Pseudogenization, genome streamlining and specific gene repertoire landmark the genomes of *Carnobacterium maltaromaticum* isolated from diseased sharks

Authors: Laura Martinez Steele[1*], Christopher G Lowe[1¶], Mark S Okihiro[2&], Jesse G. Dillon[1&], Renaud Berlemont[1¶]

Affiliation:
[1]California State University Long Beach, Department of Biological Sciences, Long Beach, CA, USA.

[2]California Department of Fish and Wildlife, Fisheries Branch, Wildlife and Fisheries Division, Vista, CA, USA

*Corresponding author:

E-mail: Lmartinezsteele@gmail.com

¶ These authors contributed equally to this work

& These authors contributed equally to this work

Abstract

Carnobacterium maltaromaticum is a well-known pathogen of bony fish. More recently, *C. maltaromaticum* have been isolated from the brain and inner ear of disorientated and stranded common thresher (*Alopias vulpinus*) and salmon shark (*Lamna ditropis*). While thresher shark strandings are recent, salmon sharks have been stranding for decades, suggesting a long-term association between *C. maltaromaticum* and sharks. Interestingly, some strains of *C. maltaromaticum* are used by the food industry for their probiotic and antimicrobial activity. Here, we sequenced the genome of 9 *C. maltaromaticum* strains (SK-isolates) from diseased common thresher and salmon sharks and compared them to other *C. maltaromaticum* strains in order to identify the genomic signatures that differentiate the disease-associated from the innocuous *C. maltaromaticum* isolates. SK strains formed a monophyletic clade, with a conserved gene repertoire, and shared a high degree of pseudogenization even though isolates were from different shark species, locations, and across years. In addition, these strains displayed few virulence associated genes and unique genomic regions, some resulting from horizontal gene transfer. The association of diseased sharks and SK strains suggests their role as potential pathogens. Although the high degree of pseudogenization suggests a transition to a host-adapted lifestyle, a set of conserved functional genes highlights the need of essential functions required for a host-independent life style. Globally, this work identifies specific genomic signatures of *C. maltaromaticum* strains isolated from infected sharks, provides the framework to elucidate the role of SK strains in the development of the disease in sharks, and further investigate the dissemination of SK strains in populations of wild fish.

Introduction

Occurrences of stranded salmon sharks (*Lamna ditropis*) [1] and common thresher sharks (*Alopias vulpinus)* [2] have been happening along the West Coast of North America. These stranding events involve disoriented, sick juveniles that swim onto the beach, and occur throughout the year although mostly reported from July to September [3]. Both thresher and salmon sharks belong to the order of Lamniformes, and are top predators that migrate between the coastal and offshore waters of the Northeast Pacific Ocean, and use the continental shelf as nurseries when juveniles [4]. Histology of the brain and inner ear of stranded sharks revealed severe meningitis [1] and acute otitis, most probably caused by *Carnobacterium maltaromaticum* (phylum Firmicutes) identified in bacterial cultures obtained from damaged tissues [1, 2]. *C. maltaromaticum* is a facultative anaerobic, psychrotrophic, lactic acid bacterium, of biotechnological interest due to its antimicrobial properties inhibiting *Listeria monocytogenes* [5, 6]. It is also detected in many natural environments including the gastrointestinal tract of many teleost [7-11], where it is thought to stimulate the immune system, and thus is sometimes used as a probiotic in aquaculture [12, 13]. However, *C. maltaromaticum* is also a pathogen in cold-water teleost fishes exposed to various stressors (e.g., spawning, handling events) [11, 14]. In diseased fish, *C. maltaromaticum* is associated with pseudokidney disease, septicemia, splenomegaly, internal hemorrhages, muscular abscesses, visceral congestion, and thickening of the swim-bladder walls [11, 14, 15].

As of June 2018, the genus *Carnobacterium* comprises 11 species and 42 genomes are publicly accessible through the NCBI database (S1 Table, S2 Table). Genomes range from 1.5 to 4.0 Mbp and have GC content ranging from 34 to 39.4%. Among sequenced genomes, *C. maltaromaticum* strains have the largest, yet variable, genomes (3.3 ± 0.7 Mbp) mirroring the

74 gain and loss of massive DNA fragments [16]. Additionally, *C. maltaromaticum* is the only
75 species to be reported in multiple environments including processed food, human skin, and the
76 digestive system of some teleost species [6].
77 Here, we characterized the genome of 9 *C. maltaromaticum* (SK strains) isolated from the brain
78 and inner ear of stranded diseased-sharks (i.e., common thresher shark and salmon shark) from
79 along the coast of North America between 2013 and 2016. Despite varying sample origin and
80 collection times, we hypothesized that these strains would be phylogenetically related and would
81 display specific genomic signatures discriminating them from strains originating from other
82 sources. Thus, we first investigated the phylogenetic placement of the newly sequenced strains
83 relative to publicly accessible *Carnobacterium* genomes using 16S rRNA sequence,
84 tetranucleotide correlation search (TCS), and based on single nucleotide polymorphisms (SNPs)
85 distribution in the core genome of all sequenced *C. maltaromaticum*. Next, we compared the
86 functional potential, based on annotated genes, in the nine SK strains and in all the sequenced *C.*
87 *maltaromaticum* (n_{total}=17) in relation to their known habitat and lifestyle (i.e., food-derived and
88 diseased teleost or sharks).
89 The consistent identification of the SK strains in stranded sharks using different techniques such
90 as cytology, histology and bacterial culture [1, 2], suggests the potential role of *C.*
91 *maltaromaticum* in the sharks' disorientation and stranding. Thus, we specifically investigated
92 the distribution of potential genes involved in pathogenicity. Furthermore, we searched for
93 differences in the proportion of functional genes discriminating diseased-shark isolates and food-
94 associated strains, to identify niche-specific functions associated with the origin of the strains
95 [17, 18]. We also identified regions of lateral transfer such as "Genomic Islands" (GIs) [19, 20]
96 and bacteriophages [21-23]. Finally, we determined the accumulation of pseudogenes [24, 25] in

97 SK strains, relative to the other strains, as an indication of the reductive evolutionary processes
98 often observed in pathogens [26-28].
99 For the first time, this study investigates the genomic signatures of *C. maltaromaticum* isolated
100 from diseased-sharks in comparison to diseased-teleost and food relatives with the purpose of
101 elucidating evolutionary forces that shaped the genome of strains associated with disease in wild-
102 fish populations.

103 # Materials and methods

104 ## Strains

105 Nine strains of *C. maltaromaticum* (SK strains) were isolated from the brain and inner ear of
106 stranded common thresher sharks and salmon sharks and sequenced as described before (S1
107 Table) [2]. For phylogenetic placement, genomic DNA extracts were PCR-amplified using the
108 primers GM3/GM4 [29] targeting the 16S rRNA gene and the products were Sanger-sequenced
109 and deposited to the NCBI (S1 Table). Sequences were BLASTed against the NCBI database
110 and were confirmed as *Carnobacterium maltaromaticum* (>99% identity).
111 For the comparative genome analysis, we first used 41 publicly accessible genomes from NCBI
112 database covering different species from the genus *Carnobacterium,* and five pathogenic
113 outgroups (S2 Table). Then we characterized 17 *C. maltaromaticum* genomes including 9 SK
114 strain genomes (Table 1) and 8 isolated from various environments. The strain 757 CMAL
115 (human isolate) was excluded from the *C. maltaromaticum* analysis because of its divergence
116 from the rest of the strains, which might potentially interfere with the analysis. For consistency,
117 contigs less than 3 kb were removed from all 46 genomes and were consecutively (re)annotated
118 with the Rapid Annotation using Subsystem Technology (RAST) online server (v 2.0) [30, 31].
119

120

121 **Table 1. Analyzed *C. maltaromaticum* genomes.**

Strain	Size (Mbp)	CDS	GC (%)	Contigs	Contigs > 3 Kb	N50	Pseudo genes	Source/ Life-style	Ref.
3-18	3.5	3245	34.3	160	71	82,960	41	Pork	[32]
ATCC 35586	3.5	3333	34.5	74	63	106,236	70	Diseased trout	[33]
DSM 20342	3.7	3481	34.3	132	54	79,358	48	Milk with malty flavor	[34]
DSM 20342 MX5	3.9	3636	34.4	5	5	3,620,667	71	Milk with malty flavor	[34]
DSM 20722	3.6	3301	34.3	60	22	133,412	41	Vacuumed meat	[35]
DSM 20730	3.5	3254	34.4	143	28	75,250	19	Diseased trout	[35]
LMA 28	3.7	3472	34.5	1	1	NA	155	Cheese	[36]
ML 1 97	3.2	3108	34.4	229	176	26,642	283	Fresh salmon	[32]
SK AV1	3.3	3220	34.3	42	18	312,952	262	Thresher shark 1 brain	[2] This study
SK AV2	3.3	3219	34.3	123	19	312,934	264	Thresher shark 1 ear	
SK AV3	3.3	3222	34.3	44	18	312,857	263	Thresher shark 8 brain	
SK AV4	3.3	3220	34.3	44	18	312,883	262	Thresher shark 8 ear	
SK AV5	3.4	3300	34.4	3,070	33	312,904	263	Thresher shark 11 brain	
SK AV6	3.3	3225	34.3	116	19	312,881	263	Thresher shark 11 ear	
SK LD1	3.3	3220	34.3	70	18	389,499	262	Salmon shark 1 Brain	
SK LD2	3.3	3218	34.3	79	17	312,874	265	Salmon shark 1 ear	
SK LD3	3.3	3219	34.3	76	19	312,880	263	Salmon shark 2 brain	

122 Genome characteristics are extracted from the RAST except for the pseudogenes, which were

123 predicted using DFAST, and contigs which were taken from the NCBI.

