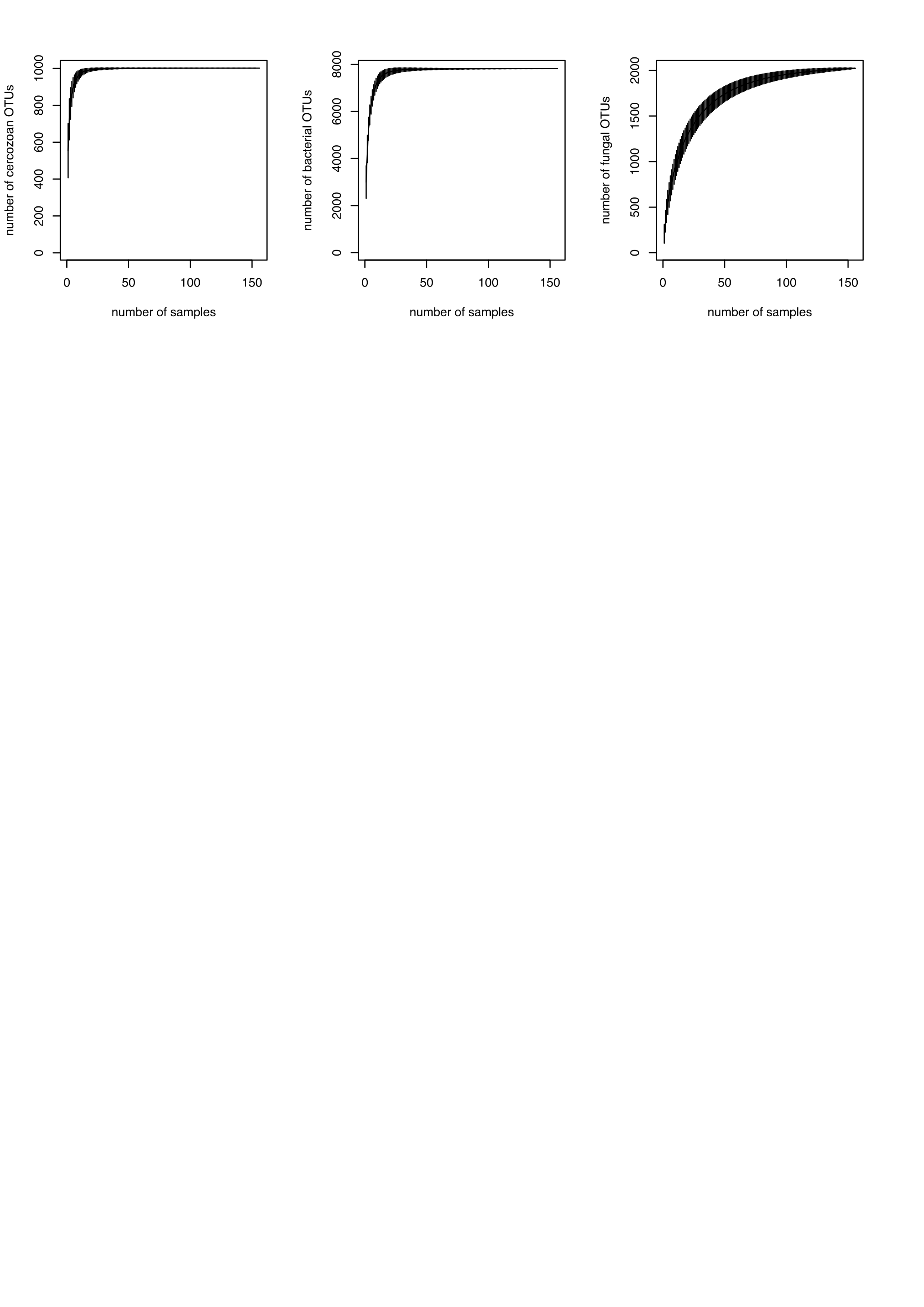
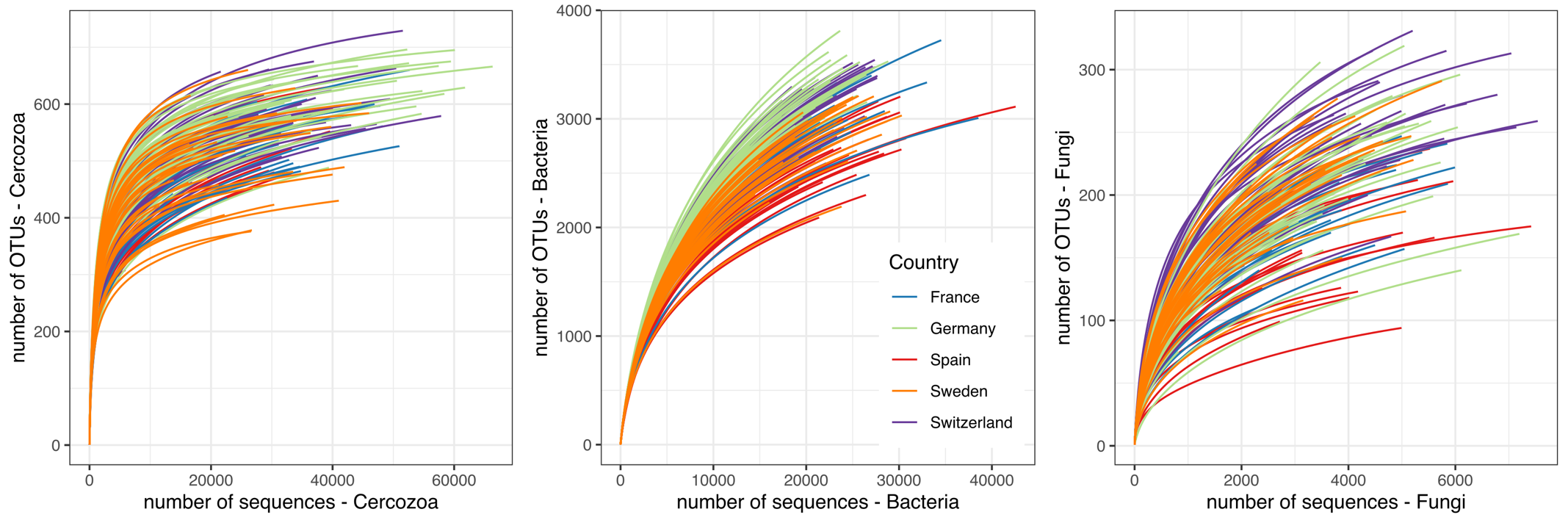
****

