

Supplementary Information

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

Asimenia Gavriilidou, Johanna Gutleben, Dennis Versluis, Francesca Forgiarini, Mark W.J. van Passel, Colin J. Ingham, Hauke Smidt and Detmer Sipkema.

Table of contents

Table S1	3
Table S2	4
Table S3	4
Table S4	7
Table S5	7
Table S6	10

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

Strain ID	Isolation source	Growth Medium	References
DN50	<i>A. aerophoba</i>	Mucin & Marine agar	Versluis et al. (2017)
DN105	<i>A. aerophoba</i>	Marine agar	Versluis et al. (2017)
DN112	<i>A. aerophoba</i>	Marine agar	Versluis et al. (2017)
Da_B9	<i>D. avara</i>	Marine agar	unpublished
Aa_F7	<i>A. aerophoba</i>	Mucin (50x)	Gutleben et al. (submitted)
Aa_C5	<i>A. aerophoba</i>	All amino acids (50x)	Gutleben et al. (submitted)
Aa_D4	<i>A. aerophoba</i>	All amino acids (50x)	Gutleben et al. (submitted)

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

	Aa_C5	Aa_D4	Aa_F7	Da_B9	DN50	DN105	DN112
16S rRNA gene length (bp)	1362	1349	1257	1510	1121	1190	1391
Closest NCBI hit (Accession number)	<i>Eudoraea chungangensis</i> (NR_148299.1)	<i>Eudoraea chungangensis</i> (NR_148299.1)	<i>Flagellimonas</i> sp. (MF287795.1)	<i>Arenibacter</i> sp. (MH753600.1)	<i>Spongiimicrobium salis</i> (NR_146706.1)	<i>Gaetbulibacter</i> sp. (KY463498.1)	<i>Lacinutrix</i> sp. (KX398615.1)
Identity (%)	99.7	99.7	97.6	93.5	97.7	96.6	97.2

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*; F, *Flavobacterium*.

Function name (Pfam ID)	Relative Abundance (%)				Contribution (%)			p-values (< 0.05)		
	M	C	T	F	M-C	M-T	M-F	M-C	M-T	M-F
CarboxypepD_reg-like domain (pfam13715)	1.20	1.78	1.28	0.95	1.06	0.51	0.68	0.011	0.999	0.398
TonB-dependent Receptor Plug Domain (pfam07715)	1.08	1.65	1.16	0.97	0.93	0.64	0.57	0.01	0.984	0.731
SusD family (pfam07980)	0.30	0.81	0.26	0.17	0.90	0.36	0.41	0.002	1	0.616
Starch-binding associating with outer membrane (pfam14322)	0.28	0.78	0.24	0.17	0.71	0.38	0.43	0.002	1	0.591

TonB dependent receptor (pfam00593)	0.53	0.90	0.47	0.45	0.62	0.36	0.43	0.01	1	0.438
Sigma-70, region 4 (pfam08281)	0.53	0.62	0.52	0.26	0.42	0.23	0.39	0.04	0.999	0.027
FecR protein (pfam04773)	0.12	0.31	0.08	0.04	0.41	0.17	0.18	0.015	0.977	0.492
Domain of unknown function (DUF4974) (pfam16344)	0.12	0.30	0.08	0.05	0.40	0.17	0.19	0.016	0.978	0.469
Glycosyl hydrolases family 43 (pfam04616)	0.06	0.28	0.02	0.03	0.40	0.10	0.11	0.001	0.998	0.931
Phage integrase family (pfam00589)	0.19	0.32	0.19	0.27	0.33	0.15	0.21	0.015	0.999	0.112
Outer membrane efflux protein (pfam02321)	0.25	0.44	0.28	0.31	0.32	0.13	0.14	0.001	0.993	0.432
Barrel-sandwich domain of CusB or HlyD membrane-fusion (pfam16576)	0.20	0.37	0.22	0.29	0.31	0.16	0.16	0.001	0.676	0.252
AcrB/AcrD/AcrF family (pfam00873)	0.16	0.33	0.19	0.19	0.31	0.13	0.12	0.001	0.888	0.606
ABC transporter (pfam00005)	0.88	0.79	0.86	0.96	0.30	0.22	0.23	0.039	0.955	0.202
Tetratricopeptide repeat (pfam13181)	0.33	0.16	0.27	0.24	0.29	0.20	0.30	0.022	0.938	0.574
Phage integrase SAM-like domain (pfam13102)	0.09	0.23	0.09	0.06	0.28	0.15	0.09	0.003	0.685	0.996

Integrase core domain (pfam00665)	0.03	0.18	0.01	0.11	0.27	0.07	0.16	0.016	0.999	0.141
Glycosyl hydrolases family 2, sugar binding domain (pfam02837)	0.08	0.20	0.06	0.05	0.25	0.17	0.15	0.04	0.733	0.514
Transposase DDE domain (pfam01609)	0.01	0.15	0.00	0.15	0.25	0.02	0.23	0.048	0.999	0.027
Glycosyl hydrolases family 2 (pfam00703)	0.07	0.19	0.06	0.06	0.24	0.17	0.14	0.033	0.669	0.567
IstB-like ATP binding protein (pfam01695)	0.00	0.14	0.00	0.00	0.24	0.00	0.00	0.001	0.999	1
Starch-binding associating with outer membrane (pfam12771)	0.08	0.18	0.08	0.06	0.20	0.10	0.08	0.005	0.968	0.992
Heavy-metal-associated domain (pfam00403)	0.13	0.23	0.19	0.08	0.20	0.16	0.10	0.024	0.409	0.932
C-terminal domain of CHU protein family (pfam13585)	0.41	0.26	0.26	0.37	0.27	0.30	0.19	0.061	0.045	0.461
Response regulator receiver domain (pfam00072)	1.11	1.01	1.08	1.04	0.46	0.39	0.52	0.529	0.995	0.037
Sigma-70 region 2 (pfam04542)	0.66	0.71	0.60	0.32	0.45	0.23	0.43	0.063	1	0.003
Helix-turn-helix (pfam01381)	0.18	0.27	0.18	0.33	0.21	0.16	0.30	0.386	0.994	0.005
Enoyl-(Acyl carrier protein) reductase (pfam13561)	0.33	0.34	0.27	0.19	0.20	0.21	0.26	0.834	0.938	0.024

