

Supplementary Information for

Adaption mechanism and ecological role of CPR bacteria in

brackish-saline groundwater

Authors: Jiawen Wang^{1,2,3†}, Haohui Zhong^{1,2†}, Qian Chen^{1,2*}, Jinren Ni^{1,2}

Author affiliations:

¹Key Laboratory of Water and Sediment Sciences, Ministry of Education, Beijing 100871,
P. R. China

²State Environmental Protection Key Laboratory of All Material Fluxes in River
Ecosystems, Beijing 100871, China

³School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an,
Shaanxi Province 710049, China

[†]These authors contributed equally: Jiawen Wang, Haohui Zhong

***Correspondence:** Qian Chen, College of Environmental Sciences and Engineering,
Peking University, Beijing 100871, P. R. China. E-mail: qianchen@pku.edu.cn

Supplementary Files includes:

I. Supplementary File 1:

Supplementary Figure 1 to16.

II. Supplementary File 2:

Supplementary Table 1 to 12.

III. Supplemental File 3:

Supplementary alignment file for Figure 1.

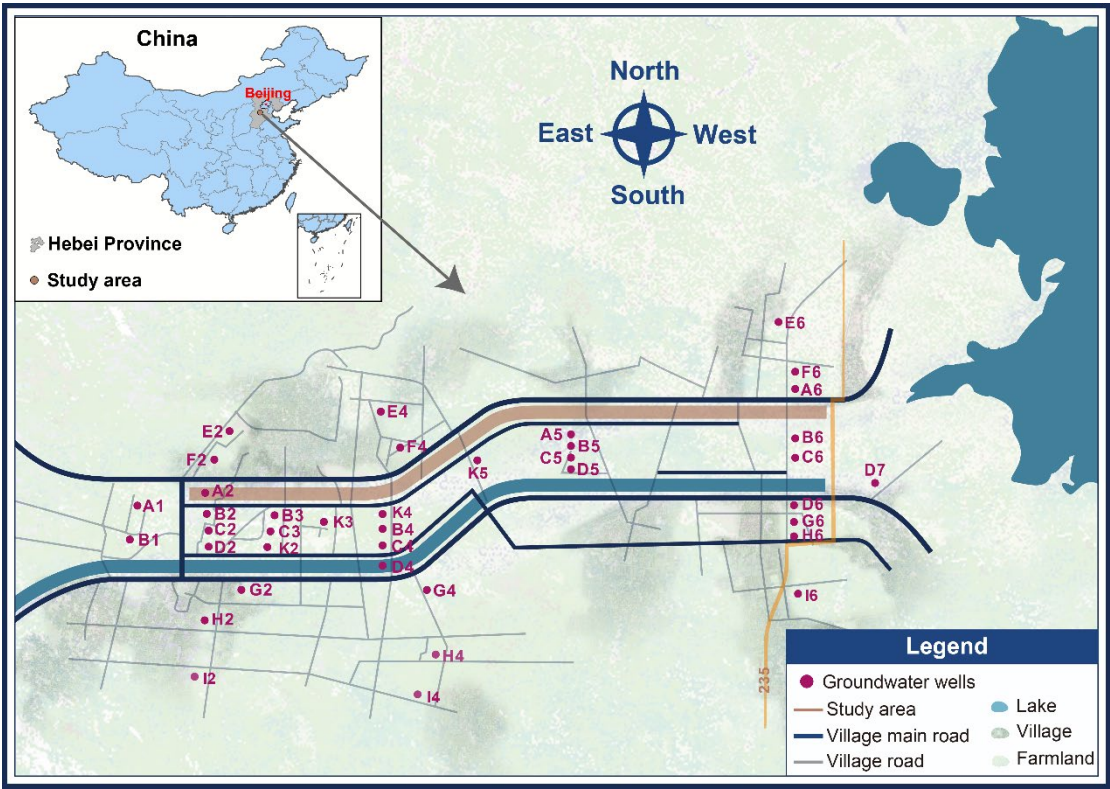
IV. Supplementary File 4:

Supplementary alignment file for Supplementary Figure 2.

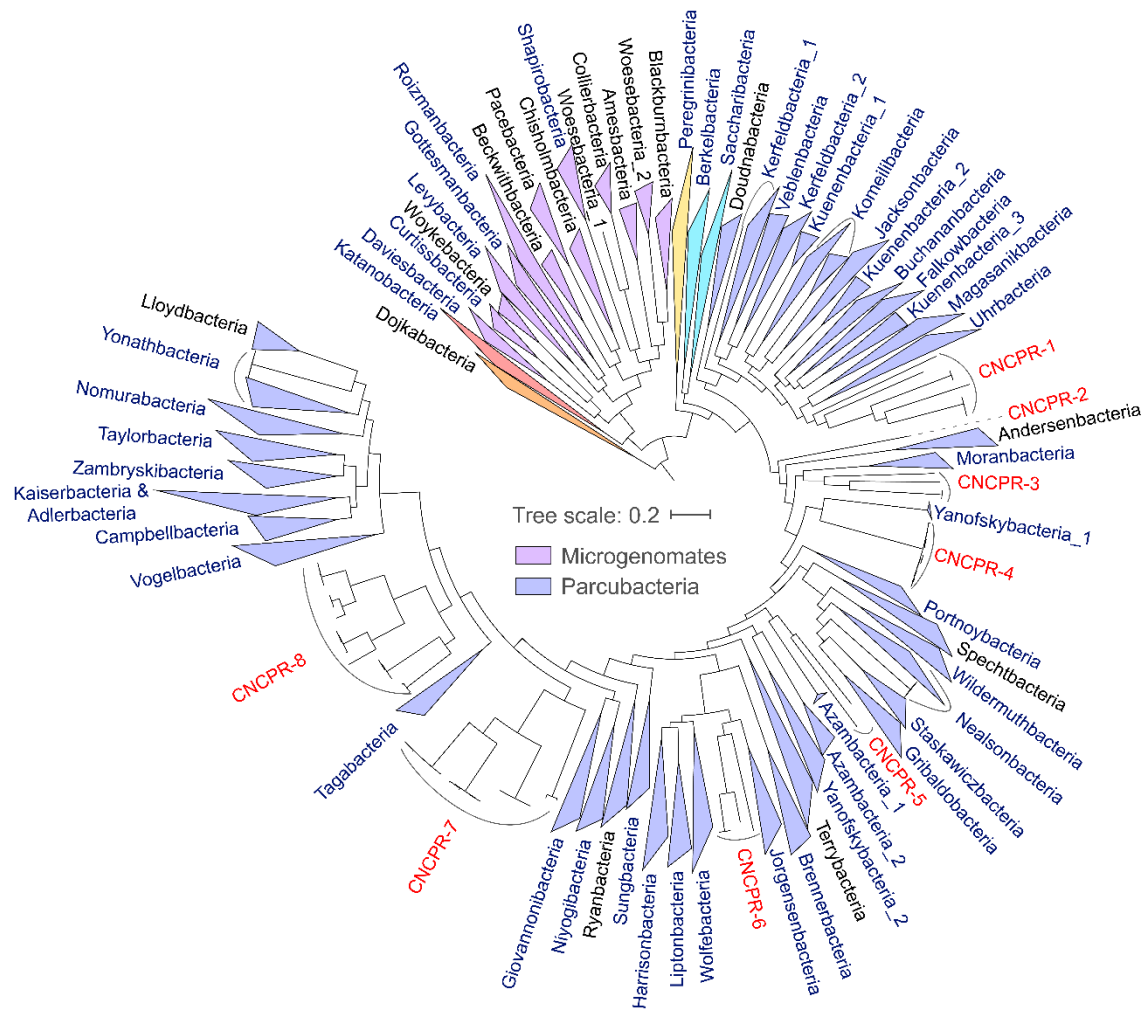
V. Supplementary File 5:

Supplementary alignment file for Supplementary Figure 3.

