Supplementary Materials for

Seasonal temperature changes shape microbial community patterns and carbon-fixing gene fluctuations in tidal flats

Kuo-Jian Ma *et al.*

*Corresponding author. Email: xuxw@sio.org.cn, michael_sc@sina.com

This PDF file includes:

Supplementary Methods
Figs. S1–S23

Other Supplementary Materials for this manuscript include the following:

Supplementary data 1–16

Supplementary Methods

Total Protein Extraction

Sediment samples were kept frozen and processed on ice. Samples were suspended in borax/polyvinylpolypyrrolidone/phenol buffer (BPP buffer, 100 mM EDTA, 50 mmol/L borax, 50 mmol/L vitamin C, 30 % sucrose (w/v), 10 mM Tris-base, 1 % Triton-100 (w/v), 5 mM dithiothreitol, 1 % polyvinylpolypyrrolidone (w/v)) and homogenized using a high-flux tissue grinder for three cycles (40 s each). After centrifugation (12,000 g, 20 min, 4℃), the supernatant was mixed with an equal volume of Tris-saturated phenol and vortexed for 10 min at 4℃. The phenol phase was collected after centrifugation (12,000 g, 20 min, 4℃), washed with an equal volume of BPP buffer, vortexed, and centrifuged again. Five volumes of pre-cooled ammonium acetate methanol solution were added to precipitate proteins overnight at −20℃. Next day, the pellet was washed twice with 90% pre-cooled acetone after centrifugation (12,000 g, 20 min, 4℃).

The precipitate was washed and dissolved in lysis buffer containing 8 M urea, 1% SDS, and protease inhibitor cocktail, followed by 2 min of sonication on ice and centrifugation (12,000 g, 20 min, 4℃). Protein concentration was measured using the ProteoAnalyzer (M5350AA). Protein quality was assessed by Bicinchoninic acid assay kit (BCA kit, Thermo, USA) and Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE).

Protein Digestion, Desalting, and Quantification

A total of 100 μg of protein was resuspended in Triethylammonium bicarbonate buffer (TEAB buffer) with the final concentration of 100 mM. Proteins were reduced with 10 mM TCEP (final concentration) at 37℃ for 60 min and alkylated with iodoacetamide (IAM) which with the final concentration of 40 mM at room temperature for 40 min in the dark. After centrifugation (10,000 g, 20 min, 4℃), the pellet was resuspended in 100 μL TEAB (final concentration 100 mM). Trypsin was added at a 1:50 enzyme-to-protein ratio, and digestion was carried out overnight at 37℃. After digestion, peptides were vacuum dried and reconstituted in 100 μL 0.1% trifluoroacetic acid (TFA, w/v), desalted using HLB cartridges, and dried again. Peptide concentrations were measured using a NanoDrop One spectrophotometer (Thermo, USA) by UV absorption value.

DIA mass detection

Based on the peptide quantification results, peptides were analyzed on a VanquishNeo coupled with an Orbitrap Astral mass spectrometer (Thermo, USA) at Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China). Briefly, a 75 μm×5.5 cm uPAC High Throughput column (Thermo, USA) was equilibrated by solvent A (water with 2% ACN and 0.1% formic acid) and solvent B (water with 80% ACN and 0.1% formic acid). The chromatography run time was set to 8 minutes. Data-independent acquisition (DIA) data were acquired using an Orbitrap Astral mass spectrometer operated in DIA mode. The detection was carried out over a mass range of 70-1050 m/z (MS1), and 150-2000 m/z (MS2).

Protein identification

The DIA raw data were processed using Spectronaut v19 against a non-redundant metagenomic protein database constructed from the corresponding sediment samples. Search parameters were set as follows: the peptide length range was restricted to 7–52 amino acids; the enzyme specificity was trypsin/P with up to two missed cleavages; carbamidomethylation of cysteine was specified as a fixed modification, while methionine oxidation and protein N-terminal acetylation were set as variable modifications. Protein and peptide false discovery rates (FDR) were controlled at $\leqslant$1%, with peptide confidence $\geqslant$99% and XIC width $\leqslant$75 ppm. Protein quantification was performed using the MaxLFQ algorithm. All identified proteins were subject to taxonomic assignment, functional classification annotation using NCBI non-redundant protein database and Kyoto Encyclopedia of Genes and Genomes database (KEGG database).

Supplementary Figures

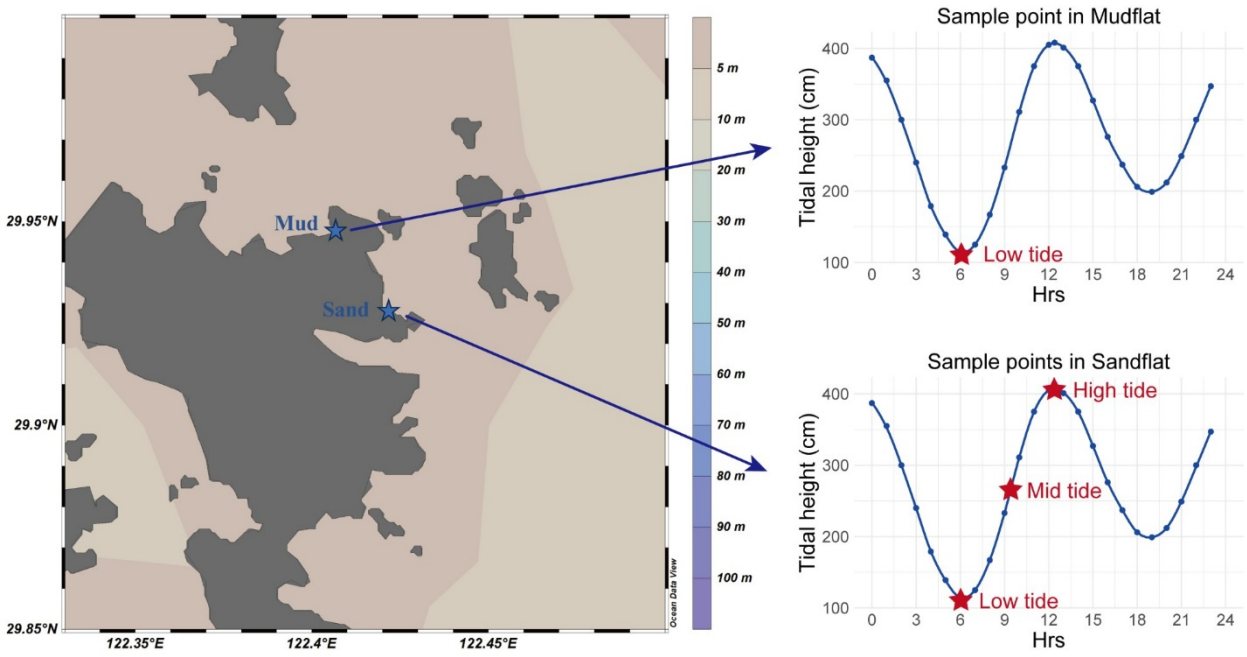

Fig. S1. Distribution of sampling points in Zhoushan Archipelago (left) and sampling strategies for mudflat and sandflat sediments (right). Sampling points on the sandflat were categorized based on tidal levels: low tide (SL), mid tide (SM), and high tide (SH). Mudflats were only accessible during low tide and samples were collected during low tide. Sediment cores of 30 cm in length were collected using acrylic plexiglass and were stratified into three depths: 1–5 cm, 5–15 cm, and 15–25 cm. The marigram on the right is illustrative and does not represent the exact times of sampling, and all samples were collected during daylight whenever possible to minimize the impact of light exposure.

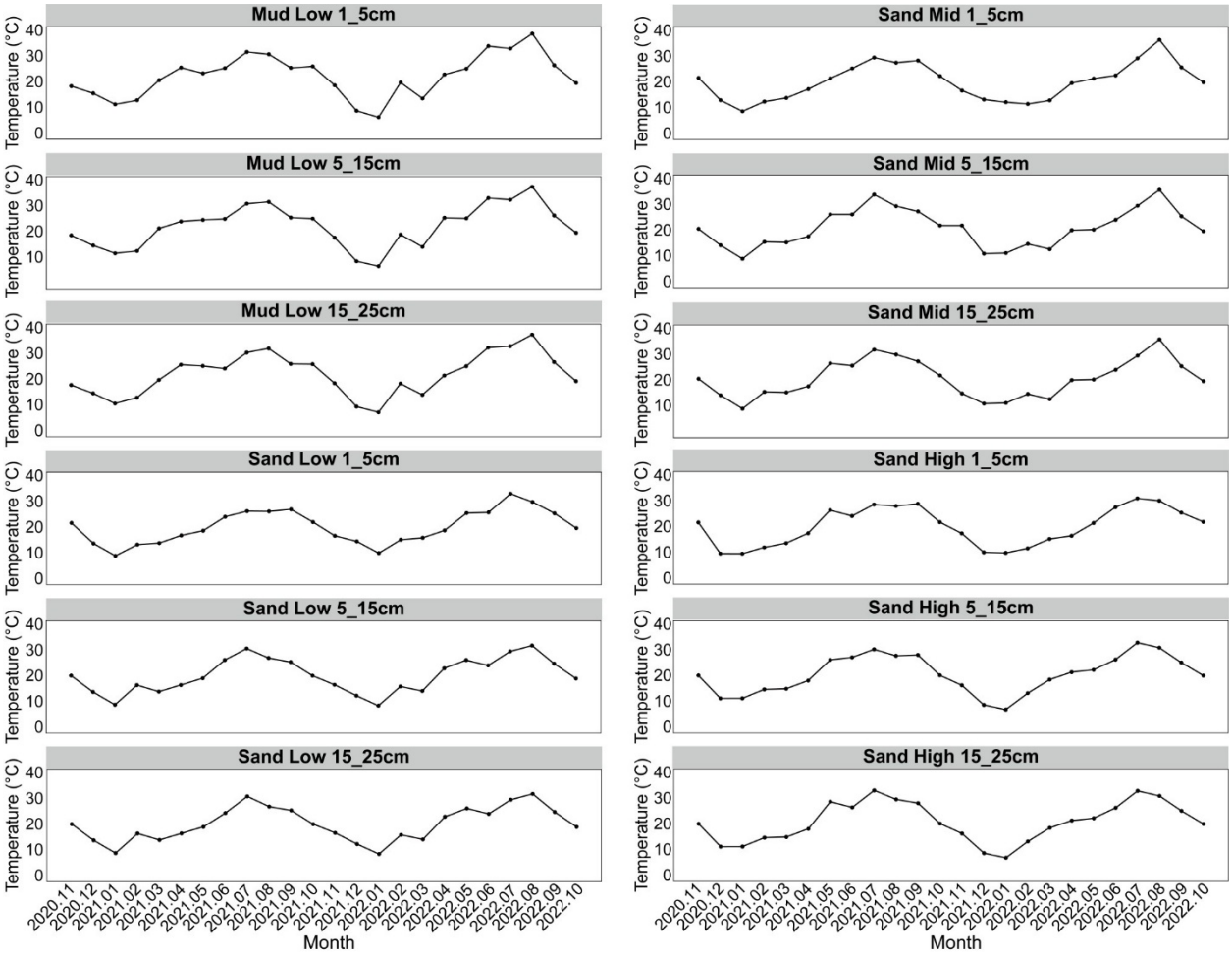

Fig. S2. Temperature fluctuations of sediment pore water across 12 groups of sediments.

