SUPPLEMENTARY INFORMATION for

A High-Resolution LC-MS-Based Secondary Metabolite Fingerprint Database of Marine Bacteria

Liang Lu[1,7], Jijie Wang[2,7], Ying Xu[3], Kailing Wang[4], Yingwei Hu[5], Renmao Tian[1], Bo Yang[3], Qiliang Lai[6], Yongxin Li[3], Weipeng Zhang[1], Zongze Shao[6], Henry Lam[2,5,*], Pei-Yuan Qian[1,3,*]

[1] Environmental Science Program, School of Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong 999077,China.

[2] Division of Biomedical Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong 999077, China.

[3] Division of Life Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong 999077, China.

[4] School of Medicine and Pharmacy, Ocean University of China, Qingdao 266003, China.

[5] Department of Chemical and Biomolecular Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong 999077, China.

[6] Third Institute of Oceanography, State Oceanic Administration, Xiamen 361005, China.

[7] These authors contributed equally to this work.

* Correspondence should be addressed to P.Y.Qian (boqianpy@ust.hk) or H. Lam (kehlam@ust.hk)

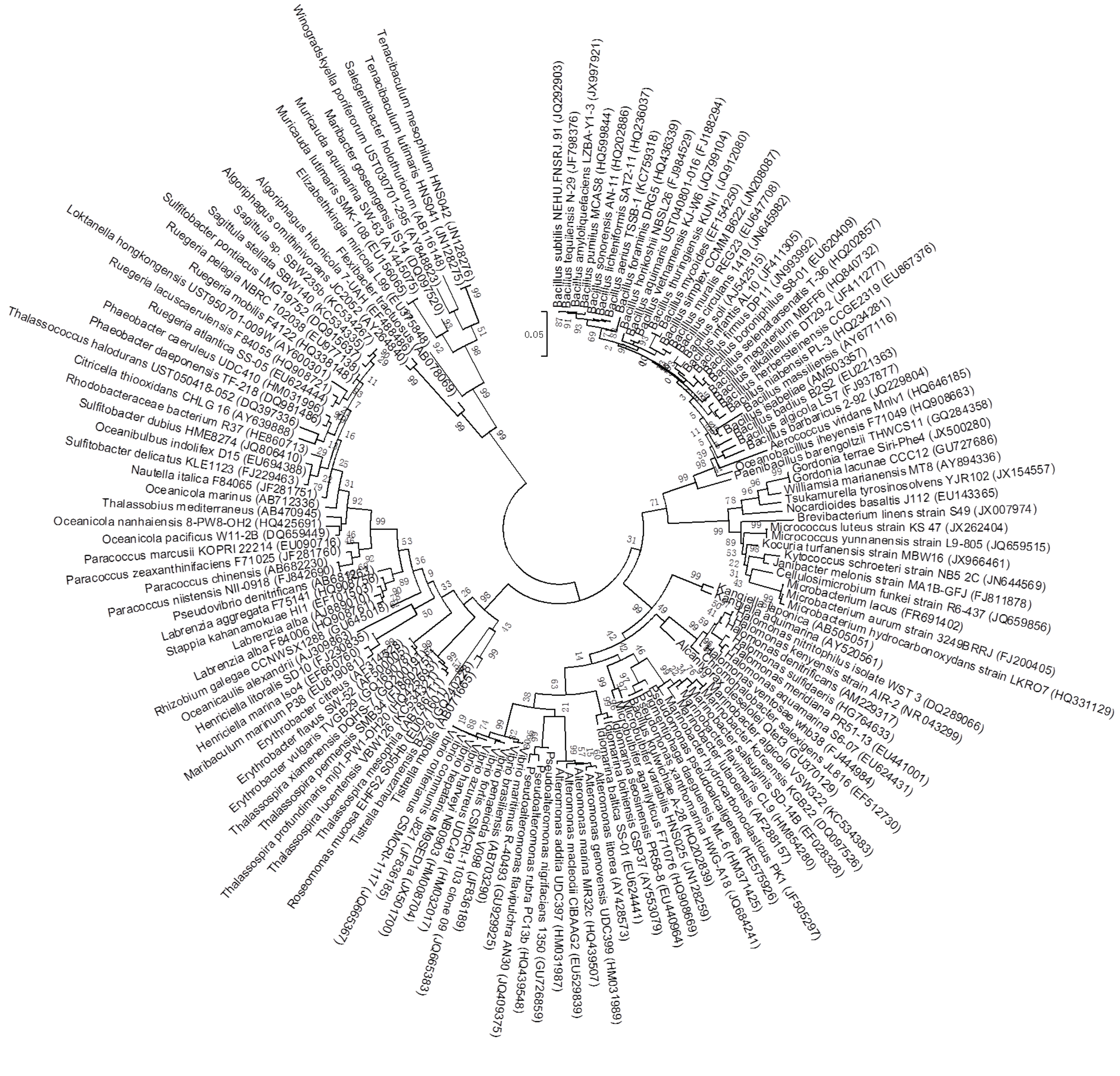
Table of Contents

Supplementary Figures

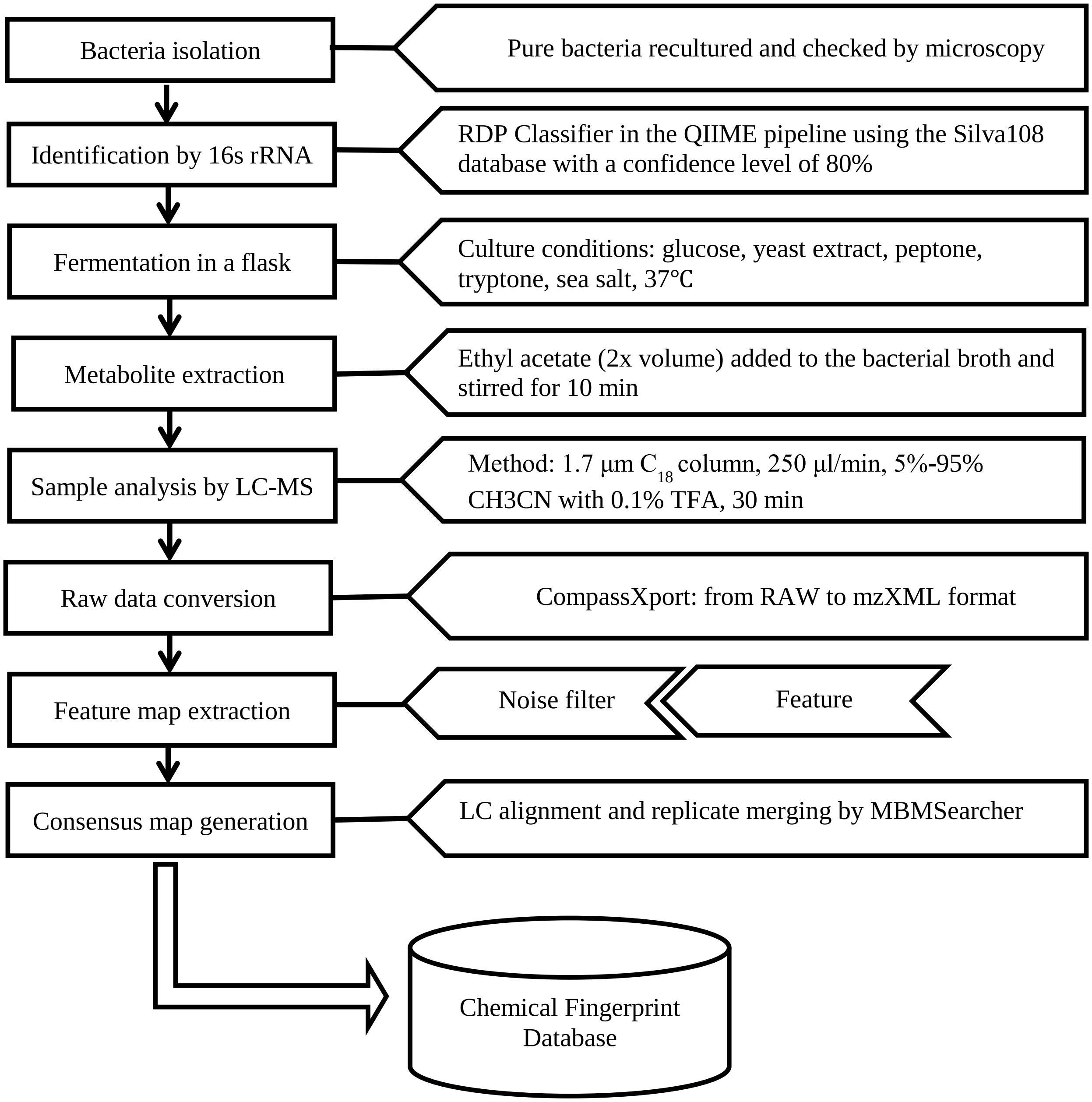
Supplementary Tables

Supplementary Method

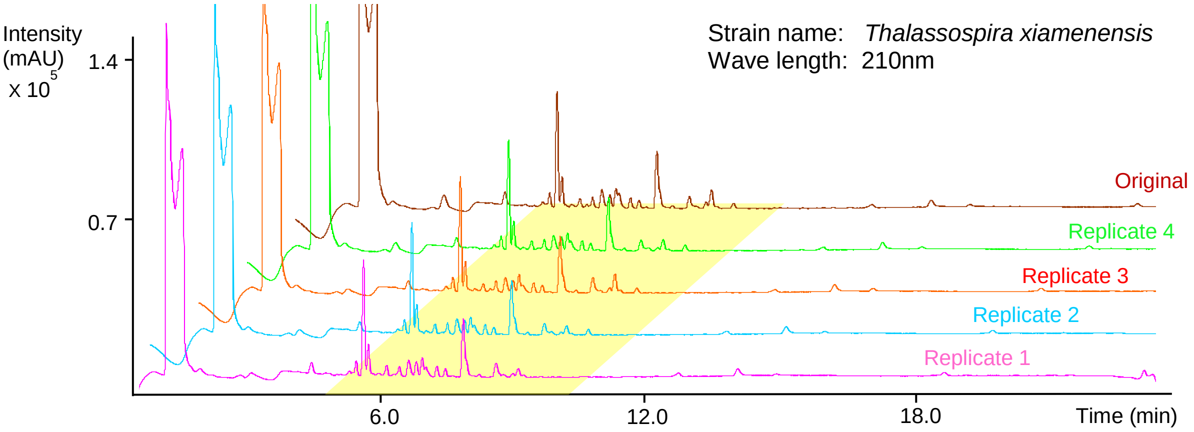
Supplementary Figures


Bacillus subtilis NEHU.FNSRJ.91 (JQ292903)
Bacillus tequilensis N-29 (JF798376)
Bacillus amyloliquefaciens LZBA-Y1-3 (JX997921)
Bacillus pumilus MCAS8 (HQ599844)
Bacillus sonorensis AN-11 (HQ202886)
Bacillus licheniformis SAT2-11 (HQ236037)
Bacillus aerius TSSB-1 (KC759318)
Bacillus foraminis DRG5 (HQ436339)
Bacillus horikoshii NBSL26 (FJ984529)
Bacillus aquimaris UST040801-016 (FJ188294)
Bacillus vietnamensis KJ-W6 (JQ799104)
