

Supplementary Information

*Hologenomic insights demonstrate co-evolution between an intestinal
Mycoplasma and its salmonid host*

Tags: Comparative genomics, Host adaptation, Mycoplasma, Salmonids, Hologenomic evolution

Quality Overview of sequencing data

Table S1) Metagenomic sequencing output and processing

	READS (MILLIONS)			PERCENTAGE	
	Raw	Mapped to host	Unmapped (Bacteria)	Mapped to host	Unmapped (Bacteria)
Mean	66.35	35.40	27.90	53.82%	40.74%
Minimum	0.52	0.31	0.01	6.84%	0.10%
Maximum	108.93	100.12	73.39	99.74%	93.75%
SD	24.96	25.44	20.71	27.72%	27.85%

Remapping of eukaryotic gut content post host filtering

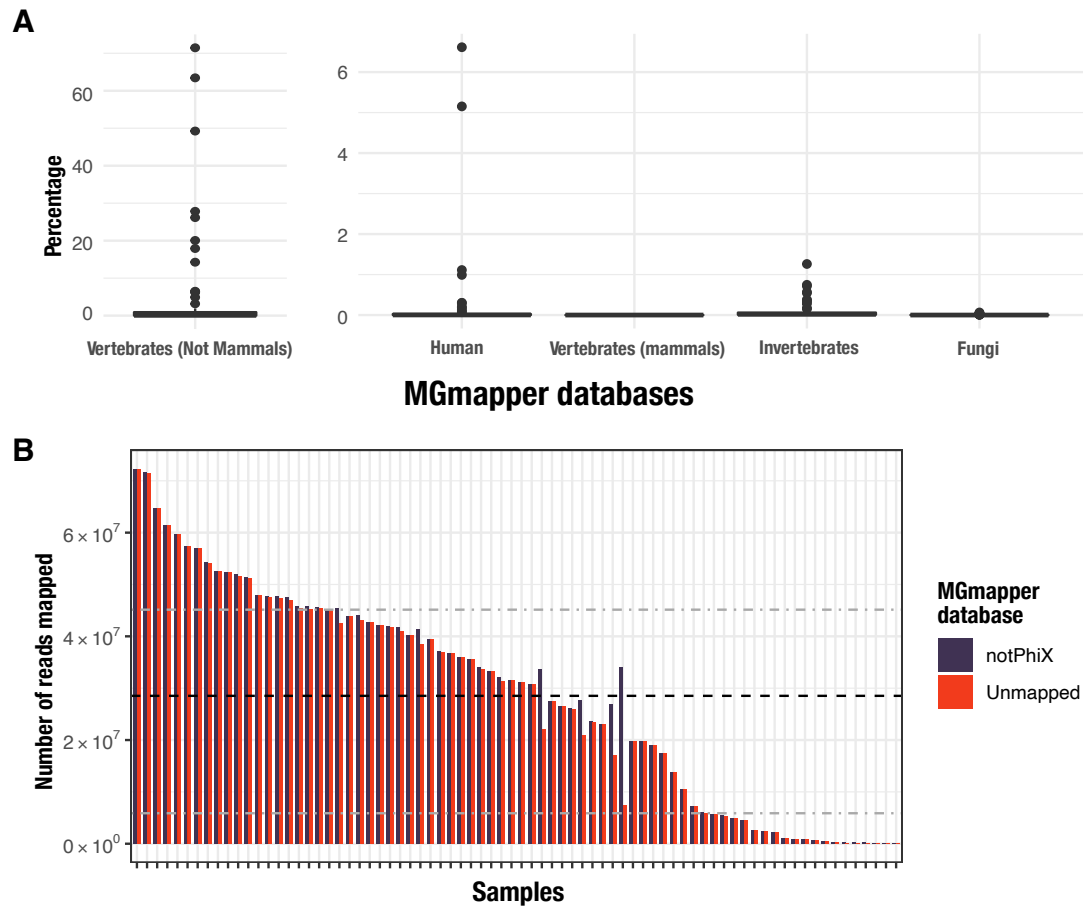
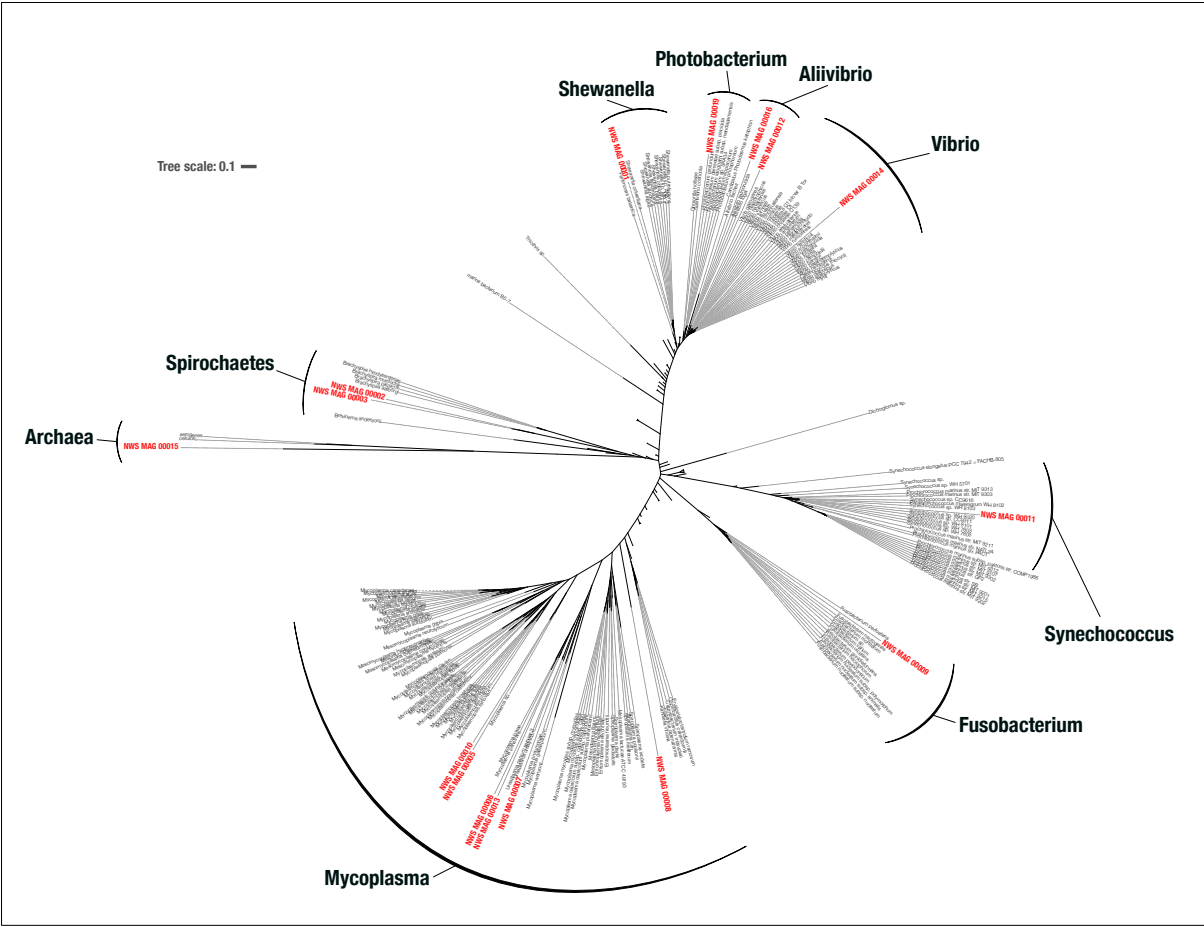