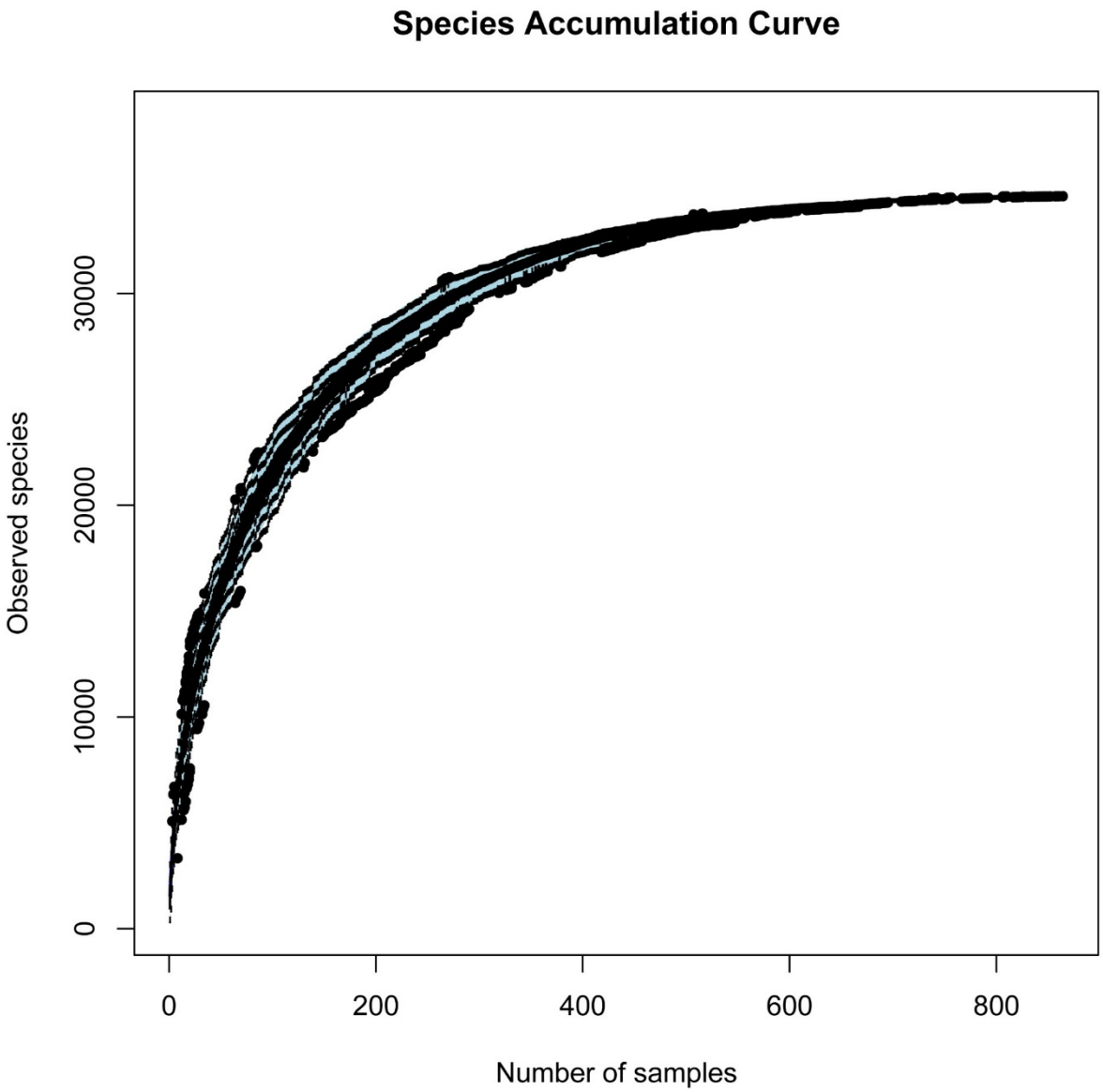

Fig. S3. Species accumulation curves plotted against the number of samples, illustrating the observed species richness.

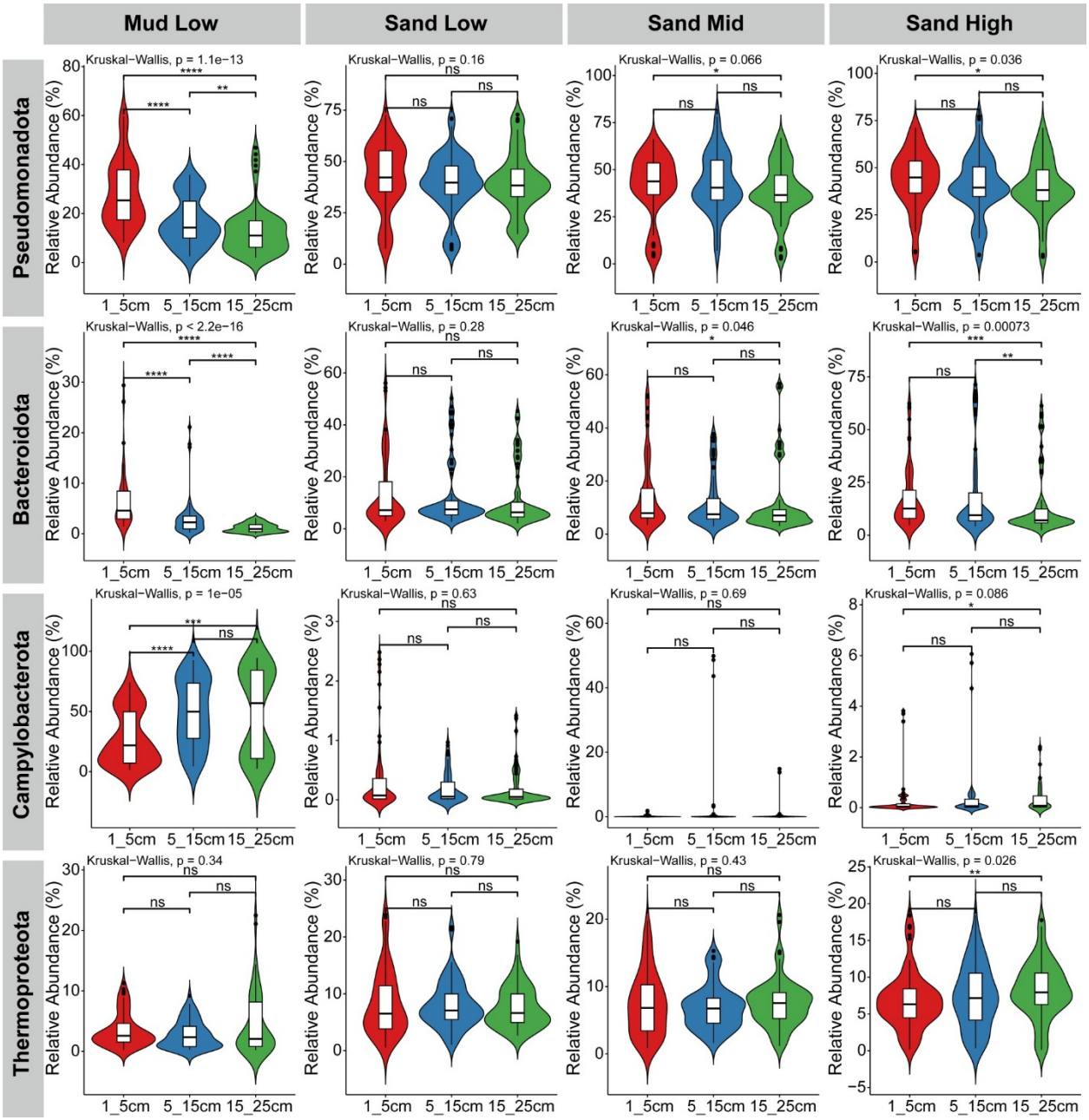

Fig. S4. Variations in the abundance of the four most abundant phyla in unvegetated tidal flat sediments at different depths. The statistical significance of median differences between groups was assessed using the Wilcoxon rank-sum test. Significance levels are indicated as ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns, no significant difference.

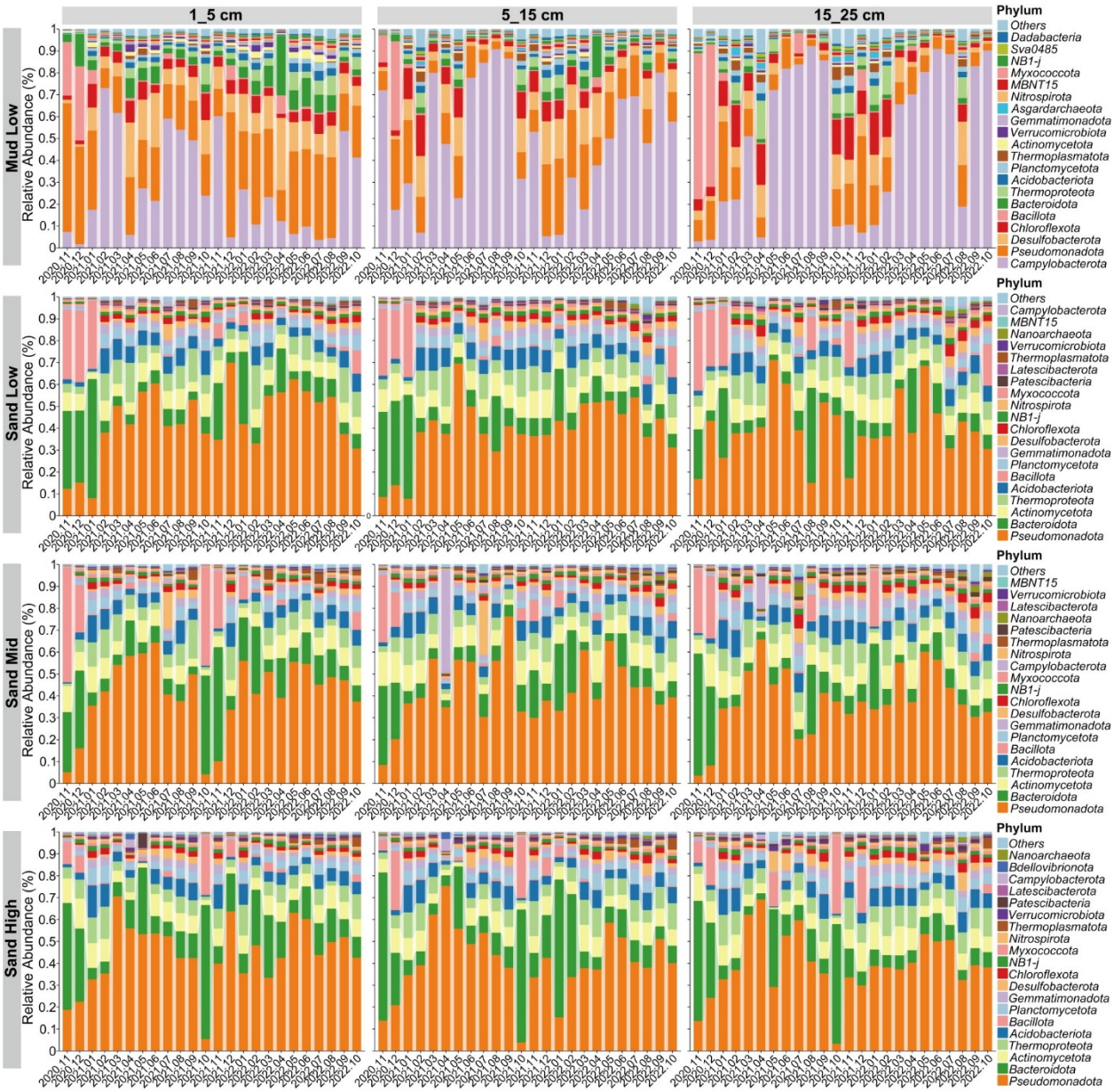

Fig. S5. Microbial composition of the top 20 phyla in unvegetated tidal flat sediments over 24 months across 12 groups of sediments.

