

ARTICLE OPEN



Co-diversification of an intestinal *Mycoplasma* and its salmonid host

Jacob A. Rasmussen^{1,2}, Pia Kiilerich³, Abdullah S. Madhun⁴, Rune Waagbø^{4,7}, Erik-Jan R. Lock⁴, Lise Madsen⁴, M. Thomas P. Gilbert^{2,5}, Karsten Kristiansen^{1,6} and Morten T. Limborg²

© The Author(s) 2023

Understanding the evolutionary relationships between a host and its intestinal resident bacteria can transform how we understand adaptive phenotypic traits. The interplay between hosts and their resident bacteria inevitably affects the intestinal environment and, thereby, the living conditions of both the host and the microbiota. Thereby this co-existence likely influences the fitness of both bacteria and host. Whether this co-existence leads to evolutionary co-diversification in animals is largely unexplored, mainly due to the complexity of the environment and microbial communities and the often low host selection. We present the gut metagenome from wild Atlantic salmon (*Salmo salar*), a new wild organism model with an intestinal microbiota of low complexity and a well-described population structure, making it well-suited for investigating co-evolution. Our data reveal a strong host selection of a core gut microbiota dominated by a single *Mycoplasma* species. We found a clear co-diversification between the population structure of Atlantic salmon and nucleotide variability of the intestinal *Mycoplasma* populations conforming to expectations from co-evolution between host and resident bacteria. Our results show that the stable microbiota of Atlantic salmon has evolved with its salmonid host populations while potentially providing adaptive traits to the salmon host populations, including defence mechanisms, biosynthesis of essential amino acids, and metabolism of B vitamins. We highlight Atlantic salmon as a novel model for studying co-evolution between vertebrate hosts and their resident bacteria.

The ISME Journal (2023) 17:682–692; <https://doi.org/10.1038/s41396-023-01379-z>

INTRODUCTION

Co-evolutionary interactions between animals and their microbiota have become a key topic in evolutionary biology because of their adaptive relevance for understanding basic eco-evolutionary processes [1]. However, vertebrates and their associated microbial communities have yet to be systematically studied. One major drawback to understanding evolutionary host-microbiota relationships in many vertebrates, such as mammals, is that they are confounded by the high complexity of their microbial communities, challenging the detection of specific host-microbe interactions [2]. Dynamics in these complex systems are also often shaped by environmental factors limiting the ability to answer basic questions about broader evolutionary trends between host and resident bacteria, such as the simultaneous diversification of both host and microbe, which is often referred to as co-diversification [3].

One way to overcome these limitations is to investigate microbiotas of low diversity related to a host with a well-known population structure [3]. Teleosts offer an attractive option as more than 33,000 described species represent nearly half of all vertebrate species. Many teleosts have well-characterised population structures due to their commercial, cultural, and recreational importance [4]. However, despite representing the largest

vertebrate group, very few host–bacteria systems have been investigated in teleosts compared to terrestrial species, such as hibernating brown bears [5], plant-eating rodents [6], and production animals such as pigs [7] and ruminants [8, 9]. Due to its significant commercial, cultural, and recreational importance, the Atlantic salmon (*Salmo salar*) provides a well-studied model system. Atlantic salmon has a well-described genome, evolutionary history, and population structure throughout its North Atlantic distribution [10–12]. These assets make Atlantic salmon an ideal model for studying general evolutionary aspects of host-microbe interactions in vertebrates.

Adult salmon are piscivorous, and their associated microbiota is thus expected to be relatively simpler than microbiomes originating from fish being omnivorous and herbivorous [13]. Recent surveys of intestinal microbial communities in salmonid species have revealed a general dominance of a single unique *Mycoplasma* species suggesting a close relationship with the salmon host [14–16]. Salmonid-related *Mycoplasmas* are hypothesised to be vertically transmitted between generations since i) they have not yet been discovered in the environment [17], ii) are prominent in both wild and farmed salmon [15–18], and iii) do not follow neutral processes in both wild and farmed individuals [16].

¹Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Copenhagen, Denmark. ²Center for Evolutionary Hologenomics, GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ³Danish Center for Neonatal Screening, Department of Congenital Disorders, Statens Serum Institut, 2300 Copenhagen, Denmark. ⁴Institute of Marine Research, Bergen, Norway. ⁵Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. ⁶Institute of Metagenomics, Qingdao-Europe Advanced Institute for Life Sciences, Qingdao, China. ⁷Deceased: Rune Waagbø. ✉email: jacob.rasmussen@bio.ku.dk; morten.limborg@sund.ku.dk

Received: 14 July 2022 Revised: 27 January 2023 Accepted: 2 February 2023

Published online: 17 February 2023

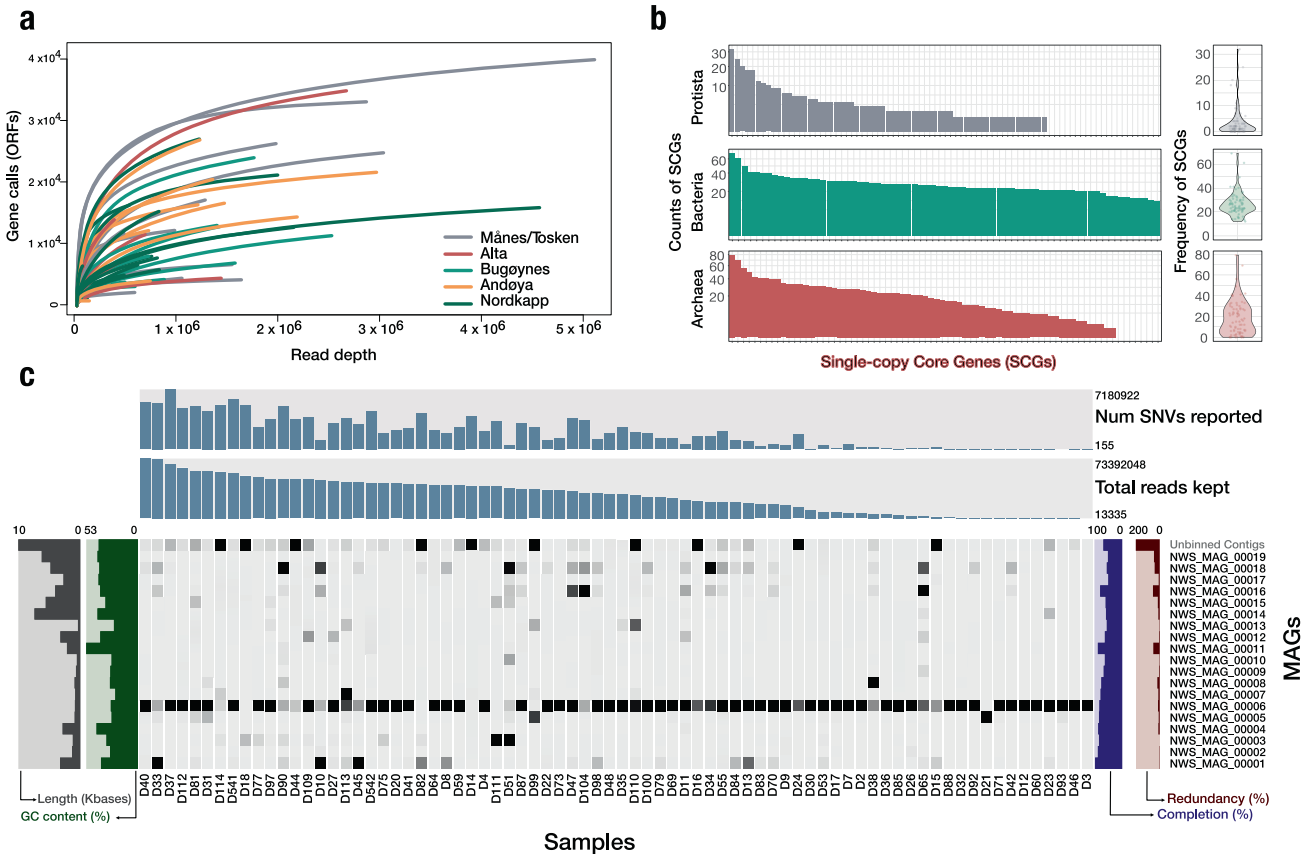


Fig. 1 Overview of the Atlantic salmon metagenome. **a** Rarefaction curves of gene calls across all samples from the co-assembled metagenome. **b** Single-copy core gene estimates present protists, bacteria, and archaea. **c** Coverage heatmap of recovered MAGs from the gut microbiota of Atlantic salmon. The top barplot reports the number of single nucleotide variants (SNVs) reported per sample, and the bar plot below reports read depth per sample. From the left, grey bars indicate the log₁₀ length of each MAG and unbinned contigs, green bars indicate GC content (%) per MAG, dark grey bars show the length of MAGs (kbases), and the heatmap of the MAGs is based on mean coverage per sample. White indicates low coverage, whereas black indicates high coverage. The heatmap is normalised according to the highest coverage of the most dominant MAG across each sample. Blue bars indicate completion (%) of each MAG, whereas burgundy indicates redundancy (%).

This consistent trend of intestinal microbial communities characterised by a low diversity makes Atlantic salmon a practical model for studying co-evolutionary relationships between vertebrate hosts and their core bacteria [19, 20].

