Supplementary Information

Unraveling Plasmid Contributions to Phosphorus Acquisition in Soil Microbiomes

Pablo Bruna[1,2], Patricio Javier Barra[2,3], Matías García[2,3], Ivan Liachko[4], María de la Luz Mora[2], Bas E. Dutilh[5,6], Michel Abanto[2*]

[1] Doctoral Program in Science, Major in Applied Cellular and Molecular Biology, Universidad de La Frontera, Av. Francisco Salazar 01145, Temuco, Chile.

[2] Scientific and Technological Bioresource Nucleus (BIOREN), Universidad de La Frontera, Avenida Francisco Salazar, 01145, Temuco, Chile.

[3] Biocontrol Research Laboratory–Center of Plant, Soil Interaction and Natural Resources Biotechnology; BIOREN, Universidad de La Frontera, Avenida Francisco Salazar, 01145, Temuco, Chile.

[4] Phase Genomics, Seattle, WA, USA

[5] Cluster of Excellence Balance of the Microverse, Institute of Biodiversity, Ecology, and Evolution, Friedrich Schiller University Jena; 07745 Jena; Germany.

6 Theoretical Biology and Bioinformatics; Science for Life; Utrecht University; 3584 CH Utrecht; the Netherlands.

*Corresponding author: mfabanto@gmail.com

Figure S1. Pairwise average nucleotide identity (ANI) among 449 dereplicated soil-plasmids retrieved from the PLSDB. The heatmap shows the pairwise ANI comparisons among plasmid sequences, highlighting degrees of nucleotide similarity. The right side displays the sequence names of the plasmids. The color scale ranges from grey (<70% similarity) to red (100% similarity), indicating increasing sequence identity between plasmid pairs.

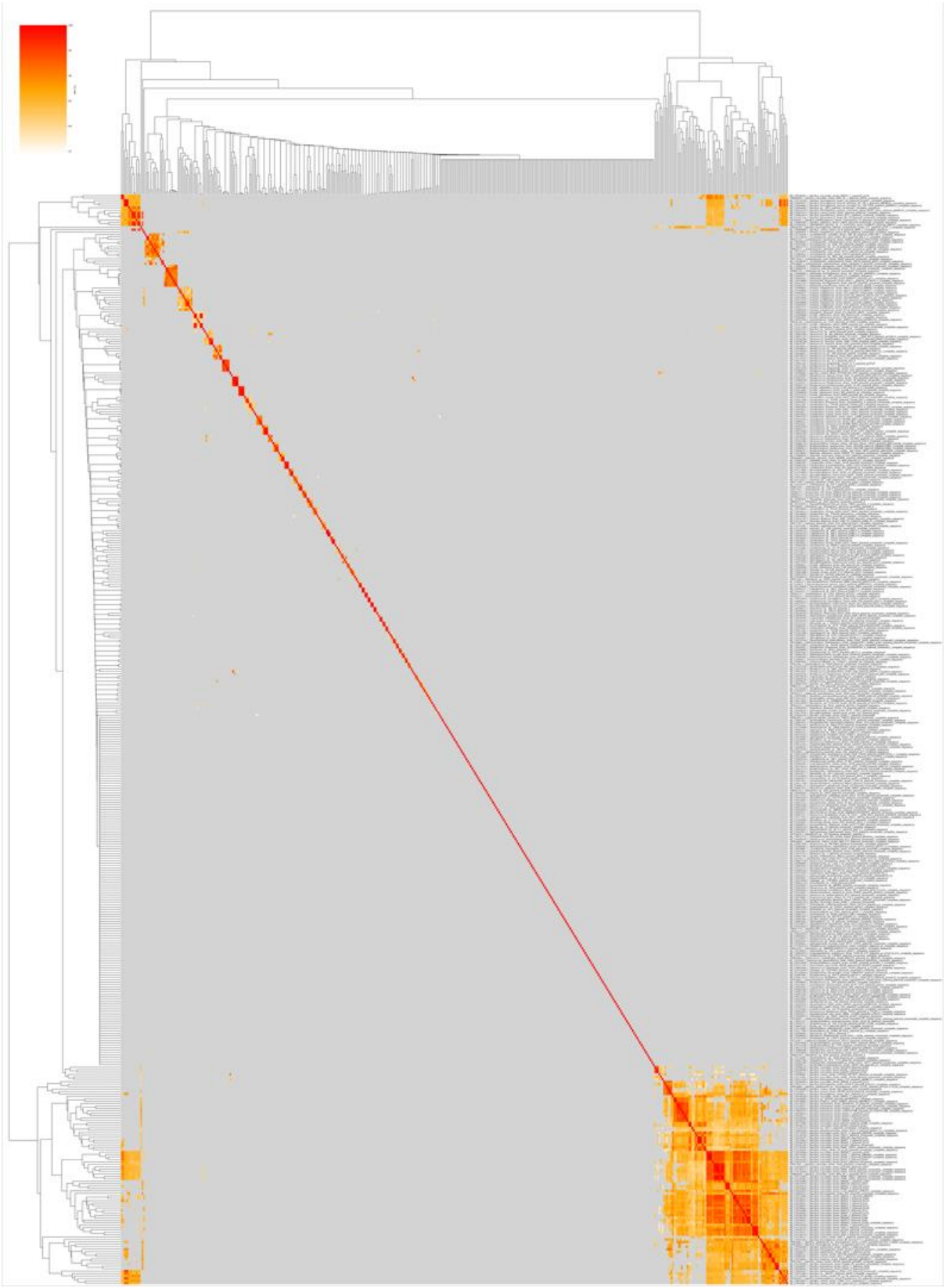

Figure S2. Dendrogram of the 449 dereplicated soil plasmids retrieved from the PLSDB. Hierarchical clustering based on pairwise ANI was used to construct the dendrogram, illustrating genomic relatedness among plasmid sequences. Different colors indicate the phylum-level taxonomy of the corresponding plasmid host.

