

Comparative genomic analysis of Flavobacteriaceae: insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis.

Asimenia Gavriilidou (✉ asimenia.gavriilidou@wur.nl)

Wageningen Universiteit en Research <https://orcid.org/0000-0003-0204-2223>

Johanna Gutleben

Wageningen Universiteit en Research

Dennis Versluis

Wageningen Universiteit en Research <https://orcid.org/0000-0001-8430-4500>

Francesca Forgiarini

Wageningen Universiteit en Research

Mark WJ van Passel

Wageningen Universiteit en Research <https://orcid.org/0000-0001-6519-641X>

Colin J Ingham

Hoekmine BV <https://orcid.org/0000-0003-1508-019X>

Hauke Smidt

Wageningen Universiteit en Research <https://orcid.org/0000-0002-6138-5026>

Detmer Sipkema

Wageningen Universiteit en Research <https://orcid.org/0000-0003-3836-219X>

Research article

Keywords: Comparative genomics, Flavobacteriaceae, marine, host-associated, free-living

Posted Date: March 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-15750/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on August 20th, 2020. See the published version at <https://doi.org/10.1186/s12864-020-06971-7>.

Abstract

Background

Members of the bacterial family Flavobacteriaceae are widely distributed in the marine environment and often found associated with algae, fish, detritus or marine invertebrates. Yet, little is known about the characteristics that drive their ubiquity in diverse ecological niches. Here, we provide an overview of functional traits common to taxonomically diverse members of the family Flavobacteriaceae from different environmental sources, with a focus on the Marine clade. We include seven newly sequenced marine sponge-derived strains that were also tested for gliding motility and antimicrobial activity.

Results

Comparative genomics revealed that genome similarities appeared to be correlated to 16S rRNA gene phylogeny, while differences were mostly associated with nutrient acquisition, such as carbohydrate metabolism and gliding motility. The high frequency and diversity of genes encoding polymer-degrading enzymes support the capacity of marine Flavobacteriaceae to utilize diverse carbon sources. Homologs of gliding proteins were widespread among all studied Flavobacteriaceae in contrast to members of other phyla, highlighting the particular presence of this feature within the Bacteroidetes. Notably, not all gliding bacteria formed spreading colonies. Genome mining uncovered a diverse secondary metabolite biosynthesis arsenal of Flavobacteriaceae with high prevalence of gene clusters encoding pathways for the production of antimicrobial, antioxidant and cytotoxic compounds. Antimicrobial activity tests showed, however, that the phenotype differed from the genome-derived predictions for the seven tested strains.

Conclusions

Our study elucidates the functional repertoire of marine Flavobacteriaceae and highlights the need to combine genomic and experimental data while using the appropriate stimuli to unlock their uncharted metabolic potential.

Background

The family Flavobacteriaceae is the largest family of the Bacteroidetes phylum, and its members thrive in a wide variety of habitats. These Gram-negative, non-spore forming, rod-shaped, aerobic bacteria are commonly referred to as flavobacteria (1, 2). The taxonomy of members of the family Flavobacteriaceae is considered controversial (3), with many members that have been renamed and uprooted several times since the first classification (1, 4). For the past two decades, the family structure has been relatively stable, even though changes in the taxonomy are still occurring due to the large number of newly isolated flavobacteria (1). There has been a vast increase in the number of described genera within the Flavobacteriaceae from ten (3) to 158 (5) in the past 20 years. Due to this large number of members, the family has been divided into the following clades: Marine, Capnocytophaga, Flavobacterium, Tenacibaculum-Polaribacter and Chryseobacterium, based on 16S rRNA gene-based phylogenetic analysis (1). Recently the genus Chryseobacterium has been reclassified and moved into the family Weeksellaceae (6). Flavobacteriaceae are common in terrestrial and freshwater environments and in many cases numerically predominant in marine habitats (7). Within the Flavobacteriaceae, bacteria isolated from marine sources are widespread throughout all clades, but a large proportion belongs to the Marine clade (1). To date, marine flavobacteria have been found either free-living or attached to detritus in the water column (8, 9). Their lifestyle

includes colonization of the surface of algae (10), but also close association with invertebrate animals such as sponges (11), corals (12) and echinoderms (13).

A large number of marine flavobacteria are well known for their capability to degrade high molecular weight macromolecules, such as complex polysaccharides and proteins, contributing to the carbon turnover in marine environments (8, 10, 14). Earlier comparative genomics analyses showed that genomes of marine flavobacteria encode a relatively large number of both glycosyl hydrolases (GHs) and peptidases compared to other marine bacteria. Moreover, a similar number of peptidases was found as compared to other proteolytic specialists, suggesting the important role of flavobacteria in the degradation of both complex carbohydrates and proteins (15). The capacity of flavobacteria for utilization of macromolecules varies considerably, which is reflected by differences in a broad spectrum of enzymes known as carbohydrate-active enzymes (CAZymes) for the whole family. Genes that encode the protein machinery for polysaccharide binding, hydrolysis and transport are often encoded in distinct polysaccharide utilization loci (PULs) (16) that so far have been confined to the Bacteroidetes phylum. The first described PUL was the starch utilization system (Sus) of the human gut symbiont *Bacteroides thetaiotaomicron* (17). PULs in other saccharolytic Bacteroidetes encode proteins homologous to SusC and SusD. The SusC-like proteins act as TonB-dependent receptor transporters while the SusD-like proteins are carbohydrate-binding lipoproteins (18). Within PULs, genes encoding these homologs can be found in close proximity to genes coding for CAZymes, such as GHs, polysaccharide lyases (PLs), carbohydrate esterases (CEs) and carbohydrate binding modules (CBMs). So far, sequenced genomes of marine Flavobacteriaceae feature high proportions of CAZymes and PULs (8, 10, 14, 15, 19, 20), reinforcing their adaptations towards biopolymer degradation.

Besides their specialization as degraders of complex organic matter, many members of the Bacteroidetes, and particularly Flavobacteriaceae, share another distinct feature, known as 'gliding motility' (21). In general, most gliding species are members of the phyla Cyanobacteria, Proteobacteria and Bacteroidetes, however, each group utilizes different motility machineries (22, 23). Many flavobacterial cells exhibit rapid gliding movement by pivoting, flipping or crawling over solid surfaces without the aid of flagella or pili (21, 24). Gliding motility can be observed both microscopically by cells moving on glass slides, as well as on agar by colony spreading (1, 21). Cell movement is attributed to a unique gliding motor that is composed of several groups of proteins (Gld, Spr and Rem) and powered by the proton motive force (25–27). Recent studies revealed that this form of motility involves the rapid movement of adhesins delivered to the cell surface by a novel secretion system, the type 9 secretion system (T9SS), which is unrelated to other known bacterial secretion systems (21, 23, 25, 28). This system has been mostly studied in the motile aquatic or soil-derived *Flavobacterium johnsoniae* and the non-motile human oral pathogen *Porphyromonas gingivalis* (29). In *F. johnsoniae*, 27 proteins are involved in gliding motility, protein secretion, signalling and regulation (Additional File 3). Orthologs of the *F. johnsoniae* proteins involved in type 9 secretion are also found in *P. gingivalis* (30, 31). They are common within (21), but apparently limited to members of the Bacteroidetes (31). Nevertheless, the exact nature of the gliding motility mechanism and its relationship with T9SS remain under debate.

Gliding motility has previously been linked to the production of secondary metabolites likely due to their common purpose (predation/defence) (22). Several antibiotics (β -lactams, quinolones) have been previously isolated from Flavobacteriaceae strains (32–35), with some of them acting against recalcitrant targets such as methicillin-resistant *Staphylococcus aureus* (36, 37). Other bioactive molecules derived from flavobacteria include cell growth-promoting (38) and antitumor (39) compounds, von Willebrand factor receptor antagonists (40), protease (41) and topoisomerase (42) inhibitors, as well as antioxidative and neuroprotective myxols (43). Even though numerous intriguing molecules are flavobacterial products, the metabolic potential of the family has been poorly investigated.

Most of the microbially derived bioactive molecules belong to polyketides, non-ribosomal peptides, saccharides, alkaloids or terpenes and are related through common, highly conserved biosynthetic gene clusters (BGCs) (44, 45). Nowadays, computational detection of these BGCs and structural prediction of their products allow microbial genomes to be mined for metabolites (44). Further exploration of the secondary metabolism of marine flavobacteria might therefore unlock a vast resource of novel bioactive compounds.

Here, we performed a comparative genomics analysis to investigate characteristic biological features of the Flavobacteriaceae family, such as their complex carbohydrate metabolism and gliding motility mechanism (1). We furthermore addressed their functional dissimilarity to other dominant marine groups (Cyanobacteria and Proteobacteria) and differences between the distinct Flavobacteriaceae clades. Moreover, to discern traits that highlight niche-adaptation the same analysis was conducted, comparing host-associated and non-host associated flavobacteria. Of the many recently identified members of the Marine clade of the Flavobacteriaceae, only a few have been studied in detail beyond their isolation and initial physiological characterization. In this study, we obtained the individual genomes of seven presumably novel flavobacteria isolated from marine sponges (46, 47) and associated their genomic content with phenotypic features in terms of their gliding motility and antimicrobial activity. Finally, *in silico* genome mining for BGCs was performed to elucidate the secondary metabolic potential of bacteria belonging to dominant marine phyla (Cyanobacteria and Proteobacteria) and particularly, to members of the Flavobacteriaceae.

Results

Genome properties and phylogeny

To determine the genomic characteristics of the seven Flavobacteriaceae isolates recently obtained from the sponges *Aplysina aerophoba* and *Dysidea avara* and to allow for comparison with publicly available genomes from other members of the Flavobacteriaceae, their genomes were elucidated using Illumina sequencing. Genome sizes of the seven strains ranged from 3.7 to 5.3 Mbp with an average GC content of 38% (Table 1). The assembly generated genomes with 13–83 contigs and N50 ranging from 0.1 to 1.3 Mbp. Estimated completeness was more than 98%, and redundancy was lower than 2% for all assembled genomes. Coverage per base was above 200x for all draft genomes, except for DN50 for which it was 77x. Total gene count varied between 3317 and 4738, while the percentage of coding sequences (CDS) in all the assembled genomes was more than 98%. On average, more than 75% of the total genes could be assigned to protein families (Pfams).

A phylogenetic tree (Fig. 1) was built using the 16S rRNA gene sequences of the genomes included in the comparative genomics analysis. Based on their near complete 16S rRNA gene sequences, all seven newly sequenced strains were affiliated with the Marine clade of the family Flavobacteriaceae (Fig. 1). All strains isolated in this study were placed in separate branches of the corresponding tree (Fig. 1), except for Aa_D4 and Aa_C5 which were almost identical showing 99.6% similarity and 98% query cover of their 16S rRNA gene sequences. Among the newly sequenced isolates, two subclades within the Marine clade were represented, where DN112 and DN105 were phylogenetically adjacent, but distant from the rest of the isolates that grouped together. A BLASTN search against the NCBI database showed that both Aa_C5 and Aa_D4 were most closely related to strains of the genus *Eudoraea* (Additional File: Table S2). BLASTN returned *Flagellimonas* sp. as the closest hit to Aa_F7, *Arenibacter* sp. to Da_B9 and *Spongiimicrobium salis* to DN50. The strains DN105 and DN112 were found to be members of the *Gaetbulibacter* and *Lacinutrix* genera, respectively (Additional File: Table S2).

Functional profiling

Pfam profiles of 66 genomes representing three taxonomic groups (the family Flavobacteriaceae, and the phyla Cyanobacteria and Proteobacteria), all four clades of the Flavobacteriaceae (Marine, Capnocytophaga, Tenacibaculum-Polaribacter, Flavobacterium) and two different types of isolation sources (host-associated and non-host associated) were used as input to assess the correlation of taxonomy and life strategy with genome-derived functional traits. Flavobacteriaceae, Cyanobacteria and Proteobacteria had highly divergent functional profiles based on the relative abundances of Pfam annotations (PERMANOVA, $p = 0.001$) (Fig. 2a). Flavobacteriaceae and Cyanobacteria showed higher overall dissimilarity (57.4%) at the functional level, compared to Flavobacteriaceae and Proteobacteria (50%). Similarly, functional profiles were significantly different between the four different flavobacterial clades (PERMANOVA, $p = 0.001$) (Fig. 2b). In contrast, no significant differences were observed between host-associated and non-host associated flavobacteria at the functional level (PERMANOVA, $p > 0.05$) (data not shown).