Genome-resolved metagenomics offers a detailed resolution of microbial communities and microbial species. However, the often-reported low biomass of intestinal microbiota in salmonids has limited the broader application of shotgun metagenomic sequencing in these species [14, 15]. Therefore, previous studies have yet to use detailed genome-resolved metagenomics to uncover the functional and genomic legacy of the intestinal microbiome in wild populations of Atlantic salmon. Here, we use genome-resolved metagenomics and comparative genomics to investigate the metagenomic dynamics and hologenome of more than 70 adult foraging wild Atlantic salmon. We further compare the intestinal microbiota from Atlantic salmon with gut bacteria of other sympatric teleost species to elucidate environmental and host-related dynamics. Our study presents the first genome-resolved metagenome from wild Atlantic salmon, extending decades of PCR-based gene surveys, and highlights the value of applying a hologenomics approach [3] to study how host-microbiota relationships shape adaptive evolutionary processes.

RESULTS

Metagenomes were recovered from 75 wild Atlantic salmon individuals from five different regions in northern Norway, including

Alta, Andøya, Bugøynes, Månes/Torsken, and Nordkapp, spanning more than 700 km across the North Atlantic Ocean. Metagenomic sequencing of all individuals resulted in 5806 million reads. Re-mapping of the host genome revealed that 53.8% (SD ± 27.7%) of the reads mapped to the host genome, a percentage lower than in previous studies [14, 15]. Removal of low-quality and non-microbial data resulted in a recovery of 40.7% (SD ± 27.8%) of the reads (Suppl. Table S1, Suppl. Fig. S1). We identified 2,241,791 non-redundant genes in scaffolds longer than 1000 nucleotides. Investigation of sequencing depth and gene calls' rarefaction indicated sufficient metagenomic data saturation for a representative metagenomic analysis (Fig. 1a). Automatic binning and manual curation of co-assembly resulted in 19 non-redundant metagenome-assembled genomes (MAGs) containing 50,153,577 bases; see the summary of MAGs (Suppl. Table S2). The MAG catalogue represented 99.9% of the size-normalised bacterial data. Counting 71 bacterial single-copy core genes (SCGs) across the metagenome revealed that 23 bacterial genomes were potentially present in the metagenome, indicating that we recovered the majority of the MAGs (82.6 %) and that the metagenome of wild Atlantic salmon is characterised by a relatively low diversity (Fig. 1b, c).

Genome-resolved metagenomics reveals a low diversity gut microbiota of the Atlantic salmon, indicating strong host selection

Diversity analysis of MAGs revealed a gut microbiome of low diversity during the marine foraging life stage. Phylogenomics of

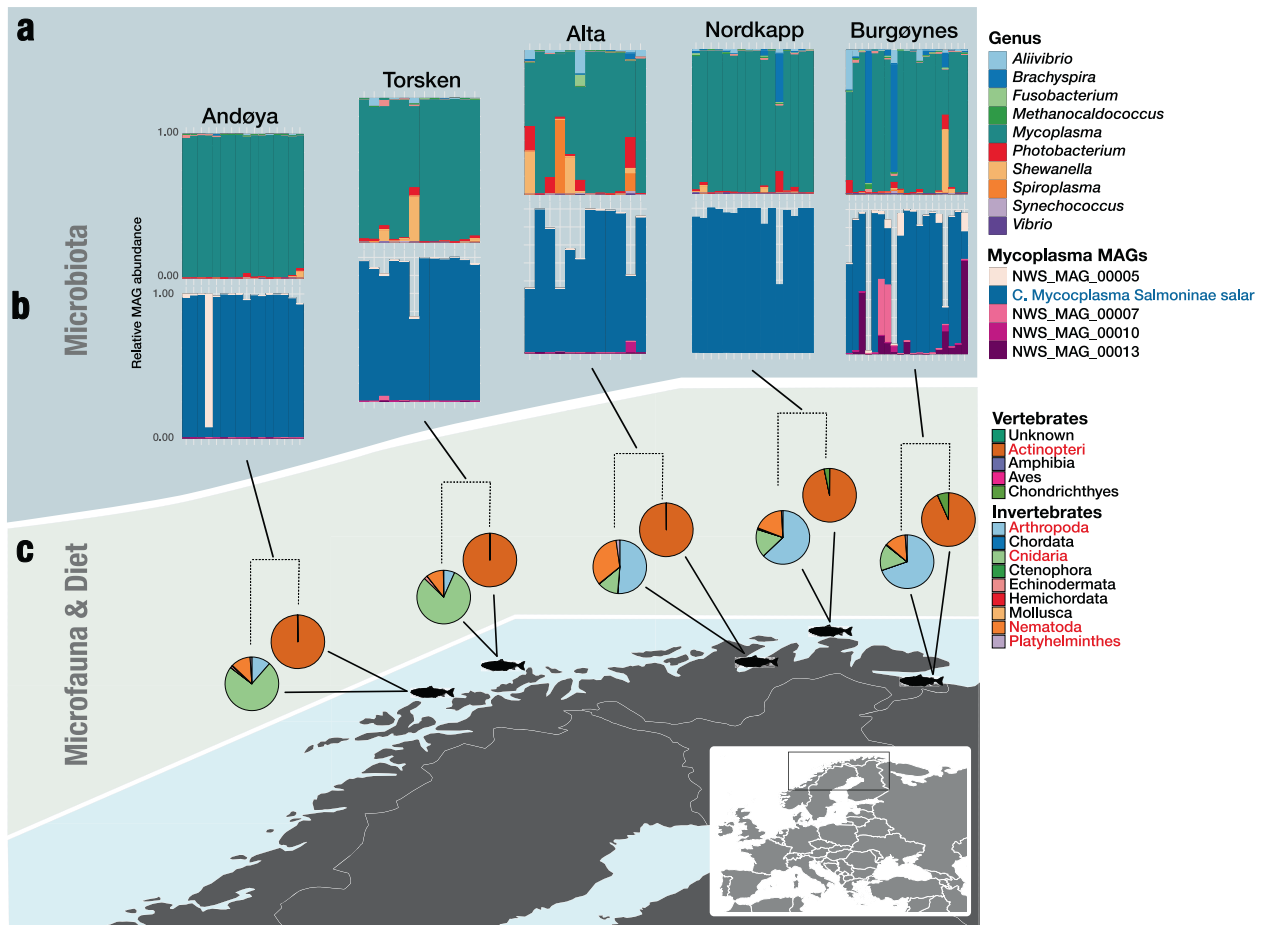


Fig. 2 Overview of gut microbiota and microfauna composition of wild Atlantic salmon. Atlantic salmon were sampled across five locations in northern Norway, including Andøye, Torsken, Alta, Nordkapp, and Bugøyenes, as indicated by black fish images. **a** illustrates the relative MAG abundance of all 19 MAGs, including all genera, and **b** focuses on the relative MAG abundance within the *Mycoplasma* genus. **c** Pie charts indicate microfauna and diet composition. The right pie chart indicates vertebrate composition, and the left indicates invertebrate composition.

recovered MAGs together with 3,207 bacterial reference genomes resulted in the taxonomic placement of MAGs, which included especially Tenericutes and Proteobacteria, but also Spirochaetae, Cyanobacteria, and Pseudomonadota (Suppl. Fig. S2). Our analysis revealed no significant differences in microbiota composition between sampling locations (Fig. 2a, b, Suppl. Fig. S3). *Mycoplasma* was highly dominant (90.4% of the MAG profiled reads), consistent with previous findings [14, 15, 17, 18, 21, 22]. We found an intermittent high abundance of *Photobacterium phosphoreum*, a species of *Brachyspira* (formerly classified as *Brevinema andersonii*), and *Shewanella* MAGs across eight samples (Fig. 2a). Subsequently, we recovered low amounts of *Synechococcus*, *Vibrio*, *Aliivibrio salmonicida*, and *Methanocaldococcus* across the 75 Atlantic salmon.

A large proportion of the data represented *Mycoplasma*, which dominated overall and represented over 90.4% of all host-filtered reads. Of the 19 MAGs recovered, five MAGs were within the family of Mycoplasmataceae (Fig. 2b). Subsequently, one MAG of *Mycoplasma*, *Candidatus Mycoplasma salmoninae salar* (MSS), alone accounted for 83.4% of the host filtered reads. Our analysis revealed a stable MSS dominance across recorded environmental factors, suggesting an ecological adaptation of high-abundance *Mycoplasma* (Suppl. Figs. S4, S5). The richness of *Mycoplasma* MAGs followed a latitudinal pattern showing a mixture between *Mycoplasma* MAGs (NWS_MAG_00006 and NWS_MAG_00013), where NWS_MAG_00013 became more abundant in northern regions, indicating that a higher richness of *Mycoplasma* species could originate from northern Norway or Russia and that more

Mycoplasma species are to be discovered. We further analysed TARA Oceans data to investigate the putative origin of *Mycoplasma*. Still, no *Mycoplasma* MAGs were found across the Arctic Ocean, suggesting that MSS was unlikely to be obtained from the ocean during the adult life stages of Atlantic salmon.

Reference-based mapping and taxonomy annotation of metagenomics sequence data revealed eukaryotic gut content consisting of diet or gut microfauna (e.g., tapeworms, nematodes, and Myxozoa). Vertebrate diet content in the gut included Atlantic herring (*Clupea harengus*) (Fig. 2c). Gut content of invertebrates was comprised of Arthropoda, Cnidaria, Nematoda, and Platyhelminthes, where arthropods were thought to originate from the diet like krill. Cnidaria, Nematoda, and Platyhelminthes were considered parasitic microfauna, including tapeworms (*Eubothrium*), anisakid nematodes (*Anisakis simplex*), and salmon-related Myxozoa (Fig. 2c). The composition of invertebrates differed according to locations, where cnidarian content dominated the gut content in the south and Platyhelminthes was the dominating phylum in the gut content of northern individuals (Fig. 2c). Despite the compositional differentiation of microfauna and diet, no effect on high-abundance *Mycoplasma* (NWS_MAG_00006) was detected.

