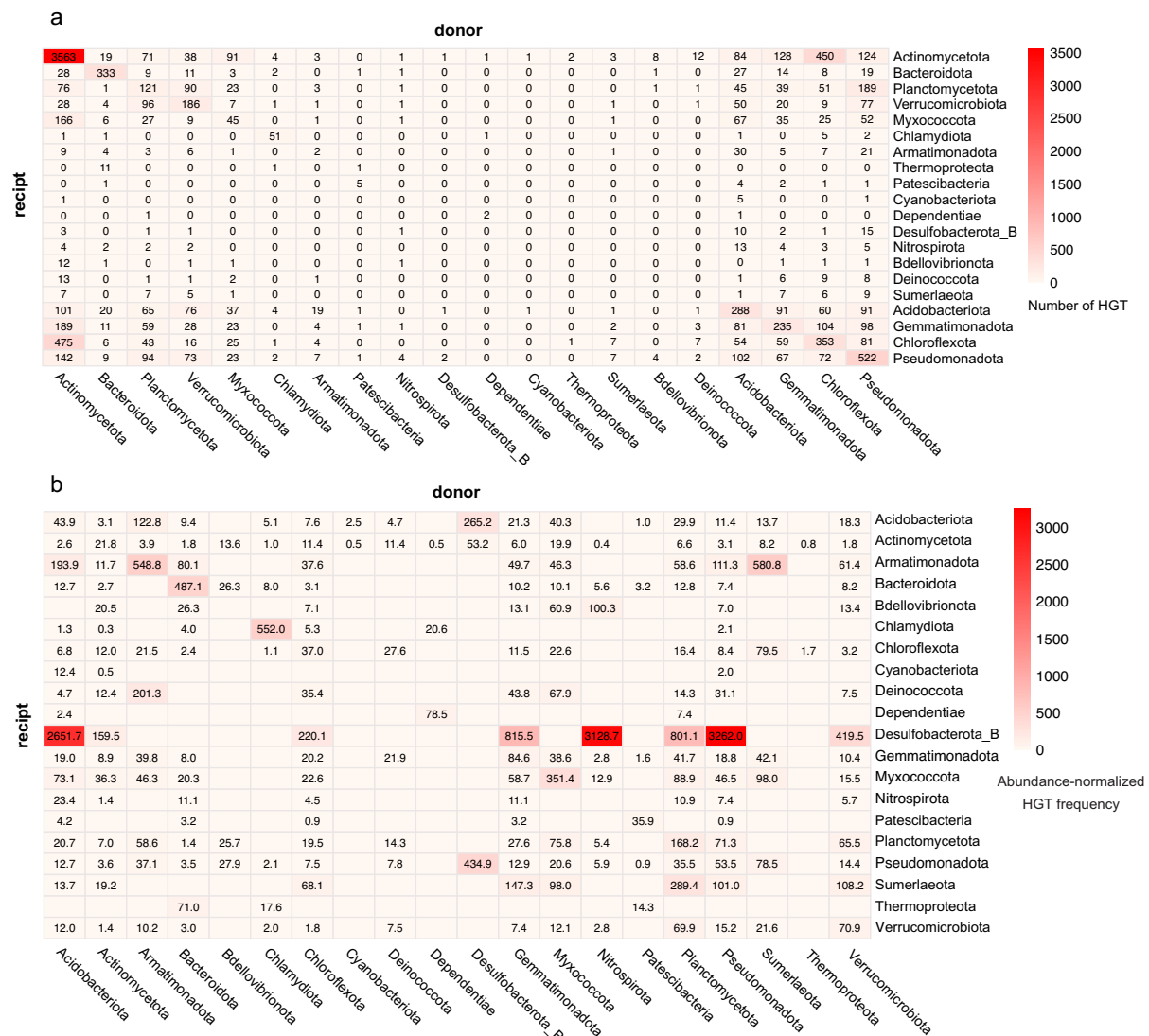
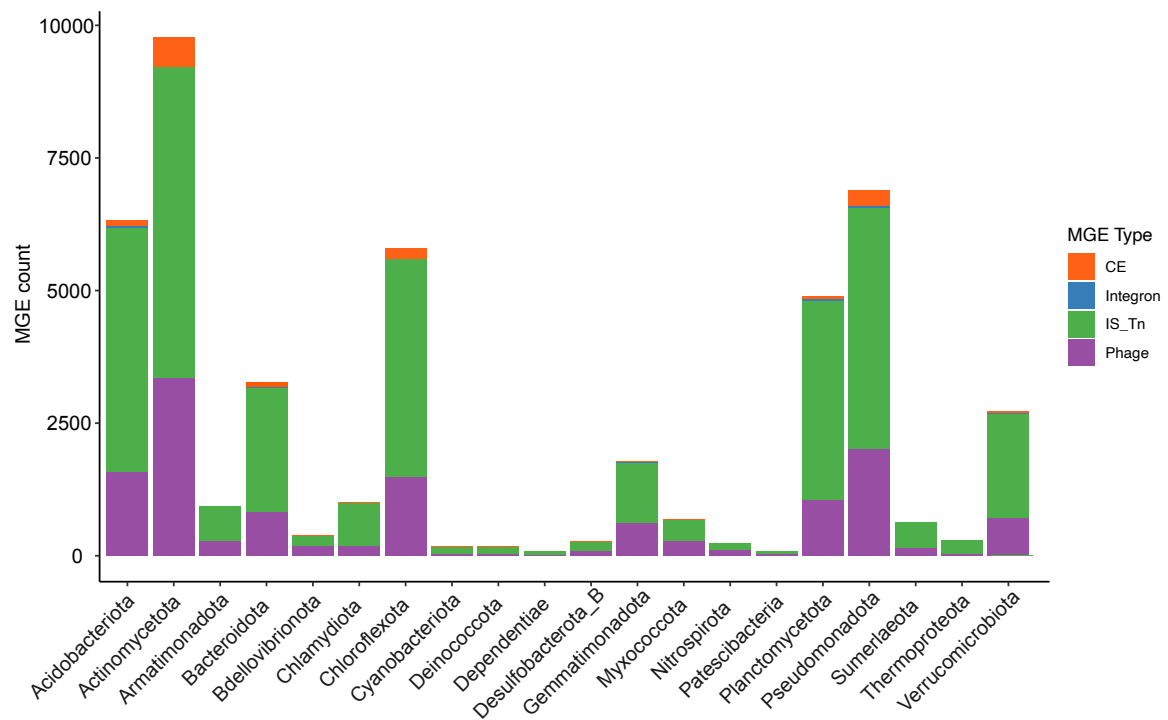




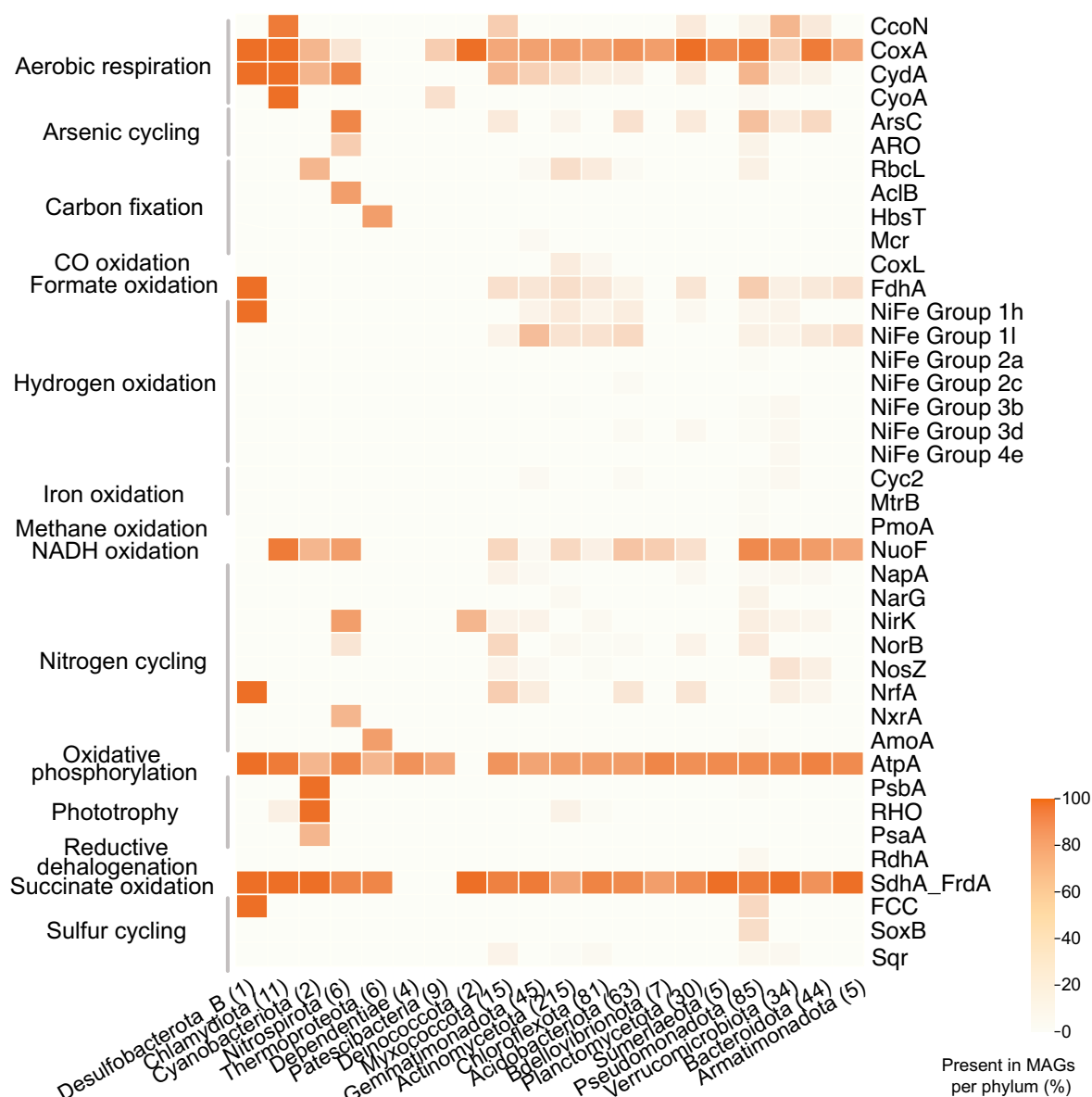
Supplementary Figure 2. HGT events among different microbial phyla in Antarctic soils.