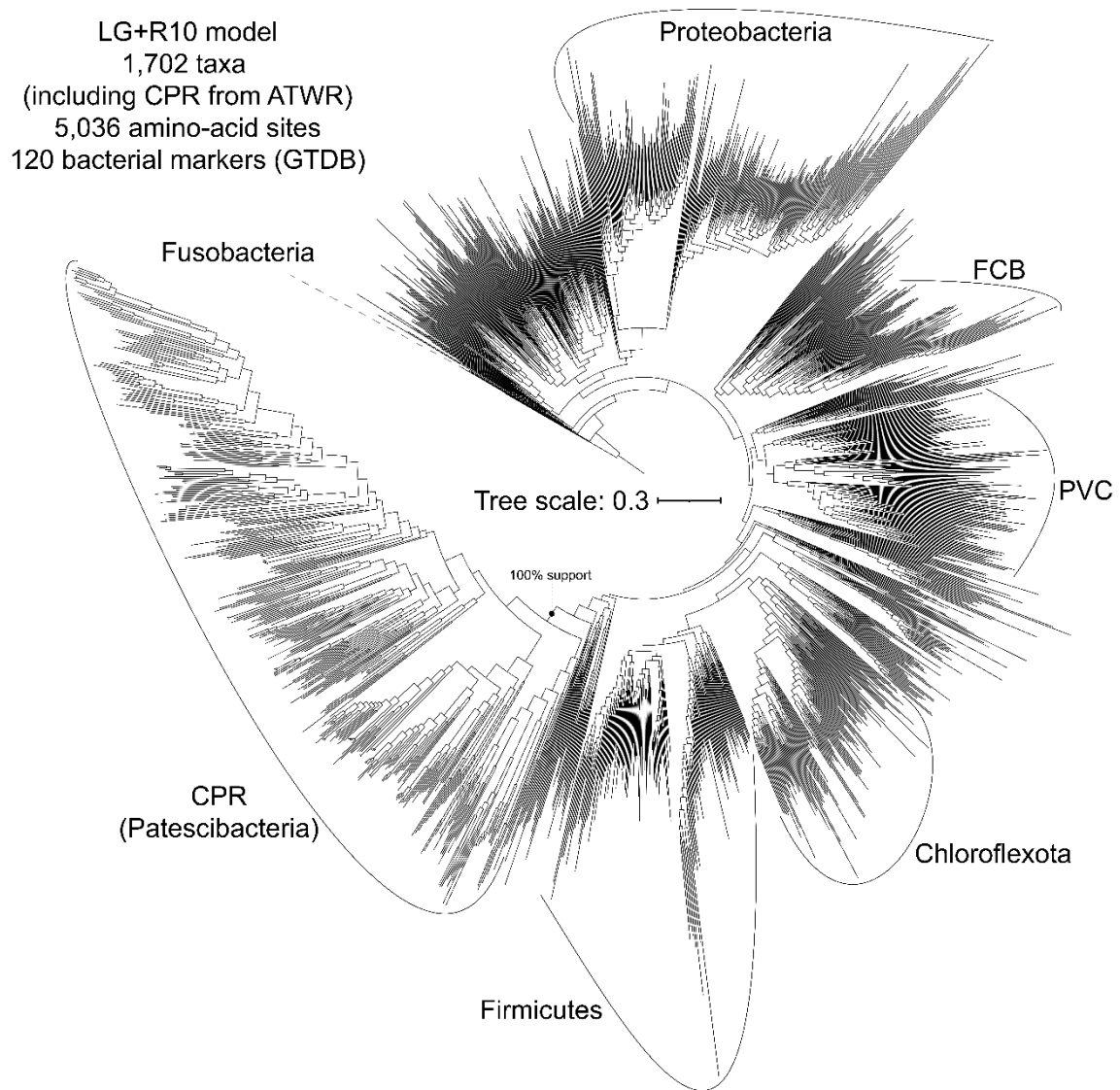
I. Supplementary File 1:



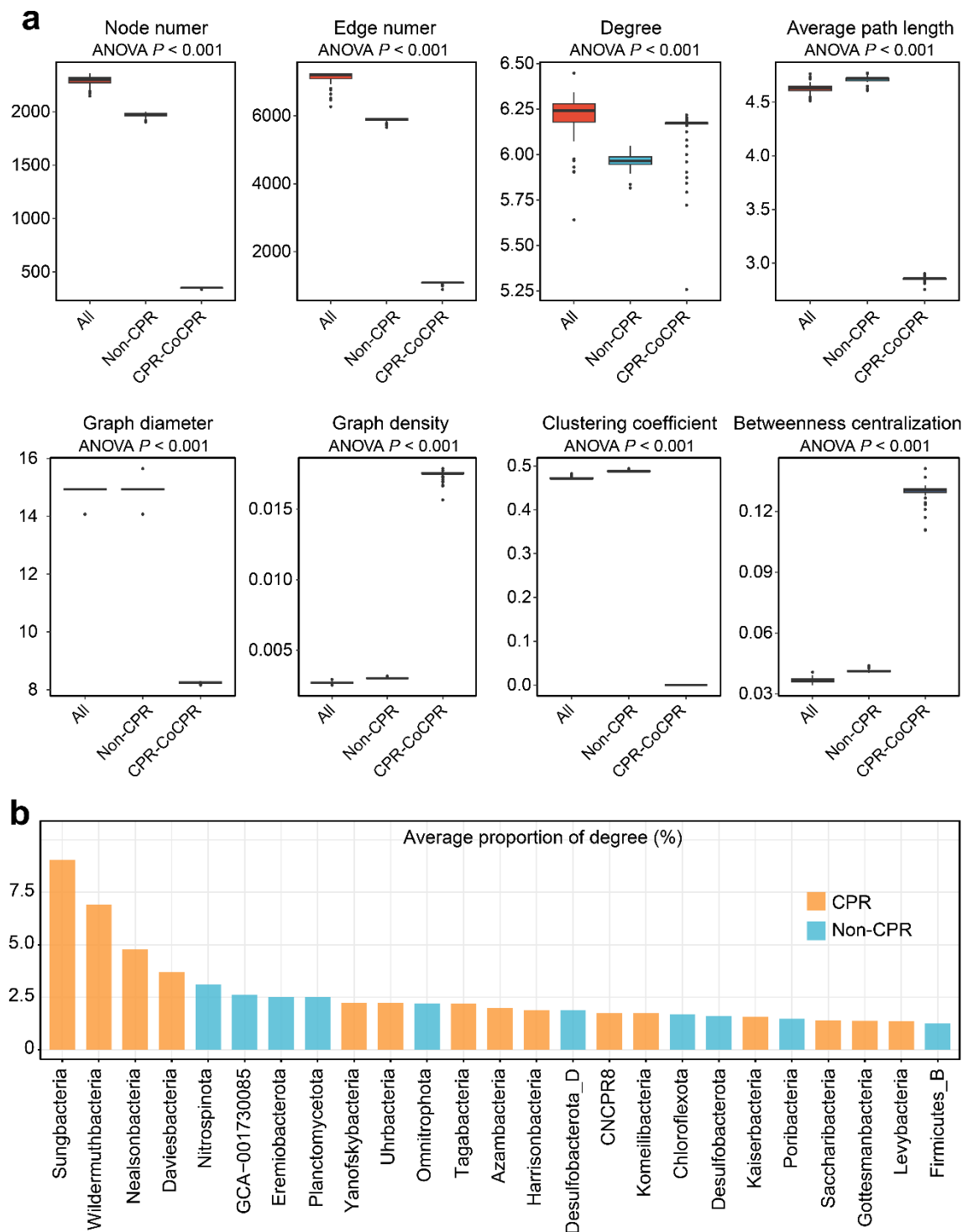
Supplementary Figure 1. Location and sampling wells in this study.