Meta-pangenomics reveals high host specificity of *Mycoplasma*

Our analysis revealed that the high-abundance *Mycoplasma* MAG was closely related to previously recovered salmonid-related *Mycoplasma* MAGs from aquaculture strains (Fig. 3a).

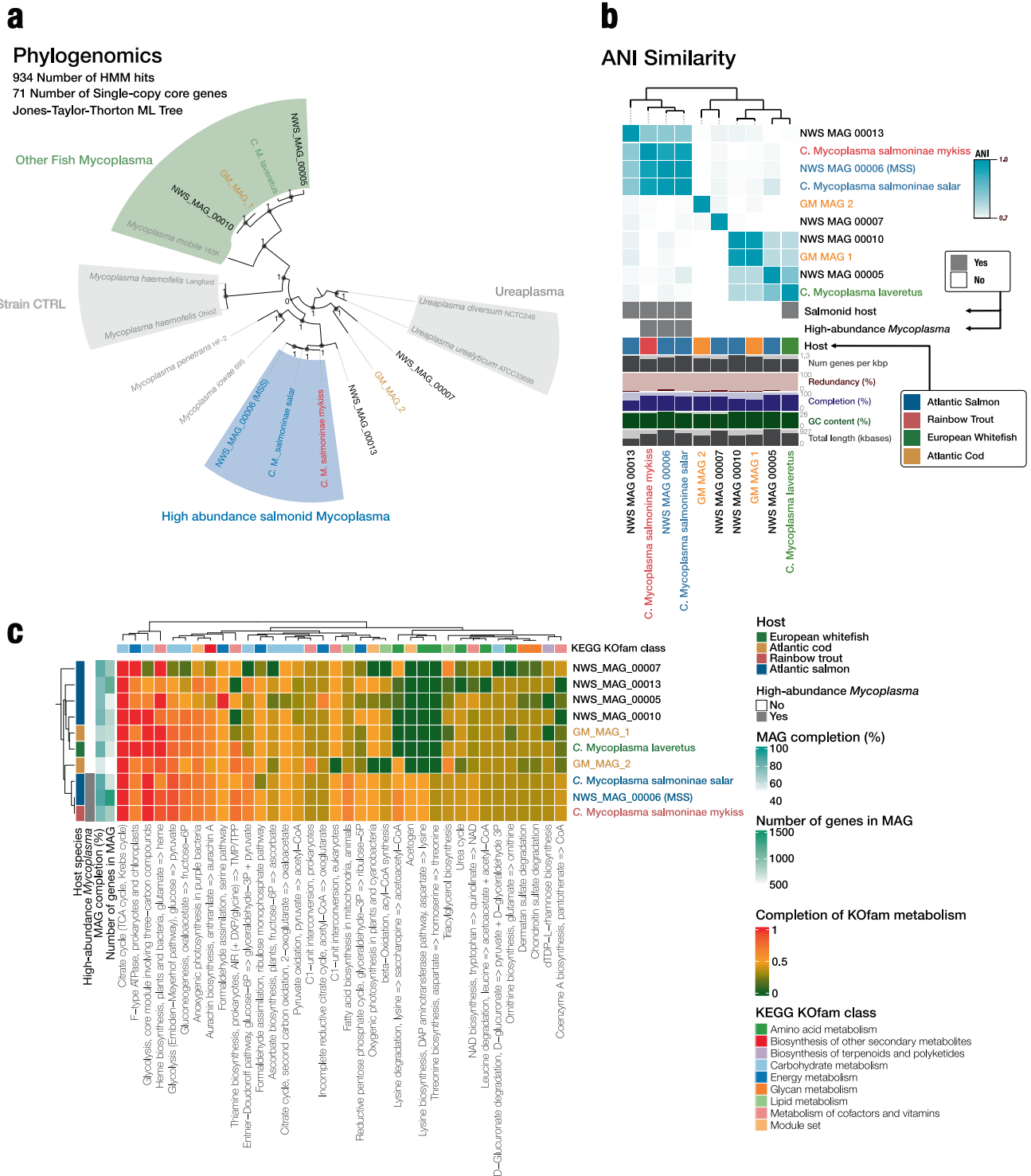


Fig. 3 Phylogenomics and comparative genomics of co-assembled MAGs from wild Atlantic salmon and closely related *Mycoplasma* species and candidate genomes. **a** unrooted phylogenomic maximum likelihood tree constructed based on 934 HMM hits from 71 single-copy core genes across 17 genomes and MAGs. **b** Heatmap of Average Nucleotide Identity (ANI) similarity among salmonid-related *Mycoplasma* MAGs from this study and farmed salmonids [15], and *Mycoplasma* MAGs related to Atlantic cod from the same environment [75]. **c** Heatmap of the metabolic reconstruction and comparison of salmonid-related *Mycoplasma* versus other fish-related *Mycoplasma*, based on completion of KEGG pathways.

Our phylogenomic and comparative analysis revealed that NWS_MAG_00006 and *Candidatus* *Mycoplasma salmoninae* salar (MSS) were more closely related to each other than the *Mycoplasma*, *Candidatus* *Mycoplasma salmoninae* mykiss (MSM), associated with farmed rainbow trout (*Oncorhynchus*

mykiss), which underpins the specificity of these salmonid related *Mycoplasma* MAGs to their salmonid host regardless of a farmed or wild origin (Fig. 3a, b). The four other *Mycoplasma*-related MAGs were all found to have low abundance across all individuals and clustered with various salmonid-related clades

like NWS_MAG_00013 as a sister group to the clade of Atlantic salmon, rainbow trout, and NWS_MAG_00006. We found two new MAGs related to *Candidatus Mycoplasma lavaretus* (ML) and *Mycoplasma mobile* 163 K. Further, we found one MAG (NWS_MAG_00007) associated with *Ureaplasma*. Analysis of the gene clusters among all *Mycoplasma* MAGs further confirmed the differentiation between MAGs from rainbow trout and Atlantic salmon compared to the other MAGs (Fig. 3a, b). These findings reveal a clade of abundant *Mycoplasma* highly prevalent in salmonids and living in coexistence with other low abundant *Mycoplasma* species.

We found that all high-abundance MAGs in the salmonid clade were not only phylogenetically separated from the other *Mycoplasma* MAGs but also exhibited distinct metabolic traits (Fig. 3c). We reconstructed the metabolic pathways across the fish-related *Mycoplasma* MAGs to decipher the potential functional differences between high and low abundant *Mycoplasma* MAGs for comparison with our previously described phylogenetic differentiation. We found KOfams in the high-abundance clade, which could be potentially beneficial for its host, including genes encoding enzymes involved in thiamine (B1 vitamin) biosynthesis (Fig. 3c). Furthermore, genes involved in the metabolism of nicotinamide/niacin (B3 vitamin) and pantothenate (B5 vitamin) were also found in the high-abundance *Mycoplasma* MAGs (Fig. 3c, Fig. Suppl. S6). Synthesis of two essential amino acids for Atlantic salmon (lysine and threonine), suggesting a potential beneficial role of MSS to its salmon host. Furthermore, we found that the high-abundance clade could function by an acetogenic lifestyle, a function lacking in the other fish-related *Mycoplasma* MAGs. Together, these observations follow a mutualistic relationship where MSS provides this essential feature of acetate fermentation in the predominantly anoxic gut ecosystem of salmonid hosts (Fig. 3c). However, it remains to be shown whether this directly affects the fitness of the salmon host.

Mirrored patterns of population structures show co-diversification of Atlantic salmon and *Candidatus Mycoplasma salmoninae salar*

To study the evolutionary relationship of wild Atlantic salmon across our sampled locations, we compared our data with Atlantic salmon genotypes of already known origins across north- and southern Norway from publicly available genomes [12]. Principal component analysis (PCA) of host genotype likelihoods revealed three clusters explained by PC1 and PC2 (Fig. Suppl. S7). These clusters were mainly defined by latitude consistent with previous work [11, 12]. Our PCA resulted in three clusters: a cluster (Cluster 1), where the investigated individuals and known individuals were from northern Norway. A cluster (Cluster 2) where the investigated individuals together with known individuals were from southern Norway. Lastly, a cluster with known individuals from the Baltic population. Of the individuals in our data, we detected 24 (32%) individuals from Norwegian populations clustered by latitude. Our analysis revealed that far from all individuals were closely related to the Norwegian genotypes, indicating that several foraging individuals were from genotypes originally reported outside of Norway.

Based on their genotypes, these Norwegian lineages were used to infer the possible co-phylogeny between MSS and Atlantic salmon. We chose only to use the Norwegian individuals since the population structure was better described for these individuals. Of the 24 Norwegian individuals, we recovered single nucleotide variants (SNVs) from the associated MSS genomes of 19 individuals (Fig. 4). Analysing SNVs in MSS revealed a total of 136 SNVs present in 90% of the 19 individual metagenomes. Analysis of the Norwegian salmon genotype and MSS showed co-clustering between host populations of Atlantic salmon and the variability of single nucleotide variants (SNVs) in MSS, further supporting co-diversification between MSS and Atlantic salmon.