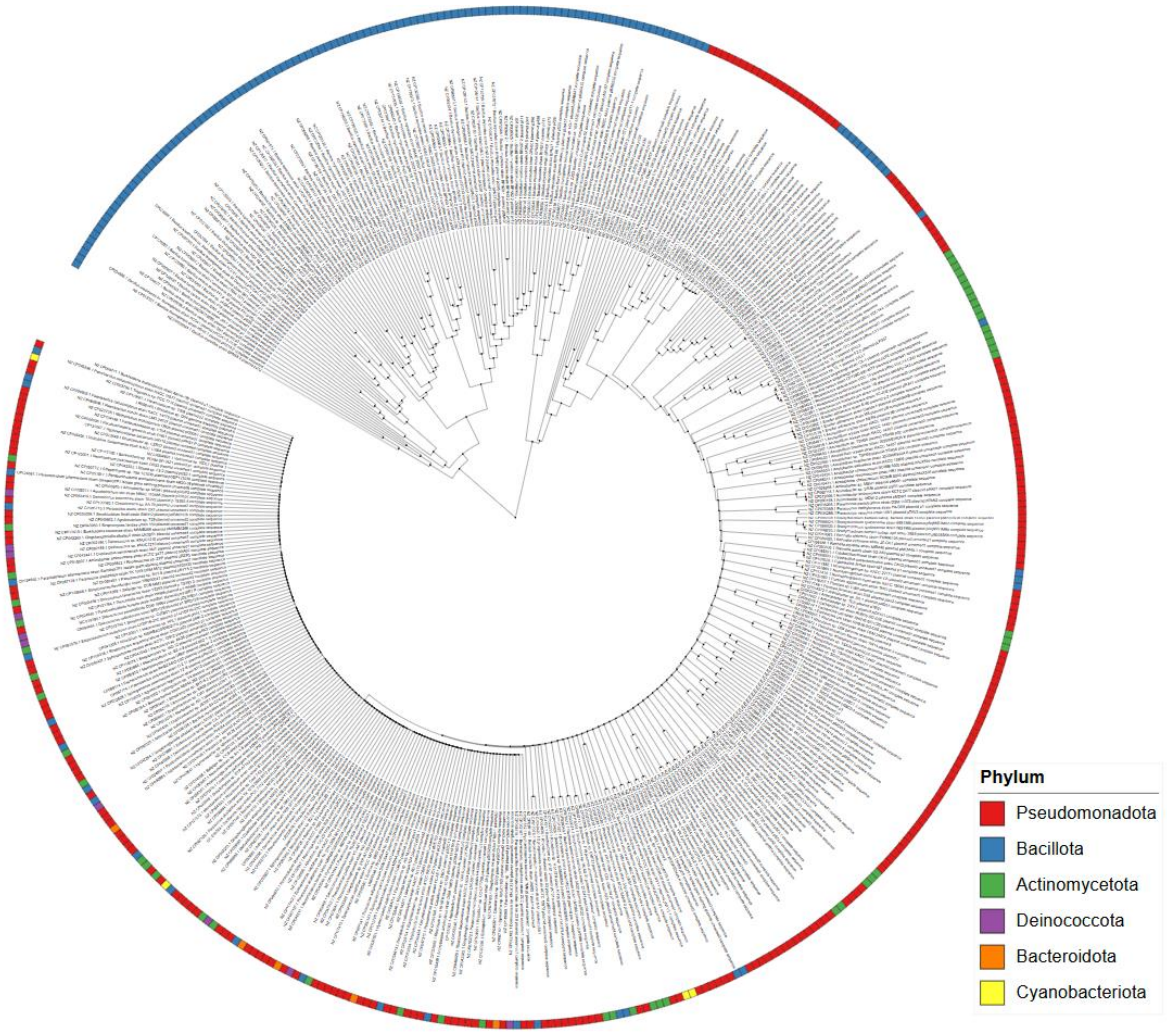

Figure S3. GC content of host chromosomes (black) and associated plasmids encoding phosphorus-acquisition genes from PLSDB (mobilizable, red; non-mobilizable, blue), per host genome. Per-host/plasmid metadata are provided in Table Supplementary 2.