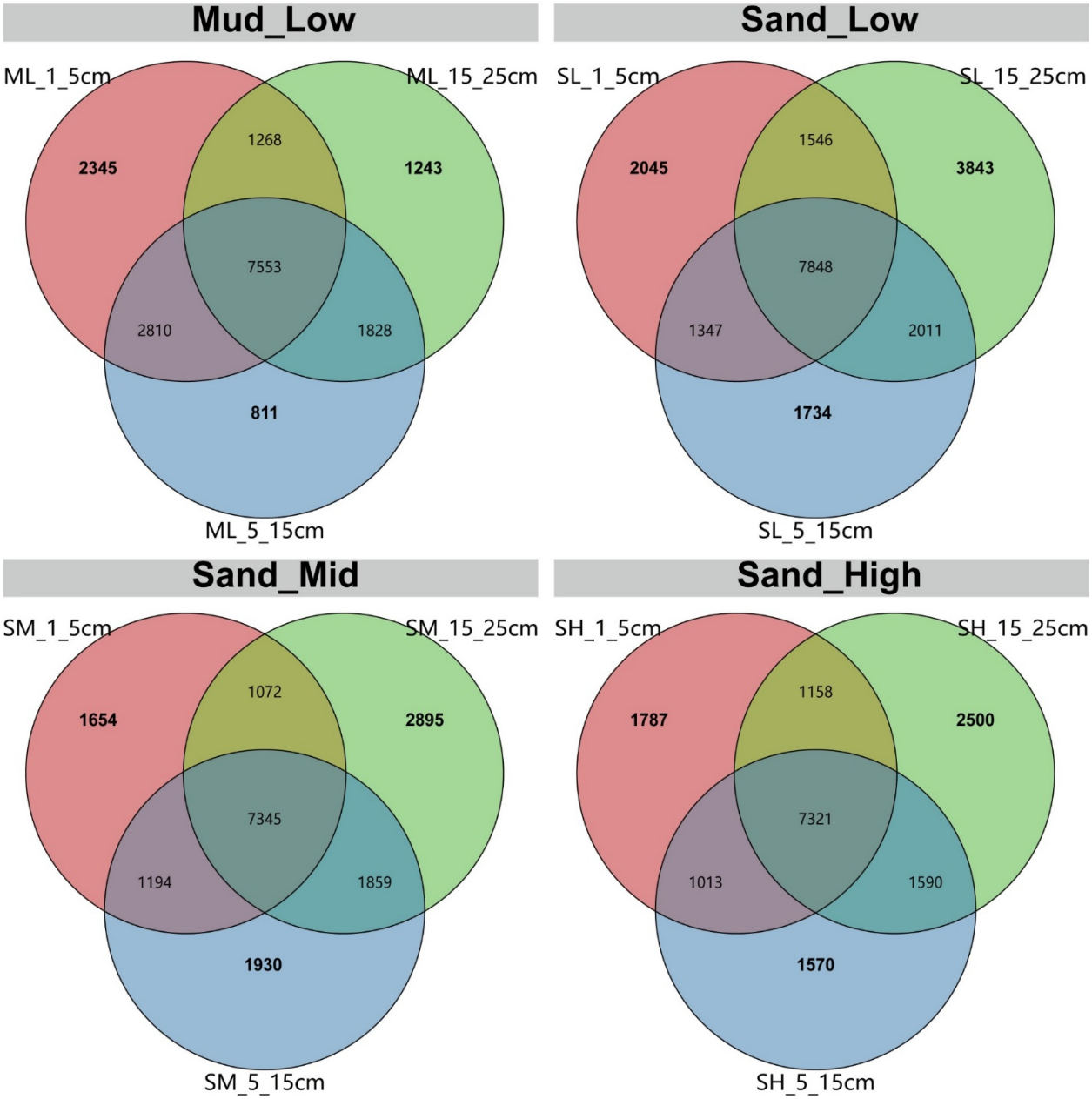

Fig. S6. Venn diagrams showing the number of unique and shared ASVs across different groups. The 12 groups of sediments represent samples collected from mudflat low tide (ML), sandflat low tide (SL), sandflat mid tide (SM), and sandflat high tide (SH) at three sediment depths (1–5 cm, 5–15 cm, and 15–25 cm).

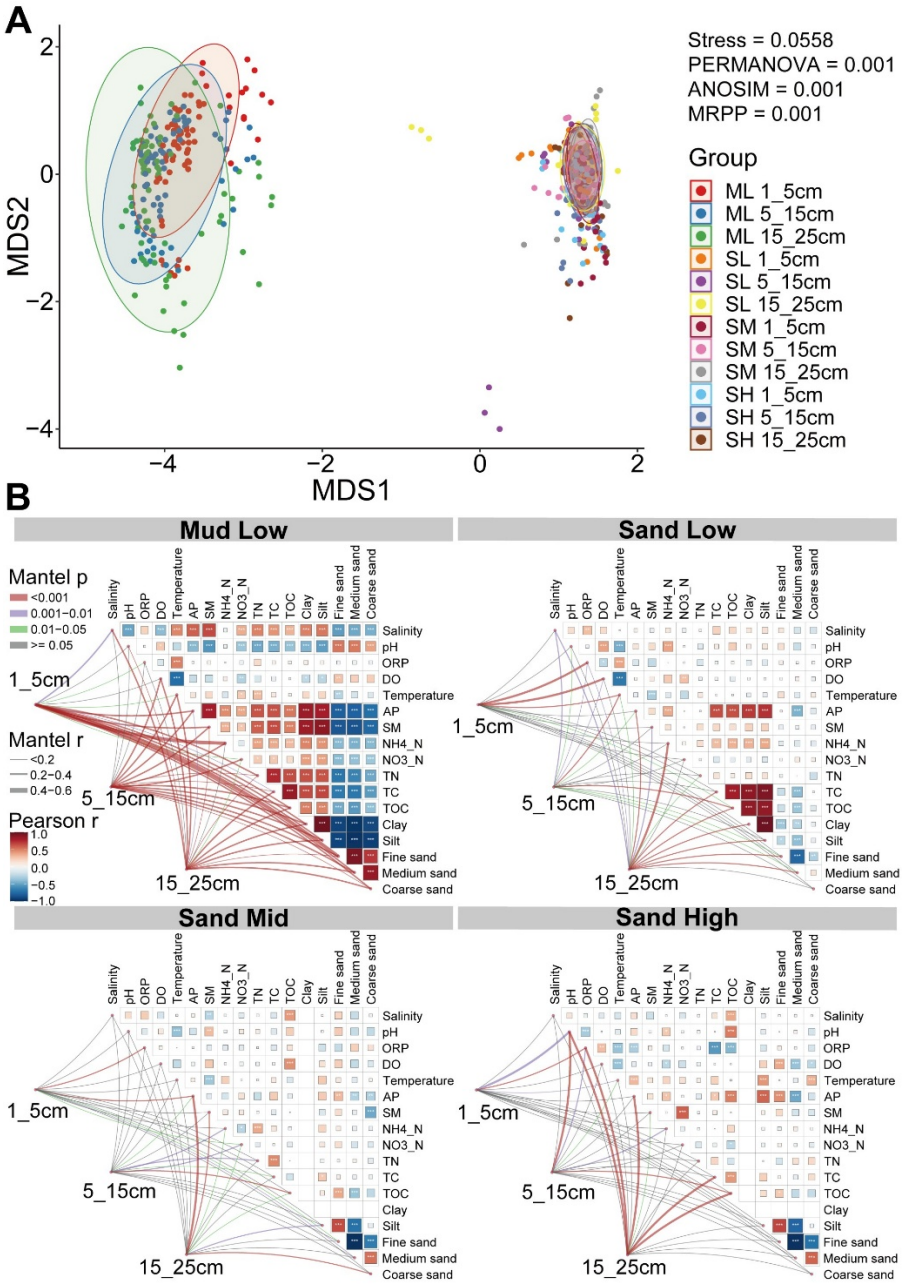

Fig. S7. Variations and driving factors of microbial community structure in unvegetated tidal flat sediments. A, Non-multidimensional scaling (NMDS) ordination based on Bray–Curtis distance reveals compositional variations of microbial communities across 12 groups of sediments. The 12 groups of sediments represent samples collected from mudflat low tide (ML), sandflat low tide (SL), sandflat mid tide (SM), and sandflat high tide (SH) at three sediment depths (1–5 cm, 5–15 cm, and 15–25 cm). **B**, Pairwise comparisons of physicochemical factors and Mantel tests assessing the correlation between physicochemical factors and microbial compositions. The color gradient represents Pearson's correlation coefficients. Significance levels are indicated as ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, unmarked, no significant difference. Edge width corresponds to the Mantel's r value, and edge color denotes the statistical significance of the correlation.

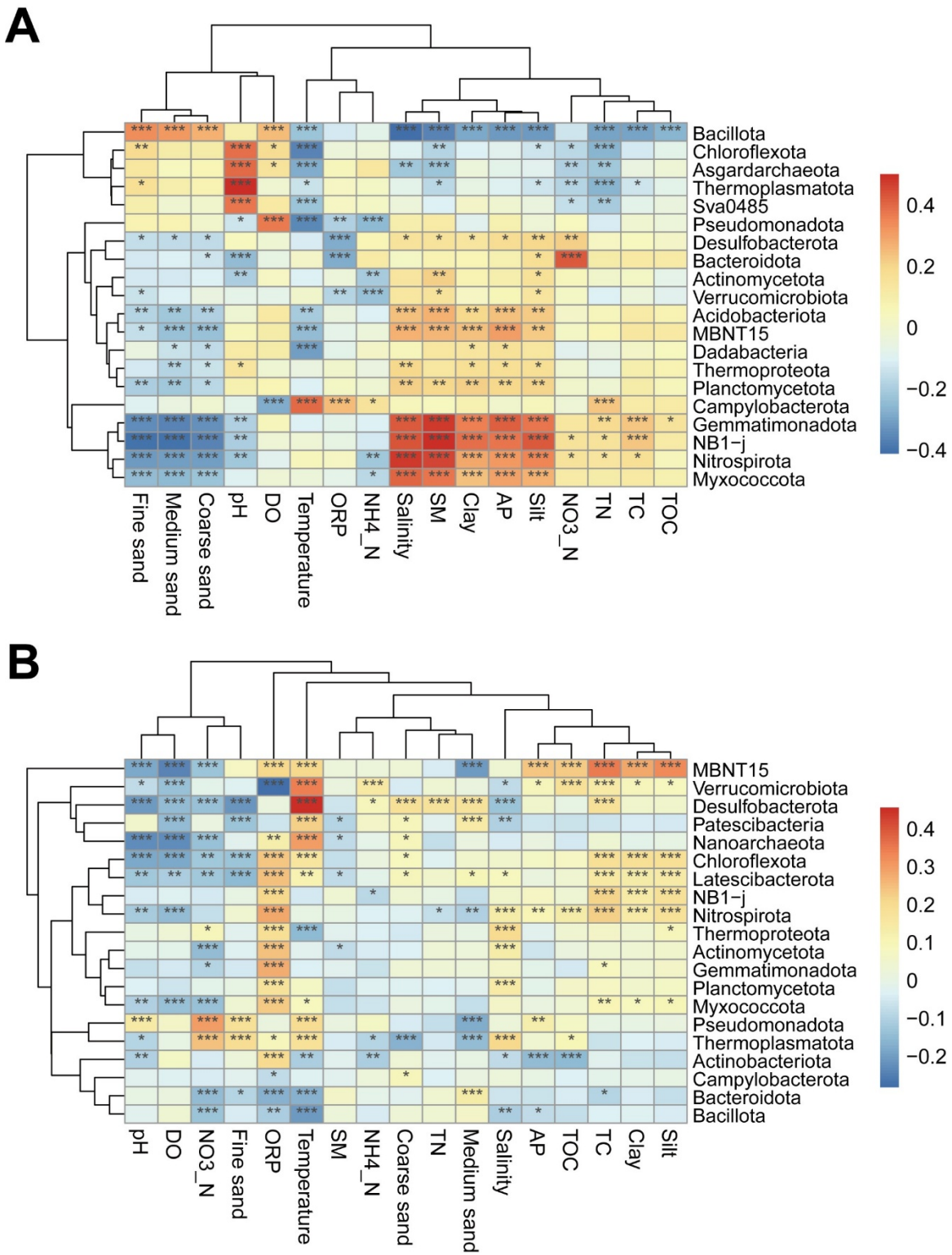

Fig. S8. Correlation between the top 20 phyla and physicochemical factors in mudflats (A) and sandflats (B). Clustering of species and physicochemical parameters was performed using Euclidean distance. The color gradient represents Pearson's correlation coefficients. Significance levels are indicated as ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, unmarked, no significant difference.