The upper heatmap (a) displays the number of inferred HGT events between pairs of microbial phyla, while the lower heatmap (b) shows HGT frequencies normalized by the product of phylum relative abundances. Detailed statistics for HGT events across 676 MAGs are provided in Supplementary Data 3 and 4. Source data are provided as a Source Data file.

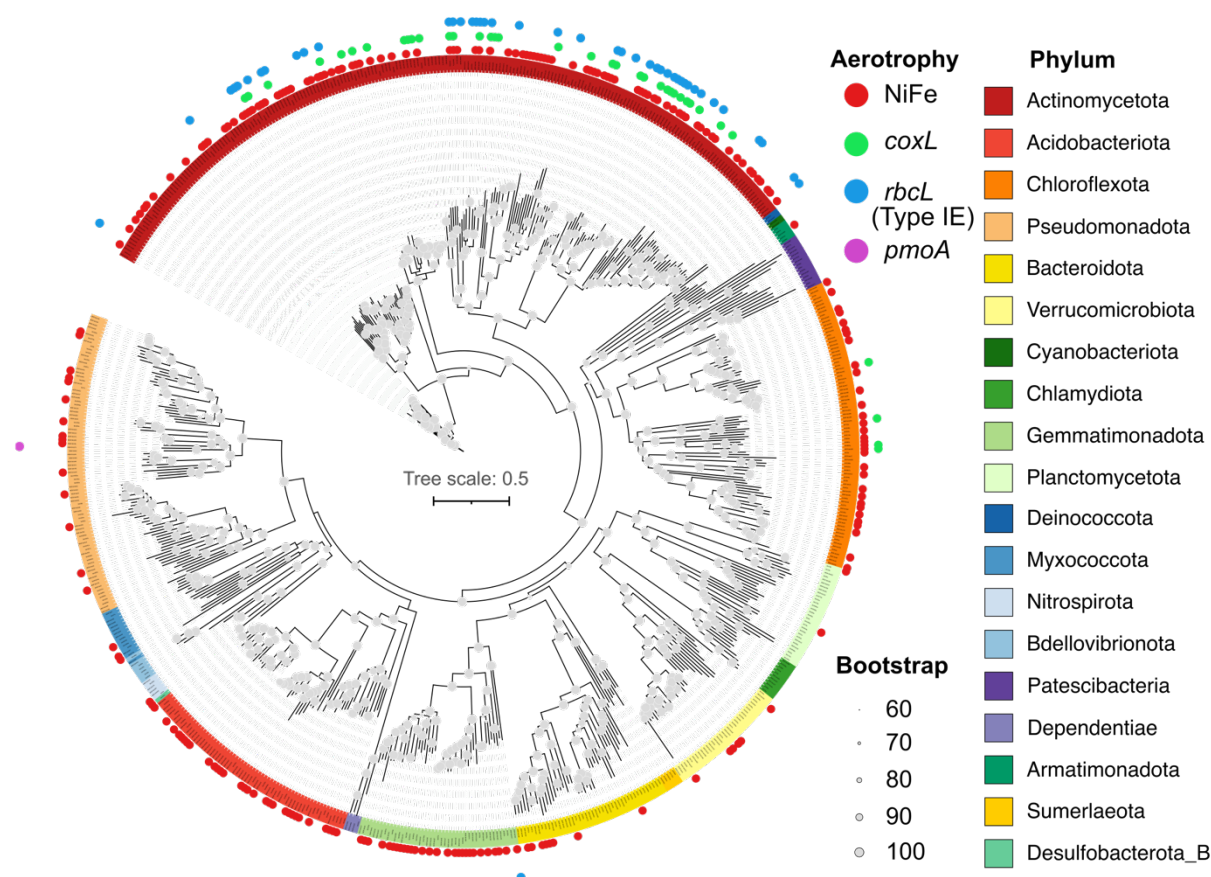


Supplementary Figure 3. The number of MGEs across microbial phyla in Antarctic soils.

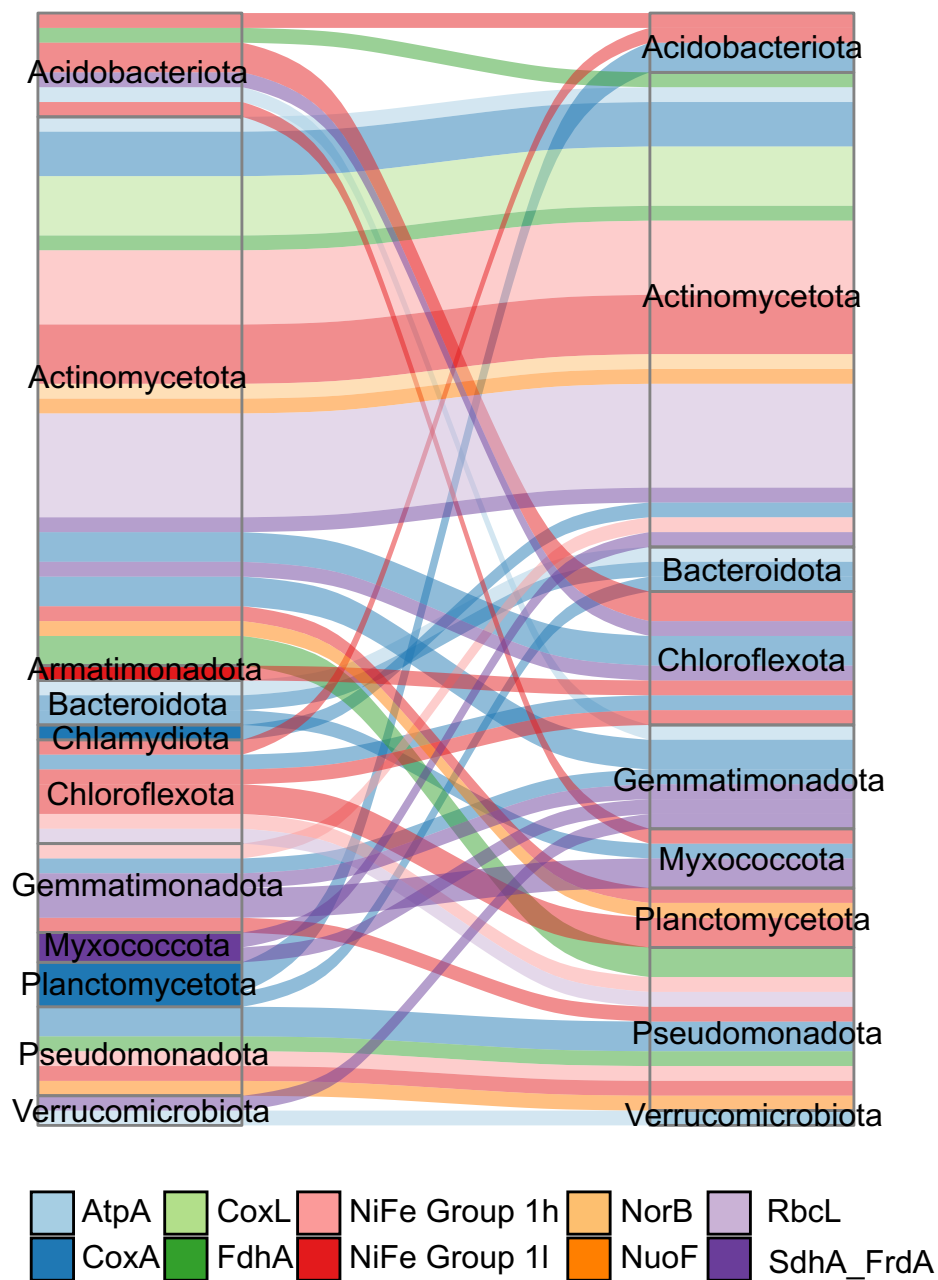
The bar plot shows the total number of MGEs identified in each phylum, with different MGE types represented by distinct colors. CE refers to conjugative elements and IS_Tn refers to insertion sequences (IS) and transposons. Detailed annotations of MGE-related genes are provided in Supplementary Data 9. Source data are provided as a Source Data file.



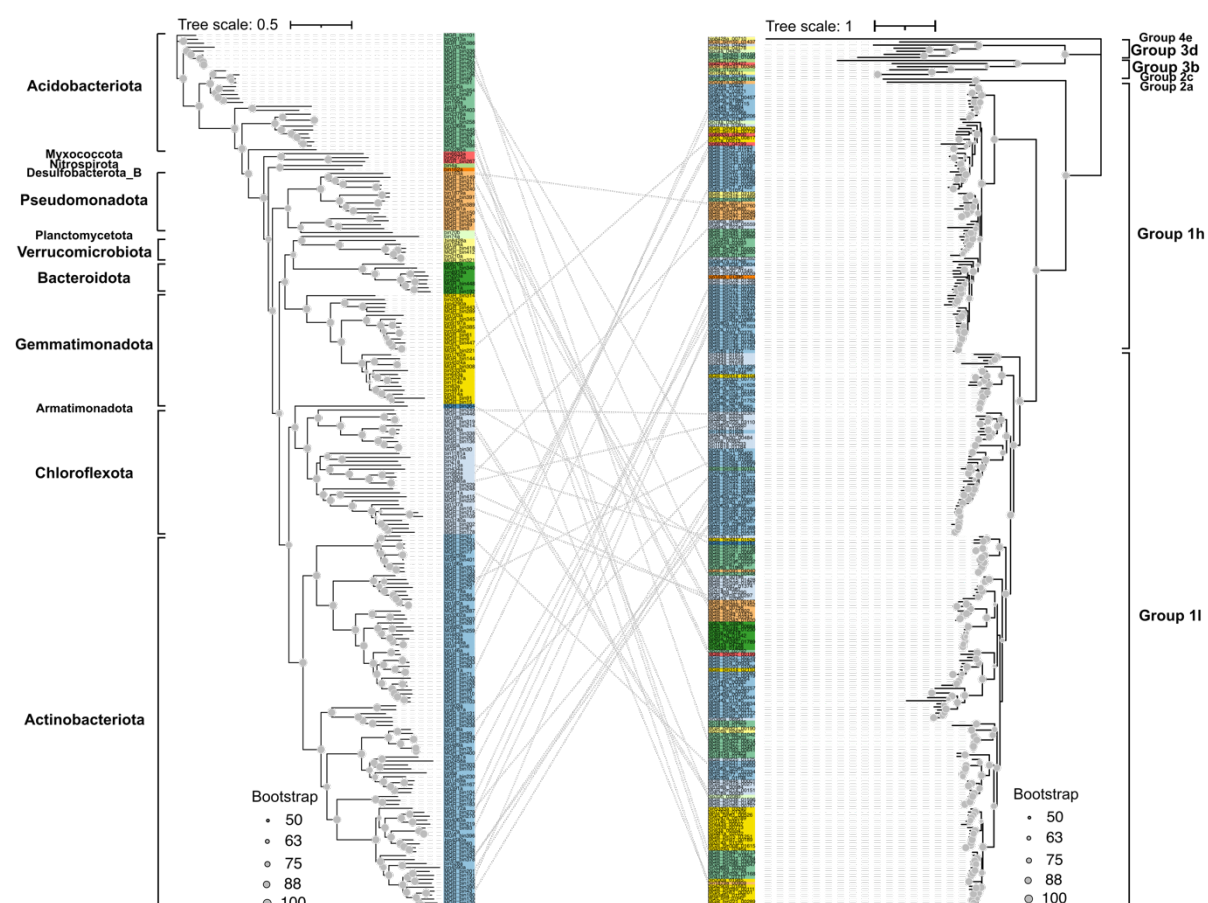
Supplementary Figure 4. The proportion of key metabolic proteins detected across Antarctic microbial phyla. The heatmap shows the proportion of MAGs within each microbial phylum that contain specific key metabolic proteins. Color intensity reflects the frequency of protein presence. Detailed annotations are available in Supplementary Data 8. Source data are provided as a Source Data file.



