

Supplementary material
Host-associated phages disperse across the extraterrestrial analogue
Antarctica

**Janina Rahlff^{a, #}, Till L.V. Bornemann^a, Anna Lopatina^b,
Konstantin Severinov^b, Alexander J. Probst^a**

^aGroup for Aquatic Microbial Ecology, University of Duisburg-Essen, Department of
Chemistry, Environmental Microbiology and Biotechnology (EMB), Universitätsstraße 5,
45141 Essen, Germany

^bInstitutes of Molecular Genetics and Gene Biology of the Russian Academy of Sciences,
Moscow, Russia

#Corresponding author: Janina Rahlff, Janina.rahlff@uol.de

Present address: Centre for Ecology and Evolution in Microbial Model Systems (EEMiS),
Department of Biology and Environmental Science, Linnaeus University, SE-39182, Kalmar,
Sweden

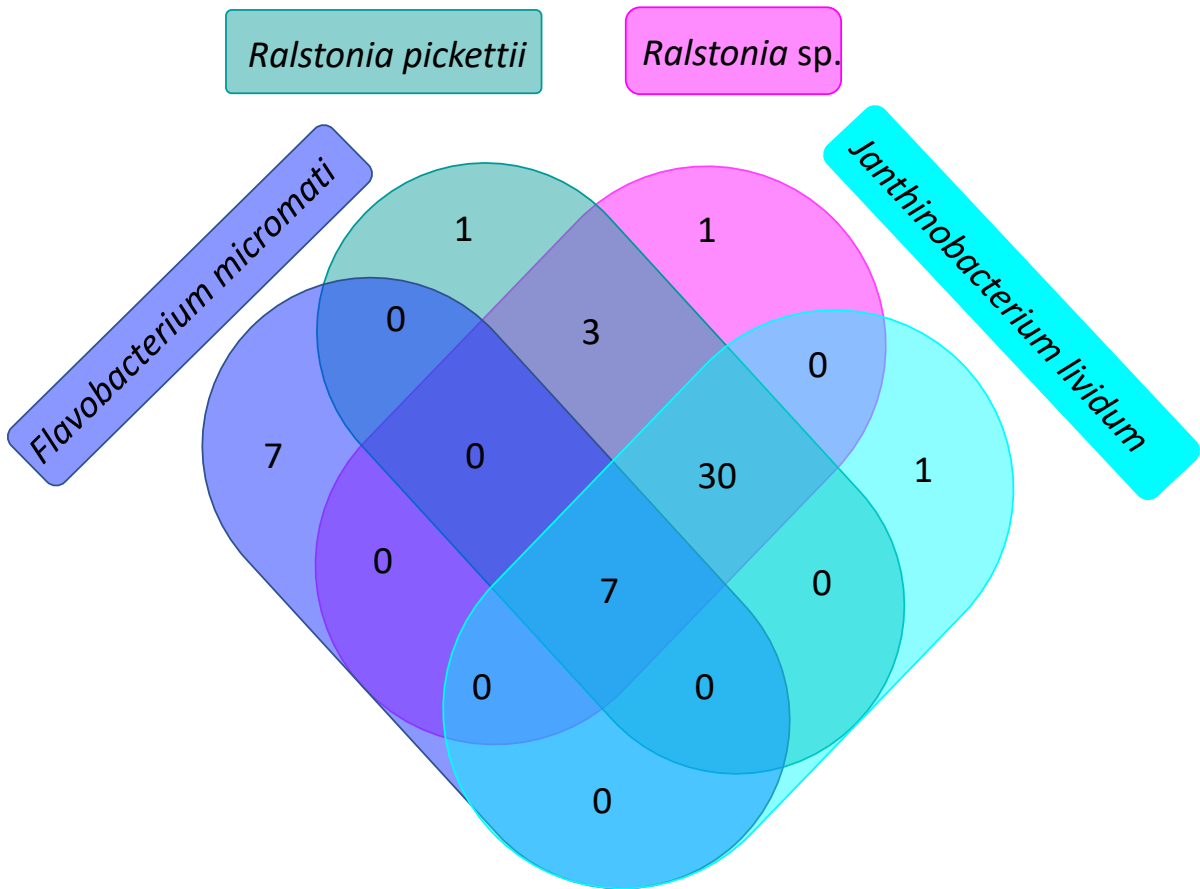
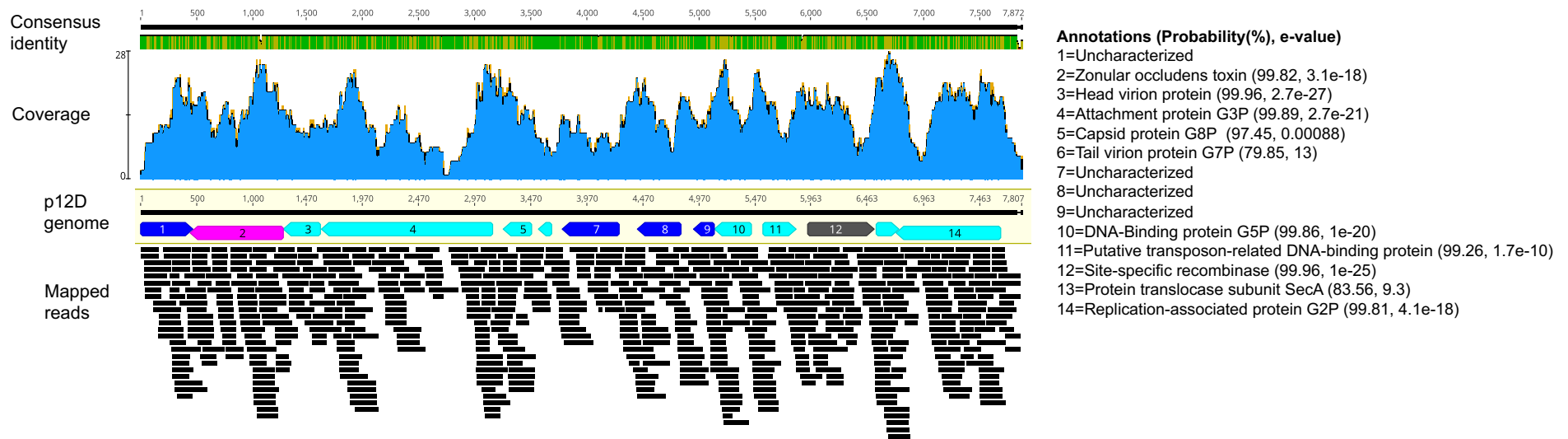


