ, et al (2005) Variation in the strength of selected codon usage bias among bacteria. Nucleic acids research 33(4):1141–1153

Sharp PM, Emery LR, Zeng K (2010) Forces that influence the evolution of codon bias. Philosophical Transactions of the Royal Society B: Biological Sciences 365(1544):1203–1212

Simonsen AK (2022) Environmental stress leads to genome streamlining in a widely distributed species of soil bacteria. The ISME Journal 16(2):423–434

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

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

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

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, et al (2012) The molecular diversity of adaptive convergence. Science 335(6067):457–461

Tiedje JM, Bruns MA, Casadevall A, et al (2022) Microbes and climate change: a research prospectus for the future. Mbio pp e00,800–22

Trivedi P, Anderson IC, Singh BK (2013) Microbial modulators of soil carbon storage: integrating genomic and metabolic knowledge for global prediction. Trends in Microbiology 21(12):641–651

Vaser R, Sović I, Nagarajan N, et al (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome research 27(5):737–746

Wagg C, Bender SF, Widmer F, et al (2014) Soil biodiversity and soil community composition determine ecosystem multifunctionality. Proceedings of the National Academy of Sciences 111(14):5266–5270

Wallenstein MD, Hall EK (2012) A trait-based framework for predicting when and where microbial adaptation to climate change will affect ecosystem functioning. Biogeochemistry 109:35–47

Weisburg WG, Barns SM, Pelletier DA, et al (1991) 16s ribosomal dna amplification for phylogenetic study. Journal of bacteriology 173(2):697–703

Wick RR, Judd LM, Gorrie CL, et al (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS computational biology 13(6):e1005,595

Wieder W, Grandy A, Kallenbach C, et al (2014) Integrating microbial physiology and physio-chemical principles in soils with the microbial-mineral

carbon stabilization (mimics) model. Biogeosciences 11(14):3899–3917

Wieder W, Grandy A, Kallenbach C, et al (2015) Representing life in the earth system with soil microbial functional traits in the mimics model. Geoscientific Model Development 8(6):1789–1808

Wieder WR, Bonan GB, Allison SD (2013) Global soil carbon projections are improved by modelling microbial processes. Nature climate change 3(10):909–912

Wright F (1990) The 'effective number of codons' used in a gene. Gene 87(1):23–29

Zhang H, Yohe T, Huang L, et al (2018) dbcan2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic acids research 46(W1):W95–W101

Zheng J, Ge Q, Yan Y, et al (2023a) dbcan3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Research p gkad328

Zheng J, Hu B, Zhang X, et al (2023b) dbcan-seq update: Cazyme gene clusters and substrates in microbiomes. Nucleic Acids Research 51(D1):D557–D563

