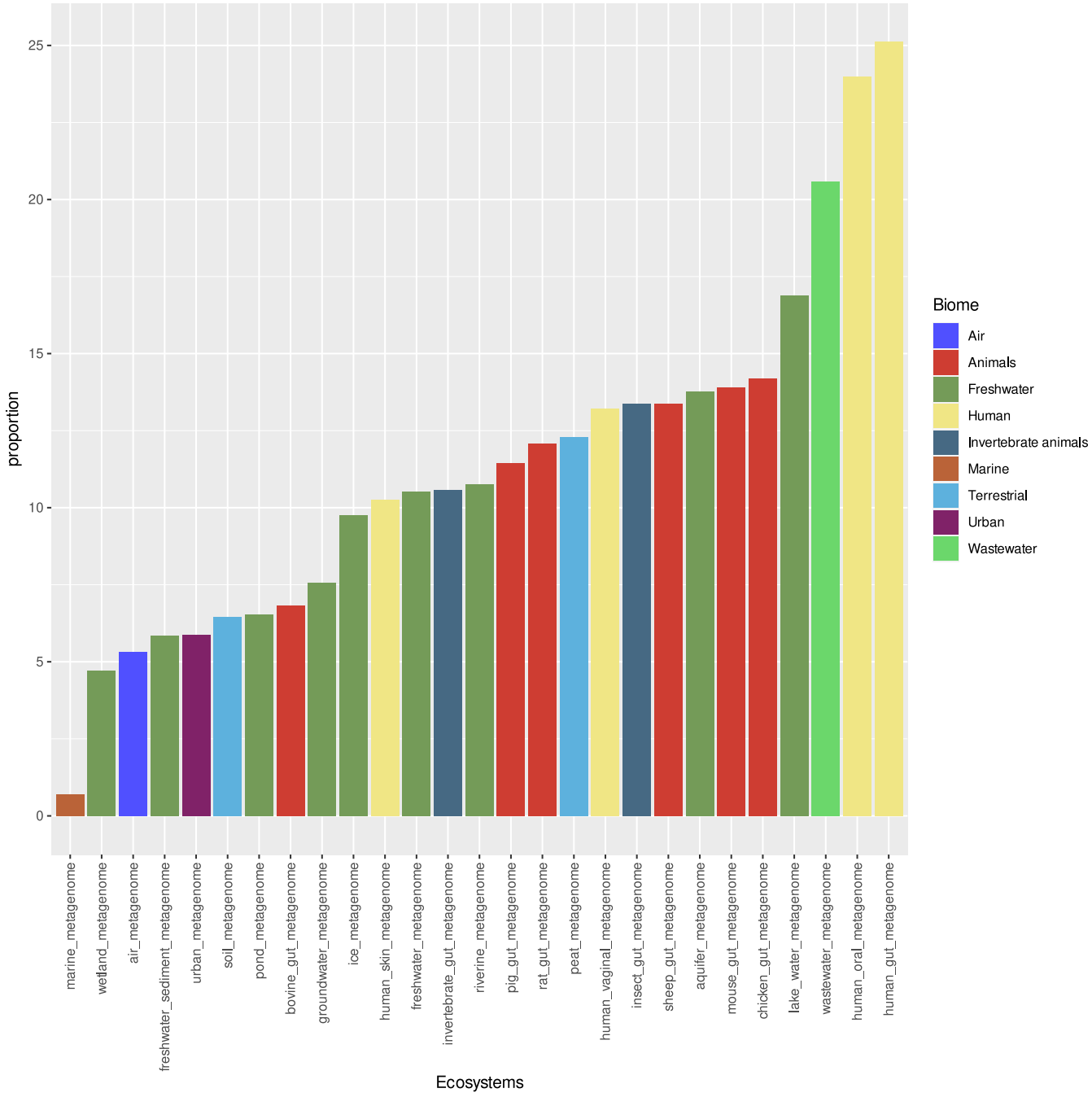
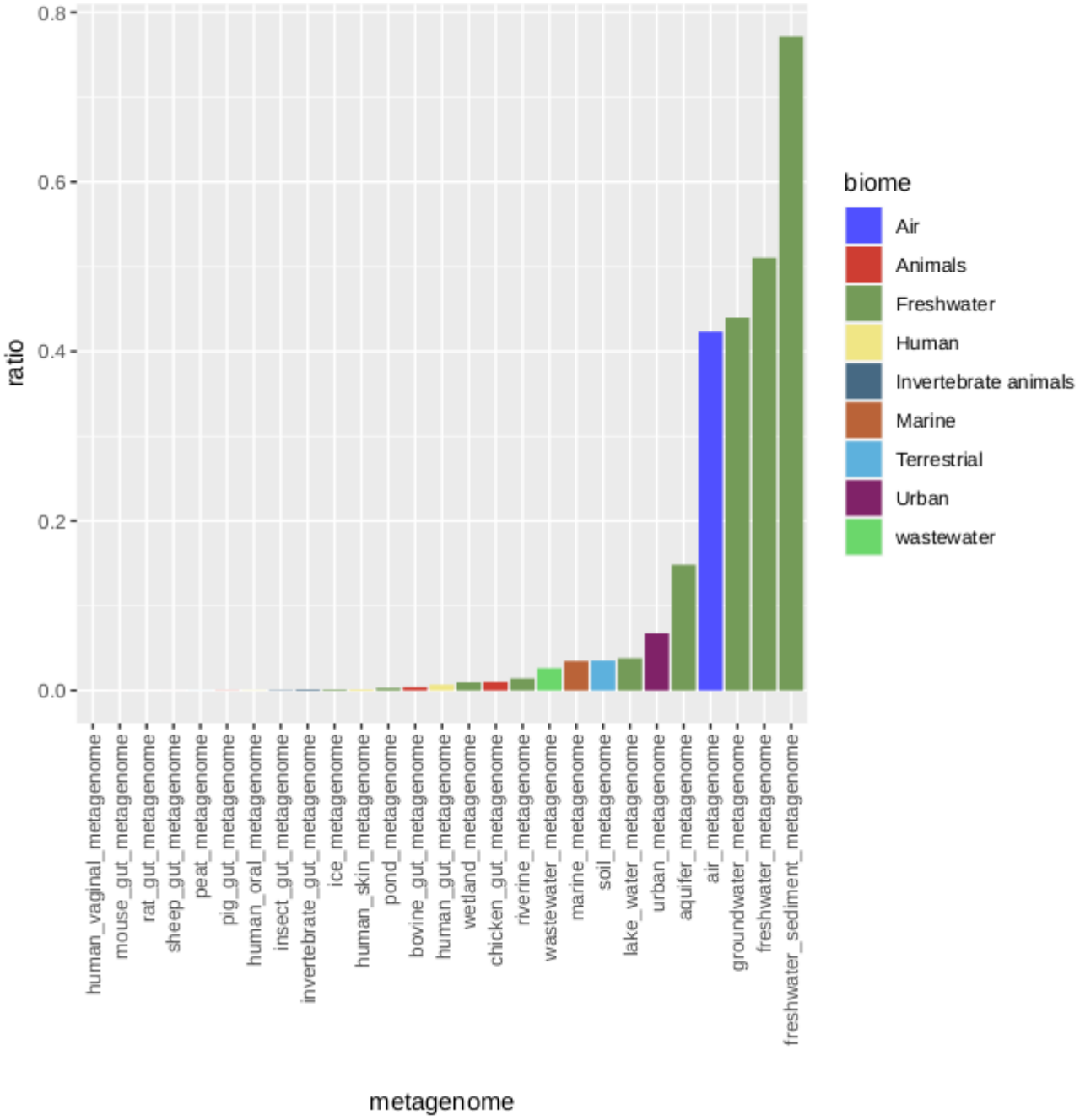
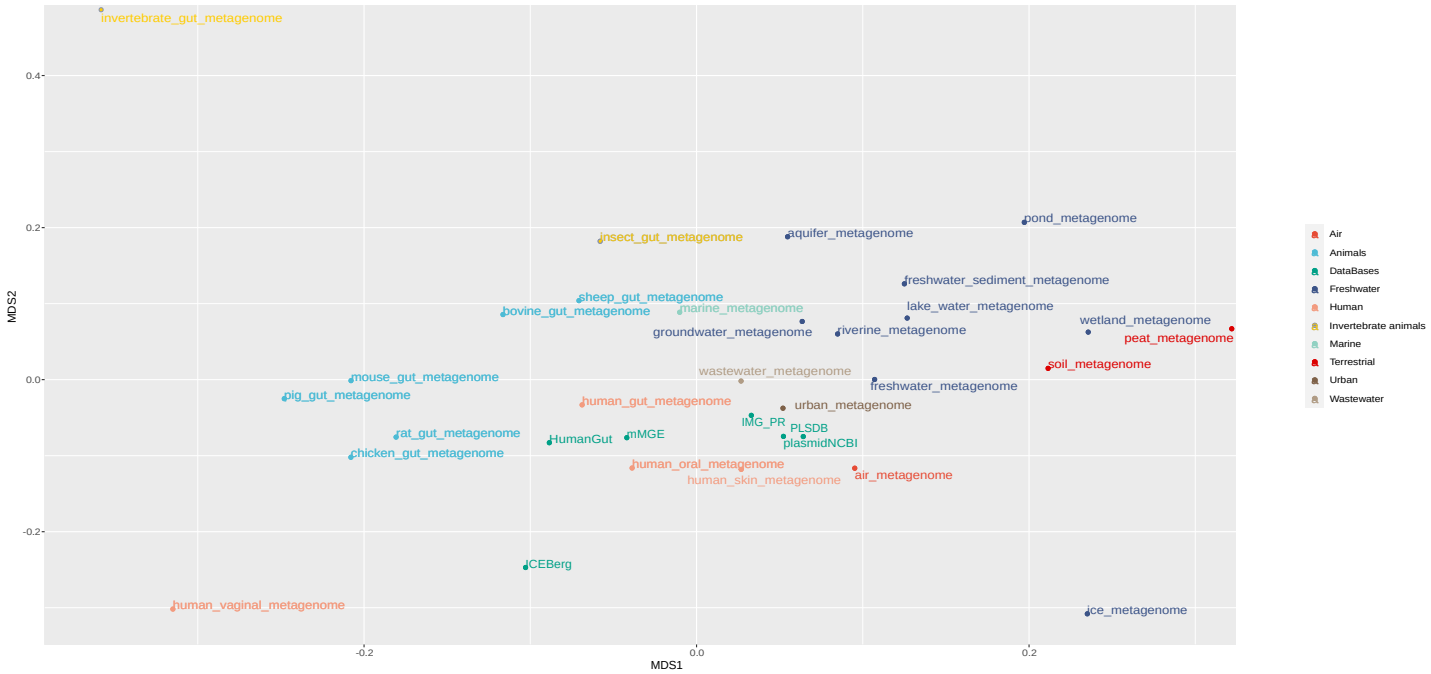