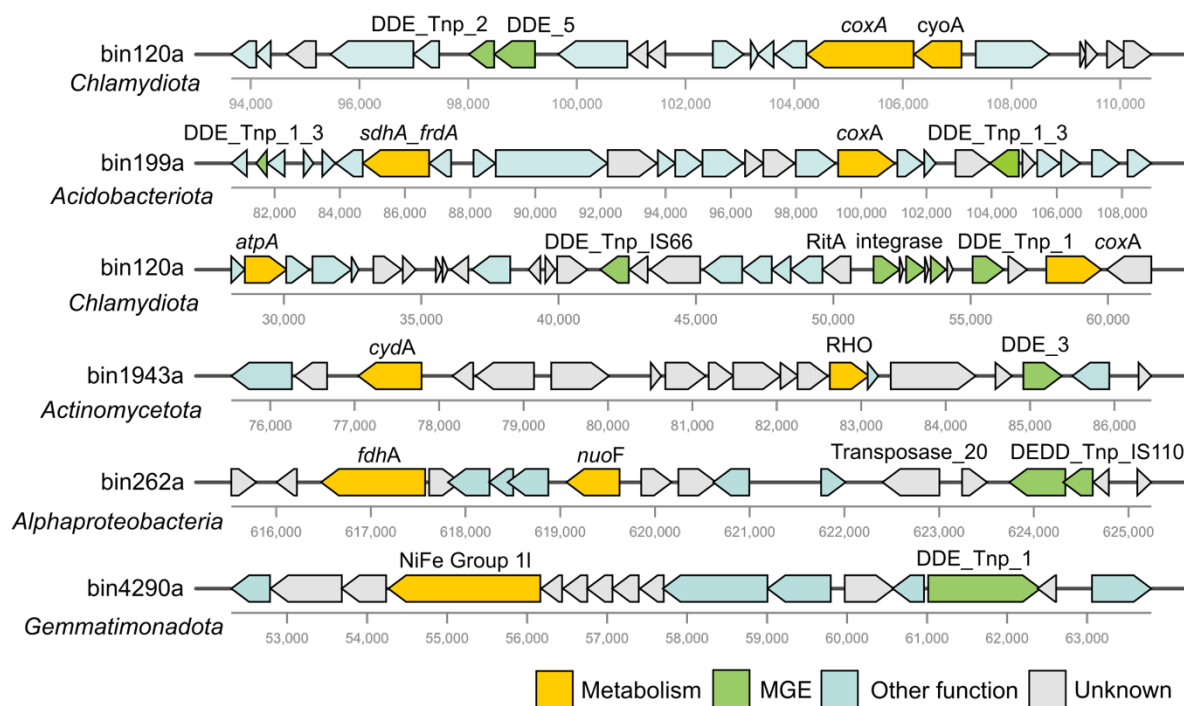
Supplementary Figure 5. Maximum-likelihood phylogenetic tree of bacterial MAGs in Antarctic soils, annotated with phylum-level taxonomy and aerotrophy gene content. The phylogenomic tree was inferred based on concatenation of conserved bacterial single-copy genes that were used as phylogenetic markers in GTDB-Tk, constructed using IQ-TREE. MAGs containing aerotrophy genes (group 1h and 1l [NiFe]-hydrogenase, *coxL*, type IE *rbcL*, and *pmoA*) are indicated by colored circles. Bootstrap support values (60-100%) are indicated by the size of the grey dots at each node. Scale bars represent the number of substitutions per site. Details of these MAGs and aerotrophy genes are available in Supplementary Data 8.



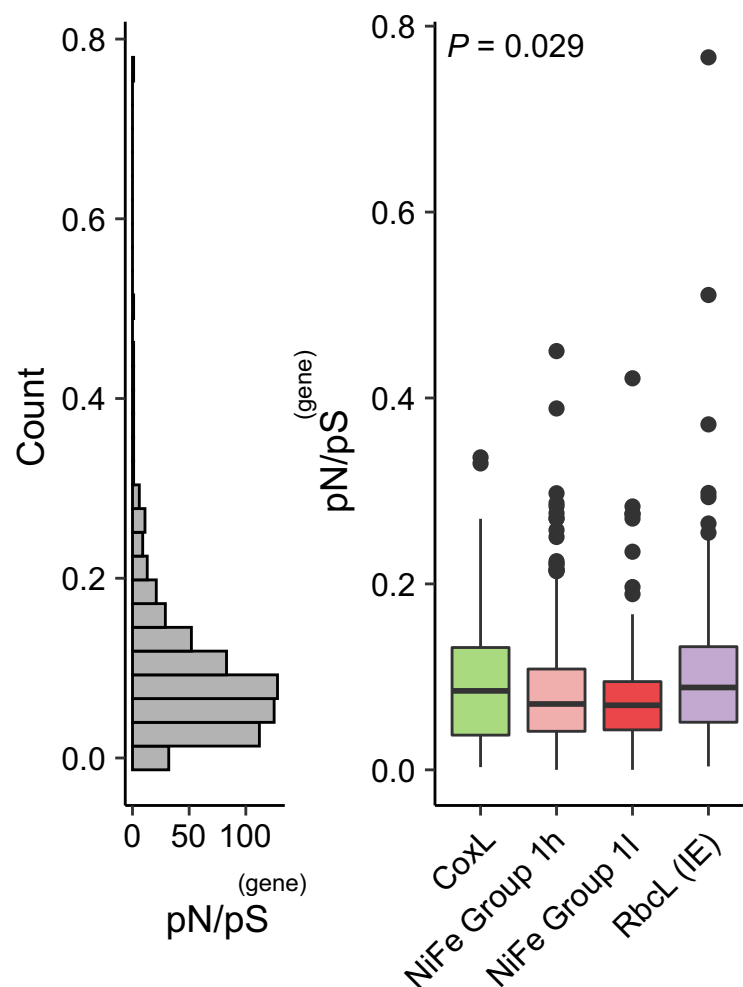
Supplementary Figure 6. Predicted gene flow of key metabolic genes within Antarctic microbial communities. Bands connect donors (left) and recipients (right), with the width of the band correlating to the number of HGTs and the colour corresponding to the gene types. Detailed information on HGT events of key metabolic genes is provided in Supplementary Data 8. Source data are provided as a Source Data file.



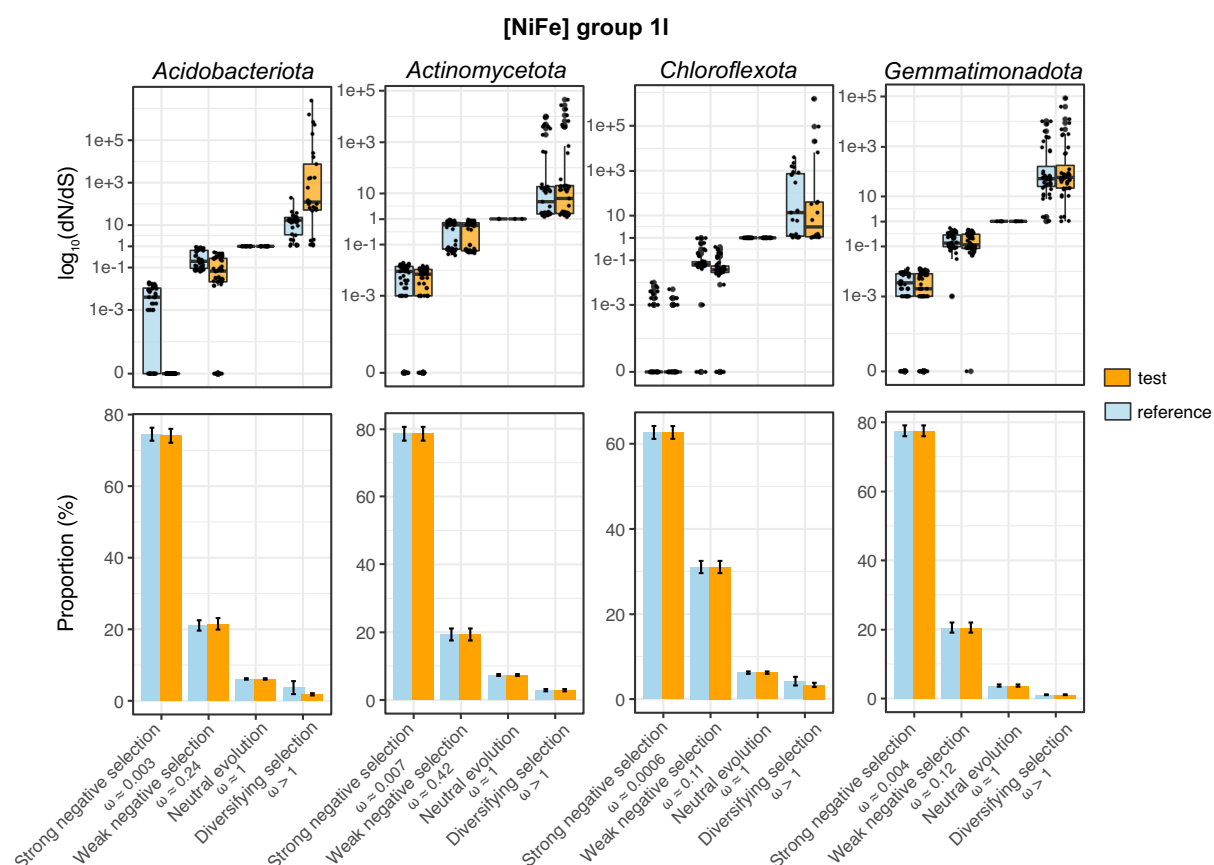
Supplementary Figure 7. Maximum-likelihood phylogenetic trees of hydrogenase-bearing MAGs and their hydrogenase protein sequences. Left panel shows a phylogenomic tree based on concatenated conserved marker genes from 222 bacterial MAGs encoding hydrogenases, with phylum-level taxonomy annotated. The right panel displays a phylogenetic tree of the hydrogenase protein sequences derived from the same MAGs. Hydrogenases predicted to be horizontally transferred are linked to their corresponding MAGs with dashed lines. Sequences from the same taxonomic groups are color-coded consistently across both trees. Bootstrap support values (50-100%) are indicated by the size of the grey dots at each node. Scale bars represent the number of substitutions per site. Detailed statistics for representative MAGs and hydrogenase genes are available in Supplementary Data 2 and 8.