From the 5,173 Pfams detected, 1,648 were present in all three groups (Flavobacteriaceae, Cyanobacteria and Proteobacteria), whereas 1,132 were specific to Flavobacteriaceae (Fig. 3a). The total number of predicted Pfams in the included genomes of the Flavobacteriaceae was 3,843 with 17% of them being specific to the Marine clade. Overall, a high degree of functional conservation, was observed at the Pfam-level among the different clades of the Flavobacteriaceae, with 48% of all annotated Pfams being shared by the four clades (Fig. 3b). All predicted Pfams that significantly contributed most to the dissimilarity between Flavobacteriaceae, Cyanobacteria and Proteobacteria (SIMPER analysis, $> 0.2\%$ contribution, $p < 0.05$) showed higher abundances in Flavobacteriaceae. Particularly, most functional attributes strongly selected for in flavobacteria were related to proteins involved in carbohydrate metabolism and transport (Table 2). These include a series of protein domains found in TonB-dependent receptors (pfam13715, pfam07715, pfam00593) and two-component regulatory systems (pfam00072, pfam08281, pfam04542, pfam04397). The majority of the genes coding for TonB-dependent receptors were found adjacent to genes that code for SusD/RagB family proteins (pfam07980), and which are restricted to the phylum Bacteroidetes. Moreover, among the most differentiating Pfam entries was the CHU_C family (pfam13585) that was found only in Flavobacteriaceae. Similarly, within the Flavobacteriaceae, the functional differences between the Marine, the Capnocytophaga and the Flavobacterium clade were mainly due to contribution of Pfam entries related to the attachment and degradation of polymeric compounds (Additional File 1: Table S3). There were no obvious functions distinguishing the Marine and the Tenacibaculum-Polaribacter clades, except for a gliding motility protein domain family (pfam13585) that was significantly more abundant in the Marine clade members.

Table 1
Genome properties and quality metrics of the strains sequenced in this study.

	Aa_C5	Aa_D4	Aa_F7	Da_B9	DN50	DN105	DN112
Size (Mbp)	3.7	3.7	4.2	4.3	5.3	4.8	4.7
Contigs	45	13	43	19	38	83	19
%GC	37	37	41	45	39	36	31
N50 (Mbp)	0.8	0.7	0.6	0.6	0.3	0.1	1.3
Per base coverage (x)	1331	357	298	248	77	207	261
Completeness (%)	98.7	98.7	99.3	99.7	99.7	99.7	99.3
Contamination	0.1	0.1	0.2	0.7	1.8	0.6	1.0
Total gene count	3317	3306	3776	4079	4738	4331	4069
CDS genes (%)	98.7	98.7	98.8	98.7	98.7	98.8	98.0
Number of 16S rRNA genes	1	1	2	1	3	1	1
Number of tRNA genes	36	36	37	45	50	40	63
Genes in Pfams (%)	80.9	80.9	77.8	74.8	70.2	73.2	77.8
Isolation source	Aplysina aerophoba	Aplysina aerophoba	Aplysina aerophoba	Dysidea avara	Aplysina aerophoba	Aplysina aerophoba	Aplysina aerophoba

Table 2
Pfams contributing most significantly ($> 0.2\%$, $p < 0.05$) to differences across major taxonomic groups.

Function name (Pfam ID)	Relative Abundance (%)			Contribution (%)		p-values (< 0.05)	
	F	C	P	F-C	F-P	F-C	F-P
CarboxypepD_reg-like domain (pfam13715)	1.23	0	0.004	0.97	0.97	0.001	0.001
TonB-dependent Receptor Plug Domain (pfam07715)	1.13	0.01	0.44	0.77	0.76	0.001	0.016
TonB dependent receptor (pfam00593)	0.55	0.01	0.29	0.47	0.42	0.001	0.04
Response regulator receiver domain (pfam00072)	1.09	0.85	0.95	0.41	0.49	0.014	0.015
Sigma-70, region 4 (pfam08281)	0.51	0.04	0.11	0.41	0.37	0.001	0.002
LytTr DNA-binding domain (pfam04397)	0.40	0	0.06	0.35	0.34	0.005	0.029
Sigma-70 region 2 (pfam04542)	0.63	0.24	0.21	0.33	0.40	0.004	0.001
C-terminal domain of CHU protein family (pfam13585)	0.38	0	0	0.31	0.35	0.001	0.001

F, Flavobacteriaceae; C, Cyanobacteria; P, Proteobacteria.

Carbohydrate metabolism and transport

The ability to degrade and transform complex carbohydrates was assessed by mining the genomic data for carbohydrate-active enzymes (CAZymes). Comparison of the CAZyme repertoire (average number of CAZymes per Mbp) across the different groups showed significantly higher frequencies of glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and carbohydrate binding modules (CBMs) in Flavobacteriaceae (Kruskal-Wallis test, $p < 0.001$), whereas glycosyl transferases (GTs) were more frequent in Cyanobacteria (Kruskal-Wallis test, $p < 0.05$) (Fig. 4a). No statistically significant differences were observed in the frequency of CAZymes between the different Flavobacteriaceae clades (Fig. 4b). In the Marine clade, GH families 74 (GH74) and 109 (GH109) were the most abundant families (Additional File 4). GH74 comprises many xyloglucan-hydrolysing enzymes, while enzymes belonging to GH109 have α -N-acetylgalactosaminidase activity (48). In the case of CBMs in the Marine clade, most hits belonged to the chitin- or peptidoglycan-binding family (CBM50) and cellulose- and xyloglucan-binding family (CBM44) (48).

Gliding Motility And Type 9 Secretion

Based on the studies of Veith et al. (2017) and McBride (2019), we defined a set of 27 proteins that are related to gliding motility and T9SS, e.g. as reported for *F. johnsoniae* and *P. gingivalis*. Moreover, these above mentioned proteins were categorized as i) essential, ii) nearly essential, when their absence resulted in defective but not complete loss of motility and secretion, and iii) non-essential, according to previous reports on *F. johnsoniae* (23, 30) (Additional File 3). Forty-nine publicly available genomes (Additional File 2) of gliding and non-gliding members of the different clades of the Flavobacteriaceae (Marine, Capnocytophaga, Flavobacterium, Tenacibaculum-Polaribacter) (Fig. 5) isolated from different environments (host-associated and non-host associated) were

screened for the presence of homologs to these 27 proteins, using amino acid sequences. All 29 flavobacteria reported in literature as 'gliding' had homologs for 18 of the motility and T9SS-related proteins (GldA, GldB, GldD, GldH, GldI, GldJ, GldK, GldL, GldM, GldN, SprA, SprE, SprF, SprT, PorV, SigP, PorX and PorY), except for *Cellulophaga tyrosinoxydans* which was lacking GldF and GldG (Fig. 5). The 18 proteins encoded by the genomes of all other gliding flavobacteria include all the Gld and Spr proteins that were considered essential or nearly essential for motility in *F. johnsoniae* (Additional File 3) (21). Homologs to the motility adhesins RemA were not detected in five out of the 29 genomes of gliding members of the Bacteroidetes phylum, while SprB homologs were present in all gliding bacteria. In our dataset, there were two strains for which no information regarding motility was available in literature (*Winogradskyella jejuensis* CP32^T and *Winogradskyella* sp. J14-2). Here, we report that both strains have a complete set of homologs to the T9SS-based gliding proteins. For the non-gliding *P. gingivalis*, homologs for the major T9SS components (GldK, GldL, GldM, GldN, SprA, SprE, SprT) were detected in the genome (Fig. 5). Four proteins (GldF, GldG, GldD, GldI) that are required for gliding motility of *F. johnsoniae* were not detected in *P. gingivalis*. The same was the case for the motility adhesin RemA and the SprB supporting proteins, SprC and SprD. No obvious differences in the presence of gliding and T9SS proteins were found among genomes of host-associated and non-host associated strains (data not shown). Similarly, no specific pattern was observed in terms of the gliding motility of the different Flavobacteriaceae clades (Fig. 5).

Apart from the Bacteroidetes, proteins encoded by genomes from two other phyla (Cyanobacteria and Proteobacteria) were searched by BLASTP (E-value 1e-5) for the presence of homologs of gliding and T9SS-related proteins (Fig. 5). In general, homologs to *F. johnsoniae* and *P. gingivalis* motility proteins were scarce among these two phyla. From the gliding motility proteins, only GldA homologs were present in all cyanobacterial and proteobacterial genomes. *Anabaena* sp. and *Trichodesmium erythraeum* strains were the only cyanobacteria in our dataset previously reported as gliding (49, 50), and both genomes were found to encode homologs for five gliding proteins (GldA, GldF, GldG, GldJ, GldK) and three regulatory proteins (SigP, PorX and PorY). The *Anabaena* sp. genome also encoded one Spr ortholog (SprE) that was absent from *T. erythraeum*. Among the genomes tested in this study, those of *Prochlorococcus marinus* and "Candidatus Pelagibacter ubique" encoded the lowest number (three) of homologs to the motility and T9SS proteins (Fig. 5).

In addition to examining the genomes of the seven flavobacterial sponge isolates, their gliding motility was also tested based on the formation of spreading colonies on agar. All newly sequenced genomes encode homologs for each of the proteins considered essential for gliding motility and type 9 secretion in *F. johnsoniae* (Additional File 3). However, after examination of the colony morphology, most of the isolated strains (DN105, Da_B9, Aa_C5, Aa_D4, DN112) were non-motile on agar, and only Aa_F7 and DN50 formed spreading colonies. The edges of the colonies of both Aa_F7 and DN50 formed slender peninsulas spreading on 1% (w/v) agar plates (Fig. 6a, b). On 1.8% (w/v) agar plates, DN50 exhibited a colony surface contour pattern probably attributed to gliding motility and secondary colony formation (Fig. 6c). This pattern was not observed for any of the other strains.

Homologs for all gliding and T9SS proteins were identified in the genomes of the spreading strains (Aa_F7 and DN50), except for PorQ and PorU. Consequently, the chromosomal neighbourhood of these genes was investigated for the genomes of Aa_F7 and DN50 and compared to that of *F. johnsoniae* UW101 (Fig. 7). In the *F. johnsoniae* genome, porU (Fjoh_1556) lies immediately upstream of porV (Fjoh_1555) and adjacent to gldJ (Fjoh_1557), being encoded on the opposite strand. For Aa_F7 and DN50, the gene arrangement was similar to that found in *F. johnsoniae* with a gene coding for a ligase downstream of gldJ. No homolog to porU was identified in the genomes of Aa_F7 and DN50, and the genes directly adjacent to gldJ were annotated as hypothetical proteins (Fig. 7).

Sequence alignments of these hypothetical proteins and *F. johnsoniae* PorU by BLASTP showed amino acid identities of 24.4% and 26.3% and a query cover of 8% and 14% for Aa_F7 and DN50, respectively. In the case of porQ, it was not possible to use the gene neighbourhood because it is not adjacent to any of the known gliding or T9SS-related genes in *F. johnsoniae*. Instead, a BLASTP search with a lower cut-off (E-value 1e-0) revealed weak hits for PorQ in both Aa_F7 and DN50. Interestingly, in the genome of DN105 all components of *F. johnsoniae* and *P. gingivalis* gliding and T9SS machinery were identified, even though it did not form spreading colonies on agar under the examined conditions (Fig. 6d). Similarly, *Bizionia echini*, *Lacinutrix venerupis* and *Tamlana sedimentorum* were previously reported as non-motile, but they carry a complete set of orthologous gliding genes in their genomes (Fig. 5).

Secondary Metabolome

The genomic repertoire for secondary metabolite synthesis of the 66 studied bacteria was assessed using the 'Antibiotics and secondary metabolite analysis shell' (antiSMASH) (51). In Flavobacteriaceae, the vast majority of BGCs was predicted to encode terpene synthases (44%), followed by type III polyketide synthases (T3PKS) (9%), non-ribosomal peptide synthetases (NRPS) (7%), lanthipeptide (6%) and aryl polyene (6%) clusters. In total, the percentage of genes putatively involved in the biosynthesis of secondary metabolites in relation to the total number of protein-coding genes in each genome ranged between 0.02 and 0.3%. Cyanobacteria showed the highest average number of annotated BGCs ($n = 7$), compared to Flavobacteriaceae ($n = 4$) and Proteobacteria ($n = 4$) (Additional File 1: Table S4). Gene clusters encoding the proteins involved in the biosynthesis of non-ribosomal peptides and polyketides were identified in genomes from all three major groups (Additional File 1: Table S5). Terpene BGCs were detected in all analysed genomes and showed the highest relative abundance in Flavobacteriaceae and Cyanobacteria. In contrast, the most abundant gene cluster in Proteobacteria was affiliated with the biosynthesis of homoserine lactones (HSL), a quorum sensing (QS) molecule in Gram-negative bacteria (Table 3). Interestingly, only one flavobacterial strain (*Muriicola jejuensis* DSM 21206) harboured HSL BGCs.

Only BGC types with mean abundance of more than 0.1 BGCs/strain in the complete dataset are shown. A more detailed table of the BGC distribution can be found in Additional File 1: Table S5. M, Marine; C, Capnocytophaga; F, Flavobacterium; T, Tenacibaculum-Polaribacter.

Within the Flavobacteriaceae, strains from the Marine and Flavobacterium clades contained the largest average number of BGCs (Additional File 1: Table S4), whereas on average only 2 BGCs were detected in the Tenacibaculum-Polaribacter genomes. In general, the different flavobacterial clades exhibited a similar composition in terms of the types of predicted BGCs being enriched in their genomes, including terpene, lanthipeptide, NRPS and aryl polyene BGCs.

