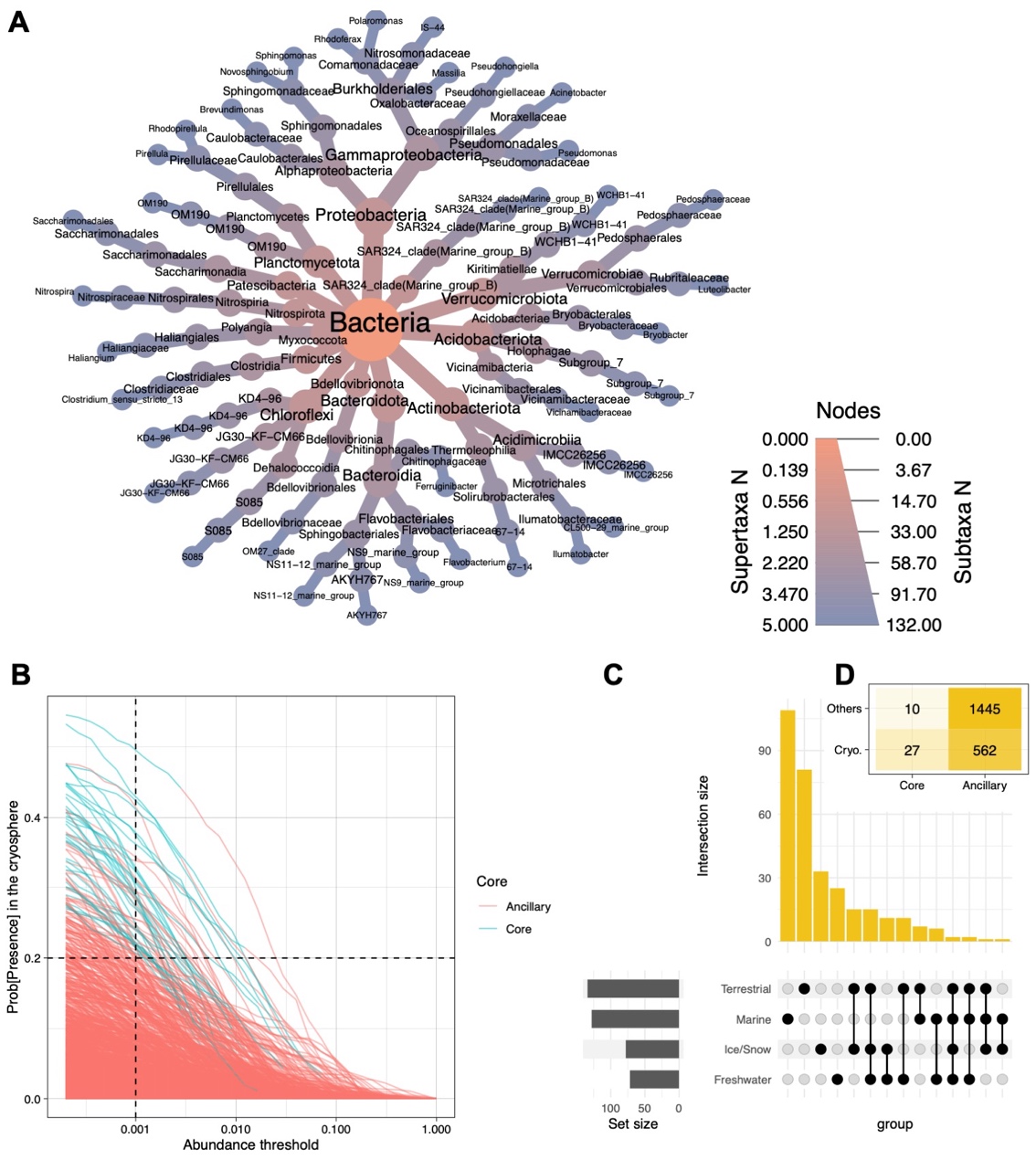
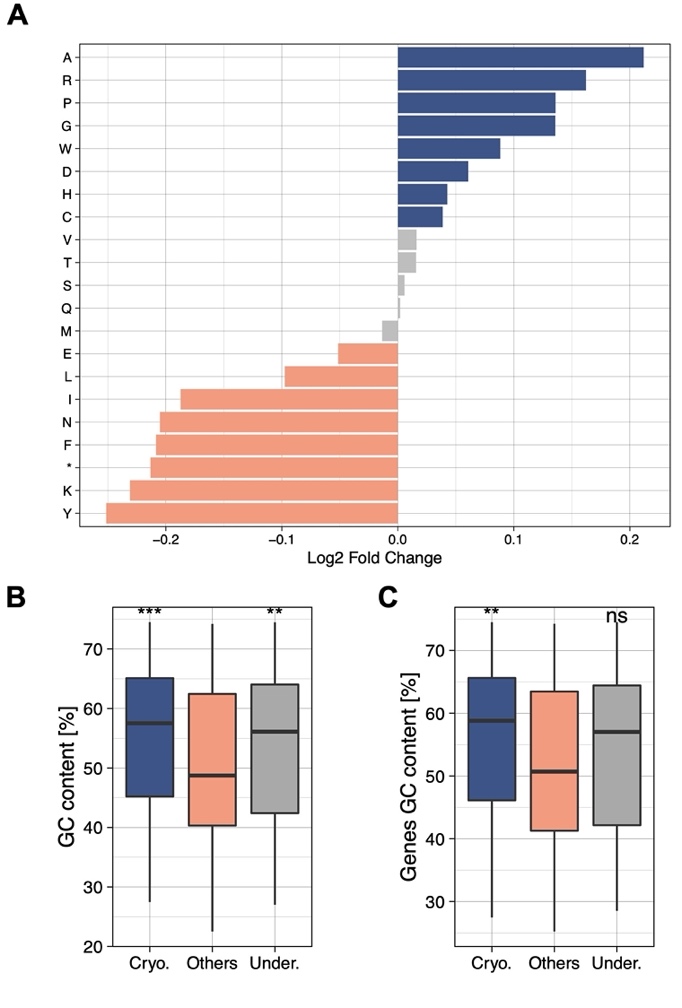