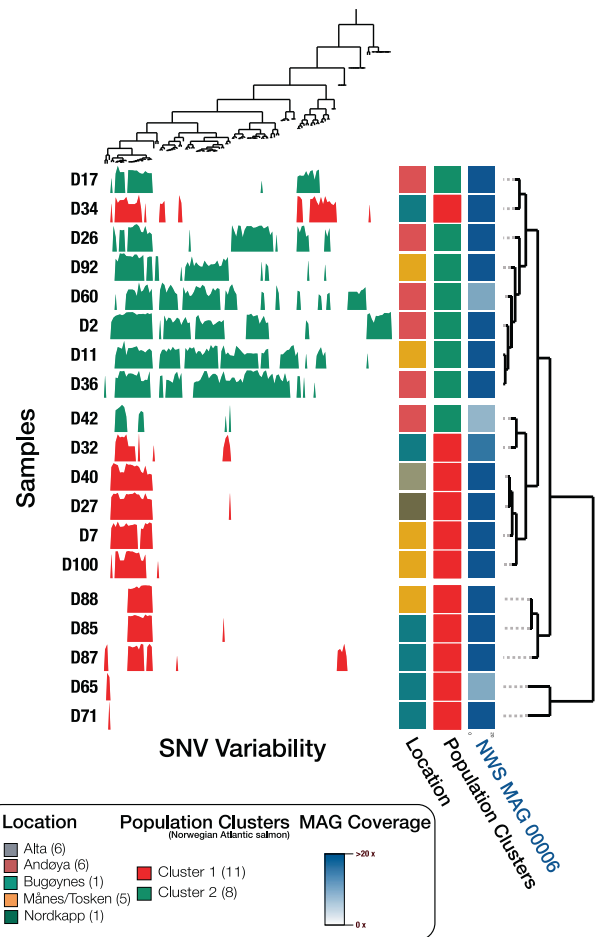


Fig. 4 Analysis of host genotypes in Norwegian Atlantic salmon (Host origin) and SNV variability from high-abundance *Mycoplasma* (*Candidatus Mycoplasma salmoninae salar*). Variability of Single Nucleotide Variants (SNVs) in the high-abundance *Mycoplasma* ordered along the x-axis. Only SNVs with a minimum of 20X coverage in at least 90% of all individual metagenomes were considered. Each layer represents an individual MSS genotype recovered from a host metagenome. Layers have been hierarchically clustered (dendrogram) based on the similarity of SNV variability (departure from consensus) in MSS between samples using euclidean distance and ward ordination. For each layer, bars for each SNV illustrate variability from zero to one between competing nucleotides (NTs). The SNV variability across samples has been hierarchical clustered (dendrogram) on the x-axis, using euclidean distance and ward ordination. The right bar plots represent the host genotype clusters, location, and coverage of MSS. The location information is coloured as indicated by the legend.

The clustering of MSS was primarily driven by the presence of high SNV variability in Cluster 2, which was absent in Cluster 1, indicating that either Cluster 1 lost its variants or Cluster 2 acquired new variants (Fig. 4). Overall, the mirrored population structure between MSS and Norwegian Atlantic salmon and the constant domination of MSS across any measured environmental factor suggest a high selection pressure on the microbiota. We additionally evaluated patterns of co-diversification with the Parafit analysis, a permutational method that assesses the similarity of principal coordinates derived from the host and the phylogeny of a specific symbiont. The global Parafit test was significant (Parafit, $p < 0.001$), confirming our findings of co-diversification between Atlantic salmon and MSS.

We looked for evidence that the salmonid MSS population had been subjected to high selection pressure by calculating the

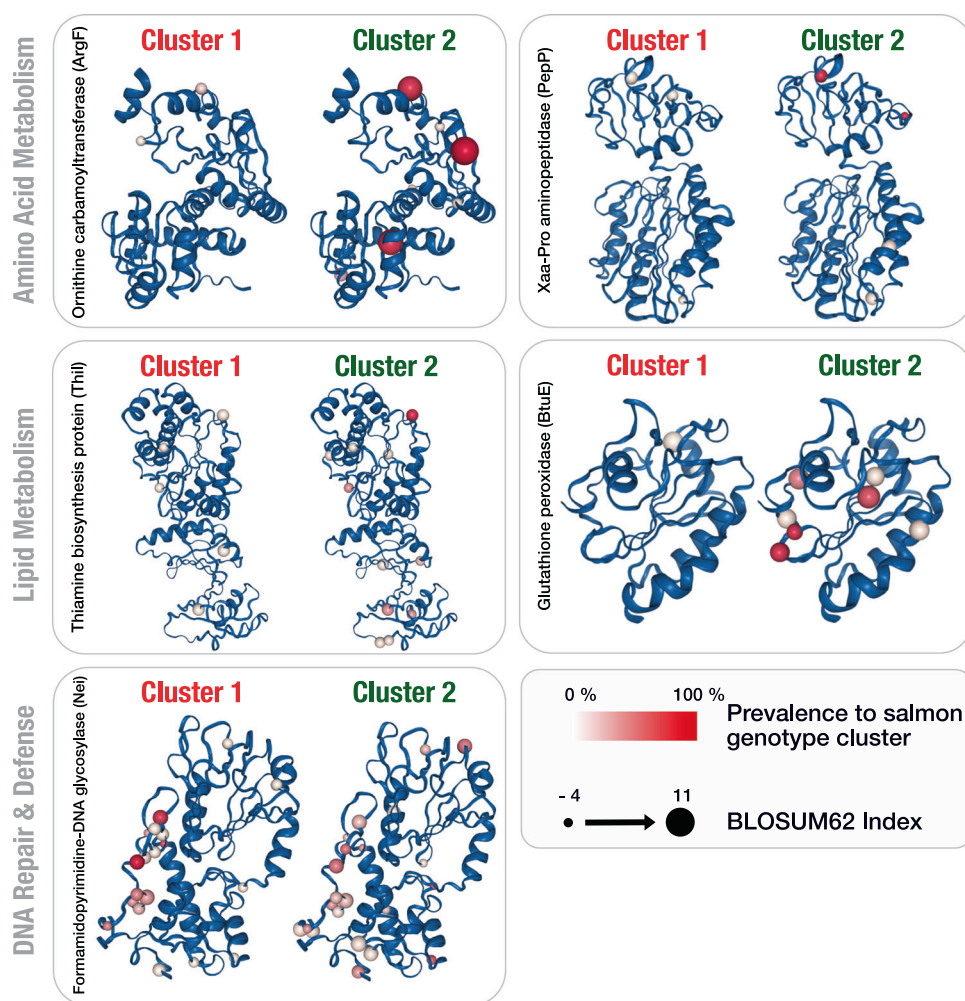


Fig. 5 Examples of predicted protein structure variants in *Candidatus Mycoplasma salmoninae salar* related to the host genotype. Single amino acid variants (SAAVs) on the predicted protein structures of five example genes from MSS across 19 metagenomes from the two Atlantic salmon genotypic clusters. The sphere in each predicted protein structure illustrates a SAAV, where colour indicates prevalence concerning the specific host genotype and size of the sphere indicates the interchangeability of two amino acids using BLOcks SUBstitution Matrix (BLOSUM62 index) as noted in the legend.

average polymorphism rates of non-synonymous to synonymous mutations (pN/pS) for each gene within the MSS population. The genes with pN/pS values higher than one suggested that MSS individuals had been subjected to positive selection pressure (Suppl. Fig. S8). Most genes with a high pN/pS ratio were unknown. Still, this group also included genes involved in defence mechanisms, translation, ribosomal structure and biogenesis that may have been involved in the adaptation of MSS to the intestinal environment of its salmon host.

We also inferred single amino acid variants (SAAVs) to characterise non-synonymous mutations between MSS variants and thereby investigate the putative phenotypic variation in MSS prevalent to its host genotype. Before exploring SAAVs, we predicted protein structures from 141 genes containing SNVs resulting in 51 (36.2%) predicted protein structures with 1813 SAAVs across all individuals. Of the 141 genes containing SNVs, we recovered 51 predicted proteins from the MSS genome. Several of the predicted proteins included SAAVs that were highly prevalent in the related genotype cluster of host salmon (Fig. 5). Three predicted proteins were outstanding in prevalence and relation to Atlantic salmon and the intestinal environment. These proteins included i) ornithine carbamoyl transferase, which is central to the urea cycle and arginine biosynthesis, ii) glutathione peroxidase, involved in peroxide-related defence mechanism in intestinal

environments in response to lipid peroxidation [23], and iii) thiamine biosynthesis protein, related to cofactor functions important for lipid metabolism. The SAAVs in these MSS encoded proteins showed a high association with the host genotypic pattern, and our results enable us to predict the genotype of Atlantic salmon by the predicted protein structure variation within MSS conforming with co-diversification of not only synonymous mutations but also non-synonymous mutations between Atlantic salmon and the dominating *Mycoplasma*.