Genome mining for BGCs of the seven newly isolated flavobacteria resulted in the detection of five BGCs, on average, belonging to six different classes based on the antiSMASH database. In total, 63% of the identified BGCs were encoding pathways for the production of terpenes, lanthipeptides and aryl polyenes. Less abundant but widely distributed across these genomes were bacteriocin, T3PKS and NRPS BGCs. Strain DN50 harboured the largest number of different BGC types and the highest absolute BGC abundance (Fig. 8). A more detailed comparative analysis of the BGCs showed that the necessary proteins predicted to synthesize carotenoid pigments were present in all seven flavobacterial genomes. Similarly, DN50 harboured a BGC related to the production of flexirubin, another type of bacterial pigment (43). Notably, strain DN112 was predicted to possess two BGCs with 8%

homology to gene clusters involved in the biosynthesis of the antitumor agents Ecteinascidin-743 (ET-743) (52) and oocydin A (53). To accompany the genome mining results, antimicrobial activity of the seven sponge-associated flavobacterial strains was tested against six indicator strains (*Escherichia coli*, *Bacillus subtilis*, *Staphylococcus simulans*, *Aeromonas salmonicida*, *Candida oleophila*, *Saprolegnia parasitica*) belonging to the three categories Gram-positive, Gram-negative bacteria and fungi-oomycetes. Even though the genome predictions revealed a variety of secondary metabolite BGCs, none of the studied isolates showed antimicrobial activity under the examined conditions (data not shown).

Discussion

The family Flavobacteriaceae in the phylum Bacteroidetes currently contains more than 150 described genera (5) that are widespread in marine and non-marine ecosystems (7). Thriving in such diverse habitats could indicate a high genetic plasticity. Disentangling the phylogeny of the family was controversial and challenging in the past, yet it has remained relatively stable for the past two decades (1). In this study, we followed the definition of flavobacterial clades as proposed by McBride (2014) with a focus on the Marine clade. We compared representative genomes of phylogenetically diverse flavobacteria inhabiting various niches to reveal novel insights into the characteristic properties of this group.

The functional profiling based on Pfam entries revealed a high degree of distinction between the Flavobacteriaceae in comparison with other taxa dominant in marine environments, i.e. Proteobacteria and Cyanobacteria. This was also true for the genomic content of flavobacteria from different clades. However, comparison of functional profiles of host-associated and non-host-associated flavobacteria did not show significant dissimilarity. These results indicate that the functional traits within the Flavobacteriaceae were congruent with the phylogeny (54, 55), rather than the habitat type, i.e. host-associated or free-living (55). In marine environments, a considerable number of host-associations might be coincidental as many microbes show metabolic versatility of a biphasic lifestyle-strategy that includes seawater particles ("free-living") and animal hosts (56–59), most likely in the search of nutrient sources (59). Previous studies provided evidence with respect to the phylogenetic conservation of traits associated with organic substrate utilization in different microorganisms (60, 61) and particularly in members of the Flavobacteriaceae (54, 55). Interestingly, functional traits appeared to be highly conserved within the different flavobacterial clades (48% shared Pfams), even though the isolation sources of the strains were largely diverse. The key functions strongly enriched in Flavobacteriaceae, in general, and also in the Marine clade, were related to nutrient acquisition in regard to complex carbohydrate metabolism as well as gliding motility.

These two features have been repeatedly linked to the important role of Flavobacteriaceae in the degradation of high molecular weight organic matter within the marine environment (7, 15). Investigation of the repertoire of polymer-degrading enzymes among dominant marine bacterial groups revealed that Flavobacteriaceae had significantly more GHs, PLs, CEs and CBMs per Mbp compared to Cyanobacteria and Proteobacteria, supporting their pronounced capacity as polymer degraders and perhaps as key players in ocean carbon cycling (8, 15, 19, 54, 55). Within the Marine flavobacterial clade, the GH74 family occurs with the highest frequency and contains various enzymes that hydrolyse β-1,4 linkages of glucans. Thus, these enzymes assist in the degradation of different oligo- and polysaccharides, including xyloglucans (48), which is a hemicellulose polysaccharide present in plant cell walls and green algae (62). GH109, predicted as α-N-acetylgalactosaminidase, was the second most frequent GH family in the Marine clade genomes. These enzymes cleave N-acetylgalactosamine residues from various substrates such as glycolipids, glycopeptides and glycoproteins. It is highly likely that these substrates are derived from dissolved and/or particulate organic matter found in seawater, but also from the matrix of sponges

(63, 64). In addition, high frequencies of CBM50 and CBM44 genes were observed in the Marine clade. CBM50 proteins are found attached to enzymes that might cleave either bacterial peptidoglycan or animal chitin, and CBM44 are modules related to enzymes that bind both cellulose and xyloglucan (8, 64). Taken together, our results reflect that the Marine clade flavobacteria have the genetic repertoire for utilizing a large diversity of carbon sources derived from algae, plants, bacteria and animals. This feature common in these microbes derived from numerous source habitats underlines the essential role of substrate utilization in the colonization of both host-associated and free-living niches.

Many members of the Bacteroidetes and particularly Flavobacteriaceae can glide rapidly over surfaces using two novel and intertwined machines, one gliding motor involved in cell movement and one protein secretion system, known as T9SS (21, 23, 25, 31). Gliding is common among members of the Flavobacteriaceae (65). This notion could be reinforced by the fact that genomes of all investigated gliding flavobacteria encoded homologs of the necessary components for T9SS-based gliding, with the exception of *Cellulophaga tyrosinoxydans*. This bacterium was lacking the two proteins GldF and GldG, which together with GldA are thought to be the components of an ATP-binding cassette (ABC) transporter involved in flavobacterial gliding (66). According to McBride and Zhu (2013), two other flavobacteria, *Cellulophaga algicola* and *Maribacter* sp. HTCC2170 were also actively gliding even though their genomes were lacking the genes encoding this ABC transporter (21). This might suggest either the involvement of another ABC transporter, common in most organisms, or a non-essential role of this ABC transporter in Bacteroidetes gliding motility (21, 23, 67). The T9SS exports the motility adhesins RemA and SprB which are propelled along the cell surface by the gliding motor, resulting in cell movement of *F. johnsoniae* (21, 25, 68). SprB homologs were ubiquitous in gliding flavobacteria, whereas RemA was rare. This might indicate that only SprB is essential for Bacteroidetes gliding, but might also suggest the presence of different currently unknown motility adhesins that allow other species to move over diverse surfaces (25, 67).

Gliding motility is considered as an efficient strategy to enhance survival, but the actual role in nature is still unknown. It has been previously suggested that gliding over surfaces facilitates access to nutrient sources (22), or is involved in pathogenesis (28) or symbiosis (69). Accordingly, in both free-living and host-associated bacteria, the importance of T9SS has been demonstrated for many processes such as polymer degradation, adhesion, motility and virulence (23). In this study, no specific link was found between host-association and gliding or type 9 secretion. Similarly, no difference was found when comparing the genomic repertoire of members of different taxonomic clades within the Flavobacteriaceae. These observations add to the fact that flavobacteria appear to be 'generalists', rather than 'specialists' and this might explain why they thrive in such diverse niches. Especially in the marine environment, the availability of nutrients might drive opportunistic host-microbe interactions, which might explain the presence of flavobacteria in the biomes of numerous marine hosts.

The high degree of conservation of Gld and Spr proteins in all Flavobacteriaceae, irrespective of their environmental source or taxonomic clade, supports the notion that T9SS-based gliding motility is widespread among Bacteroidetes members (21) and especially among Flavobacteriaceae. Gliding motility occurs in other bacteria apart from the phylum Bacteroidetes (myxobacteria, cyanobacteria, mycoplasmas, etc.) but these employ their own unique machineries involving secreted polysaccharides, type IV pili (T4P) or flagella movement (70–73). Filamentous cyanobacteria such as *Anabaena* sp. and *Trichodesmium erythraeum* are thought to share a non-T9SS based gliding mechanism that involves a T4P-like nanomotor and polysaccharide secretion that drive and facilitate locomotion (70, 71). Nevertheless, homologs to certain core Bacteroidetes gliding proteins (GldA, GldF, GldG, GldI, GldJ, GldK, SprE) were found to be encoded in genomes belonging to Cyanobacteria or Proteobacteria, regardless of their gliding status or mechanism. For example, GldA, GldF and GldG are similar to many putative

components of the ABC transporter system, which is ubiquitous in all microorganisms (21, 31). These proteins might serve another purpose in the respective organisms that might not be linked to gliding motility. In general, the paucity of homologs to Bacteroidetes gliding and T9SS proteins among bacteria from other phyla implies the presence of diverse and likely unrelated gliding motility mechanisms (21, 23, 74).

Gliding motility can also result in colony spreading, observed on agar plates as flat, irregularly edged colonies (1, 21, 25). The majority of the strains investigated for colony spreading (5 out of 7) were non-motile on agar under the tested conditions, even though presence of all genes encoding proteins involved in the gliding motility apparatus was predicted by the genomic analysis. Presence of a gene in a genome does not yet imply gene expression. As described above, several bacterial strains that were previously described as 'non-gliding' had all essential genes present in their genomes. For example, *P. gingivalis* is a non-motile bacterium but all T9SS components and most of the gliding proteins were found to be encoded in its genome (Fig. 5).

Moreover, previous findings support that formation of non-spreading colonies is not necessarily an indication of absence of gliding motility. For example, *Maribacter* sp. HTCC2170 could glide well on glass but its colonies did not spread on agar (21). In addition, it should be highlighted that the growth media used in our colony spreading tests were high in nutrients. Spreading under low nutrient conditions is more effective because non-metabolizable sugars tend to suppress the active cell movement on agar media (1, 75, 76). Thus, these bacteria may be motile in other conditions that were not examined here. On the other hand, the strains Aa_F7 and DN50, both carrying homologs for all T9SS-gliding proteins in their genomes, formed spreading colonies on agar displaying characteristic edges and a surface pattern indicative of secondary colony formation (Fig. 6). In contrast to the other strains, it is possible that these isolates could metabolize all sugars present in the high-nutrient media, thus surpassing the carbohydrate-mediated inhibitory effect on colony spreading. The absence of homologs to PorQ and PorU from Aa_F7 and DN50 genomes did not affect their gliding motility. Both PorQ and PorU are involved in the T9SS substrate attachment to the cell surface in *P. gingivalis* (77, 78) but in *F. johnsoniae* their deletion did not affect gliding motility (79). This might be another indication that PorQ and PorU are not essential for full motility (23, 30).

A wide variety of bioactive molecules, such as antibiotics, cell growth-promoting, antioxidative and neuroprotective compounds, have been previously isolated from members of the Flavobacteriaceae (1). Nevertheless, the metabolic potential of flavobacteria in terms of bioactivity has been underreported in literature. Across taxonomic groups, terpene BGCs were prevalent in the predicted metabolite arsenal of both Cyanobacteria and Flavobacteriaceae, particularly of the Marine clade. Terpenes are the largest class of small-molecule natural products, found in almost all life forms and performing diverse functions (80). Previous findings showed that terpene synthases are widely distributed in bacteria, the majority of which are Gram-positive (Actinomycetales) but also in Gram-negative bacteria, such as Cyanobacteria and Flavobacteriales (81). In terms of bioactivity, to date there are only a few reports on terpenes of microbial origin exhibiting antimicrobial (82–84) and antioxidant activity (43).

The almost complete absence of HSL BGCs from the Flavobacteriaceae genomes, compared to their high relative abundance in Proteobacteria (Table 3), is in accordance with previous studies on flavobacteria that suggest the existence of alternative QS signalling molecules for Bacteroidetes (85, 86). For example, strains of *Zobellia galactanivorans* (member of Flavobacteriaceae) were found to synthesize another type of QS signalling molecules called dialkylresorcinols with antibiotic and antioxidant activity (86). In addition, the detection of a HSL-degrading enzyme in the genome of *Z. galactanivorans* (14) implies the presence of a communication system distinct from the HSL-based QS system that could function as a defence mechanism with antibiotic effect (86).

Within Flavobacteriaceae, the secondary metabolite repertoire did not appear to be influenced by phylogeny, as the global composition of predicted BGCs was similar among the clades. Besides terpene synthases, BGCs responsible for the biosynthesis of lanthipeptides were also found in high abundance in flavobacterial genomes. Formerly known as lantibiotics, they are ribosomally synthesized post-translationally modified peptides that belong to class I bacteriocins (87). It has been previously suggested that bacteriocins play a critical role in mediating microbial community dynamics (88, 89). Presumably, bacteriocin-producing flavobacteria act as anti-competitors, facilitating the invasion of other closely related species into an established microbial community to successfully colonize a niche. An additional role suggested for bacteriocins is their involvement in host defence, protecting the host from pathogens (89).

Genome mining of the BGC arsenal of the sponge-associated strains sequenced in this study revealed homologs to carotenoid-producing BGCs in all seven genomes. Carotenoids are often the most abundant pigments in marine flavobacteria, and their importance lies in their strong antioxidant properties. In flavobacteria, a few structurally rare carotenoids have been identified before, such as saproxanthin and myxol that show significant antioxidative activities and neuroprotective effects (43). Flexirubins represent another type of pigment common in Bacteroidetes but not in other bacteria (1). A BGC similar to one shown to encode proteins needed for flexirubin production was identified in strain DN50. Interestingly, such pigments exhibit anticancer and antimicrobial properties (e.g. against *Mycobacterium* sp.) and are considered promising candidates for treatment and prevention of cancer and microbial infections (90, 91). Furthermore, one BGC homologous to the ET-743 BGC was identified in strain DN112. ET-743 is a cytotoxic compound synthesized by a bacterial endosymbiont that belongs to Gammaproteobacteria and the complete genome of which was recently reconstructed from the metagenome of its mangrove tunicate host (52). Subsequent analysis showed that ET-743 was the only natural product BGC found within the genome, highlighting its ecological value to the host (52). Thus, in the case of bacteria associated with sponge hosts, the cytotoxic activity of ET-743 could serve as a chemical deterrent of predators.