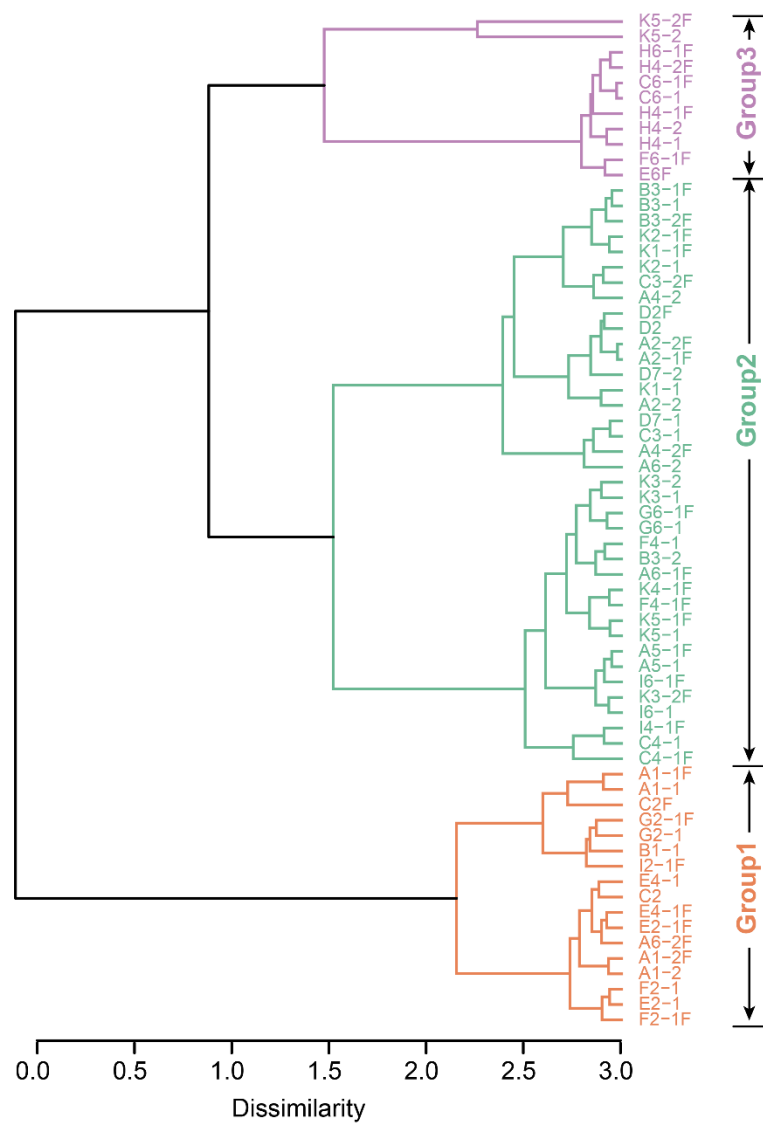
Supplementary Figure 2. The maximum likelihood tree was inferred from a concatenated alignment of 120 marker genes from GTDB-Tk for a dereplicated set of strain-representative ATWR-MAGs and publicly available reference genomes. The eight phyla colored in red are candidate novel CPR (CNCPR) affiliating to the superphylum Parcubacteria.



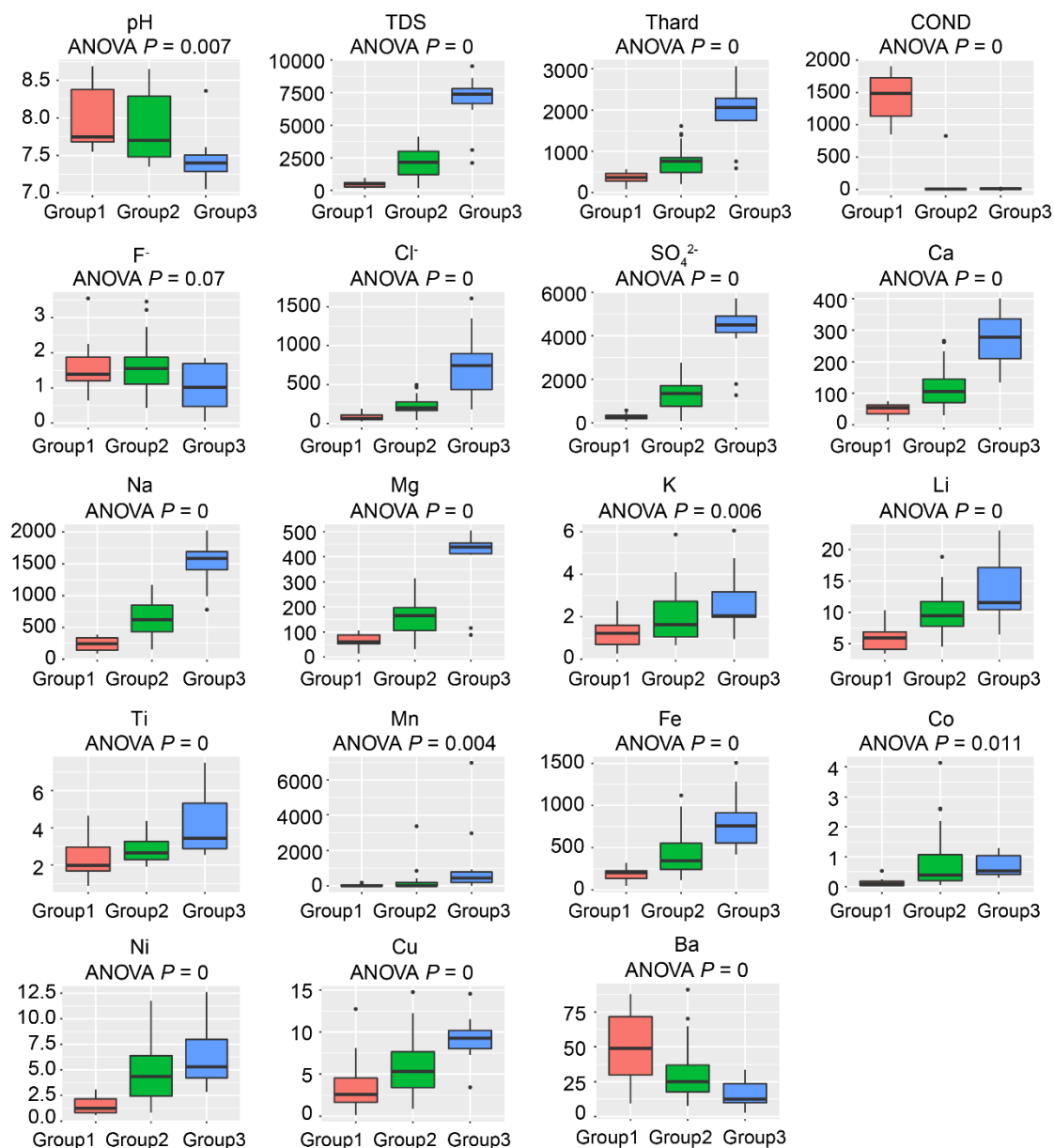
Supplementary Figure 3. The maximum-likelihood bacterial phylogeny inferred from a concatenation of 120 bacterial markers selected by GTDB based on 5,036 informative amino-acid sites under the best fitting model LG+R10. 1,702 taxa were included in this analysis. 399 of taxa were CPR genomes from ATWR. CPR superphylum is 100% supported as a monophyletic group. This tree is rooted between Fusobacteria and all other phyla according to a recent rooted bacterial evolution analysis (Coleman et al., 2021).