Supplementary Fig 1: proportion of plasmids expressed in base pairs (bp) recruited (i.e. mapping to the PLSs) related to the total bp mapped



Supplementary Fig. 2: Detection of the *att*C sites per megabase in various ecosystems categorized by biomes.



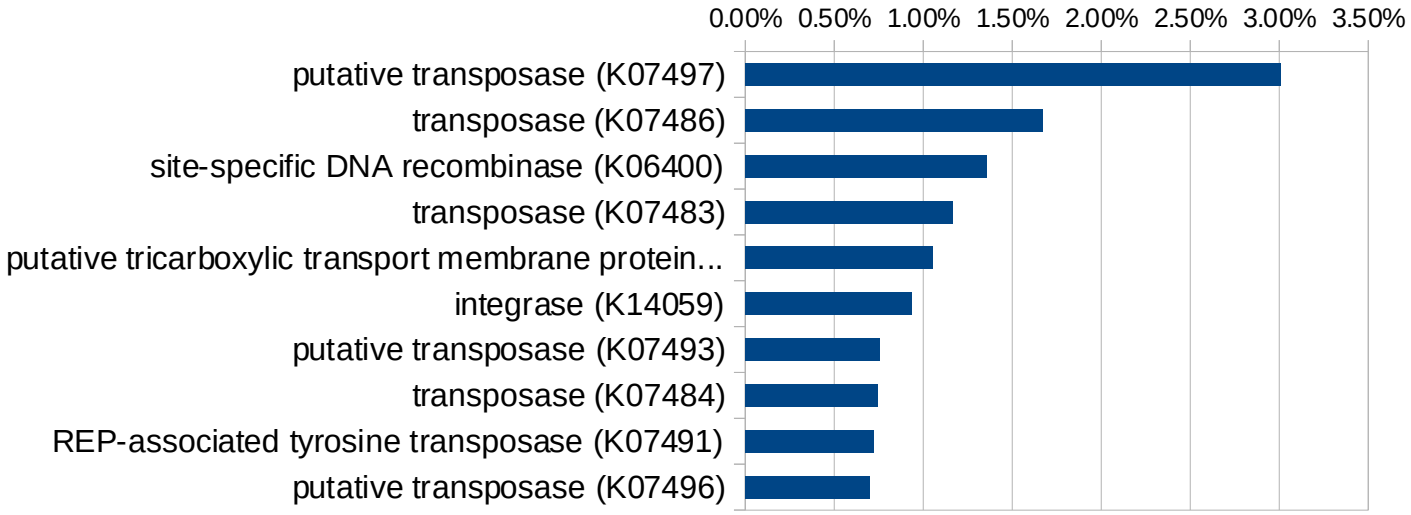
Supplementary Fig. 3: NMDS analysis of the Bray-Curtis distance between ecosystems or biomes and main databases. The databases included in this analysis were ICEberg, plasmids from RefSeq and PLSDB, HumanGut (https://gmrepo.humangut.info/data) and IMG/PR. The distances between ecosystems and reference databases were computed with simka (-max-read 1000000 -nb-kmers 10000000).

Camargo, A.P., Call, L., Roux, S., Nayfach, S., Huntemann, M., Palaniappan, K., et al. (2023) IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. *Nucleic Acids Research* gkad964.

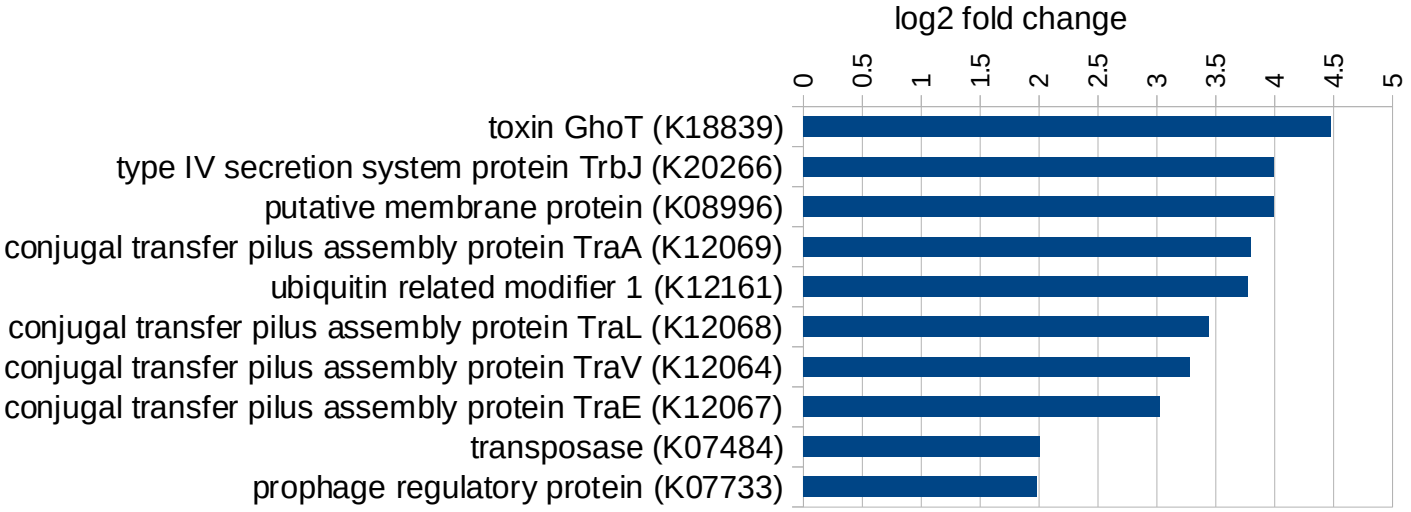
Galata, V., Fehlmann, T., Backes, C., and Keller, A. (2019) PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Research* **47**: D195–D202.

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., et al. (2019) ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* **47**: D660–D665.

