An ecological-evolutionary perspective on the genomic diversity and habitat preferences of the Acidobacteriota

Ella McReynolds, Mostafa S. Elshahed, and Noha H. Youssef*
Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA

*Corresponding author: Address: 1110 S. Innovation Way, Stillwater, OK. Email: noha@okstate.edu

Abstract

Members of the phylum Acidobacteriota inhabit a wide range of ecosystems including soils. We analyzed the global patterns of distribution and habitat preferences of various Acidobacteriota lineages across major ecosystems (soil, engineered, host-associated, marine, non-marine saline and alkaline, and terrestrial non-soil ecosystem) in 248,559 publicly available metagenomic datasets. Classes Terriglobia, Vicinamibacteria, Blastocatellia, and Thermoanaerobaculia were highly ubiquitous and showed clear preference to soil over non-soil habitats, class Polarisedimenticolia showed comparable ubiquity and preference between soil and non-soil habitats, while classes Aminicenantia and Holophagae showed preferences to non-soil habitats. However, while specific preferences were observed, most Acidobacteriota lineages were habitat generalists rather than specialists, with genomic and/or metagenomic fragments recovered from soil and non-soil habitats at various levels of taxonomic resolution. Comparative analysis of 1930 genomes strongly indicates that phylogenetic affiliation plays a more important role than the habitat from which the genome was recovered in shaping the genomic characteristics and metabolic capacities of the *Acidobacteriota*. The observed lack of strong habitat specialization and habitat transition driven lineage evolution in the Acidobacteriota suggest ready cross colonization between soil and non-soil habitats. We posit that such capacity is key to the successful establishment of Acidobacteriota as a major component in soil microbiomes post ecosystem disturbance events or during pedogenesis.

Introduction

The phylum Acidobacteriota [1] (previously, the Acidobacteria) represents one of the most prevalent phyla encountered in soils, as evident in 16S rRNA gene-based surveys, isolation efforts, and metagenomic studies [2-15]. Additionally, members of the phylum Acidobacteriota are also encountered in a wide range of habitats other than soil, e.g. hydrothermal vents [16], anoxic freshwater mud [17], hydrocarbon-contaminated aquifer [18], marine chiton microbiome [19], alkaline hot springs [20], termite nests [21], and marine sediments [22]. Several classification schemes have been proposed for the Acidobacteriota [2, 23, 24]. The Genome Taxonomy Database (GTDB, release 214) [25] classifies the phylum Acidobacteriota into 15 classes, 54 orders, and 102 families, many of which do not have a pure culture representative.

Based on isolation and culture-independent efforts, preference of specific Acidobacteriota lineages to a certain habitat has been observed. For example Acidobacteriota groups 1, 3, 4, and 6 (Barns [2] classification scheme) are the most prevalent members of the Acidobacteriota in diversity surveys [9] with several representative isolates described from soil [6, 7, 26-29], while representatives of groups 8, 10, and 23 have been isolated form non-soil habitats (e.g. hydrothermal vent, chiton microbiome, anoxic sediments, hot springs) [16-18, 30]. However, a detailed global meta-analysis to elucidate patterns of distribution and habitat preference for cultured and yet-uncultured lineages at various levels of phylogenetic resolution is currently lacking. The current availability of hundreds of thousands of metagenomic datasets, as well as a genome-based reference database and taxonomic outlines that encompass cultured as well as uncultured taxa [25] allows for assessing global ecological distribution patterns of target Acidobacteriota lineages at various levels of phylogenetic resolution using metagenomic read-mapping approaches. Such approaches are superior to assessments that rely on documenting the

habitat origin of isolates, 16S rRNA sequences, or genomes deposited in databases, where issues regarding differential amenability of lineages to culturing, sequence deposition procedures, preferential amplification of lineages, and genome assembly problems could impact the accuracy of the obtained outcome. As well, read mapping-based assessments would go beyond cursory characterization of habitat preferences by providing quantitative metrics for ubiquity (occurrence of a lineage in a specific ecosystem), preference (comparison of ubiquity levels of a specific lineage across ecosystems), and relative abundance (percentage of reads belonging to a specific lineage within a metagenomic dataset). Further, such assessment could also be used to identify whether a specific lineage is a habitat specialist (i.e. restricted to a single habitat) or a generalist (i.e. encountered in a wide range of habitats) at various taxonomic resolutions.

Progress in isolation, genome sequencing and -most importantly- generation of single cell genomes (SAGs) and metagenome-assembled genomes (MAGs) from environmental samples has greatly increased the number of publicly available Acidobacteriota genomes. Currently, the genome taxonomy database (GTDB) contains 2028 Acidobacteriota genomes from a wide array of habitats. As such, a phylum-wide comparative genomic analysis to elucidate and expand on key genomic features, metabolic capacities, and physiological preferences within various members in the Acidobacteriota is feasible. Such detailed analysis could also be used to disentangle the relative importance of phylogeny (lineage to which a genome belongs) versus habitat (origin from which the isolate, MAG, or SAG was recovered) in shaping genomic features and metabolic capacities in the Acidobacteriota.

Here, we present a detailed analysis of the global distribution patterns of the Acidobacteriota based on fragment recruitment from 248,559 metagenomic studies. We combine this analysis with a detailed comparative genomic analysis based on thousands of genomes

available in GTDB to identify key differences across lineages and habitats. Based on our results, we propose the occurrence of ready cross colonization of Acidobacteriota between soil and non-soil habitats through a continuous niche-selection process. We posit that such capacity is key to the rapid and successful establishment of Acidobacteriota as a major component of the soil microbiome during pedogenesis or during recolonization post drastic disturbance events. The implications of such findings on our understanding of how the evolution of soil as a distinct habitat on earth impacted the evolutionary trajectory of the *Acidobacteriota* is further discussed.

Materials and Methods.

Taxonomic framework for the Acidobacteriota. Several classification schemes have been proposed for the phylum Acidobacteriota. Barns et al [2] classified the phylum into 26 subgroups based on amplicons identified in subsurface sediments. Dedysh and Yelmaz built on this scheme and assigned the 26 subgroups into 15 class-level units [23]. The Genome Taxonomy Database (GTDB) [25] incorporates genomes from isolates as well as single amplified genomes (SAGs) and metagenome-assembled genomes (MAGs) into a global genome-based taxonomy, while using validly described names for isolates. The GTDB (release 214) classifies the phylum Acidobacteriota into 14 classes, 52 orders, 102 families, and 486 genera. Here, given the need for genome-based analysis, and the fact that the GTDB incorporates genomes from both cultured and uncultured lineages, we used the GTDB as our classification framework. Table S1 and [31] provide a comprehensive view of the phylum classification based on the three schemes discussed above [3, 23, 25], as well as the Silva classification database [24].

Ecological distribution of the Acidobacteriota. Identification of the taxonomy and relative abundances of key taxa in metagenomic datasets is key to deciphering their ecological distribution. We used Sandpiper, an interface that utilizes a recently developed tool (SingleM) for accurate mapping of metagenomic reads to genomes, to determine the occurrence and relative abundance of the phylum Acidobacteriota, its classes, orders, and families [32] in metagenomes. Our search comprised 248,559 metagenomic datasets in 17,617 projects with 1.3 Pbp of sequencing data. For each taxon searched, the output includes the number of datasets where the taxon was identified, along with individual accession numbers and ecological classification of each dataset, as well as relative abundance of the taxon in each dataset. We believe that this approach is superior to other approaches for metagenomic profiling, since shotgun sequencing

directly from an environment provides an unbiased view of the available community that is not prone to experimental issues, e.g. primer bias, that would be encountered with culture-independent studies. Also, searching metagenomic reads obtained with shotgun sequencing to identify taxa without assembling the reads into contigs and binning contigs into genomes (MAGs) would provide access to the fraction of the community that would otherwise not be binned into MAGs due to variability in coverage.

We used the output from Sandpiper searches to assess the distribution and prevalence patterns of the phylum Acidobacteriota, as well as its classes, orders, and families recognized in the GTDB taxonomy across six major habitats, as defined by [33]: soil (n=10,250 datasets), engineered (n=16,320 datasets), host-associated (n=165,152 datasets), marine (n=13,880 datasets), non-marine saline and alkaline ecosystems (n=1306 datasets), and terrestrial non soil ecosystems (n=5039 datasets). For each taxon (the phylum, as well as each class, order, family), three criteria were assessed: (1) ubiquity: assessed as the percentage of the total datasets from each habitat classification where a specific lineage was identified by read mapping, (2) Preference of a specific lineage for soil: assessed as the ratio of soil/non-soil datasets where a specific lineage was identified by read mapping, and (3) relative abundance of a specific lineage: assessed as the percentage of reads in a specific dataset that mapped to a target lineage. Due to the overrepresentation of host-associated (mostly human and murine) datasets in that database, and the well-established scarcity of Acidobacteriota in this habitat, all comparative analysis between soil and non-soil habitats, e.g. preference values, were conducted after exclusion of the host-associated datasets.

Habitat generalization versus specialization estimation. Habitat specialization was assessed by identifying the range of habitats where a specific lineage (class, order, family) was

encountered. The analysis was conducted for all taxa (classes, orders, and families) encountered in the GTDB database. However, it is important to note that taxa rarely encountered could erroneously be classified as specialists. Therefore, we also re-analyzed such patterns using anempirical cutoff that excludes taxa encountered in less than 250 studies.

Comparative genomic analysis. We assessed different genomic features, secretory arsenals, predicted physiological preferences, and lifestyle within Acidobacteriota genomes. We compared these characteristics across various lineages, as well as across habitats in a lineage-agnostic manner (i.e. comparing all genomes recovered from soil to those recovered from non-soil habitats). In addition, we also assessed within lineage habitat effects, i.e. whether genomic features differ between the same lineage genomes recovered from soil versus non-soil habitats. Out of the 2028 Acidobacteriota genomes available in GTDB, we focused our comparative genomic analysis on 1930 genomes belonging to classes with at least 100 genomes and for which genomes were recovered from at least 5 of the 7 habitats (Figure S1). These classes are Aminicenantia, Blastocatellia, Holophagae, Terriglobia, Thermoanaerobaculia, and Vicinamibacteria.

General genomic features (including genome size, GC content, coding density, and number of genes) were retrieved from the GTDB metadata file (available at https://data.gtdb.ecogenomic.org/releases/release214/214.1/). Average gene length was calculated using a perl script. CRISPRs were predicted using CCTyper [34]. Viral contigs were predicted using Virsorter2 [35]. For secretory capacities: CAZymes were predicted using dbCAN3 [36], while biosynthetic gene clusters (BGCs) were predicted using antiSMASH [37]. Proteases were identified via comparisons to the MEROPS database [38].

We used recently developed machine-learning (ML) based bioinformatic approaches to predict physiological preferences from genomic data. To predict optimal growth temperatures, we used the Tome suite [39]. pH preferences were assessed using a newly devised ML approach [40]. To predict oxygen preference from genomic data, we used the recently developed aerobicity ML predictor [41].

To predict life history strategy (Scarcity, Ruderal, or Competitor), we used the recently developed machine-learning approach that measures the tradeoff between genomic investment in resource acquisition and regulatory flexibility to predict the ecological strategy [42]. Genomic investment in regulatory flexibility was calculated as the number of transcription factors relative to total gene number. Transcription factors were predicted using two approaches: (1) Examining the output of BlastKOALA [43] for the presence of transcription factor Kegg Orthologies (KOs) (a list is available at https://www.genome.jp/brite/ko03000), and (2) additional transcription factors were predicted using DeepTFactor [44]. Genomic investment in resource acquisition was calculated as the number of genes encoding secreted enzymes (CAZymes, proteases, and lipases/hydrolases) plus the number of BGCs divided by the total number of membrane transporters. CAZymes and proteases were predicted as detailed above. Lipases/hydrolases were identified through comparison to the ESTHER database [45]. All three groups of enzymes were then subjected to SignalP [46] analysis to identify those with a secretion signal. Finally, membrane transporters were predicted by examining the output of BlastKOALA [43] for the presence of transporters Kegg Orthologies (KOs) (a list is available at https://www.genome.jp/brite/ko02000). Both genomic investment in resource acquisition (the number of genes encoding secreted CAZymes, proteases, and lipases/hydrolases plus the number of BGCs divided by the total number of membrane transporters) and in regulatory flexibility (the

number of transcription factors divided by the total number of genes) were then used to predict the life history strategy (one of Scarcity, Ruderal, or Competitor) of all Acidobacteriota genomes via kmeans clustering using the R package flexclust [47] and the k-centroid cluster analysis. The training dataset was comprised of the 27 ^{13}C-labeled MAGs from [42].

Metabolic potential encoded in Acidobacteriota genomes was predicted using METABOLIC [48].

Statistical analysis. Analysis of variance (ANOVA) (run through the aov command in R) was used to test for the effect of phylogeny (Class), habitat (the environmental source from which the genomes were obtained), and the interaction between the two on general genomic features (genome size, GC content, coding density, number of genes, and average gene length), phage infection/immunity features (number of viral contigs, number of CRISPR occurrence in a genome), potential extracellular products arsenal (CAZymes, peptidases, BGCs), predicted physiological optima (temperature, pH, and oxygen-preferences), and predicted life history strategy (ruderal, competitor, scarcity). Factors with F-test p-value $<1x10^{-5}$ were considered significant. The % contribution of phylogeny (Class), and habitat (the environmental source from which the genomes were obtained) was calculated based on the F-test sum of squares. For all significant comparisons, the TukeyHSD command in R was used for multiple comparisons of means.