Main Figures and Tables

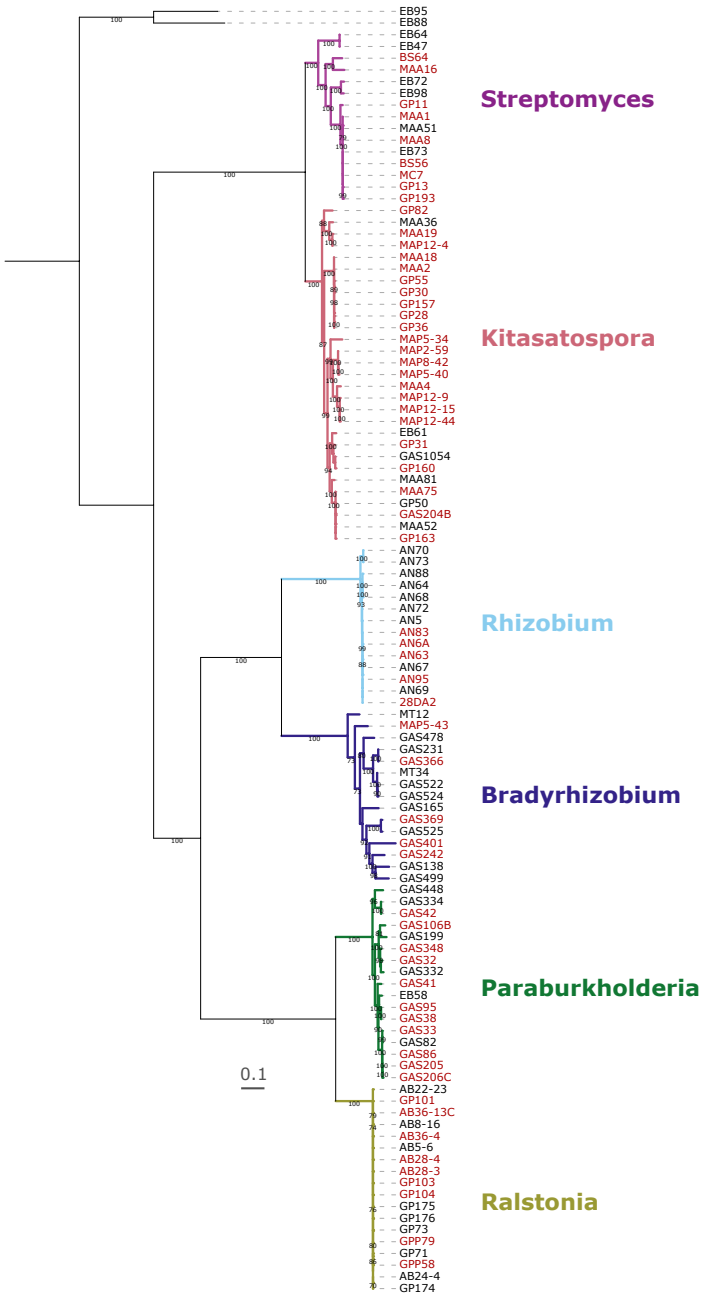

Figure 1. Ribosomal protein gene phylogeny. Tree shows the phylogenetic relationships between all isolates (Table S1) constructed from 39 concatenated ribosomal protein gene nucleotide sequences. Genome names are colored according to treatment group, with heated in red and control in black. Phylogenetic clades are also 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.