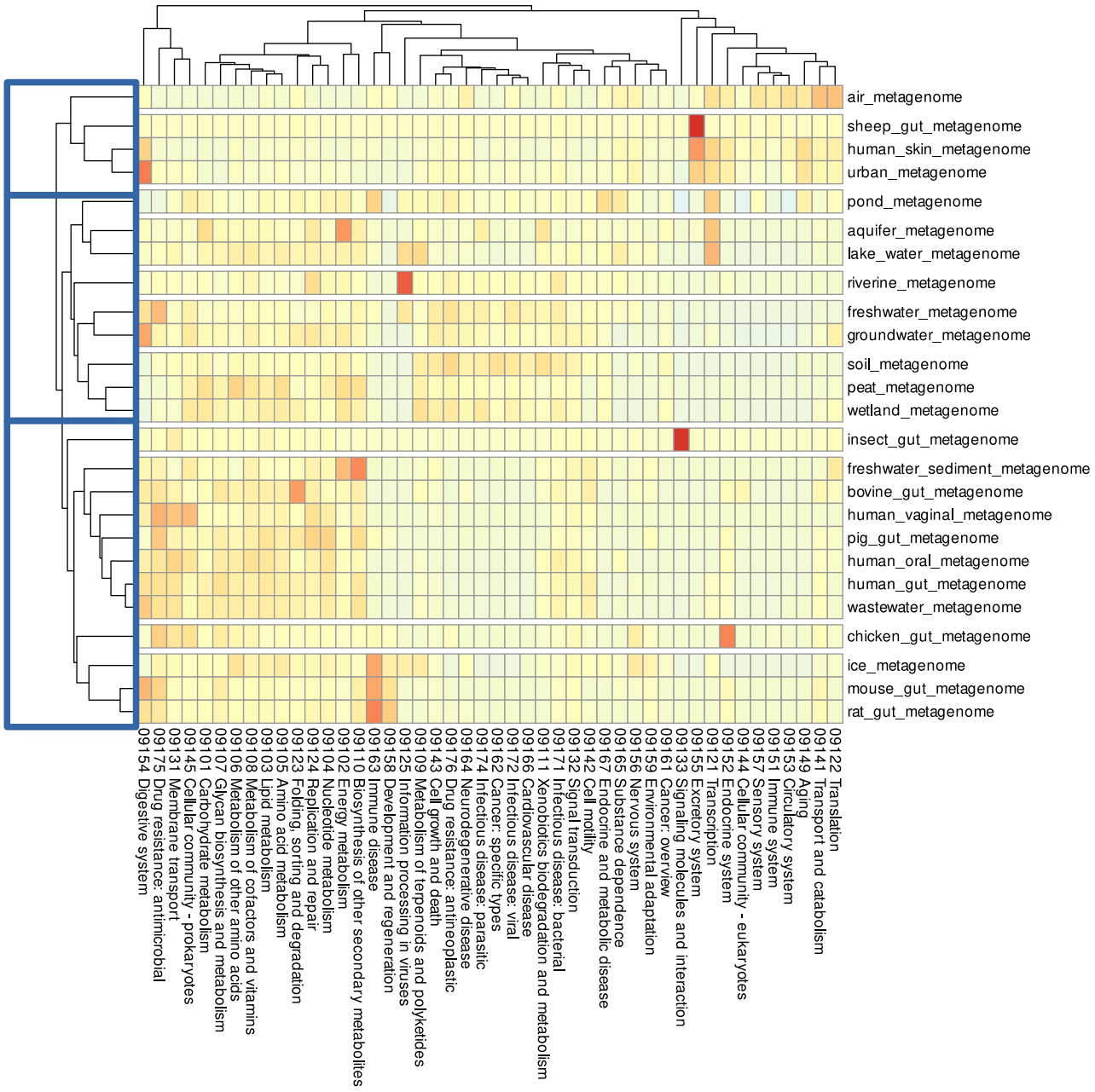
A



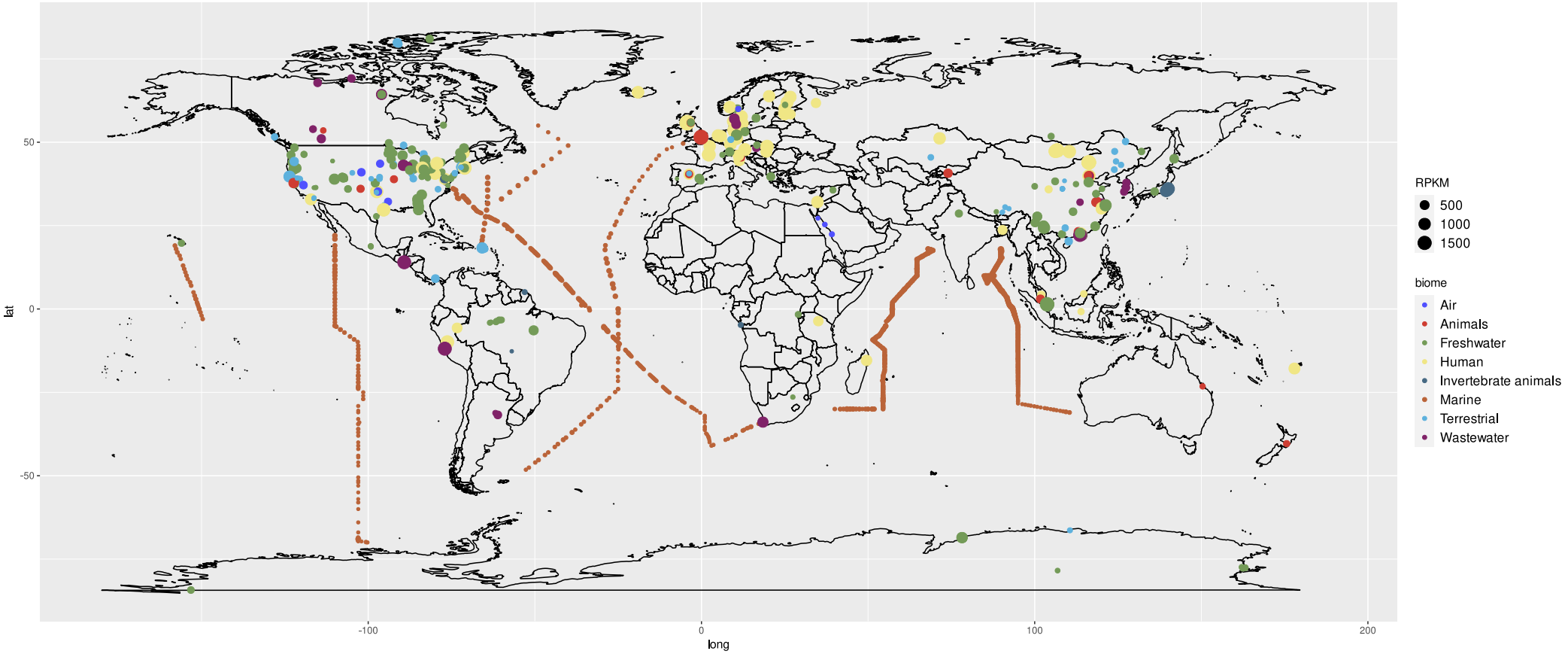
B



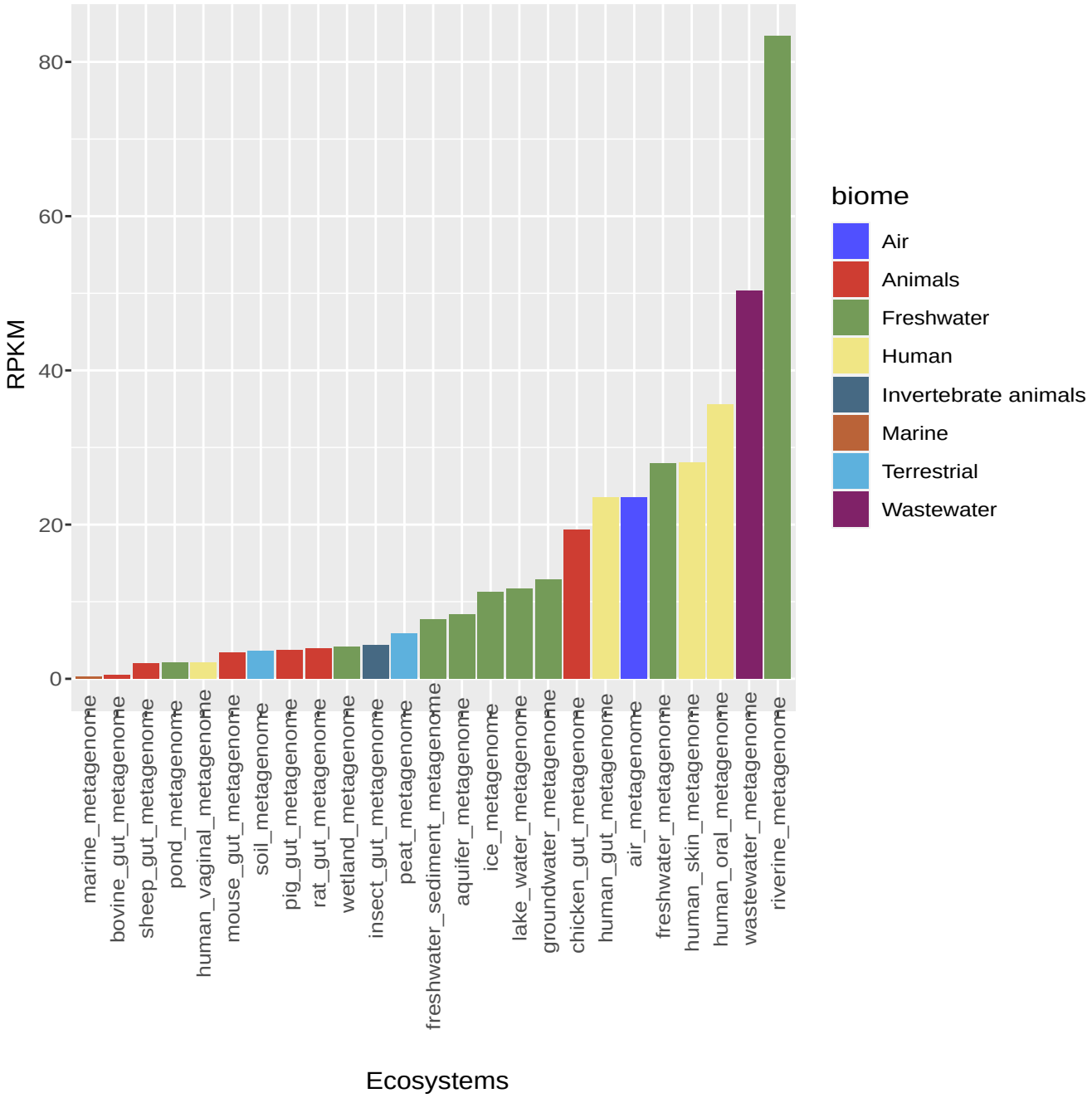
Supplementary Fig 4: Annotation of the PLCs (Plasmid-like clusters) by the KEGG database. A) The ten most abundant KEGG orthologs B) Top differentially abundant KEGG orthologs between the metaplasmidome and bacterial chromosomes, as inferred from DESeq2 analysis (Padj < 0.05).



Supplementary Fig. 5: Main kegg pathways inferred from read recruitments on the genes predicted from PLCs (Supplementary Fig. 15D and E). Three main clusters, highlighted by blue rectangles, are described in detail within this paper.