To test for the effect of phylogeny and habitat on metabolic features predicted in the genomes, we first converted the dichotomous output (Presence/absence) from METABOLIC into 1/0 numerical output (if a pathway was identified as present in a genome, a value of 1 was used, while a value of 0 was used if the pathway was absent). Following, the effect of phylogeny and habitat on the metabolic potential (either the pattern of occurrence of KEGG modules or

212 TIGRfam/Pfam/custom HMM functions as predicted by METABOLIC) was tested using

213 ANOVA as explained above for genomic features.

Results

Global patterns of Acidobacteriota distribution across biomes. At the phylum level, Acidobacteriota showed the highest level of ubiquity in soil ecosystems, with fragments mapped to 95.7% of soil-derived datasets (9814 out of 10,250 soil datasets) (Table S2, Figure 1A). In comparison, Acidobacteriota-associated metagenomic fragments were mapped to only 37.6% of non-soil-derived datasets (18,027 out of 47,953 datasets from engineered, freshwater, marine, non-marine saline and alkaline, and terrestrial non-soil habitats) (Table S2, Figure 1A). Similarly, Acidobacteriota exhibited higher mean relative abundance in soil-derived datasets, where it was encountered on average in 15.79% of reads, compared to only 3.71% of reads in non-soil-derived datasets (Table S2, Figure 1C).

Four out of the fourteen classes of Acidobacteriota were highly ubiquitous in soil (identified in >75% of soil datasets). These were classes Terriglobia (92.61%), Vicinamibacteria (85.61%), Thermoanaerobaculia (78.84%), and Blastocatellia (76.2%) (Figure 1A, Figure 2, Table S2). In addition, these four classes also showed strong preference to soil (ratio of occurrence in soil to non-soil datasets >4) (Figure 1B, Table S2). This pattern was mostly driven by higher soil ubiquity and preference values for fourteen families belonging to ten orders within these classes (highlighted in Figure 2). Interestingly, while many of these lineages have representative isolates and are currently recognized as prevalent members of the microbial communities in soil, e.g. families *Acidobacteriaceae*, *Koribacteriaceae*, *Bryobacteraceae*, *Pyrinomonadaceae*, and *Vicinamibacteraceae* corresponding to subgroups 1, 3, 4, and 6 (Barns classification system); others represent lineages mostly identified through metagenomic studies, and have, so far, no cultured representatives or recognition in prior amplicon-based [2] and amplicon- and isolates-based [23] taxonomic schemes. For example, within the Terriglobia, in

addition to families *Acidobacteriaceae* and *Koribactereaceae* (subgroup 1), and *Bryobacteraceae* (subgroup 3), members of the yet uncultured families 20CM-2-55-15, UBA7541, and SbA1 also showed high levels of ubiquity and preference to soil ecosystems (Figure 2). Within the class Thermoanaerobaculia, named based on earliest isolates from high temperature settings [30], multiple uncultured families showed high levels of ubiquity and preference to soil, e.g. families Gp7-AA8, Gp7-AA6, and UBA5704. Similarly, within the class Vicinamibacteria, in addition to the family *Vicinamibacteraceae* (subgroup 6) [7], the uncultured families Fen-336, 2-12-FULL-66-21, SCN-69-37, and UBA2999 showed high levels of ubiquity and preference to soil. Finally, within the Blastocatellia, while members of the family *Pyrinomonadaceae* (subgroup 4) [6] were ubiquitous in soil, the uncultured family UBA7656 also showed high ubiquity and preference to soil.

Acidobacteriota classes that were less ubiquitous in soil included the Polarisedimenticolia (encountered in 33.47% of soil studies), UBA6911 (encountered in 29.48% of soil studies), Holophagae (encountered in 24% of soil studies), HRBIN11 (encountered in 15.78% of soil studies), and Aminicenantia (encountered in 11.5% of soil studies) (Figure 1A, Figure 2, Table S2). Within these five classes, UBA6911, Polarisedimenticolia, and HRBIN11 showed preference to soil (soil: non-soil occurrence ratio >4, Figure 1B, Table S2), while classes Holophagae and Aminicenantia showed lower ratios (<4). Finally, the remaining classes (B3-B38, CAIWXX01, UBA4820, UBA890, and G020349885) were extremely rare (less than 4%) in all habitats examined.

We also assessed relative abundance values (proportion of reads mapped to a specific lineage within a metagenomic dataset) of various classes, orders, and families in the Acidobacteriota, an indirect measure of their niche-colonization capacities and relative

contribution to ecosystem functions within a specific habitat (Figure 1C). Classes identified as most ubiquitous in soil also showed the highest relative abundance in soil, with relative abundance values of 9.76%, 3.87%, 1.9%, and 0.92% for Terriglobia, Vicinamibacteria, Blastocatellia, and Thermoanaerobaculia, respectively (Figure 1C). Out of these four classes, Terriglobia had the highest ratio of relative abundance between soil and non-soil datasets (5.2), followed by Vicinamibacteria, Blastocatellia, and Thermoanaerobaculia with soil: non-soil relative abundance ratios of 2.43, 1.42, and 1.12, respectively (Figure 1D). On the other hand, classes with lower ubiquity in soil (Polarisedimenticolia, HRBIN11, Holophagae, Aminicenantia, and UBA6911) showed lower relative abundance in soil compared to non-soil habitats with ratios consistently <1 (Figure 1D, Table S2).

Therefore, based on the above analysis of soil ubiquity, preference, and relative abundance, we infer that members of the four classes Terriglobia, Vicinamibacteria, Blastocatellia, and Thermoanaerobaculia show high level of ubiquity, preference, and relative abundance in soil over other habitats. We refer to these lineages henceforth as "soil-preferring lineages, SPL". On the other hand, members of the two classes Holophagae and Aminicenantia show lower soil ubiquity, lower preference to soil habitats, and lower relative abundance in soil versus non soil datasets. We refer to these two lineages henceforth as "non-soil preferring lineages, NSPL". Finally, the three moderately soil-ubiquitous Acidobacteriota classes UBA6911, Polarisedimenticolia, and HRBIN11 showed moderate relative abundance values that were lower in soil, compared to non-soil habitats (values of soil: non-soil relative abundance ratio <1).

Habitat generalization versus specialization in the Acidobacteriota. We examined patterns of habitat generalization versus specialization in the Acidobacteriota at various levels of taxonomic

283 resolution. We found 78.6%, 61.5%, and 52.9% of classes, orders, and families, respectively,
284 within the Acidobacteriota to be habitat generalists (i.e. with reads belonging to these lineages
285 identified in each of the 7 habitat classifications) (Table S3). However, it is important to note that
286 such pattern could greatly be impacted by sampling depth, with taxa rarely encountered
287 erroneously classified as specialists. Exclusion of rare taxa from our datasets (empirically
288 defined as those encountered in less than 250 studies) results in even higher percentage of
289 generalist taxa (91.7%, 77.5%, and 66.7% of classes, orders, and families, respectively). Our
290 analysis indicate that preferences does not necessarily correspond to specialization, and lineages
291 identified as SPL or NSPL could still be habitat generalists and exhibit wide global multi-habitat
292 distribution patterns. For example, families *Acidobacteriaceae*, *Bryobacteraceae*,
293 *Pyrinomonadaceae*, and *Vicinamibacteraceae* while encountered in 24.76-97.32% of soil
294 datasets, were also present in 2.03-19.75% of engineered, 4-31% of freshwater, 0.61-4.93% of
295 marine, 0.53-10.8% of non-marine saline and alkaline, and 1.47-3.65% of terrestrial non-soil
296 datasets. On the other hand, NSPL families *Aminicenantaceae* and *Holophagaceae,* while
297 enriched in non-soil datasets were also present in 1.23, and 19.26%, respectively, of soil datasets.
298 To confirm the results obtained from the global metagenomic dataset utilized, we
299 examined the habitat generalization and specialization patterns in the collection of 2028
300 Acidobacterial genomes available in the GTDB database [25]. Our results (Figure S2, Table S3)
301 confirm habitat generalization of the Acidobacteriota at the class, order, and family levels, with
302 85.7%, 65.4%, and 54.91% of classes, orders, and families, respectively, within the
303 Acidobacteriota with genome representatives obtained from 2 or more habitat classifications
304 (Table S3). 24.69% of genera had genome representatives obtained from more than one habitat,
305 while 75.3% had genome representatives obtained from only one habitat. Further, when

306 removing lineages defined by less than 5 genomes (404 genera, 48 families, 15 orders, 3 classes),
307 91.67%, 84.49%, 87.04%, and 62.2% of classes, orders, families, and genera, respectively, were
308 not specialized to one type of habitat (Table S3).
309 **Comparative genomic features across lineages and habitats.** We assessed the general
310 genomic features (genome size, GC content, coding density, average gene length, and number of
311 protein-coding genes (Figure 3)), phage infection/immunity features (number of viral contigs,
312 number of CRISPR occurrence in a genome (Figure 4)), potential extracellular products arsenal
313 (CAZymes, peptidases, BGCs (Figure 5)), predicted physiological optima (temperature, pH, and
314 oxygen-preferences (Figures 6-7)), and predicted life history strategy (ruderal, competitor,
315 scarcity (Figure 8)) in the 1930 genomes belonging to six classes with at least 100 genomes and
316 for which genomes were recovered from at least 5 of the 7 habitats (Aminicenantia,
317 Blastocatellia, Holophagae, Terriglobia, Thermoanaerobaculia, and Vicinamibacteria (Table 1,
318 Figure S1, Table S4)). Comparative analysis was conducted to determine whether significant
319 differences could be identified at a phylogenetic level (i.e. between different lineages within the
320 Acidobacteriota), as well as at the habitat level (i.e. between genomes recovered from soil versus
321 genomes recovered from non-soil habitats, regardless of their phylogenetic affiliation), and the
322 relative contribution of phylogeny versus habitat in shaping Acidobacteriota genomes.
323 Our analysis demonstrated a significant role played by phylogenetic affiliation in shaping
324 all fourteen examined criteria in Acidobacteriota genomes (F test p-value $<1.2 \times 10^{-10}$,
325 percentage lineage contribution 3.6-45.3%, Figures 3-8, Table S5). Interestingly, a trend in some
326 of these criteria was observed where SPL lineages were more similar to each other, compared to
327 NSPL lineages. Specifically, SPL lineages possessed larger genomes (F test p-value = 2×10^{-16},
328 phylogeny percentage contribution=14.96%), higher GC content F test p-value = 2×10^{-16},

329 phylogeny percentage contribution=45.3%), lower gene density F test p-value = $2x10^{-16}$,
330 phylogeny percentage contribution= 26.31%), shorter average gene length F test p-value = $2x10^{-16}$, phylogeny percentage contribution= 10.31%), and a higher number of protein-coding genes F
332 test p-value = $2x10^{-16}$, phylogeny percentage contribution= 18.01%) (Figure 3); as well as a
333 significantly larger number of viral contigs (F test p-value = $4.3x10^{-13}$, phylogeny percentage
334 contribution= 4.18%) (Figure 4) and a significantly expanded CAZymes (F test p-value = $2x10^{-16}$, phylogeny percentage contribution= 36.42%), peptidases (F test p-value = $2x10^{-16}$, phylogeny
336 percentage contribution= 27.83%), and BGCs (F test p-value = $2.6x10^{-10}$, phylogeny percentage
337 contribution= 3.53%) (Figure 5), and a higher proportion of genes with a predicted aerobic
338 oxygen preference (F test p-value = $2x10^{-16}$, phylogeny percentage contribution= 41.26%)
339 (Figure 7). For the remaining criteria, while phylogeny played a significant role in shaping the
340 Acidobacteriota genome examined, the SPL lineage: NSPL lineage dichotomy across classes was
341 not observed. For example, while phylogeny had a significant effect on predicted optimal growth
342 temperature (OGT) (F test p-value =$2x10^{-16}$, phylogeny percentage contribution=34.3%), and in
343 general SPL were predicted to have lower optimal pH (Figure 6), this pattern was mainly due to
344 the higher predicted OGT for the Aminicenantia genomes, while Holophagae genomes were
345 predicted to have lower OGT than genomes from SPL lineages (Figure 6). Similarly, while
346 phylogeny had a significant effect on predicted optimal growth pH (F test p-value =$2x10^{-16}$,
347 phylogeny percentage contribution=43.2%), this pattern was mainly due to the lower predicted
348 pH for Terriglobia genomes, while Blastocatellia and Vicinamibacteria genomes were predicted
349 to have higher pH than genomes from NSPL lineages. Finally, for life history strategies (ruderal,
350 competitor, scarcity), phylogeny played a significant role in shaping the predicted life history
351 strategy (F test p-value =$2x10^{-16}$, phylogeny percentage contribution=9.13%), and in general SPL

were predicted to have more competitor strategy. However, this pattern was mainly due to the higher number of predicted competitors for Terriglobia genomes, while other SPL classes showed no significant difference in terms of the distribution of predicted life history strategies from the two NSPL classes (Figure 8).