Even though the genome predictions revealed a large number and variety of secondary metabolite BGCs, none of the studied isolates showed antimicrobial activity under the conditions we examined. The majority of these biosynthetic loci are frequently dormant or expressed at low constitutive levels under laboratory conditions, keeping the true biosynthetic potential of microorganisms hidden and thus hampering the discovery of novel bioactive compounds (92, 93). This 'silent' state can be reversed by inducers of gene expression, such as environmental cues, nutrients or signal molecules (92, 93). Unlocking the full metabolic potential encoded by the studied strains might require high-throughput screening of various growth conditions in combination with a large number of indicator strains.

Conclusions

The comparative genomics analysis performed in this study demonstrated that 16S rRNA gene-based phylogeny, rather than life strategy (host-associated or free-living) of the organisms is the main factor correlated to the functional profile of Flavobacteriaceae. The traits responsible for the functional divergence between phyla investigated here were found to be associated with gliding motility and nutrient acquisition through the catabolism of carbohydrates. Marine flavobacteria appear to be potent utilizers of a large variety of carbon polymers from algae, bacteria, plants and animals, confirming their role in the ocean carbon cycling as exceptional degraders of particulate organic matter. Additionally, inspection of the gene content revealed the occurrence of homologs for all major components of the T9SS-gliding motility apparatus in all Flavobacteriaceae in contrast to members of other phyla (Cyanobacteria and Proteobacteria) that are known to use different mechanisms for gliding. Phenotypic

assays showed the formation of spreading colonies for some of the tested flavobacteria that had the complete set of T9SS-gliding homologs, confirming that not all potentially gliding bacteria are forming spreading colonies on agar. In terms of their secondary metabolic potential, a large diversity of BGCs was identified in the studied genomes, with terpene BGCs being highly prevalent in both Flavobacteriaceae and Cyanobacteria. Other BGCs that potentially encode proteins required for the production of compounds with known antimicrobial, antioxidant and anticancer properties were found in Flavobacteriaceae. Nevertheless, bioactivity tests did not reflect these genomic findings supporting the fact that the true biosynthetic potential of microorganisms remains hidden due to the 'dormant' state of gene clusters under laboratory conditions. Hence, while this study provides the required broad overview of the genomic content of Flavobacteriaceae in terms of their carbon metabolism, gliding motility and secondary metabolite biosynthetic potential, further studies are essential to enhance our current understanding of these distinct features and how flavobacteria implement them in their natural environment.

Methods

Sample Collection and Isolation of Strains

Samples from the sponges *Aplysina aerophoba* and *Dysidea avara* were collected in January 2012 and June 2014, respectively, from Cala Montgó, Spain ($42^{\circ} 06' 52.20''$ N, $03^{\circ} 10' 06.52''$ E) by SCUBA diving at a depth of approximately 12 m (46, 47). The collection of sponge samples was conducted in strict accordance with Spanish and European regulations within the rules of the Spanish National Research Council with the approval of the Directorate of Research of the Spanish Government. The study was found exempt from ethics approval by the ethics commission of the University of Barcelona since, according to article 3.1 of the European Union directive (2010/63/UE) from the 22/9/2010, no approval is needed for sponge sacrifice, as they are the most primitive animals and lack a nervous system. Moreover, the collected sponges are not listed in CITES. Tissue preparation and cryopreservation were performed as previously described (94). Cryopreserved samples were stored at -80°C . Initial cultivation and isolation of the Flavobacteriaceae strains from the sponges *Aplysina aerophoba* (DN50, DN105, DN112, Aa_C5, Aa_D4 and Aa_F7) and *Dysidea avara* (Da_B9) were described in Versluis et al. (2017) and Gutleben et al. (2019) (46, 47) (Additional File 1, Table S1).

DNA extraction, Identification and Sequencing

Glycerol stocks of the original strains were initially used as inoculum for regrowth on the original solid isolation media at 20°C (Additional File 1, Table S1). Single colonies were picked and cultured in marine broth 2216 (Difco, Detroit, USA) at 30°C . Genomic DNA was extracted using the MasterPure DNA Purification Kit (Epicentre, Madison, USA). The quality, purity and concentration of the extracted DNA was estimated by gel electrophoresis, spectrophotometric analysis using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA) and Qubit dsDNA BR Assay kit (Molecular Probes, Life Technologies) used with the DS-11 FX Fluorometer (DeNovix, USA).

To confirm the identity of the strains, 16S ribosomal RNA (rRNA) gene amplicons were generated by PCR using primers 27F (5'-AGAGTTGGATCMTGGCTCAG-3') and 1492R (5'-CGGTTACCTTGTACGACTT-3'). The PCR reaction mixture contained 10 μL 5X GoTaq reaction buffer (Promega), 1 μL dNTPs (10 mM), 2.5 μL primer 27F (10 μM), 2.5 μL primer 1492R (10 μM), 0.5 μL GoTaq Polymerase (5 U/ μL) (Promega) and 1 μL of the extracted DNA. Nuclease-free water (Promega) was added to reach a total reaction volume of 50 μL . The following conditions were used for the bacterial 16S rRNA gene amplification: initial denaturation at 98°C for 10 min followed by 35 cycles of

denaturation at 98 °C for 20 sec, annealing at 52 °C for 20 sec, elongation at 72 °C for 45 sec and a final extension step at 72 °C for 5 min. PCR products were purified using the GeneJET PCR purification kit (Thermo Fisher Scientific, USA) and quantified using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, USA). The purified PCR products were sent for Sanger sequencing with primers 27F and 1492R (GATC Biotech, Cologne, Germany; now part of Eurofins Genomics Germany GmbH). Trimming (99% good bases, quality value > 20, 25-base window) and contig assembly were conducted with DNA Baser (version 3.5.4.2).

Genome sequencing of strains DN50, DN105 and DN112 was performed using the Illumina MiSeq platform (paired end, 2 × 300 bp reads) (46). The genomes of strains Da_B9, Aa_C5, Aa_D4 and Aa_F7 were sequenced with Illumina HiSeq (paired end, 2 × 150 bp reads). All genomes were sequenced at GATC Biotech (Konstanz, Germany; now part of Eurofins Genomics Germany GmbH).

Genome Assembly and Quality Control

The quality of the reads was assessed with FASTQC 0.11.4 (95). Trimmomatic 0.32 was used to remove the Illumina TruSeq adapter sequences and to perform quality filtering (96). A sliding window trimming approach was employed where part of the read in the window (4 bases) was cropped if the average Phred quality in the window was lower than 20. Any raw reads shorter than 20 bases were discarded. Genome sequences generated by the Illumina MiSeq platform were de novo assembled with the A5-miseq assembler (version 20160825) (97) using default settings. In the case of DN50, the A5-miseq assembler generated a highly fragmented assembly, and the SPAdes 3.11.1 assembler (98) was used instead. For the HiSeq data, the best k-mer size was automatically selected by KmerGenie 1.6741 (99). SPAdes 3.11.1 was employed for the de novo assembly of the Illumina HiSeq reads using the selected k-mer. BLASTN (100) with default settings was employed to investigate the assemblies for contamination. All contigs assigned to sequencing artifacts and contamination (e.g. Enterobacteria phage phiX174) were discarded prior to downstream analysis. Bowtie2 2.2.5 (101) was used to map the quality-filtered reads to the assembled contigs resulting in a sequence alignment map (SAM) file. The SAM file was converted into a binary alignment map (BAM) file that was sorted and indexed using SAMtools 0.1.19 (102). The BAM file was used as input to improve the draft assemblies using Pilon 1.13 (103). In addition, coverage per base was calculated using the 'genomcov' command of BedTools 2.17.0 (104). The quality of the draft assemblies was evaluated using QUAST 4.6.3 (105). Completeness and contamination of all analysed genomes were estimated using CheckM 1.07 with the default set of marker genes (106).

Genome Annotation and Comparative Genomic Analysis

The draft assemblies of the seven strains sequenced in this study were uploaded to the Integrated Microbial Genomes and Microbiomes (IMG/M version 5.0) system (107), and metadata was submitted to the Genomes OnLine Database (GOLD) (108). For the comparative analysis, 59 genomes of strains belonging to the phyla Bacteroidetes (family Flavobacteriaceae), Cyanobacteria and Proteobacteria (Additional File 2) were selected that were publicly available at IMG/M (107). Protein functional families were automatically assigned via the IMG/M pipeline by comparing predicted proteins to Pfam-A (109) Hidden Markov Models using HMMER 3.0b2 (110). Gliding motility and T9SS proteins of the gliding *F. johnsoniae* and non-gliding *P. gingivalis* (Additional File 3) were used to identify homologs in the protein sequences of the studied strains by BLASTP searches with an E-value cut-off of 1e-5 using the IMG BLAST Tool. All selected genomes were downloaded via the IMG/M website for further annotation. Predictions of the protein sequences were obtained using Prokka 1.13 (111). CAZymes were annotated based on HMMER searches (HMMER 3.0b) (110) against the dbCAN database release 6.0 (112). The online server antiSMASH 5.0 (51) was used for the identification of secondary metabolite BGCs with "relaxed" detection

strictness. The ClusterBlast and KnownClusterBlast modules integrated into antiSMASH 5.0 (51) were also used for comparative gene cluster analysis based on the NCBI GenBank (113) and the ‘Minimum Information about a Biosynthetic Gene Cluster’ (MIBiG) (114) data standard, respectively.

Phylogenetic Analysis

BLASTN (100) against the NCBI database was performed using the 16S rRNA gene sequences of the seven sponge-derived strains sequenced in this study. For the reconstruction of the phylogeny, the closest relative of each isolate in NCBI was selected based on the availability of genomes in IMG/M (107). Similarly, representatives of the other clades of the family Flavobacteriaceae were chosen according to the genome availability in IMG/M (107), but also the phylogenetic position of the isolates (based on their 16S rRNA gene sequences) in ARB (115) using the SILVA SSU Ref NR 99 database (release 132) (116) as reference. To compare the functional traits of the Flavobacteriaceae members, representatives of two other phyla that are dominant in the marine environment (Cyanobacteria and Proteobacteria) were included in the analysis and were also used as outgroup. Multiple alignments of the 16S rRNA gene sequences were performed using the SINA Aligner 1.2.11 (117). A Maximum Likelihood tree was constructed in ARB (118) employing RAxML 7.0.3 (119) with 1,000 iterations of rapid bootstrapping. Further editing of the phylogenetic tree was done using the Interactive Tree of Life (iTOL) version 3 (120).

Phenotypic Assays

Gliding Motility Tests

To determine colony spreading, two different types of marine agar plates were prepared using marine broth 2216 (Difco, Detroit, USA) solidified with either 1% or 1.8% of Noble Agar (Sigma-Aldrich). The cells were first grown in marine broth at 30 °C until they reached mid-exponential phase. Subsequently, a 5 µL sample of the cell suspension was spotted at the centre of each test plate using an Eppendorf pipette. After observing growth, the edges of the colonies were viewed with an Axio Scope.A1 (Zeiss) phase-contrast microscope at a magnification of 10X. Pictures of the colony edges were taken using an AxioCam ICc3 (Zeiss) attached to the phase-contrast microscope.

Antimicrobial Activity Tests

Disc diffusion assays were carried out according to the Kirby-Bauer susceptibility method (121). To determine the antimicrobial activity of the seven isolates (Additional File 1: Table S1), a number of indicator strains were used (Additional File 1: Table S6). Prior to the screening, broth cultures of the indicator strains were prepared using 200 µL of the stock cultures to inoculate 3 mL of the respective media and incubated overnight at the corresponding temperatures (Additional File 1: Table S6). Subsequently, 200 µL of the cultures (adjusted to OD₆₀₀ 0.5) were uniformly spread on agar plates using a sterile L-shape spreader. The tested isolates were cultured in marine broth 2216 (Difco, Detroit, USA) at 30 °C until they reached stationary phase. After harvesting, the cultures were centrifuged at 1.657 x g for 20 min at room temperature. The supernatant was sterile-filtered using a 0.2 µm syringe filter to remove the bacterial cells. For screening, sterile, 6-mm diameter filter paper discs were impregnated with 60 µL of the cell-free supernatant and air-dried for 1 h. The paper discs were then transferred onto the agar plates covered with a lawn of the indicator strains. As positive controls, antibiotics known to be effective against the indicator strains were used (Additional File 1: Table S6). The uninoculated growth medium of each of the tested strains was used as negative control, after receiving the same treatment as described for the cultures of the isolates. The plates were incubated for 24 to 48 h at different temperatures depending on the indicator strains used

for the assay. After the incubation, the plates were examined for the presence of inhibition zones that were measured using a calliper. All assays were performed in triplicate.