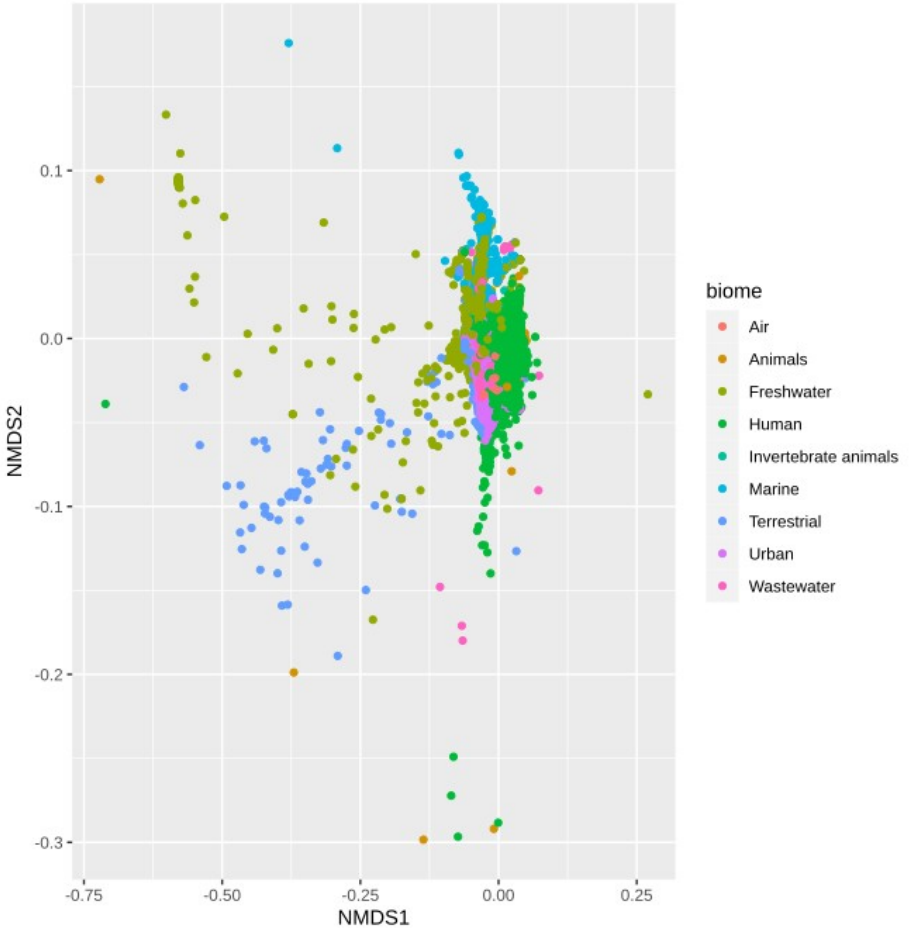


Supplementary Fig. 6: ARGs abundances harbored by metaplasmidome in the different ecosystem investigated across the world . The quantification was assessed after mapping reads against genes on PLCs (Supplementary Fig. 15D)

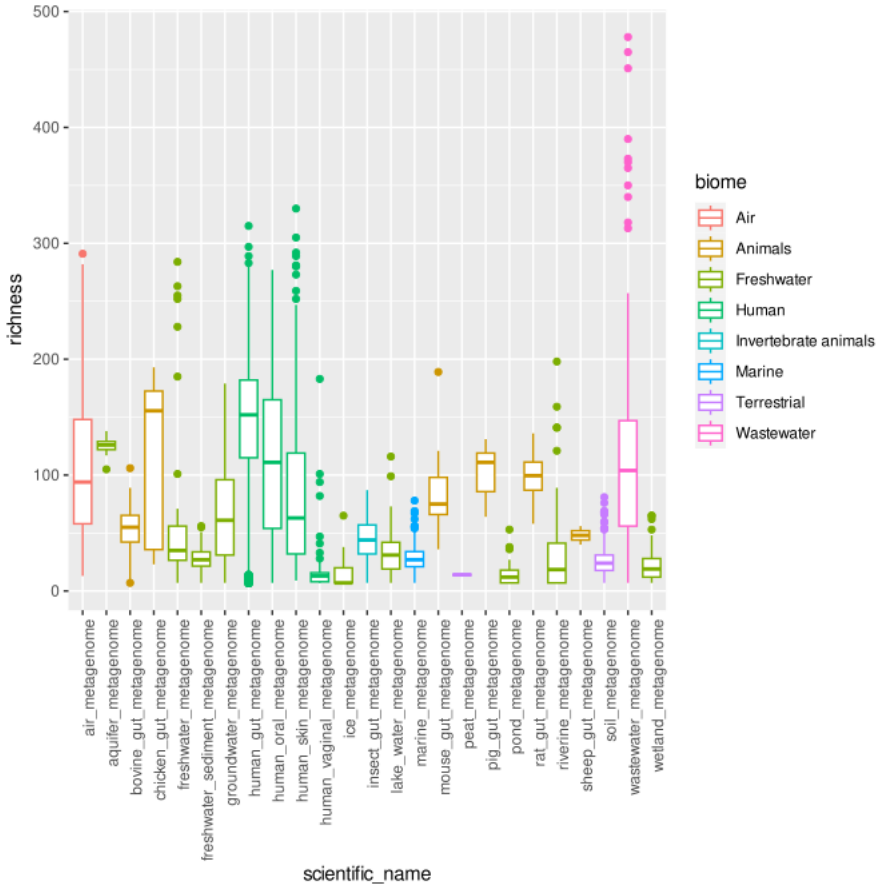


Supplementary Fig. 7: MRGs abundances harbored by metaplasmidome in the different ecosystem investigated (A) and heat-map of the resistance categories. The quantification was assessed after mapping reads against genes on PLCs (Supplementary Fig. 15D).