On the other hand, habitat-specific, but lineage-agnostic, comparisons (i.e. comparing genomes originating from soil to those originating from non-soil sources, regardless of their phylogenetic affiliation) identified habitat-specific strong significant difference in 7/14 criteria (F test p-value $<10^{-5}$), with genomes from soil predicted to produce more peptidases, and to have lower coding density, higher number of protein-coding genes, shorter average gene length, lower predicted OGT and lower predicted pH, and predicted to be more competitor. Weak but significant differences were observed for GC percentage (lower in soil genomes), genome size (larger in soil genomes), and number of viral contigs (lower in soil genomes), while the number of predicted BGCs, and CAZymes did not significantly differ by habitat, nor did the total predicted number of CRISPRs, or the preference for O_2 (Table S5).

While the habitat from which the genome was obtained played a role in shaping genomes in 7/14 criteria; major contribution in these criteria was from the lineage (values of lineage contribution ranging from 3.6-45.3%, as opposed to 0.3-12.3% for the habitat) (Table S5). To further examine differences on a finer scale, we compared within lineages genomes originating from soil versus non-soil habitats. The SPL classes Terriglobia, Vicinamibacteria, Blastocatellia, and Thermoanaerobaculia showed more within-lineage significant differences between genomes originating from soil versus non-soil habitats with 9, 4, 2, and 1, respectively, genomic features identified as within-lineage habitat-specific. These included lower coding density, shorter genes, lower GC percentage, higher number of protein-coding genes, more predicted peptidases, less

predicted viral contigs, lower predicted OGT, lower predicted pH, and more predicted competitor lifestyle in Terriglobia soil genomes compared to non-soil genomes, larger genomes, shorter genes, more predicted peptidases and CAZymes in Vicinamibacteria soil genomes compared to non-soil genomes, shorter genes and lower coding density in in Blastocatellia soil genomes compared to non-soil genomes, and shorter genes in Thermoanaerobaculia soil genomes compared to non-soil genomes. On the other hand, NSPL classes showed less within-lineage significant differences between genomes originating from soil versus non-soil habitats with only 1 genomic feature (GC percentage) identified as within-lineage habitat-specific for class Aminicenantia (with soil Aminicenantia genomes having higher GC percentage than non-soil genomes) (Table S5).

Comparative genomic analysis identifies lineage-specific rather than habitat-specific metabolic differences. We conducted detailed comparative genomic analysis for the metabolic capacities predicted for the 1930 genomes in the SPL classes Blastocatellia, Terriglobia, Thermoanaerobaculia, and Vicinamibacteria, as well as the two NSPL classes Aminicenantia, and Holophagae. We aimed to determine whether significant differences could be identified at a phylogenetic level (i.e. between different lineages within the Acidobacteriota), as well as at the habitat level (i.e. between genomes recovered from soil versus genomes recovered from non-soil habitats, regardless of their phylogenetic affiliation), and the relative contribution of phylogeny versus habitat in shaping Acidobacteriota metabolic capacities.

We identify 73 features where SPL classes differ significantly from NSPL classes (Table S6). These features were mainly distributed among catabolic and anabolic pathways. On the catabolic front, genomes from SPL classes encoded significantly higher O_2 respiration genes (both low affinity and high affinity cytochrome oxidase), as well as dissimilatory sulfate

reduction genes (Table S6). On the other hand, genomes from NSPL classes encoded significantly higher nitrate reduction to ammonium and dissimilatory nitrate reduction genes. Genomes from SPL classes encoded significantly higher genes for ethanol fermentation from acetyl-CoA, while those from NSPL genomes encoded significantly higher phosphate acetyltransferase-acetate kinase pathway for acetate fermentation with concomitant substrate level phosphorylation, and significantly more hydrogenases belonging to the fermentative hydrogenogenic classes [FeFe] hydrogenase group b, [NiFe] hydrogenase group 3b-3d, and [NiFe] hydrogenase group 4a-g, as well as the hydrogenotrophic hydrogenase classes in classes [NiFe] group 1, and [NiFe] group 3c. However, the lack of clear autotrophic potential in these NSPL genomes cast some doubt on these results. It has previously been suggested that atmospheric H_2 can be scavenged by the thermophilic *Pyrinomonas methylaliphatogenes* K22 in absence of organic carbon as a means of operating the respiratory chain when organic electron donors are scarce [49]. Recently, more Acidobacteriota genomes were shown to encode such hydrogenases [50]. A similar role for these hydrogenotrophic hydrogenases can be predicted for these NSPL classes.

Further, SPL classes were enriched in some central metabolic pathways that are usually associated with the higher O_2 tension conditions in soil (Table S6). These included the oxidative branch of the PPP, the oxidative TCA cycle, and the semi-phosphorylative Entner-Doudoroff pathway (which joins the oxidative branch of the PPP to glycolysis). SPL genomes are also enriched in the degradation of the sugars/sugar acids galactose, glucuronate, and galacturonate, amino acids Pro, Trp, Met, and hydroxyproline (an amino acid rich in plants cell wall proteins and glycoproteins, and so would be abundant in soil constituting a rich source of C and N for soil

bacteria [51]), and in nucleotide degradation (uracil to β-alanine, thymine to 3-amino-isobutanoate, and xanthine to urea) (Table S6).

On the anabolic side, SPL classes exhibited a higher biosynthetic capacity, where biosynthesis of seven amino acids (Lys, Arg, Thr, Val, Ile, His, and Trp) and nine cofactors (Menaquinone, heme, pyridoxal-P, NAD, pantothenate, THF, Lipoic acid, riboflavin, and molybdenum cofactor) were significantly enriched in the four SPL classes (F-test p-value <1e-5). On the other hand, the biosynthesis of only one amino acid (Pro), and two cofactors (thiamine, and biotin) were enriched in the two NSPL classes. Interestingly, some of the cofactors with enriched biosynthesis in the SPL classes are functional during the operation of the electron transport chain for O_2 respiration (e.g. menaquinone and heme), a trait enriched in SPL lineages (Figure 7). SPL classes were also more enriched in trehalose biosynthesis. The disaccharide trehalose acts as an osmolyte whose biosynthesis was shown to be upregulated in dry soil mainly contributing to tolerance to water deficits [52-55]. On the other hand, NSPL classes were enriched in polyamine biosynthesis. Polyamines are aliphatic polycations with diverse function related to cell growth, biofilm formation, as well as secondary metabolite production [56]. Recently, [57] demonstrated a link between increasing intracellular concentration of polyamines and improved oxidative stress tolerance in *Pseudomonas*. A similar role for polyamine can be speculated in the NSPL classes.

On the other hand, habitat-specific comparisons identified only 45 features where genomes from soil differ significantly from genomes from non-soil sources (Table S6). The majority of these features (42 out of 45) were identified as significantly affected by phylogeny, with only three features (fatty acids beta oxidation pathway genes, N_2 fixation genes, and high affinity O_2 respiration using cytochrome oxidase cbb3) identified as habitat-specific (but not

443 lineage-specific). Statistical analysis of features where both phylogeny and habitat were
444 significant (n=42) showed lineage to be more important (lineage percentage contribution: 2-
445 50.8%), with habitat contributing < 7.2%. Of the 42 features where both phylogeny and habitat
446 were significant, 39 followed the same trend (where the feature was higher/lower in SPL as well
447 as genomes sourced from soil regardless of their lineage), with only 3 features with a divergent
448 trend. These later features included phosphatidyl choline biosynthesis, tetrahydrofolate
449 biosynthesis, and nitrite oxidation (significantly higher in SPL genomes and in genomes from
450 non-soil sources).

Discussion

Here, we examined Acidobacteriota global distribution patterns using a metagenomic read mapping approach, and identified salient differences in genomic characteristics and metabolic capacities across lineages and habitats in the Acidobacteriota. Four classes (Blastocatellia, Terriglobia, Thermoanaerobaculia, and Vicinamibacteria) were shown to have clear preference to soil over non-soil habitats, based on their high soil ubiquity (encountered in >75% of soil datasets), soil preference (ratio of occurrence in datasets from soil versus non soil datasets >4), and higher relative abundance in soil datasets compared to non-soil datasets (Figure 2, Table S2). These four classes encompass many Acidobacteriota taxa previously isolated from soil or shown to occur in high abundance in soil in amplicon-based surveys. The class Terriglobia encompasses subgroups 1 (Families *Acidobacteriaceae* and *Koribacteraceaea*), and 3 (family *Bryobacteraceae*). Class Vicinamibacteria encompasses subgroup 6 (family *Vicinamibacteraceae*). Class Thermoanaerobaculia encompasses subgroups 10, and 23 (family *Thermoanaerobaculaceae*). Class Blastocatellia encompasses subgroup 4 (Families *Pyrinomonadaceae*, *Blastocatellaceae*, and *Chloracidobacteriaceae*). More importantly, we identify multiple additional poorly characterized yet uncultured lineages that are ubiquitous and show high preference and high relative abundance in soil datasets (Figure 2, Table S2). These include the uncultured families UBA7541 (subgroup 2), and SbA1 in Terriglobia, and UBA2999 in Vicinamibacteria. Most of these families are defined based on MAGs recovered from metagenomic studies and are not currently recognized in amplicon- and isolation-based classification. The recognition of the prevalence of such lineages should spur efforts towards more detailed –omics-based characterization and assessment of their importance and potential

473 role in elemental cycling and contribution to ecosystem functioning in soil, as well as stimulate
474 endeavors towards obtaining them in pure cultures.

475 Organisms encountered in multiple niches are referred to as generalists, while those
476 restricted to a single habitat are specialists. Studies usually utilize occurrence patterns as an
477 empirical measurement of a generalist versus specialist pattern either within a target habitat [58],
478 through a habitat transition gradient, or on a global scale [59]. Assignment of microbial taxa as
479 generalists or specialists could be affected by the level of phylogenetic resolution employed,
480 number of datasets examined, sequencing depth of datasets, and detection threshold employed.
481 Prior research provided some insights on such patterns in the Acidobacteriota. For example, a
482 study examining a large collection of datasets from farmland soils suggested that Acidobacteriota
483 harbored a larger fraction of generalists than specialists at the species OTUs level [58]. On the
484 other hand, studies assessing Acidobacteriota generalist-specialist patterns on a global scale are
485 quite sparse. A recent analysis, defining specialists and generalists using the level of community
486 similarity between datasets where a specific lineage is encountered (with high community
487 similarity indicating a specialist pattern and low community similarity indicating a generalist
488 pattern) as a substitute for occurrence patterns, concluded that Acidobacteriota encompassed a
489 high proportion of specialized genera [59].

490 Our analysis of Acidobacteriota distribution patterns strongly suggests that a habitat-
491 generalist rather than a habitat-specialist pattern is more common in the Acidobacteriota (Table
492 S3). Using two datasets (publicly available metagenomes via Sandpiper, and publicly available
493 genomes in GTDB), criteria (with and without exclusion of rare taxa, with and without exclusion
494 of taxa with less than 5 genomes), and taxonomic thresholds (class, order, family), we estimate

that 73.3-91.7% of classes, 61.5-86.49% of orders, and 52.9-87.04% of families are habitat generalists, with documented ability to inhabit multiple environments.

We attribute the prevalence of habitat-generalist over habitat-specialist pattern in the Acidobacteriota to two possible reasons: the nature of its preferred habitat(s), and the genomic features and metabolic capacities of its members. Broadly, habitats that are extreme, restricted, and drastically different from their surroundings favor specialists, e.g. chemolithotrophic hyperthermophiles in hydrothermal vents [60], strict halophilic Archaea in hypersaline ecosystems [61], anaerobic gut fungi in the herbivorous alimentary tracts [62]. On the other hand, generalists thrive across more temperate, less restricted, and more complex habitats. Broadly, the habitats where Acidobacteriota appears to thrive, e.g. soil, and freshwater, are temperate, with fluctuating temperature, pH, salinity, as well as a complex variable influx of substrates. Such conditions allow for generalists, organisms usually exhibiting a wider arsenal of substrate utilization capacities and response to environmental fluctuations, to thrive as integral components of a complex habitat.

Regarding metabolic abilities, habitat-specialists' genomes are usually streamlined and either mediate a specific function or encode exceptional capacity to survive and adapt to a unique ecosystem [63]. Habitat-generalists' genomes, on the other hand, are more metabolically versatile to enable adaptation, e.g. *Pseuodomonas* [64], *Burkholderia* [65] thriving under so many substrates, *Shewanella* thriving under many electron acceptors [66]. Within the Acidobacteriota, genomic analysis identified moderate to large genomes and suggested a general prevalence of heterotrophy, and lack of auxotrophies. Our machine-learning-based analysis suggested prevalence of moderate predicted optimal growth temperature and pH, as well as

517 variable predicted oxygen preferences, all of which would explain the prevalence of habitat-
518 generalist patterns in the phylum.