**Extended data Figure 1.** **A)** Phylogenetic tree based on the abundant ASVs (relative abundance > 0.5%) found in PP2, demonstrating the cryosphere microbiome diversity across the bacterial tree of life. The highlighted colors represent the ASVs detected in cryospheric ecosystems, and the barplot represents the coefficient for the logistic classification analysis. The number of ASVs at **B)** the phylum-level, and **C)** genus-level taxonomy with an odds-ratio greater than 1 in the logistic classification is shown for PP1 and PP2. Only taxa with the highest numbers are shown.



**Extended data Figure 2.** **A)** Heat-tree showing the taxonomic classification of the 37 bacterial genera representing the core microbiome of the cryosphere. They represent all bacterial genera with a probability of presence of 20% (calculated in the binomial model analysis) in the cryosphere, and present in all four ecosystem types. **B)** Line graphs depicting the probability of presence of a given bacterial genus in the cryosphere and their respective abundance, to identify the 'core' (blue) and 'ancillary' (red) genera. The dashed lines represent the chosen thresholds for the core microbiome definition. **C)** Upset plot showing the overlap across the ecosystem types core microbiome, defined at the genus-level, with a prevalence of 20% at an abundance threshold of 0.1% relative abundance. **(D)** Heat plot showing the number of genera per group, highlighting the large overlap between the core microbiome and the cryospheric genera.