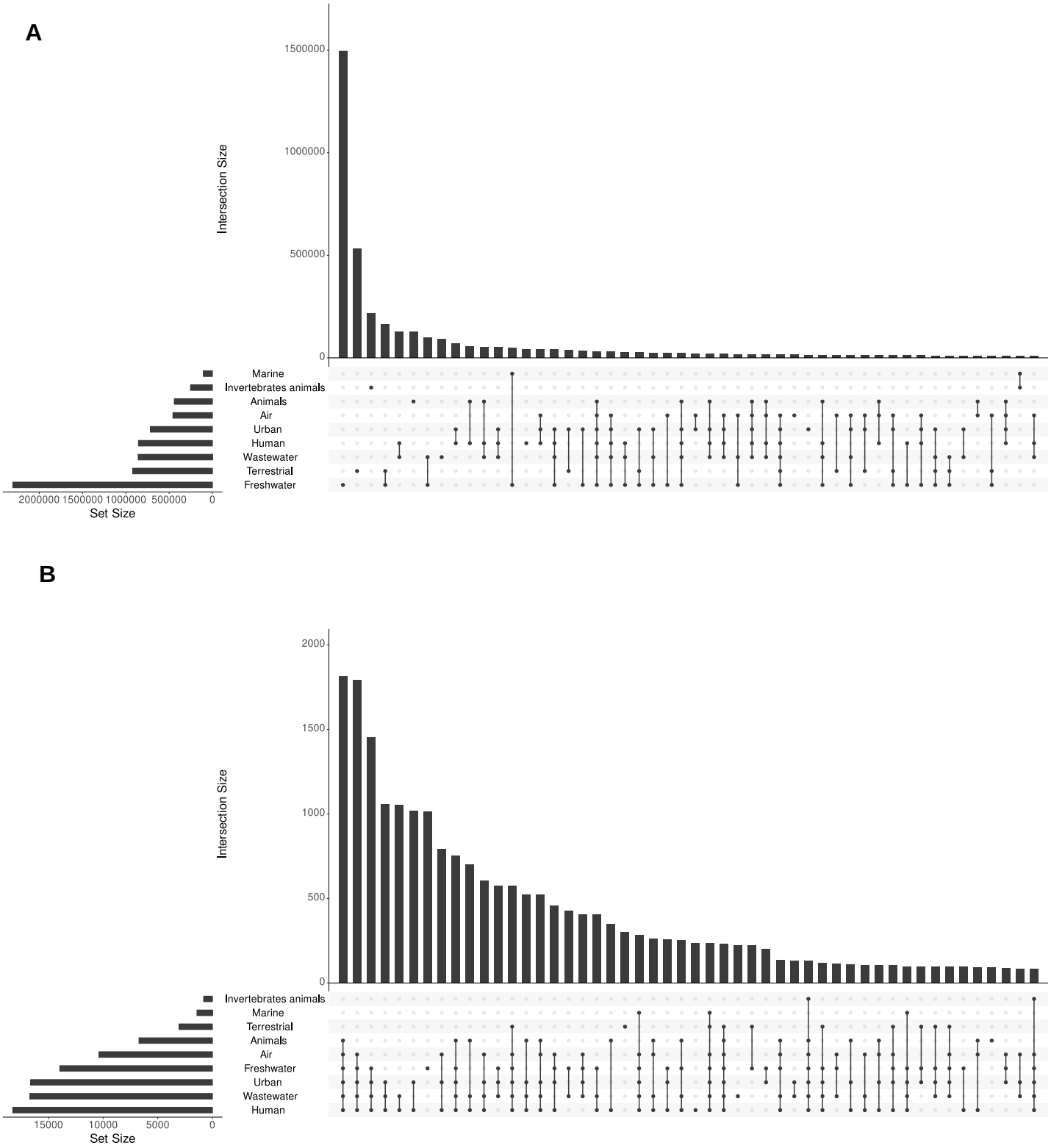
A



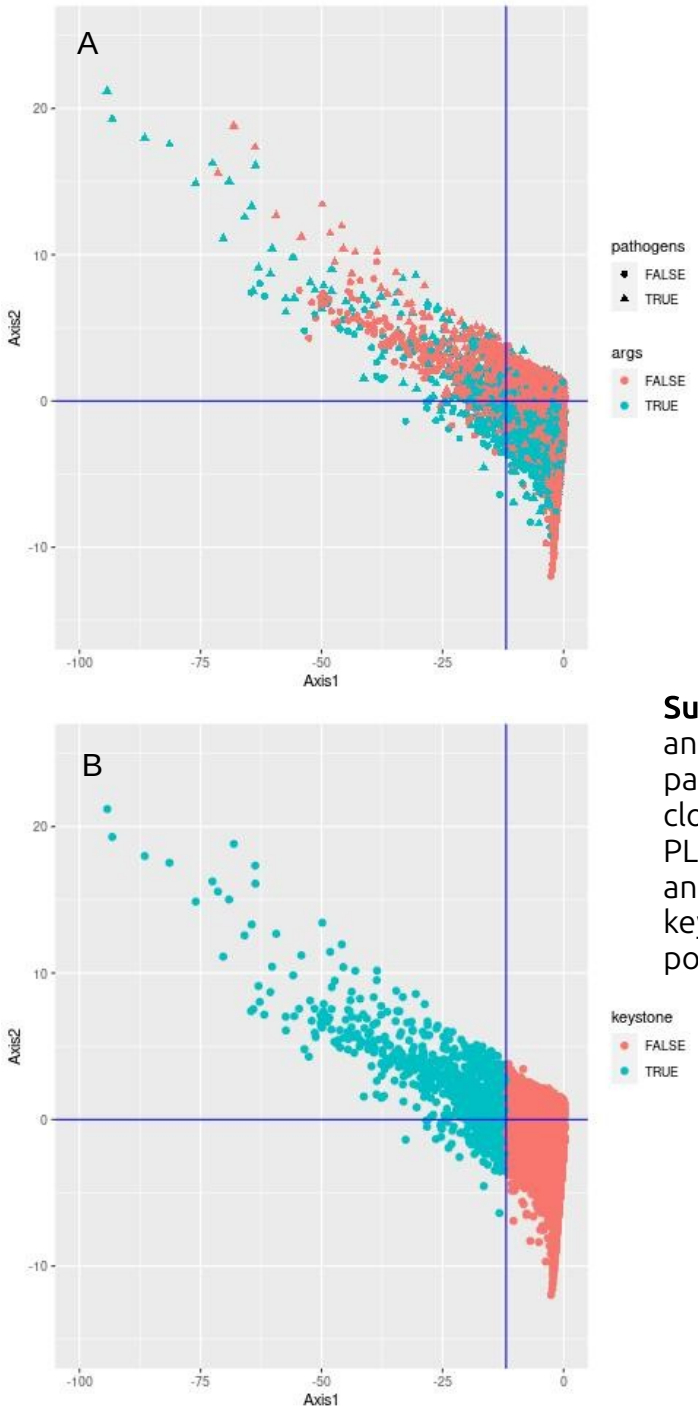
B



Supplementary Fig. 8: NMDS ordination based on Bray-Curtis distances calculated from taxonomic profiles (A) and microbial richness across ecosystems (B). Taxonomic profiling for each metagenome was performed using MetaPhlAn3 on a dataset of 20,000,000 reads.