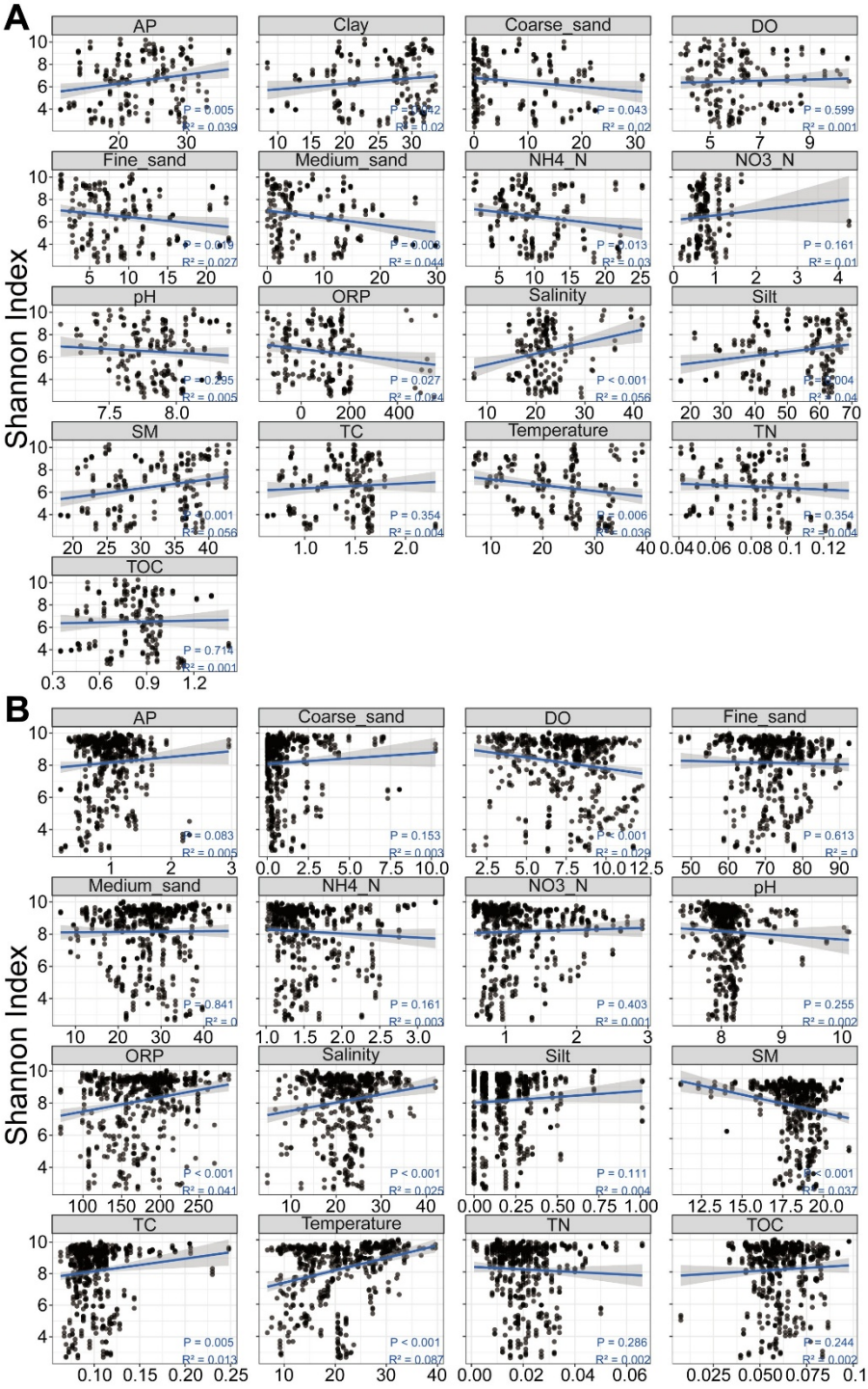

Fig. S9. Relationships between physicochemical factors and Shannon index in mudflats (A) and sandflats (B). Each subplot represents a different physicochemical factor, with data points indicating the observed values. Notably, the relationship between clay content and Shannon index was not plotted in sandflats as sandflat sediments lack of clay. A linear regression line is fitted to each dataset, with the R^2 and p values displayed in the bottom right of each subplot. p values less than 0.001 are denoted as $p < 0.001$.

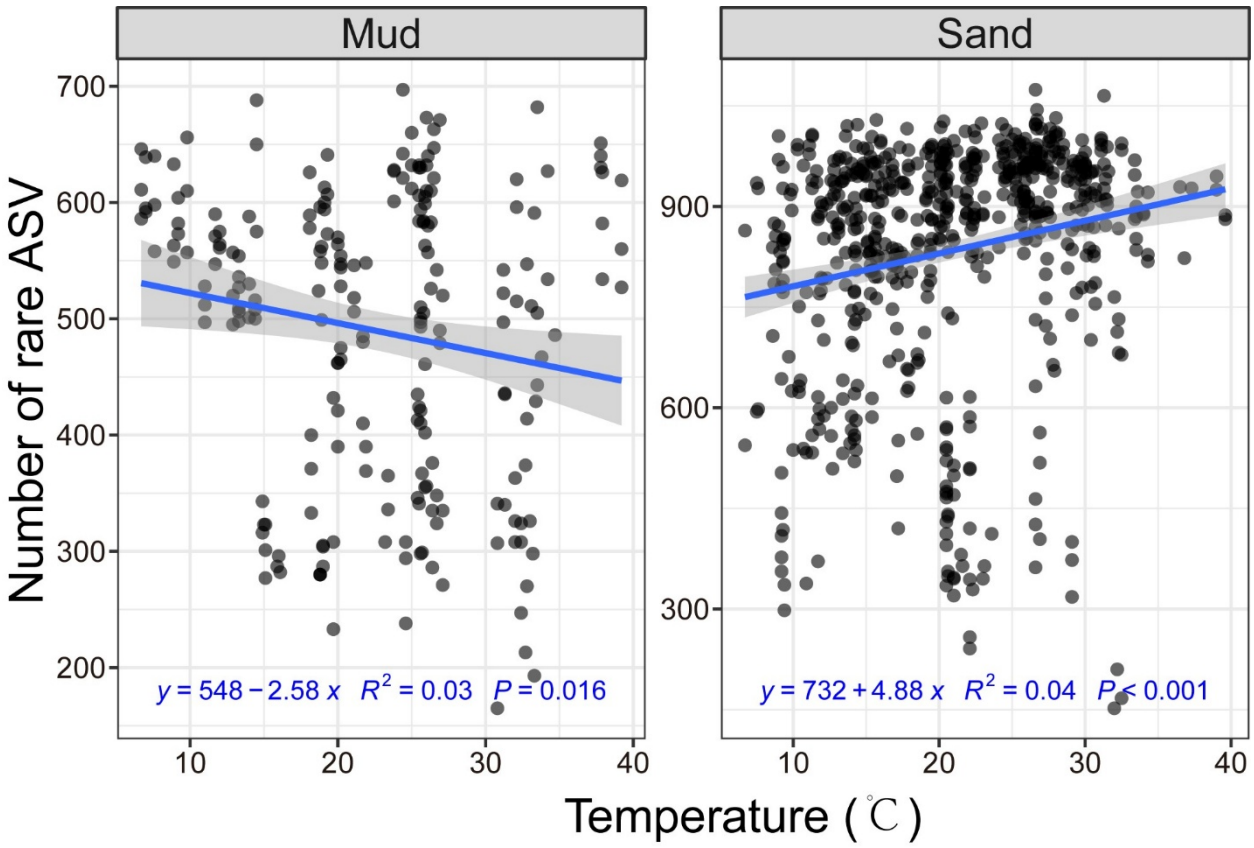

Fig. S10. Relationships between temperature and the number of rare ASVs in mudflats (left) and sandflats (right). A linear regression line is fitted to each dataset, with the R^2 and *p* values displayed in the bottom of each subplot. Data points indicating the observed values. *p* values less than 0.001 are denoted as $p < 0.001$.

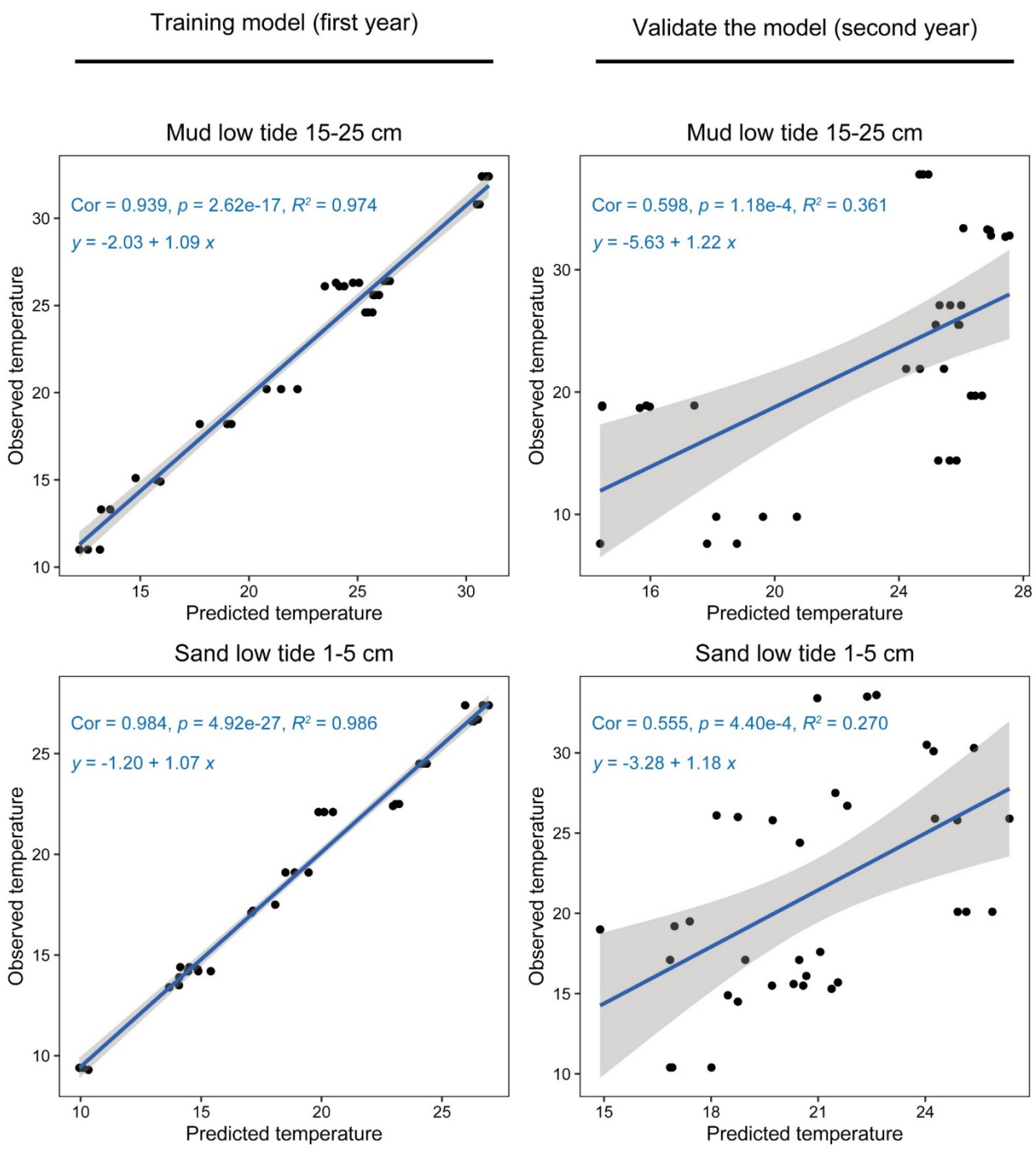

Fig. S11. Predicted temperatures from the random forest regression model compared to actual temperatures in ML15–25 and SL1–5 sediments. Linear regression analyses between predicated and observed temperatures were conducted to evaluate the accuracy of the random forest predictions in the training dataset (first year) and the test dataset (second year) for both mudflats and sandflats. A blue linear regression line is fitted to each dataset.