Statistical analysis

Data were analysed and visualized in R 3.5.0 using vegan 2.5-2 (122), phyloseq 1.26.1 (123), ggplot2 3.1.1 (124), VennDiagram 1.6.20 (125), and ComplexHeatmap (126). Functional comparisons between genomes were performed using Pfam annotations as input data. A Bray-Curtis dissimilarity distance matrix was calculated based on the relative abundance (Gene counts/Total genes) of Pfam profiles with the 'vegdist' function in the vegan R package. Variation in the functional profiles was assessed by non-metric multidimensional scaling (NMDS) ordination using the Bray-Curtis dissimilarity matrix with the 'ordiplot' function in the phyloseq R package. NMDS plots were created using the 'plot_ordination' function of the ggplot2 R package. The significance of the differences in functional profiles across major taxonomic groups and within the Flavobacteriaceae was tested by non-parametric permutational analysis of variance (PERMANOVA) on the Bray-Curtis dissimilarity matrix using the 'adonis' function in the vegan R package, with the number of permutations set at 999. To rank the Pfams contributing the most to the differentiation of the functional profiles between genomes, Similarity Percentage (SIMPER) analysis was employed with the 'simper' function of the vegan R package. For the pairwise comparisons, only Pfam entries with the highest significant contribution to the dissimilarity are shown ($> 0.2\%$, $p < 0.05$).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The raw sequencing data of the genomes (127-133) and the partial 16S rRNA gene sequences (134) of the strains isolated in this study were deposited in the European Nucleotide Archive (ENA) under accession number PRJEB35092 (135). The draft assemblies and metadata of the strains sequenced in this study are publicly available at the Joint Genome Institute (JGI) Genome Portal under IMG Taxon IDs 2806311042 (Aa_C5), 2806311043 (Aa_D4), 2806311049 (Aa_F7), 2806311064 (Da_B9), 2808606303 (DN50), 2808606304 (DN105), 2808606305 (DN112) and GOLD Study ID Gs0135980. Additional information on all assemblies used in the comparative genomics analysis are included in Additional file 2. All codes and script used for the analyses are available at GitHub (https://github.com/mibwurrepo/Gavriilidou_et_al_Flavos_ComparativeGenomics).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was financially supported by the European Commission through the MarPipe (Grant agreement ID: 721421), SponGES (Grant agreement ID: 679849), BluePharm Train (Grant agreement ID: 607786), and EvoTar

(Grant agreement ID: 282004) projects. Further support was obtained from the Netherlands Organisation for Scientific Research through the UNLOCK project (NRGWI.obrug.2018.005).

Authors' Contributions

AG, CJI, MWJvP, HS, DS conceived the study. Sampling was carried out by DS and JG. Isolation of the bacterial strains and their characterization by 16S rRNA gene analysis was done by DV and JG. FF performed the phenotypic assays. The bioinformatics analysis was performed by AG. AG wrote the first draft and all authors contributed to the preparation of the final manuscript.

Acknowledgements

We thank Bart Nijssse, Nikolaos Strepis and Sudarshan Shetty from the Laboratory of Microbiology, Wageningen University & Research, Wageningen, Netherlands for their advice regarding data analysis.

References

1. McBride MJ. The Family *Flavobacteriaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. *The Prokaryotes*. 4th ed. Berlin, Heidelberg: Springer 2014. p. 643-76.
2. Bernardet JF. Family I: *Flavobacteriaceae*. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, et al., editors. *Bergey's Manual of Systematic Bacteriology*. 4. 2nd ed. New York: Springer 2011. p. 106-314.
3. Jooste PG, Hugo CJ. The taxonomy, ecology and cultivation of bacterial genera belonging to the family *Flavobacteriaceae*. *Int J Food Microbiol*. 1999;53:81-94.
4. Bernardet JF, Nakagawa Y. An Introduction to the Family *Flavobacteriaceae*. *Prokaryotes*. 7. New York: Springer; 2006. p. 455-80.
5. List of prokaryotic names with standing in nomenclature (LPSN). <http://www.bacterio.net>. Accessed 10 May 2019.
6. Silva database. <https://www.arb-silva.de/>. Accessed 10 May 2019.
7. Kirchman DL. The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol*. 2002;39(2):91-100.
8. Bennke CM, Kruger K, Kappelmann L, Huang S, Gobet A, Schuler M, et al. Polysaccharide utilisation loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic. *Environ Microbiol*. 2016;18(12):4456-70.
9. DeLong EF, Franks DG, Alldredge AL. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr*. 1993;38(5):924-34.
10. Mann AJ, Hahnke RL, Huang S, Werner J, Xing P, Barbeyron T, et al. The Genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol*. 2013;79(21):6813-22.
11. Yoon BJ, Oh DC. *Spongiibacterium flavum* gen. nov., sp. nov., a member of the family *Flavobacteriaceae* isolated from the marine sponge *Halichondria oshoro*, and emended descriptions of the genera *Croceitalea* and *Flagellimonas*. *Int J Syst Evol Microbiol*. 2012;62:1158-64.
12. Sweet MJ, Croquer A, Bythell JC. Bacterial assemblages differ between compartments within the coral holobiont. *Coral Reefs*. 2010;30(1):39-52.

13. Romanenko LA, Uchino M, Frolova GM, Mikhailov VV. *Marixanthomonas ophiurae* gen. nov., sp. nov., a marine bacterium of the family *Flavobacteriaceae* isolated from a deep-sea brittle star. *Int J Syst Evol Microbiol.* 2007;57:457-62.
14. Barbeyron T, Thomas F, Barbe V, Teeling H, Schenowitz C, Dossat C, et al. Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium *Zobellia galactanivorans* Dsij(T). *Environ Microbiol.* 2016;18(12):4610-27.
15. Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM, et al. Ecology of marine *Bacteroidetes*: a comparative genomics approach. *ISME J.* 2013;7(5):1026-37.
16. Bjursell MK, Martens EC, Gordon JI. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J Biol Chem.* 2006;281(47):36269-79.
17. Shipman JA, Berleman JE, Salyers AA. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J Bacteriol.* 2000;182(19):5365-72.
18. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. *J Biol Chem.* 2009;284(37):24673-7.
19. Kappelmann L, Kruger K, Hehemann JH, Harder J, Markert S, Unfried F, et al. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J.* 2019;13(1):76-91.
20. Xing P, Hahnke RL, Unfried F, Markert S, Huang S, Barbeyron T, et al. Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J.* 2015;9(6):1410-22.
21. McBride MJ, Zhu Y. Gliding motility and Por secretion system genes are widespread among members of the phylum *Bacteroidetes*. *J Bacteriol.* 2013;195(2):270-8.
22. Nett M, König GM. The chemistry of gliding bacteria. *Nat Prod Rep.* 2007;24(6).
23. McBride MJ. *Bacteroidetes* Gliding Motility and the Type IX Secretion System. *Microbiol Spectr.* 2019;7(1).
24. McBride MJ. BACTERIAL GLIDING MOTILITY: Multiple Mechanisms for Cell Movement over Surfaces. *Annu Rev Microbiol.* 2001;55:49-75.
25. McBride MJ, Nakane D. *Flavobacterium* gliding motility and the type IX secretion system. *Curr Opin Microbiol.* 2015;28:72-7.
26. McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG, et al. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microbiol.* 2009;75(21):6864-75.
27. Shrivastava A, Lele PP, Berg HC. A rotary motor drives *Flavobacterium* gliding. *Curr Biol.* 2015;25(3):338-41.
28. Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, McBride MJ, et al. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *PNAS.* 2010;107(1):276-81.
29. Lasica AM, Ksiazek M, Madej M, Potempa J. The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. *Front Cell Infect Microbiol.* 2017;7:215.
30. Veith PD, Glew MD, Gorasia DG, Reynolds EC. Type IX secretion: the generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers. *Mol Microbiol.* 2017;106(1):35-53.
31. Johnston JJ, Shrivastava A, McBride MJ. Untangling *Flavobacterium johnsoniae* Gliding Motility and Protein Secretion. *J Bacteriol.* 2018;200(2).

32. Evans JR, Napier EJ, Fletton RA. G1499-2, a new quinoline compound isolated from the fermentation broth of *Cytophaga johnsonii*. *J Antibiot (Tokyo)*. 1978;31:952-8.
33. Hida T, Tsubotani S, Katayama N, Okazaki H, Harada S. Formadicins, new monocyclic beta-lactam antibiotics of bacterial origin. II. Isolation, characterization and structures. *J Antibiot (Tokyo)*. 1985;38(9):1128-40.
34. Kato T, Hinoo H, Shoji J, Matsumoto K, Tanimoto T, Hattori T, et al. PB-5266 A, B and C, new monobactams: I. Taxonomy, fermentation and isolation. *J Antibiot (Tokyo)*. 1987;55:135-8.
35. Cooper R, Bush K, Principe PA, Trejo WH, Wells JS, Sykes RB. Two new monobactam antibiotics produced by a *Flexibacter* sp. I. Taxonomy, fermentation, isolation and biological properties. *J Antibiot (Tokyo)*. 1983;36(10):1252-7.
36. Funabashi Y, Tsubotani S, Koyama K, Katayama N, Harada S. A New Anti-MRSA Dipeptide, TAN-1057-A. *Tetrahedron*. 1993;49(1):13-28.
37. Katayama N, Fukusumi S, Funabashi Y, Iwahashi T, Ono H. TAN-1057 A-D, new antibiotics with potent antibacterial activity against methicillin-resistant *Staphylococcus aureus*. Taxonomy, fermentation and biological activity. *J Antibiot (Tokyo)*. 1993;46(4):606-13.
38. Imai S, Fujioka K, Furihata K, Fudo R, Yamanaka S, Seto H. Studies on cell growth stimulating substances of low molecular weight. Part 3. Resorcinin, a mammalian-cell growth-stimulating substance produced by *Cytophaga johnsonae*. *J Antibiot (Tokyo)*. 1993;46:1319-22.
39. Umezawa H, Okami Y, Kurasawa S, Ohnuki T, Ishizuka M, Takeuchi T, et al. Marinactan, antitumor polysaccharide produced by marine bacteria. *J Antibiot (Tokyo)*. 1983;XXXVI(5):471-7.
40. Kamiyama T, Umino T, Satoh T, Sawairi S, Shirane M, Ohshima S, et al. Sulfobacins A and B, novel von Willebrand factor receptor antagonists. I. Production, isolation, characterization and biological activities. *J Antibiot (Tokyo)*. 1995;48:924-8.
41. Fujita T, Hatanaka H, Hayashi K, Shigematsu N, Takase S, Okamoto M, et al. FR901451, a novel inhibitor of human leukocyte elastase from *Flexibacter* sp. I. Producing organism, fermentation, isolation, physico-chemical and biological properties. *J Antibiot (Tokyo)*. 1994;47(12):1359-64.
42. Nemoto T, Ojika M, Takahata Y, Andoh T, Sakagami Y. Structures of topostins, DNA topoisomerase I inhibitors of bacterial origin. *Tetrahedron*. 1998;54(12):2683-90.
43. Shindo K, Kikuta K, Suzuki A, Katsuta A, Kasai H, Yasumoto-Hirose M, et al. Rare carotenoids, (3R)-saproxanthin and (3R,2'S)-myxol, isolated from novel marine bacteria (*Flavobacteriaceae*) and their antioxidative activities. *Appl Microbiol Biotechnol*. 2007;74(6):1350-7.
44. Medema MH, Fischbach MA. Computational approaches to natural product discovery. *Nat Chem Biol*. 2015;11(9):639-48.
45. Milshteyn A, Schneider JS, Brady SF. Mining the metabiome: identifying novel natural products from microbial communities. *Chem Biol*. 2014;21(9):1211-23.
46. Versluis D, McPherson K, van Passel MWJ, Smidt H, Sipkema D. Recovery of Previously Uncultured Bacterial Genera from Three Mediterranean Sponges. *Mar Biotechnol* 2017;19:454-68.
47. Gutleben J, Loureiro C, Ramírez Romero LA, Shetty S, Wijffels RH, Smidt H, et al. Cultivation of Bacteria from *Aplysina aerophoba*: Effects of Oxygen and Nutrient Gradients. Submitted. 2019.
48. The CAZypedia Consortium. Ten years of *CAZypedia*: a living encyclopedia of carbohydrate-active enzymes. *Glycobiology*. 2018;28(1):3-8.

49. Hoiczyk E. Gliding motility in cyanobacteria: observations and possible explanations. *Arch Microbiol.* 2000;174(1-2):11-7.
50. Lee MD, Walworth NG, McParland EL, Fu FX, Mincer TJ, Levine NM, et al. The *Trichodesmium* consortium: conserved heterotrophic co-occurrence and genomic signatures of potential interactions. *ISME J.* 2017;11(8):1813-24.
51. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 2019;47(W1):W81-W7.
52. Schofield MM, Jain S, Porat D, Dick GJ, Sherman DH. Identification and analysis of the bacterial endosymbiont specialized for production of the chemotherapeutic natural product ET-743. *Environ Microbiol.* 2015;17(10):3964-75.
53. Matilla MA, Stockmann H, Leeper FJ, Salmond GP. Bacterial biosynthetic gene clusters encoding the anti-cancer heterumalide class of molecules: biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A. *J Biol Chem.* 2012;287(46):39125-38.
54. Liu J, Xue CX, Sun H, Zheng Y, Meng Z, Zhang XH. Carbohydrate catabolic capability of a *Flavobacteriia* bacterium isolated from hadal water. *Syst Appl Microbiol.* 2019;42(3):263-74.
55. Silva SG, Blom J, Keller-Costa T, Costa R. Comparative genomics reveals complex natural product biosynthesis capacities and carbon metabolism across host-associated and free-living *Aquimarna* (*Bacteroidetes*, *Flavobacteriaceae*) species. *Environ Microbiol.* 2019.
56. Bondarev V, Richter M, Romano S, Piel J, Schwedt A, Schulz-Vogt HN. The genus *Pseudovibrio* contains metabolically versatile bacteria adapted for symbiosis. *Environ Microbiol.* 2013;15(7):2095-113.
57. Versluis D, Nijssse B, Naim MA, Koehorst JJ, Wiese J, Imhoff JF, et al. Comparative Genomics Highlights Symbiotic Capacities and High Metabolic Flexibility of the Marine Genus *Pseudovibrio*. *Genome Biol Evol.* 2018;10(1):125-42.
58. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, et al. Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable *Alphaproteobacteria*. *Sci Rep.* 2019;9(1):1999.
59. Wollenberg MS, Ruby EG. Phylogeny and fitness of *Vibrio fischeri* from the light organs of *Euprymna scolopes* in two Oahu, Hawaii populations. *ISME J.* 2012;6(2):352-62.
60. Martiny AC, Treseder K, Pusch G. Phylogenetic conservatism of functional traits in microorganisms. *ISME J.* 2013;7(4):830-8.
61. Zimmerman AE, Martiny AC, Allison SD. Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J.* 2013;7(6):1187-99.
62. Del Bem LEV, Vincentz MGA. Evolution of xyloglucan-related genes in green plants. *BMC Evol Biol.* 2010;10.1(341).
63. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J.* 2013;7(12):2287-300.
64. Bayer K, Jahn MT, Slaby BM, Moitinho-Silva L, Hentschel U. Marine Sponges as *Chloroflexi* Hot Spots: Genomic Insights and High-Resolution Visualization of an Abundant and Diverse Symbiotic Clade. *mSystems.* 2018;3(6).

