**Supplementary Table 1. Information of MG-RAST soil metagenomes.** Data are extracted from publications used in this study including study ID, MG-RAST ID, sample name, publication, bp, reads, latitude and longitude, climate zone, climate code, mean annual temperature (MAT, °C), mean annual precipitation (MAP, mm yr-1), soil C, pH, Sand (%), Silt (%), Clay (%), and hits in RefSeq databases for “broad” and “narrow” functions.

Available in a separated Excel spreadsheet file.

**Supplementary Table 2. Distance between broad and narrow functions.** PERMANOVA pairwise test results (pseudo-t statistics) showing the significant difference between microbial taxonomic compositions for “broad” and “narrow” functions#.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Broad | | | | | Narrow | | | | |
|  | AAD | CHO | CBS | CVPGP | Protein | N | P | K | S | Fe |
| AAD |  |  |  |  |  |  |  |  |  |  |
| CHO | 8.81 |  |  |  |  |  |  |  |  |  |
| CBS | 9.24 | 9.99 |  |  |  |  |  |  |  |  |
| CVPGP | 7.73 | 9.33 | 7.52 |  |  |  |  |  |  |  |
| Protein | 12.96 | 13.31 | 8.70 | 11.25 |  |  |  |  |  |  |
| N | 16.68 | 18.80 | 17.67 | 15.79 | 21.15 |  |  |  |  |  |
| P | 12.54 | 14.31 | 11.85 | 12.52 | 14.65 | 15.42 |  |  |  |  |
| K | 18.90 | 20.20 | 18.83 | 18.52 | 21.86 | 15.99 | 14.64 |  |  |  |
| S | 18.14 | 19.65 | 19.52 | 18.01 | 22.97 | 15.56 | 17.40 | 15.53 |  |  |
| Fe | 20.50 | 22.18 | 21.70 | 20.85 | 23.77 | 21.08 | 19.21 | 18.08 | 19.11 |  |

#Broad functions include AAD (Amino Acids and Derivatives), CHO (Carbohydrates), CBS (Clustering-based Subsystems), CVPGP (Cofactors, Vitamins, Prosthetic Groups, Pigments), and Protein (Protein Metabolism). Narrow functions contain N (Nitrogen Metabolism), P (Phosphorous Metabolism), K (Potassium Metabolism), S (Sulfur Metabolism), and Fe (Iron Acquisition and Metabolism).

**Supplementary Table 3. Relations between broad and narrow functions.** Pairwise sample statistic (Rho) of RELATE test using Spearman rank correlation method showing the relationships between microbial taxonomic compositions for “broad” and “narrow” functions#.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Broad | | | | | Narrow | | | | |
|  | AAD | CHO | CBS | CVPGP | Protein | N | P | K | S | Fe |
| AAD |  |  |  |  |  |  |  |  |  |  |
| CHO | 0.987 |  |  |  |  |  |  |  |  |  |
| CBS | 0.987 | 0.982 |  |  |  |  |  |  |  |  |
| CVPGP | 0.986 | 0.980 | 0.979 |  |  |  |  |  |  |  |
| Protein | 0.967 | 0.971 | 0.973 | 0.963 |  |  |  |  |  |  |
| N | 0.895 | 0.893 | 0.884 | 0.924 | 0.885 |  |  |  |  |  |
| P | 0.917 | 0.902 | 0.900 | 0.938 | 0.888 | 0.944 |  |  |  |  |
| K | 0.838 | 0.834 | 0.830 | 0.865 | 0.818 | 0.878 | 0.911 |  |  |  |
| S | 0.856 | 0.849 | 0.840 | 0.894 | 0.840 | 0.953 | 0.936 | 0.888 |  |  |
| Fe | 0.819 | 0.794 | 0.802 | 0.847 | 0.773 | 0.876 | 0.911 | 0.853 | 0.906 |  |

#Broad functions include AAD (Amino Acids and Derivatives), CHO (Carbohydrates), CBS (Clustering-based Subsystems), CVPGP (Cofactors, Vitamins, Prosthetic Groups, Pigments), and Protein (Protein Metabolism). Narrow functions contain N (Nitrogen Metabolism), P (Phosphorous Metabolism), K (Potassium Metabolism), S (Sulfur Metabolism), and Fe (Iron Acquisition and Metabolism).