Bacillus thuringiensis KUNi1 (JQ912080)
Bacillus mycoides (EF154250)
Bacillus simplex CCMM B622 (JN208087)
Bacillus muralis REG23 (EU647708)
Bacillus circulans 1419 (JN645982)
Bacillus soli (AJ542515)
Bacillus infantis AL10 (JF411305)
Bacillus firmus DP11 (JN993992)
Bacillus boroniphilus S8-01 (EU620409)
Bacillus selenatarsenatis T-36 (HQ202857)
Bacillus megaterium MBFF6 (HQ840732)
Bacillus alkalitelluris DT29-2 (JF411277)
Bacillus herbersteinensis CCGE2319 (EU867376)
Bacillus niabensis PL-3 (HQ234281)
Bacillus massiliensis (AY677116)
Bacillus isabeliae (AM503357)
Bacillus badius B2S2 (EU221363)
Bacillus algicola LS7 (FJ937877)
Bacillus barbaricus 2-92 (JQ229804)
Aerococcus viridans Mnlv1 (HQ646185)
Oceanobacillus iheyensis F71049 (HQ908663)
Paenibacillus barengoltzii THWCS11 (GQ284358)
Gordonia terrae Siri-Phe4 (JX500280)
Gordonia lacunae CCC12 (GU727686)
Williamsia marianensis MT8 (AY894336)
Tsukamurella tyrosinosolvens YJR102 (JX154557)
Nocardioides basaltis J112 (EU143365)
Brevibacterium linens strain S49 (JX007974)
Micrococcus luteus strain KS 47 (JX262404)
Micrococcus yunnanensis strain L9-805 (JQ659515)
Kocuria turfanensis strain MBW16 (JX966461)
Kytococcus schroeteri strain NB5 2C (JN644569)
Janibacter melonis strain MA1B-GFJ (FJ811878)
Cellulosimicrobium funkei strain R6-437 (JQ659856)
Microbacterium lacus (FR691402)
Microbacterium aurum strain 3249BRRJ (FJ200405)
Microbacterium hydrocarbonoxydans strain LKRO7 (HQ331129)
Kangiella japonica (AB505051)
Kangiella aquimarina (AY520561)
Halomonas nitritophilus isolate WST 3 (DQ289066)
Halomonas kenyensis strain AIR-2 (NR 043299)
Halomonas denitrificans (AM229317)
Halomonas sulfidaeris (HG764633)
Halomonas aquamarina PR51-13 (EU441001)
Halomonas meridiana S6-07 (EU624431)