519 Prior genomic analysis of members of the Acidobacteriota has been conducted on pure
520 culture isolates [14, 50, 67, 68] or assembled MAGs [4, 22, 31, 69, 70]. These efforts have
521 yielded valuable insights into the salient genomic features of various members of the phylum.
522 Our analysis aimed for a broader and more comprehensive view of Acidobacteriota genomes via
523 a global comparative analysis approach. In addition to expanding on known capacities within the
524 phylum, we aimed to identify variations in genomic features, physiological optima, and
525 metabolic capacities within members of the Acidobacteriota, and to determine whether the
526 observed differences are habitat-specific (e.g. present in all genomes of soil Acidobacteriota
527 regardless of the lineage they belong to), lineage-specific (e.g. present in all genomes of a
528 lineage regardless of the environmental source), or a combination of both. Our results suggest
529 that lineage is more important than habitat in describing the observed differences, with very few
530 exceptions. Within general genomics features and physiological optima prediction, lineage
531 significantly affected all 14 criteria compared (F test P-value $<2.6 \times 10^{-10}$), while habitat only
532 significantly affected 7 criteria, and the interaction between the two only significantly affected 4
533 criteria (GC percentage, number of predicted peptidases, predicted optimal growth temperature,
534 and predicted optimal growth pH) (Table S5, Figures 3, 5, 6). In all these comparisons, lineage
535 explained 3.5-45.3% of the variability, while habitat only explained 0.91-12.34%. Detailed
536 analysis of metabolic identified 73 functions that were significantly different between SPL and
537 NSPL classes, while habitat only significantly affected 45 functions (42 of which were also
538 significantly affected by lineage). In all these comparisons, lineage explained 2.02-50.8% of the
539 variability, while habitat % contribution never exceeded 7.2%. Indeed, many of the metabolic

features previously shown to underpin success of Acidobacteriota in soil, were found to be lineage-specific (significantly encountered more in SPL genomes), rather than habitat-specific (present in all soil Acidobacteriota regardless of their phylogeny). For examples genes encoding CAZymes and BGCs were significantly higher in SPL genomes, but these numbers did not significantly vary in genomes derived from soil versus genomes derived from non-soil environments. However, while clearly lineage plays a more important role, some habitat-specific features are of note. Soil as a habitat appears to select for genomes with lower coding density, higher number of protein-coding genes, shorter average gene length, larger number of peptidases, and for organisms with lower predicted OGT and lower predicted pH, and with a predicted competitor lifestyle.

Finally, the observed prevalence of a generalist patten and higher importance of lineage compared to habitat in shaping Acidobacteriota genomes strongly suggest a ready cross-colonization of Acidobacteriota across major habitats. Under this scenario, while differential preference of certain Acidobacteriota lineages to specific habitat exists, the presence of a ready reservoir, albeit in minor quantities, in other environments would allow ready cross-colonization when needed, e.g. as a mechanism for repopulation post-disturbances (e.g. fire, extensive pollution), or during the process of pedogenesis (soil formation). Nevertheless, our results also advocate for a role played by the environment, specifically soil studied here, in shaping Acidobacteriota genomes post-acquisition. Specifically, larger number of encoded peptidases, lower predicted OGT and pH, and a predicted competitor lifestyle would allow organisms to survive better in the soil ecosystem. Genomes from soil also encoded pathways beneficial to surviving in such ecosystem, e.g. N_2 fixation, purine degradation to urea, and beta lactam resistance.

563 In conclusion, our results identify clear soil preferences for four classes (Terriglobia,
564 Vicinamibacteria, Blastocatellia, and Thermoanaerobaculia) of Acidobacteriota. We also
565 demonstrate that such preferences are driven not only by taxa previously recognized as
566 prominent soil-dwellers in prior isolation and amplicon efforts, but also by multiple yet-
567 uncultured orders and families in these four classes that are ubiquitous and abundant yet poorly
568 characterized. Further, our analysis indicates that despite the observed preference patterns, most
569 Acidobacteriota classes, orders, families, and genera are habitat generalists rather than
570 specialists. As well, our global comparative genomic analysis provides new insights into the
571 genomic features, predicted physiological optima, and metabolic repertoires of members of the
572 Acidobacteriota, and disentangles the role played by phylogeny versus habitat in shaping
573 Acidobacteriota genomic and predicted metabolic features.
574
575 **Acknowledgments.** This work has been supported by the US National Science Foundation
576 (NSF) grant 2016423 to N.H.Y. and M.S.E., and the US National Institute of Health (NIH) grant
577 1P20GM152333 to MSE.

578 **Table 1.** Genomes compared in this study. Total number of genomes compared belonging to each of the 6 Acidobacterioat classes is
579 shown. Number of genomes originating from soil, engineered, freshwater, host-associated, marine, non-marine saline and alkaline,
580 and terrestrial non-soil are also shown for each class.

Class	Total number of genomes	Number of genomes from soil	Number of genomes from non-soil environments					
			Engineered	Freshwater	Host-associated	Marine	Non marine saline and alkaline	Terrestrial non-soil
Blastocatellia	215	64	103	35	8	1	0	4
Terriglobia	1006	537	179	197	50	16	0	27
Thermoanaerobaculia	153	25	49	22	37	17	0	3
Vicinamibacteria	231	44	35	59	47	46	0	0
Aminicenantia	119	7	27	49	0	35	1	0
Holophagae	206	16	30	151	4	5	0	0

581 **Figure legends**

582 **Figure 1.** Ecological distribution of the phylum Acidobacteriota and its 14 classes in 248,559
583 metagenomic datasets available through the web interface Sandpiper [32]. (A) Percentage
584 occurrence of members of the phylum Acidobacteriota and its 14 classes in datasets originating
585 from soil (■), and non-soil (■) ecosystems. Classes are shown on the X-axis. The dotted line
586 represents 75% occurrence, the cutoff used to define a lineage as ubiquitous in an ecosystem. (B)
587 The ratio of percentage occurrence of the phylum and each of its 14 classes in soil versus non-
588 soil ecosystems. The dotted line represents a ratio of 4, the cutoff used to define a lineage as soil-
589 preferring in an ecosystem. (C) Average percentage abundance of members of the phylum
590 Acidobacteriota and its 14 classes in datasets originating from soil (■), and non-soil (■)
591 ecosystems. Classes are shown on the X-axis. The dotted line represents 1% occurrence, the
592 cutoff used to define a lineage as abundant in an ecosystem. (D) The ratio of percentage
593 abundance of the phylum and each of its 14 classes in soil versus non-soil ecosystems. The
594 dotted line represents a ratio of 1, the cutoff used to define a lineage as relatively more abundant
595 in an ecosystem.

596 **Figure 2.** Phylogenomic tree constructed using the GTDB concatenated alignment of 120 single
597 copy marker gene. The tree was constructed in FastTree [71] and is wedged at the family level.
598 Family names are shown on the right. Wedges are color coded by order (shown to the left of the
599 tree). Class names are shown to the left. Families and orders with higher soil ubiquity and
600 preference values are highlighted by an asterisk (*). The values corresponding to the percentage
601 occurrence of each family in soil (■), and non-soil (■) ecosystems, as well as the ratio (■) of
602 percentage occurrence in soil versus non-soil ecosystems are shown as horizontal bars to the
603 direct right of the tree. Families with >25% occurrence in soil datasets were considered

604 ubiquitous in soil, while families with a ratio of percentage occurrence in soil versus non-soil
605 ecosystems >4 were considered soil-preferring. For ease of visualization, dashed vertical lines
606 are shown for the 20% and 50% occurrence in soil datasets, and for the soil: non-soil %
607 occurrence ratios 4, 20, 50, and 100. Horizontal bars to the right of the thick vertical line
608 represent the average percentage abundance of each family in soil (), and non-soil ()
609 ecosystems, as well as the ratio () of percentage abundance in soil versus non-soil ecosystems.
610 Families with a ratio of percentage abundance in soil versus non-soil ecosystems >1 were
611 considered more abundant in soil. For ease of visualization, dashed vertical lines are shown for
612 the 1% abundance in soil datasets, 1% and 20% abundance in non-soil datasets, and for the soil:
613 non-soil % abundance ratio of 1.
614 **Figure 3.** Box plots for the distribution of several general genomic features in the 1930 genomes
615 of the 6 Acidobacteriota classes compared in this study. Features shown are genome size (top
616 row), genome percentage GC (second row from top), genome coding density (third row from
617 top), average gene length (second to last row), and protein count (bottom row). Results for two-
618 tailed ANOVA followed by Tukey for pairwise comparisons are shown on top of the box plots
619 only for significant comparisons. *, $0.01 < p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p <$
620 0.0001; *****, $p < 0.00001$. The first two columns show results for lineage-specific comparisons
621 at the Acidobacteriota Class level, as well as when combining Acidobacteriota SPL (Terriglobia,
622 Vicinamibacteria, Thermoanaerobaculia, and Blastocatellia) versus NSPL (Holophagae and
623 Aminicenantia). The third column shows results for habitat-specific comparisons (genomes
624 originating from soil versus non-soil environments regardless of phylogeny). Results of within-
625 lineage habitat-specific comparisons are shown in the fourth column, where genomes from soil

626 origin are shown in cyan, while genomes from non-soil origin are shown in red. The exact p-
627 values for all comparisons are shown in Table S4 and S5.

628 **Figure 4.** Box plots for the distribution of phage infection/immunity features in the 1930
629 genomes of the 6 Acidobacteriota classes compared in this study. Features shown are number of
630 CRISPRs (top row), and number of viral contigs (bottom row). Results for two-tailed ANOVA
631 followed by Tukey for pairwise comparisons are shown on top of the box plots only for
632 significant comparisons. *, $0.01 < p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$;
633 *****, $p < 0.00001$. The first two columns show results for lineage-specific comparisons at the
634 Acidobacteriota Class level, as well as when combining Acidobacteriota SPL (Terriglobia,
635 Vicinamibacteria, Thermoanaerobaculia, and Blastocatellia) versus NSPL (Holophagae and
636 Aminicenantia). The third column shows results for habitat-specific comparisons (genomes
637 originating from soil versus non-soil environments regardless of phylogeny). Results of within-
638 lineage habitat-specific comparisons are shown in the fourth column, where genomes from soil
639 origin are shown in cyan, while genomes from non-soil origin are shown in red. The exact p-
640 values for all comparisons are shown in Table S4 and S5.

641 **Figure 5.** Box plots for the distribution of potential extracellular products arsenal in the 1930
642 genomes of the 6 Acidobacteriota classes compared in this study. Features shown are number of
643 biosynthetic gene clusters (BGCs, top row), number of CAZymes (middle row), and number of
644 proteases (bottom row). Results for two-tailed ANOVA followed by Tukey for pairwise
645 comparisons are shown on top of the box plots only for significant comparisons. *, $0.01 < p <$
646 0.05; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; *****, $p < 0.00001$. The first two columns
647 show results for lineage-specific comparisons at the Acidobacteriota Class level, as well as when
648 combining Acidobacteriota SPL (Terriglobia, Vicinamibacteria, Thermoanaerobaculia, and

649 Blastocatellia) versus NSPL (Holophagae and Aminicenantia). The third column shows results
650 for habitat-specific comparisons (genomes originating from soil versus non-soil environments
651 regardless of phylogeny). Results of within-lineage habitat-specific comparisons are shown in
652 the fourth column, where genomes from soil origin are shown in cyan, while genomes from non-
653 soil origin are shown in red. The exact p-values for all comparisons are shown in Table S4 and
654 S5.

655 **Figure 6.** Box plots for the distribution of predicted physiological optima in the 1930 genomes
656 of the 6 Acidobacteriota classes compared in this study. Features shown are predicted optimal
657 growth temperature (OGT, top row), and predicted optimal pH (bottom row). Results for two-
658 tailed ANOVA followed by Tukey for pairwise comparisons are shown on top of the box plots
659 only for significant comparisons. *, $0.01 < p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p <$
660 0.0001; *****, $p < 0.00001$. The first two columns show results for lineage-specific comparisons
661 at the Acidobacteriota Class level, as well as when combining Acidobacteriota SPL (Terriglobia,
662 Vicinamibacteria, Thermoanaerobaculia, and Blastocatellia) versus NSPL (Holophagae and
663 Aminicenantia). The third column shows results for habitat-specific comparisons (genomes
664 originating from soil versus non-soil environments regardless of phylogeny). Results of within-
665 lineage habitat-specific comparisons are shown in the fourth column, where genomes from soil
666 origin are shown in cyan, while genomes from non-soil origin are shown in red. The exact p-
667 values for all comparisons are shown in Table S4 and S5.

668 **Figure 7.** Bar plots for the distribution of the predicted aerobic (■) versus anaerobic (■) lifestyle
669 in the 1930 genomes of the 6 Acidobacteriota classes compared in this study. Results are shown
670 by class (A), Soil preference (SPL versus NSPL, B), habitat from which the genome originated

(C), and within lineage broken down by habitat (D). Number of genomes in each category is shown on top of the bar.

Figure 8. Bar plots for the distribution of the predicted life history strategy (ruderal (■), competitor (■), or scarcity (■)) in the 1930 genomes of the 6 Acidobacteriota classes compared in this study. Results are shown by class (A), Soil preference (SPL versus NSPL, B), habitat from which the genome originated (C), and within lineage broken down by habitat (D). Number of genomes in each category is shown on top of the bar.