65. Bernardet JF. Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *Int J Syst Evol Microbiol.* 2002;52(3):1049-70.
66. Hunnicutt DW, Kempf MJ, McBride MJ. Mutations in *Flavobacterium johnsoniae* gldF and gldG disrupt gliding motility and interfere with membrane localization of GldA. *J Bacteriol.* 2002;184(9):2370-8.
67. Zhu Y, McBride MJ. Comparative Analysis of *Cellulophaga algicola* and *Flavobacterium johnsoniae* Gliding Motility. *J Bacteriol.* 2016;198(12):1743-54.
68. Shrivastava A, Rhodes RG, Pochiraju S, Nakane D, McBride MJ. *Flavobacterium johnsoniae* RemA is a mobile cell surface lectin involved in gliding. *J Bacteriol.* 2012;194(14):3678-88.
69. Raina JB, Fernandez V, Lambert B, Stocker R, Seymour JR. The role of microbial motility and chemotaxis in symbiosis. *Nat Rev Microbiol.* 2019;17(5):284-94.
70. Wilde A, Mullineaux CW. Motility in cyanobacteria: polysaccharide tracks and Type IV pilus motors. *Mol Microbiol.* 2015;98(6):998-1001.
71. Khayatan B, Meeks JC, Risser DD. Evidence that a modified type IV pilus-like system powers gliding motility and polysaccharide secretion in filamentous cyanobacteria. *Mol Microbiol.* 2015;98(6):1021-36.
72. Luciano J, Agrebi R, Le Gall AV, Wartel M, Fiegna F, Ducret A, et al. Emergence and modular evolution of a novel motility machinery in bacteria. *PLoS Genet.* 2011;7(9):e1002268.
73. Jarrell KF, McBride MJ. The surprisingly diverse ways that prokaryotes move. *Nat Rev Microbiol.* 2008;6(6):466-76.
74. Braun TF, Khubbar MK, Saffarini DA, McBride MJ. *Flavobacterium johnsoniae* gliding motility genes identified by mariner mutagenesis. *J Bacteriol.* 2005;187(20):6943-52.
75. Wolkin RH, Pate JL. Translocation of motile cells of the gliding bacterium *Cytophaga johnsonae* depends on a surface component that may be modified by sugars. *J Gen Microbiol.* 1984;130:2651-69.
76. Gorski L, Godchaux WI, Leadbetter ER. Structural specificity of sugars that inhibit gliding motility of *Cytophaga johnsonae*. *Arch Microbiol.* 1993;160:121-5.
77. Glew MD, Veith PD, Chen D, Gorasia DG, Peng B, Reynolds EC. PorV is an Outer Membrane Shuttle Protein for the Type IX Secretion System. *Sci Rep.* 2017;7(1):8790.
78. Glew MD, Veith PD, Peng B, Chen YY, Gorasia DG, Yang Q, et al. PG0026 is the C-terminal signal peptidase of a novel secretion system of *Porphyromonas gingivalis*. *J Biol Chem.* 2012;287(29):24605-17.
79. Kharade SS, McBride MJ. *Flavobacterium johnsoniae* PorV is required for secretion of a subset of proteins targeted to the type IX secretion system. *J Bacteriol.* 2015;197(1):147-58.
80. Tholl D. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr Opin Plant Biol.* 2006;9(3):297-304.
81. Yamada Y, Kuzuyama T, Komatsu M, Shin-Ya K, Omura S, Cane DE, et al. Terpene synthases are widely distributed in bacteria. *PNAS.* 2015;112(3):857-62.
82. Gallucci MN, Oliva M, Casero C, Dambolena J, Luna A, Zygadlo J, et al. Antimicrobial combined action of terpenes against the food-borne microorganisms *Escherichia coli*, *Staphylococcus aureus* and *Bacillus cereus*. *Flavour Frag J.* 2009;24(6):348-54.
83. Zhao B, Lin X, Lei L, Lamb DC, Kelly SL, Waterman MR, et al. Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). *J Biol Chem.* 2008;283(13):8183-9.
84. Song C, Schmidt R, de Jager V, Krzyzanowska D, Jongedijk E, Cankar K, et al. Exploring the genomic traits of fungus-feeding bacterial genus *Collimonas*. *BMC Genomics.* 2015;16:1103.

85. Hudson J, Kumar V, Egan S. Comparative genome analysis provides novel insight into the interaction of *Aquimarina* sp. AD1, BL5 and AD10 with their macroalgal host. *Mar Genomics*. 2019;46:8-15.
86. Harms H, Klockner A, Schror J, Josten M, Kehraus S, Crusemann M, et al. Antimicrobial Dialkylresorcins from Marine-Derived Microorganisms: Insights into Their Mode of Action and Putative Ecological Relevance. *Planta Med*. 2018;84(18):1363-71.
87. Knerr PJ, van der Donk WA. Discovery, biosynthesis, and engineering of lantipeptides. *Annu Rev Biochem*. 2012;81:479-505.
88. Desriac F, Defer D, Bourgougnon N, Brillet B, Le Chevalier P, Fleury Y. Bacteriocin as weapons in the marine animal-associated bacteria warfare: inventory and potential applications as an aquaculture probiotic. *Mar Drugs*. 2010;8(4):1153-77.
89. Riley MA, Wertz JE. Bacteriocins: evolution, ecology, and application. *Annu Rev Microbiol*. 2002;56:117-37.
90. Shim JS, Liu JO. Recent advances in drug repositioning for the discovery of new anticancer drugs. *Int J Biol Sci*. 2014;10(7):654-63.
91. Mojib N, Philpott R, Huang JP, Niederweis M, Bej AK. Antimycobacterial activity *in vitro* of pigments isolated from Antarctic bacteria. *Antonie Van Leeuwenhoek*. 2010;98(4):531-40.
92. Hertweck C. Hidden biosynthetic treasures brought to light. *Nat Chem Biol*. 2009;5(7):450-2.
93. Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol*. 2015;13(8):509-23.
94. Sipkema D, Schippers K, Maalcke WJ, Yang Y, Salim S, Blanch HW. Multiple approaches to enhance the cultivability of bacteria associated with the marine sponge *Haliclona (gellius)* sp. *Appl Environ Microbiol*. 2011;77(6):2130-40.
95. Andrews S. FASTQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. 2010.
96. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20.
97. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*. 2015;31(4):587-9.
98. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-77.
99. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics*. 2014;30(1):31-7.
100. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403-10.
101. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-9.
102. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
103. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963.
104. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*. 2014;47:11 2 1-34.

105. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-5.
106. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25(7):1043-55.
107. Chen IA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res*. 2019;47(D1):D666-D77.
108. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Katta HY, Mojica A, et al. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res*. 2019;47(D1):D649-D59.
109. 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(D1):D427-D32.
110. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. *Nucleic Acids Res*. 2015;43(W1):W30-8.
111. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-9.
112. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42(Database issue):D490-5.
113. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res*. 2013;41(Database issue):D36-42.
114. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol*. 2015;11(9):625-31.
115. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadukumar, et al. ARB: a software environment for sequence data. *Nucleic Acids Res*. 2004;32(4):1363-71.
116. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(Database issue):D590-6.
117. Pruesse E, Peplies J, Glockner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*. 2012;28(14):1823-9.
118. Westram R, Bader K, Pruesse E, Kumar Y, Meier H, Glockner FO, et al. ARB: a software environment for sequence data. In: de Bruijn FJ, editor. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*: John Wiley & Sons; 2011. p. 399-406.
119. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol*. 2008;57(5):758-71.
120. 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 Res*. 2016;44(W1):W242-5.
121. Bauer AW, Kirby WM, Sherris JC, Turck M. Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol*. 1966;45(4):493-6.
122. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: Community Ecology Package. R package version 2.5-2. <https://CRAN.R-project.org/package=vegan>. 2018.
123. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217.
124. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Verlag New York: Springer; 2016.

125. Chen H. VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20. <https://CRAN.R-project.org/package=VennDiagram>. 2018.
126. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847-9.
127. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Ingham CJ, Smidt H, et al. Raw DNA sequences of *Flavobacteriaceae* isolated from marine sponges. ENA; 2020. ERS3924874, <http://www.ebi.ac.uk/ena/data/view/ERS3924874>, Access date: 7 Jan 2020
128. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Ingham CJ, Smidt H, et al. Raw DNA sequences of *Flavobacteriaceae* isolated from marine sponges. ENA; 2020. ERS3924875, <http://www.ebi.ac.uk/ena/data/view/ERS3924875>, Access date: 7 Jan 2020.
129. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Ingham CJ, Smidt H, et al. Raw DNA sequences of *Flavobacteriaceae* isolated from marine sponges. ENA; 2020. ERS3924876, <http://www.ebi.ac.uk/ena/data/view/ERS3924876>, Access date: 7 Jan 2020.
130. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Ingham CJ, Smidt H, et al. Raw DNA sequences of *Flavobacteriaceae* isolated from marine sponges. ENA; 2020. ERS3924877, <http://www.ebi.ac.uk/ena/data/view/ERS3924874>, Access date: 7 Jan 2020.
131. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Ingham CJ, Smidt H, et al. Raw DNA sequences of *Flavobacteriaceae* isolated from marine sponges. ENA; 2020. ERS3924878, <http://www.ebi.ac.uk/ena/data/view/ERS3924878>, Access date: 7 Jan 2020.
132. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Ingham CJ, Smidt H, et al. Raw DNA sequences of *Flavobacteriaceae* isolated from marine sponges. ENA; 2020. ERS3924879, <http://www.ebi.ac.uk/ena/data/view/ERS3924879>, Access date: 7 Jan 2020.
133. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Ingham CJ, Smidt H, et al. Raw DNA sequences of *Flavobacteriaceae* isolated from marine sponges. ENA; 2020. ERS3924880, <http://www.ebi.ac.uk/ena/data/view/ERS3924880>, Access date: 7 Jan 2020.
134. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Ingham CJ, Smidt H, et al. Partial 16S rRNA gene sequences of *Flavobacteriaceae* isolated from marine sponges. ENA; 2020. LR736666-LR736669, <http://www.ebi.ac.uk/ena/data/view/LR736666-LR736669>, Access date: 7 Jan 2020.
135. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Van Passel MW, Ingham CJ, et al. Comparative genomic analysis of *Flavobacteriaceae*: insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis. ENA. <https://www.ebi.ac.uk/ena/data/view/PRJEB3509>.

Supplementary Files Legend

Additional files

Additional file 1:

Supplementary Information. (.pdf 317 Kb)

Table S1. Strains and growth media. Details on the preparation of the media and the cultivation conditions can be found in the References section.

Table S2. Identification of flavobacterial strains sequenced in this study.

Table S3. Pfam entries most strongly contributing to differentiating genomes from different *Flavobacteriaceae* clades. Pfam entries with the highest significant contribution ($> 0.2\%$, $p < 0.05$) to the dissimilarity are shown. M, Marine; C, *Capnocytophaga*; T, *Tenacibaculum-Polaribacter*; F, *Flavobacterium*.

Table S4. Number of BGCs, average gene counts and % of genes in BGCs per group.

Table S5. Distribution of identified BGCs across different groups and *Flavobacteriaceae* clades. M, Marine; C, *Capnocytophaga*; F, *Flavobacterium*; T, *Tenacibaculum-Polaribacter*.

Table S6. Cultivation conditions of indicator strains and antibiotics used in the antimicrobial activity tests.