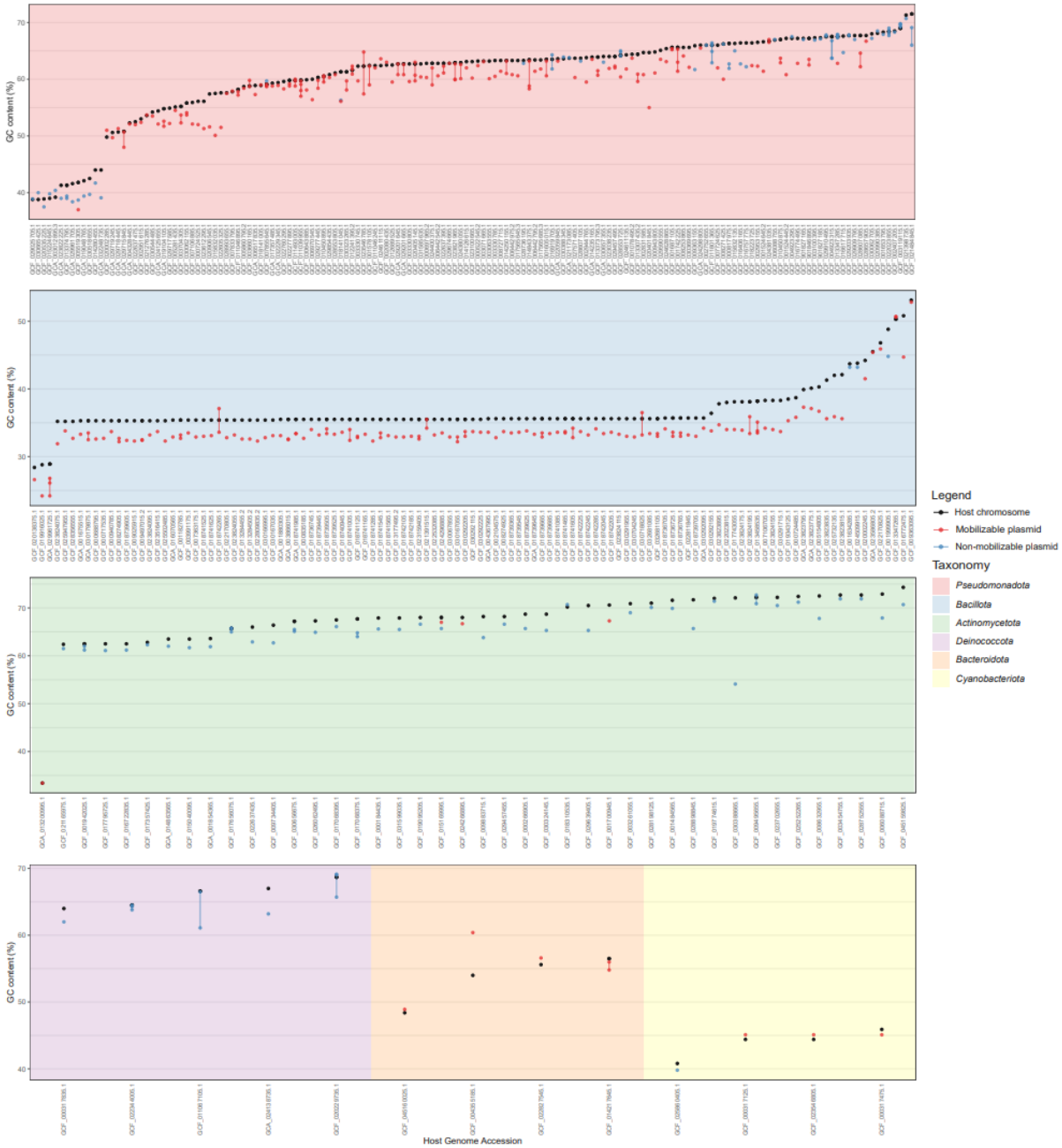

Figure S4. GC content of host chromosomes (grey) and associated plasmids encoding phosphorus-acquisition genes from PLSDB, shown as boxplots for the three most represented phyla: (A) *Pseudomonadota*, (B) *Bacillota*, and (C) *Actinomycetota*. Individual points are hosts (chromosome) or plasmids, which are classified as mobilizable (red) or non-mobilizable (blue). Asterisks indicate statistical significance: * = $p < 0.1$, ** = $p < 0.01$, **** = $p < 0.0001$. Non-significant comparisons are omitted.

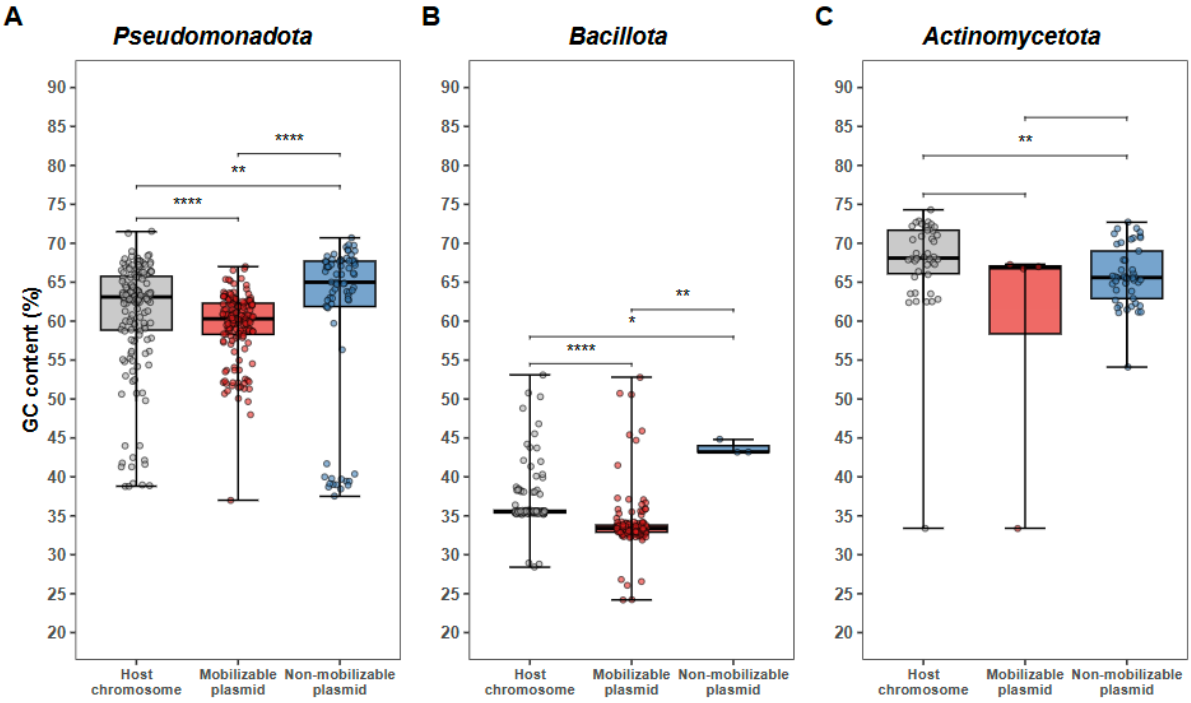

Figure S5. Relationship between the number of P-acquisition genes and plasmid length for sequences retrieved from PLSDB-soil database. Pearson and Spearman correlation analyses are shown for the 449 unique dereplicated plasmids.