DISCUSSION

Previous investigations of associations between vertebrate hosts and their microbiota in wild-living species, including teleosts, have mainly been covered by amplicon-based approaches with a low taxonomic resolution and limited functional inference [21, 22, 24]. Moving beyond gene amplicon surveys, we provide the first metagenomic description related to the gastrointestinal environment of multiple wild Atlantic salmon ranging in size, observed stomach content, sequenced diet, genotype, and location. We acknowledge that MAG analysis has limitations, including low sensitivity of low abundant organisms and variable reliability of MAG generation. To increase the reliability of our MAGs, we applied binning methods, combining automatic binning and

manual curation with *anvi'o*, and used good practice for the reliable generation of MAGs [25]. To ensure suitable sequencing depth, we estimated the saturation of ORFs across our samples and applied co-assembly to increase the microbial data for assemblies. Moving beyond gene amplicon surveys of wild Atlantic salmon gut microbiota, we provide a metagenomic description, giving genomic information for lineages missing a functional context and allowing us to search for strain variation within a large pool of microbial populations associated with salmonid hosts. Our results indicate that investigations of salmonid *Mycoplasma* are more complex than previously anticipated and that 16S rRNA gene investigations at the genus level will often simplify the actual strain-level variation in the microbiota.

Our data confirm a low-diverse and *Mycoplasma*-dominated gut microbiota in the Atlantic salmon [13–15]. Furthermore, our analysis also found abundant members of Proteobacteria, including *Allivibrio*, *Photobacterium*, and *Vibrio*, previously known as part of the gut microbiota in salmonids [26–28]. The microbial variation was low across all 75 investigated individuals despite being exposed to different conditions for multiple environmental factors, indicating high host selection pressure favouring a high relative abundance of *Mycoplasma*. Therefore, we hypothesise that this *Mycoplasma*-dominated microbiota is characteristic of natural and undisturbed Atlantic salmon. Thus, our data should serve as a valuable resource for future microbiota studies in salmonids, including aquaculture applications where gut microbiota investigations have been a research field of increasing interest [26–32].

While numerous cases of co-adaptation have been documented in mammals [33–36], other warm-blooded organisms such as vultures [37], and invertebrates [38, 39], very little is known about the adaptive potential of the associated microbiota in teleosts representing more than half of all vertebrate species. Despite the detailed environmental and life-history metadata for each sampled individual in our study, the environment explained no concluding variation in the abundance of MSS. We found that the SNVs variation of MSS was better explained by the genetic background of the host individual rather than the sampling location, indicating an intriguing host-bacteria co-diversification.

Groussin et al. recently formulated a guideline model to separate co-phylogeny from co-evolution by rejecting the geographic isolation of two host populations [2]. While demonstrations of co-evolution in mammalian and bacterial genomes appear to be demanding [2, 35], our study illustrates that salmon can serve as a valuable future model for expanding our understanding of co-evolution between vertebrate host species and associated bacteria. Indeed, according to Groussin et al., three independent prerequisites are needed for co-evolution to occur with high stability, including i) strong fitness dependencies, ii) stable transmission across generations and iii) strong host selectivity [2]. First, for mammals, they argue that it is non-trivial to convincingly show a direct fitness of symbiotic microbe on its host as the same function can easily be provided by other related species. However, we argue that this is less likely in salmonids due to the low diversity of the microbiota and the low biomass observed. Second, the phylogenomic observation that *Mycoplasma* strains cluster according to both host species, despite being farmed or wild, and population clusters strongly suggest vertical transmission across generations. However, the exact mechanism remains to be elucidated. Though evidence for vertical transmission has not yet been established, other observations besides our study indicate such transmissions, such as the lack of *Mycoplasma* in the surrounding rivers [17] or the Arctic oceans and that *Mycoplasma* species are highly host-dependent due to their intracellular lifestyle. Furthermore, it should be noted that examples of co-evolution do exist based on horizontal transfer, like the bobtail squid-*Vibrio* symbiosis [40] and bean bug-*Burkholderia* symbiosis [41]. These examples indicate that the

evolution of mutualism and the emergence of co-phylogenetic patterns do not strictly rely on strict vertical inheritance [42], but rather a strict host control [43]. Third, our results are in line with previous observations that independently showed positive correlations between *Mycoplasma* abundance and fitness of the salmonid host following strong host selectivity on *Mycoplasma* abundance [26, 44–47]. This is following recent findings showing that *Mycoplasma* abundance does not follow a neutral model in Atlantic salmon, supporting a non-neutral explanation for the observed co-diversification [16]. Our SNV analysis revealed that Cluster 2 of MSS had acquired a higher amount of variable SNV compared to Cluster 1, which is driving the mirrored separation between MSS and Atlantic salmon. Some of these SNVs might come from ecological processes. Overall, our results follow established expectations for co-evolution, presenting salmon as a new model for understanding holobiont evolution [2, 3]. However, our study could not show evidence for reciprocal selection and resulting change. Therefore, we cannot rule out that the co-diversification pattern of MSS and Atlantic salmon could originate from more neutral processes, including ecological processes and genetic drift within the populations of MSS and Atlantic salmon.

Our data provide additional insights into the dominance of *Mycoplasma* and the overall gut microbiota of Atlantic salmon by revealing that several novel, lowly abundant, and dynamic MAGs of *Mycoplasma* co-exist with high-abundance *Mycoplasma*. Furthermore, our data show a clear phylogenomic and functional differentiation between *Mycoplasma* clades of high and low abundance, which will be difficult to decipher using an amplicon-based approach. Specifically, we found a higher completion of pathways associated with the biosynthesis of essential amino acids (lysine and threonine) and biosynthesis of thiamine (B1 vitamin). Wild Atlantic salmon may be deficient in B1 vitamins, suggesting that MSS could be a beneficial source of B1 vitamins in Atlantic salmon [48]. A recent study revealed that the Baltic population of Atlantic salmon lacked the presence of *Mycoplasma*, a population also shown to suffer from B1 vitamin deficiency during early life stages [49] and thereby thiamine deficiency-related reproduction [50–52].

Using untargeted metabolomics, we previously found a higher amount of B vitamins in rainbow trout associated with an increased presence of *C. Mycoplasma salmoninae mykiss*, further supporting some fitness dependency of *Mycoplasma* to its salmonid host [53]. Furthermore, we analysed SAAVs with a high prevalence to host genotype, indicating an apparent fixation of mutations within the MSS populations related to their host. Specifically, we did find SAAVs coupled to thiamine biosynthesis protein and glutathione peroxidase, which were prevalent in the same host population cluster. Both proteins are related to lipid peroxidase, indicating that MSS might be important for lipid metabolism in salmonids or that the diet is important for Atlantic salmon and *Mycoplasma* adaptations. Both scenarios could align with a previous report that revealed alterations in lipid efficiency when *Mycoplasma* was missing in the gut microbiota [53]. In addition, we found SAAVs related to the host in the ornithine carbamoyl transferase protein structure. This protein utilises ammonia and has previously been associated with *Mycoplasma*, indicating a putative adaptation in the intestinal environment [15, 53].

Understanding the functional interdependence between hosts and their microbiota by studying holobionts represents a rewarding field in evolutionary biology [3, 54, 55]. The case of Atlantic salmon studied here has not only furthered our evolutionary understanding of this species, but the findings also hold potential for further discoveries towards feed or health optimisation resulting in more sustainable aquaculture practices. We envisage that our study may inspire similar investigations in systems previously investigated using amplicon-based markers to reveal the intriguing functional host-microbiota interactions.

METHODS

Ethical approval

Atlantic salmon included in this study were sacrificed immediately upon the catch resulting in instant death before tissue and gut samples were taken. While no separate licence is required for such sampling, we emphasise that all fish handling was supervised by experienced and trained staff following standard and legal procedures in Norway.

Sample collection

Sample collections for Atlantic salmon were taken from five locations across the coast of northern Norway, including Månes/Torsken, Nordkapp, Alta, Bugøyenes, and Andøya. Commercial fishermen caught the Atlantic salmon with nets or traps in coastal waters, as previously described [56–58]. We froze individuals whole immediately after the catch. Sample collections of gut content and gut scrapings were carried out for midgut and hindgut at the National Institute of Nutrition and Seafood Research (NIFES). Atlantic salmon were thawed overnight. Gut sections were divided, and content was squeezed out of the gut using a sterile scalpel. All samples were snap-frozen with dry ice immediately after sample collection. While sample collection was carried out, stomach content, fish weight, length, nematodes, and tapeworm were noticed for subsequent analysis.

DNA sequencing

DNA was extracted using ZymoBionics DNA miniPrep (Zymo Research) following the manufacturer's protocol. Initial quality control of samples was performed using a Qubit 3.0 fluorometer following the manufacturer's protocol. The DNA was shipped to Novogene (Cambridge, UK) post-extraction for DNA fragmentation by sonication. Library preparation was carried out using Novogene NGS DNA Library Prep Set (Cat No.PT004). Further quality control included qPCR for quantification. Size distribution was detected using Bioanalyzer 2100 (Agilent), and quantified libraries were pooled and sequenced on a NovaSeq 6000 (Illumina) with a 150 bp paired-end strategy.