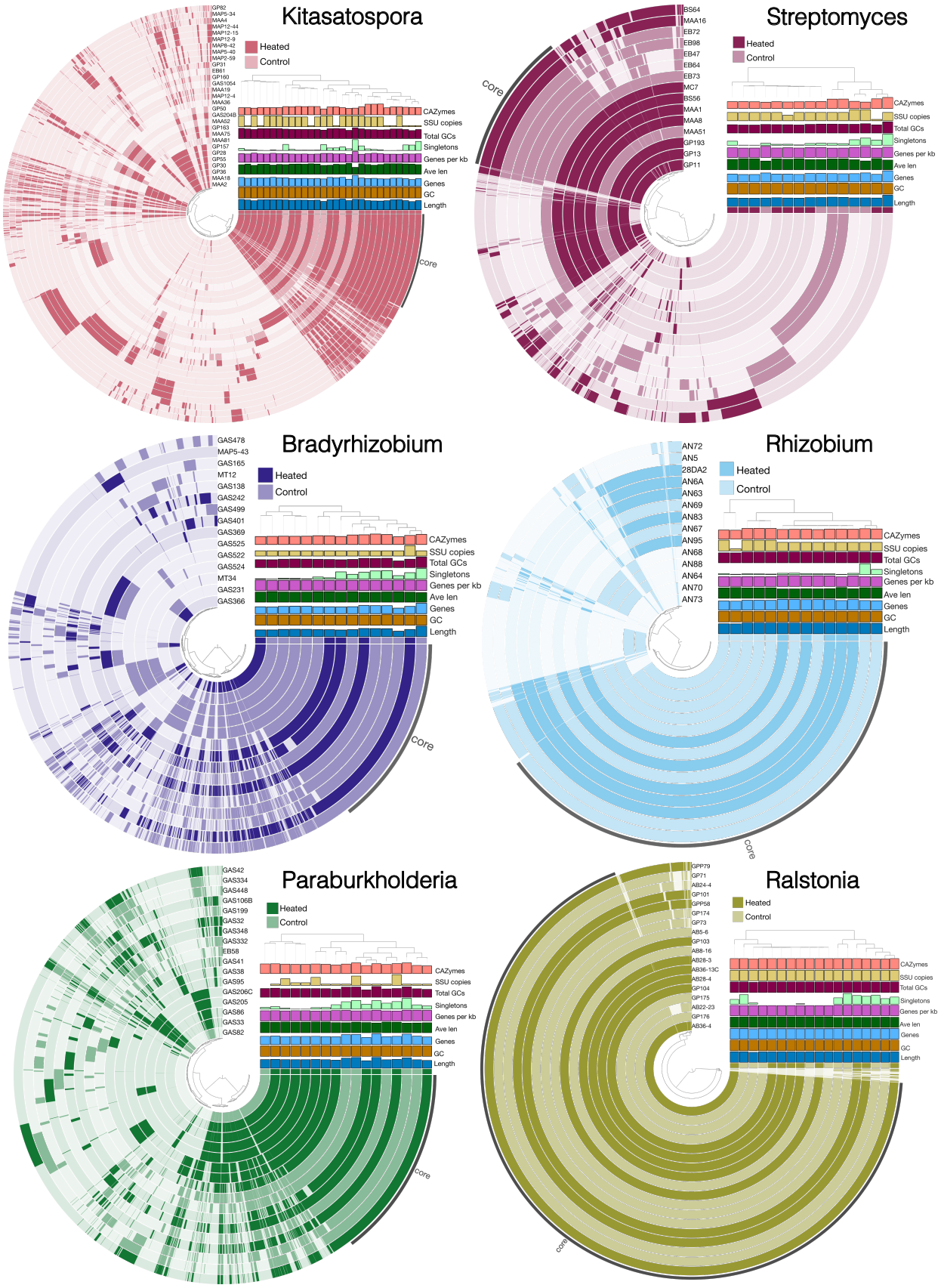
Kitasatospora
Heated
Control
GP82
MAP5-34
MAA4
MAP12-44
MAP12-15
MAP12-9
MAP8-42
MAP5-40
MAP2-59
GP31
EB81
GP160
GAS1054
MAA19
MAP12-4
MAA36
GP50
GAS204B
MAA52
GP163
MAA75
MAA81
GP157
GP28
GP55
GP30
GP36
MAA18
MAA2
CAZymes
SSU copies
Total GCs
Singletons
Genes per kb
Ave len
Genes
GC
Length
core
Streptomyces
Heated
Control
BS64
MAA16
EB72
EB98
EB47
EB64
EB73
MC7
BS56
MAA1
MAA8
MAA51
GP193
GP13
GP11
core
CAZymes
SSU copies
Total GCs
Singletons
Genes per kb
Ave len
Genes
GC
Length
Bradyrhizobium
Heated
Control
GAS478
MAP5-43
GAS165
MT12
GAS138
GAS242
GAS499
GAS401
GAS369
GAS525
GAS522
GAS524
MT34
GAS231
GAS366
CAZymes
SSU copies
Total GCs
Singletons
Genes per kb
Ave len
Genes
GC
Length
core
Rhizobium
Heated
Control
AN72
AN5
28DA2
AN6A
AN63
AN69
AN83
AN67
AN95
AN68
AN88
AN64
AN70
AN73
CAZymes
SSU copies
Total GCs
Singletons
Genes per kb
Ave len
Genes
GC
Length
core
Paraburkholderia
Heated
Control
GAS42
GAS334
GAS448
GAS106B
GAS199
GAS32
GAS348
GAS332
EB58
GAS41
GAS38
GAS95
GAS206C
GAS205
GAS86
GAS33
GAS82
CAZymes
SSU copies
Total GCs
Singletons
Genes per kb
Ave len
Genes
GC
Length
core
Ralstonia
Heated
Control
GPP79
GP71
AB24-4
GP101
GPP58
GP174
GP73
AB5-6
GP103
AB8-16
AB28-3
AB36-13C
AB28-4
GP104
GP175
AB22-23
GP176
AB36-4
CAZymes
SSU copies
Total GCs
Singletons
Genes per kb
Ave len
Genes
GC
Length
core