124 Number of contigs are shown before and after contigs < 3 Kb were removed, while the rest of the

125 features shown are after contigs < 3 Kb were removed

Phylogenetic placement

The nine SK strains were phylogenetically identified by comparing the complete 16S rRNA gene sequences from 28 *Carnobacterium* genomes with entire 16S rRNA sequence and 5 pathogenic outgroups (n_{total}= 42 strains, S2 Table). Sequences were aligned using ClustalW [37] and we constructed a phylogenetic tree using the PHYLIP-phylogeny interference package (version 3.695, F84 distance method, and neighbor-joining tree) [38]. The tree was visualized using the interactive tree of life (iTOL v4) [39].

Phylogenomic placement

We first performed TCS [40] against the JSpeciesWS database [41], and Average Nucleotide Identity (ANIb) among the newly sequenced genomes [42]. Next, we used the Pangenome Analysis Tool from Panseq [43] to identify SNPs in the core genome of the 17 *C. maltaromaticum* species only. We used default identity cut-off of 85% to exclude genes acquired recently through horizontal gene transfer. Next, a tree based on identified SNPs was built as previously described in the phylogenetic analysis of this study. Finally, we computed a clustering analysis of the genomes based on their functional distribution (FigFam-based annotation [30, 31]), using the pairwise Bray-Curtis dissimilarity index and visualized using complete linkage clustering [44] (n_{total}= 46 genomes, S2 Table).

Orthologous clustering and pan-genome functional annotation

Anvi'o pangenomic workflow [45] was used to identify and annotate gene clusters with the 2014 COG database [46]. Briefly similarities between genes were performed using BLASTp and weak matches were eliminated with *minibit heuristic* [45]. Next, clustering was performed with the Markov Cluster algorithm (MCL) [47] using default parameters with the exception for a mcl-

148 inflation of 10 (used for closely related strains). The core, accessory and unique genes and their
149 COG function and category distribution were extracted and used for further analysis. For this
150 study core genome was defined as the number of genes present in all genomes, while accessory
151 genes were those gene clusters present in more than 1 but less than 17 genomes, and finally
152 singletons were genes clusters just found in one genome.
153 Because changes in functional distribution mirrors adaptation to different environments, we
154 performed Wilcoxson Rank Sum test with Bonferroni correction for multiple comparison to
155 detect differences in the COG categorical functions between isolates from food (n=6), diseased
156 trout (n=2), and diseased shark (n=9).

157 ## Pan-genome profile

158 A panmatrix containing gene clusters presence in 17 *C. maltaromaticum* genomes was imported
159 from Anvi'o [45] to the Micropan package implemented in R [48] to generate pan-genome
160 visualizations [49]. The Heaps law model [50] and the Binomial Mixture Model, using the
161 Bayesian Information Criterion (BIC) [51], implemented in the package were used to determined
162 openness of the pan-genome and to estimate pan and core-genome size when envisioning an
163 "infinite" number of genomes. Rarefaction curve using 100 permutations helped visualize the
164 increase of the pan-genome size with the addition of new genomes.

165 ## Virulent associated genes

166 We searched for virulence factors, within the 17 *C. maltaromaticum* genomes, using protein
167 sequences provided by the RAST annotation, and by performing a BLAST search (60% coverage
168 and 60% identity) against the Virulence Factor Database (VFDB, [52]). These results were
169 combined with genes under the *Virulence, Disease and Defense* category from the corresponding
170 FigFam-annotated genomes [30, 31].

Genomic plasticity

GIs were identified and visualized with Islandviewer v.4 [53] using annotation files retrieved from the RAST. *C. maltaromaticum* LMA28 was used as a reference to re-order contigs in draft genomes with default parameters. To determine unique islands in the SK strains we compared them to islands detected in the other strains using BLAST (60% coverage and 60% identity). Prophage sequences were predicted in re-ordered genomes using PHASTER [54]. This tool scores the results according the completeness of the sequence taking into account length, gene content, GC content and attachment sites. Results were labeled as intact, questionable and incomplete. Intact phage sequences were than scan for virulence genes as previously explained. The script "rod_finder" [55] was used to determine regions of difference (RODs) in each genome in order to determine DNA sequences unique to the SK clade. This script was used between SK_AV1 strain and non-SK strains with a minimum size of 5,000 bp. RODs were then BLASTed against each other (60% coverage and 60% identity cut-off) to determine their presence in the other genomes. Finally, RODs, phages, and GIs where visualized in BRIG [56], using the strain SK_AV1 as reference genome for the alignment and all their genes were re-annotated with WebMGA (e-value of 10^{-5} [57]).

Genome degradation

DFAST [58] was used to determine pseudogenization resulting from frameshift mutation and nonsense-mutation in the genes of the 17 *C. maltaromaticum* genomes using the LMA 28 strain as a reference. Pseudogenes where annotated with WebMGA [57].

Results

Strains identification

First, we investigated the phyletic affiliation of the nine *C. maltaromaticum* SK strains isolated from the infected brain and inner ear of 5 stranded sharks using phylogenetic and phylogenomic approaches (Table 1). All the nine strains were identified as *C. maltaromaticum* according to complete 16SrRNA gene analysis (Fig 1A), the Tetra-nucleotide correlation analysis (TCS >0.99), and the Average Nucleotide Identity (ANIb >98%). Furthermore, 16S rRNA sequences from all publicly accessible genomes of *Carnobacterium* (n=37) (S2 Table) revealed little phylogenetic divergence among sequenced genomes of *C. maltaromaticum* with the exception of the *C. maltaromaticum* 757 CMAL, derived from human skin (Fig 1A). However, both the identification of 91,236 identified SNPs distributed throughout the core genomes of the 17 *C. maltaromaticum* genomes (Fig 1B) and the clustering based on functional genes (FigFams) distribution (S1 Fig), produced a more distantly branched cluster with the nine SK strains, suggesting the shared evolution of the SK strains and their divergence from other known isolates. The 17 *C. maltaromaticum* strains were classified into three groups, diseased-shark (n=9), diseased-trout (n=2), and food-associated isolates (n=6).

Fig 1. *C. maltaromaticum* phylogeny according to the 16S rRNA gene and SNPs.
(**A**) Neighbor-joining tree of the 16S rRNA gene of 37 *Carnobacterium sp.* along with 5 pathogenic non-*Carnobacterium* strains as outgroup. Scale represents the frequency of substitution per site. (**B**) Unrooted neighbor-joining tree according to the SNPs in the core genome of the 17 C. maltaromaticum strains. Isolates from diseased-sharks: red, diseased-trout: purple, food: blue, and human skin: orange.

Genomic plasticity & Gene content

Considering the 17 *C. maltaromaticum* genomes (Table 1), the pan-genome neared completion (Heaps law model, $\alpha = 0.93$, S2 Fig) and consisted of 4,746 gene clusters (3,074.6 ± 136.9

clusters per genome) annotated with 1,444 different functions (Table S3). Conversely, the binomial mixture model, anticipated a final core and pan-genome size of 2,285 and 5,903 gene-clusters, respectively, suggesting potential missing gene clusters from the *C. maltaromaticum* pan-genome. As shown by the actual gene distribution among the *C. maltaromaticum* (Fig 2), there was relatively high gene diversity among the 17 genomes, reflecting their diverse life-styles and isolation sources. However, most of the genes in the SK strains belonged to the SK-clade's core-genome, with the accessory genes representing a smaller proportion of the genome (Fig 2). Thus, most of the genome diversity observed within the *C. maltaromaticum* group reflects variation in the gene content outside the SK clade.