Figure S1: Venn-diagram showing overlapping phage-host matches of the 50 unique matching Antarctic PGFs as determined by VirHostMatcher (1).



1
2
3
4
5
6
7
8

Figure S2: Open reading frames and protein annotations of the phage genome of Antarcophage49_Dr_7823_circ. Since this complete phage genome was found integrated in the chromosome of *Ralstonia* 12D, we propose the name *Ralstonia* phage p12D. Annotations of proteins were performed with HHpred (2, 3) against PDB, Pfam, UniProt-SwissProt-viral and NCBI_Conserved Domains using a probability threshold of 70%. Uncharacterized proteins are colored dark blue, the zot ORF in pink, the resolvase/recombinase in grey and remaining ORF in light blue. Reads from Druzhnaja were mapped against the genome using bowtie -sensitive mode (4). Genes that do not align by height overlap in their sequence. Visualization was performed in Geneious Prime version 11.1.5 (5).

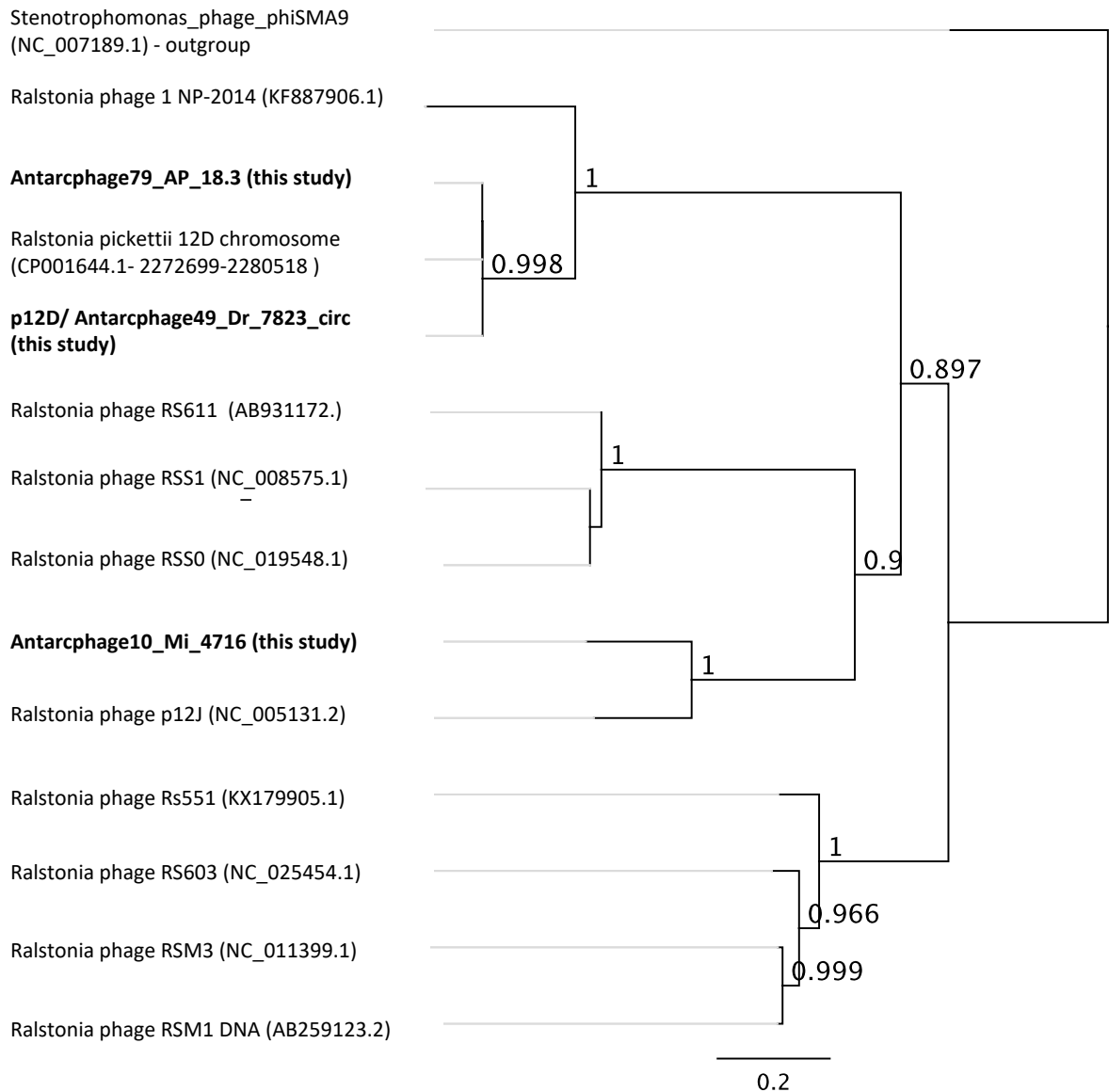


Figure S3: Phylogenetic tree constructed with FastTree 2.1.11 (6) in Geneious 11.1.5 under default settings for MUSCLE-aligned (7) nucleic acid sequence of the Zot protein of known *Ralstonia* phages and Antarctic phage genome fragments.