**Figure S1:** Species accumulation curve showing the sampling effort of cercozoan, bacterial and fungal OTUs. The diversity was covered already after the collection of a few samples.

****

**Figure S2:** Rarefaction curves for the 156 samples and showing the sequencing effort to cover the diversity of Cercozoa, Bacteria and Fungi. Curves are color-coded by country.

**Table S1:** Descriptive statistics (n=156) of retrieved OTUs and number of sequences for Cercozoa, fungi, bacteria, omnivorous and bacterivorous Cercozoa.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Total number** | **Mean** | **sd** | **Median** | **Minimum** | **Maximum** |
| **Cercozoa** | *OTUs* | 1,001 | 553.9 | 73.7 | 563 | 308 | 729 |
| *Sequences* | 5,016,415 | 32,156.5 | 12,291.8 | 31,141.5 | 5,875 | 66,330 |
| **Bacteria** | *OTUs* | 7,811 | 2,998.1 | 3,46.8 | 3,013 | 1,438 | 3,811 |
| *Sequences* | 3,562,803 | 22,838.5 | 4,582.3 | 22,683 | 5,115 | 42,552 |
| **Fungi** | *OTUs* | 2,023 | 207.1 | 50.9 | 208.5 | 75 | 331 |
| *Sequences* | 620,218 | 3,975.8 | 1,354 | 3,875.5 | 821 | 7,541 |
| **Omnivores** | *OTUs* | 239 | 150.1 | 19.2 | 152.5 | 79 | 188 |
| *Sequences* | 1,695,995 | 10,871.8 | 4,469.9 | 10,250.5 | 2,304 | 26,498 |
| **Bacterivores** | *OTUs* | 400 | 242.4 | 30.1 | 247 | 140 | 316 |
| *Sequences* | 2,003,878 | 12,845.4 | 4,809.6 | 12,371 | 2,391 | 26,402 |