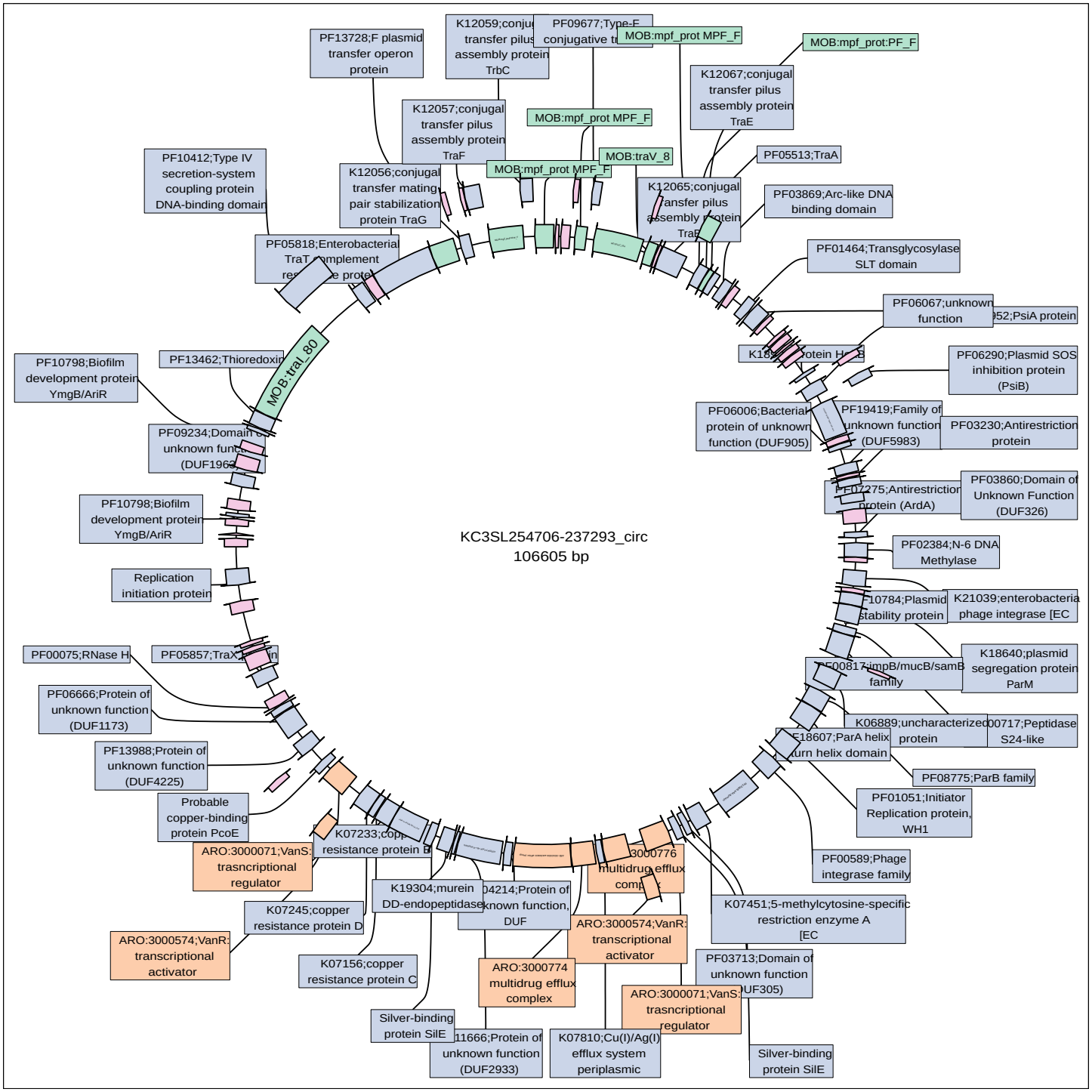
Supplementary Fig. 9: Intersection size between different biomes for A) total PLCs and B) PLCs restricted to known PLCs. These last PLCs were referenced in plasmid RefSeq database or carried plasmid markers.



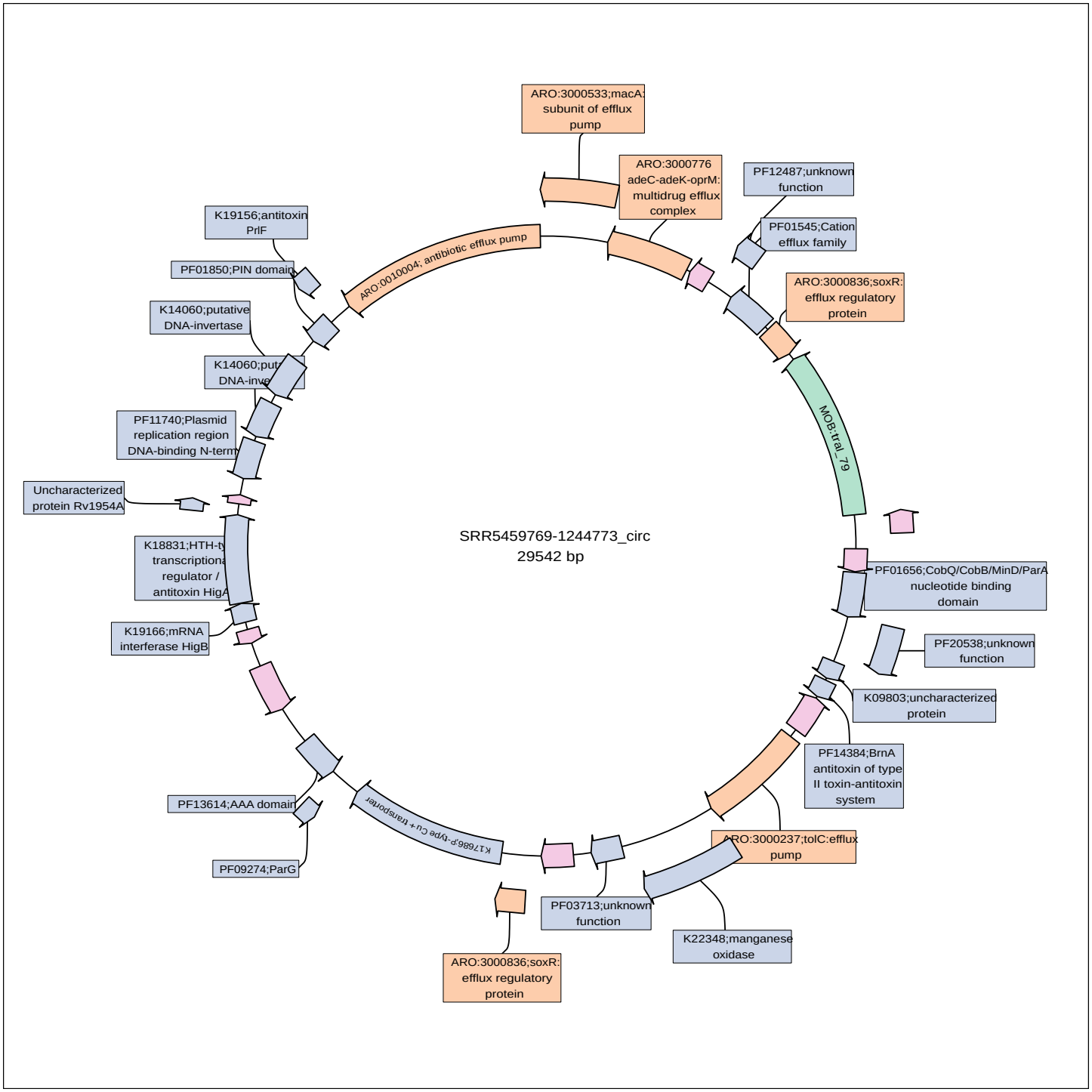
	Comp1	Comp2
degree	-9.560090e-01	1.326978e-01
betweenness	-6.697485e-01	1.956913e-01
closeness	-1.523366e-13	7.843911e-13
strength	-9.560090e-01	1.326978e-01
coverage	-2.153939e-01	-8.332381e-01
frequence	-2.503849e-01	-8.199779e-01

Supplementary Fig. 10: Principal component analysis (PCA) processed from the main centrality parameters of the network (degree, betweenness, closenesss), locations numbers (frequence) and PLCs coverages. A) PLCs linked to pathogens and/or harbored ARGs in the network and B) keystone PLCs (keyPLCs) selected from tipping points on the Axis 1.