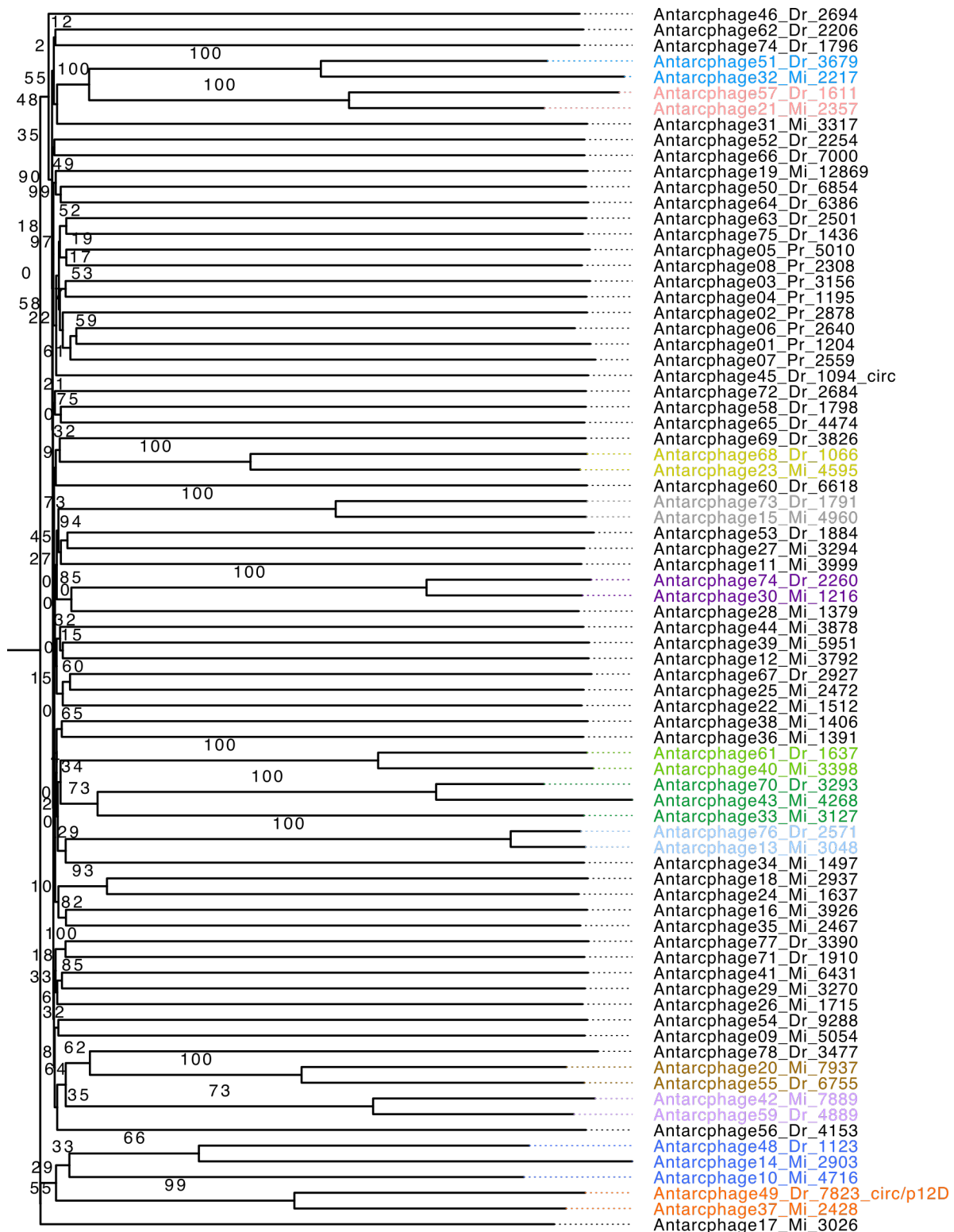


Figure S4: Nucleic acid-based clustering of Antarctic PGFs according to VICTOR (8). PGFs clustering on genus level (12 viral clusters) are color-coded. Branch support was inferred from 100 pseudo-bootstrap replicates each. Circ=circular, Dr=Druzhnaja, Mi=Mirni (station), Pr=Progress (station). Visualization was performed with FigTree v.1.4.4. (9)