**Table S2:** Results of the random forest analysis run on each response variable. The table provides the R2-fitting and R2-validation of each model, which represent the overall predictability of each model. Individual R2 and associated p-value are provided which represent the prediction power of each predictor in the model. The data can be visualized in Fig. 2 (A&B) in the manuscript. The beta diversity (community composition) is represented by the first component of the principal coordinate analysis (pco1) based on Bray-Curtis dissimilarity index computed on the Hellinger-transformed OTUs table (see the method for more details). The R script is available on GitHub: https://github.com/flopy007/Degrune\_et\_al\_2023

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **predictors (X)** | **response variables (Y)** | | | | | | | |
| **cercozoa (pco1\_c)** | | **fungi (pco1\_f)** | | **omnivore (pco1)** | | **bacterivore (pco1)** | |
| ***R2-fitting*** | ***R2-validation*** | ***R2-fitting*** | ***R2-validation*** | ***R2-fitting*** | ***R2-validation*** | ***R2-fitting*** | ***R2-validation*** |
| 0.938 | 0.906 | 0.971 | 0.954 | 0.931 | 0.901 | 0.896 | 0.834 |
| ***R2*** | ***P-value*** | ***R2*** | ***P-value*** | ***R2*** | ***P-value*** | ***R2*** | ***P-value*** |
| **SM** | 0.036 | 0.000 | 0.023 | 0.000 | 0.002 | 0.000 | 0.045 | 0.000 |
| **Clay** | 0.021 | 0.000 | 0.007 | 0.000 | 0.001 | 0.04805 | 0.023 | 0.000 |
| **Ctot** | 0.002 | 0.03604 | 0.001 | 0.12 | 0.002 | 0.01201 | 0 | 0.13213 |
| **Ntot** | 0.019 | 0.000 | 0.002 | 0.000 | 0 | 0.40841 | 0.024 | 0.000 |
| **Ptot** | 0.007 | 0.000 | 0 | 0.000 | 0 | 0.13213 | 0.011 | 0.000 |
| **pH** | 0.013 | 0.000 | 0.01 | 0.000 | 0.059 | 0.000 | 0.015 | 0.000 |
| **Lat** | 0.048 | 0.000 | 0.071 | 0.000 | 0.03 | 0.000 | 0.046 | 0.000 |
| **Long** | 0.067 | 0.000 | 0.159 | 0.000 | 0.077 | 0.000 | 0.065 | 0.000 |
| **MAT** | 0.102 | 0.000 | 0.204 | 0.000 | 0.202 | 0.000 | 0.051 | 0.000 |
| **MAP** | 0.215 | 0.000 | 0.034 | 0.000 | 0.001 | 0.000 | 0.33 | 0.000 |
| **pco1\_b** | 0.011 | 0.000 | 0.046 | 0.000 | 0.179 | 0.000 | 0.006 | 0.000 |
| **pco1\_f** | 0.396 | 0.000 | NA | NA | 0.379 | 0.000 | 0.281 | 0.000 |
| **pco1\_c** | NA | NA | 0.414 | 0.000 | NA | NA | NA | NA |