Fig 2. Core, accessory and singleton genes within the *C. maltaromaticum* pan-genome. (A) Distribution of gene clusters in the 17 *C. maltaromaticum* genomes. (B) Distribution of gene clusters in the 9 *C. maltaromaticum* SK genomes. Bar plot shows core genome (red), singletons (blue) and accessory gene clusters found in more than 1 and < 17 genomes (black bars). Comparison of both Figs reveals that SK strains have few genes different from each other while the accessory genome in all 17 strains represents a big proportion of the pan-genome.

The core, accessory, and singleton genes in the 17 strains represented 50.9%, 34.5% and 14.5% of the identified genes, respectively (Fig 2). In addition, 33.1% of the identified gene clusters, mostly in the accessory and singleton category, remained with unknown function. Only 106 clusters unique to the SK strains had functional annotation. Among these, only 19 COG functions were not shared by the other *C. maltaromaticum* genomes (Table S4), and only 9 were present in all the SK strains. Apart from the *unknown functions*, predominant COG functions in the core genome were associated to housekeeping genes such as *Translation, ribosomal structure and biogenesis* (12.35 %)*, Amino acid transport and metabolism* (9.35 %). Conversely, the accessory genes and singletons, involved in the adaptation to the environment, had a higher occurrence of functions involved in *Mobilome: prophages, transposons* (accessory 12.13 %), *Cell wall/membrane/envelope biogenesis* (singletons 11.11%), *Defense mechanisms* (singletons

246 11.11 %), *Extracellular structures* (singletons 8.33%) and *Cell motility* (singletons 8.33%) (S3
247 Fig).

248 The systematic comparison of the 17 *C. maltaromaticum* genomes highlighted 6 GIs conserved
249 in all the SK genomes, four of them being unique to the SK clade (Fig 3). A complete phage
250 sequence (i.e., PHAGE_Strept_phiARI0746_NC_031907 from *Streptococcus pneumoniae*
251 10B04751, Phage 4) was also identified. Finally, two incomplete phages (Phages 2 and 3) and
252 one potential phage (Phage 1) sequences were also found in all SK genomes (Fig 3). However,
253 no virulent gene was found in the phages or GIs when aligning the corresponding sequences
254 against the VFDB. In total 35 RODs were identified, but only 5 were specific of the diseased-
255 shark strains and 4 were also shared by at least 1 of the diseased-trout strains (Fig 3). Although
256 most of the genes identified in the phages, GIs, and RODs had unknown functions (70%, 71%,
257 and 79%), many genes with annotated function in the RODs were associated with cell wall
258 synthesis and capsular polysaccharide production (S5 Table).

259 **Fig 3. Alignment of *C. maltaromaticum* genomes against SK AV1 strain.**
260 The outer light blue ring contains all the SK clade (besides the AV1, which was used as a
261 reference for the alignment). RODs in red are unique to shark and trout pathogen, whereas the
262 black RODs are unique to the SK clade. Green represent phage-derived sequences identified by
263 PHASTER and pink represent GIs unique to the SK clade identified by IslandViewer.
264

265 # Functional variation between groups

266 When comparing the distribution of predicted functions in the analyzed genomes, we identified
267 significant differences between the SK strains (n=9 genomes) and the food-associated strains
268 (n=6 genomes) whereas no significant difference was observed when considering diseased-trout
269 strains (n=2 genomes, data not shown). More precisely, the SK strains had significantly less
270 genes for *Replication, recombination, and repair* and *Mobilome: prophages, transposons*. In
271 addition, the SK strains consistently displayed reduced number of genes for *Cell motility*

although not significantly. This mirrored the variable number of genes for this function in the non-SK strains. Conversely, the SK strains were significantly enriched in genes for *Secondary metabolites biosynthesis, transport and catabolism*, *Post-translational modification*, and *Energy production and conversion*, among others (Fig 4).

Fig 4. Functional comparison between strains isolated from food and diseased sharks. X axis represents the log ratio of the COG categories relative abundance. Asterisks indicate level of significant (one asterisk $p < 0.05$, two asterisk $p < 0.01$) when performing a Wilcoxson Rank Sum test with Bonferroni correction for multiple comparison. Variations between these groups indicate their adaptation to different environments associated different functional needs.

The SK strains and the other *C. maltaromaticum* genomes contained 29.3 ± 1 and 29.3 ± 4 virulent associated genes, respectively. However, seven were only found in the SK strains. These genes coded for a potential enzyme involved in capsular polysaccharide synthesis, a dTDP-4-dehydrorhamnose reductase, and a prolipoprotein diacylglyceryl transferase all found in the 9 SK strains. In addition, a PixD protein, a bifunctional aldehyde-alcohol dehydrogenase, a cytotoxic necrotizing factor 1, and a fimbriae were identified only in some of the SK strains (S6 Table). Using the FigFam annotations, we compared and identified unique functions inside the *Virulence, Disease and Defense* category (S4 Fig). SK strains had an overrepresentation of genes for the *Invasion and intracellular resistance* (with a SK-unique internalin like protein *Lmo0327*), and *Bacitracin* (with a SK-unique bacitracin stress response regulator). Furthermore, the SK clade showed unique presence of fosfomycin resistance and beta-lactamase along with other genes involved in the synthesis of cell wall and capsule, phages, iron metabolism, RNA and nucleotide metabolism, respiration and stress response (S7 Table).

Pseudogenization

Within the 17 *C. maltaromaticum* genomes, the SK strains displayed a high frequency of pseudogenes (263 ± 1) resulting mostly from non-sense mutations (Table 1). Exceptionally, the

298 *C. maltaromaticum* ML_1_97, isolated from fresh salmon, also had a high number of
299 pseudogenes, perhaps as a consequence of the high number of contigs of this genome assembly
300 (Table 1). Pseudogene functional annotation in SK strains showed high frequency of genes
301 annotated under the categories of *Carbohydrate transport and metabolism* (13.1%), *Amino acid*
302 *transport and metabolism* (10.6%), and *Energy production and conversion* (9.7%) (S5 Fig).
303 Pseudogenization potentially inactivated 115 functions in the SK strains and caused the strains
304 from food and diseased-trout to have a total of 166 and 126 functions absent from the SK cluster,
305 respectively.

306 # Discussion

307 *Carnobacterium* are environmental microbes [6] frequently identified in dairy products [59],
308 whereas some species are able to cause disease in fishes [11, 14, 15]. Members of this genus are
309 hard to identify and discriminate using standard techniques (e.g., phenotypic characterization,
310 16S rRNA phylotyping) because of their phylogenetic proximity and functional plasticity [11,
311 60]. However, a recent study using Multilocus Sequence Typing revealed that *C.*
312 *maltaromaticum* causing disease in teleosts are non-clonal and vary from isolates from dairy
313 products [61].
314 Recently, *C. maltaromaticum* isolates have been isolated from the brain and inner ear of stranded
315 salmon sharks [1] and common thresher sharks [2] found with severe brain and inner ear
316 infection. To date, both the detailed epidemiology and the exact etiology of these infections
317 remain largely unknown; however, the consistent presence of *C. maltaromaticum* in infected
318 areas highly suggests its role in the development of the pathology, as it is involved in the
319 development of disease in bony fishes [11, 14, 15]. Here, we performed a thorough comparison
320 of *C. maltaromaticum* sequenced genomes in order to identify the genomic features associated

with niche adaptation in this clade. As expected, the systematic comparison of sequenced genomes from the *Carnobacterium* genus highlights the phylogenetic cohesion of all the *C. maltaromaticum* isolates displaying less than 1% 16S rRNA variation [11, 60]. The phylogenetically coherent *C. maltaromaticum* cluster, containing the SK strains, is related with the *C. divergens* and *C. gallinarum* species [59], two species frequently detected across environments [6], whereas the other species formed a distinct and more distantly related cluster. Interestingly, the clustering of *C. maltaromaticum* SK strains according to phylogenomic markers (e.g., SNPs, functional gene distribution) correlated with their phylogenetic clustering [44]. Thus, all the SK isolates are distantly related to other lineages of *C. maltaromaticum* used in this study, and share a unique specific gene repertoire. This suggests that SK strains share the same evolutionary history and do not seem to originate from food or teleost contamination sources [12, 13].