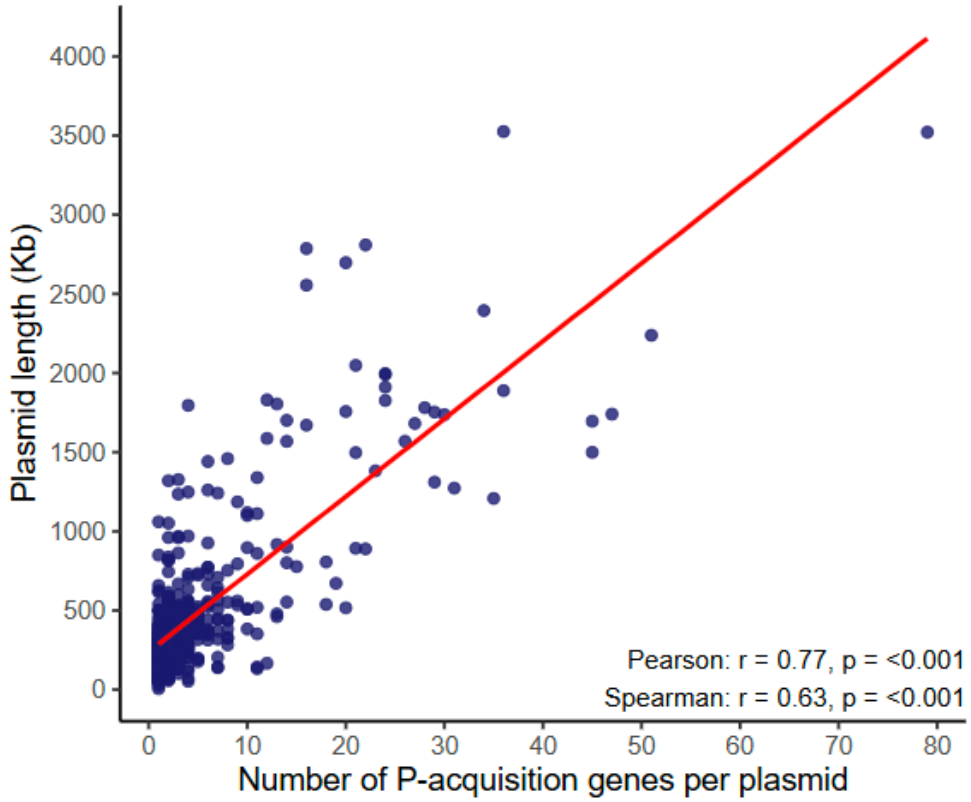

Figure S6. P-acquisition genes encoded in plasmids retrieved from the PLSDB-soil database. Bar plots show the number of P-acquisition genes identified in plasmids, grouped by P-acquisition functional categories. Bar colors indicate the presence or absence of mobility genes within each plasmid.

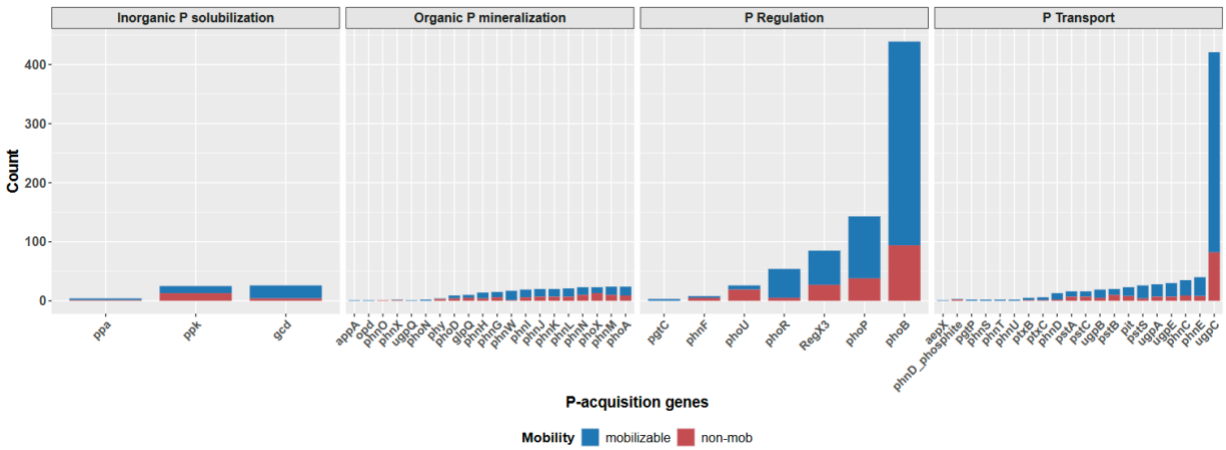

Figure S7. Pairwise average nucleotide identity (ANI) among 1,966 dereplicated plasmids identified from metagenomic assemblies. The heatmap shows the pairwise ANI comparisons among plasmid sequences, highlighting degrees of nucleotide similarity. The right side displays the sequence names of the plasmids. The color scale ranges from grey (<70% similarity) to red (100% similarity), indicating increasing sequence identity between plasmid pairs.