**Table S3**. Output of the structural equation modeling to explain the variation in the community composition of Cercozoa (pco1\_c). The SEM was run in R using the lavaan function from lavaan package (version 0.6.9). pco1\_c and pco1\_f are the community composition of cercozoa and fungi (calculated based on bray-curtis dissimilarity index on helliger transformed OTU table). The analysis can be reproduced with the code available here: <https://github.com/biodiversaCercozoa_2023>.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| lhs | op | rhs | estimate | se | pvalue | ci.lower | ci.upper | std.nox |
| pco1\_c | ~ | MAP | 0.289 | 0.031 | 0.000 | 0.229 | 0.349 | 0.290 |
| pco1\_c | ~ | Clay | -0.126 | 0.022 | 0.000 | -0.170 | -0.082 | -0.126 |
| pco1\_c | ~ | pH | 0.220 | 0.029 | 0.000 | 0.163 | 0.277 | 0.221 |
| pco1\_c | ~ | pco1\_f | 0.829 | 0.036 | 0.000 | 0.759 | 0.898 | 0.829 |
| pco1\_f | ~ | MAP | 0.539 | 0.031 | 0.000 | 0.478 | 0.600 | 0.541 |
| pco1\_f | ~ | MAT | -0.598 | 0.037 | 0.000 | -0.671 | -0.525 | -0.600 |
| pco1\_f | ~ | Clay | -0.112 | 0.031 | 0.000 | -0.172 | -0.051 | -0.112 |
| pco1\_f | ~ | Ctot\_mg | -0.166 | 0.034 | 0.000 | -0.233 | -0.099 | -0.166 |
| pco1\_f | ~ | pH | -0.121 | 0.039 | 0.002 | -0.198 | -0.045 | -0.122 |
| pco1\_c | ~~ | pco1\_c | 0.075 | 0.008 | 0.000 | 0.058 | 0.091 | 0.075 |
| pco1\_f | ~~ | pco1\_f | 0.135 | 0.015 | 0.000 | 0.105 | 0.165 | 0.136 |
| pco1\_c | r2 | pco1\_c | 0.925 | NA | NA | NA | NA | NA |
| pco1\_f | r2 | pco1\_f | 0.864 | NA | NA | NA | NA | NA |

**Table S4**. Output of the sparse partial least square regression analysis (spls). Predictors (X) are fungal phyla (relative abundance of reads) and explanatory variables (Y) are the ten selected cercozoan clades (relative abundance of reads). The details of the spls anlaysis is provided in the main text. The analysis can be reproduced with the code available here: <https://github.com/biodiversaCercozoa_2023>.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Rhogostomidae | Assulinidae | Euglyphidae | Trinematidae | Viridiraptoridae | Fiscullidae | Sphenoderiidae | Cercomonadidae | Plasmodiophoridae | Tremulidae |
| Ascomycota | -0.263 | -0.14 | -0.207 | 0.053 | 0.019 | -0.241 | -0.238 | 0.186 | -0.006 | 0.191 |
| unclassified | 0.266 | 0.145 | 0.108 | 0.029 | 0.005 | 0.023 | -0.012 | -0.173 | -0.126 | -0.333 |
| Basidiomycota | 0.248 | 0.133 | 0.153 | -0.015 | -0.008 | 0.135 | 0.119 | -0.169 | -0.049 | -0.238 |
| Entomophthoromycota | 0.082 | 0.043 | 0.067 | -0.018 | -0.006 | 0.08 | 0.08 | -0.058 | 0.005 | -0.056 |
| Blastocladiomycota | -0.161 | -0.087 | -0.066 | -0.017 | -0.003 | -0.016 | 0.004 | 0.105 | 0.075 | 0.2 |
| Chytridiomycota | -0.076 | -0.041 | -0.05 | 0.007 | 0.003 | -0.048 | -0.044 | 0.052 | 0.011 | 0.069 |
| Olpidiomycota | -0.004 | -0.002 | 0.006 | -0.007 | -0.002 | 0.017 | 0.02 | 0.002 | 0.012 | 0.016 |
| Mucoromycota | -0.035 | -0.019 | -0.019 | 0 | 0.001 | -0.014 | -0.011 | 0.024 | 0.01 | 0.037 |
| Monoblepharomycota | -0.036 | -0.02 | -0.014 | -0.004 | -0.001 | -0.002 | 0.003 | 0.023 | 0.018 | 0.046 |
| GS19 | -0.057 | -0.031 | -0.02 | -0.009 | -0.002 | 0.002 | 0.011 | 0.036 | 0.031 | 0.076 |
| Glomeromycota | 0.007 | 0.002 | 0.073 | -0.056 | -0.016 | 0.152 | 0.174 | -0.015 | 0.087 | 0.087 |
| Zoopagomycota | -0.127 | -0.07 | -0.006 | -0.051 | -0.013 | 0.088 | 0.119 | 0.076 | 0.119 | 0.221 |
| Mortierellomycota | -0.059 | -0.034 | 0.05 | -0.066 | -0.019 | 0.155 | 0.186 | 0.027 | 0.123 | 0.174 |
| Rozellomycota | -0.125 | -0.07 | 0.02 | -0.071 | -0.019 | 0.143 | 0.182 | 0.071 | 0.151 | 0.253 |