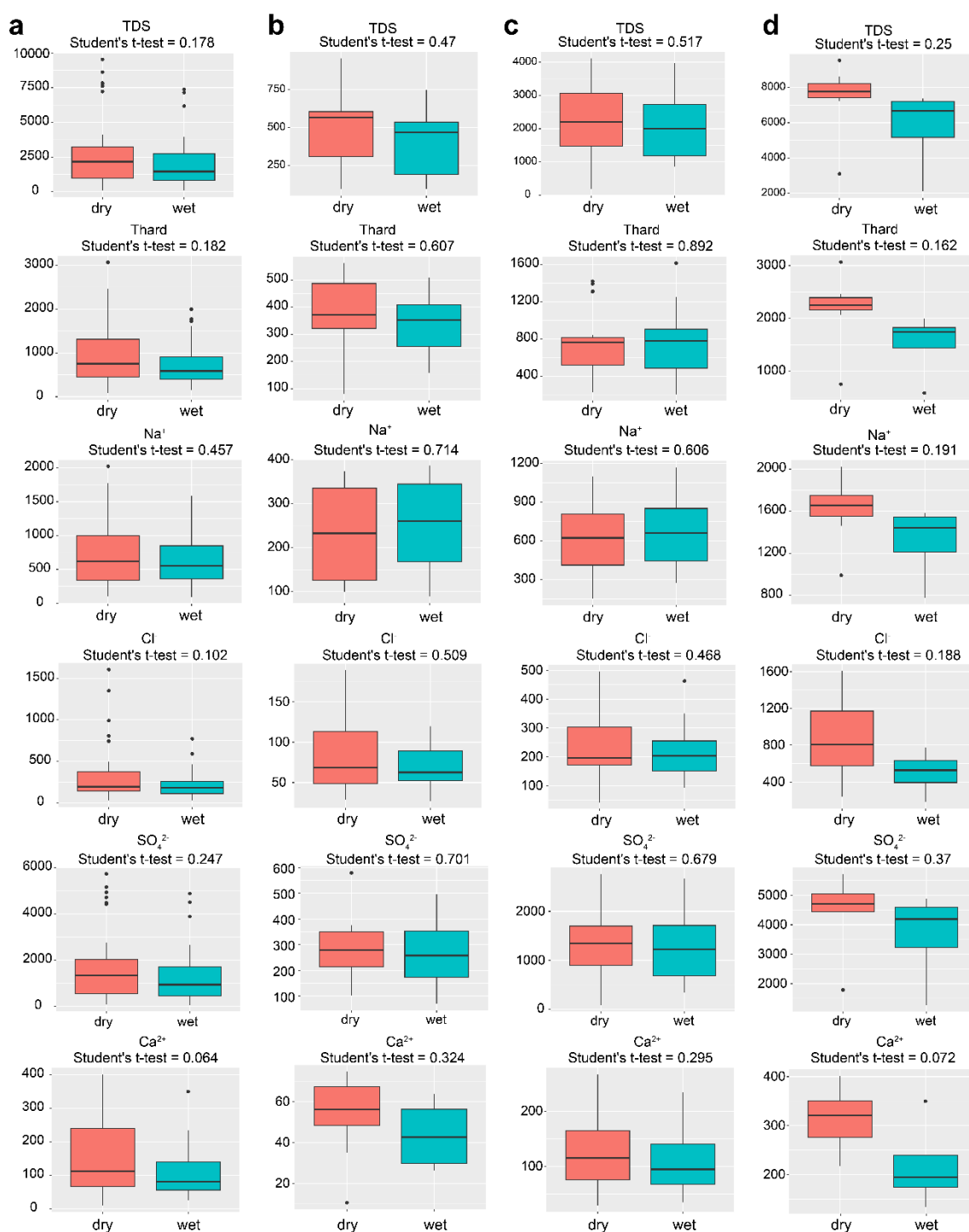
Supplementary Figure 4. Co-occurrence network characteristics. (a) Topological properties of the co-occurrence networks for all bacteria (All), non-CPR bacteria (Non-CPR), as well as CPR and Co-CPR (CPR-CoCPR). (b) Proportion of the average node degree for the dominant phyla in the co-occurrence networks for all bacteria.



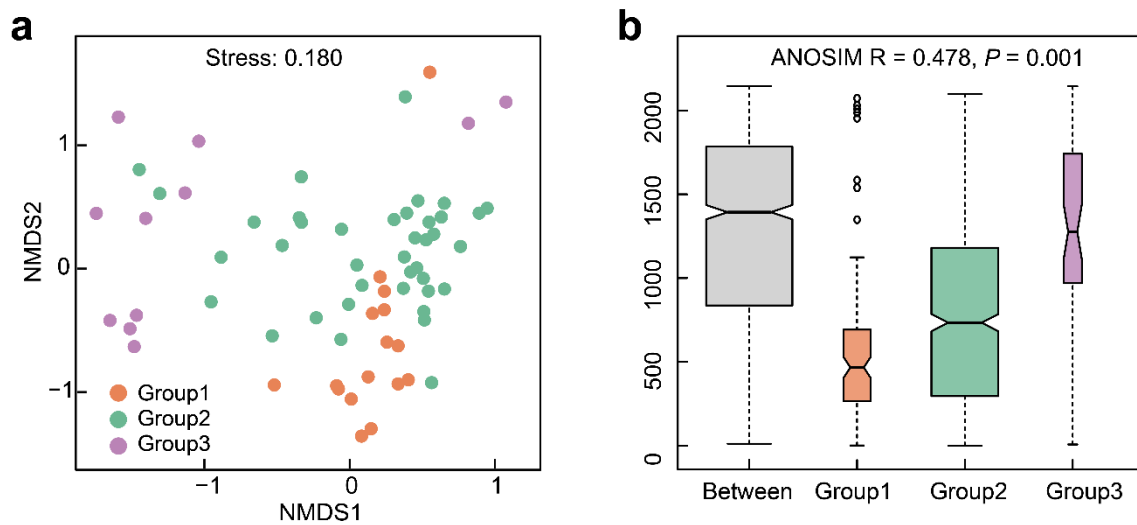
Supplementary Figure 5. Hierarchical clustering of environmental conditions for each groundwater sample based on the physicochemical parameters.