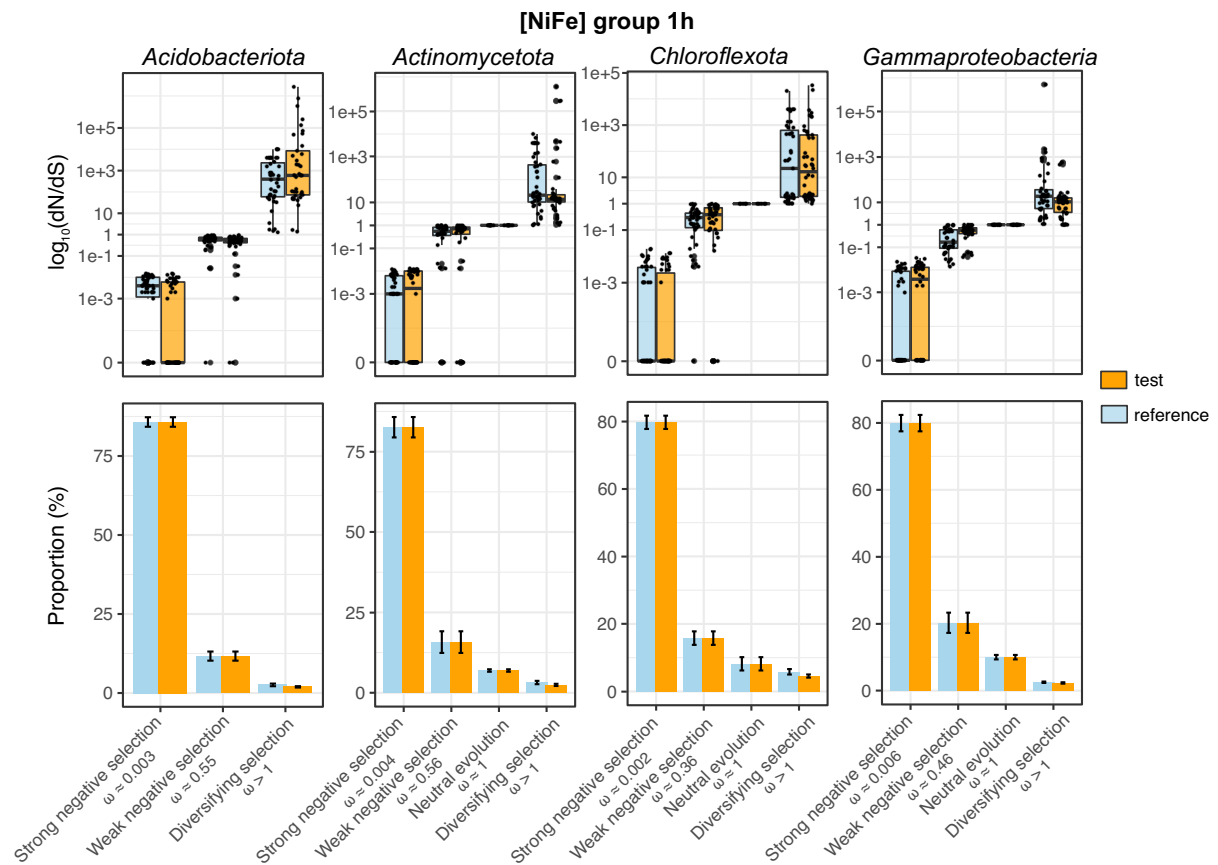
Supplementary Figure 8. Representative gene clusters containing metabolic marker genes involved in energy conservation. Genes are color-coded by functional category: yellow, metabolic genes; green, MGEs; blue, other functional genes; grey, unknown function. Detailed statistics for MGEs are provided in Supplementary Data 9.



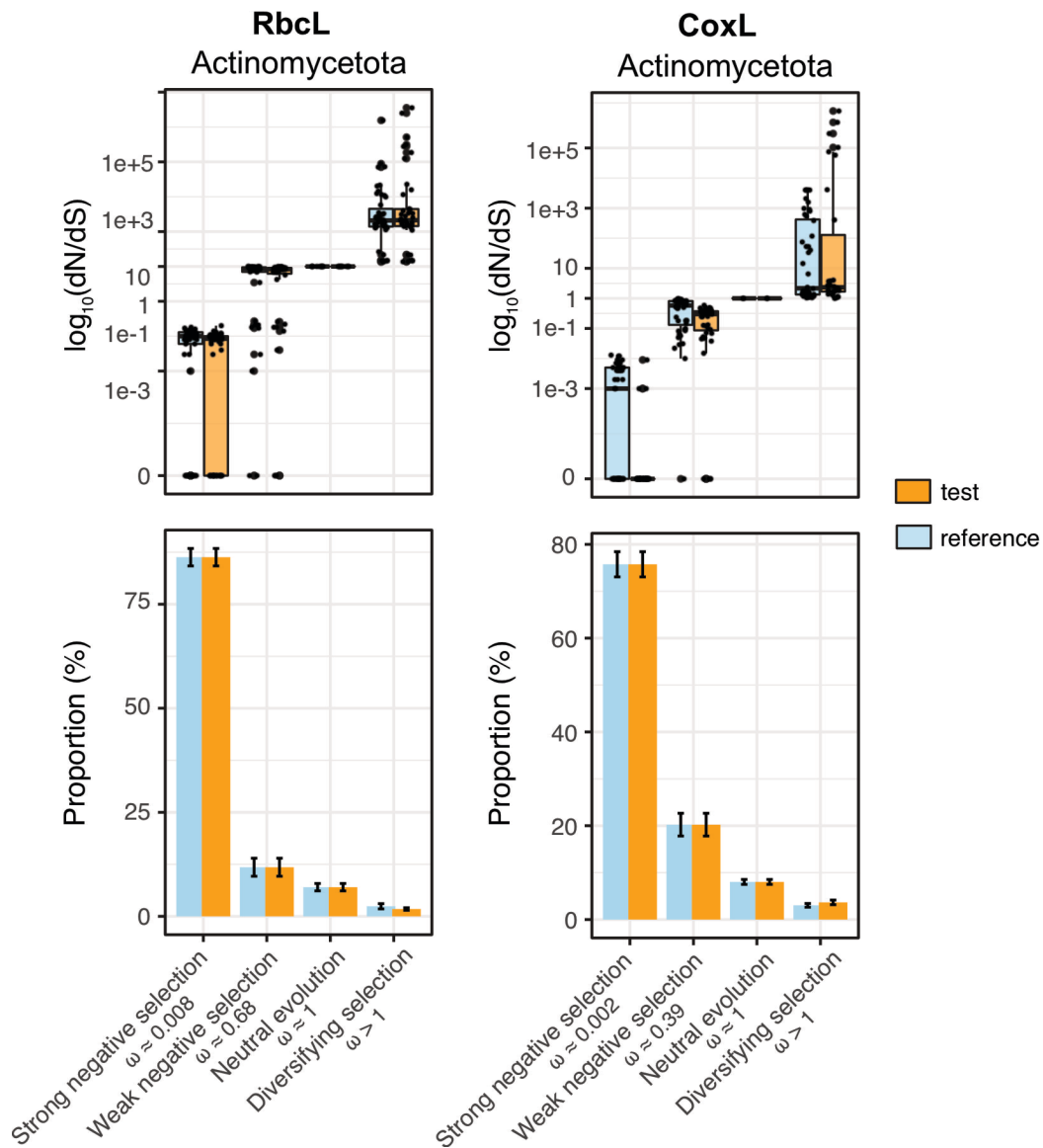
Supplementary Figure 9. Selection pressures of aerotrophy genes in an independent set of nine Antarctic metagenomes from geographically distinct regions (Dronning Maud Land, Robinson Ridge, and Bunger Hills). (a) Distributions of $pN/pS^{(gene)}$ for aerotrophy genes across nine samples. (b) Boxplot of $pN/pS^{(gene)}$ for two groups of hydrogenases, CO dehydrogenases, and Type IE RuBisCO. P-value of differences as calculated using a Kruskal-Wallis test. Source data are available in Supplementary Data 11. Box plots show the median (centre line), interquartile range (box), and $1.5 \times IQR$ whiskers. Source data are provided as a Source Data file.