681 References

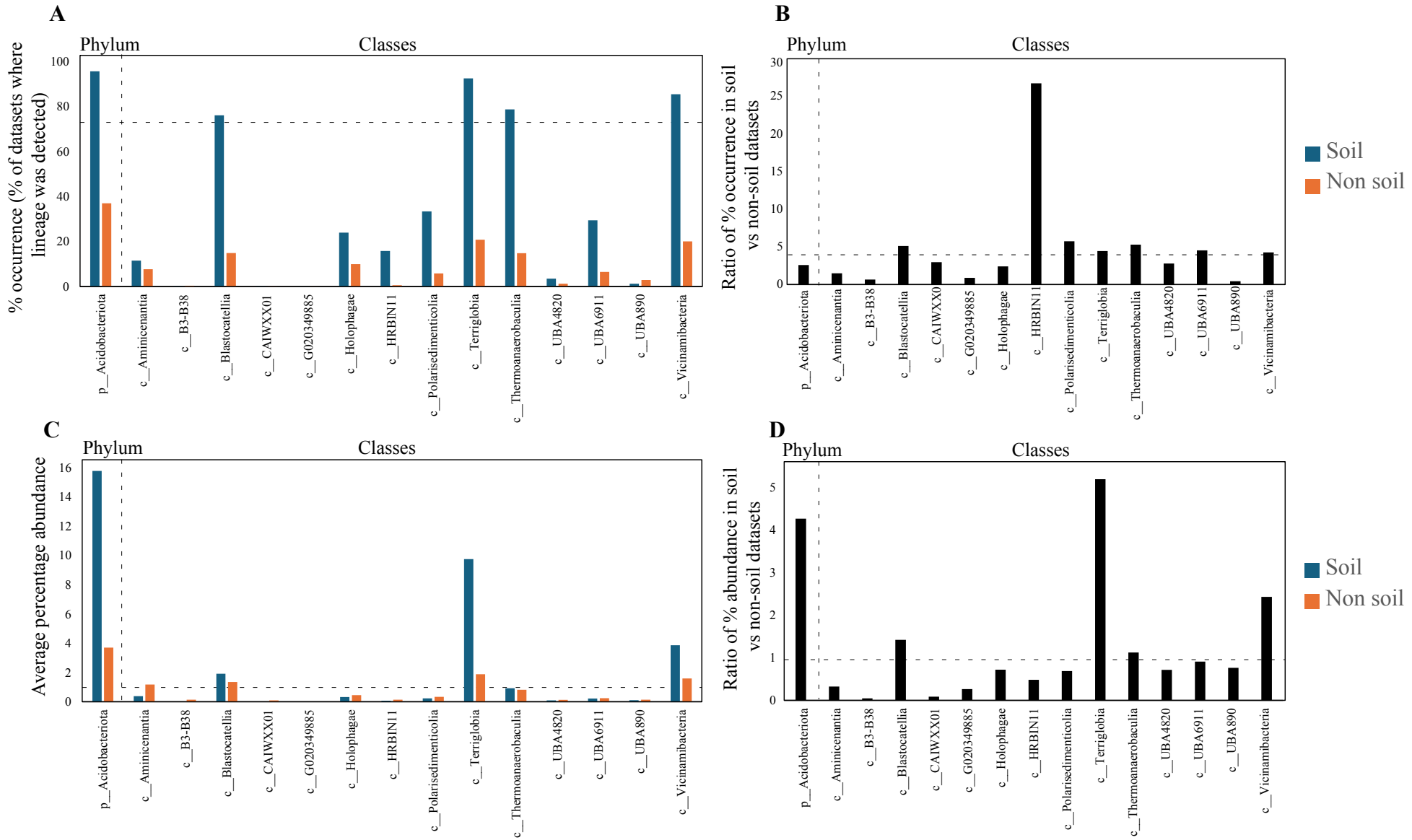
682 1. Oren A, Garrity GM. Valid publication of the names of forty-two phyla of prokaryotes.
683 *Int J Syst Evol Microbiol* 2021;**71**:005056. https://doi.org/10.1099/ijsem.0.005056
684 2. Barns SM, Cain EC, Sommerville L *et al.* Acidobacteria phylum sequences in uranium-
685 contaminated subsurface sediments greatly expand the known diversity within the
686 phylum. *Appl Environ Microbiol* 2007;**73**:3113-6. https://doi.org/10.1128/aem.02012-06
687 3. Barns SM, Takala SL, Kuske CR. Wide distribution and diversity of members of the
688 bacterial kingdom Acidobacterium in the environment. *Appl Environ Microbiol*
689 1999;**65**:1731-7. https://doi.org/10.1128/aem.65.4.1731-1737.1999
690 4. Crits-Christoph A, Diamond S, Al-Shayeb B *et al.* A widely distributed genus of soil
691 Acidobacteria genomically enriched in biosynthetic gene clusters. *ISME Communications*
692 2022;**2**:70. https://doi.org/10.1038/s43705-022-00140-5
693 5. Eichorst SA, Kuske CR, Schmidt TM. Influence of plant polymers on the distribution and
694 cultivation of bacteria in the phylum Acidobacteria. *Appl Environ Microbiol*
695 2011;**77**:586-96. https://doi.org/10.1128/aem.01080-10
696 6. Huber KJ, Geppert AM, Groß U *et al. Aridibacter nitratireducens* sp. nov., a member of
697 the family *Blastocatellaceae*, class *Blastocatellia*, isolated from an African soil. *Int J Syst*
698 *Evol Microbiol* 2017;**67**:4487-93. https://doi.org/10.1099/ijsem.0.002318
699 7. Huber KJ, Geppert AM, Wanner G *et al.* The first representative of the globally
700 widespread subdivision 6 Acidobacteria,*Vicinamibacter silvestris* gen. nov., sp. nov.,
701 isolated from subtropical savannah soil. *Int J Syst Evol Microbiol* 2016;**66**:2971-79.
702 https://doi.org/10.1099/ijsem.0.001131
703 8. Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S
704 rRNA genes. *Appl Environ Microbiol* 2006;**72**:1719-28.
705 https://doi.org/10.1128/aem.72.3.1719-1728.2006
706 9. Jones RT, Robeson MS, Lauber CL *et al.* A comprehensive survey of soil acidobacterial
707 diversity using pyrosequencing and clone library analyses. *ISME J* 2009;**3**:442-53.
708 https://doi.org/10.1038/ismej.2008.127
709 10. Joseph SJ, Hugenholtz P, Sangwan P *et al.* Laboratory cultivation of widespread and
710 previously uncultured soil bacteria. *Appl Environ Microbiol* 2003;**69**:7210-5.
711 https://doi.org/10.1128/aem.69.12.7210-7215.2003
712 11. Kuske CR, Barns SM, Busch JD. Diverse uncultivated bacterial groups from soils of the
713 arid southwestern United States that are present in many geographic regions. *Appl*
714 *Environ Microbiol* 1997;**63**:3614-21. https://doi.org/10.1128/aem.63.9.3614-3621.1997
715 12. Lauber CL, Hamady M, Knight R *et al.* Pyrosequencing-based assessment of soil pH as a
716 predictor of soil bacterial community structure at the continental scale. *Appl Environ*
717 *Microbiol* 2009;**75**:5111-20. https://doi.org/10.1128/aem.00335-09
718 13. Ludwig W, Bauer SH, Bauer M *et al.* Detection and in situ identification of
719 representatives of a widely distributed new bacterial phylum. *FEMS Microbiol Lett*
720 1997;**153**:181-90. https://doi.org/10.1111/j.1574-6968.1997.tb10480.x
721 14. Rawat SR, Männistö MK, Bromberg Y *et al.* Comparative genomic and physiological
722 analysis provides insights into the role of Acidobacteria in organic carbon utilization in
723 Arctic tundra soils. *FEMS Microbiol Ecol* 2012;**82**:341-55.
724 https://doi.org/10.1111/j.1574-6941.2012.01381.x

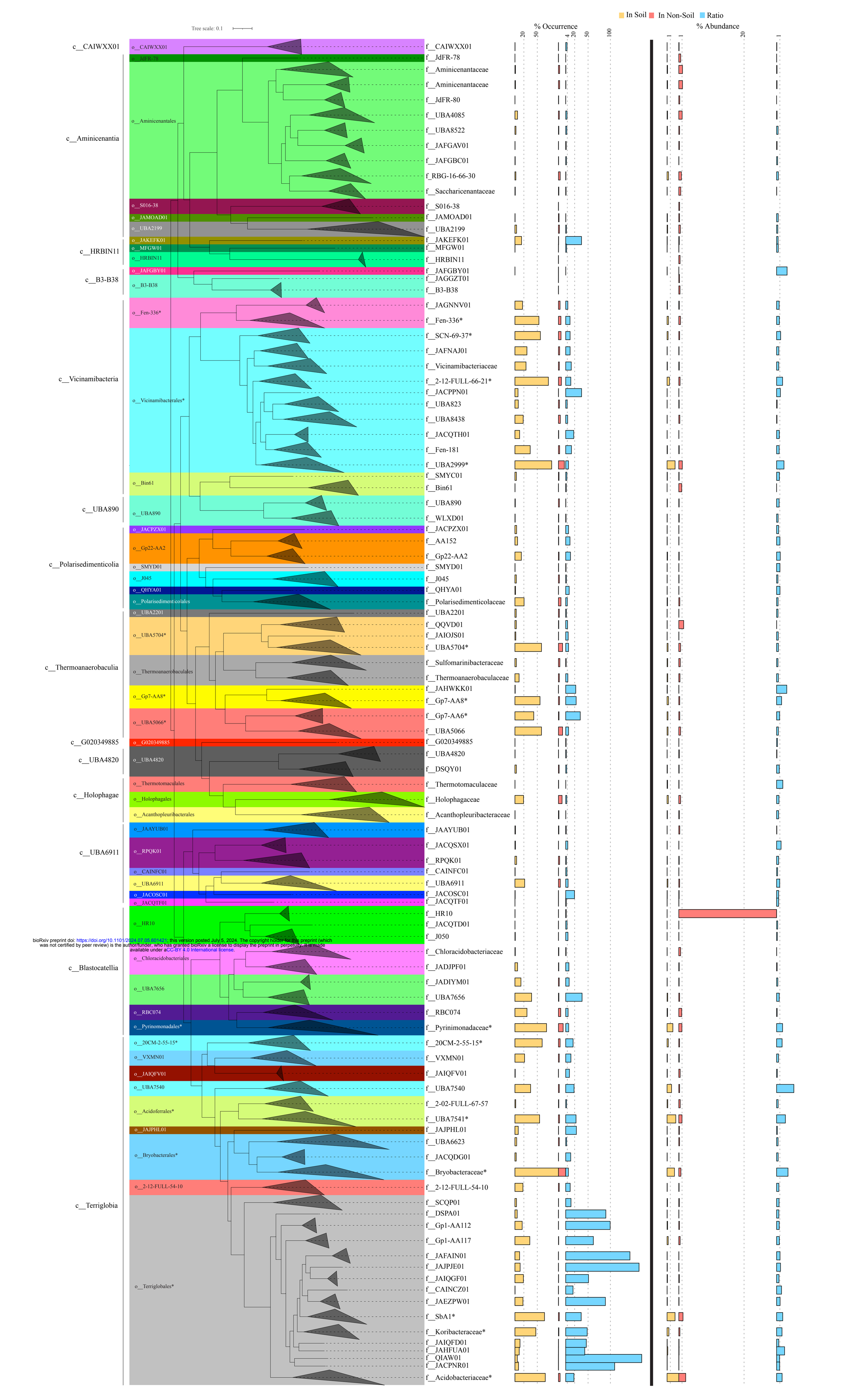
15. Sait M, Hugenholtz P, Janssen PH. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ Microbiol* 2002;**4**:654-66. https://doi.org/10.1046/j.1462-2920.2002.00352.x
16. Izumi H, Nunoura T, Miyazaki M *et al. Thermotomaculum hydrothermale* gen. nov., sp. nov., a novel heterotrophic thermophile within the phylum Acidobacteria from a deep-sea hydrothermal vent chimney in the Southern Okinawa Trough. *Extremophiles* 2012;**16**:245-53. https://doi.org/10.1007/s00792-011-0425-9
17. Liesack W, Bak F, Kreft JU *et al. Holophaga foetida* gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. *Arch Microbiol* 1994;**162**:85-90. https://doi.org/10.1007/bf00264378
18. Coates JD, Ellis DJ, Gaw CV *et al. Geothrix fermentans* gen. nov., sp. nov., a novel Fe(III)-reducing bacterium from a hydrocarbon-contaminated aquifer. *Int J Syst Bacteriol* 1999;**49 Pt 4**:1615-22. https://doi.org/10.1099/00207713-49-4-1615
19. Fukunaga Y, Kurahashi M, Yanagi K *et al. Acanthopleuribacter pedis* gen. nov., sp. nov., a marine bacterium isolated from a chiton, and description of *Acanthopleuribacteraceae* fam. nov., *Acanthopleuribacterales* ord. nov., *Holophagaceae* fam. nov., *Holophagales* ord. nov. and *Holophagae* classis nov. in the phylum 'Acidobacteria'. *Int J Syst Evol Microbiol* 2008;**58**:2597-601. https://doi.org/10.1099/ijs.0.65589-0
20. Bryant DA, Costas AM, Maresca JA *et al. Candidatus* Chloracidobacterium thermophilum: an aerobic phototrophic Acidobacterium. *Science* 2007;**317**:523-6. https://doi.org/10.1126/science.1143236
21. Oberpaul M, Zumkeller CM, Culver T *et al.* High-throughput cultivation for the selective isolation of Acidobacteria from termite nests. *Front Microbiol* 2020;**11**:597628. https://doi.org/10.3389/fmicb.2020.597628
22. Flieder M, Buongiorno J, Herbold CW *et al.* Novel taxa of Acidobacteriota implicated in seafloor sulfur cycling. *ISME J* 2021;**15**:3159-80. https://doi.org/10.1038/s41396-021-00992-0
23. Dedysh SN, Yilmaz P. Refining the taxonomic structure of the phylum Acidobacteria. *Int J Syst Evol Microbiol* 2018;**68**:3796-806. https://doi.org/10.1099/ijsem.0.003062
24. Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:D590-6. https://doi.org/10.1093/nar/gks1219
25. Parks DH, Chuvochina M, Rinke C *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;**50**:D785-d94. https://doi.org/10.1093/nar/gkab776
26. Eichorst SA, Breznak JA, Schmidt TM. Isolation and characterization of soil bacteria that define *Terriglobus* gen. nov., in the phylum Acidobacteria. *Appl Environ Microbiol* 2007;**73**:2708-17. https://doi.org/10.1128/aem.02140-06
27. Kishimoto N, Kosako Y, Tano T. *Acidobacterium capsulatum* gen. nov., sp. nov.: An acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. *Curr Microbiol* 1991;**22**:1-7. https://doi.org/10.1007/BF02106205
28. Koch IH, Gich F, Dunfield PF *et al. Edaphobacter modestus* gen. nov., sp. nov., and *Edaphobacter aggregans* sp. nov., acidobacteria isolated from alpine and forest soils. *Int J Syst Evol Microbiol* 2008;**58**:1114-22. https://doi.org/10.1099/ijs.0.65303-0