Additional file 2:

Publicly available genomes used in this study. (.pdf 65.7 Kb)

Additional file 3:

List of proteins involved in gliding motility and type 9 secretion in *F. johnsoniae* UW101 and *P. gingivalis* W83 used in the comparative genomics analysis. (.xlsx 26.4 Kb)

Additional file 4:

Average and total number of CAZymes in different taxonomic groups and *Flavobacteriaceae* clades. (.xlsx 45.5 Kb)

Figures

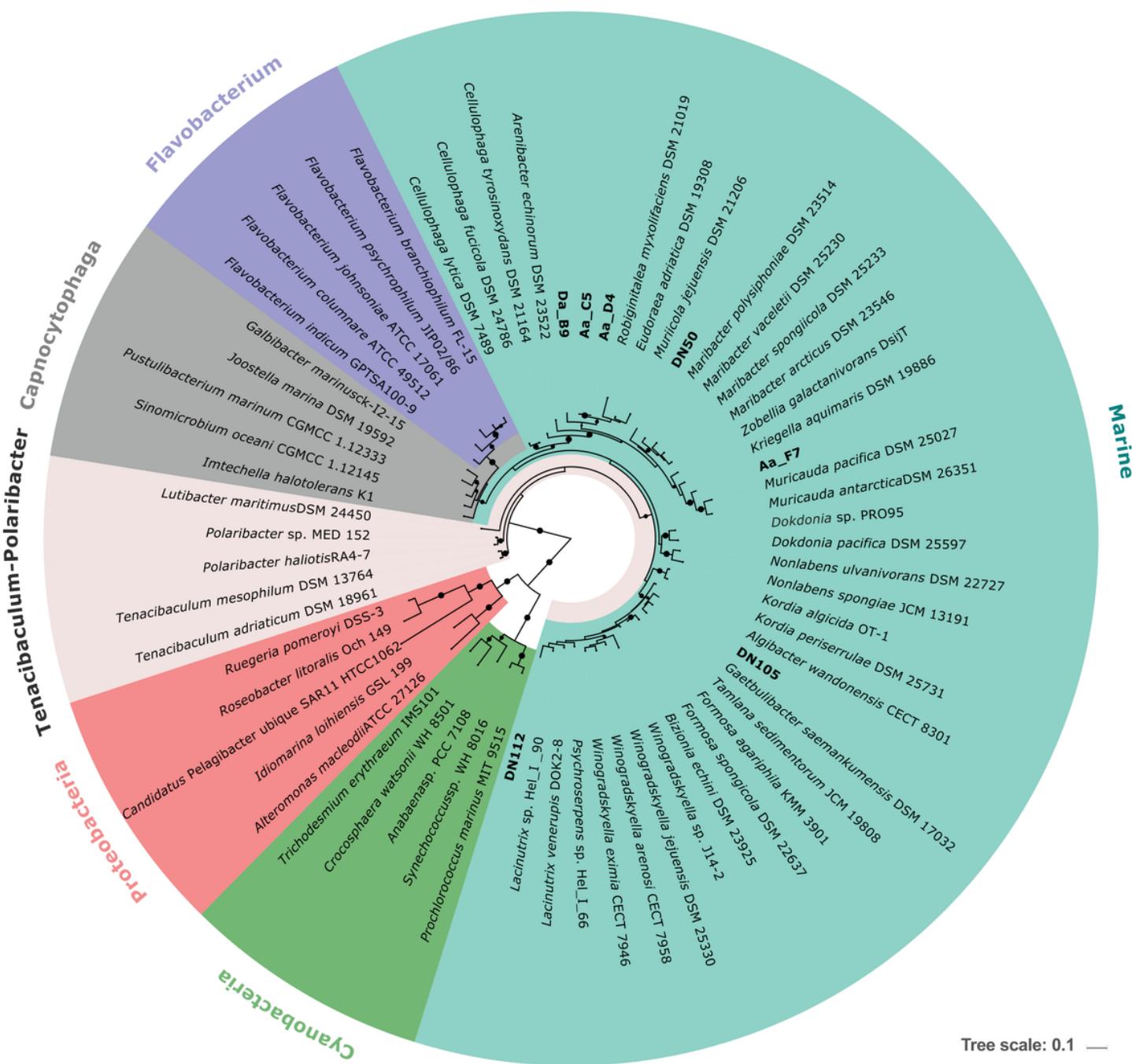


Figure 1

16S rRNA gene phylogenetic tree of microorganisms whose genomes were compared in this study. Phylogeny was inferred using a maximum likelihood method. Bootstrap values (>70%) are displayed as black circles on the middle of the branches. Colour annotations represent the different Flavobacteriaceae clades. Sequences belonging to Cyanobacteria and Proteobacteria were used as outgroups. Names in bold indicate sequences generated in the current study. The scale bar indicates 0.1 substitutions per site.

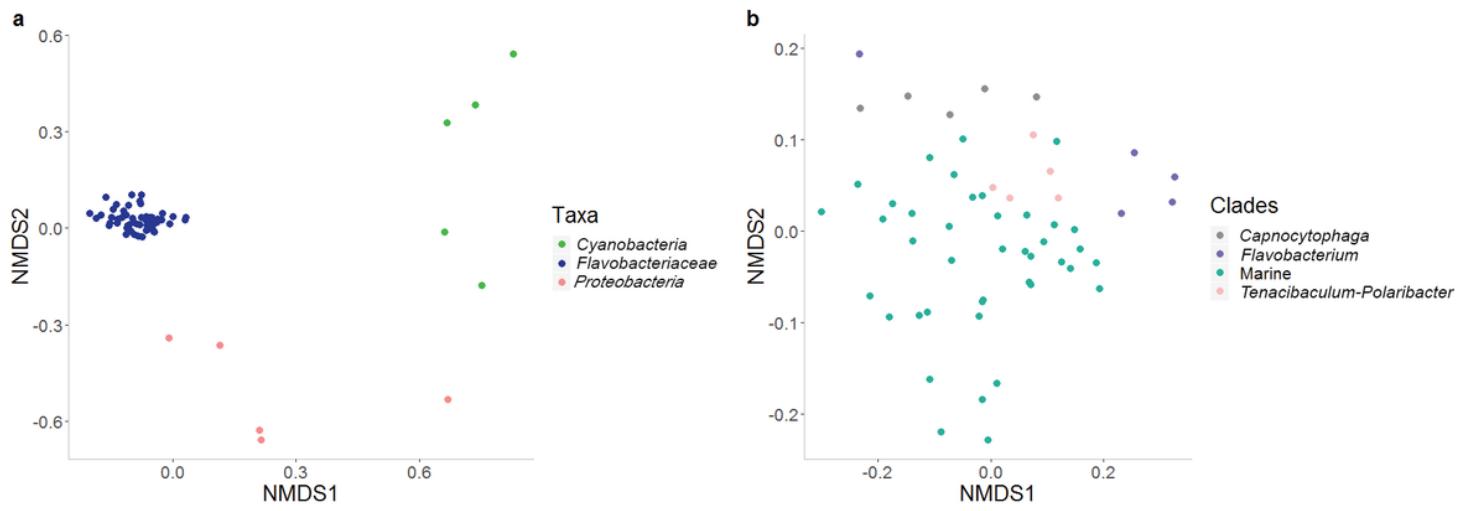


Figure 2

NMDS clustering of genomes based on similarity of functional groups according to Pfam annotations. Genomes are plotted following Bray-Curtis (dis)similarity values calculated from Pfam abundance profiles between major taxonomic groups (a) and different Flavobacteriaceae clades (b).

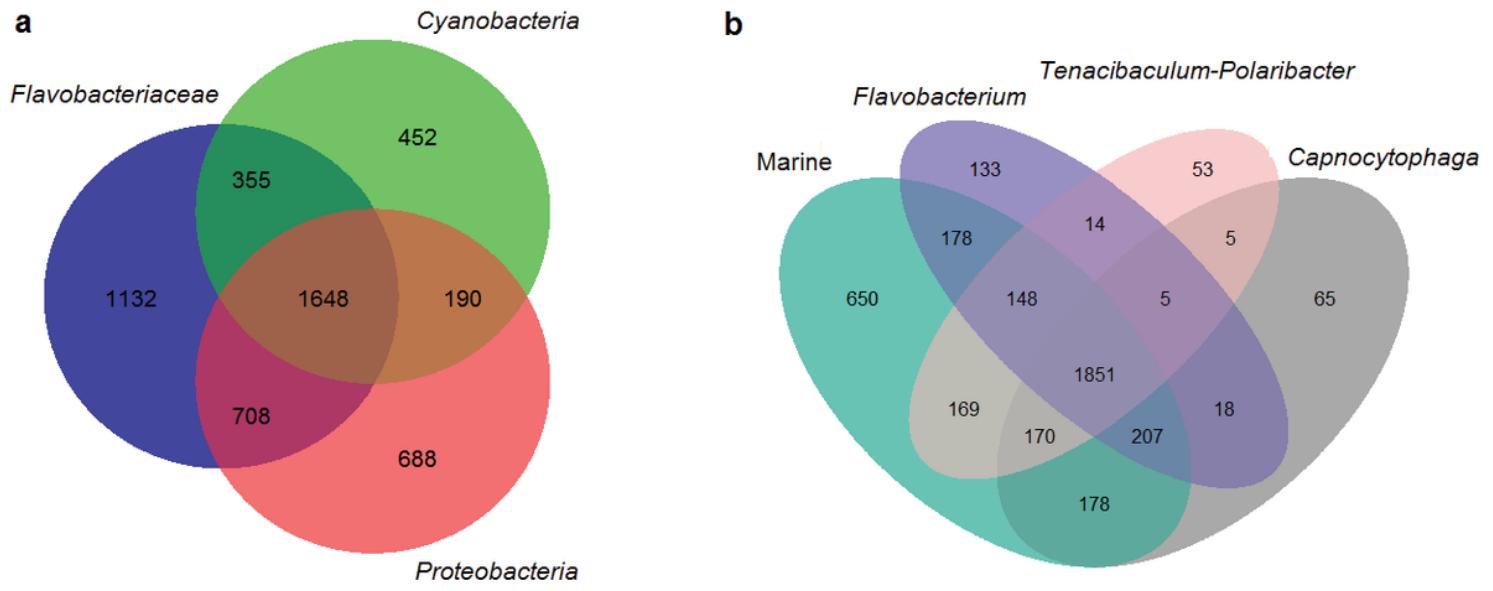


Figure 3

Shared and unique protein families (Pfams) between the analysed genomes. Venn diagrams illustrate shared and unique Pfams amongst Flavobacteriaceae, Cyanobacteria and Proteobacteria (a) and within the Flavobacteriaceae (b).

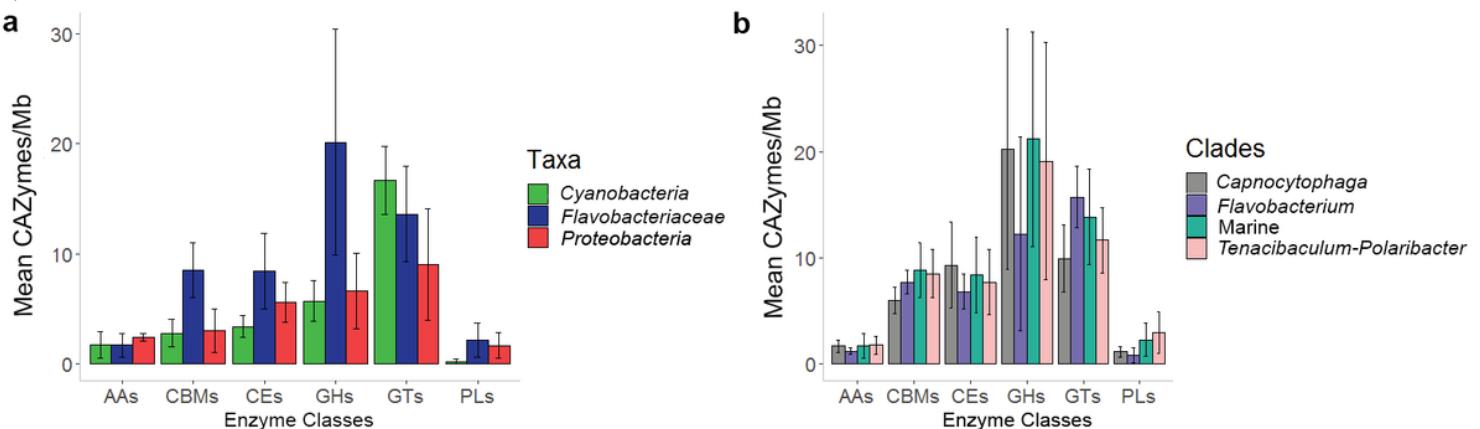


Figure 4

Mean number of CAZymes per Mb in the analysed genomes. Frequency of CAZymes normalized by total gene counts is shown between different taxonomic groups (a) and Flavobacteriaceae clades (b). GHs, glycoside hydrolases; PLs, polysaccharide lyases; CEs, carbohydrate esterases; GTs, glycosyl transferases; CBMs, carbohydrate-binding modules.