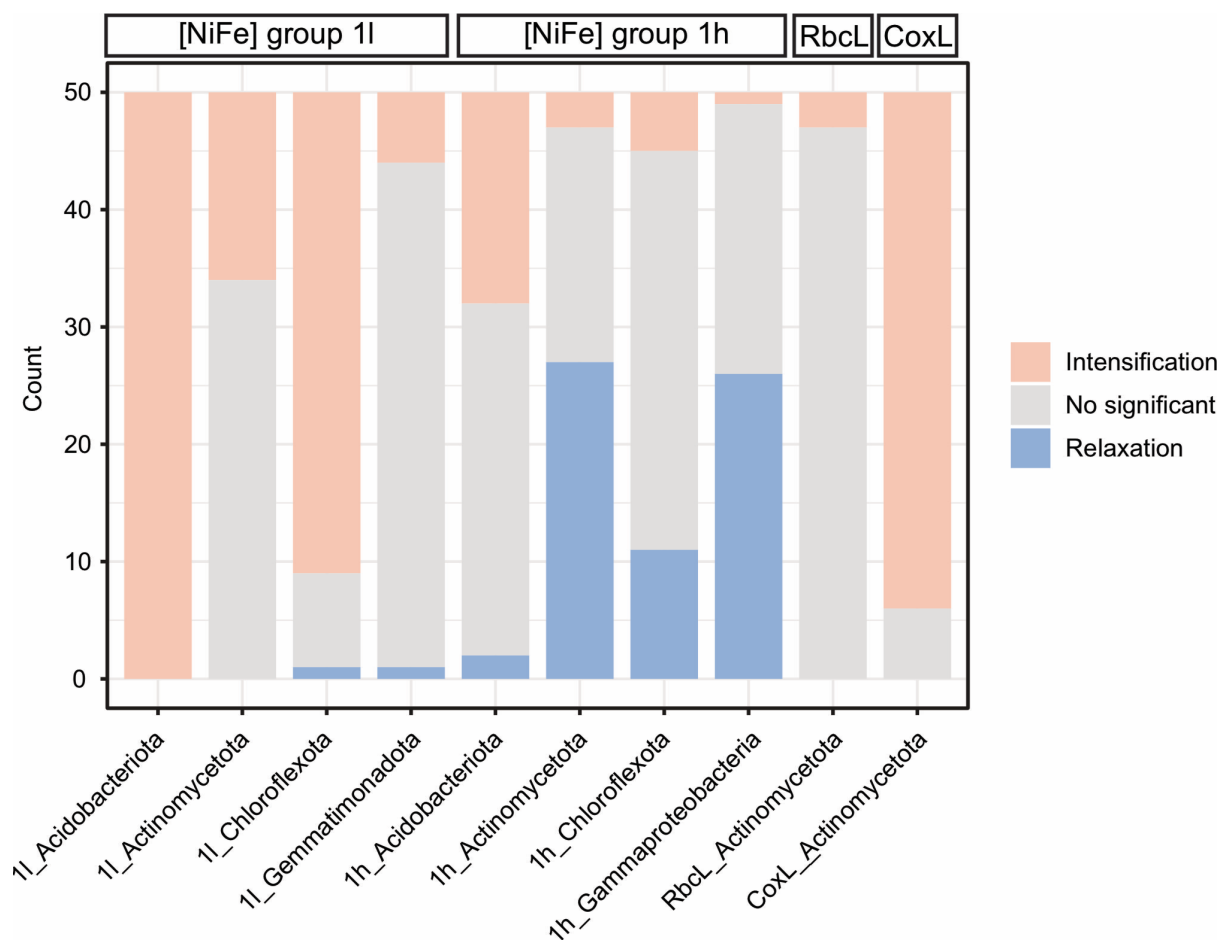
Supplementary Figure 10. Codon-based substitution model (dN/dS) results for [NiFe] group 1I hydrogenases comparing test (Antarctic) and reference (non-Antarctic) groups using RELAX in HyPhy, based on 50 independent subsampling replicates. The upper panels show boxplots of dN/dS (ω) values across four phyla, categorized into four selection groups: strong negative selection, weak negative selection, neutral evolution, and diversifying selection. The y-axis represents log₁₀-transformed dN/dS values. The lower panels show the mean proportion of sites assigned to each selection group, with error bars indicating standard error across subsampling replicates. Selection categories are annotated with their corresponding mean dN/dS (ω) values. Test branches are shown in orange and reference branches in blue. Relevant data are available in Supplementary Data 12. Box plots show the median (centre line), interquartile range (box), and $1.5 \times \text{IQR}$ whiskers. Source data are provided as a Source Data file.



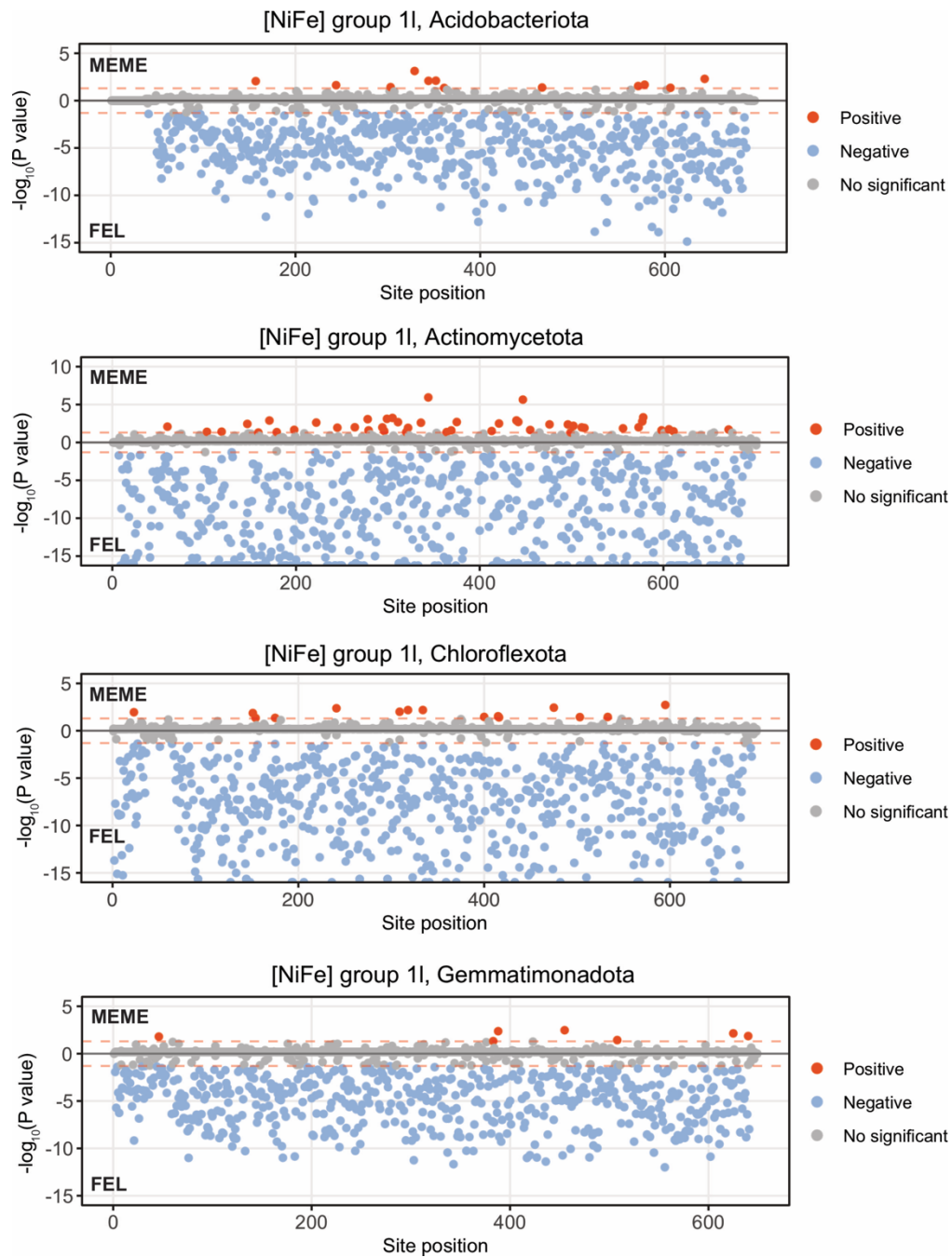
Supplementary Figure 11. Codon-based substitution model (dN/dS) results for [NiFe] group 1h hydrogenases comparing test (Antarctic) and reference (non-Antarctic) groups using RELAX in HyPhy, based on 50 independent subsampling replicates. The upper panels show boxplots of dN/dS (ω) values across four phyla, categorized into four selection groups: strong negative selection, weak negative selection, neutral evolution, and diversifying selection. The y-axis represents log₁₀-transformed dN/dS values. The lower panels show the mean proportion of sites assigned to each selection group, with error bars indicating standard error across subsampling replicates. Selection categories are annotated with their corresponding mean dN/dS (ω) values. Test branches are shown in orange and reference branches in blue. Relevant data are available in Supplementary Data 12. Box plots show the median (centre line), interquartile range (box), and $1.5 \times \text{IQR}$ whiskers. Source data are provided as a Source Data file.