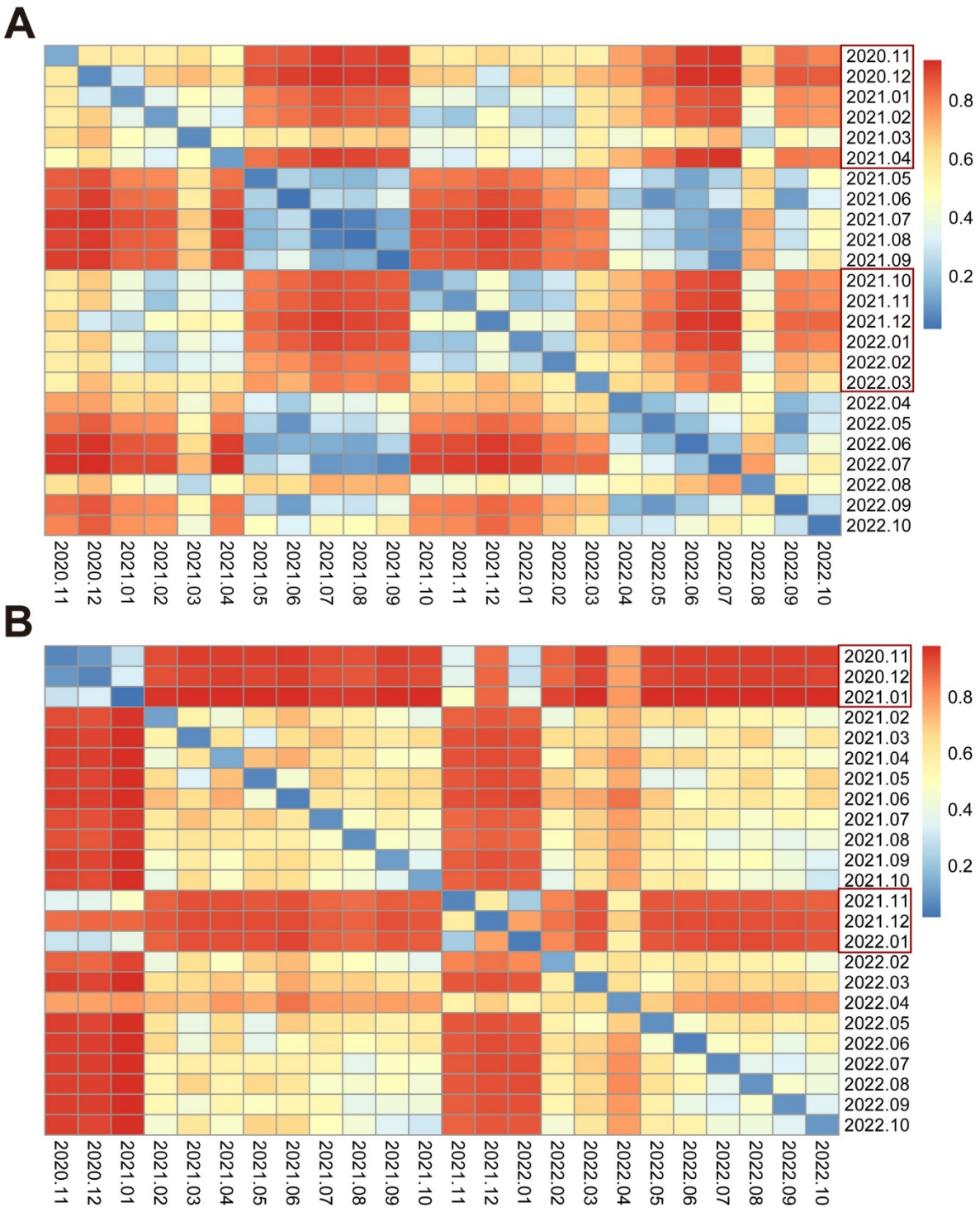

Fig. S12. Dissimilarity of microbial community structures across 24 months in mudflat low tide 15–25 cm sediments (ML15–25, A) and sandflat low tide 1–5 cm (SL1–5, B) sediments based on Bray–Curtis distance. Three replicate samples were collected each month to calculate the Bray–Curtis distance. The color gradient represents community dissimilarity between each group.

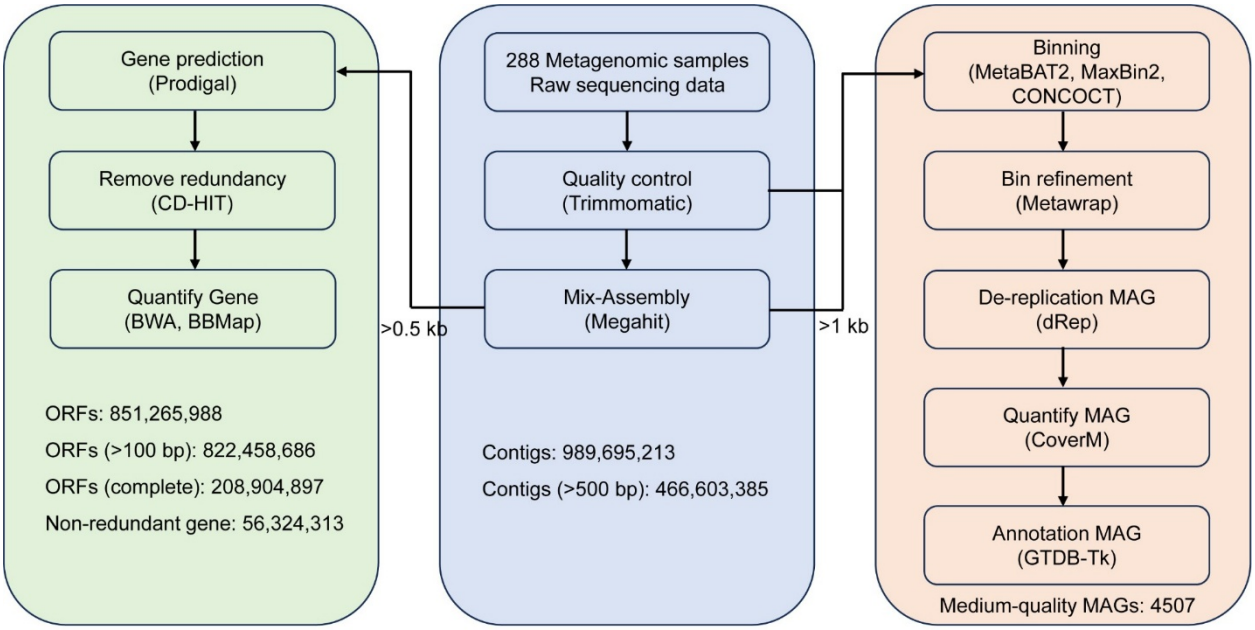

Fig. S13. Computational pipeline for assembly-based metagenomic analysis. Each panel illustrates the number of contigs, ORFs, and MAGs generated at key steps of the process.

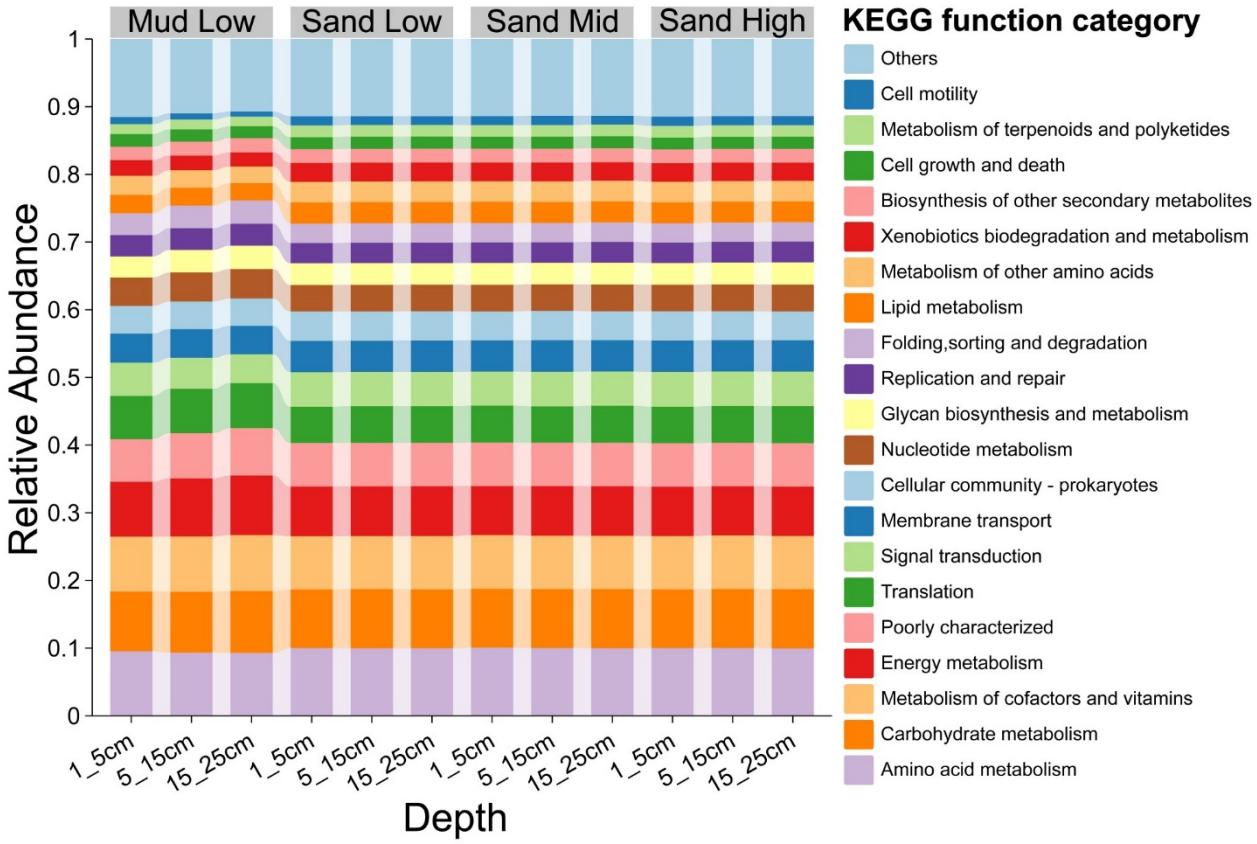

Fig. S14. Functional classification of the genes across 12 groups of sediments based on the annotation results of KEGG functional categories at the second hierarchy level. Only the top 20 categories from each group are displayed.