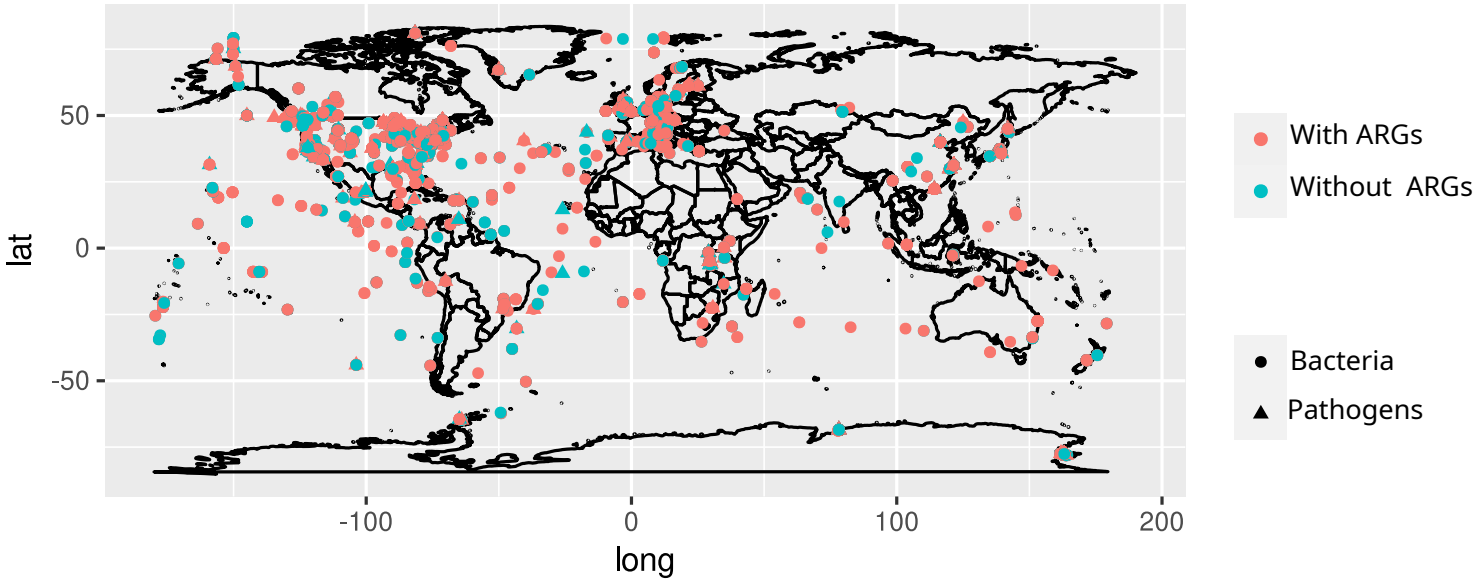
The PCA, processed by ADE4 under R software, showed that betweenness, strength and degree are negatively correlated with the first axis, while closeness is not correlated with the axes of this analysis. The second axis enables discrimination of highly frequent PLCs (present in numerous ecosystems) with correspondingly high coverages. The most influential PLCs in this network are therefore located on the left side of the factorial plan, with the highest values of betweenness, strength and degree. To define the keystone PLCs (keyPLCs), a coordinate threshold on the first axis was determined by the R package “changepoint”. The keyPLCs therefore defined (1022) are displayed in green on the panel B.



Supplementary Fig. 11: Annotation of a plasmid belonging to generalist PLC (i.e. keyPLC)

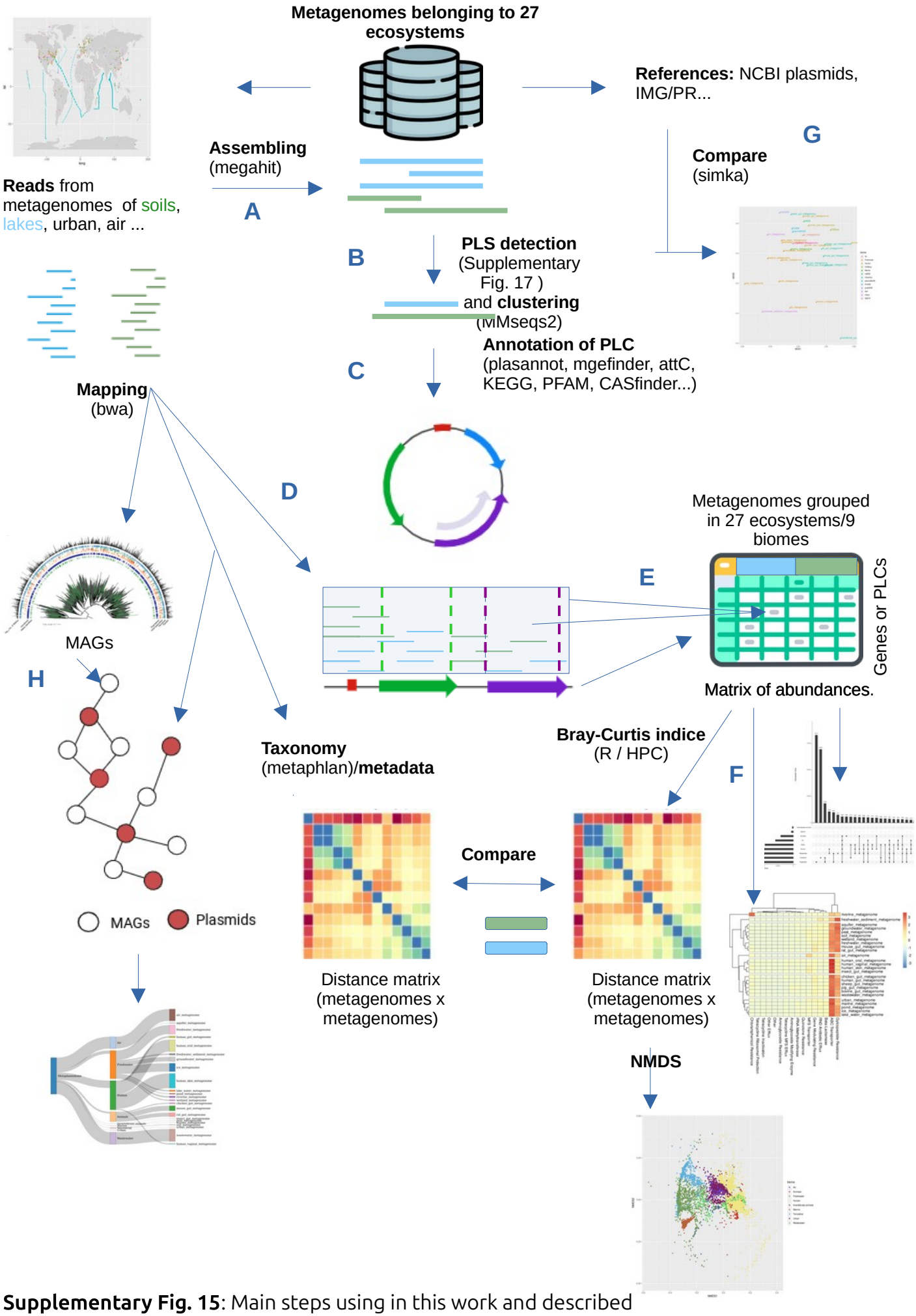


Supplementary Fig. 12: Annotation of a plasmid belonging to generalist PLC (i.e. keyPLC)