**Extended data Figure 3.** **A)** The enriched abundance of aminoacids in the cryosphere (blue; positive fold change) compared to those found in the non-cryospheric ecosystems are depicted. The “\*” represents stop codons. **B)** The overall GC% of the genomes belonging to the cryospheric, others and underrepresented genera are depicted. **C)** The GC% of the genes predicted in the genomes belonging to the cryospheric, others and underrepresented genera are depicted. The median, 25% and 75% quartiles are represented in the boxplots. The significant difference detected for both **B)** and **C)** panels was computed using Wilcoxon tests. (\*\*\*: 0-0.001, \*\*: 0.001-0.01, \*: 0.01-0.05)



**Extended data Figure 4.** **A)** Boxplots indicate the overall identity percentage for representative sequences for each of the gene clusters that matched with UniProt sequences. The figures demonstrate these values for the Cryosphere-only genes, non-cryosphere (Others) and those 'shared' between the two habitats. **B)** The GC content % of all the genes within each cluster is shown. **C)** The pairwise identity of all sequences within each cluster is represented on the boxplots. **D)** Odds-ratio estimations of the UniProt matches with respect to the annotation level, and the presence/or not in the cryospheric metagenomes. (\*\*\*: 0-0.001, \*\*: 0.001-0.01, \*: 0.01-0.05)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | Accuracy [%] | Precision [%] | | Recall [%] | AUC [%] |
| PP1 | 98.16±1.19 | 99.96±0.09 | 96.36±2.39 | | 99.92±0.05 |
| PP2 | 97.49±1.38 | 99.92±0.16 | 95.06±2.76 | | 99.93±0.04 |

**Extended data Table 1.** Cryospheric bacterial communities’ logistic classification models performance summary.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Group** | **MedianSorensen’s index** | **Median**ß**-MNTD** |
| **PP1** | Cryo-Cryo | 0.284 | 0.400 |
| Cryo-Others | 0.250 | 0.415 |
| Other-Others | 0.245 | 0.398 |
| **PP2** | Cryo-Cryo | 0.290 | 0.341 |
| Cryo-Others | 0.245 | 0.371 |
| Other-Others | 0.233 | 0.360 |

**Extended data Table 2.** ß-diversity phylogenetics.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Metric** | | **α-MPD** | | **α-MNTD** | | **α-PD** | |
| **Estimate** | ***p-value*** | **Estimate** | ***p-value*** | **Estimate** | ***p-value*** |
| **Coefficients** | **Intercept** | 0.389±0.014 | < 0.001 | 0.307±0.006 | < 0.001 | 0.450±0.047 | < 0.001 |
| **Cryosphere** | 0.077±0.005 | < 0.001 | 0.015±0.003 | < 0.001 | 0.532±0.049 | < 0.001 |
| **log(SR)** | 0.062±0.004 | < 0.001 | -0.053±0.002 | < 0.001 | 0.110±0.001 | < 0.001 |
| **DatasetPP2** | -0.038±0.004 | < 0.001 | 0.020±0.002 | < 0.001 | 0.398±0.037 | < 0.001 |
| **Model** | **Adj. R2** | 0.108 | | 0.191 | | 0.664 | |
| **df** | 4240 | | 4240 | | 4240 | |