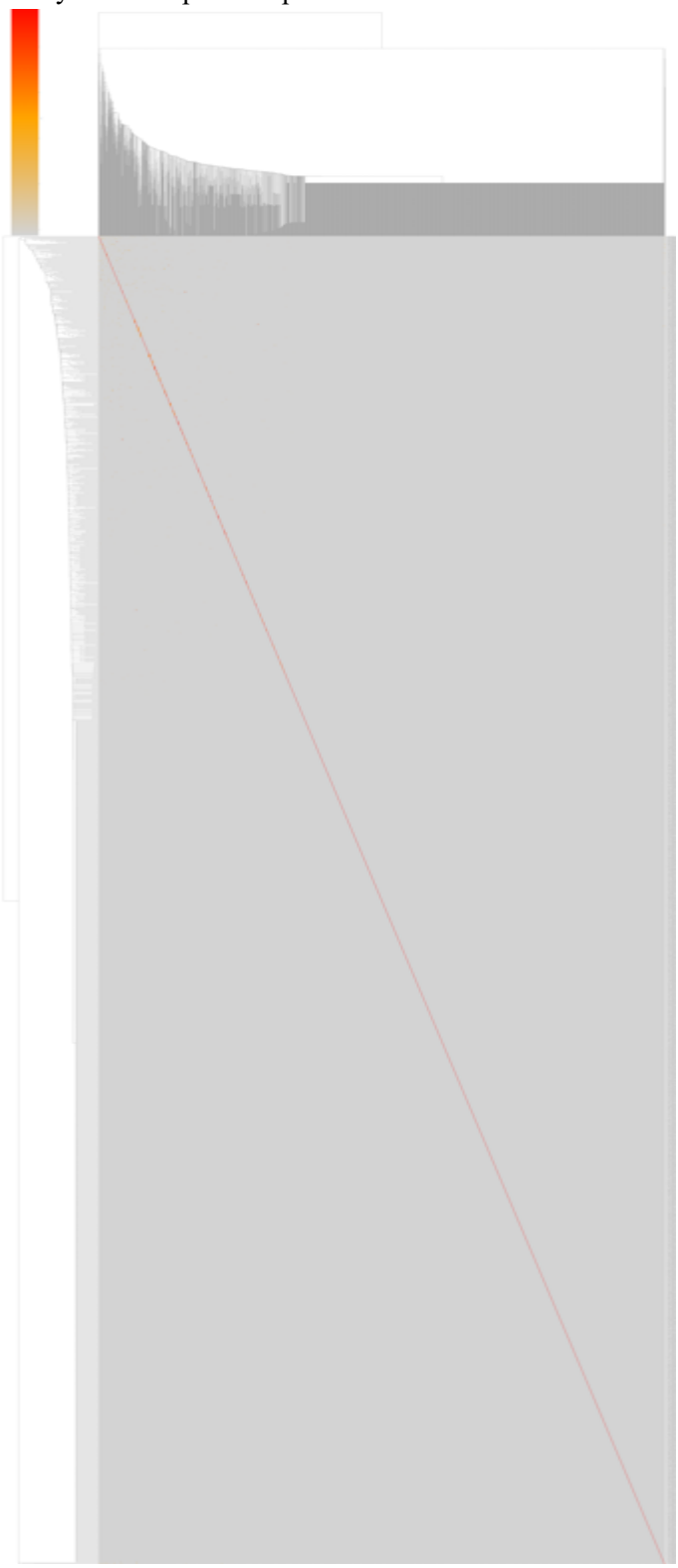

Figure S8. Sequence length of plasmids containing P-acquisition genes identified in soil metagenomes. Plasmids are grouped by mobility class, and their distribution is shown across different soil environments. Asterisks indicate statistical significance: ns = not significant, ** = p < 0.01, *** = p < 0.001.