Since the genomic plasticity of bacteria underlines the ability of phylogenetically related microbes to colonize distinct environments [62] and sometimes to be involved in pathologies [19, 21, 63], we expected the SK strains to display specific genomic adaptation of the host-adapted lifestyle. Frequently identified features associated with this life-style include genomic reduction mirrored by the reduced genome and pseudogenization [24, 25]. Furthermore, the presence of specific genes, such as the virulence genes, can aid in the invasion process [52]. Genome streamlining is a gradual process, which erodes genes and functions potentially supported by the host, and eventually results in a reduced genome with only essential functions conserved. Similar genome reduction is observed in well-known pathogens *Shigella* sp. and *Salmonella enterica* [64, 65]. Many of the pseudogenes identified in the SK strains most likely affected the functionality of genes as a consequence of the introduction frameshift, indels, and

344 random stop codons in the coding regions (Table 1). The accumulation of pseudogenes in
345 sequences likely results in the loss of function, suggesting a stronger connection between the
346 microbe and its host [25, 66]. Here, the pseudogenization of SK genomes affected mostly genes
347 involved in central metabolism (e.g., *Amino acid transport and metabolism*). Similar
348 pseudogenization has been identified in pathogens with the host's cellular machinery supporting
349 the loss of function [67, 68]. However, as genomic degradation reduces the fitness in the open
350 environment, losing too many functions or of essential functions could have a detrimental effect
351 on the ability of the strain to propagate [69, 70]. Thus, unlike some highly specialized pathogens,
352 many diseased-causing agents rely on their ability, even reduced, to survive and potentially
353 propagate in the environment [71].
354 Interestingly, apart from the described loss of accessory genes and the pseudogenization of some
355 important functions, the SK strains displayed a specific set of intact genes potentially associated
356 with the development of the pathology. Among others, SK strains have two extra genes encoding
357 ABC transporter for polysaccharide export and potentially involved in exopolysaccharide
358 production that could promote the evasion of the immune system of the shark [72-74]. In
359 addition, *CapO* (S6 Table), a gene essential in the for capsule synthesis [75], is only found in SK
360 strains. Furthermore, some peroxiredoxin genes, potentially involved in protecting the bacteria
361 from oxidative stress mediated by the host immune system [76-78], were also more abundant in
362 the SK strains (S6 Table). Finally, an internalin-gene, supporting the colonization of the host by
363 facilitating the crossing of the intestinal epithelium and blood-brain barrier [79] was also unique
364 to the SK strains (i.e., *Lmo0327*) (S6 Table).
365 The low frequency of genes from Mobile Genetic Elements in the SK strains relative to the other
366 food-derived and diseased teleost-derived genomes (Fig 4), could be due their limited interaction

367 with environmental microbial communities, which limits the potential to gain new functions
368 through horizontal gene transfer and further reduce the genome diversity. This phenomenon is
369 often found in pathogens [80]. Nonetheless, a complete phage and four GIs (three of likely phage
370 origin) were specific to the SK clade (Fig 3). Additionally, the SK strains displayed nine RODs
371 of unknown origin, with four being shared with the trout pathogen (i.e., *C. maltaromaticum*
372 ATTC 35586). The phages and the RODs sequences, although containing many genes with
373 unknown function, displayed several genes involved in cell-wall and exopolysaccharide
374 biosynthesis.
375 In total, we compared 17 *C. maltaromaticum* bacterial genomes in order to identify the genomic
376 features associated with the 9 strains isolated from diseased-sharks. However, we recognize that
377 many of these genomes are not complete and that the nature of the missing information could
378 affect our results and interpretation in different ways, for example by inferring the wrong number
379 of genes (potentially adding or subtracting genes due to the fragmentation of the genome into
380 multiple contigs, and the presence of gaps) [81-83]. However, most of the genomes used in this
381 study had low contig number (19.9±5 for SK strains and 34.9±28 for remaining seven, with the
382 exception of ML 1 97 which had 176 contigs), and large contig size (N50 > 312,000 for SK
383 strains and N50 > 75,000, for all genomes with the exception of strain ML 1 97 with N50 =
384 26,642) (Table 1), which decrease the potential for missing or false produced genes [81].
385 Nonetheless, over or under-representation of gene numbers could be affecting the identification
386 of specific genomic features. However, sequencing 9 SK genomes confirmed the presence of
387 features of interest in all strains and reduced the possibility of missing genes. Additionally, for
388 consistency, all the publicly accessible genomes included in this study were re-annotated and

389 treated in the same way in order to minimize systemic bias introduced by different
390 bioinformatics pipelines.
391 Contrary to high diversity among teleost pathogens [61], the strains isolated from these diseased-
392 sharks have very similar genomes, with only few identified accessory genes. The conservatism
393 of identified genomic features such as the functional genes and the pseudogenization in the SK
394 strains supports a monophyletic origin for the nine SK strains rather than a converging evolution.
395 Indeed, the nine SK strains, despite being derived from different tissues (brain and inner ear),
396 and sharks (i.e., common thresher vs. salmon shark), collected in different years (from 2013 to
397 2016) and locations (throughout California) (S1 Table), are all highly similar. Thus, the
398 consistent presence of *C. maltaromaticum* in diseased sharks, and the proximity among the SK
399 strains (Fig 1B) further suggest that *C. maltaromaticum* SK strains are involved in periodic
400 thresher and salmon shark stranding events.
401 Bacterial pathogens have a range of possible classifications from environmental and commensal
402 organisms that occasionally cause infection to obligate pathogens not existing outside of their
403 host [63]. According to the SK genomic profiles and its consistent association with shark disease,
404 we can hypothesize this clade as opportunistic pathogen, as it likely maintains the capability to
405 survive and propagate while outside the host. This would benefit the spread of *C.*
406 *maltarmaticum*, as host-to-host transmission seems unfeasible in these species of shark, which do
407 not interact with each other apart from mating and during gestation periods.
408 While few sharks strand every year, the exact proportion of the populations affected by the
409 disease is unknown, as only those that make it to the beach are accessible. However, at least two
410 species of shark seem to be affected by *C. maltaromaticum* SK strains, thus suggesting other
411 sharks could possibly host and be affected by some of these strains. To further our knowledge on

412 the environmental source of this bacterium and its implication in the development of the
413 pathology, future studies should focus on determining the presence of SK strains in different
414 environmental settings and hosts. In this context, the identification of conserved SK-specific
415 genomic regions will help track SK-isolates in sequenced microbiomes. Additionally, using
416 those unique regions, SK-specific sets of primers will help identify SK strains in heathy/diseased
417 sharks and their environment, identify the geographic distribution, the host range, and the
418 mechanistic implication of SK strains in shark stranding.

419 # References

420 1. Schaffer P, Lifland B, Van Sommeran S, Casper D, Davis C. Meningoencephalitis associated with
421 *Carnobacterium maltaromaticum*–like bacteria in stranded juvenile salmon sharks (*Lamna ditropis*).
422 Veterinary Pathology Online. 2013;50(3):412-7.
423 2. Martinez-Steele L, Lowe CG, Okihiro MS, Berlemont R. Draft Genome Sequences of Nine New
424 Carnobacterium maltaromaticum Strains Isolated from Diseased Sharks. Genome announcements.
425 2018;6(18):e00354-18.
426 3. Carlisle AB, Litvin SY, Hazen EL, Madigan DJ, Goldman KJ, Lea RN, et al. Reconstructing habitat
427 use by juvenile salmon sharks links upwelling to strandings in the California Current. Marine Ecology
428 Progress Series. 2015;525:217-28. doi: 10.3354/meps11183. PubMed PMID: WOS:000354393700017.
429 4. Cartamil DP. Movement patterns, habitat preferences, and fisheries biology of the common
430 thresher shark (Alopias vulpinus) in the Southern California Bight. 2009.
431 5. Martin-Visscher LA, Yoganathan S, Sit CS, Lohans CT, Vederas JC. The activity of bacteriocins
432 from *Carnobacterium maltaromaticum* UAL307 against Gram-negative bacteria in combination with
433 EDTA treatment. FEMS microbiology letters. 2011;317(2):152-9.
434 6. Leisner JJ, Laursen BG, Prévost H, Drider D, Dalgaard P. *Carnobacterium*: positive and negative
435 effects in the environment and in foods. FEMS microbiology reviews. 2007;31(5):592-613.
436 7. Ringø E, Bendiksen H, Wesmajervi M, Olsen R, Jansen P, Mikkelsen H. Lactic acid bacteria
437 associated with the digestive tract of Atlantic salmon (*Salmo salar L.*). Journal of Applied Microbiology.
438 2000;89(2):317-22.
439 8. Kim DH, Austin B. Characterization of probiotic carnobacteria isolated from rainbow trout
440 *(Oncorhynchus mykiss)* intestine. Letters in applied microbiology. 2008;47(3):141-7.
441 9. Skrodenyte-Arbaciauskiene V, Sruoga A, Butkauskas D, Skrupskelis K. Phylogenetic analysis of
442 intestinal bacteria of freshwater salmon *Salmo salar* and sea trout *Salmo trutta trutta* and diet. Fisheries
443 Science. 2008;74(6):1307-14.
444 10. Al-Hisnawi A, Ringø E, Davies SJ, Waines P, Bradley G, Merrifield DL. First report on the
445 autochthonous gut microbiota of brown trout (*Salmo trutta Linnaeus*). Aquaculture Research. 2014.
446 11. Loch T, Kumar R, Xu W, Faisal M. *Carnobacterium maltaromaticum* infections in feral
447 *Oncorhynchus spp.* (Family Salmonidae) in Michigan. J Microbiol. 2011;49(5):703-13. doi:
448 10.1007/s12275-011-0527-1.
449 12. Kim D-H, Austin B. Innate immune responses in rainbow trout (Oncorhynchus mykiss, Walbaum)
450 induced by probiotics. Fish & shellfish immunology. 2006;21(5):513-24.