Figure 2. Pangenome structure. Panels illustrate pangenome composition and structure across six clades spanning three phyla. Concentric circles show gene cluster presence/absence for each isolate genome, and internal dendrograms order gene clusters by Euclidean distance. Note, for visualization singletons were removed from pangenomes before plotting. Genomes are ordered by ANI similarity. Finally, core gene clusters are labeled for each pangenome. Pangenomes are colored by clade, and genomes are colored by treatment as shown in the legend. Bar graphs show the relative comparisons of genome attributes between strains including genome size (Length), G+C content (GC), total number of genes (Genes), average gene length (Ave len), genes per kb, number of singleton gene clusters, total number of gene clusters, 16S rRNA gene copies (SSU copies), and total number of carbohydrate-activated enzymes (CAZymes). Actual values for all genome attributes are reported in Table S2.

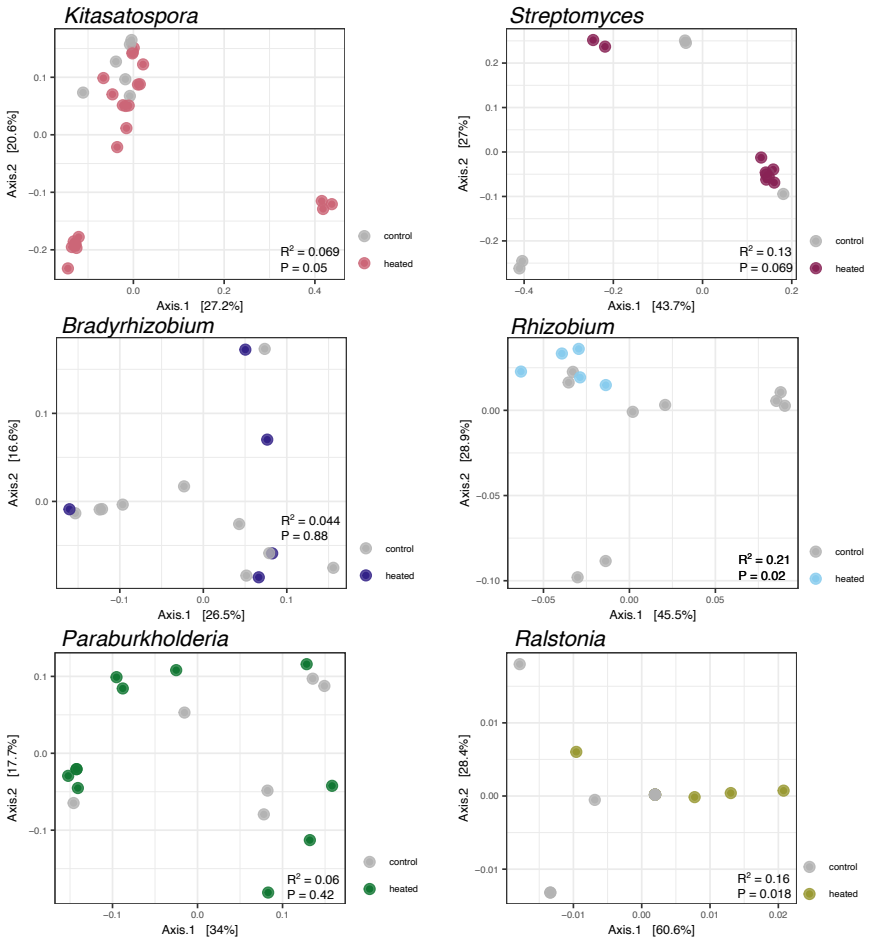

Figure 3. Ordination plot of CAZyme sub-family composition across clades. Plots show PCoA ordinations of CAZyme sub-family abundance and composition (Bray-curtis distances) for each clade. Each point represents the collection of carbohydrate-active enzymes classified at the sub-family level (Zheng *et al.* 2023). for each genome. Heated genomes are colored according to legend. R^2 and P-values report PERMANOVA results asking if treatment explains significant variation in CAZyme sub-family composition.