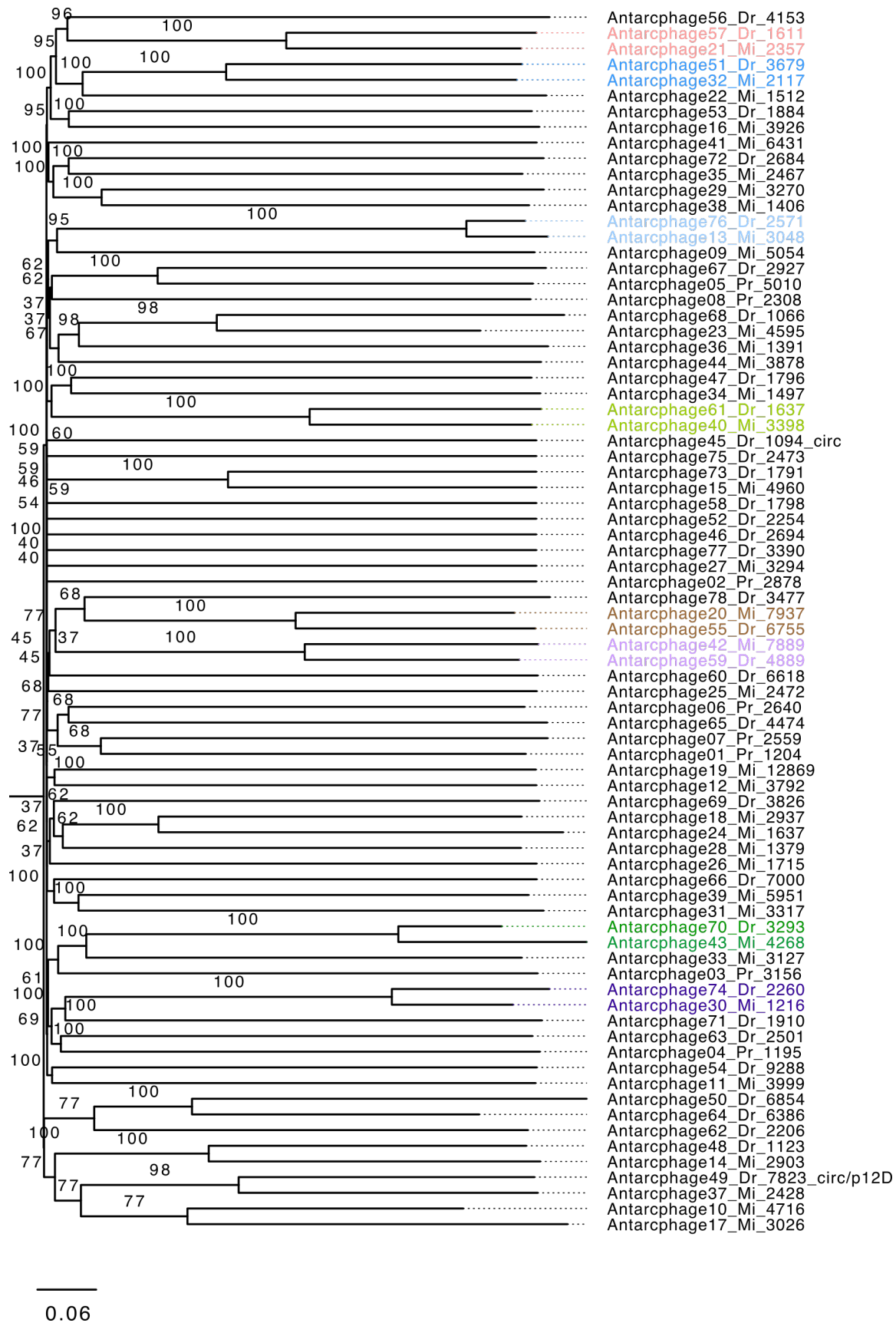


Figure S5: Amino acid-based clustering of Antarctic PGFs according to VICTOR (8). PGFs clustering on genus level (8 viral clusters) are color-coded. Branch support was inferred from

100 pseudo-bootstrap replicates each. Circ=circular, Dr=Druzhnaja, Mi=Mirni (station), Pr=Progress (station). Visualization was performed with FigTree v.1.4.4. (9).

References

1. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. 2017. Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res* 45:39-53.
2. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33:W244-8.
3. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430:2237-2243.
4. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357-9.
5. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647-9.
6. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490.
7. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
8. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* 33:3396-3404.
9. Rambaut A. 2006. FigTree, a graphical viewer of phylogenetic trees and as a program for producing publication-ready figures.