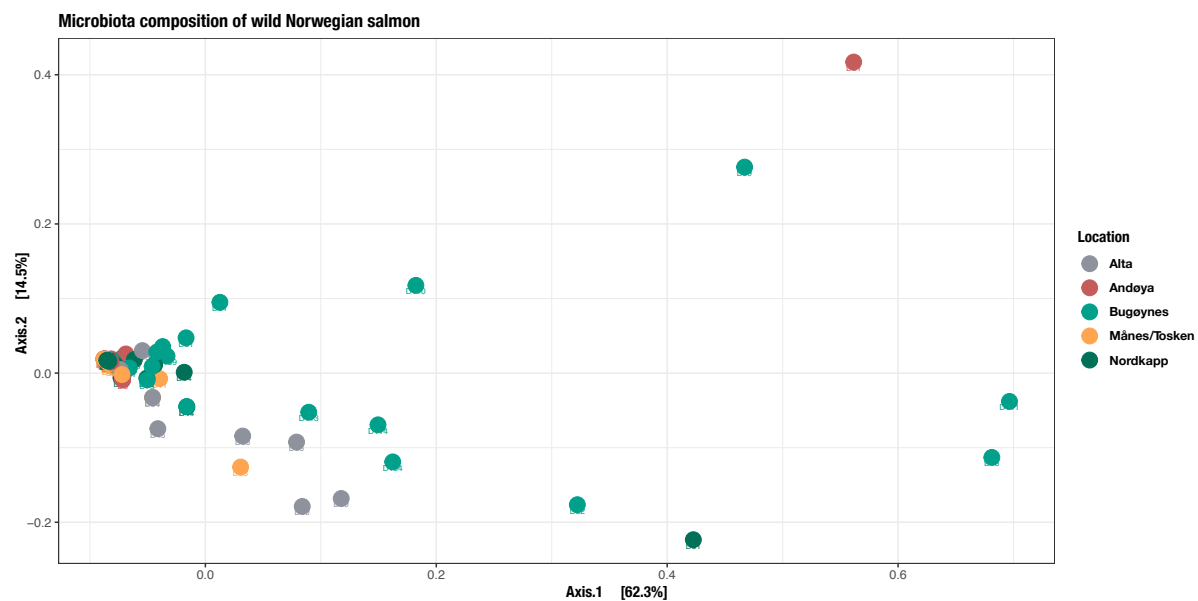
Figure S1) Overview of data quantities before and after eukaryotic mapping, using MGmapper. A) Box plots of mapped reads (percentage) to each database, including vertebrates (not mammals), human, vertebrates, invertebrate, and fungi. B) Bar plot illustrating the number of reads before mapping to the MGmapper database (notPhiX) and after extracting all database-related reads, indicating bacterial reads (Unmapped). The black dashed line indicates the mean of unmapped reads, whereas the grey dashed lines indicate the upper and lower 25% quantiles of unmapped reads.