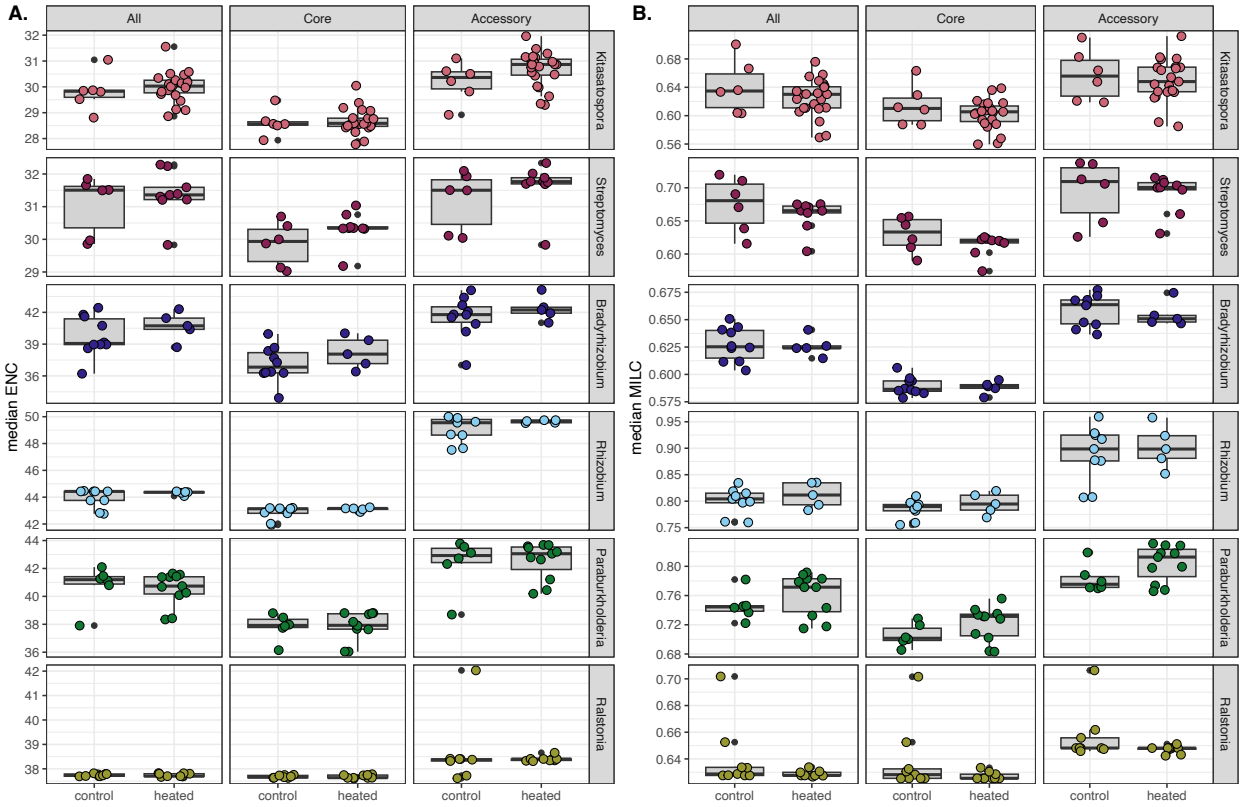

Figure 4. Strain-level median CUB values across gene pools. Plots illustrate the strain-level (i.e. summarized per genome) median codon usage between treatment groups across clades and across gene pools. Columns are codon usage summary for all genes (All), core genes (Core), or accessory genes (Accessory). Core genes belong to gene clusters that are present in all genomes within a pangenome clade. Accessory genes belong to gene clusters that are shared by at least two, but not all genomes within a pangenome clade. Note, singletons genes were not included in the accessory gene pool for this analysis. Boxplots show the interquartile range of codon usage values for control and heated genomes, and the line is the median, whiskers are 1.5 interquartile range, and black points are outliers. Circles show median codon usage values for each strain and are colored by clade. Figure 4A shows ENC (Wright 1990) values, and Figure 4B shows MILC (Supek *et al.* 2005) values. We calculated MILC values against a set of conserved ribosomal protein genes.