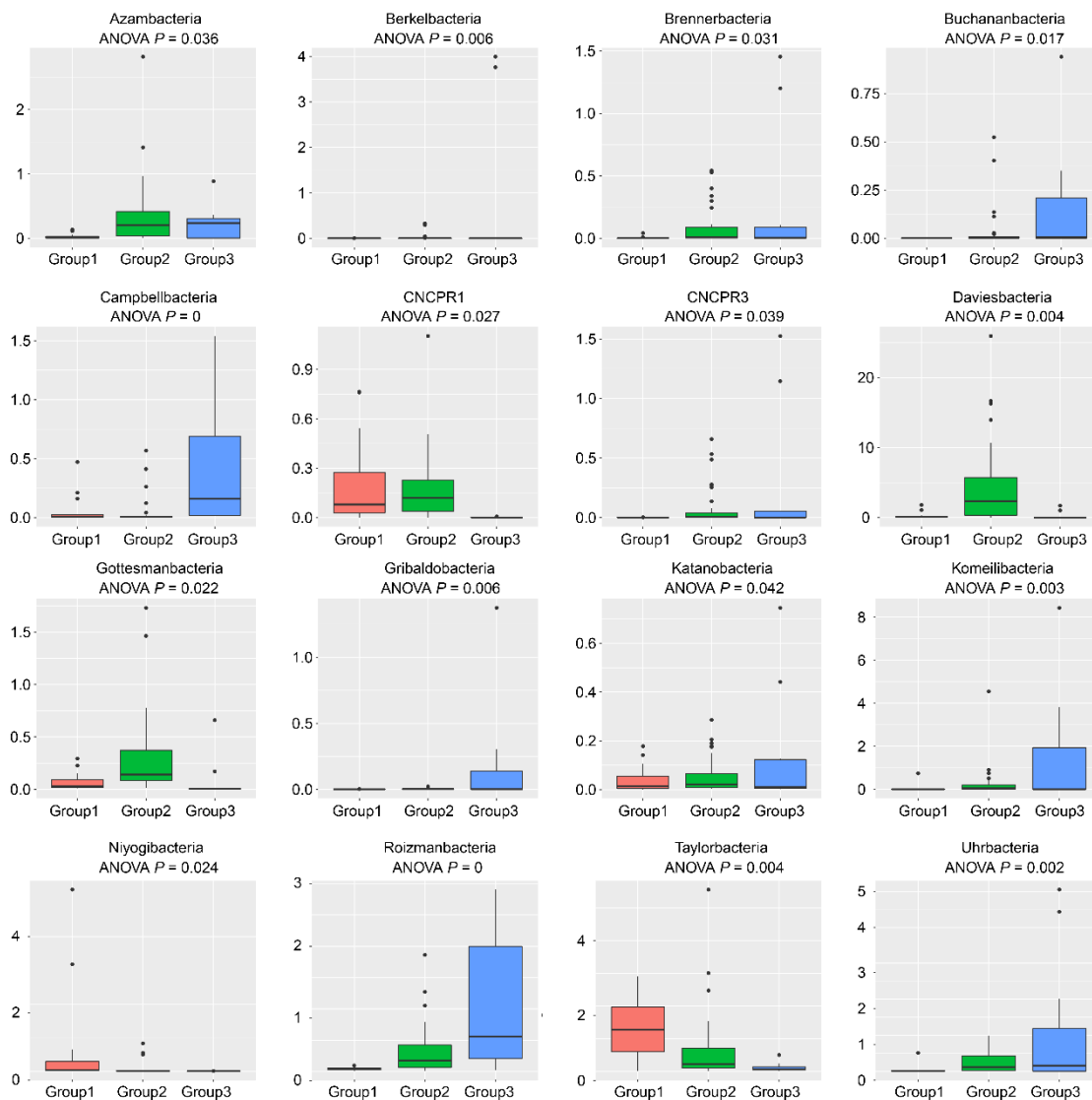
Supplementary Figure 6. Significant differences in distributions of physicochemical parameters (mg/L, except for pH) among the three groundwater groups.



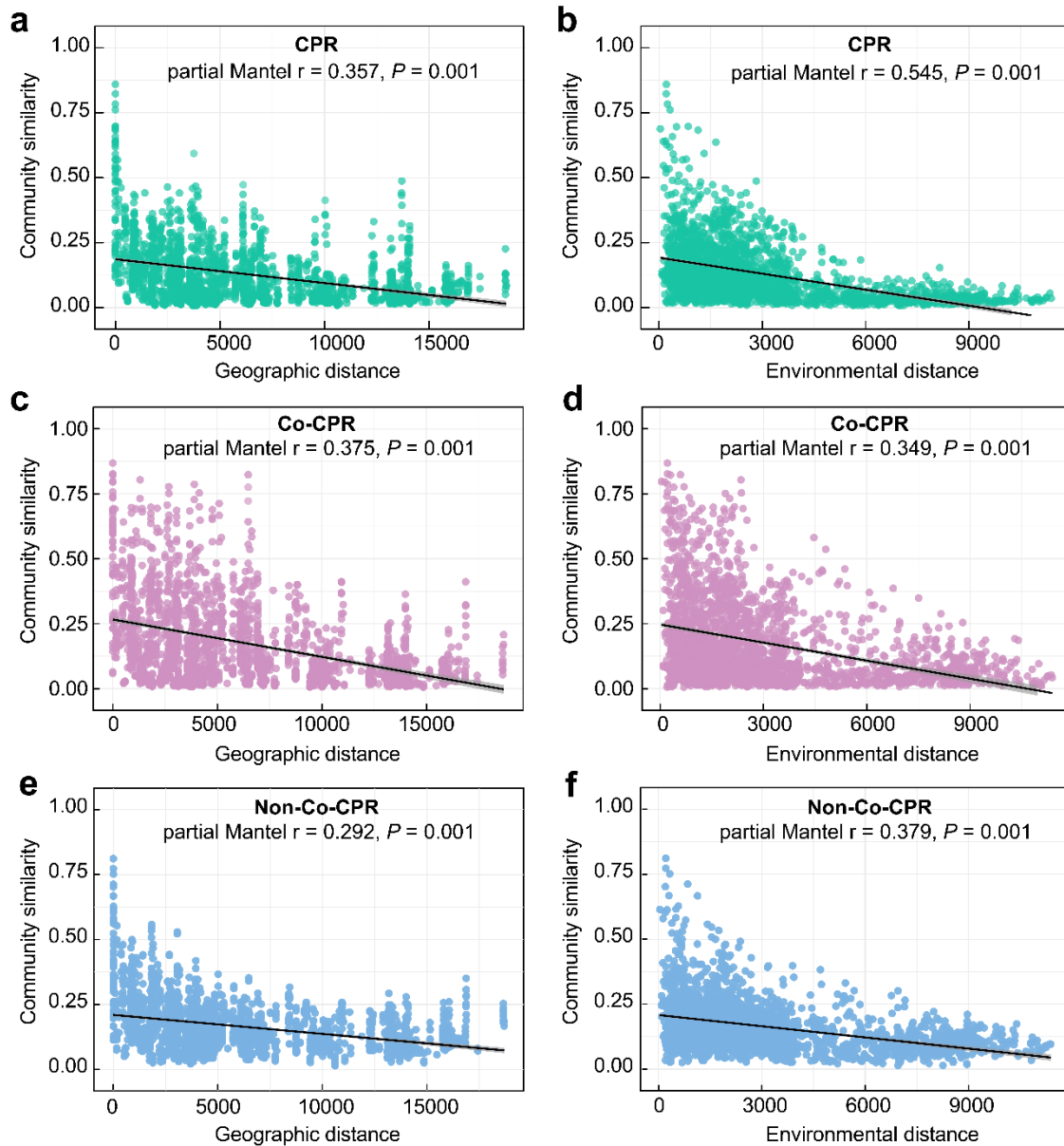
Supplementary Figure 7. Properties of physicochemical parameters between dry and wet seasons. The significant differences of physicochemical parameters between dry and wet seasons for all groundwater samples (a), samples in the lowest-salinity Group1 (b), samples in the moderate-salinity Group2 (c), as well as samples in the highest-salinity Group3 (d).