18 Taxonomic inferring of microbiota in wild Atlantic salmon

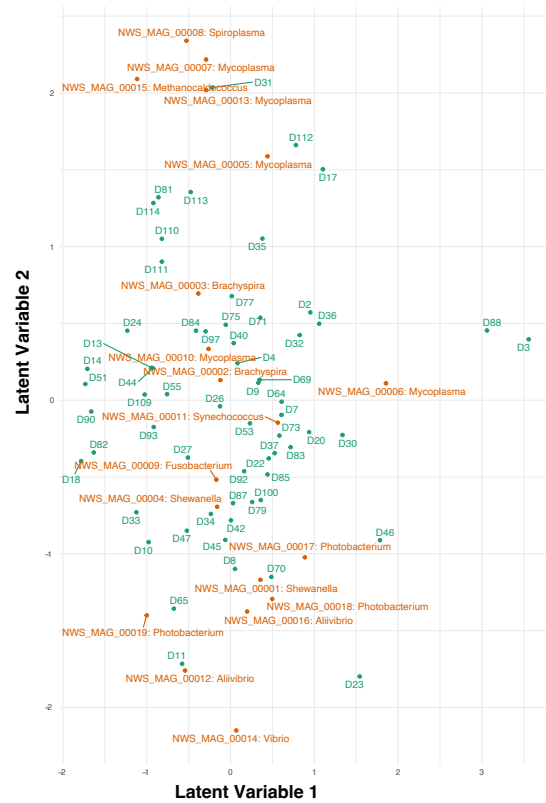


19
20 **Figure S2) Phylogenomic analysis of wild Atlantic salmon-related Metagenomic Assembled Genomes**
21 **(MAGs).** Unrooted maximum likelihood Phylogenomic tree of 3207 bacteria and the 19 MAGs from the study.
22 Multiple branches without MAGs were collapsed due to illustrative purposes. MAGs are highlighted with red
23 labels, and the closest known taxonomy of all MAGs is labelled with black.