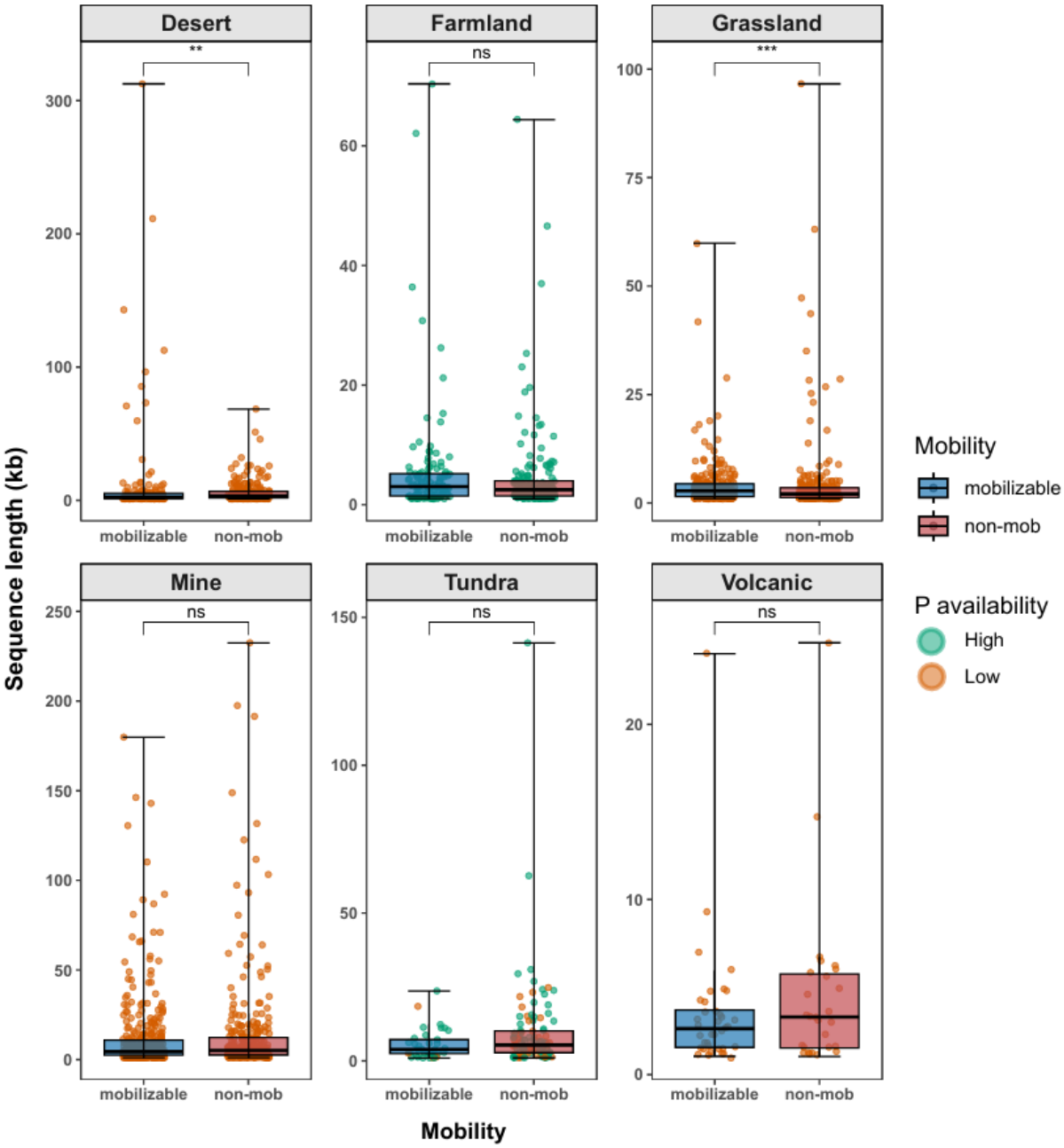

Figure S9. Relationship between the number of P-acquisition genes and plasmid length for sequences identified from metagenomic assemblies. Pearson and Spearman correlation analyses are shown for the 1,966 unique dereplicated plasmids.

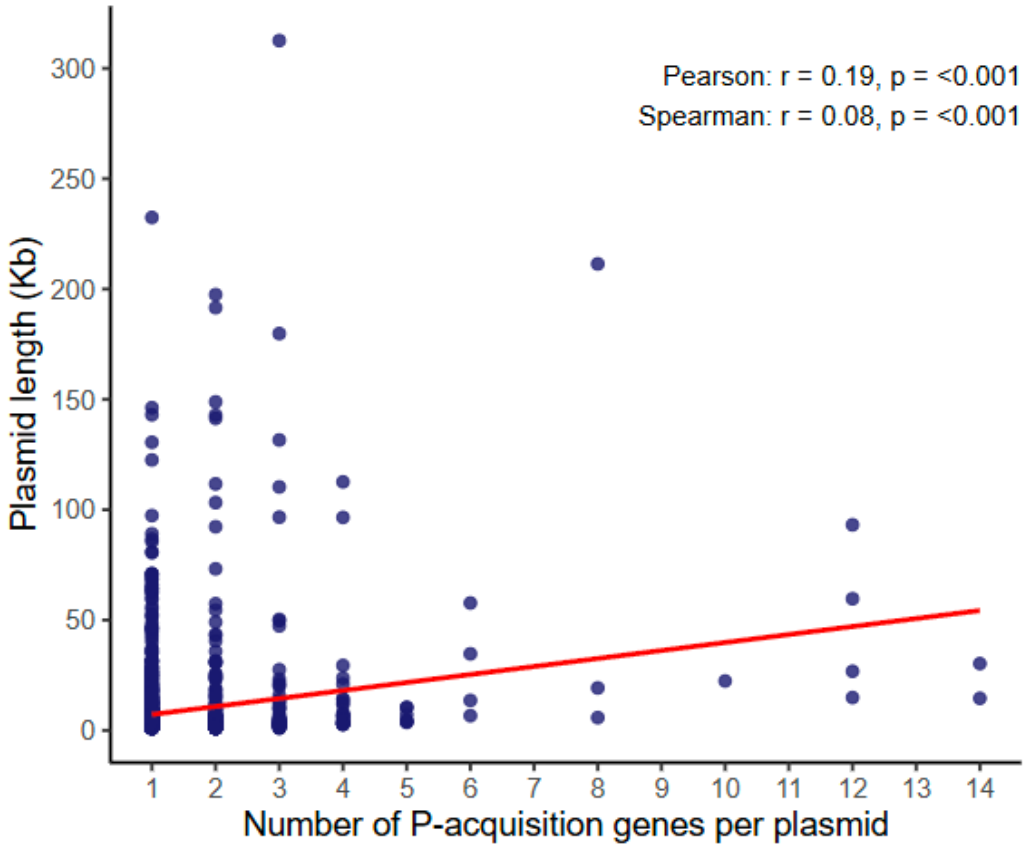

Figure S10. P-acquisition genes encoded in plasmids identified from metagenomic assemblies. Bar plots show the number of P-acquisition genes identified in plasmids, grouped by P-acquisition functional categories. Bar colors indicate the presence or absence of mobility genes within each plasmid.