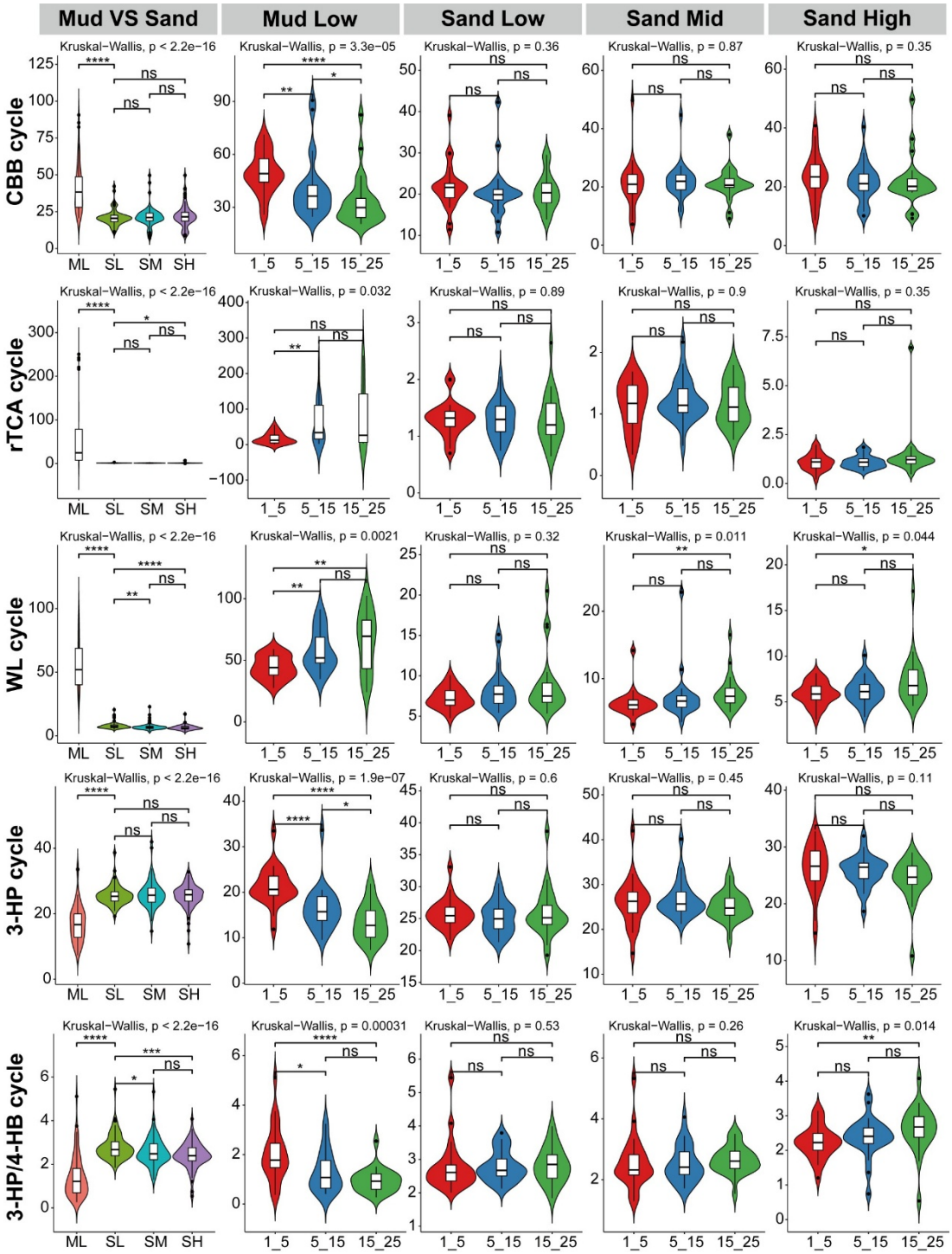

Fig. S15. Variations in the abundance of five CFPs observed in unvegetated tidal flat sediments. The average abundance of key genes within each CFP was used to represent the CFP abundance. ML, mudflat low tide; SL, sandflat low tide, SM, sandflat mid tide, SH, sandflat high tide. The statistical significance of median differences between groups was assessed using the Wilcoxon rank-sum test. Significance levels are indicated as ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns, no significant difference.

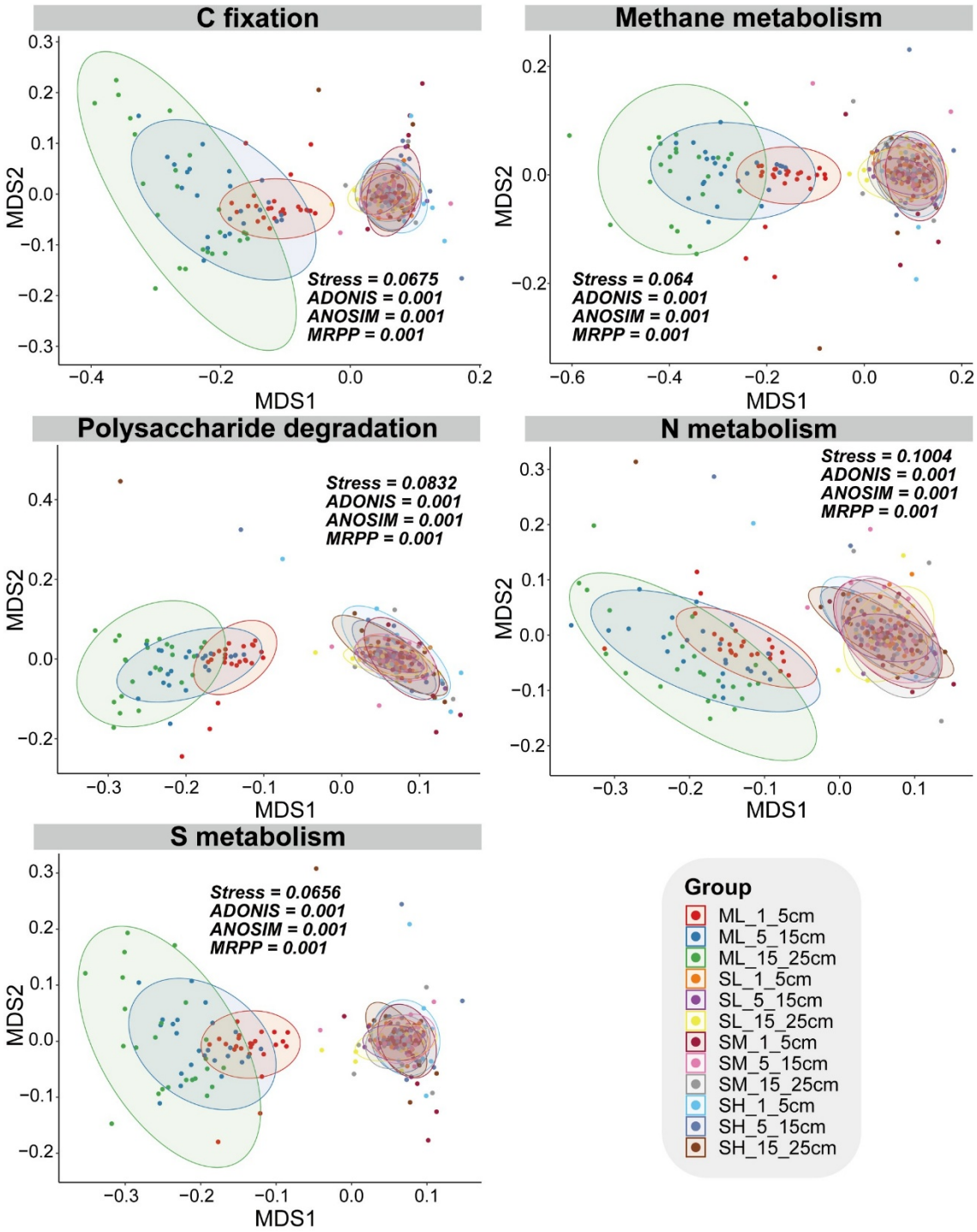

Fig. S16. Non-multidimensional scaling (NMDS) ordination based on Bray–Curtis distance reveals compositional variations of genes associated with carbon fixation, methane metabolism, polysaccharide degradation, nitrogen metabolism, and sulfur metabolism across 12 groups of sediments. The 12 groups of sediments represent samples collected from mudflat low tide (ML), sandflat low tide (SL), sandflat mid tide (SM), and sandflat high tide (SH) at three sediment depths (1–5 cm, 5–15 cm, and 15–25 cm).

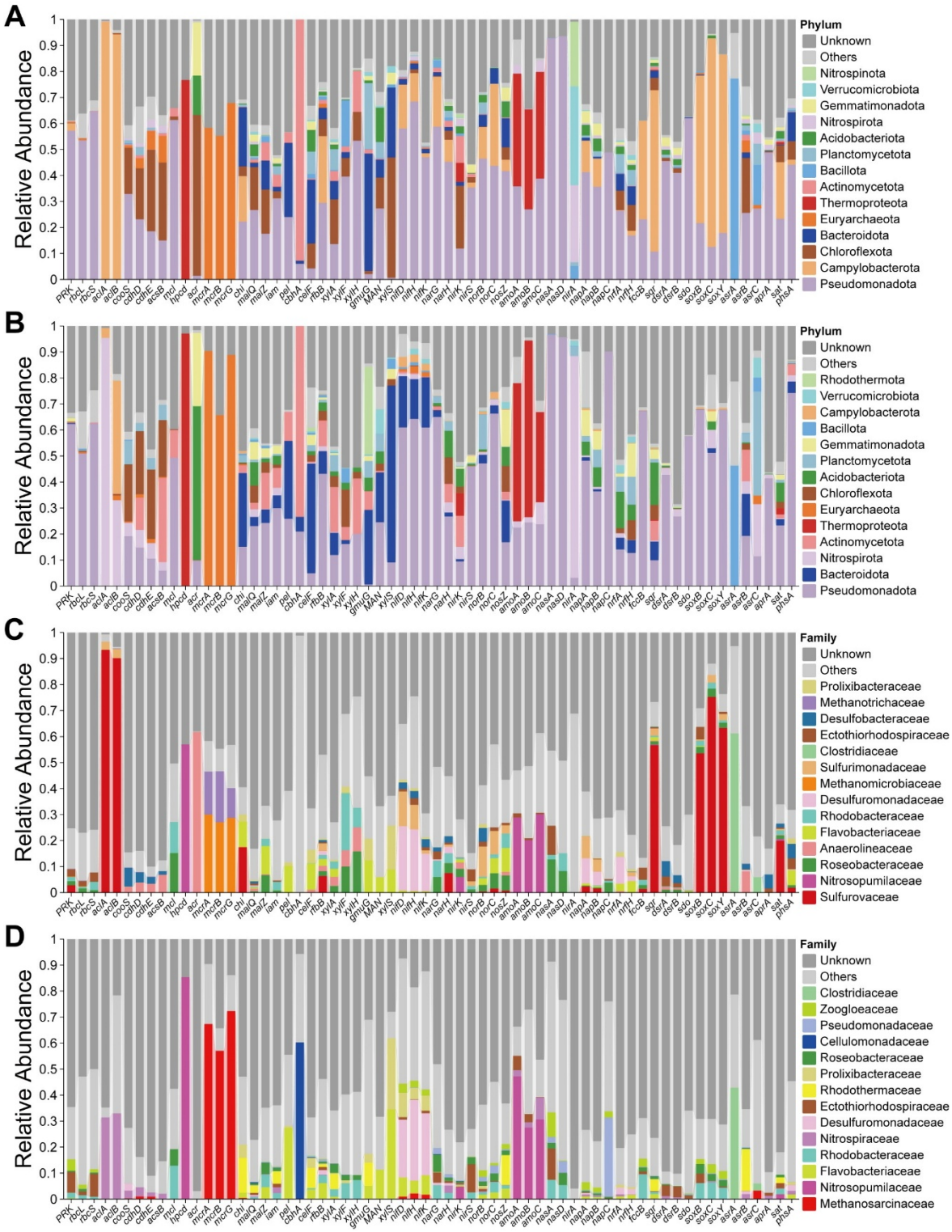

Fig. S17. Taxonomic assignments for key genes involved in carbon, nitrogen, and sulfur cycling in mudflats (A, phylum level; C, family level) and sandflats (B, phylum level; D, family level). The colored bars represent the relative abundance of key genes assigned to each taxon. Only the top 15 taxa are displayed.