Genome-resolved metagenomics

Raw sequence reads were quality controlled, using FastQC/v0.11.8 to assess filtering and quality steps. Adapters and low-quality reads were removed with AdapterRemoval/v2.2.4, with a quality base of 30 and a minimum length of 50 bases. We removed duplicates, and reads were repaired to remove singletons using bbmap/v.38.35. We filtered data for the host genome using bwa mem [59] to increase assembly efficiency by reducing eukaryotic contaminants. Subsequently, we applied a reference-based mapping and taxonomy annotation to i) remove unknown eukaryotic contaminants and ii) analyse eukaryotic gut content (diet and microfauna) using MGmapper [60]. Filtered data were co-assembled using MEGAHIT with metagenomic sensitive pre-sets [61]. Assembled contigs were quality assessed with Quast/v.5.0.2. Filtering for a minimal length of 1000 bases per scaffold was applied. To increase effective metagenomics binning, we used the anvi'o pipeline [62, 63] with automatic binning, using CONCOCT, followed by manual curation with the anvi'o platform. Briefly, i) anvi'o was used to identify genes in the scaffolds using Prodigal/v2.6.3 [64] with default parameters. Subsequently, HMMER/v.3.3 [65] were used to identify genes matching archaeal, protists, and bacterial [66] single-copy core gene (SCGs) collections with hidden Markov models (HMMs). Also, ribosomal RNA-based HMMs were identified using available scripts from <https://github.com/tseemann/barnap>. The HMMs of SCGs were used to determine the completeness and redundancy of metagenome-assembled genomes (MAGs); ii) read recruitment of the metagenome to the scaffolds was carried out using BWA/v0.7.1596 (minimum identity of 95%) and samtools [67]. We binned contigs automatically using CONCOCT [68]. Each CONCOCT bin was manually curated using the anvi'o interactive interface to ensure high completion and low redundancy. The interface considers each scaffold's sequence composition, differential coverage, GC content, and taxonomic signal [62, 69]. We defined all bins with >50 % completeness as MAGs. Calculating the potential number of genomes is based on counts of SCGs, following the tutorial outlined at <https://merenlab.org/2015/12/07/predicting-number-of-genomes/>.

Identification, refinement, taxonomic, and functional inference of MAGs

We used anvi'o to infer the taxonomy of MAGs based on the proximity of single-copy gene markers based on the Genome Taxonomy Database (GTDB) [70]. Subsequently, we applied Kaiju [71] with NCBI's non-

redundant protein database 'nr' to infer the taxonomy of genes (as described in <http://merenlab.org/2016/06/18/importing-taxonomy/>). For functional inference, we used clusters of orthologous (COGs) [72], Kyoto encyclopedia of genes and genomes (KEGG) [73], and protein families (Pfam) [74], which were annotated through the anvi'o platform. A summary of the MAGs generated for this study is available at <https://doi.org/10.6084/m9.figshare.20043452>.

Phylogenomic and comparative analysis of MAGs

Mycoplasma genomes from different species were selected based on relatedness to MAGs from this study. Selected genomes are referred to as external genomes. These external genomes were compared with wild salmonid MAGs using anvi'o/v7.1 for phylogenomic and comparative analysis. The phylogenomic analysis was carried out for all MAGs using an in-house database of 3207 bacterial genomes. Subsequently, we carried out a phylogenomic analysis for *Mycoplasma*, explicitly using the recovered MAGs from this study and MAGs recovered from other studies of salmonids [74] and Atlantic cod (*Gadus morhua*) [75].

Furthermore, outgroups of *Mycoplasma Haemofelis* Langford, *Mycoplasma Haemofelis* Ohio, *Ureaplasma diversum*, and *Ureaplasma urealyticum* ATCC33699 were included. The phylogenomic analysis was carried out based on bacterial SCGs using an anvi'o bacterial database. Amino acid sequences were extracted from HMMs of SCGs and concatenated into aligned amino acid sequences. Concatenated amino acid sequences were used to generate a Newick-based maximum-likelihood phylogeny using FastTree2 [76]. The comparative analysis was carried out as in the previous studies [76], where similarities of each amino acid sequence in every genome were calculated against every other amino acid sequence across all genomes using BLASTp. We implemented Minbit heuristics of 0.5 to eliminate weak matches between two amino acid sequences [77] and an MCL inflation of 2. We used the MCL algorithms to identify gene clusters in amino acid sequence similarity [78]. We calculated ANI using PyANI [79]. Euclidean distance and ward linkage were used to organise gene clusters and genomes. A summary of the pan-genome generated for this study is available at <https://doi.org/10.6084/m9.figshare.21183415>.

Metabolic reconstruction of compared MAGs was based on KOfams and was carried out using the anvi'o platform. We calculated the level of completeness for a given KEGG module [80, 81] in our genomes using the programme anvi-estimate-metabolism, which leveraged the previous annotation of genes with KEGG orthologs (KOs). The URL <https://merenlab.org/m/anvi-estimate-metabolism> serves as a tutorial for this programme which details the modes of usage and output file formats. The Heatmap of completion scores was illustrated using the ComplexHeatmap [82] package for R. MAGs were clustered based on similarity across the completion of pathways.

The statistical approach for enrichment analysis is previously defined [83]. Briefly, the programme anvi-compute-functional-enrichment determined enrichment scores for KOfams genomes of low- and high-abundance *Mycoplasma* by fitting a binomial generalised linear model (GLM) to the occurrence of each KOfam in each group and then computing a Rao test statistic. We considered any KOfam with a q-value less than 0.05 to be 'enriched' in its associated group. The volcano plot was visualised using the EnhancedVolcano package for R.

Presence of *Candidatus Mycoplasma salmoninae salar* in the Arctic Ocean

We searched for any detected Tenericutes (a phylum of *Mycoplasma*) across the TARA Ocean data to investigate the presence of *Mycoplasmas* in the Arctic Ocean with an eye toward *Candidatus Mycoplasma salmoninae salar*, using Supplementary information (Table S2) from Delmont et al., which holds the taxonomic information of 1887 recovered MAGs from the global ocean [84].

Population genetics of *Candidatus Mycoplasma salmoninae salar*

We use the term population to describe an assemblage of co-existing microbial genomes in an environment that can map to the context of the same reference genome [85]. SNVs, single codon variants (SCVs), and SAAVs were used to infer synonymous and non-synonymous mutations in the populations of *C. Mycoplasma salmoninae salar*.

The use of SNVs was related to interest in single nucleotide positions or non-coding regions, whereas SCVs were applied to resolve codon variants, synonymy, and calling of pNpS sites. Lastly, we analysed the structural

differences in the encoded proteins by inferring SAAVs. Analyses of SNVs, SCVs, and SAAVs were conducted using *anvi'o*/7.1, following the tutorial: <https://merenlab.org/2015/07/20/analyzing-variability/>. Briefly, to study the extent of variation of the MSS genes across all metagenomes, we instructed *anvi'o* to report positions with more than 1% variation at the nucleotide level (i.e., at least 1% of recruited reads differ from the consensus nucleotide). Here we only reported SNVs with a variation across 90% of the metagenomes to minimise zero inflation. For SCV and SAAV, the programme 'anvi-gen-variability-profile' (with an additional '-engine AA' flag and '-engine SCV' flag) reported variability tables describing the allele frequencies for each SAAV and SCV. *Anvi'o* only considers short reads that cover the entire codon to determine amino acid frequencies at a given codon position in a metagenome, as previously described [85]. Further inferring of SCVs, pNpS, and SAAVs was based on previous implementations testing for potential selective processes having shaped protein evolution [86]. Prediction of protein structures was carried out with MODELLER [87, 88], embedded in the *anvi'o* platform. Visualisation of SAAV on predicted protein structures was based on the *anvi'o* programme "anvi-display-structure" using the following tutorial: <https://merenlab.org/2018/09/04/getting-started-with-anvio-structure/>.

Inferring Atlantic salmon population structure

All host-related data were remapped to the *Ssal_v3.1* Atlantic salmon genome (Genbank accession GCA_905237065.2) using BWA mem. To estimate the origin of our samples, we spiked data from Atlantic salmon originating from southern and northern Norway, as detailed in Supplementary data S1. To assess the population structure of wild Atlantic salmon across Norway, we applied a genotype likelihood approach to handle low and medium-coverage data using ANGSD [89]. For the estimation of genotype likelihoods, we discarded reads with mapping quality below 20 and bases with base quality below 30.

For the PCA analysis, we used PCAnsd based on genotype likelihoods from variable sites [90]. We filtered genotypic sites for all population structure measurements to minimise false positives related to low coverage and varying sequencing depth across genomes. Filters included filtering for polymorphic sites with a SNP *p* value below 1e-6 and a minimum major/minor allele frequency of 0.05. Furthermore, genotypic sites should be present in at least 50% of the investigated individuals. For clustering of host genotypes, we based our approach on a previous latitudinal description, where the salmon in Norway originate from two main genetic lineages. One lineage from the Barents–White Sea refugium that recolonised northern Norwegian and adjacent Russian rivers, and one from the eastern Atlantic that recolonised the southern and western Norway [11, 12]. The division of the populations into two groups was based on their origin related to Målselva, which was used as a proxy for the middle of the transition zone between populations.

Statistical test for host-symbiont co-diversification

Parafit [91] was performed based on the Atlantic salmon distance matrix and MSS SNV distance matrix. The MSS SNV distance matrix was based on the SNVs occurring in MSS. The presence/absence of SNV among MSS from different host individuals was used for host-symbiont linkage. The Cailliez correction for negative eigenvalues was applied for the Parafit test [92], and a total of thousand permutations were used.

World maps

We used the *tidyverse* [93] and the *ggplot2* [94] packages for R to visualise the metagenomic data sets, eukaryotic gut compositions, and relative distribution of MAGs in the world map.

Statistical analyses

Rarefaction curves were estimated using the *r-package vegan* [95] to infer suitable sequencing depth. The abundance of MAGs was based on mean coverage across each MAG. Data were normalised based on MAG length and were sum-normalised before analysis of ecological dynamics based on TMP normalisation, using the *r-package ADImpute* [96]. Microbial composition analysis was carried out using the *r-package phyloseq* [97]. Correlated response models were used to infer the variation of each MAG related to environmental factors using the *r-package BORAL* [98] as previously described in ref. [44]. Statistical assumptions for parametric and non-parametric analyses were tested using Levene's test for homogeneity and the Shapiro-Wilk test for normality of data.