Halomonas ventosae whb38 (FJ444984)
Chromohalobacter salexigens JL816 (EF512730)
Alcanivorax dieselolei Qtet3 (GU370129)
Marinobacter algicola VSW322 (KC534383)
Marinobacter koreensis KGB22 (DQ097526)
Marinobacter salsuginis SD-14B (EF028328)
Marinobacter flavimaris CL9 (HM854280)
Marinobacter lutaoensis (AF288157)
Marinobacter hydrocarbonoclasticus PK1 (JF505297)
Pseudomonas pseudoalcaligenes (HE575926)
Pigmentiphaga daeguensis ML-6 (HM371425)
Pseudomonas xanthomarina HWG-A18 (JQ684241)
Bacillus krulwichiae A-28 (HQ202839)
Microbulbifer variabilis HNS025 (JN128259)
Microbulbifer agarilyticus F71076 (HQ908669)
Microbulbifer seosinensis PR58-8 (EU440964)
Idiomarina loihiensis GSP37 (AY553079)
Idiomarina baltica SS-01 (EU624441)
Idiomarina litorea (AY428573)
Alteromonas genovensis UDC399 (HM031989)
Alteromonas marina MR32c (HQ439507)
Alteromonas macleodii CIBAAG2 (EU529839)
Alteromonas addita UDC397 (HM031987)
Pseudoalteromonas nigrifaciens 1350 (GU726859)
Pseudoalteromonas rubra PC13b (HQ439548)
Pseudoalteromonas flavipulchra AN30 (JQ409375)
Vibrio maritimus R-40493 (GU929925)
Vibrio brasiliensis (AB703290)
Vibrio penaeicida V098 (JF836189)
Vibrio fortis CSMCRI-1103 clone 09 (JQ665383)
Vibrio azureus UDC491 (HM032017)
Vibrio harveyi NB0903 (HM008704)
Vibrio hepatarius M9SED1a (JX501700)
Vibrio communis J821 (JF836185)
Vibrio rotiferianus CSMCRI-1117 (JQ665367)
Tistrella mobilis (AB071665)
Tistrella bauzanensis BZ78
Roseomonas mucosa EHFS2 S05Hb
Thalassospira mesophila
Thalassospira lucentensis VBW126
Thalassospira profundimaris mj01-PW1-OH20
Thalassospira permensis SMB34
Thalassospira xiamenensis DQHS-7
Erythrobacter vulgaris TVGB29
Erythrobacter flavus SW-52
Erythrobacter citreus
Maribaculum marinum P38 (EU819081)
Henriciella marina Iso4 (EF660760)