27 Analysis of ecological dynamics and effect of environmental factors
28 on gut microbiota



A



B

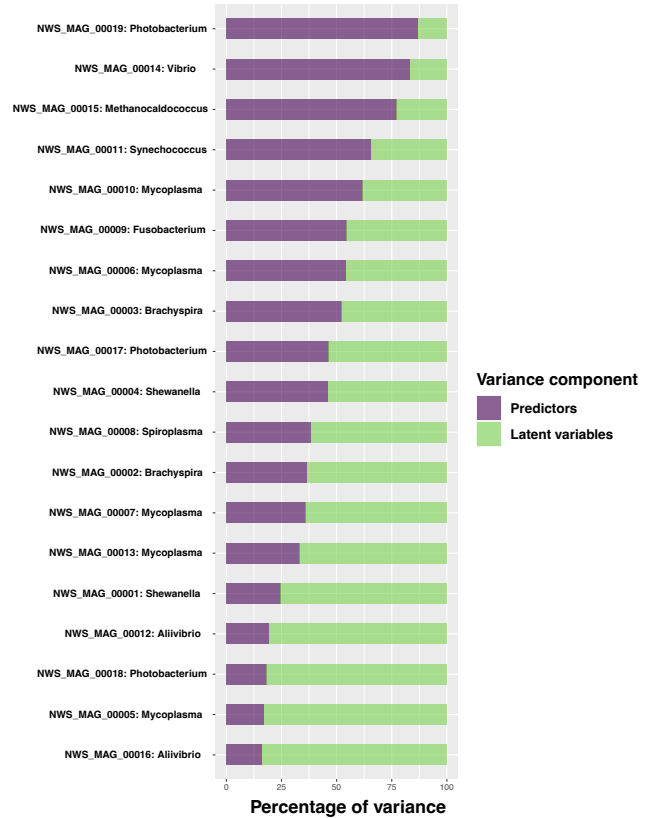


Figure S4) Bayesian constrained ordination of environmental factors related to MAGs. A) Constrained ordination of MAGs, where all environmental factors noted in our study are considered. MAGs are labelled orange, whereas samples are labelled grey. B) Stacked bar plots indicating the proportion of variance explained by known and latent predictors, as indicated in the legend.