Supplementary Figure 8. Compositions of CPR MAGs among the three groundwater groups. (a) Non-metric multidimensional scaling (NMDS) showing the compositional dissimilarity of CPR MAGs in 66 groundwater samples. **(b)** Analysis of similarity statistics (ANOSIM) displaying the dissimilarities of CPR MAGs among the three groundwater groups.



Supplementary Figure 9. Significant differences in CPR at the phylum level among the three sampling groups.



Supplementary Figure 10. Distance-decay patterns of CPR and non-CPR bacteria.

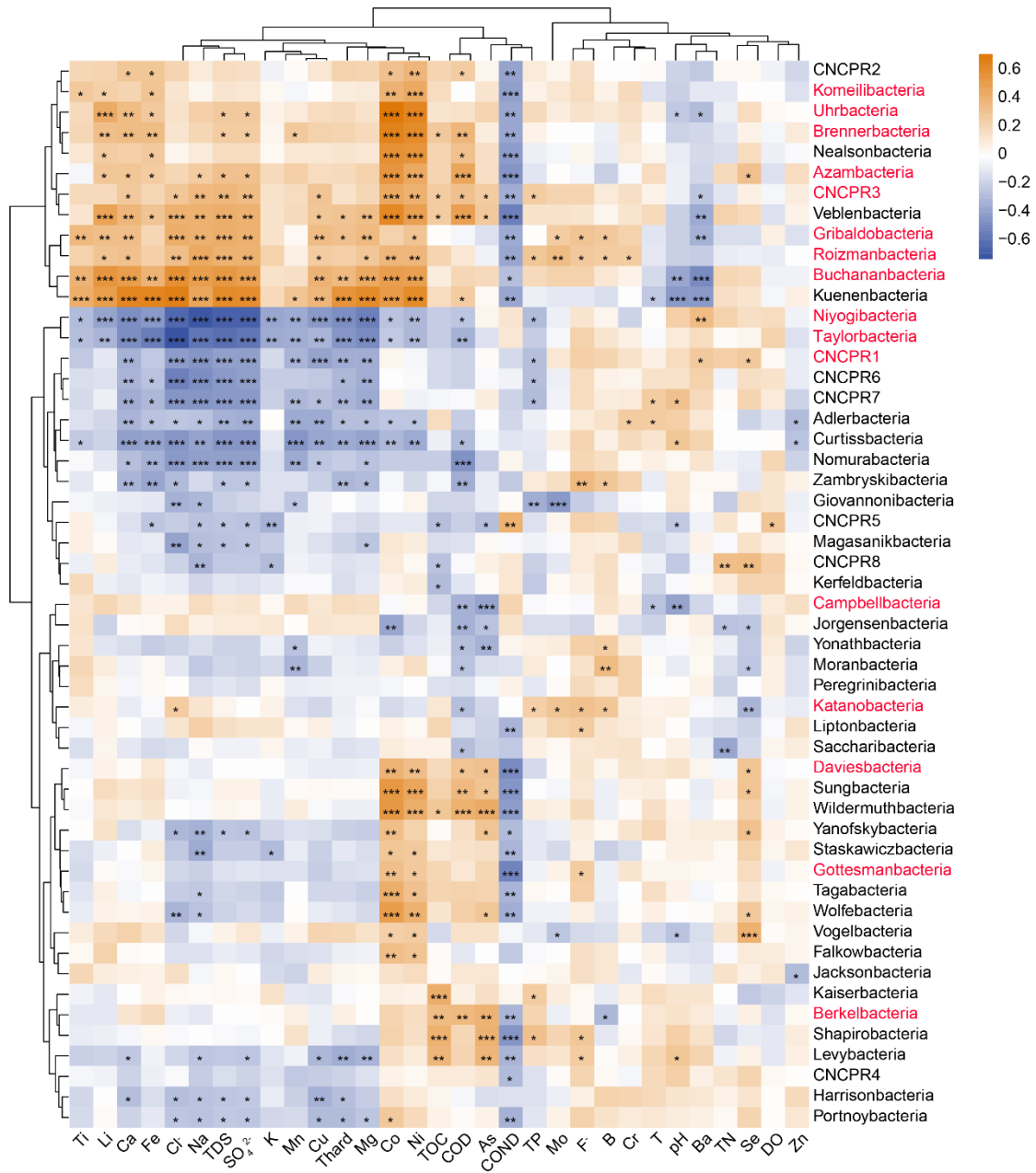
Distance-decay patterns of CPR along the geographic (**a**) and environmental distance (**b**).

Distance-decay patterns of non-CPR bacteria that co-occurred with CPR (Co-CPR) along

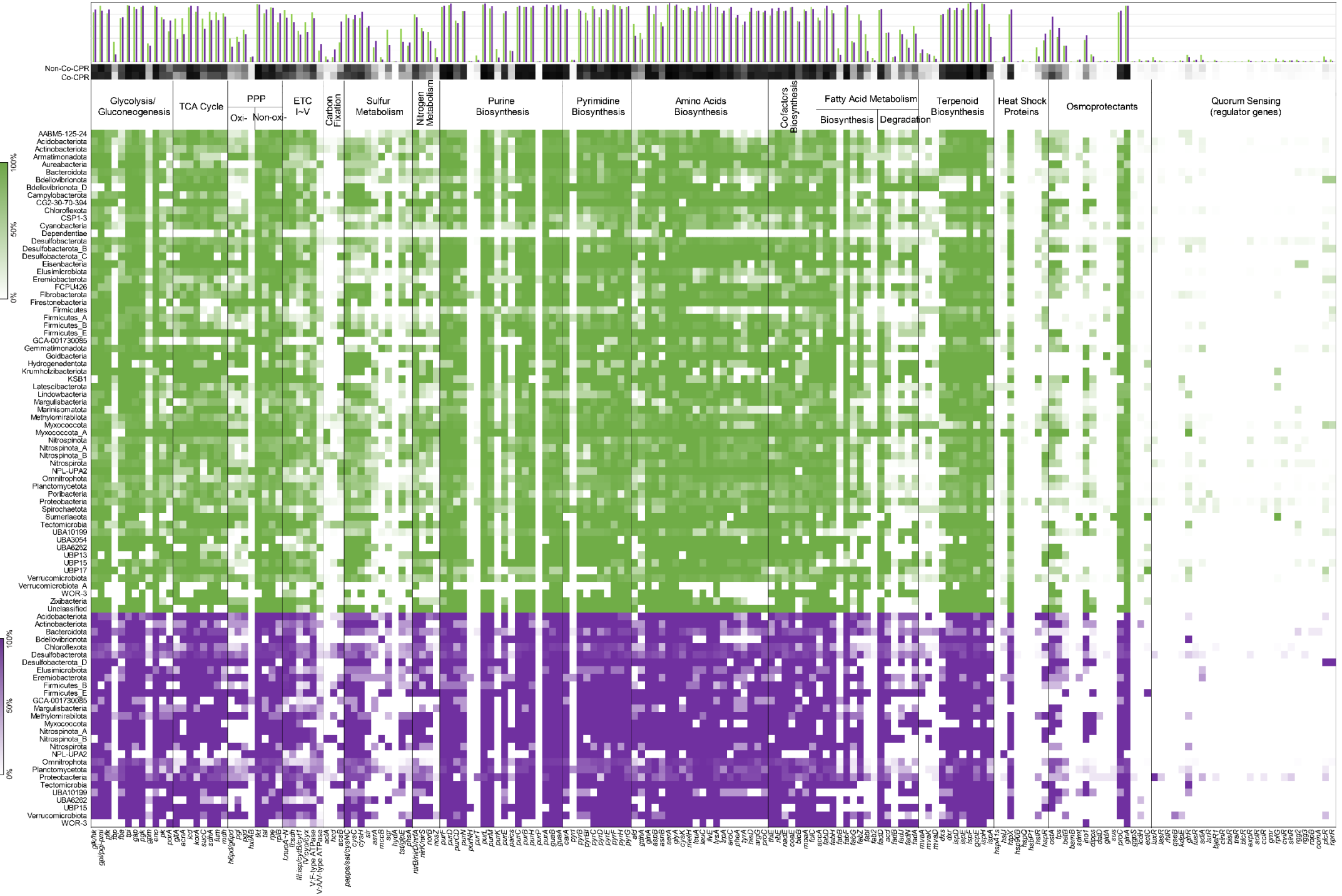
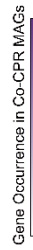
the geographic (**c**) and environmental distance (**d**). Distance-decay patterns of non-CPR

bacteria that not co-occurred with CPR (Non-Co-CPR) along the geographic (**e**) and

environmental distance (**f**).