13. Gómez GD, Balcázar JL. A review on the interactions between gut microbiota and innate immunity of fish. FEMS Immunology & Medical Microbiology. 2008;52(2):145-54.
14. Loch TP, Xu W, Fitzgerald SM, Faisal M. Isolation of a *Carnobacterium maltaromaticum*-like bacterium from systemically infected lake whitefish (*Coregonus clupeaformis*). FEMS microbiology letters. 2008;288(1):76-84.
15. Hiu S, Holt R, Sriranganathan N, Seidler R, Fryer J. *Lactobacillus piscicola*, a new species from salmonid fish. International Journal of Systematic Bacteriology. 1984;34(4):393-400.
16. Iskandar CF, Borges F, Taminiau B, Daube G, Zagorec M, Remenant B, et al. Comparative Genomic Analysis Reveals Ecological Differentiation in the Genus Carnobacterium. Frontiers in Microbiology. 2017;8(357). doi: 10.3389/fmicb.2017.00357.
17. Georgiades K, Raoult D. Comparative Genomics Evidence That Only Protein Toxins are Tagging Bad Bugs. Frontiers in Cellular and Infection Microbiology. 2011;1(7). doi: 10.3389/fcimb.2011.00007.
18. Georgiades K, Raoult D. Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. PloS one. 2011;6(3):e17962.
19. Hacker J, Kaper JB. Pathogenicity islands and the evolution of microbes. Annual Reviews in Microbiology. 2000;54(1):641-79.
20. Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Molecular microbiology. 1997;23(6):1089-97.
21. Bishai W. Bacteriophage gene products that cause human disease. The bacteriophages. 1988;2:683-724.
22. Miao EA, Miller SI. Bacteriophages in the evolution of pathogen–host interactions. Proceedings of the National Academy of Sciences. 1999;96(17):9452-4.
23. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, et al. Origins of the E. coli strain causing an outbreak of hemolytic–uremic syndrome in Germany. New England Journal of Medicine. 2011;365(8):709-17.
24. Zomorodipour A, Andersson SG. Obligate intracellular parasites: Rickettsia prowazekii and Chlamydia trachomatis. FEBS letters. 1999;452(1-2):11-5.
25. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biology Direct. 2009;4(1):13. doi: 10.1186/1745-6150-4-13.
26. Moran NA. Microbial minimalism: genome reduction in bacterial pathogens. Cell. 2002;108(5):583-6.
27. Blanc G, Ogata H, Robert C, Audic S, Suhre K, Vestris G, et al. Reductive genome evolution from the mother of Rickettsia. PLoS Genetics. 2007;3(1):e14.
28. Andersson SG, Kurland CG. Reductive evolution of resident genomes. Trends in microbiology. 1998;6(7):263-8.
29. Muyzer G, Teske A, Wirsen CO, Jannasch HW. Phylogenetic relationships ofThiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. Archives of Microbiology. 1995;164(3):165-72.
30. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic acids research. 2013;42(D1):D206-D14.
31. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC genomics. 2008;9(1):75.
32. Iskandar CF, Cailliez-Grimal C, Rahman A, Rondags E, Remenant B, Zagorec M, et al. Genes associated to lactose metabolism illustrate the high diversity of Carnobacterium maltaromaticum. Food microbiology. 2016;58:79-86.

33. Hiu S, Holt R, Sriranganathan N, Seidler R, Fryer J. Lactobacillus piscicola, a new species from salmonid fish. International Journal of Systematic and Evolutionary Microbiology. 1984;34(4):393-400.
34. Miller III A, Morgan M, Libbey L. Lactobacillus maltaromicus, a New Species Producing a Malty Aroma1. International Journal of Systematic and Evolutionary Microbiology. 1974;24(3):346-54.
35. Sun Z, Harris HM, McCann A, Guo C, Argimón S, Zhang W, et al. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. Nature communications. 2015;6:8322.
36. Cailliez-Grimal C, Chaillou S, Anba-Mondoloni J, Loux V, Afzal MI, Rahman A, et al. Complete chromosome sequence of Carnobacterium maltaromaticum LMA 28. Genome announcements. 2013;1(1):e00115-12.
37. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research. 1994;22(22):4673-80.
38. Felsenstein J. {PHYLIP}(Phylogeny Inference Package) version 3.6 a3. 2002.
39. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research. 2016;44(W1):W242-W5.
40. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. 2015;32(6):929-31.
41. Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO. Application of tetranucleotide frequencies for the assignment of genomic fragments. Environmental microbiology. 2004;6(9):938-47.
42. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. International journal of systematic and evolutionary microbiology. 2007;57(1):81-91.
43. Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, et al. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. BMC bioinformatics. 2010;11(1):461.
44. Snel B, Bork P, Huynen MA. Genome phylogeny based on gene content. Nature genetics. 1999;21(1):108.
45. Eren AM, 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.
46. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research. 2000;28(1):33-6.
47. Van Dongen S, Abreu-Goodger C. Using MCL to extract clusters from networks. Bacterial Molecular Networks: Springer; 2012. p. 281-95.
48. Team R. RStudio: integrated development for R. RStudio, Inc, Boston, MA URL http://www rstudio com. 2015.
49. Snipen L, Liland KH. Micropan: An R-package for microbial pan-genomics. BMC bioinformatics. 2015;16(1):79.
50. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences. 2005;102(39):13950-5.
51. Snipen L, Almøy T, Ussery DW. Microbial comparative pan-genomics using binomial mixture models. BMC genomics. 2009;10(1):385.
52. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. Nucleic acids research. 2015;44(D1):D694-D7.
53. Bertelli C, Laird MR, Williams KP, Group SFURC, Lau BY, Hoad G, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic acids research. 2017;45(W1):W30-W5.