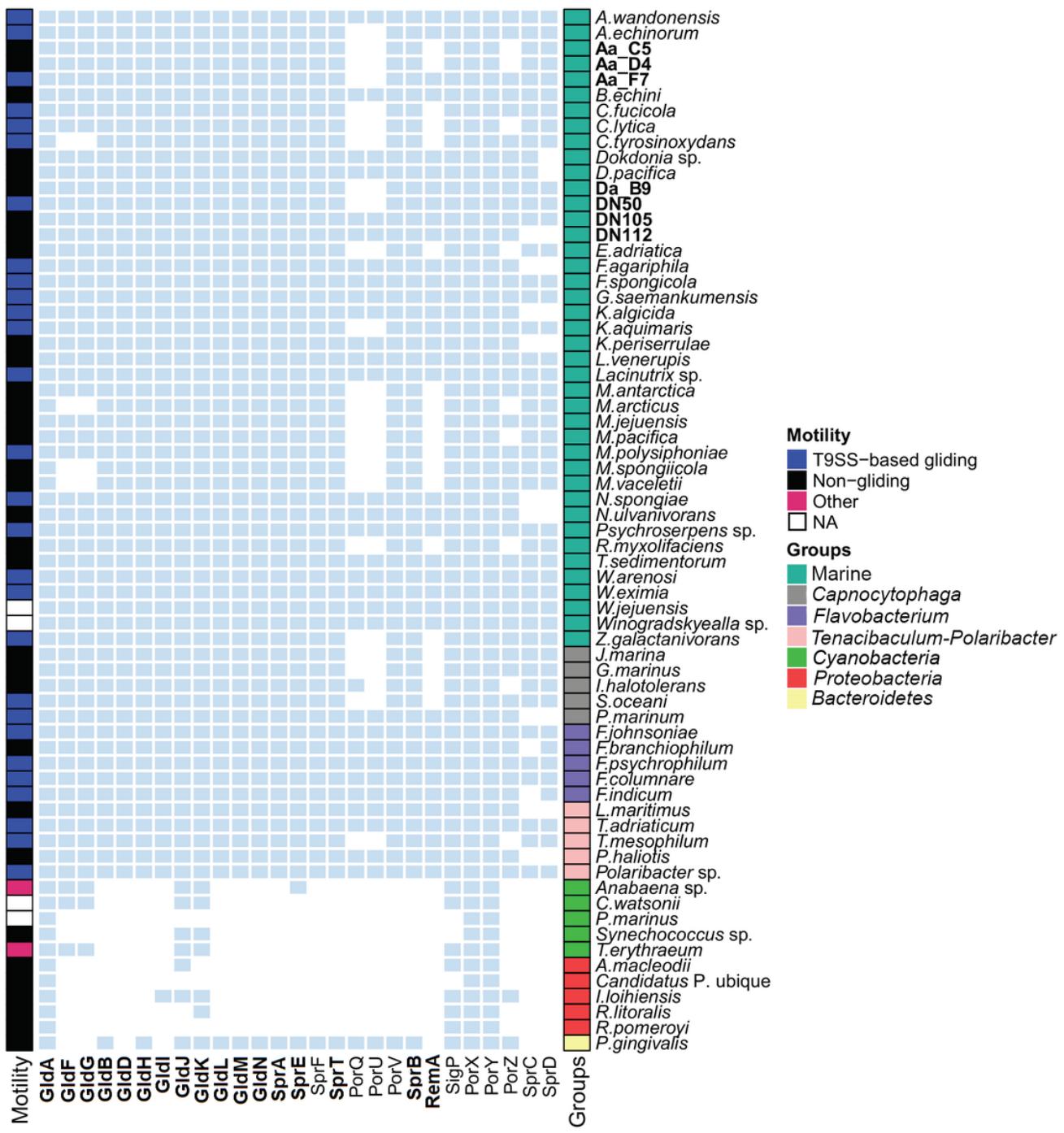


Figure 5

Presence of homologs of gliding motility and type 9 secretion proteins in the predicted proteomes. Presence of homologs is indicated by light blue squares and absence by white squares. Strains sequenced in the framework of this study are indicated in bold. Each species was annotated based on its taxonomy (Groups) and motility as reported in literature. Protein names in bold are considered essential or nearly essential for *F. johnsoniae* gliding. The rest of the proteins are either non-essential or their role in *F. johnsoniae* gliding has not been determined (23, 30). Other, gliding motility mechanism other than T9SS-based; NA, no information available (see also Additional File 3).



Figure 6

Colony morphologies of the studied flavobacterial strains. (a) and (b) Spreading colonies of strains DN50 and Aa_F7, respectively, that form slender peninsulas on 1% agar; (c) Surface contour pattern of DN50 colonies on 1.8% agar; (d) Non-spreading colonies of DN105. Photos were taken with an AxioCam ICc3 attached to an Axio Scope.A1 (Zeiss) phase-contrast microscope.

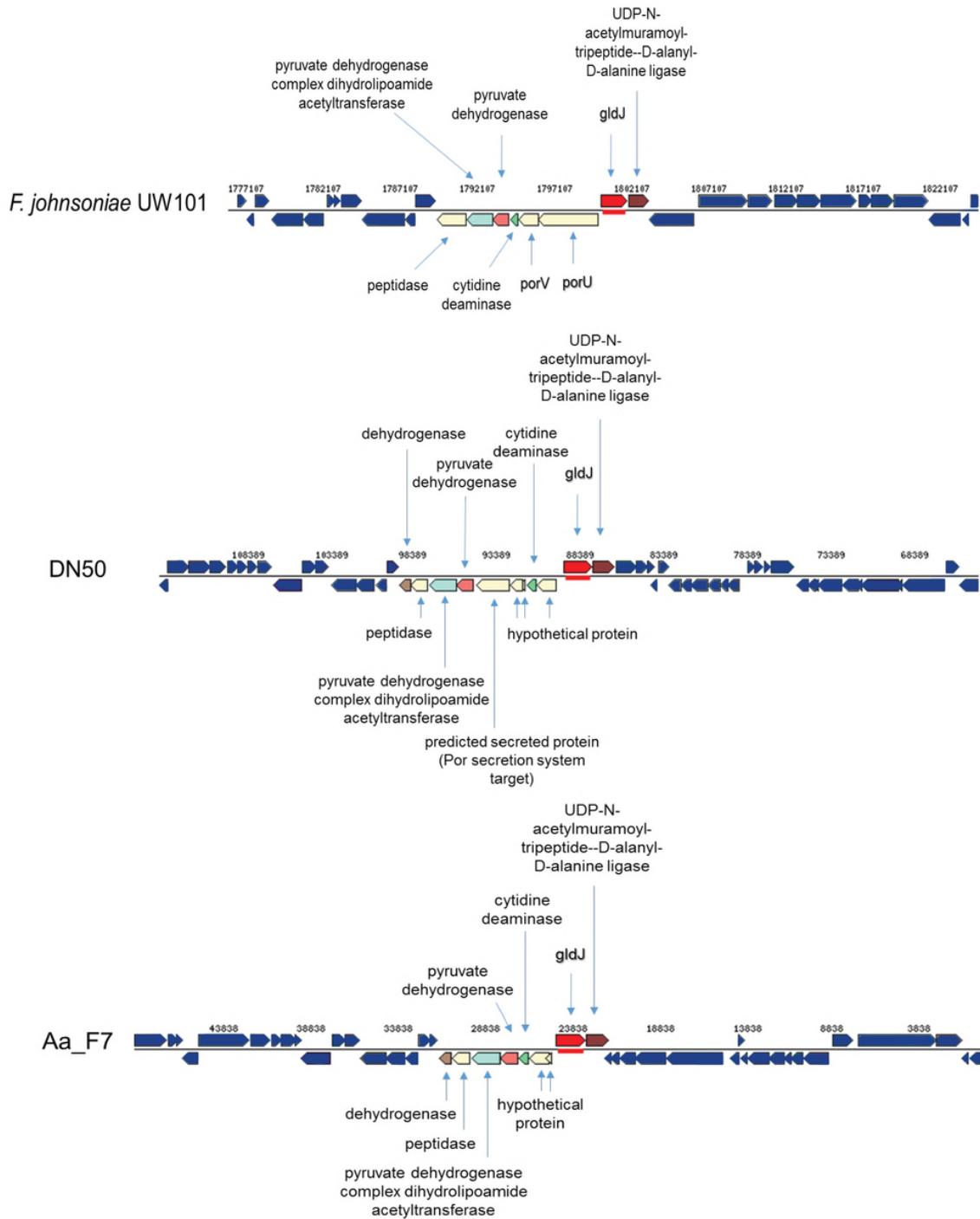


Figure 7

Gene neighbourhoods of *gldJ* in Aa_F7 and DN50 compared to *F. johnsoniae* UW101 genome. Genes of the same colour are from the same orthologous group. Light yellow genes had no Cluster of Orthologous Group (COG) assignment. (Chromosome viewer-IMG/MER).

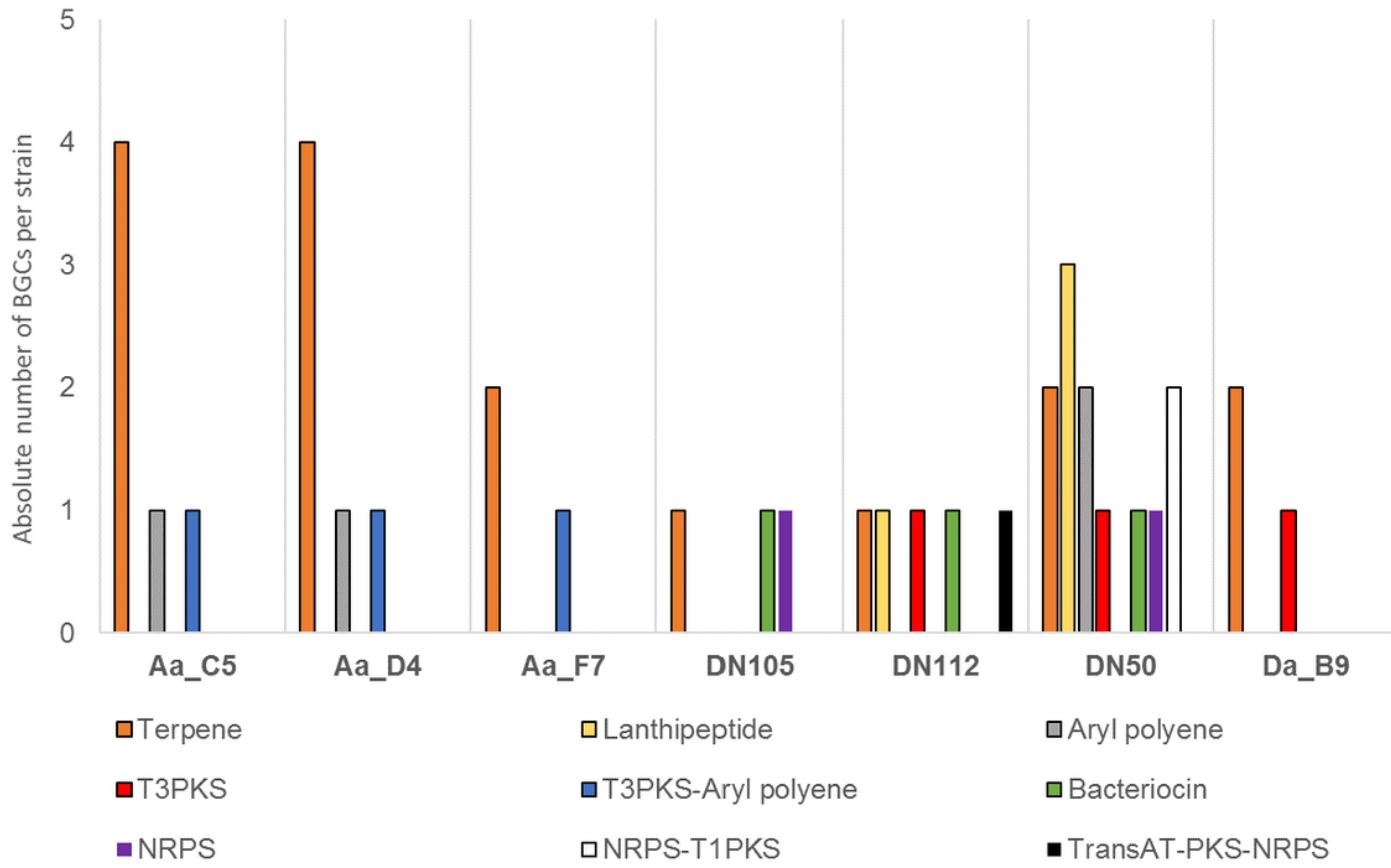


Figure 8

Composition of BGCs identified in the genomes newly sequenced in this study. The absolute number of BGCs per strain assigned to each BGC class is shown. Different colours indicate different BGC classes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile4.xlsx
- AdditionalFile1.pdf
- AdditionalFile3.xlsx
- AdditionalFile2.pdf

Table 3
 Distribution of BGCs (mean number of BGCs) across major taxonomic groups and different
Flavobacteriaceae clades.

Gene cluster type (antiSMASH)	Flavobacteriaceae				Cyanobacteria	Proteobacteria
	M	C	F	T		
Terpene	1.63	1.20	1.40	1.20	1.60	0.40
T3PKS	0.39	0	0.20	0.20	0.20	0
NRPS	0.29	0	0.20	0	1.00	0
Lanthipeptide	0.22	0.40	0	0.20	0	0
Aryl polyene	0.17	0	0.60	0.20	0.60	0.20
Aryl polyene-Resorcinol	0.17	0	0.60	0	0	0
Bacteriocin	0.12	0.20	0.20	0	1.40	0.80
NRPS-like	0.05	0.40	0	0	0.60	0
Siderophore	0	0.20	0.60	0.20	0	0.20
Hserlactone	0.02	0	0	0	0	1.40
Betalactone	0	0	0.20	0	0	0.40