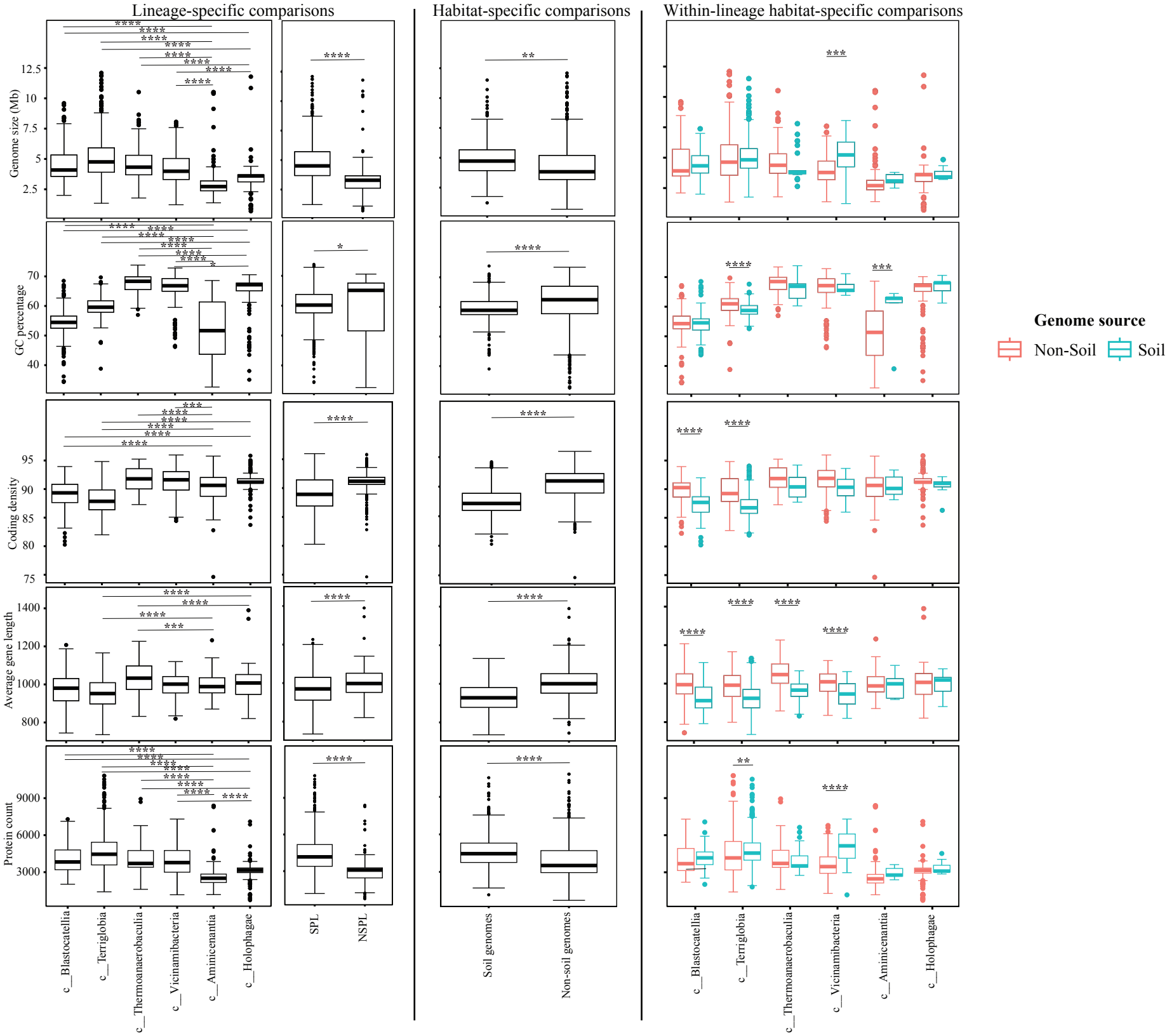
29. Kulichevskaya IS, Suzina NE, Liesack W *et al. Bryobacter aggregatus* gen. nov., sp. nov., a peat-inhabiting, aerobic chemo-organotroph from subdivision 3 of the Acidobacteria. *Int J Syst Evol Microbiol* 2010;**60**:301-06. https://doi.org/10.1099/ijs.0.013250-0
30. Losey NA, Stevenson BS, Busse HJ *et al. Thermoanaerobaculum aquaticum* gen. nov., sp. nov., the first cultivated member of Acidobacteria subdivision 23, isolated from a hot spring. *Int J Syst Evol Microbiol* 2013;**63**:4149-57. https://doi.org/10.1099/ijs.0.051425-0
31. Yadav A, Borrelli JC, Elshahed MS *et al.* Genomic analysis of family UBA6911 (group 18 acidobacteria) expands the metabolic capacities of the phylum and highlights adaptations to terrestrial habitats. *Appl Environ Microbiol* 2021;**87**:e0094721. https://doi.org/10.1128/aem.00947-21
32. Woodcroft BJ, Aroney STN, Zhao R *et al.* SingleM and Sandpiper: Robust microbial taxonomic profiles from metagenomic data. *bioRxiv* 2024:2024.01.30.578060. https://doi.org/10.1101/2024.01.30.578060
33. Ivanova N, Tringe SG, Liolios K *et al.* A call for standardized classification of metagenome projects. *Environ Microbiol* 2010;**12**:1803-5. https://doi.org/10.1111/j.1462-2920.2010.02270.x
34. Russel J, Pinilla-Redondo R, Mayo-Muñoz D *et al.* CRISPRCasTyper: Automated identification, annotation, and classification of CRISPR-Cas loci. *CRISPR J* 2020;**3**:462-69. https://doi.org/10.1089/crispr.2020.0059
35. Guo J, Bolduc B, Zayed AA *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 2021;**9**:37. https://doi.org/10.1186/s40168-020-00990-y
36. Zheng J, Ge Q, Yan Y *et al.* dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucl Acids Res* 2023;**51**:W115-W21. https://doi.org/10.1093/nar/gkad328
37. Blin K, Shaw S, Augustijn HE *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucl Acids Res* 2023;**51**:W46-W50. https://doi.org/10.1093/nar/gkad344
38. Rawlings ND, Barrett AJ, Thomas PD *et al.* The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucl Acids Res* 2017;**46**:D624-D32. https://doi.org/10.1093/nar/gkx1134
39. Li G, Rabe KS, Nielsen J *et al.* Machine learning applied to predicting microorganism growth temperatures and enzyme catalytic optima. *ACS Synth Biol* 2019;**8**:1411-20. https://doi.org/10.1021/acssynbio.9b00099
40. Ramoneda J, Stallard-Olivera E, Hoffert M *et al.* Building a genome-based understanding of bacterial pH preferences. *Sci Adv* 2023;**9**:eadf8998. https://doi.org/doi:10.1126/sciadv.adf8998
41. Davín AA, Woodcroft BJ, Soo RM *et al.* An evolutionary timescale for Bacteria calibrated using the Great Oxidation Event. *bioRxiv* 2023:2023.08.08.552427. https://doi.org/10.1101/2023.08.08.552427
42. Barnett SE, Egan R, Foster B *et al.* Genomic features predict bacterial life history strategies in soil, as identified by metagenomic stable isotope probing. *mBio* 2023;**14**:e0358422. https://doi.org/10.1128/mbio.03584-22

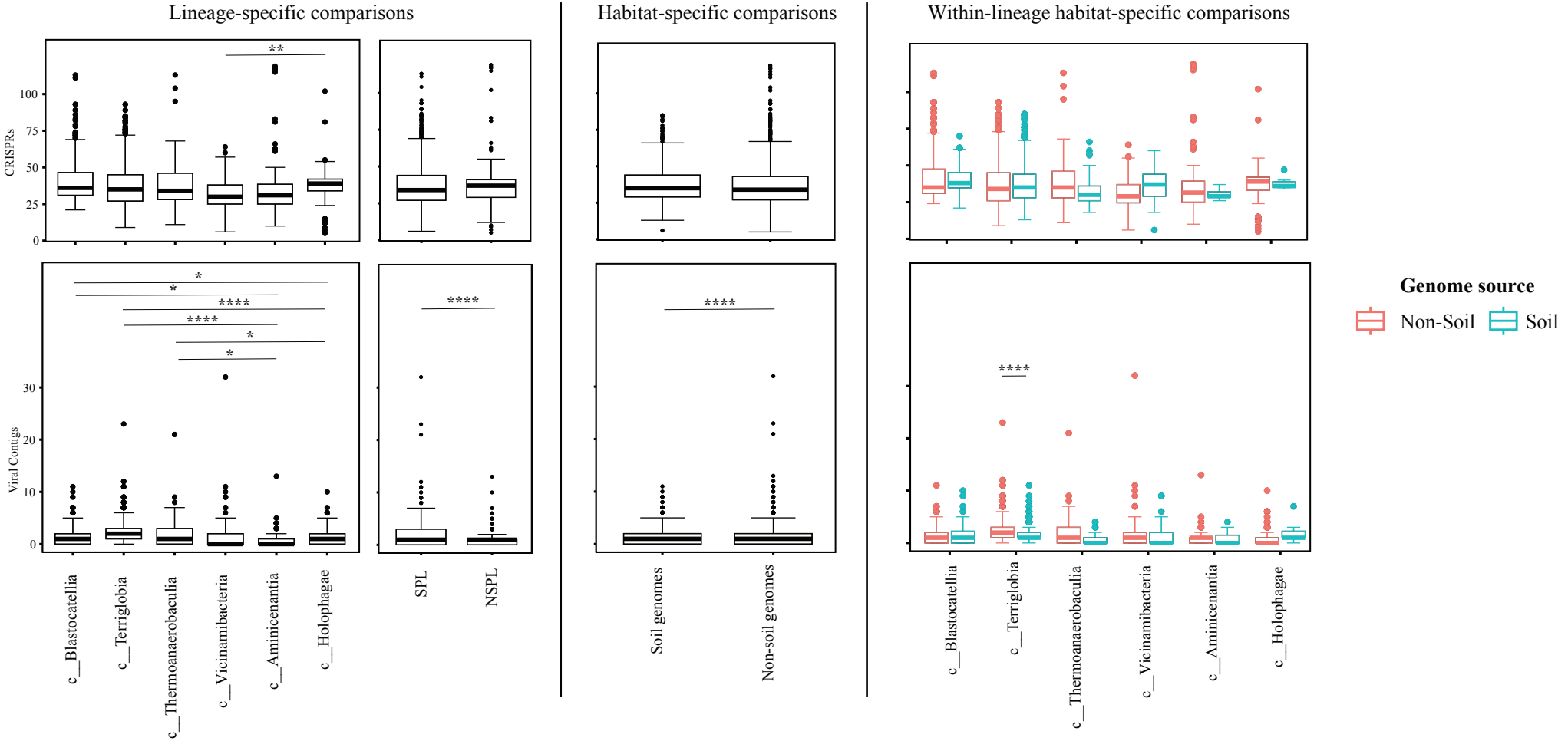
43. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;**428**:726-31. https://doi.org/10.1016/j.jmb.2015.11.006
44. Kim GB, Gao Y, Palsson BO *et al.* DeepTFactor: A deep learning-based tool for the prediction of transcription factors. *Proc Natl Acad Sci USA* 2021;**118**:e2021171118. https://doi.org/doi:10.1073/pnas.2021171118
45. Lenfant N, Hotelier T, Velluet E *et al.* ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucl Acids Res* 2013;**41**:D423-9. https://doi.org/10.1093/nar/gks1154
46. Teufel F, Almagro Armenteros JJ, Johansen AR *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 2022;**40**:1023-25. https://doi.org/10.1038/s41587-021-01156-3
47. Leisch F. A toolbox for K-centroids cluster analysis. *Comput Stat Data Anal* 2006;**51**:526-44. https://doi.org/https://doi.org/10.1016/j.csda.2005.10.006
48. Zhou Z, Tran PQ, Breister AM *et al.* METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* 2022;**10**:33. https://doi.org/10.1186/s40168-021-01213-8
49. Greening C, Carere CR, Rushton-Green R *et al.* Persistence of the dominant soil phylum Acidobacteria by trace gas scavenging. *Proc Natl Acad Sci USA* 2015;**112**:10497-502. https://doi.org/10.1073/pnas.1508385112
50. Eichorst SA, Trojan D, Roux S *et al.* Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. *Environ Microbiol* 2018;**20**:1041-63. https://doi.org/10.1111/1462-2920.14043
51. Chen S, White CE, diCenzo GC *et al.* L-hydroxyproline and D-proline catabolism in *Sinorhizobium meliloti*. *J Bacteriol* 2016;**198**:1171-81. https://doi.org/10.1128/jb.00961-15
52. Warren CR. Response of osmolytes in soil to drying and rewetting. *Soil Biol Biochem* 2014;**70**:22-32. https://doi.org/https://doi.org/10.1016/j.soilbio.2013.12.008
53. Warren CR. Do microbial osmolytes or extracellular depolymerisation products accumulate as soil dries? *Soil Biol Biochem* 2016;**98**:54-63. https://doi.org/https://doi.org/10.1016/j.soilbio.2016.03.021
54. Warren CR. Pools and fluxes of osmolytes in moist soil and dry soil that has been re-wet. *Soil Biol Biochem* 2020;**150**:108012. https://doi.org/https://doi.org/10.1016/j.soilbio.2020.108012
55. Warren CR, Manzoni S. When dry soil is re-wet, trehalose is respired instead of supporting microbial growth. *Soil Biol Biochem* 2023;**184**:109121. https://doi.org/https://doi.org/10.1016/j.soilbio.2023.109121
56. Michael AJ. Polyamine function in archaea and bacteria. *J Biol Chem* 2018;**293**:18693-701. https://doi.org/10.1074/jbc.TM118.005670
57. Solmi L, Rossi FR, Romero FM *et al.* Polyamine-mediated mechanisms contribute to oxidative stress tolerance in *Pseudomonas syringae*. *Sci Rep* 2023;**13**:4279. https://doi.org/10.1038/s41598-023-31239-x
58. Xu Q, Vandenkoornhuyse P, Li L *et al.* Microbial generalists and specialists differently contribute to the community diversity in farmland soils. *J Adv Res* 2022;**40**:17-27. https://doi.org/10.1016/j.jare.2021.12.003