**Table S5.** Total number of reads and proportion of the total per fungal phyla.

|  |  |  |
| --- | --- | --- |
| Fungal phyla | sumseq | prop |
| Ascomycota | 310975 | 0.50 |
| Basidiomycota | 80547 | 0.13 |
| Blastocladiomycota | 581 | 0.00 |
| Chytridiomycota | 6599 | 0.01 |
| Entomophthoromycota | 16 | 0.00 |
| Glomeromycota | 3458 | 0.01 |
| GS19 | 35 | 0.00 |
| Monoblepharomycota | 98 | 0.00 |
| Mortierellomycota | 133328 | 0.21 |
| Mucoromycota | 20 | 0.00 |
| Olpidiomycota | 793 | 0.00 |
| Rozellomycota | 1954 | 0.00 |
| unclassified | 81457 | 0.13 |
| Zoopagomycota | 357 | 0.00 |

**Table S6.** Total number of reads and proportion of the total per Cercozoan family

|  |  |  |
| --- | --- | --- |
|  | sumseq | prop |
| Agitatidae | 2405 | 0.05 |
| Allapsidae | 401857 | 8.01 |
| Assulinidae | 675 | 0.01 |
| Bodomorphidae | 870 | 0.02 |
| Cercomonadidae | 478085 | 9.53 |
| Euglyphidae | 51602 | 1.03 |
| Fiscullidae | 14441 | 0.29 |
| Limnofilidae | 51540 | 1.03 |
| Mesofilidae | 2564 | 0.05 |
| Novel\_Gran\_123456 | 30066 | 0.60 |
| Nudifilidae | 14571 | 0.29 |
| Paracercomonadidae | 269495 | 5.37 |
| Paulinellidae | 24252 | 0.48 |
| Peregriniidae | 170015 | 3.39 |
| Plasmodiophoridae | 9293 | 0.19 |
| Polymyxa\_lineage | 153821 | 3.07 |
| Proleptomonadidae | 40676 | 0.81 |
| Rhogostomidae | 1040229 | 20.74 |
| Sandonidae | 938837 | 18.72 |
| Sphenoderiidae | 1443 | 0.03 |
| Spongomonadidae | 62089 | 1.24 |
| Spongospora\_lineage | 101488 | 2.02 |
| Thaumatomonadidae | 63725 | 1.27 |
| Tremulidae | 38079 | 0.76 |
| Trinematidae | 185582 | 3.70 |
| unclassified | 864230 | 17.23 |
| Viridiraptoridae | 4485 | 0.09 |