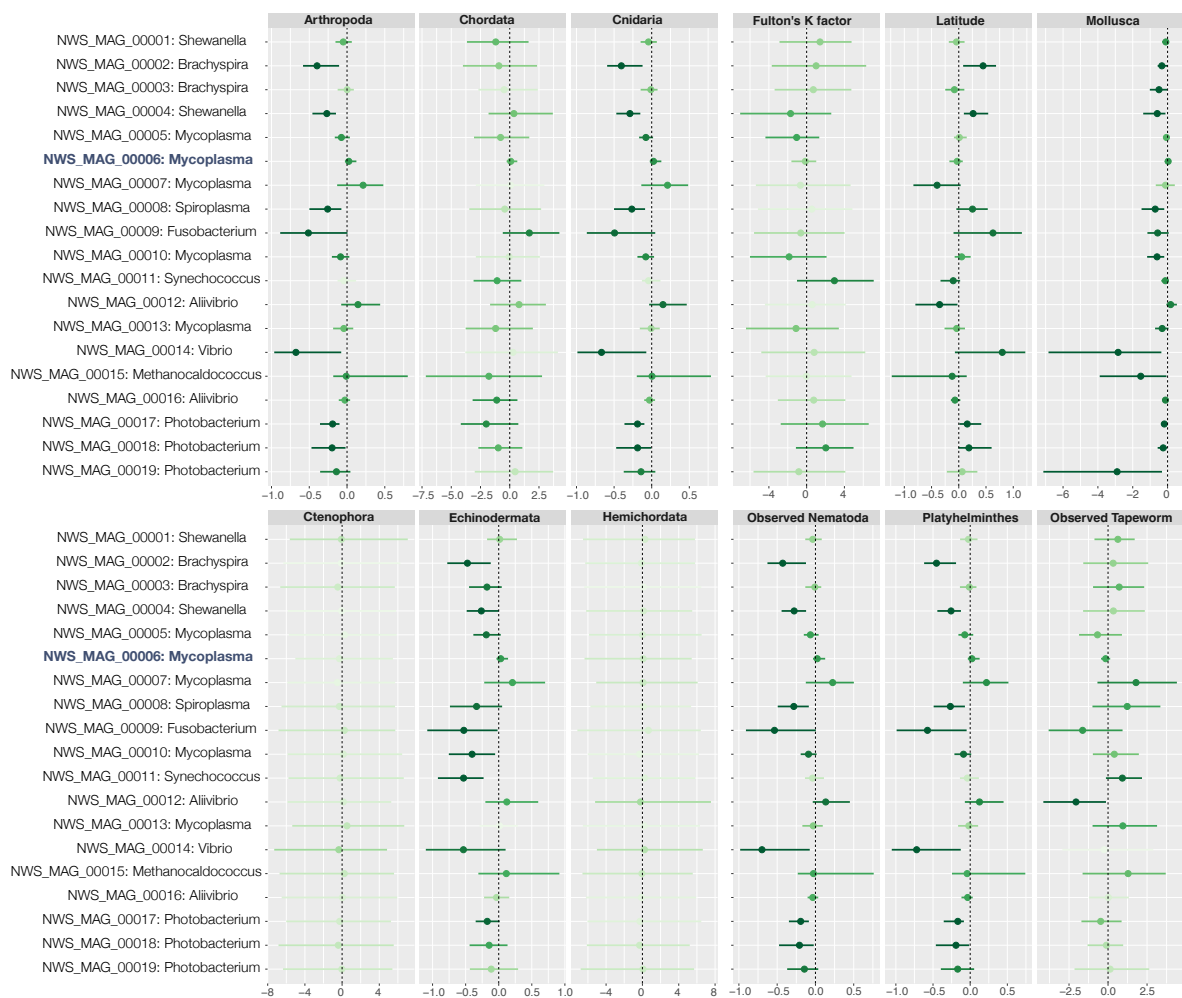


Figure S5) Regression analysis of environmental factors related to MAGs. Correlated response models consist of fitted generalised linear models. Coefficient plots are shown for individual MAGs (along the y-axis) in response to noticed environmental factors as different predictor variables, including putative diet (Arthropoda, Chordata, Mollusca, Echinodermata, Hemichordata), putative parasites (Cnidaria, Ctenophora, Platyhelminthes), observed parasites (nematodes and tapeworm), the latitude of sampling location (as an indicator for sampling location), fitness score of individual fish (Fulton's K factor). NWS_MAG_00006 (*Candidatus* Mycoplasma salmoninae salar) are highlighted in bold blue. Circles in the horizontal bars indicate the coefficient median; filled bars indicate the 95% confidence interval. The shade of green marks the relation between the coefficient, confidence intervals, and null, meaning the darkest green indicates significant correlations ($p < 0.05$).

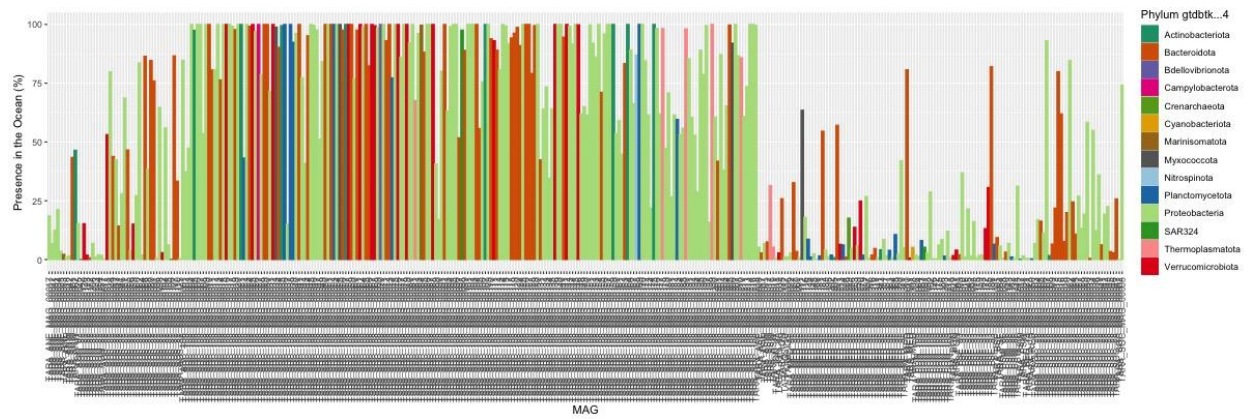


Figure S6) Bacterial presence in the Arctic ocean. Presence of 370 detectable MAGs in the Arctic ocean from the *TARA* Ocean project. The y-axis indicates the relative abundance of each MAG related to other oceans sampled by *TARA*. Each MAG is coloured based on Phylum.