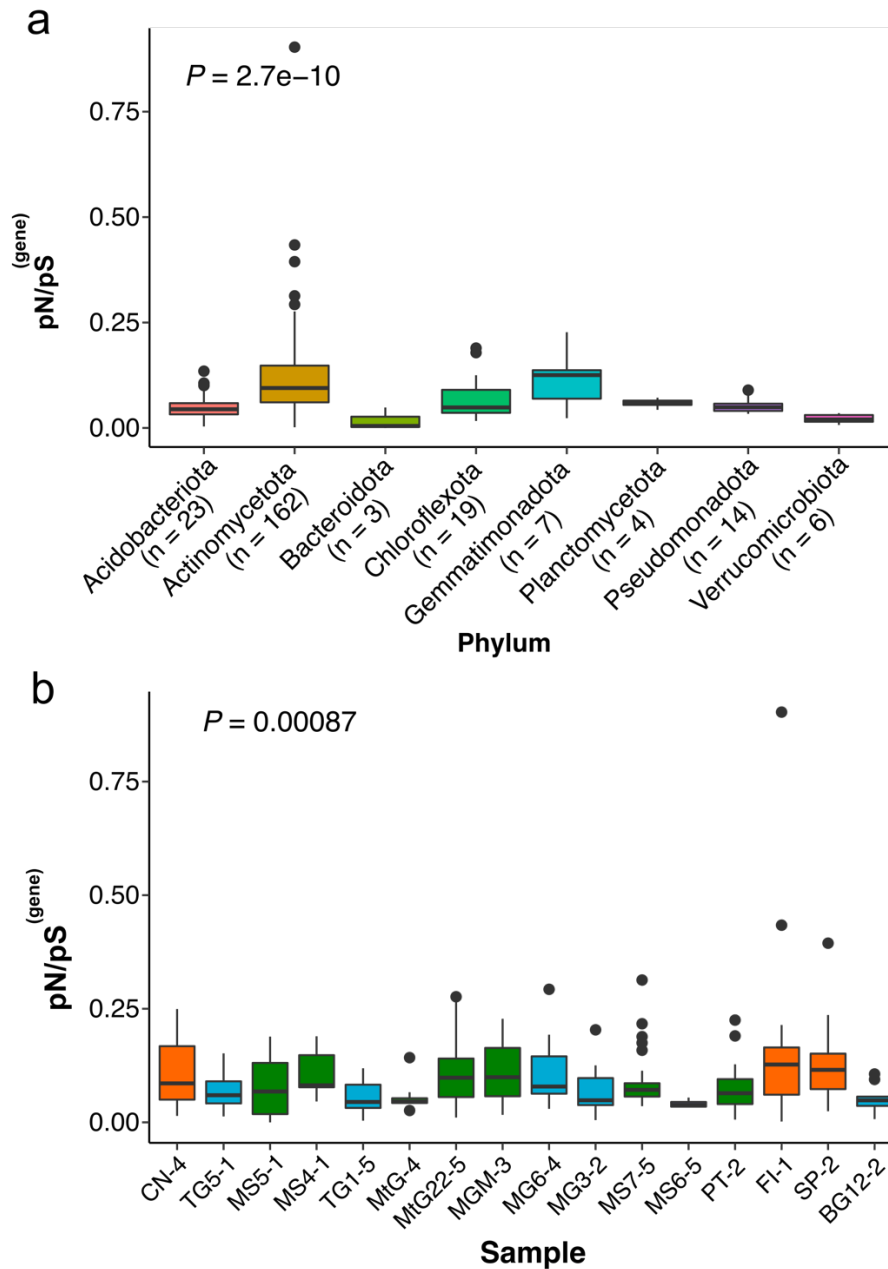
Supplementary Figure 12. Codon-based substitution model (dN/dS) results for CO dehydrogenases (CoxL) and Type IE RuBisCO (RbcL) comparing test (Antarctic) and reference (non-Antarctic) groups using RELAX in HyPhy, based on 50 independent subsampling replicates. The upper panels show boxplots of dN/dS (ω) values across four phyla, categorized into four selection groups: strong negative selection, weak negative selection, neutral evolution, and diversifying selection. The y-axis represents log₁₀-transformed dN/dS values. The lower panels show the mean proportion of sites assigned to each selection group, with error bars indicating standard error across subsampling replicates. Selection categories are annotated with their corresponding mean dN/dS (ω) values. Test branches are shown in orange and reference branches in blue. Box plots show the median (centre line), interquartile range (box), and $1.5 \times \text{IQR}$ whiskers. Relevant data are available in Supplementary Data 12. Source data are provided as a Source Data file.



Supplementary Figure 13. Selection intensity of aerotrophy genes in test (Antarctic) relative to reference (non-Antarctic) groups inferred using RELAX in HyPhy, based on 50 independent subsampling replicates. Stacked bar plots show the number of subsampling replicates supporting evidence of selection intensification (light orange), relaxation (blue), or no significant change (grey) in the test group relative to the reference group. Relevant data are available in Supplementary Data 12-13. Source data are provided as a Source Data file.

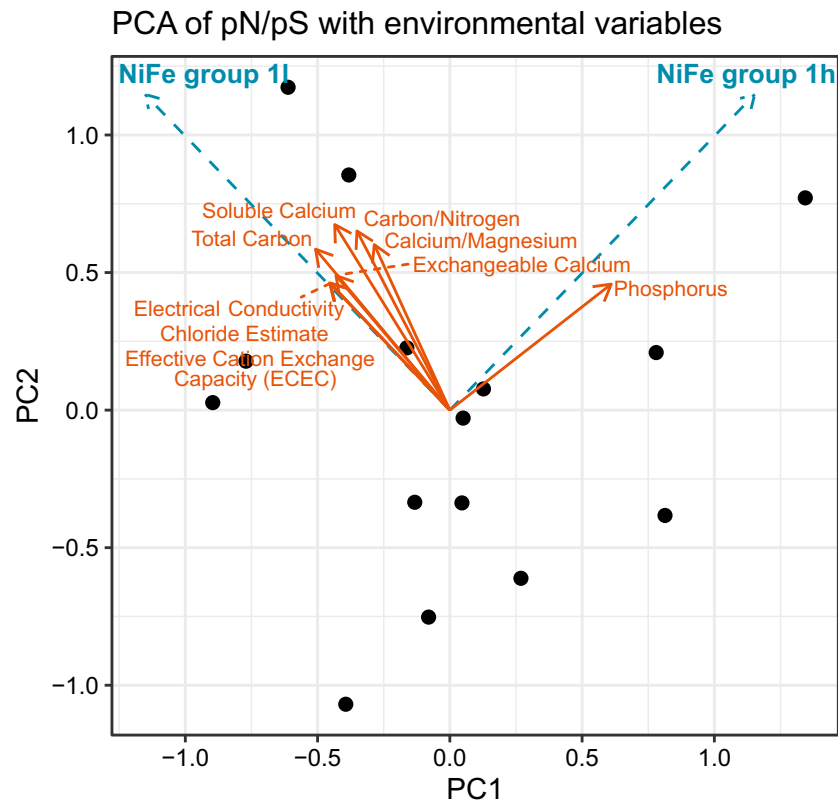


Supplementary Figure 14. Site-level selection analyses of [NiFe] group 1I hydrogenases inferred using MEME (Mixed Effects Model of Evolution) and FEL (Fixed Effects Likelihood) in HyPhy. The scatter plot shows site positions on the x-axis and $-\log_{10}(\text{P-values})$ from MEME and FEL analyses on the y-axis. Sites with MEME $P < 0.05$ (top part of the graph) are highlighted in red, indicating evidence of episodic positive selection. Sites with FEL $P < 0.05$ are considered statistically significant, providing evidence for pervasive selection: blue points indicate sites under negative selection, while red points indicate sites under positive selection. Relevant data are provided in Supplementary Data 14. Source data are provided as a Source Data file.

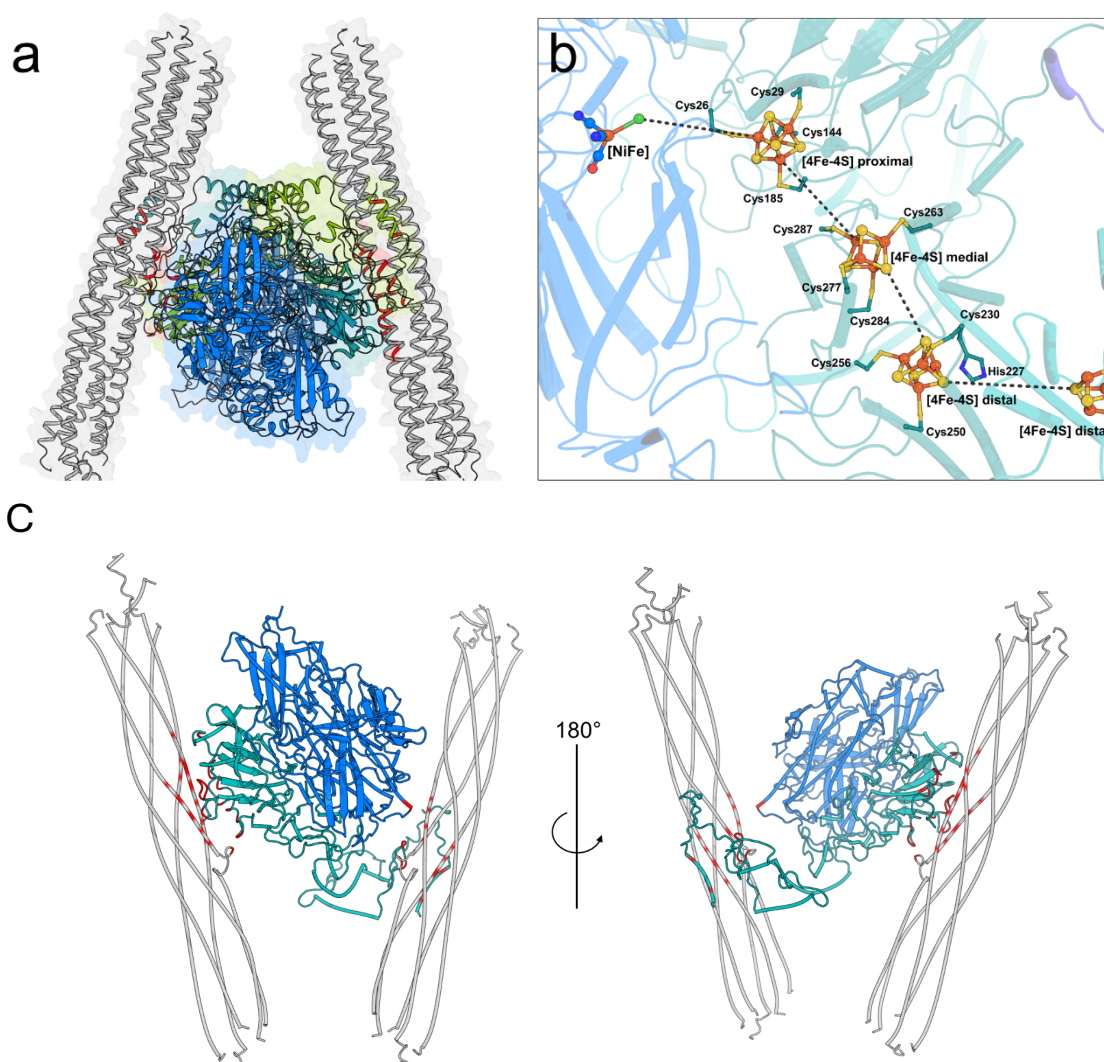


Supplementary Figure 15. Comparison of pN/pS for hydrogenase genes in Antarctic soils.