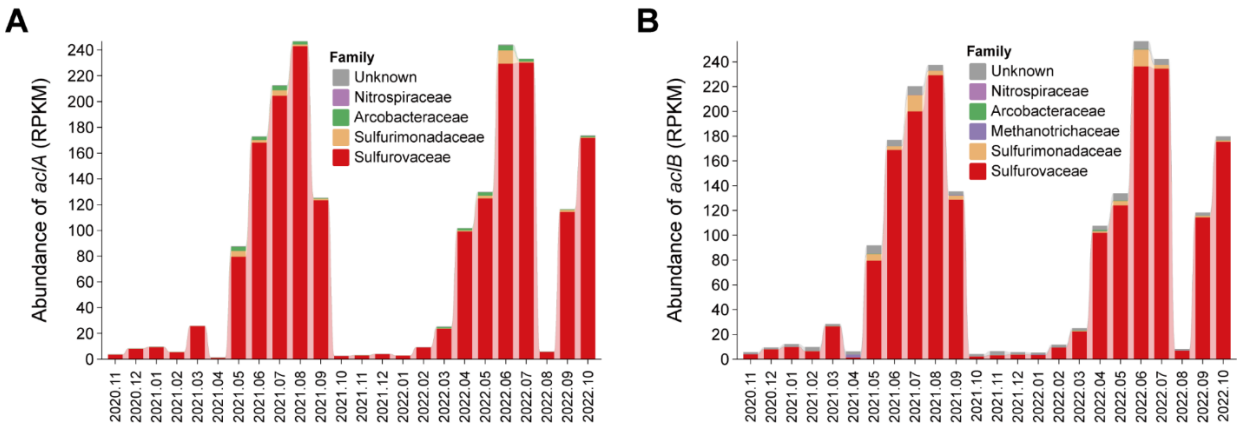

Fig. S18. Taxonomic distribution of *aclA*/*aclB* genes in ML15–25 sediments over 24 months. A,B, Taxonomic assignments and abundance fluctuations of *aclA* (A) and *aclB* (B) genes over 24 months in ML15–25 sediments. Colored bars indicate the abundance of each taxon.

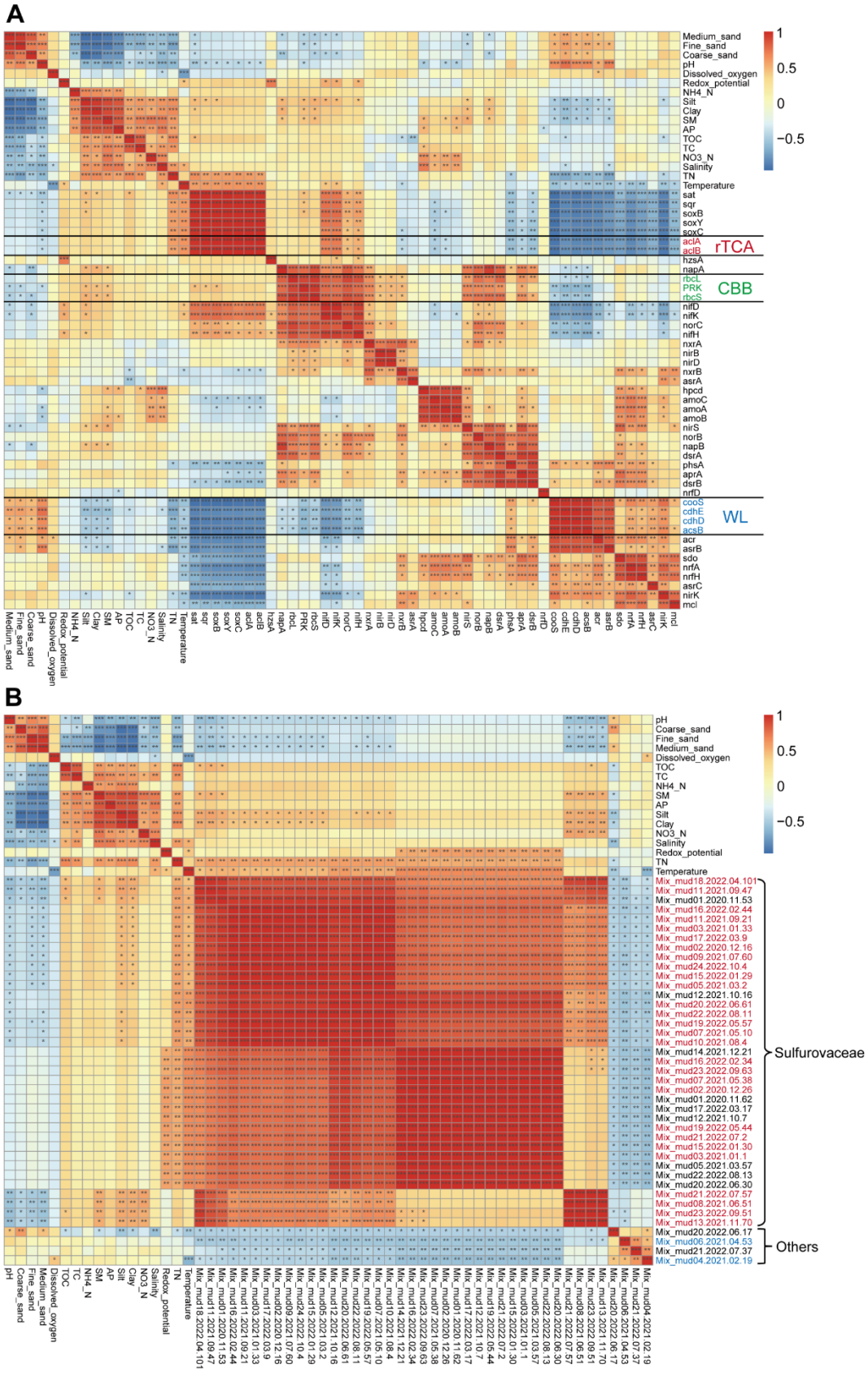

Fig. S19. Pearson correlations between physicochemical factors and C, N, S cycling genes (A)/high-abundance MAGs (B) in ML15–25 sediments. The color gradient in the heatmap tiles represents Pearson correlation coefficients as indicated in the legend. Significance levels: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$; unmarked, not significant. Genes and MAG IDs involved in the rTCA, CBB, and WL pathways are labeled in red, green, and blue, respectively, on the right side of the heatmaps.

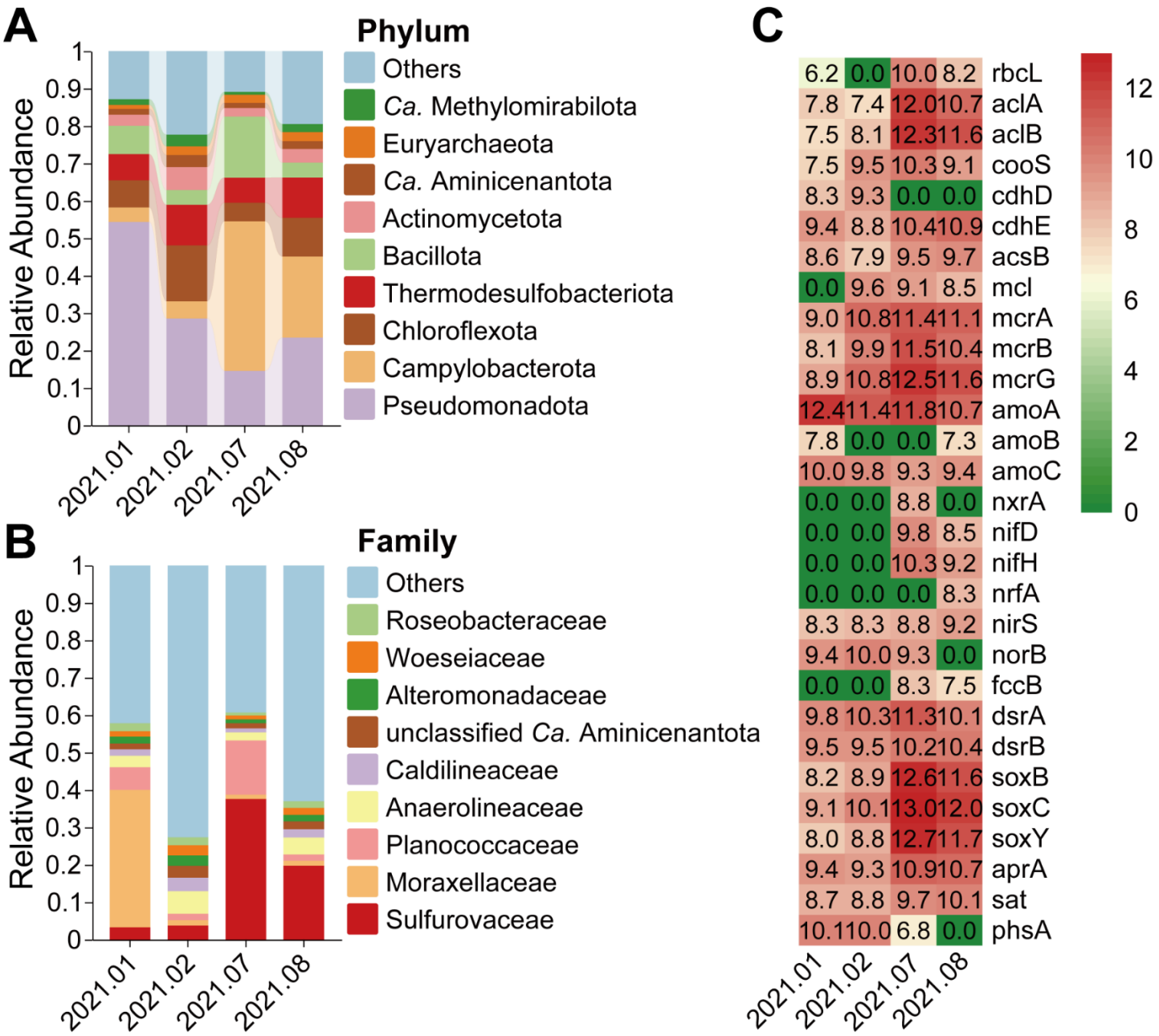

Fig. S20. Taxonomic and functional profiles of proteins detected in ML15–25 sediments during winter and summer. A,B, Protein taxonomic composition at the phylum level (A) and family level (B). **C,** Heatmap showing the abundance of proteins involved in C, N, and S cycling. Protein intensities were transformed to log2 fold-changed values (with a pseudocount of 1) as indicated in the legend.