DATA AVAILABILITY

The raw dataset generated during the current study is available in the European Nucleotide Archive (ENA) repository with project accession number PRJEB59886. Data for comparing Atlantic cod metagenomes are available in the ENA repository with project accession number PRJEB29346. Furthermore, data for estimating the origin of Atlantic salmon populations are based on a previous study [12], which is available in the ENA repository with project accession number PRJEB38061.

CODE AVAILABILITY

The code used for this study is available at: https://github.com/JacobAgerbo/Norwegian_wild_Salmon_2022.

REFERENCES

- Alberdi A, Aizpurua O, Bohmann K, Zepeda-Mendoza ML, Gilbert MTP. Do vertebrate gut metagenomes confer rapid ecological adaptation? *Trends Ecol Evol* 2016;31:689–99.
- Groussin M, Mazel F, Alm EJ. Co-evolution and co-speciation of host-gut bacteria systems. *Cell Host Microbe*. 2020;28:12–22.
- Alberdi A, Andersen SB, Limborg MT, Dunn RR & Gilbert MTP. Disentangling host-microbiota complexity through hologenomics. *Nat Rev Genet*. 2021. <https://doi.org/10.1038/s41576-021-00421-0>.
- Volff J-N. Genome evolution and biodiversity in teleost fish. *Heredity*. 2005;94:280–94.
- Sommer F, Ståhlman M, Ilkayeva O, Arnemo JM, Kindberg J, Josefsson J, et al. The gut microbiota modulates energy metabolism in the Hibernating Brown bear *Ursus arctos*. *Cell Rep*. 2016;14:1655–61.
- Kohl KD, Weiss RB, Cox J, Dale C, Dearing MD. Gut microbes of mammalian herbivores facilitate intake of plant toxins. *Ecol Lett*. 2014;17:1238–46.
- Xiao L, Estellé J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, et al. A reference gene catalogue of the pig gut microbiome. *Nat Microbiol*. 2016;1:16161.
- Delgado B, Bach A, Guasch I, González C, Elcoso G, Pryce JE, et al. Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. *Sci Rep*. 2019;9:11.
- Xie F, Jin W, Si H, Yuan Y, Tao Y, Liu J, et al. An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome*. 2021;9:137.
- Ozerov M, Vähä J, Wennevik V, Niemelä E, Svenning M, Prusov S, et al. Comprehensive microsatellite baseline for genetic stock identification of Atlantic salmon (*Salmo salar* L.) in northernmost Europe. *ICES J Mar Sci*. 2017;74:2159–69.
- Wennevik V, Quintela M, Skaala Ø, Verspoor E, Prusov S, Glover KA. Population genetic analysis reveals a geographically limited transition zone between two genetically distinct Atlantic salmon lineages in Norway. *Ecol Evol*. 2019;9:6901–21.
- Bertolotti AC, Layer RM, Gundappa MK, Gallagher MD, Pehlivanoglu E, Nome T, et al. The structural variation landscape in 492 Atlantic salmon genomes. *Nat Commun*. 2020;11:5176.
- Huang Q, Sham RC, Deng Y, Mao Y, Wang C, Zhang T. Diversity of gut microbiomes in marine fishes is shaped by host-related factors. *Mol Ecol*. 2020. <https://doi.org/10.1111/mec.15699>.
- Cheab B, Yang P, Kazlauskaitė R, Lindsay E, Heys C, Dwyer T, et al. Genome erosion and evidence for an intracellular niche—exploring the biology of mycoplasmas in Atlantic salmon. *Aquaculture*. 2021;541:736772.
- Rasmussen JA, Villumsen KR, Duchêne DA, Puetz LC, Delmont TO, Sveier H, et al. Genome-resolved metagenomics suggests a mutualistic relationship between *Mycoplasma* and salmonid hosts. *Commun Biol*. 2021;4:579.
- Heys C, Cheab B, Buseti A, Kazlauskaitė R, Maier L, Sloan WT, et al. Neutral processes dominate microbial community assembly in Atlantic Salmon, *Salmo salar*. *Appl Environ Microbiol*. 2020;86:8.
- Llewellyn MS, McGinnity P, Dionne M, Letourneau J, Thonier F, Carvalho GR, et al. The biogeography of the atlantic salmon (*Salmo salar*) gut microbiome. *ISME J*. 2016;10:1280–4.
- Holben WE, Williams P, Gilbert MA, Saarinen M, Särkilahti LK, Apajalahti JHA. Phylogenetic analysis of intestinal microflora indicates a novel *Mycoplasma* phylotype in farmed and wild salmon. *Microb Ecol*. 2002;44:175–85.
- Limborg MT, Alberdi A, Kodama M, Roggenbuck M, Kristiansen K, Gilbert MTP. Applied hologenomics: feasibility and potential in aquaculture. *Trends Biotechnol*. 2018;36:252–64.
- Nyholm L, Koziol A, Marcos S, Botnen AB, Aizpurua O, Gopalakrishnan S, et al. Holo-omics: integrated host-microbiota multi-omics for basic and applied biological research. *iScience*. 2020;23:101414.
- Fietz K, Rye H, Christian O, Skovrind M, Kjærgaard Nielsen T, Limborg MT, et al. Mind the gut: genomic insights to population divergence and gut microbial composition of two marine keystone species. *Microbiome*. 2018;6:82.