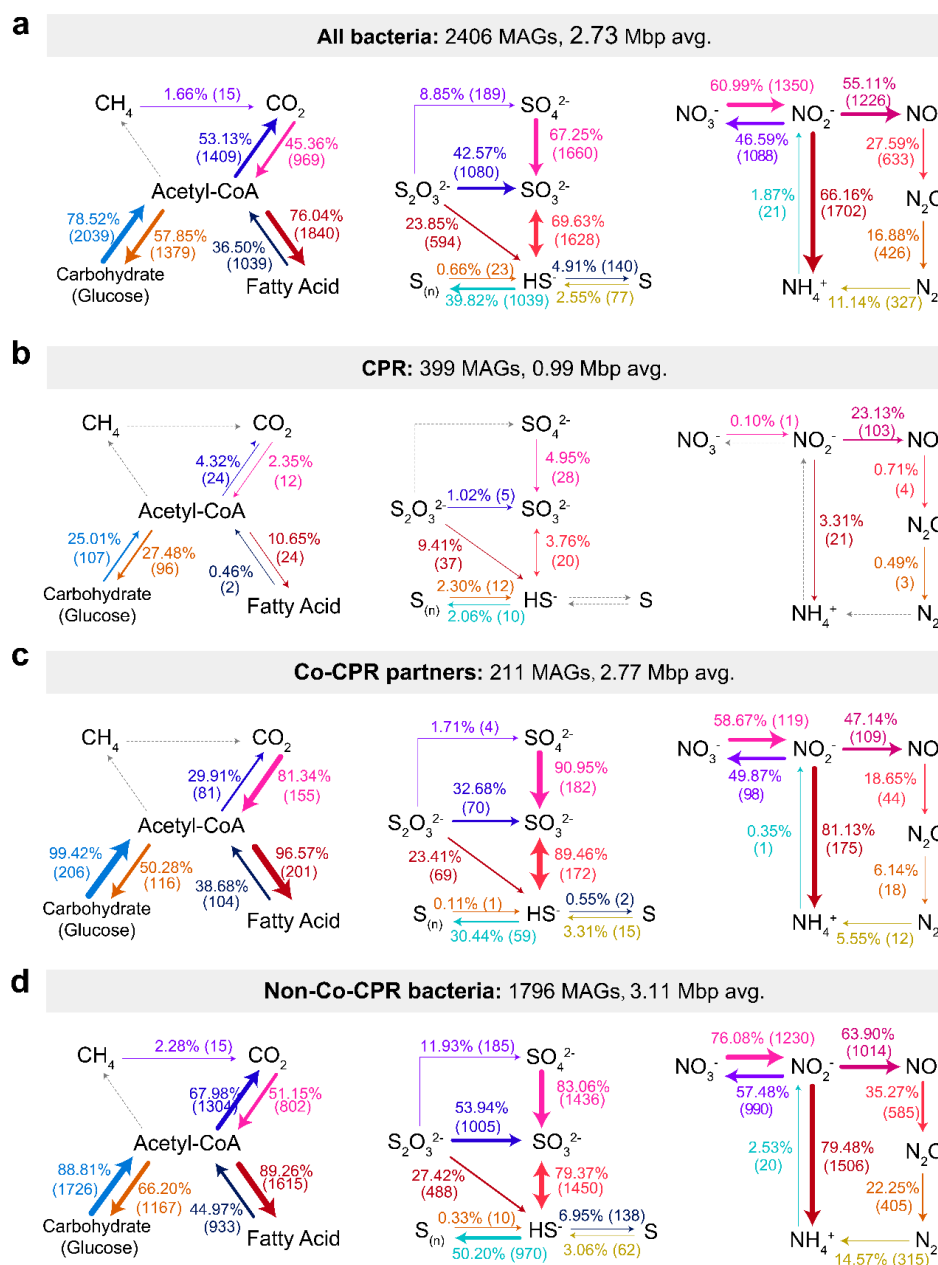
Supplementary Figure 11. Heatmap showing Spearman relationships between physicochemical parameters and CPR phyla. The phylum in red denotes the CPR with significant differences among the three sampling groups. Significant codes: *** ≤ 0.001 ** ≤ 0.01 * ≤ 0.05 .



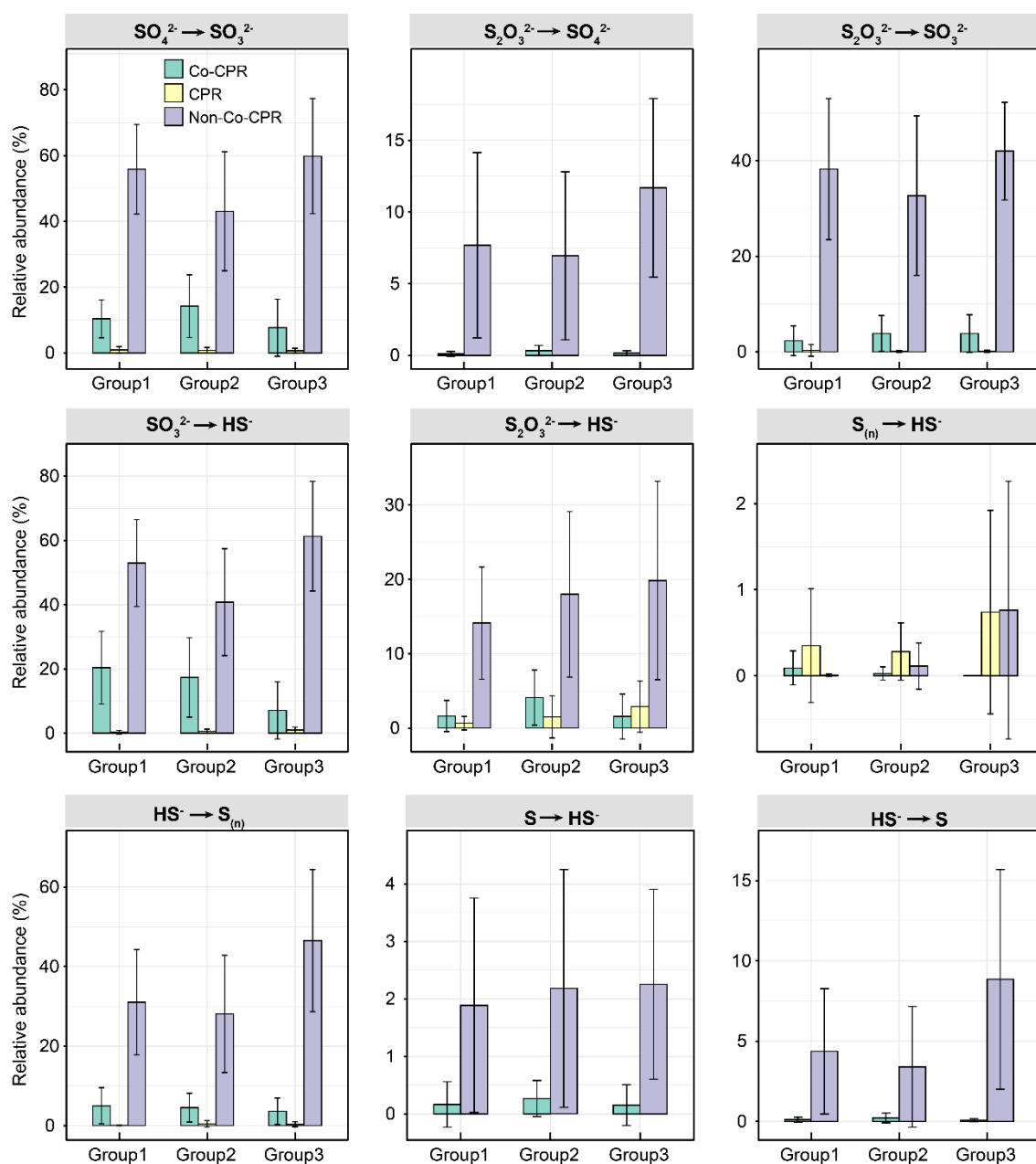
Supplementary Figure 12. Metabolic heatmap inferred for non-CPR phyla. The occurrence of genes is calculated based on the number of MAGs regardless of their abundance. Abbreviation: acetyl-CoA, acetyl-coenzyme A, a pivotal intermediate in multiple carbon pathways. TCA cycle: tricarboxylic acid cycle; PPP, pentose phosphate pathway; ETC, electron transport chain.