**Extended data Table 3.** α-diversity phylogenetics, linear models testing the influence of the cryosphere on the different metrics, with the log of species richness (SR) and the dataset as fixed effects (Intercept = non-cryospheric, and PP1).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***PERMANOVA*** | **DF** | **Sum of squares** | **r2** | **f-value** | **p-value** |
| *Ecosystem* | 3 | 55.774 | 0.18319 | 52.702 | < 0.001 |
| *Dataset* | 1 | 5.633 | 0.01850 | 15.969 | < 0.001 |
| *Residual* | 689 | 243.055 | 0.79831 |  |  |
| *Total* | 693 | 304.462 | 1.00000 |  |  |
| ***Snow/Ice – Terr.*** | **DF** | **Sum of squares** | **r2** | **f-value** | **p-value** |
| *Ecosystem* | 1 | 18.859 | 0.11779 | 54.36 | < 0.001 |
| *Dataset* | 1 | 4.906 | 0.03064 | 14.14 | < 0.001 |
| *Residual* | 393 | 136.340 | 0.85157 |  |  |
| *Total* | 395 | 160.105 | 1.00000 |  |  |
| ***Snow/Ice – Marine*** | **DF** | **Sum of squares** | **r2** | **f-value** | **p-value** |
| *Ecosystem* | 1 | 23.280 | 0.15758 | 66.918 | < 0.001 |
| *Dataset* | 1 | 4.087 | 0.02767 | 11.749 | < 0.001 |
| *Residual* | 346 | 120.370 | 0.81476 |  |  |
| *Total* | 348 | 147.737 | 1.00000 |  |  |
| ***Snow/Ice – Fresh.*** | **DF** | **Sum of squares** | **r2** | **f-value** | **p-value** |
| *Ecosystem* | 1 | 10.991 | 0.06768 | 28.6254 | < 0.001 |
| *Dataset* | 1 | 3.195 | 0.01967 | 8.3202 | < 0.001 |
| *Residual* | 386 | 148.211 | 0.91265 |  |  |
| *Total* | 388 | 162.397 | 1.00000 |  |  |
| ***Marine – Terr.*** | **DF** | **Sum of squares** | **r2** | **f-value** | **p-value** |
| *Ecosystem* | 1 | 25.402 | 0.20705 | 85.019 | < 0.001 |
| *Dataset* | 1 | 7.052 | 0.05748 | 23.604 | < 0.001 |
| *Residual* | 302 | 90.230 | 0.73547 |  |  |
| *Total* | 304 | 122.684 | 1.00000 |  |  |
| ***Marine – Fresh.*** | **DF** | **Sum of squares** | **r2** | **f-value** | **p-value** |
| *Ecosystem* | 1 | 18.464 | 0.14665 | 53.295 | < 0.001 |
| *Dataset* | 1 | 5.240 | 0.04162 | 15.125 | < 0.001 |
| *Residual* | 295 | 102.202 | 0.81173 |  |  |
| *Total* | 297 | 125.906 | 1.00000 |  |  |
| ***Terrestrial - Fresh.*** | **DF** | **Sum of squares** | **r2** | **f-value** | **p-value** |
| *Ecosystem* | 1 | 16.200 | 0.11536 | 47.332 | < 0.001 |
| *Dataset* | 1 | 7.175 | 0.05109 | 20.963 | < 0.001 |
| *Residual* | 342 | 117.056 | 0.83355 |  |  |
| *Total* | 344 | 140.431 | 1.00000 |  |  |

**Extended data Table 4.** Model summaries for the PERMANOVA and all pairwise.adonis comparisons of cryospheric ecosystems.

|  |  |  |  |
| --- | --- | --- | --- |
| ***WTS*** | **Test statistic** | **df** | **p-value WTPS** |
| *Ecosystem* | 112.0236 | 3 | 0 |
| *Interaction* | 103.1681 | 4 | 0 |

**Extended data Table 5.** Shannon’s index *H* α-diversity (calculated at the genus taxonomic level) Wald-Type Statistic (WTS) summary. This test was chosen as a non-parametric alternative to ANOVA for non-normally distributed data. Interaction represents the interaction between the Ecosystem and Dataset parameters. The computed means are 2.987, 2.856, 3.245, and 3.669 for freshwater, snow/ice, marine and terrestrial, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **MedianOthers (n=265)** | **MedianCryo (n=197)** | **MedianUnder.(n=198)** |
| *GC content [%]* | 48.7 | 57.5 \*\*\* | 56.1 \*\* |
| *Genes GC content [%]* | 50.7 | 58.8 \*\* | 57.0 |
| *Genome size [mbp]* | 3.97 | 4.19 | 4.06 |
| *Growth doubling time [d]* | 4.48 | 4.12 | 4.55 |
| *Codon usage bias [CUBHE]* | 0.627 | 0.627 | 0.623 |
| *Consistency [HE]* | 0.527 | 0.519 | 0.529 |
| *Codon pair bias [CPB]* | -0.375 | -0.370 | -0.380 |