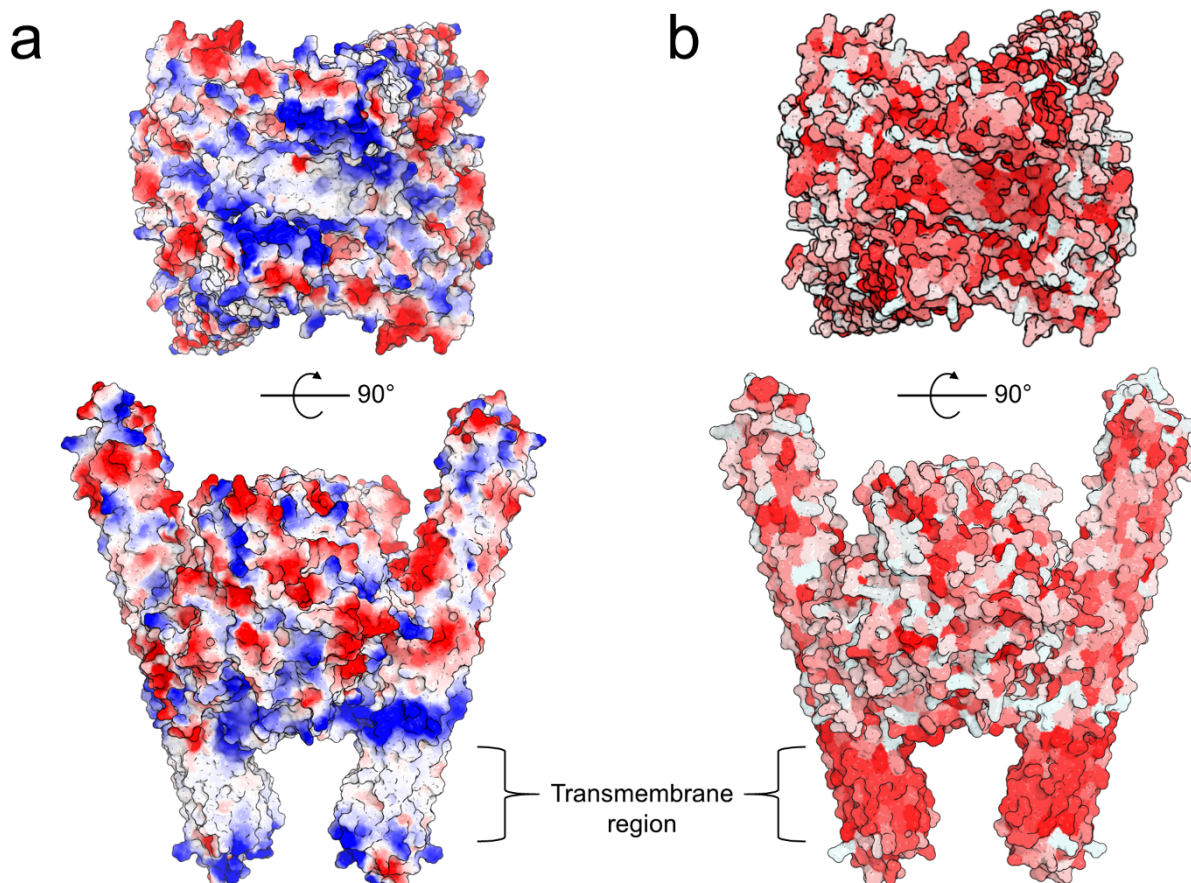
(a) pN/pS values of hydrogenase genes across different microbial phyla (n = number of gene-sample data points analysed per phylum). (b) pN/pS values of hydrogenase genes across different sampling sites. Boxes are coloured by sample type: mountain (green), glacier (blue), and miscellaneous (orange). P-values of differences across different taxonomic groups and sample sites were calculated using Kruskal-Wallis Rank Sum test. Box plots show the median (centre line), interquartile range (box), and $1.5 \times \text{IQR}$ whiskers. Source data can be found in Supplementary Data 1 and 10.



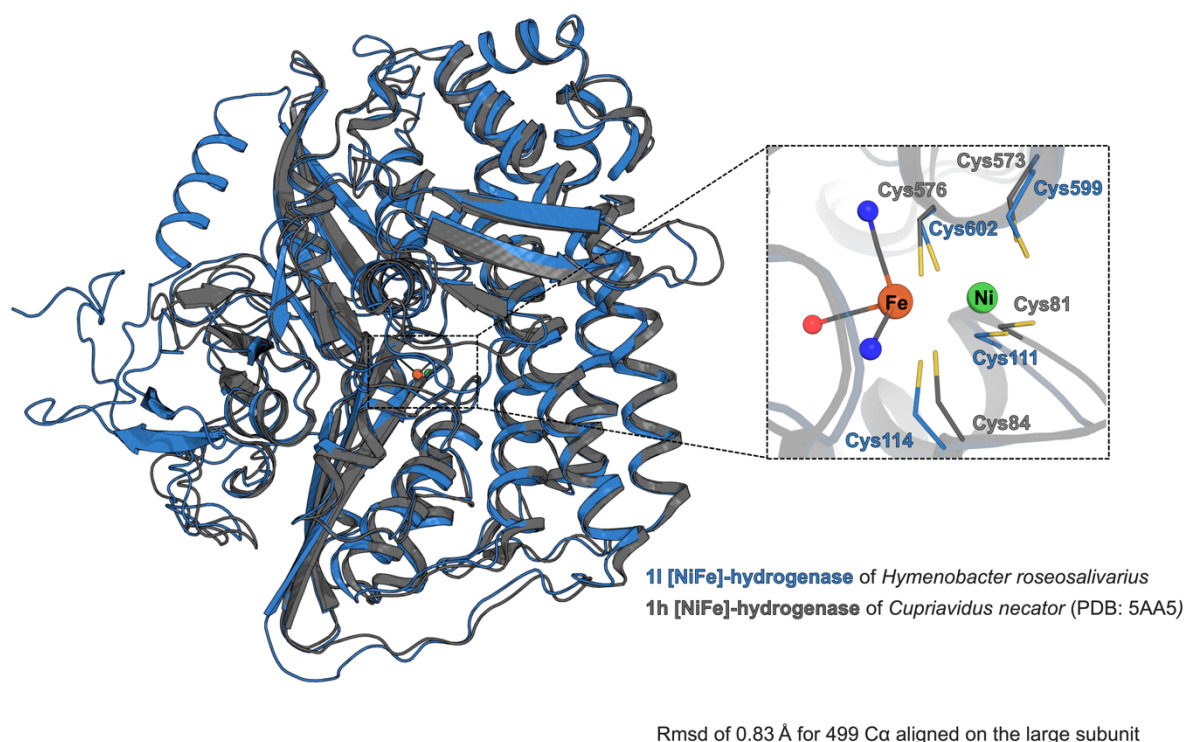
Supplementary Figure 16. Associations between sample-averaged $pN/pS^{(gene)}$ of hydrogenases and environmental parameters across samples. Principal component analysis (PCA) based on pN/pS values of [NiFe] Group 1h and 1l across samples. Arrows in blue represent the ordination loadings of the two gene groups, while red arrows indicate environmental variables significantly associated with the ordination (envfit, permutation test, $P < 0.05$). The length and direction of arrows reflect the strength and direction of correlations with the ordination axes. Environmental parameters were obtained from Ortiz et al. (2021, PNAS) and PCA loadings are provided in Supplementary Data 15.



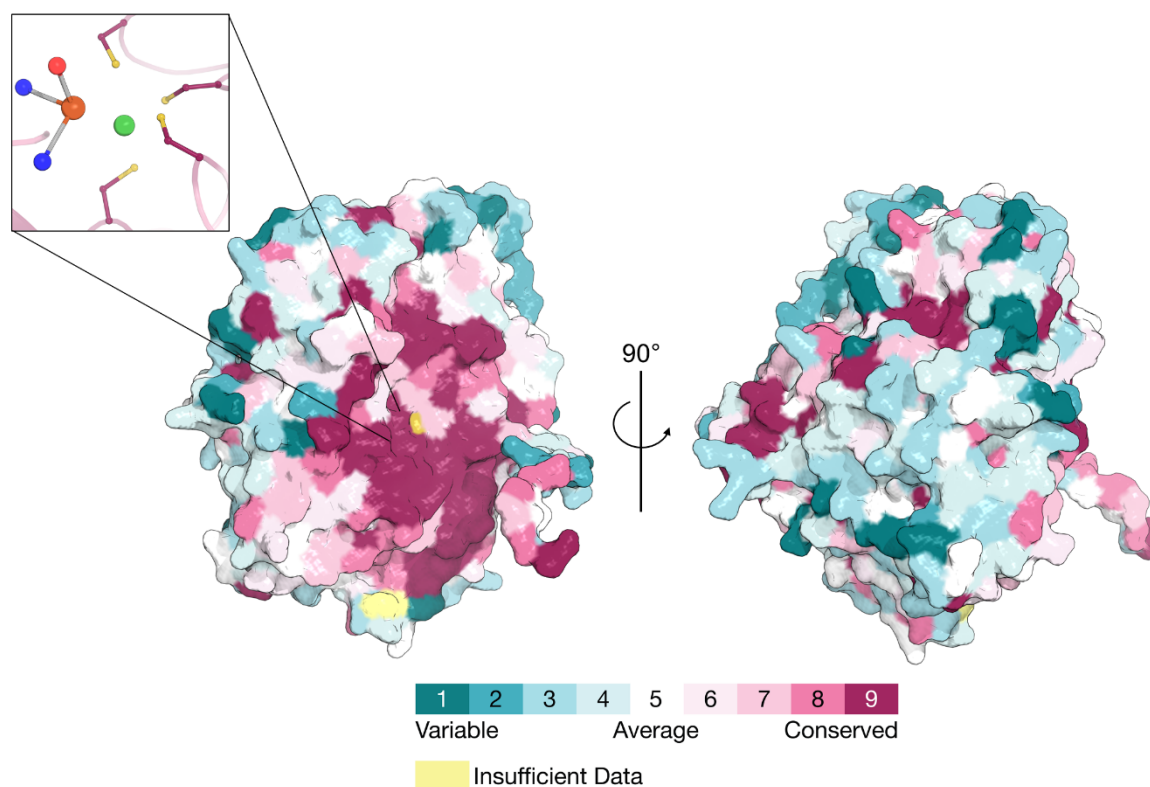
Supplementary Figure 17. Structural organization, membrane interactions, and electron transfer pathway of the Hyl hydrogenase complex. (a) The interface of Hyl with the transmembrane helices is primarily established by HylS. Contacts involved in the oligomerization of the Hyl dimer ((HylSU)₂) and the two membrane tubes (HylTM1-5) are highlighted in red. (b) Electron transfer relay within one Hyl dimer. The [NiFe] active site and the [4Fe-4S] clusters are shown as balls and sticks, with the proposed electron flow from the [NiFe] active site indicated by black dashed lines. (c) Predicted contacts involved in the oligomerization of the 11 [NiFe]-hydrogenase with its membrane stalks are highlighted in red. The cartoon shows one large subunit (HylL, blue) and one small subunit (HylS, teal) and both transmembrane stalks (HylTM1-5, grey). The interface between the two stalks (HylTM1-4) and the hydrogenase is predominantly established by the small subunit. The predicted structural model is available on Figshare (DOI: 10.6084/m9.figshare.30000151).