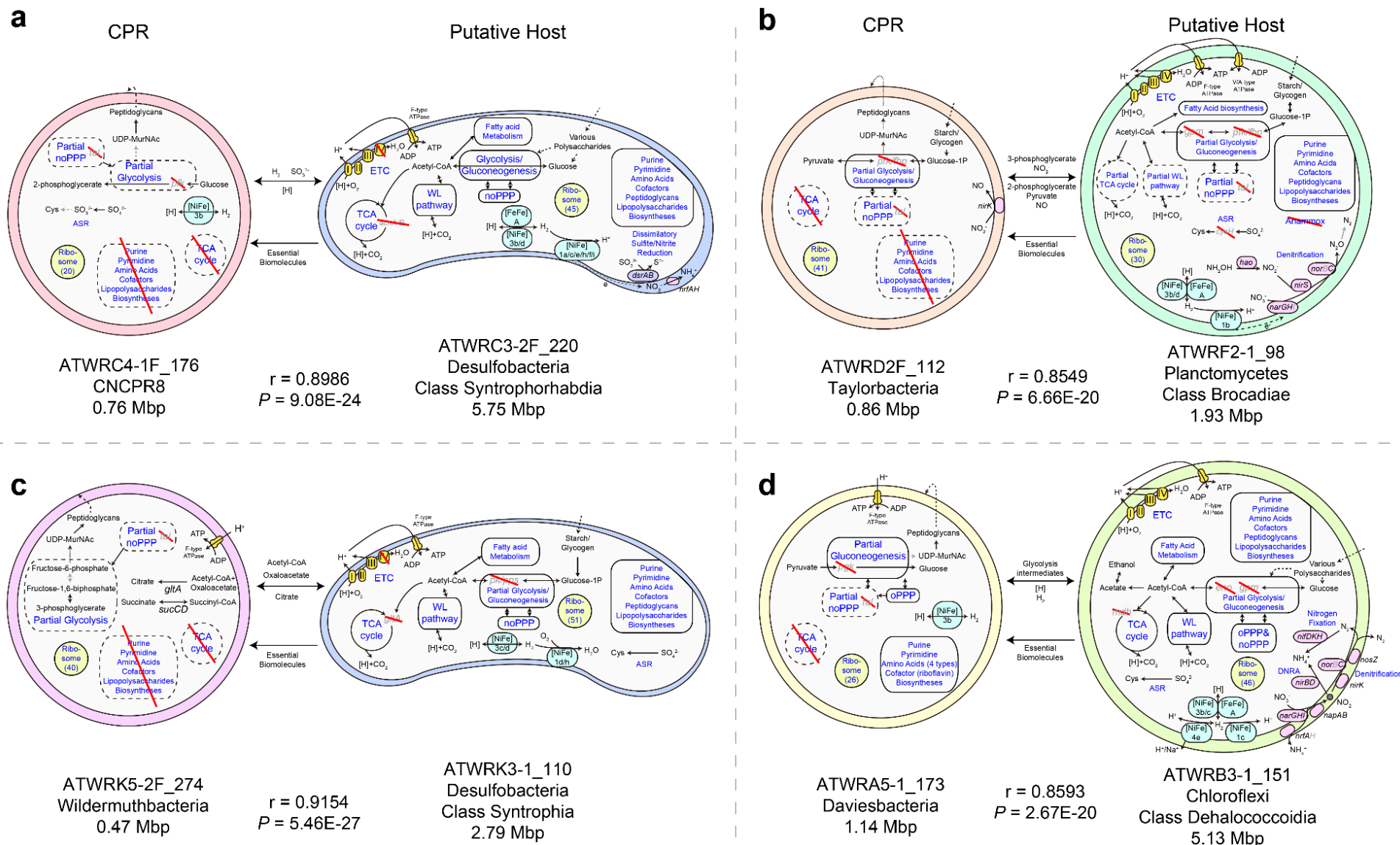
Supplementary Figure 13. Heatmap showing the completeness of the core metabolic pathways for CPR MAGs at the phylum level.



Supplementary Figure 14. The abundance of genes related to sulfur, nitrogen, and carbon cycles. Detailed information on these genes inferred for all bacteria (a), CPR (b), Co-CPR (c), or Non-Co-CPR (d). The percentage near each arrow indicates the MAG capacity for catalyzing metabolic steps, and the percentage was calculated by (the sequencing depth of genomes that possess the genes for certain steps / total sequencing depths of the selected genomes). The number in the bracket near each arrow indicates the number of genomes that possess the gene for each metabolic step.



Supplementary Figure 15. The abundance of genomes with sulfur-transformation genes relevant to CPR, Co-CPR, and Non-Co-CPR bacteria in different groups.



Supplementary Figure 16. Metabolic linkages of selected CPR bacteria and their putative hosts. (a) Intermediates might be exchanged between the two microorganisms to bypass the missing step of the TCA cycle in Desulfobacteria (*gltA* encoding citrate synthase). (b) Intermediates of the glycolysis/gluconeogenesis and nitrite reduction might be exchanged. Nitrite reductase genes *nirK* and *nirS* located in different subcellular locations (periplasm and cytoplasm) (c) Molecules related to hydrogen and sulfur metabolisms might be exchanged. (d) Intermediates of glycolysis/gluconeogenesis and hydrogen

metabolism might be exchanged. Abbreviation: noPPP: non-oxidative pentose phosphate pathway; oPPP: oxidative pentose phosphate pathway; TCA cycle: tricarboxylic acid cycle; ETC: electron transport chain; WL pathway: Wood-Ljungdahl pathway; ASR: assimilatory sulfate reduction; DNRA: dissimilatory nitrate reduction to ammonium; [H]: electron carriers, such as NADH and reduced ferredoxin.