Sodium:solute symporter family (pfam00474)	0.18	0.20	0.20	0.05	0.11	0.13	0.20	0.995	0.88	0.001
--	------	------	------	------	------	------	------	-------	------	-------

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

Groups	Total BGCs	Average BGCs	Average genes	% genes in BGCs
<i>Flavobacteriaceae</i> (n = 56)	197	4	4447	0.1
Marine (n = 41)	151	4	3668	0.1
<i>Capnocytophaga</i> (n = 5)	15	3	3591	0.1
<i>Flavobacterium</i> (n = 5)	20	4	3200	0.1
<i>Tenacibaculum-Polaribacter</i> (n = 5)	11	2	3055	0.1
<i>Cyanobacteria</i> (n = 5)	33	7	3397	0.3
<i>Proteobacteria</i> (n = 5)	20	4	3565	0.1

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

Gene cluster type (antiSMASH)	<i>Flavobacteriaceae</i>								<i>Cyanobacteria</i>		<i>Proteobacteria</i>	
	M		C		F		T					
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Aryl polyene	7	4.6	0	0.0	3	15.0	1	9.1	3	9.09	1	5.0
Aryl polyene-Resorcinol	7	4.6	0	0.0	3	15.0	0	0.0	0	0.00	0	0.0

Aryl polyene-Resorcinol-T3PKS	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0
Bacteriocin	5	3.3	1	6.7	1	5.0	0	0.0	7	21.21	4	20.0
Bacteriocin-Lanthipeptide	0	0.0	0	0.0	0	0.0	0	0.0	1	3.03	0	0.0
Betalactone	0	0.0	0	0.0	1	5.0	0	0.0	0	0.00	2	10.0
Ectoine	0	0.0	0	0.0	0	0.0	0	0.0	0	0.00	1	5.0
Hserlactone	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	7	35.0
Ladderane	0	0.0	1	6.7	0	0.0	0	0.0	0	0.00	0	0.0
Lanthipeptide	9	6.0	2	13.3	0	0.0	1	9.1	0	0.00	0	0.0
LAP-Bacteriocin-Cyanobactin	0	0.0	0	0.0	0	0.0	0	0.0	1	3.03	0	0.0
Lasso peptide	0	0.0	0	0.0	0	0.0	0	0.0	0	0.00	1	5.0
Linaridin	2	1.3	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0
Microviridin	0	0.0	1	6.7	0	0.0	0	0.0	0	0.00	0	0.0
NRPS	12	7.9	0	0.0	1	5.0	0	0.0	5	15.15	0	0.0
NRPS-like	2	1.3	2	13.3	0	0.0	0	0.0	3	9.09	0	0.0
NRPS-like-T1PKS	0	0.0	0	0.0	0	0.0	0	0.0	0	0.00	1	5.0
NRPS-T1PKS	8	5.3	0	0.0	0	0.0	0	0.0	2	6.06	0	0.0
Resorcinol	2	1.3	1	6.7	0	0.0	0	0.0	0	0.00	0	0.0
Siderophore	0	0.0	0	0.0	3	15.0	1	9.1	0	0.00	1	5.0
T1PKS	1	0.7	1	6.7	0	0.0	0	0.0	1	3.03	0	0.0
T1PKS-PUFA-hg1E-KS	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0
T3PKS	16	10.6	0	0.0	1	5.0	1	9.1	1	3.03	0	0.0
T3PKS-Aryl polyene	8	5.3	0	0.0	0	0.0	1	9.1	0	0.00	0	0.0
Terpene	67	44.4	6	40.0	7	35.0	6	54.5	8	24.24	2	10.0
Terpene-Bacteriocin	0	0.0	0	0.0	0	0.0	0	0.0	1	3.03	0	0.0
TransAT-PKS-like, TransAT-PKS, PKS-like	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0

TransAT-PKS-NRPS	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0
------------------	---	-----	---	-----	---	-----	---	-----	---	------	---	-----

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

Reference strains	Temperature (°C)	Media	Antibiotics (Positive control)
<i>Bacillus subtilis</i> (DSM 402)	30	Nutrient broth	Tetracycline
<i>Staphylococcus simulans</i> (DSM 20037)	37	Tryptic Soy broth (TSB)	Vancomycin
<i>Escherichia coli</i> (MIB, WUR)	37	Luria-Bertani broth (LB)	Tetracycline
<i>Aeromonas salmonicida</i> (DSM 19634)	30	Nutrient broth	Vancomycin
<i>Saprolegnia parasitica</i> (CBS 223.65)	20	Potato Dextrose agar (PDA)	Delvocid
<i>Candida oleophila</i> (DSM 70763)	25	Yeast Malt Broth (YMB)	Nystatin