546 54. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of
547 the PHAST phage search tool. Nucleic acids research. 2016;44(W1):W16-W21.
548 55. Leimbach A. bac-genomics-scripts: Bovine E. coli mastitis comparative genomics edition.
549 Zenodo; 2016.
550 56. Alikhan N-F, Petty NK, Zakour NLB, Beatson SA. BLAST Ring Image Generator (BRIG): simple
551 prokaryote genome comparisons. BMC genomics. 2011;12(1):1.
552 57. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic
553 sequence analysis. BMC genomics. 2011;12(1):444.
554 58. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline
555 for faster genome publication. Bioinformatics. 2017;34(6):1037-9.
556 59. Afzal MI, Jacquet T, Delaunay S, Borges F, Millière J-B, Revol-Junelles A-M, et al. Carnobacterium
557 maltaromaticum: identification, isolation tools, ecology and technological aspects in dairy products.
558 Food microbiology. 2010;27(5):573-9.
559 60. Hammes WP, Hertel C. The genera lactobacillus and carnobacterium. The prokaryotes: Springer;
560 2006. p. 320-403.
561 61. Ramia NE, El Kheir SM, Taha S, Mangavel C, Revol-Junelles AM, Borges F. MultiLocus Sequence
562 Typing of Carnobacterium maltaromaticum strains associated with fish disease and dairy products.
563 Journal of applied microbiology. 2018.
564 62. Lande R. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and
565 genetic assimilation. Journal of evolutionary biology. 2009;22(7):1435-46.
566 63. Bentley SD, Parkhill J. Genomic perspectives on the evolution and spread of bacterial pathogens.
567 Proc R Soc B. 2015;282(1821):20150488.
568 64. Wei J, Goldberg M, Burland V, Venkatesan M, Deng W, Fournier G, et al. Complete genome
569 sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T. Infection and
570 immunity. 2003;71(5):2775-86.
571 65. Bliven KA, Maurelli AT. Antivirulence genes: insights into pathogen evolution through gene loss.
572 Infection and immunity. 2012:IAI. 00740-12.
573 66. Medina M, Sachs J. Symbiont genomics, our new tangled bank. Genomics. 2010;95(3):129-37.
574 67. Parkhill J, Dougan G, James K, Thomson N, Pickard D, Wain J, et al. Complete genome sequence
575 of a multiple drug resistant Salmonella enterica serovar Typhi CT18. Nature. 2001;413(6858):848.
576 68. Thomson NR, Howard S, Wren BW, Holden MT, Crossman L, Challis GL, et al. The complete
577 genome sequence and comparative genome analysis of the high pathogenicity Yersinia enterocolitica
578 strain 8081. PLoS genetics. 2006;2(12):e206.
579 69. Casadevall A. Evolution of intracellular pathogens. Annu Rev Microbiol. 2008;62:19-33.
580 70. Merikanto I, Laakso J, Kaitala V. Invasion Ability and Disease Dynamics of Environmentally
581 Growing Opportunistic Pathogens under Outside-Host Competition. PLoS ONE. 2014;9(11):e113436.
582 71. Anttila J, Ruokolainen L, Kaitala V, Laakso J. Loss of competition in the outside host environment
583 generates outbreaks of environmental opportunist pathogens. PLoS One. 2013;8(8):e71621.
584 72. Willis LM, Whitfield C. Structure, biosynthesis, and function of bacterial capsular polysaccharides
585 synthesized by ABC transporter-dependent pathways. Carbohydrate research. 2013;378:35-44.
586 73. Cuthbertson L, Kimber MS, Whitfield C. Substrate binding by a bacterial ABC transporter
587 involved in polysaccharide export. Proceedings of the National Academy of Sciences.
588 2007;104(49):19529-34.
589 74. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence
590 factors in host respiratory colonization and disease. Nature Reviews Microbiology. 2008;6(4):288.
591 75. Portolés M, Kiser KB, Bhasin N, Chan KH, Lee JC. Staphylococcus aureus Cap5O has UDP-ManNAc
592 dehydrogenase activity and is essential for capsule expression. Infection and immunity. 2001;69(2):917-
593 23.

76. Guimaraes BG, Souchon H, Honoré N, Saint-Joanis B, Brosch R, Shepard W, et al. Structure and mechanism of the alkyl hydroperoxidase AhpC, a key element of the Mycobacterium tuberculosis defense system against oxidative stress. Journal of Biological Chemistry. 2005.
77. Cosgrove K, Coutts G, Jonsson M, Tarkowski A, Kokai-Kun JF, Mond JJ, et al. Catalase (KatA) and alkyl hydroperoxide reductase (AhpC) have compensatory roles in peroxide stress resistance and are required for survival, persistence, and nasal colonization in Staphylococcus aureus. Journal of bacteriology. 2007;189(3):1025-35.
78. Baillon M-LA, Van Vliet AH, Ketley JM, Constantinidou C, Penn CW. An iron-regulated alkyl hydroperoxide reductase (AhpC) confers aerotolerance and oxidative stress resistance to the microaerophilic pathogen Campylobacter jejuni. Journal of Bacteriology. 1999;181(16):4798-804.
79. Lecuit M. Understanding how Listeria monocytogenes targets and crosses host barriers. Clinical Microbiology and Infection. 2005;11(6):430-6.
80. Merhej V, Georgiades K, Raoult D. Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors. Briefings in functional genomics. 2013;12(4):291-304.
81. Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW. Extensive error in the number of genes inferred from draft genome assemblies. PLoS computational biology. 2014;10(12):e1003998.
82. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Molecular biology and evolution. 2013;30(8):1987-97.
83. Nourdin-Galindo G, Sánchez P, Molina CF, Espinoza-Rojas DA, Oliver C, Ruiz P, et al. Comparative pan-genome analysis of Piscirickettsia salmonis reveals genomic divergences within genogroups. Frontiers in cellular and infection microbiology. 2017;7:459.
84. Kunttu HM, Valtonen ET, Jokinen EI, Suomalainen L-R. Saprophytism of a fish pathogen as a transmission strategy. Epidemics. 2009;1(2):96-100.
85. Ewald PW. Evolution of infectious disease: Oxford University Press on Demand; 1994.

Supporting information

S1 Fig. Complete-linkage clustering of 41 *Carnobacterium sp*. and 5 outgroups.
Clustering was performed from a Bray Curtis dissimilarity matrix according to FigFam annotation. *C. maltaromaticum* strains derive from diseased sharks (red), diseased trout (purple), food (blue) and human skin (orange).

S2 Fig. *C. maltaromaticum* pan-genome accumulation curve.
Curve shows the increment of unique gene clusters with the addition of new genomes to the pan-genome analysis. Blue area represents confidence intervals from standard deviation.

S3 Fig. Functional distribution of the core, accessory and unique genes of the *C. maltaromaticum* pan-genome.
Gene clustering and annotation are according to Anvi'o. Plot shows that high proportion of the core genome are house-keeping genes (like metabolism), while non-essential genes represent a higher proportion of the variable genome indicating difference in adaptations between strains.

S4 Fig. Scaled heat map of the genes categorized as virulence by the RAST annotation server in all *C. maltaromaticum*.

SK strains highlighted in bold show higher proportion of genes that represent invasion and resistance, perhaps indicating their tendency towards a more pathogenic lifestyle.

S5 Fig. Functional gene distribution of pseudogenes extracted from the SK genomes. Pseudogenes were predicted according to COG annotation using WebMGA software. High proportion of the genes are associated with metabolism, indicating the potential loss of metabolic functions that are no longer essential when pathogens adapt towards an intra-cellular lifestyle.

S1 Table. *C. maltaromaticum* strains isolated from diseased sharks. The column "Sharks" indicates the shark species and region the strain was isolated from. Numbers following shark species name used to discriminate between different sharks sampled. Column "Location" indicates where the shark was found stranded.

S2 Table. Strains used in the phylogenetic analysis.

S3 Table. Pan-genomic distribution of the *C. maltaromaticum* strains. Table shows total number and percentage of gene clusters forming the pan-genome, core genome, accessory genome and singletons. Number of clusters and relative abundance of annotated and unclassified genes are also shown along with the number of different functions in each category. The pan-genome of just the SK clade is shown separately with their gene annotation and functional distribution.

S4 Table. Unique COG functions to the SK clade. The column number of genomes represent presence of the function in just one or all SK genomes.

S5 Table. **Assigned functions to the genes found in the RODs.** Table shows genes in RODs unique to SK strains and unique to SK and trout pathogen genomes. Table reveals all genes with known function, while most of the genes were unclassified (not shown here).

S6 Table. Virulent associated genes found in all *C. maltaromaticum* genomes according to the VFDB.

S7 Table. Unique functions to the SK clade according to RAST FigFam-based annotation.