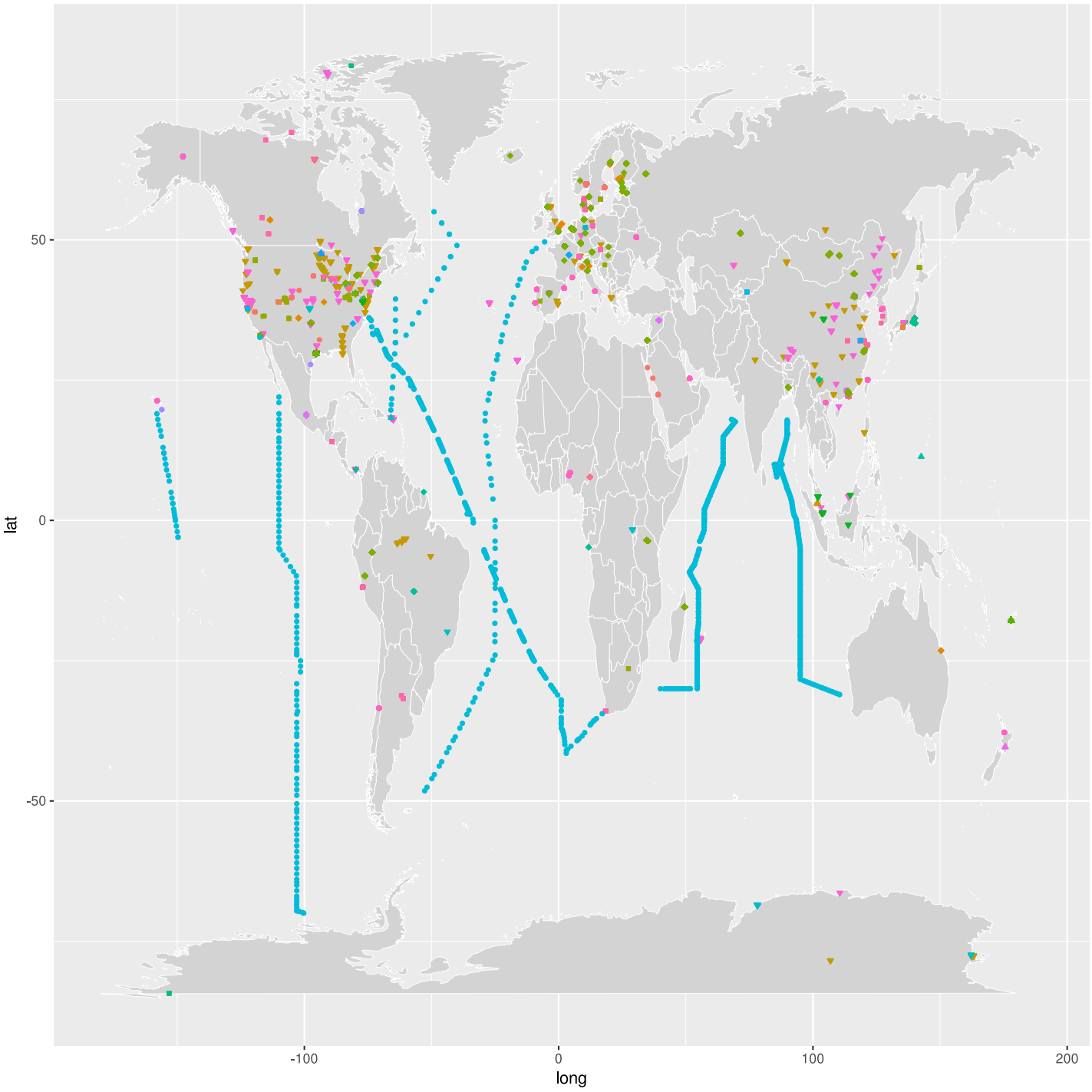


Supplementary Fig. 13: Global distribution of micro-organisms (MAGs) in relation to their putative pathogenicity and their association with antibiotic resistance genes (ARGs) carried by plasmids The pathogens were inferred from the taxonomy defined by Nayfach et al. (2021), while the ARGs characteristics were determined based on the relationship between plasmids harbouring ARGs and MAGs defined by bipartite network analysis.

Supplementary Fig. 14: Taxonomic composition of bacteria linked to keyPLCs found in human metagenomes and at least in 4 other biomes.



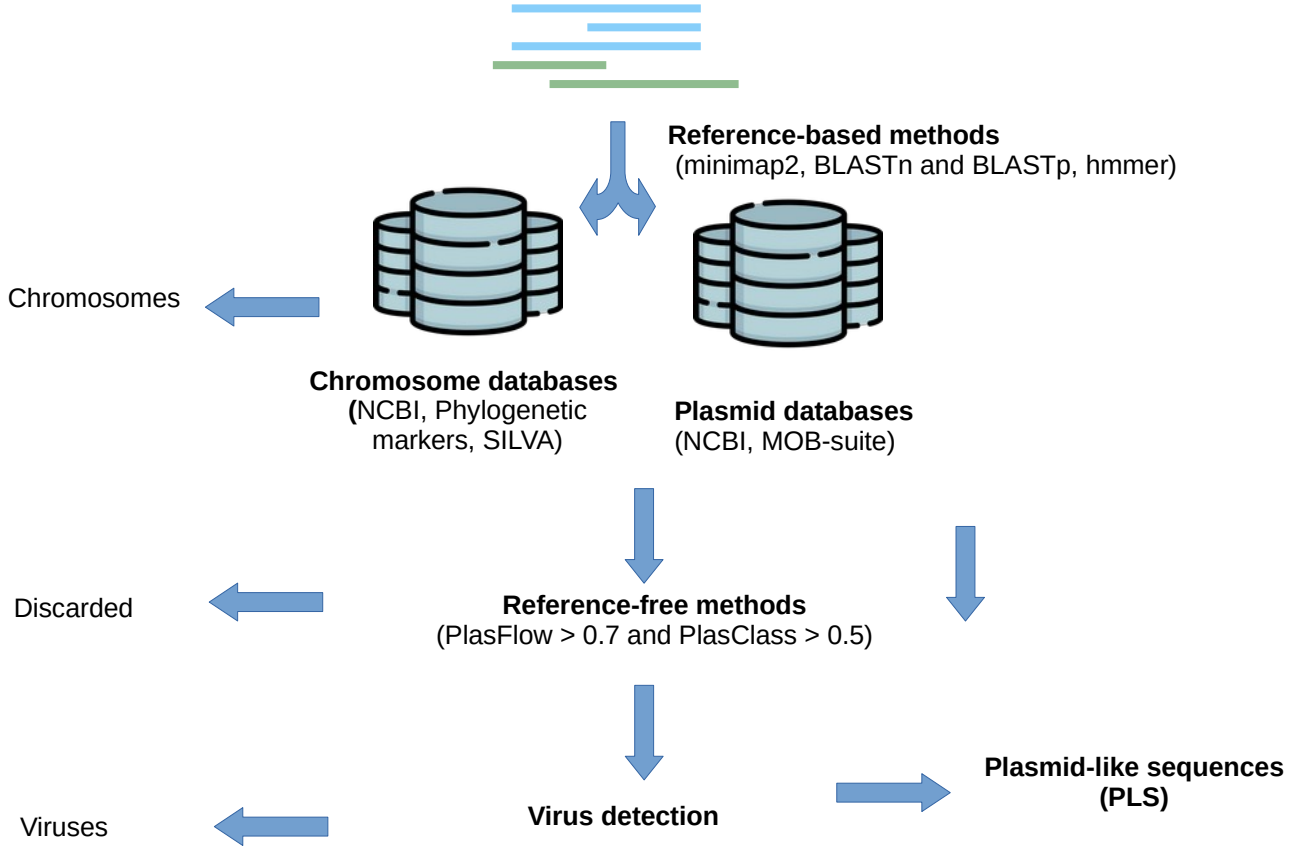
Supplementary Fig. 15: Main steps using in this work and described In the Materials and methods part of this work



scientific_name

- air metagenome
- aquifer metagenome
- bovine gut metagenome
- chicken gut metagenome
- freshwater metagenome
- freshwater sediment metagenome
- groundwater metagenome
- human gut metagenome
- human oral metagenome
- human skin metagenome
- human vaginal metagenome
- ice metagenome
- insect gut metagenome
- invertebrate gut metagenome
- lake water metagenome
- marine metagenome
- mouse gut metagenome
- Other
- peat metagenome
- pig gut metagenome
- pond metagenome
- rat gut metagenome
- riverine metagenome
- sheep gut metagenome
- soil metagenome
- urban metagenome
- wastewater metagenome
- wetland metagenome

Supplementary Fig. 16: Geographic distribution of metagenomes categorised by ecosystems (i.e. scientific names)



Supplementary Fig. 17: Plasmids were predicted for each assembly (Supplementary Fig. 15B) by using scripts describing in-depth in Hilpert et al. (2021) and available in github website (https://github.com/meb-team/PlasSuite/). In the first step, the contigs were compared to various databases and were retained, discarded or in absence of significant results analysed by a reference-free method. The contigs with plasmid markers were retained in the final results regardless of the outcomes of the free-reference tools. Only the contigs detected as plasmids by PlasFlow (>0.7) and PlasClass (> 0.5) were selected. Combining both methods reduces the false positive. Viruses were checked by ViralVerify that can classified also plasmids and chromosomes. The viral detection creates computational difficulties since some MGEs termed 'phage-like plasmids' encode viral structural genes to facilitate the transduction and replicating within bacteria as a plasmid. Among the tools available at the beginning of this work (e.g. PPR-Meta, VirusSorter2), ViralVerify outperformed other methods to avoid false positives against a plasmid database (results no shown)

Hilpert, C., Bricheux, G., and Debroas, D. (2021) Reconstruction of plasmids by shotgun sequencing from environmental DNA: which bioinformatic workflow? *Briefings in Bioinformatics* **22**: bbaa059.
Krawczyk, P.S., Lipinski, L., and Dziembowski, A. (2018) PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res* **46**: e35.
Pellow, D., Mizrahi, I., and Shamir, R. (2020) PlasClass improves plasmid sequence classification. *PLOS Computational Biology* **16**: e1007781.
Antipov, D., Raiko, M., Lapidus, A. & Pevzner, P. A. MetaviralSPAdes: assembly of viruses from metagenomic data. *Bioinformatics* **36**, 4126–4129 (2020).
Fang, Z. *et al.* PPR-Meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. *GigaScience* **8**, (2019).
Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).