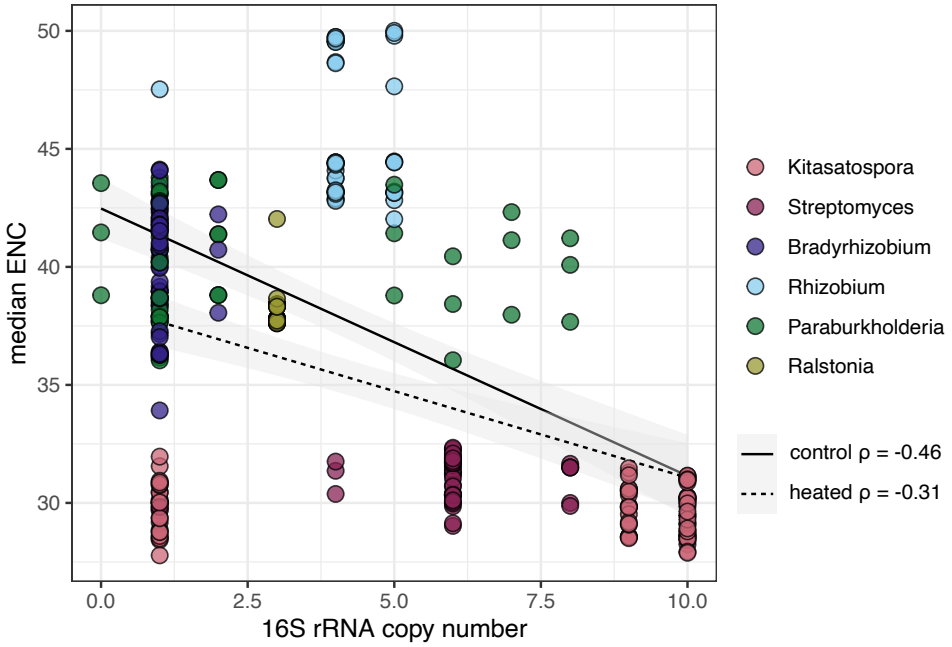

Figure 5. CUB versus 16S rRNA gene copy number. Plot illustrates the relationships between global codon usage bias (y-axis) and 16S rRNA gene copy number (x-axis) (Table S2). Circles show median values of strain-level ENC codon usage for each genome and are colored according to pangenome clade. Linear regression is plotted for heated (dashed line) and control (solid line) treatment. Shaded areas depict 95% confidence intervals. Spearman's rho for heated and control treatments are included in the legend.

Table 1. Pangenome Summary

Clade	Phylum	n Genomes	ANI	16S rRNA Similarity	MCL Inflation	Total GCs	Core GCs	p_Core	Accessory	p_Accessory	Singletons	p_Singletons
Kitasatospora	Actinobacteria	29	0.795–0.999	0.978–1.00	6	32,646	2,051	0.063	30,595	0.937	11,544	0.354
Streptomyces	Actinobacteria	15	0.769–0.999	0.956–1.00	6	29,515	2,107	0.071	27,408	0.929	14,590	0.494
Bradyrhizobium	Alphaproteobacteria	15	0.788–0.999	0.976–1.00	6	25,935	2,583	0.100	23,352	0.900	12,486	0.481
Rhizobium	Alphaproteobacteria	14	0.903–0.999	0.995–1.00	10	7,984	3,714	0.465	4,270	0.535	1,001	0.125
Paraburkholderia	Betaproteobacteria	17	0.799–0.999	0.977–1.00	8	25,952	2,696	0.104	23,256	0.896	12,106	0.466
Ralstonia	Betaproteobacteria	18	0.999	0.995–1.00	10	5,164	4,517	0.875	647	0.125	180	0.035

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