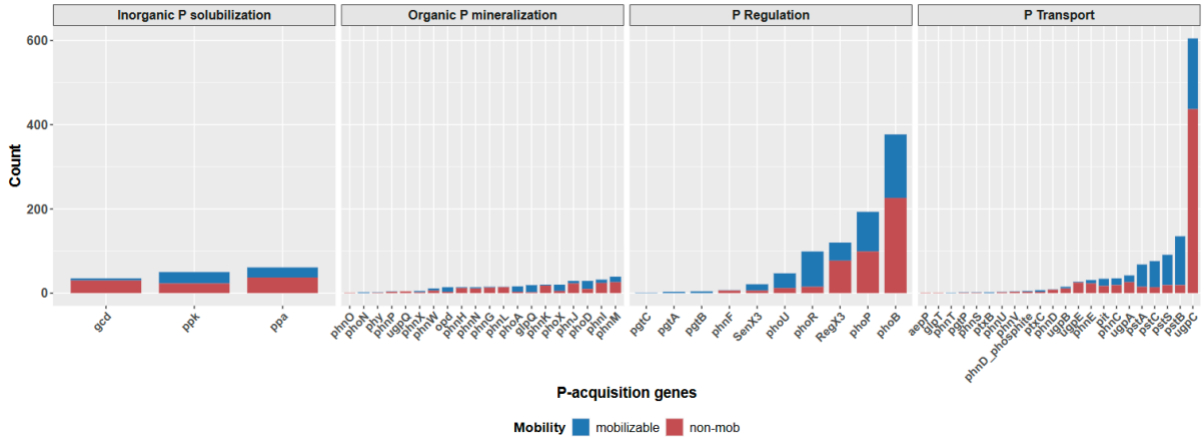

Figure S11. Heatmap of P-acquisition gene abundance in plasmids across soil environments with varying P availability. This heatmap presents untransformed gene abundance values across six soil types (desert, farmland, grassland, mine, tundra, and volcanic), classified as low or high in P availability. Unlike Figure 4, no Z-score transformation was applied, allowing for direct interpretation of raw abundance patterns between environments.

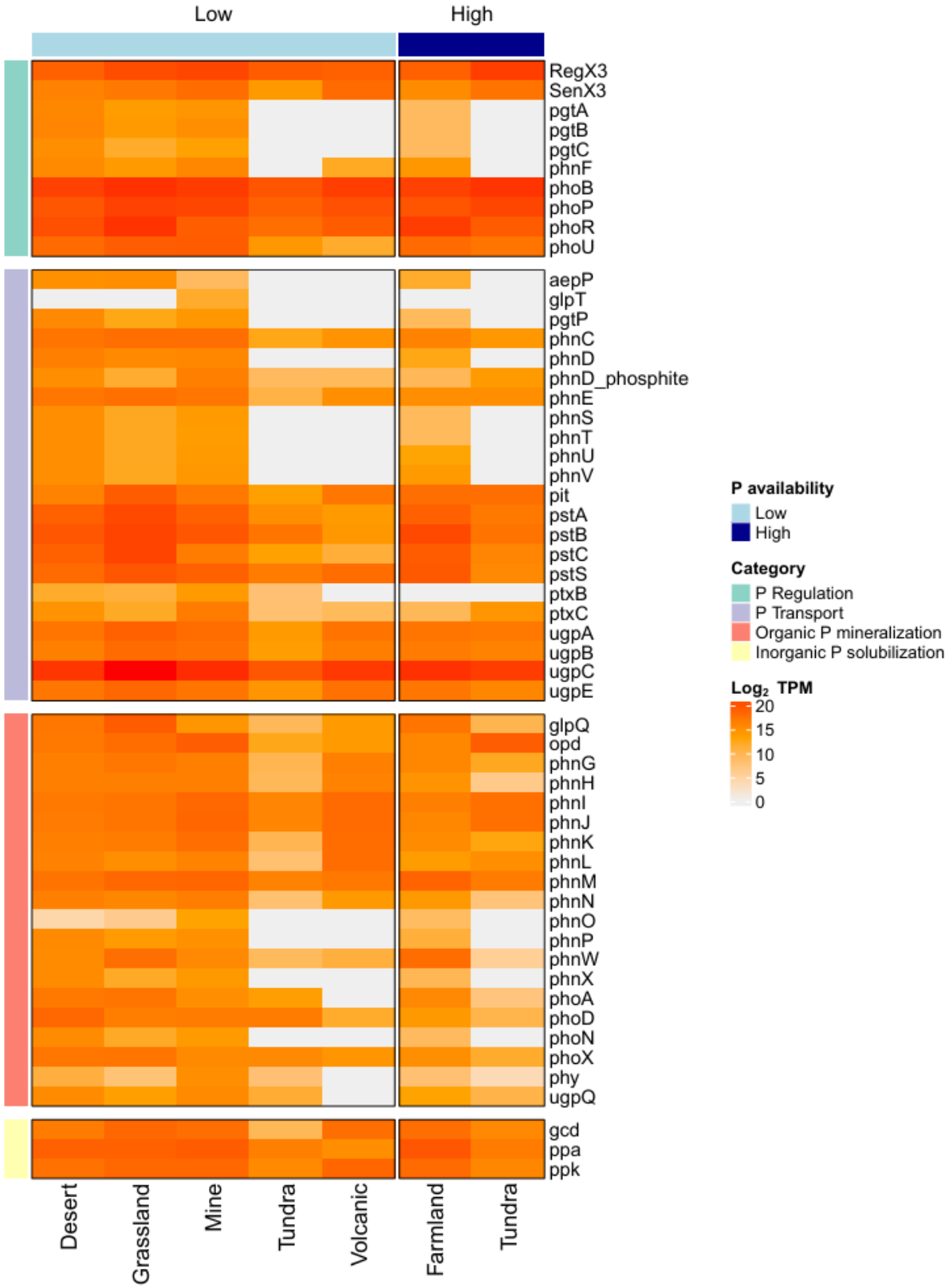

Figure S12. Overall abundance of P-acquisition genes (including all P-acquisition functional categories) in plasmids identified from metagenomic assemblies, comparing high- and low-P environments. Asterisk indicate statistical significance: * = $p < 0.1$.