22. Sevellec M, Derome N, Bernatchez L. Holobionts and ecological speciation: the intestinal microbiota of lake whitefish species pairs. *Microbiome*. 2018;6:47.
23. Knaus UG, Hertzberger R, Pircalabioru GG, Yousefi SPM, Branco Dos Santos F. Pathogen control at the intestinal mucosa—H₂O₂ to the rescue. *Gut Microbes*. 2017;8:67–74.
24. Smith CCR, Snowberg LK, Gregory Caporaso J, Knight R, Bolnick DI. Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota. *ISME J*. 2015;9:2515–26.
25. Meziti A, Rodriguez-R LM, Hatt JK, Peña-Gonzalez A, Levy K, Konstantinidis KT. The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample. *Appl Environ Microbiol*. 2021;87:e02593–20.
26. Bozzi D, Rasmussen JA, Carøe C, Sveier H, Nordøy K, Gilbert MTP, et al. Salmon gut microbiota correlates with disease infection status: potential for monitoring health in farmed animals. *Anim Microbiome*. 2021;3:1–17.
27. Zhao R, Symonds JE, Walker SP, Steiner K, Carter CG, Bowman JP, et al. Effects of feed ration and temperature on Chinook salmon (*Oncorhynchus tshawytscha*) microbiota in freshwater recirculating aquaculture systems. *Aquaculture*. 2021;543:736965.
28. Ciric M, Waite D, Draper J, Jones J, Brian. Characterization of mid-intestinal microbiota of farmed Chinook salmon using 16S rRNA gene metabarcoding. *Arch Biol Sci*. 2019;71:577–87.
29. Wang J, Jaramillo-Torres A, Li Y, Kortner TM, Gajardo K, Brevik Ø, et al. Microbiota in intestinal digesta of Atlantic salmon (*Salmo salar*), observed from late freshwater stage until one year in seawater, and effects of functional ingredients: a case study from a commercial sized research site in the Arctic region. *Anim Microb*. 2021;3:14.
30. Zhao R, Symonds JE, Walker SP, Steiner K, Carter CG, Bowman JP, et al. Salinity and fish age affect the gut microbiota of farmed Chinook salmon (*Oncorhynchus tshawytscha*). *Aquaculture*. 2020;528:735539.
31. Pérez-Pascual D, Vendrell-Fernández S, Audrain B, Bernal-Bayard J, Patiño-Navarrete R, Petit V, et al. Gnotobiotic rainbow trout (*Oncorhynchus mykiss*) model reveals endogenous bacteria that protect against *Flavobacterium columnare* infection. *PLoS Pathog*. 2021;17:e1009302.
32. Rimoldi S, Antonini M, Gasco L, Moroni F and Terova G. Intestinal microbial communities of rainbow trout (*Oncorhynchus mykiss*) may be improved by feeding a *Hermetia illucens* meal/low-fishmeal diet. *Fish Physiol Biochem*. 2021. <https://doi.org/10.1007/s10695-020-00918-1>.
33. Groussin, M, Mazel, F, Sanders, JG, Smillie, CS, Lavergne, S, Thuiller, W et al. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat Commun*. 2017;8:14319.
34. Zepeda Mendoza ML, Xiong Z, Escalera-Zamudio M, Runge AK, Thézé J, Streicker D, et al. Hologenic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nat Ecol Evol*. 2018;2:659–68.
35. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, et al. Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe*. 2019;25:656–667.e8.
36. Youngblut ND, Reischer GH, Walters W, Schuster N, Walzer C, Stalder G, et al. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat Commun*. 2019;10:2200.
37. Zepeda Mendoza ML, Roggenbuck M, Manzano Vargas K, Hansen LH, Brunak S, Gilbert MTP, et al. Protective role of the vulture facial skin and gut microbiomes aid adaptation to scavenging. *Acta Vet Scand*. 2018;60:61.
38. Moran NA. The coevolution of bacterial endosymbionts and phloem-feeding insects. *Ann Mo Bot Gard*. 2001;88:35–44.
39. Brucker RM, Bordenstein SR. The roles of host evolutionary relationships (genus: *Nasonia*) and development in structuring microbial communities. *Evolution*. 2012;66:349–62.
40. Nishiguchi MK, Ruby EG, McFall-Ngai MJ. Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in Sepiolid squid-vibrio symbioses. *Appl Environ Microbiol*. 1998;64:3209–13.
41. Kikuchi Y, Hosokawa T, Fukatsu T. Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. *Appl Environ Microbiol*. 2007;73:4308–16.
42. Roughgarden J. Holobiont evolution: mathematical model with vertical vs. horizontal microbiome transmission. *Philos Theory Pract Biol*. 2020;12:002.
43. Foster KR, Wenseleers T. A general model for the evolution of mutualisms. *J Evol Biol*. 2006;19:1283–93.
44. Rasmussen JA, Villumsen KR, von Gersdorff Jørgensen L, Forberg T, Zuo S, Kania PW, et al. Integrative analyses of probiotics, pathogenic infections, and host immune response highlight the importance of gut microbiota in understanding disease recovery in rainbow trout (*Oncorhynchus mykiss*). *J Appl Microbiol*. 2022. <https://doi.org/10.1111/jam.15433>.
45. Karlsen C, Ottem KF, Brevik ØJ, Davey M, Sørum H, Winther-Larsen HC, et al. The environmental and host-associated bacterial microbiota of Arctic seawater-farmed Atlantic salmon with ulcerative disorders. *J Fish Dis*. 2017;40:1645–63.
46. Brown RM, Wiens GD, Salinas I. Analysis of the gut and gill microbiome of resistant and susceptible lines of rainbow trout (*Oncorhynchus mykiss*). *Fish Shellfish Immunol*. 2019;86:497–506.
47. Kousoulaki K, Krasnov A, Ytteborg E, Sweetman J, Pedersen ME, Høst V, et al. A full factorial design to investigate interactions of variable essential amino acids, trace minerals and vitamins on Atlantic salmon smoltification and post transfer performance. *Aquac Rep*. 2021;20:100704.
48. Balk L, Hägerroth P, Gustavsson H, Sigg L, Åkerman G, Ruiz Muñoz Y, et al. Widespread episodic thiamine deficiency in Northern Hemisphere wildlife. *Sci Rep*. 2016;6:38821.
49. Skrodenyte-Arbačiauskienė V, et al. Gut microbiome of wild Baltic Salmon (*Salmo salar* L.) Parr. *Microb Ecol*. 2021. <https://doi.org/10.1007/s00248-021-01910-9>.
50. Amcoff P, Börjeson H, Landergren P, Vallin L, Norrgren L. Thiamine (Vitamin B₁) concentrations in Salmon (*Salmo salar*), Brown Trout (*Salmo trutta*) and cod (*Gadus morhua*) from the Baltic sea. *Ambio*. 1999;28:48–54.
51. Fisher JP, Fitzsimons JD, Combs GF Jr, Spitsbergen JM. Naturally occurring thiamine deficiency causing reproductive failure in finger lakes Atlantic salmon and great lakes lake trout. *Trans Am Fish Soc* 1996;125:167–78.
52. Keinänen M, Uddström A, Mikkonen J, Casini M, Pönni J, Myllylä T, et al. The thiamine deficiency syndrome M74, a reproductive disorder of Atlantic salmon (*Salmo salar*) feeding in the Baltic Sea, is related to the fat and thiamine content of prey fish. *ICES J Mar Sci*. 2012;69:516–28.
53. Rasmussen JA, Villumsen KR, Ernst M, Hansen M, Forberg T, Gopalakrishnan S, et al. A multi-omics approach unravels metagenomic and metabolic alterations of a probiotic and synbiotic additive in rainbow trout (*Oncorhynchus mykiss*). *Microbiome*. 2022;10:21.
54. Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev*. 2008;32:723–35.
55. Bordenstein SR, Theis KR. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol*. 2015;13:e1002226.
56. Lundebye A, Lock E, Rasinger JD, Nøstbakken OJ, Hannisdal R, Karlsbakk E, et al. Lower levels of persistent organic pollutants, metals and the marine omega 3-fatty acid DHA in farmed compared to wild Atlantic salmon (*Salmo salar*). *Environ Res*. 2017;155:49–59.
57. Madhun AS, Isachsen CH, Omdal LM, Einen ACB, Maehle S, Wennevik V, et al. Prevalence of piscine orthoreovirus and salmonid alphavirus in sea-caught returning adult Atlantic salmon (*Salmo salar* L.) in northern Norway. *J Fish Dis*. 2018;41:797–803.
58. Madhun AS, Maehle S, Wennevik V, Karlsbakk E. Prevalence and genotypes of infectious salmon anaemia virus (ISAV) in returning wild Atlantic salmon (*Salmo salar* L.) in northern Norway. *J Fish Dis* 2019;42:1217–21.
59. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* 2013.
60. Petersen TN, Lukjancenko O, Thomsen MCF, Maddalena SM, Lund O, Møller Aarestrup F, et al. MGMapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads. *PLoS One*. 2017;12:e0176469.
61. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31:1674–6.
62. Murat EA, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015;3:e1319.
63. Eren AM, Kiehl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol*. 2021;6:3–6.
64. Hyatt D, Chen G, Locascio, Philip F, Land ML, Larimer FW, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma*. 2010;11:119.
65. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–37.
66. Lee MD. GTOTree: a user-friendly workflow for phylogenomics. *Bioinformatics*. 2019;35:4162–4.
67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
68. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods*. 2014;11:1144–6.
69. Delmont TO, Eren AM. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ*. 2018;6:e4320.
70. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. 2018;36:996–1004.
71. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun*. 2016;7:11257.
72. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*. 2000;28:33–36.

73. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
74. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47:D427–D432.
75. Riiser ES, Haverkamp THA, Varadharajan S, Borgan Ø, Jakobsen KS, Jentoft S, et al. Switching on the light: using metagenomic shotgun sequencing to characterize the intestinal microbiome of Atlantic cod. *Environ Microbiol.* 2019;21:2576–94.
76. Price MN, Dehal PS, Arkin AP. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5:e9490.
77. Benedict MN, Henriksen JR, Metcalf WW, Whitaker RJ, Price ND. ITEP: an integrated toolkit for exploration of microbial pan-genomes. *BMC Genom.* 2014;15:8.
78. van Dongen S, Abreu-Goodger C. Using MCL to extract clusters from networks. *Methods Mol Biol.* 2012;804:281–95.
79. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 2015;8:12–24.
80. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. et al. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 2014;42:D199–205.
81. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45:D353–D361.
82. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32:2847–9.
83. Shaiber A, Willis AD, Delmont TO, Roux S, Chen L, Schmid AC, et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biol.* 2020;21:292.
84. Delmont TO, Pierella Karlusich JJ, Veseli I, Fuessel J, Eren AM, Foster RA, et al. Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean. *ISME J.* 2022;16:927–36.
85. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, et al. Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. 245;2019;8.
86. Kiefl E, Esen OC, Miller SE, Kroll KL, Willis AD, Rappé MS, et al. Structure-informed microbial population genetics elucidate selective pressures that shape protein evolution. *bioRxiv.* 2022.03.02.482602. 2022, <https://doi.org/10.1101/2022.03.02.482602>.
87. Sánchez R, Šali A. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins.* 1997;29:50–58.
88. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M, et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinform.* 2006;5:5.6.
89. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* 2014;15:356.
90. Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen A, et al. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics.* 2013;195:979–92.
91. Legendre P, Desdevises Y, Bazin E. A statistical test for host-parasite coevolution. *Syst Biol.* 2002;51:217–34.
92. Cailliez F. The analytical solution of the additive constant problem. *Psychometrika.* 1983;48:305–8.
93. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. *J Open Source Softw.* 2019;4:1686.
94. Ginestet C. ggplot2: elegant graphics for data analysis. *J R Stat Soc Ser A.* 2011;174:245.
95. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927–30.
96. Xu L, Xu Y, Xue T, Zhang X, Li J. AdImpute: an imputation method for single-cell RNA-seq data based on semi-supervised autoencoders. *Front Genet.* 2021;12:739677.
97. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One.* 2013;8:e61217.
98. Hui FK. C. boral - Bayesian ordination and regression analysis of multivariate abundance data in r. *Methods Ecol Evol.* 2016;7:744–50.

ACKNOWLEDGEMENTS

The research was funded by The Independent Research Fund Denmark (“HappyFish”, grant No. 8022-00005B) and by The Danish National, Research Foundation grant no. DNR143 to MTPG Sampling was funded by internal funding of the Institute of Marine Research (project number 15697) and in collaboration with the Kolarctic salmon project financed by the European Union (Kolarctic ENPI CBC project KO197 “Trilateral cooperation in our common resource; the Atlantic salmon in the Barents Region). During the review process of this article, Prof. Rune Waagbø passed away unexpectedly. Rune was an international authority in the field of Fish Nutrition and a beloved colleague and friend. We are grateful to have had the opportunity to work with him.

AUTHOR CONTRIBUTIONS

JAR and MTL conceived the study with input from MTPG and KK. PK, ASM, JEL, and RW organised and carried out sampling. JAR carried out laboratory work and performed the computational analysis. The authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41396-023-01379-z>.

Correspondence and requests for materials should be addressed to Jacob A. Rasmussen or Morten T. Limborg.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023