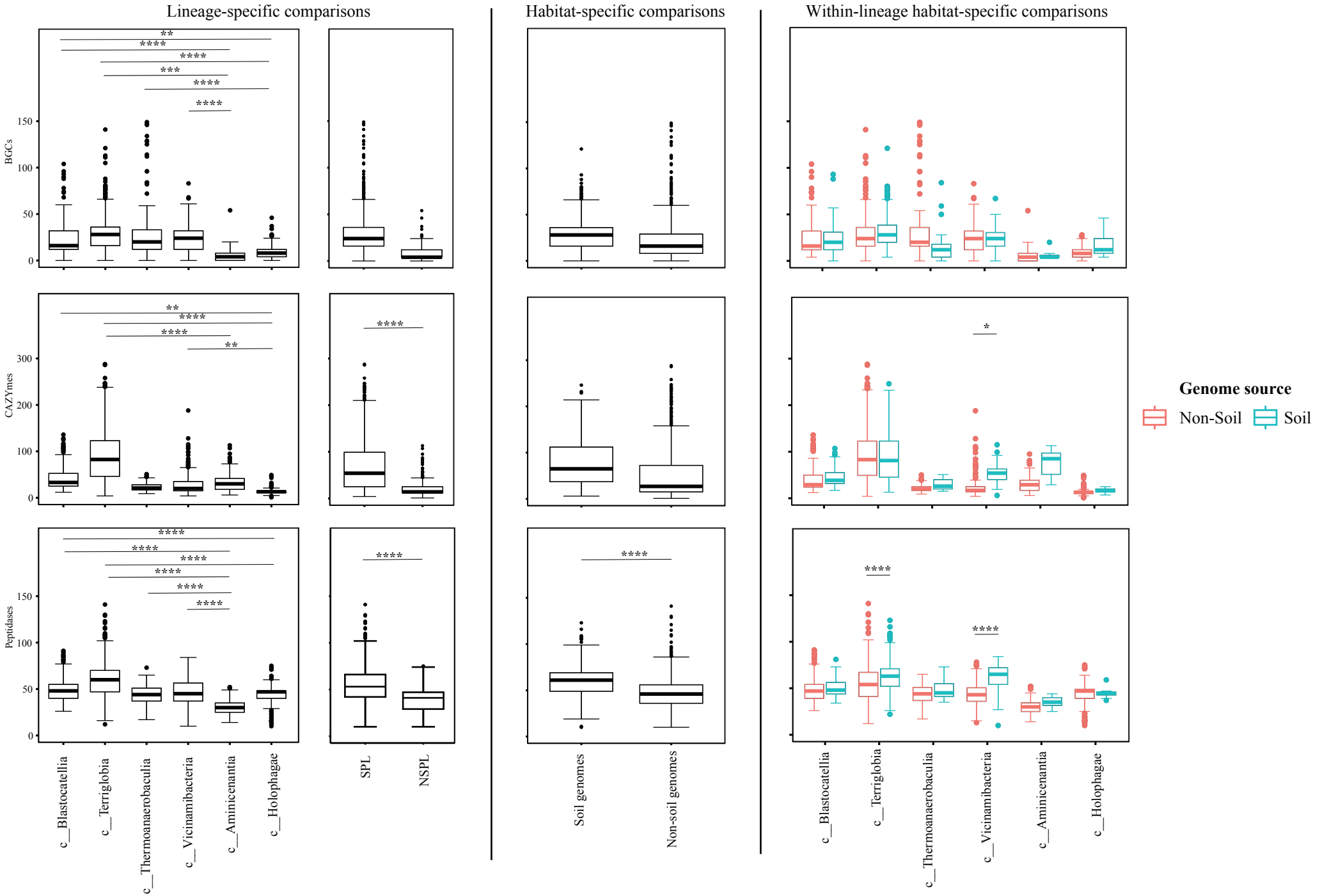
59. von Meijenfeldt FAB, Hogeweg P, Dutilh BE. A social niche breadth score reveals niche range strategies of generalists and specialists. *Nat Ecol Evol* 2023;**7**:768-81. https://doi.org/10.1038/s41559-023-02027-7
60. Zeng X, Alain K, Shao Z. Microorganisms from deep-sea hydrothermal vents. *Mar Life Sci Technol* 2021;**3**:204-30. https://doi.org/10.1007/s42995-020-00086-4
61. Oren A. Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol Ecol* 2002;**39**:1-7. https://doi.org/10.1111/j.1574-6941.2002.tb00900.x
62. Gruninger RJ, Puniya AK, Callaghan TM *et al.* Anaerobic fungi (phylum Neocallimastigomycota): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiol Ecol* 2014;**90**:1-17. https://doi.org/10.1111/1574-6941.12383
63. Bell TH, Bell T. Many roads to bacterial generalism. *FEMS Microbiol Ecol* 2020;**97** https://doi.org/10.1093/femsec/fiaa240
64. Koehorst JJ, van Dam JCJ, van Heck RGA *et al.* Comparison of 432 *Pseudomonas* strains through integration of genomic, functional, metabolic and expression data. *Sci Rep* 2016;**6**:38699. https://doi.org/10.1038/srep38699
65. Compant S, Nowak J, Coenye T *et al.* Diversity and occurrence of *Burkholderia* spp. in the natural environment. *FEMS Microbiol Rev* 2008;**32**:607-26. https://doi.org/10.1111/j.1574-6976.2008.00113.x
66. Lian Y, Yang Y, Guo J *et al.* Electron acceptor redox potential globally regulates transcriptomic profiling in *Shewanella decolorationis* S12. *Sci Rep* 2016;**6**:31143. https://doi.org/10.1038/srep31143
67. Anderson I, Held B, Lapidus A *et al.* Genome sequence of the homoacetogenic bacterium *Holophaga foetida* type strain (TMBS4(T)). *Stand Genomic Sci* 2012;**6**:174-84. https://doi.org/10.4056/sigs.2746047
68. Ward NL, Challacombe JF, Janssen PH *et al.* Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol* 2009;**75**:2046-56. https://doi.org/10.1128/aem.02294-08
69. Reji L, Zhang X. Genome-resolved metagenomics informs the functional ecology of uncultured Acidobacteria in redox oscillated sphagnum peat. *mSystems* 2022;**7**:e0005522. https://doi.org/10.1128/msystems.00055-22
70. Xu R, Sun X, Häggblom MM *et al.* Metabolic potentials of members of the class Acidobacteriia in metal-contaminated soils revealed by metagenomic analysis. *Environ Microbiol* 2022;**24**:803-18. https://doi.org/10.1111/1462-2920.15612
71. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* 2010;**5**:e9490. https://doi.org/10.1371/journal.pone.0009490