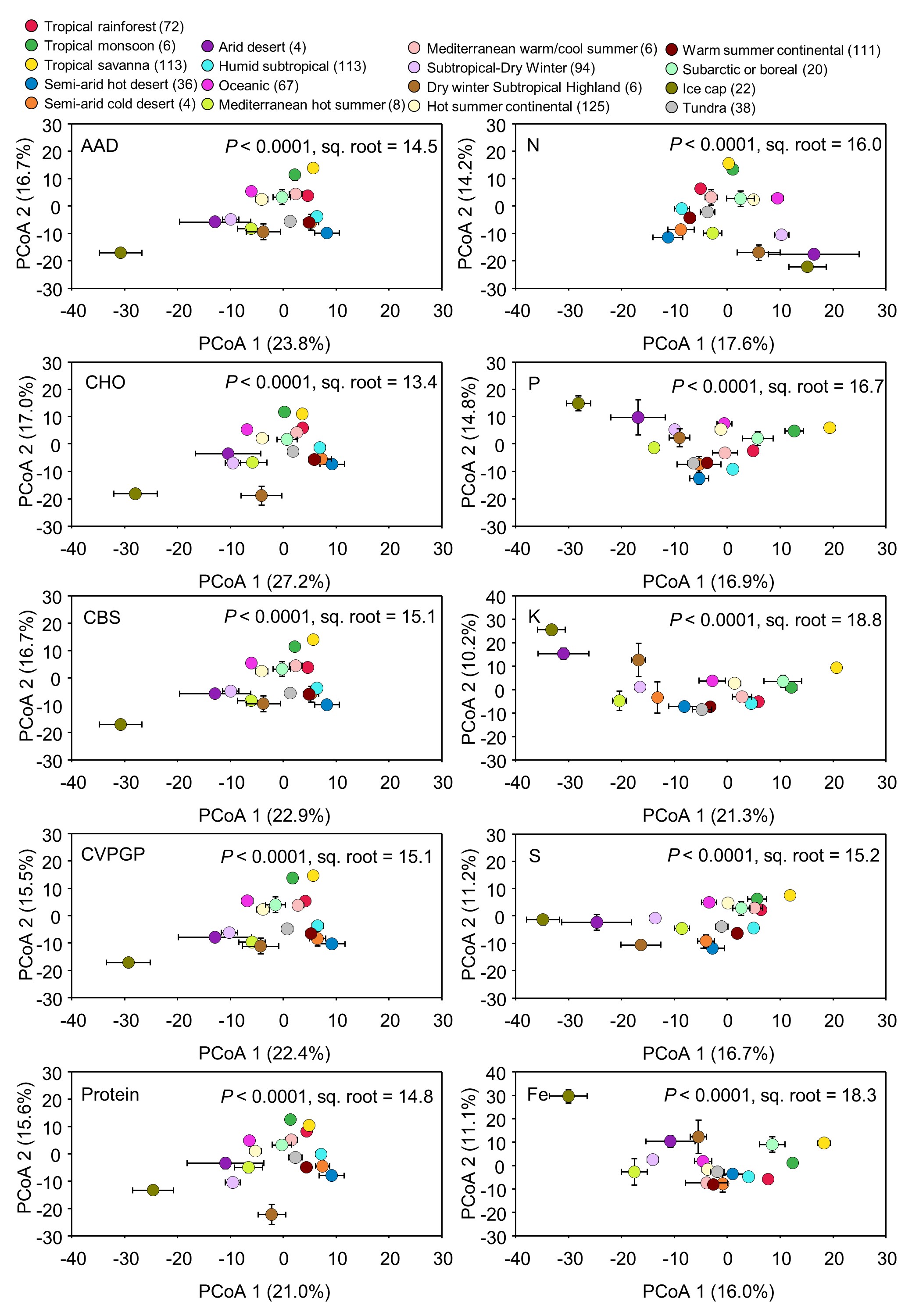
**Supplementary figure legends**

**Supplementary Fig. 1. Taxonomic compositions for “broad” versus “narrow” functions across climate zones.** Principal coordinates analysis (PCoA) showing beta-diveristy of soil metagenomes annotated using RefSeq database at genus level (Taxonomy) grouped by seventeen climate zones. The error bars represent the standard deviation of data ranges. Sample sizes of each climate zone are given in parentheses. Variations (by percentage) explained by the two principal coordinate dimensions are given in parentheses. *P* values and sq. root of PERMANOVA are also given.

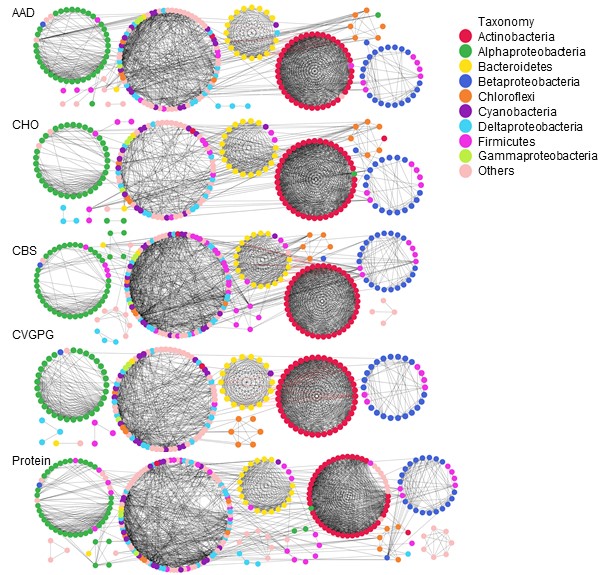
**Supplementary Fig. 2. Co-occurrence networks of microbial compositions for the five “broad” functions.** Network graphs with submodule structures of soil microbial taxonomic compositions for “broad” functions annotated using RefSeq at genus level. Node color represents classification of RefSeq database at phylum/class levels. A black edge indicates a positive interaction between two nodes, while a red edge indicates a negative interaction.

**Supplementary Fig. 3. Co-occurrence networks of microbial compositions for the five “narrow” functions.** Network graphs with submodule structures of soil microbial taxonomic compositions for “narrow” functions annotated using RefSeq at genus level. Node color represents classification of RefSeq database at phylum/class levels. A black edge indicates a positive interaction between two nodes, while a red edge indicates a negative interaction.

**Supplementary Fig. 1.**

****

**Supplementary Fig. 2.**

****

**Supplementary Fig. 3.**