**Extended data Table 6.** RefSeq genomic properties summary. P-values were computed using Wilcoxon tests implemented in the *compare\_means* function of the ggpubr R package, comparing the cryospheric and underrepresented genera against the others (\*\*\*: 0-0.001, \*\*: 0.001-0.01, \*: 0.01-0.05).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Study ID** | **Sample n.** | **Ecosystem type** | **Description** |
| *PP1* | Bergk2019 | 12 | Ice/Snow | Snow samples, Svalbard |
| *PP1* | NOMIS | 10 | Freshwater | Glacier-fed stream sediment samples, Russia and New Zealand |
| *PP1* | PRJDB9246 | 11 | Freshwater | Microbial mat and water, Antarctica |
| *PP1* | PRJEB12640 | 41 | Terrestrial | Soil chronosequence samples, Svalbard |
| *PP1* | PRJEB26163 | 62 | Marine | Marine water, Arctic Ocean |
| *PP1* | PRJEB29215 | 32 | Ice/Snow | Snow samples, Antarctica |
| *PP1* | PRJEB31938 | 26 | Marine | Sea ice, snow, water and sediment, Greenland and the Arctic Ocean |
| *PP1* | PRJEB40467 | 18 | Terrestrial | Alpine permafrost, Italy |
| *PP1* | PRJNA296475 | 3 | Ice/Snow | Cryoconite hole, Svalbard |
| *PP1* | PRJNA320505 | 1 | Ice/Snow | Cryoconite hole, Antarctica |
| *PP1* | PRJNA380676 | 1 | Terrestrial | Arctic rock, Svalbard |
| *PP1* | PRJNA418054 | 32 | Terrestrial | Permafrost, Alaska |
| *PP1* | PRJNA430179 | 11 | Ice/Snow | Glacier Ice/snow, Spain |
| *PP1* | PRJNA436954 | 8 | Frehshwater | High-arctic microbial mat |
| *PP1* | PRJNA529498 | 33 | Ice/Snow | Cryoconite hole, Antarctica |
| *PP2* | PRJEB11496 | 24 | Marine | Marine sediment, Antarctica |
| *PP2* | PRJEB23054 | 59 | Terrestrial | Permafrost, Alaska |
| *PP2* | PRJNA244335 | 31 | Freshwater | Sediment and water from subglacial lake, Antarctica |
| *PP2* | PRJNA255432 | 70 | Freshwater | Arctic lake, Canada |
| *PP2* | PRJNA278982 | 4 | Marine | Ice-shelf water cavity, Antarctica |
| *PP2* | PRJNA321351 | 9 | Freshwater | Arctic lake, Greenland |
| *PP2* | PRJNA324626 | 9 | Terrestrial | Frozen soil, China |
| *PP2* | PRJNA430887 | 26 | Ice/Snow | Glacier ice and weather crust, USA |
| *PP2* | PRJNA431087 | 1 | Terrestrial | Microbial mat, Antarctica |
| *PP2* | PRJNA432743 | 7 | Freshwater | Subglacial aquifer brine, Antarctica |
| *PP2* | PRJNA471245 | 51 | Ice/Snow | Water, ice, soil, sediment and microbial mat, Antarctica |
| *PP2* | PRJNA480849 | 1 | Ice/Snow | Cryoconite hole, Antarctica |
| *PP2* | PRJNA554442 | 11 | Terrestrial | Permafrost, Alaska |
| *PP2* | PRJNA593264 | 13 | Marine | Water, sediments and snow, Antarctica |
| *PP2* | PRJNA629965 | 6 | Freshwater, Ice/Snow | Snow and glacier melt |
| *PP2* | PRJNA744712 | 72 | Ice/Snow | Cryoconite hole, Antarctica |

**Extended data Table 7.** Summary of the cryospheric samples included in the two amplicon datasets.