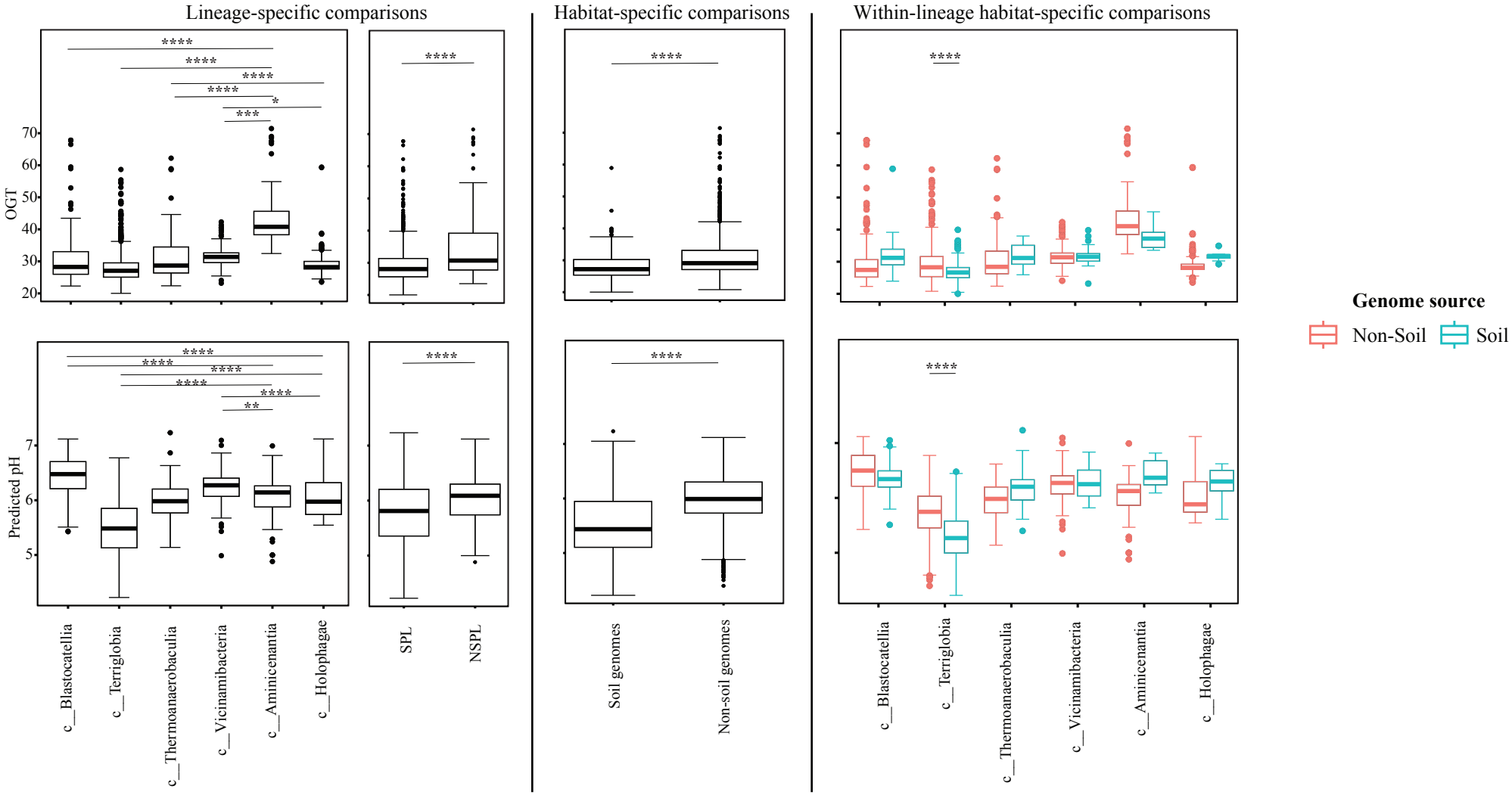

A
Phylum
Classes
% occurrence (% of datasets where lineage was detected)
100
80
60
40
20
0
p__Acidobacteriota
c__Aminicenantia
c__B3-B38
c__Blastocatellia
c__CAIWXX01
c__G020349885
c__Holophagae
c__HRBIN11
c__Polarisedimenticolia
c__Terriglobia
c__Thermoanaerobaculia
c__UBA4820
c__UBA6911
c__UBA890
c__Vicinamibacteria
B
Phylum
Classes
Ratio of % occurrence in soil vs non-soil datasets
30
25
20
15
10
5
0
p__Acidobacteriota
c__Aminicenantia
c__B3-B38
c__Blastocatellia
c__CAIWXX0
c__G020349885
c__Holophagae
c__HRBIN11
c__Polarisedimenticolia
c__Terriglobia
c__Thermoanaerobaculia
c__UBA4820
c__UBA6911
c__UBA890
c__Vicinamibacteria
Soil
Non soil
C
Phylum
Classes
Average percentage abundance
16
14
12
10
8
6
4
2
0
p__Acidobacteriota
c__Aminicenantia
c__B3-B38
c__Blastocatellia
c__CAIWXX01
c__G020349885
c__Holophagae
c__HRBIN11
c__Polarisedimenticolia
c__Terriglobia
c__Thermoanaerobaculia
c__UBA4820
c__UBA6911
c__UBA890
c__Vicinamibacteria
D
Phylum
Classes
Ratio of % abundance in soil vs non-soil datasets
5
4
3
2
1
0
p__Acidobacteriota
c__Aminicenantia
c__B3-B38
c__Blastocatellia
c__CAIWXX01
c__G020349885
c__Holophagae
c__HRBIN11
c__Polarisedimenticolia
c__Terriglobia
c__Thermoanaerobaculia
c__UBA4820
c__UBA6911
c__UBA890
c__Vicinamibacteria
Soil
Non soil

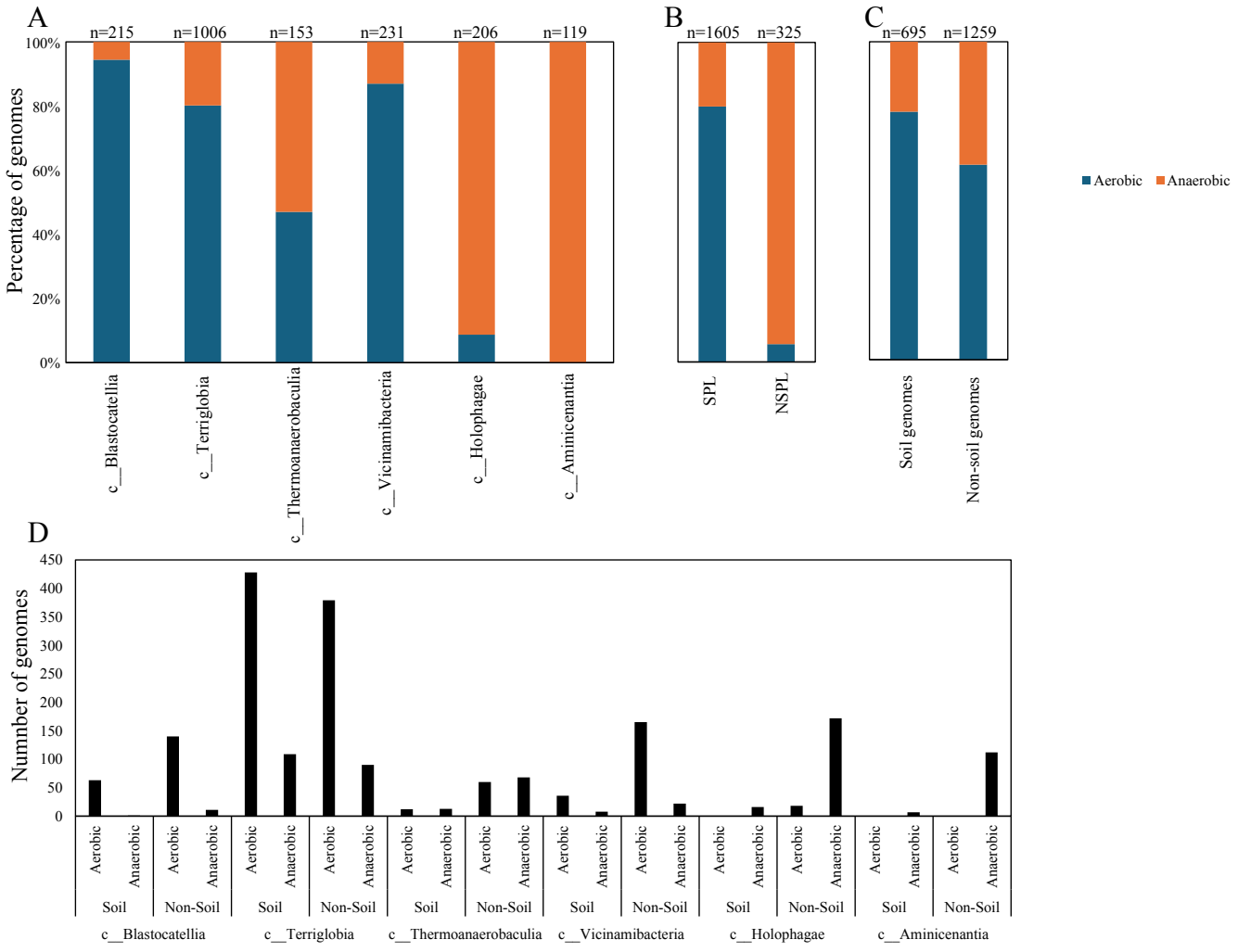


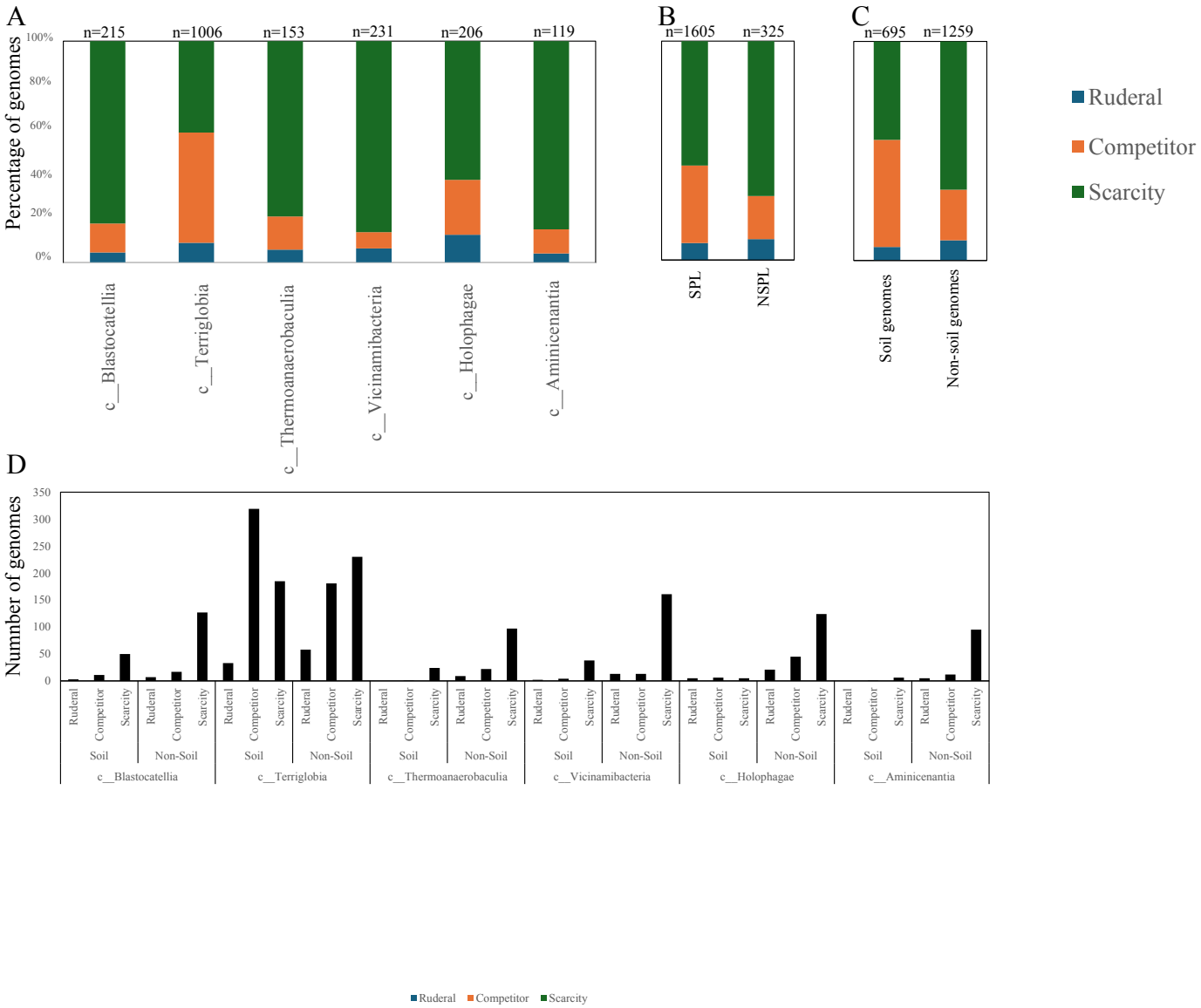

Lineage-specific comparisons
Habitat-specific comparisons
Within-lineage habitat-specific comparisons
Genome size (Mb)
GC percentage
Coding density
Average gene length
Protein count
c__Blastocatellia
c__Terriglobia
c__Thermoanaerobaculia
c__Vicinamibacteria
c__Aminicenantia
c__Holophagae
SPL
NSPL
Soil genomes
Non-soil genomes
Genome source
Non-Soil
Soil


Lineage-specific comparisons
Habitat-specific comparisons
Within-lineage habitat-specific comparisons
CRISPRs
Viral Contigs
**
*
*
****
****
*
*
****
****
****
Genome source
Non-Soil
Soil
c__Blastocatellia
c__Terriglobia
c__Thermoanaerobaculia
c__Vicinamibacteria
c__Aminicenantia
c__Holophagae
SPL
NSPL
Soil genomes
Non-soil genomes


Lineage-specific comparisons
Habitat-specific comparisons
Within-lineage habitat-specific comparisons
BGCs
CAZYmes
Peptidases
Genome source
Non-Soil
Soil
c__Blastocatellia
c__Terriglobia
c__Thermoanaerobaculia
c__Vicinamibacteria
c__Aminicenantia
c__Holophagae
SPL
NSPL
Soil genomes
Non-soil genomes


Lineage-specific comparisons
Habitat-specific comparisons
Within-lineage habitat-specific comparisons
OGT
Predicted pH
****
*
***
**
c__Blastocatellia
c__Terriglobia
c__Thermoanaerobaculia
c__Vicinamibacteria
c__Aminicenantia
c__Holophagae
SPL
NSPL
Soil genomes
Non-soil genomes
Genome source
Non-Soil
Soil


A
n=215
n=1006
n=153
n=231
n=206
n=119
100%
80%
60%
40%
20%
0%
Percentage of genomes
c__Blastocatellia
c__Terriglobia
c__Thermoanaerobaculia
c__Vicinamibacteria
c__Holophagae
c__Aminicenantia
B
n=1605
n=325
SPL
NSPL
C
n=695
n=1259
Soil genomes
Non-soil genomes
Aerobic
Anaerobic
D
Numnber of genomes
450
400
350
300
250
200
150
100
50
0
Aerobic
Anaerobic
Soil
Non-Soil
c__Blastocatellia
c__Terriglobia
c__Thermoanaerobaculia
c__Vicinamibacteria
c__Holophagae
c__Aminicenantia


A
n=215
n=1006
n=153
n=231
n=206
n=119
Percentage of genomes
100%
80%
60%
40%
20%
0%
c__Blastocatellia
c__Terriglobia
c__Thermoanaerobaculia
c__Vicinamibacteria
c__Holophagae
c__Aminicenantia
B
n=1605
n=325
SPL
NSPL
C
n=695
n=1259
Soil genomes
Non-soil genomes
Ruderal
Competitor
Scarcity
D
Numnber of genomes
350
300
250
200
150
100
50
0
Ruderal
Competitor
Scarcity
Soil
Non-Soil
c__Blastocatellia
c__Terriglobia
c__Thermoanaerobaculia
c__Vicinamibacteria
c__Holophagae
c__Aminicenantia