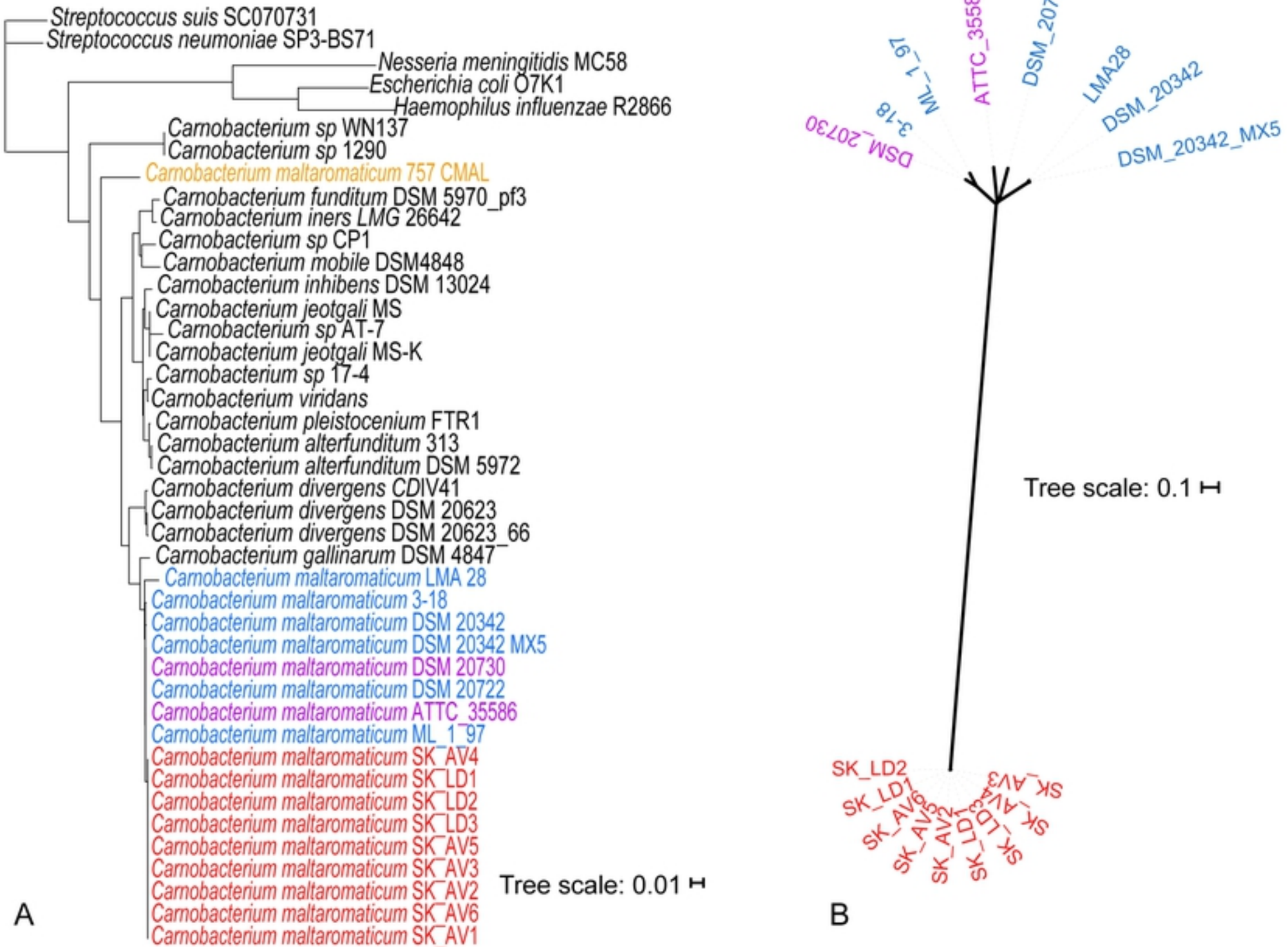
Streptococcus suis SC070731
Streptococcus neumoniae SP3-BS71
Nesseria meningitidis MC58
Escherichia coli O7K1
Haemophilus influenzae R2866
Carnobacterium sp WN137
Carnobacterium sp 1290
Carnobacterium maltaromaticum 757 CMAL
Carnobacterium funditum DSM 5970_pf3
Carnobacterium iners LMG 26642
Carnobacterium sp CP1
Carnobacterium mobile DSM4848
Carnobacterium inhibens DSM 13024
Carnobacterium jeotgali MS
Carnobacterium sp AT-7
Carnobacterium jeotgali MS-K
Carnobacterium sp 17-4
Carnobacterium viridans
Carnobacterium pleistocenium FTR1
Carnobacterium alterfunditum 313
Carnobacterium alterfunditum DSM 5972
Carnobacterium divergens CDIV41
Carnobacterium divergens DSM 20623
Carnobacterium divergens DSM 20623_66
Carnobacterium gallinarum DSM 4847
Carnobacterium maltaromaticum LMA 28
Carnobacterium maltaromaticum 3-18
Carnobacterium maltaromaticum DSM 20342
Carnobacterium maltaromaticum DSM 20342 MX5
Carnobacterium maltaromaticum DSM 20730
Carnobacterium maltaromaticum DSM 20722
Carnobacterium maltaromaticum ATTC_35586
Carnobacterium maltaromaticum ML_1_97
Carnobacterium maltaromaticum SK_AV4
Carnobacterium maltaromaticum SK_LD1
Carnobacterium maltaromaticum SK_LD2
Carnobacterium maltaromaticum SK_LD3
Carnobacterium maltaromaticum SK_AV5
Carnobacterium maltaromaticum SK_AV3
Carnobacterium maltaromaticum SK_AV2
Carnobacterium maltaromaticum SK_AV6
Carnobacterium maltaromaticum SK_AV1
Tree scale: 0.01
A
DSM_20730
3-18
ML_1_97
ATTC_35586
DSM_20722
LMA28
DSM_20342
DSM_20342_MX5
Tree scale: 0.1
SK_LD2
SK_LD1
SK_AV6
SK_AV5
SK_AV2
SK_LD3
SK_AV4
SK_AV3
B

Figure 1

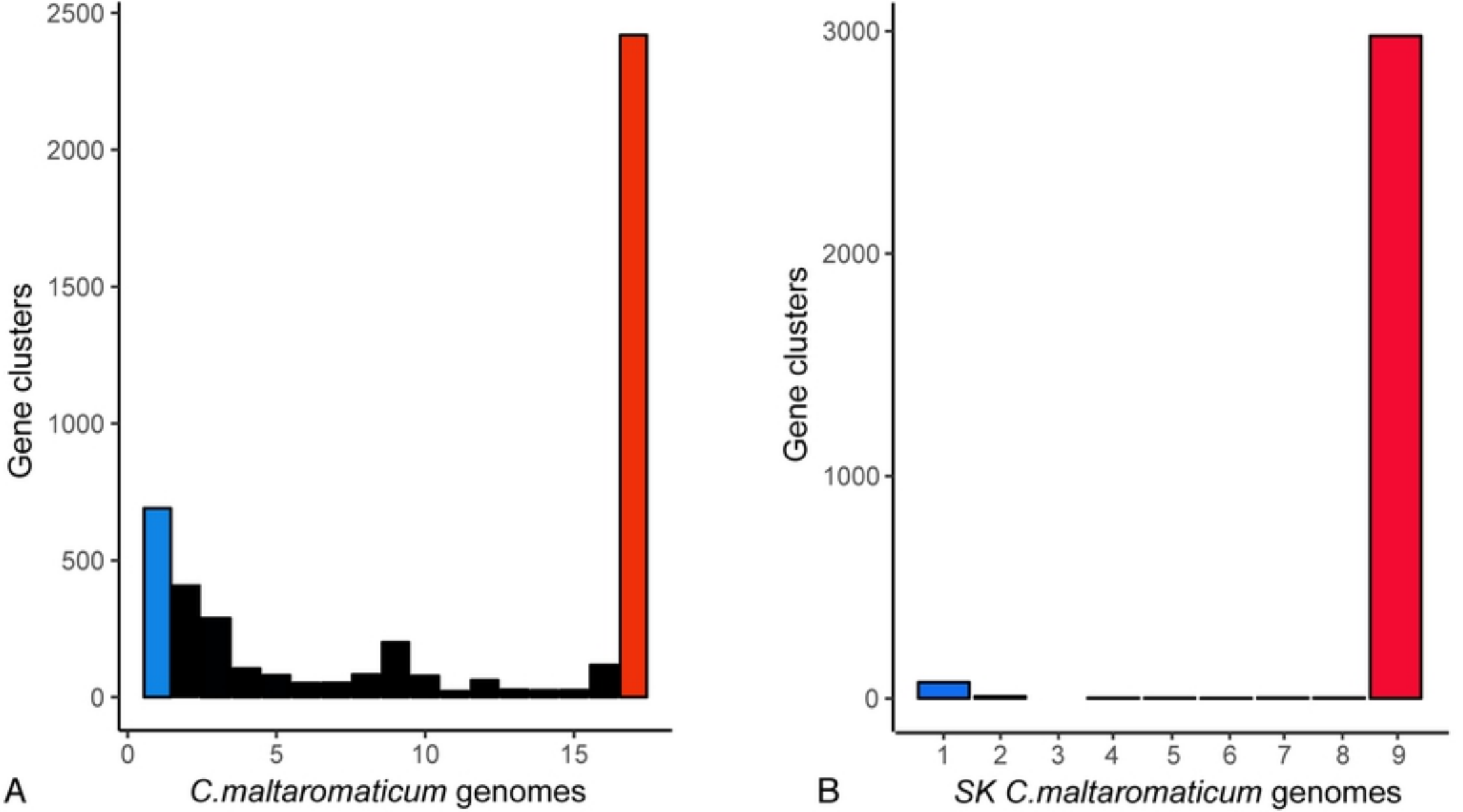
2500
2000
1500
1000
500
0
Gene clusters
0
5
10
15
A
C.maltaromaticum genomes
3000
2000
1000
0
Gene clusters
1
2
3
4
5
6
7
8
9
B
SK C.maltaromaticum genomes