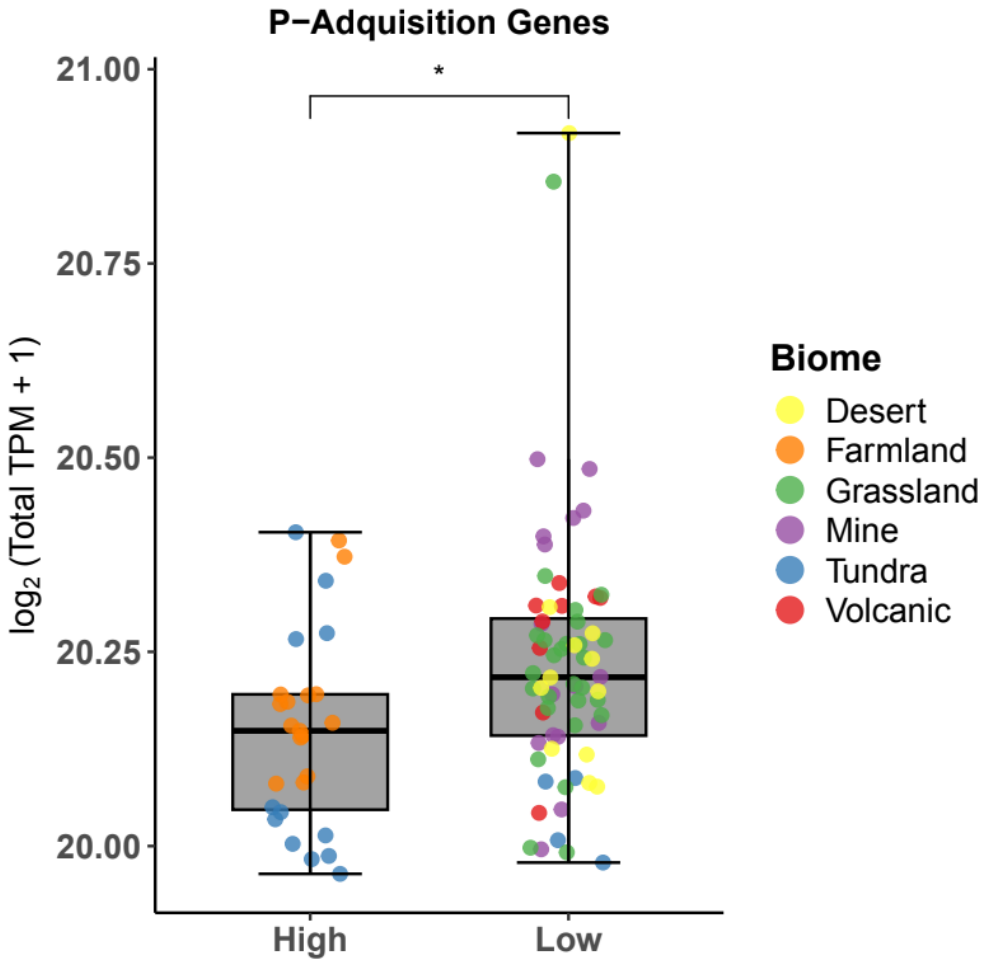

Figure S13. Relative abundance of bacterial host taxa predicted for 429 plasmids carrying P-acquisition genes identified from the metagenomic assemblies.

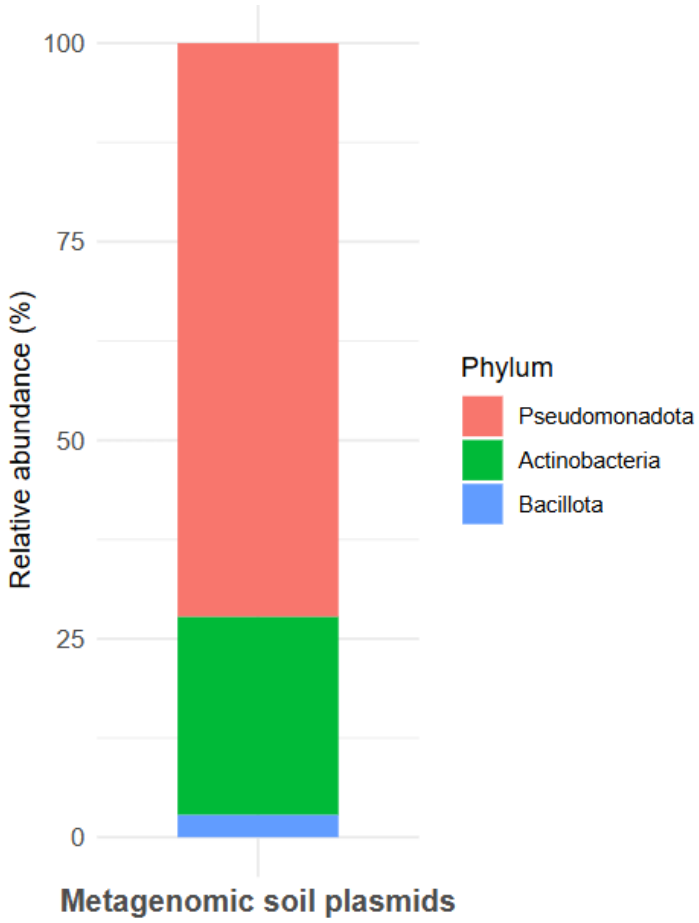

Figure S14. Heatmap visualization showing the count and distribution of 429 plasmids with unique predicted host taxa, carrying P-acquisition genes identified from the metagenomic assemblies. The genes are grouped according to their respective P-acquisition functional categories.