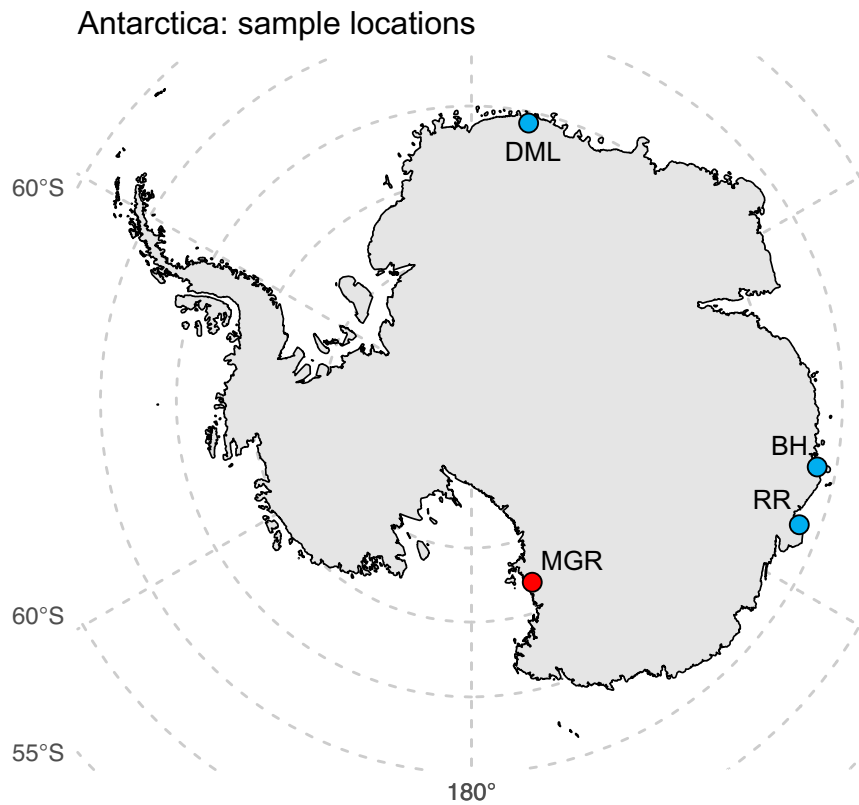
Supplementary Figure 18. Predicted vacuum electrostatics and hydrophobicity surface analysis of the group 11 [NiFe] hydrogenase complex. The predicted structural model of the group 11 [NiFe] hydrogenase was generated from the experimentally validated H₂ oxidizer *Hymenobacter roseosalivarius*. (a) Vacuum electrostatic surface analysis, shown as a gradient from red to blue for a more negative to a positive potential, respectively, with white indicating neutral surface potential. (b) Hydrophobicity surface analysis, shown in a hydrophobic gradient with non-hydrophobic amino acids in white to hydrophobic amino acids in red. The predicted structural model is available on Figshare (DOI: 10.6084/m9.figshare.30000151).



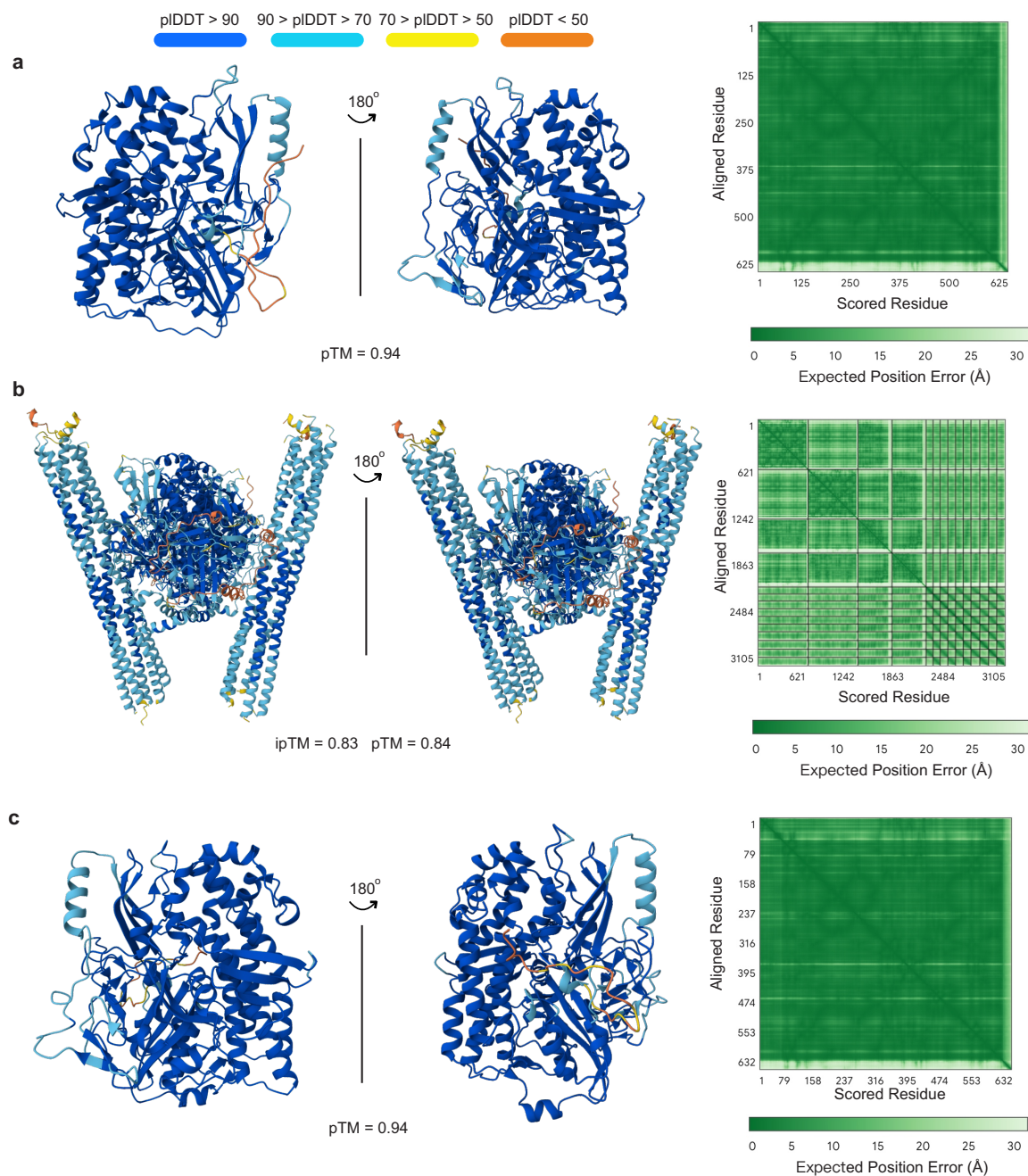
Supplementary Figure 19. Superposition of the active sites of group 1I [NiFe] hydrogenase and the group 1h [NiFe] hydrogenase. Structural alignment of the large subunit of group 1I [NiFe] hydrogenase from *Hymenobacter roseosalivarius* (blue) and group 1h [NiFe] hydrogenase (grey; PDB: 5AA5, rmsd of 0.83-Å for 499-Cα aligned). The [NiFe]-active site and coordinating ligands are shown in a close-up in stick and sphere representation. Carbon, nitrogen, oxygen, sulfur, iron and nickel are coloured in magenta/grey, blue, red, yellow, orange and green, respectively. The predicted structural model is available on Figshare (DOI: 10.6084/m9.figshare.30000151).



Supplementary Figure 20. Evolutionary conservation of amino acid positions in the catalytic subunit of a group 11 [NiFe] hydrogenase protein structure based on the phylogenetic relations between homologous sequences. The large subunit of the 11 [NiFe] structure shown in surface representation and coloured by its sequence conservation. The conservation scores are shown in a gradient from variable (turquoise), through intermediately conserved (white), to conserved positions (magenta). A close-up shows the [NiFe]-active site and the highly conserved ligating cysteine-residues, in sphere and stick representation. Position-specific conservation scores were calculated using the Bayesian method.



Supplementary Figure 21. Map of sampling locations across Antarctica using a Natural Earth basemap. Samples used in this study from the Mackay Glacier region (MGR) are shown as red points. An independent set of nine Antarctic metagenomes from geographically distinct regions, including Dronning Maud Land (DML), Bunge Hills (BH), and Robinson Ridge (RR), are shown in blue.



Supplementary Figure 22. Structural quality metrics and validation plots for the predicted protein models. Quality metrics are shown for the protein models presented in the main text: (a) the large subunit of group 11 [NiFe] hydrogenase from *Hymenobacter roseosativarius*, (b) the group 11 [NiFe] hydrogenase complex from *Hymenobacter roseosativarius*, comprising the Hyl dimer ((HylSU)₂) and the two membrane tubes (HylTM1-5), and (c) the large subunit of group 11 [NiFe] hydrogenase from an *Acidothermales* bacterium (*Sporichthyaceae* family). For each panel, the left and middle structures display the front and back views coloured by pLDDT scores to illustrate local confidence. Model prediction scores (pTM and/or ipTM) are indicated below. The right panels show the corresponding Predicted

Aligned Error (PAE) matrices, indicating the predicted relative positional error between residues in Angstroms (Å).