Figure 2

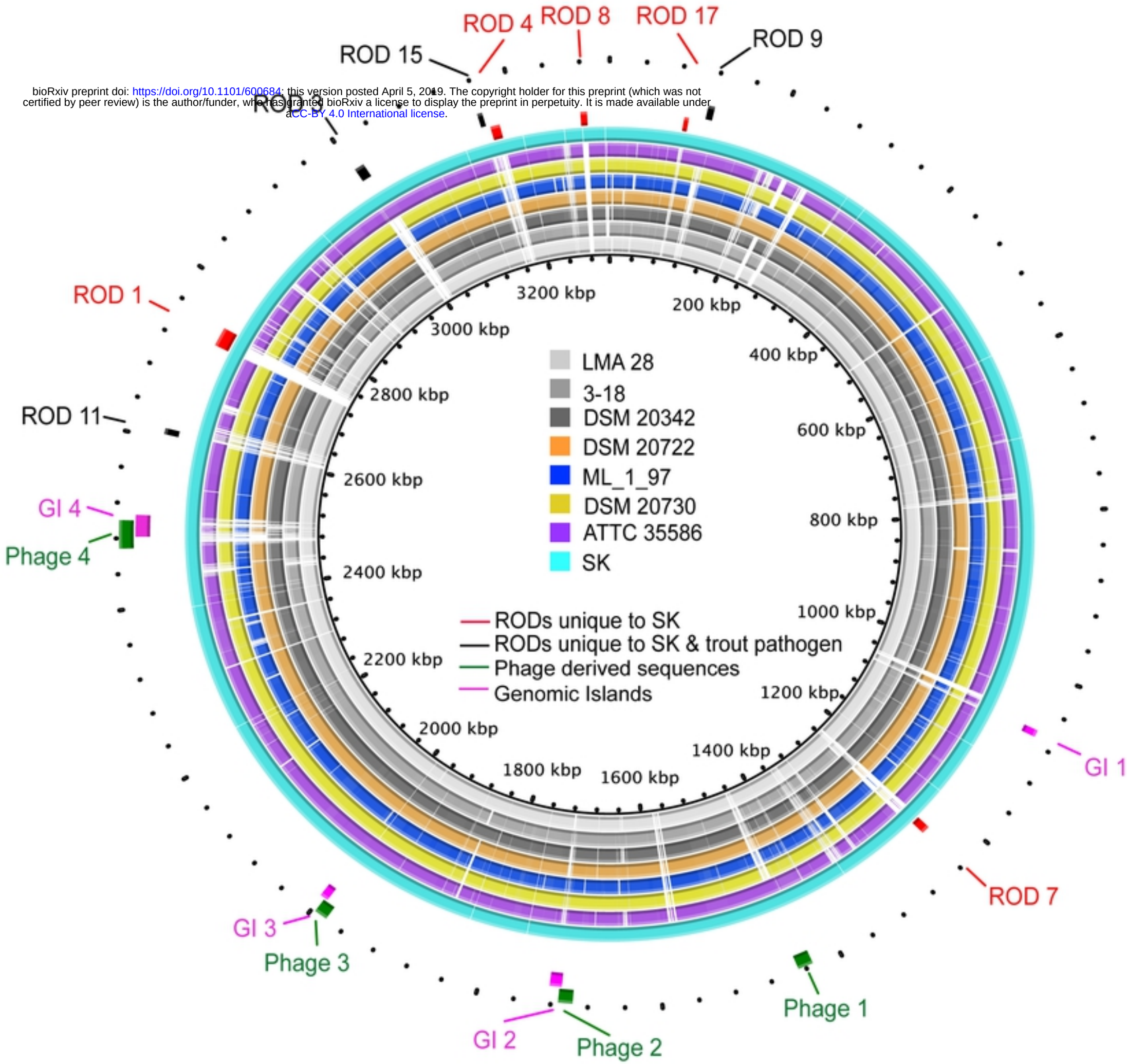

Figure 3

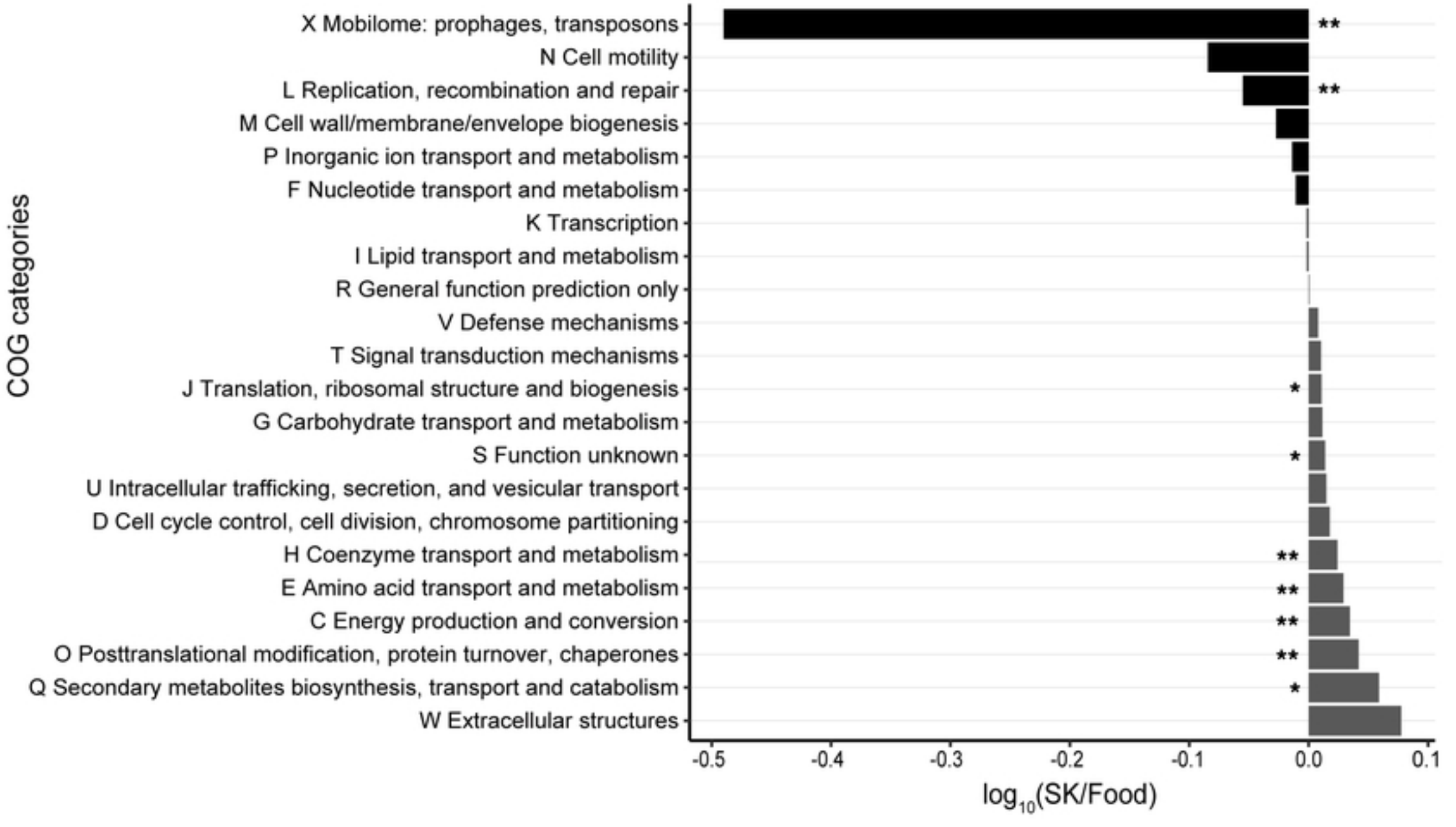
COG categories
X Mobilome: prophages, transposons
N Cell motility
L Replication, recombination and repair
M Cell wall/membrane/envelope biogenesis
P Inorganic ion transport and metabolism
F Nucleotide transport and metabolism
K Transcription
I Lipid transport and metabolism
R General function prediction only
V Defense mechanisms
T Signal transduction mechanisms
J Translation, ribosomal structure and biogenesis
G Carbohydrate transport and metabolism
S Function unknown
U Intracellular trafficking, secretion, and vesicular transport
D Cell cycle control, cell division, chromosome partitioning
H Coenzyme transport and metabolism
E Amino acid transport and metabolism
C Energy production and conversion
O Posttranslational modification, protein turnover, chaperones
Q Secondary metabolites biosynthesis, transport and catabolism
W Extracellular structures
**
**
*
*
**
**
**
**
*
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
log10(SK/Food)

Figure 4