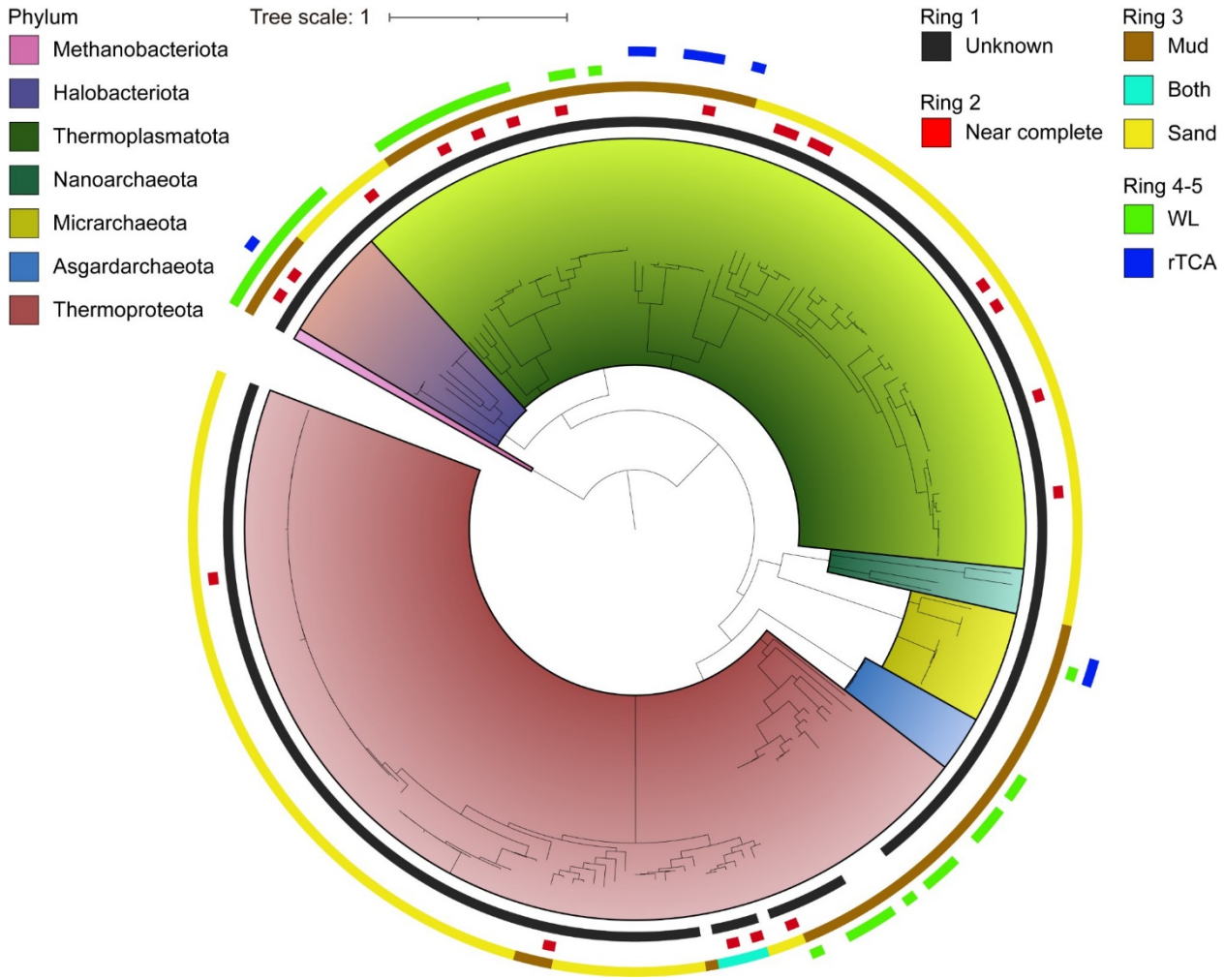

Fig. S21. Phylogenetic tree of the 211 archaeal MAGs constructed based on the 53 concatenated conserved archaeal marker genes. Clade background colors indicate different phyla. The scale bar represents 1 amino acid substitution per position. Genome information, MAG sources, and associated CFPs are displayed in the outer layers. Mud, MAGs only present in mudflats or their abundances in mudflats are more than 10 times higher than in sandflats. Sand, MAGs only present in sandflats or their abundances in sandflats are more than 10 times higher than in mudflats. Both, MAGs present in both mudflats and sandflats with abundance differences less than 10-fold.

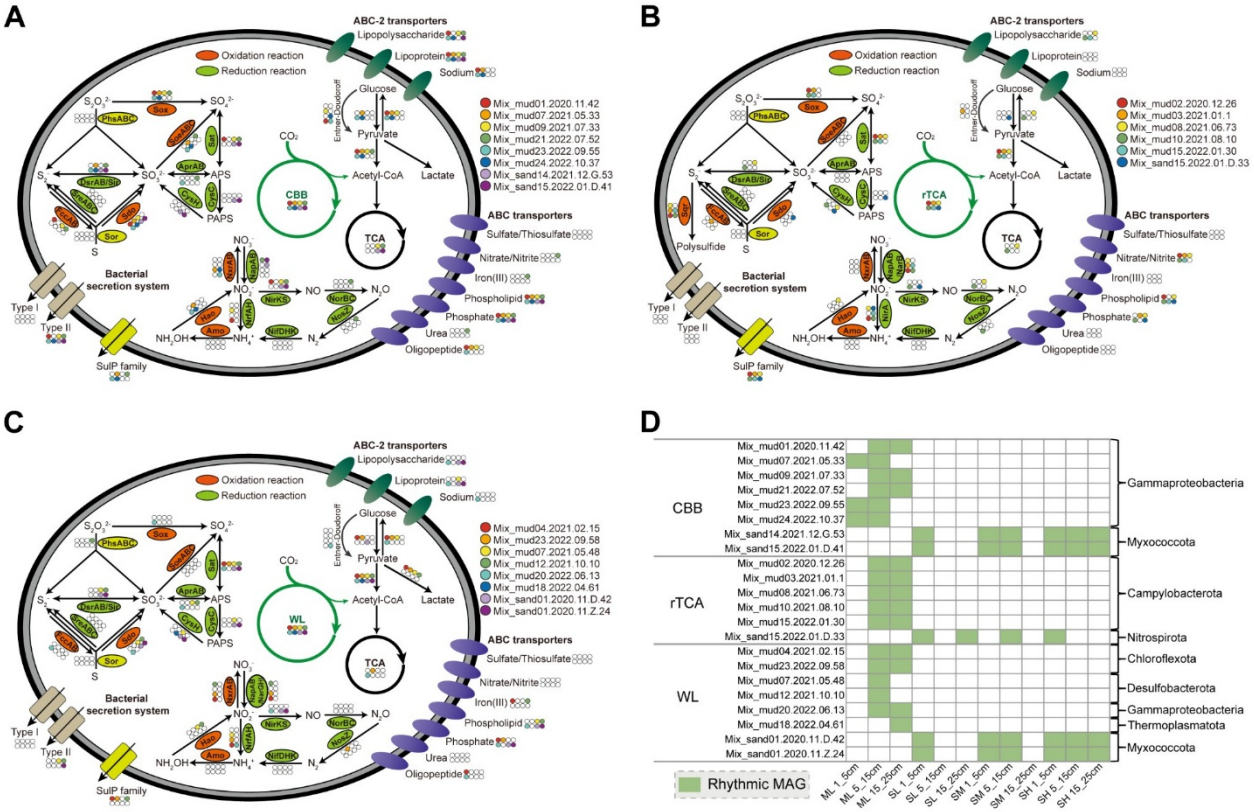

Fig. S22. Metabolic reconstructions of representative chemoautotrophic and rhythmic MAGs in unvegetated tidal flat sediments. A–**C**, Metabolic capacities of representative chemoautotrophic and rhythmic MAGs associated with the CBB cycle (**A**), the rTCA cycle (**B**), and the WL pathway (**C**). White dots indicate absence of the corresponding genes. **D**, Tables indicating the rhythmic distribution of these representative chemoautotrophic MAGs across 12 groups of sediments. The taxonomic assignments of these MAGs are shown on the right, with detailed assignments provided in **Supplementary data 13**.

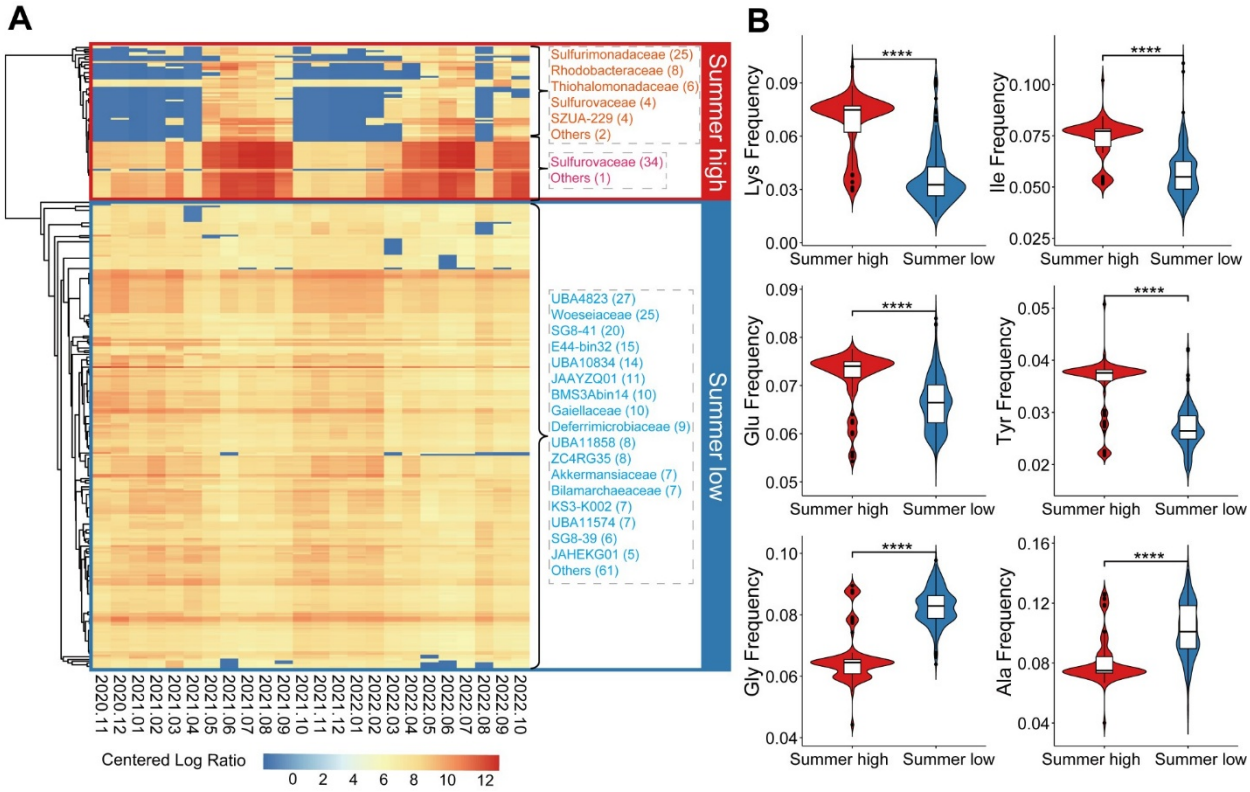

Fig. S23. Two groups of MAG with distinct rhythmic patterns and their seasonal adaptions in mudflat low tide 15–25 cm sediments (ML15–25). A, Cluster heatmaps of centered logarithm ratio abundances of 341 rhythmic MAGs in ML15–25 sediments. Clustering of MAGs was performed using Pearson distance with average linkage method. Rhythmic MAGs were categorized into two major clusters based on their monthly abundance patterns, with the number of rhythmic MAGs classified to each family within the clusters annotated on the right of the heatmaps. **B**, Divergence in temperature-adapted amino acid frequencies between the two MAG clusters in ML15–25 sediments. The statistical significance of median differences between groups was assessed using the Wilcoxon rank-sum test. Significance levels are indicated as ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns, no significant difference. Lys, Lysine; Ile, Isoleucine; Glu, Glutamate; Tyr, Tyrosine; Gly, Glycine; Ala, Alanine.