Henriciella litoralis SD10 (FJ230835)
Oceanicaulis alexandrii (AJ309863)
Rhizobium galegae CCNWSX1288 (GU645018)
Labrenzia alba F84006 (HQ908761)
Labrenzia alba (AJ889010)
Stappia kahanamokuae HI1 (EF101503)
Labrenzia aggregata F75141 (HQ908756)
Pseudovibrio denitrificans (AB681263)
Paracoccus niistensis NII-0918 (FJ842690)
Paracoccus chinensis (AB682230)
Paracoccus zeaxanthinifaciens F71025 (JF281760)
Paracoccus marcusii KOPRI 22214 (EU090716)
Oceanicola pacificus W11-2B (DQ659449)
Oceanicola nanhaiensis 8-PW8-OH2 (HQ425691)
Thalassobius mediterraneus (AB470945)
Oceanicola marinus (AB712336)
Nautella italica F84065 (JF281751)
Sulfitobacter delicatus KLE1123 (FJ229463)
Oceanibulbus indolifex D15 (EU694388)
Sulfitobacter dubius HME8274 (JQ806410)
Rhodobacteraceae bacterium R37 (HE860713)
Citricella thiooxidans CHLG 16 (AY639888)
Thalassococcus halodurans UST050418-052 (DQ397336)
Phaeobacter daeponensis TF-218 (DQ981486)
Phaeobacter caeruleus UDC410 (HM031996)
Ruegeria atlantica SS-05 (EU624444)
Ruegeria lacuscaerulensis F84055 (HQ908721)
Loktanella hongkongensis UST950701-009W (AY600301)
Ruegeria mobilis F4122 (HQ338148)
Ruegeria pelagia NBRC 102038 (EU977138)
Sulfitobacter pontiacus LMG19752 (DQ915637)
Sagittula stellata SBW140 (KC534335)
Sagittula sp. SBW235b (KC534267)
Algoriphagus ornithinivorans JC2052 (AY264840)
Algoriphagus hitonicola 7-UAH (EF488486)
Flexibacter tractuosus (AB078069)
Elizabethkingia miricola L99 (EU375848)
Muricauda lutimaris SMK-108 (EU156065)
Muricauda aquimarina SW-63 (AY445075)
Maribacter goseongensis IS14 (DQ097520)
Winogradskyella poriferorum UST030701-295 (AY848823)
Salegentibacter holothuriorum (AB116148)
Tenacibaculum lutimaris HNS041 (JN128275)
Tenacibaculum mesophilum HNS042 (JN128276)
0.05

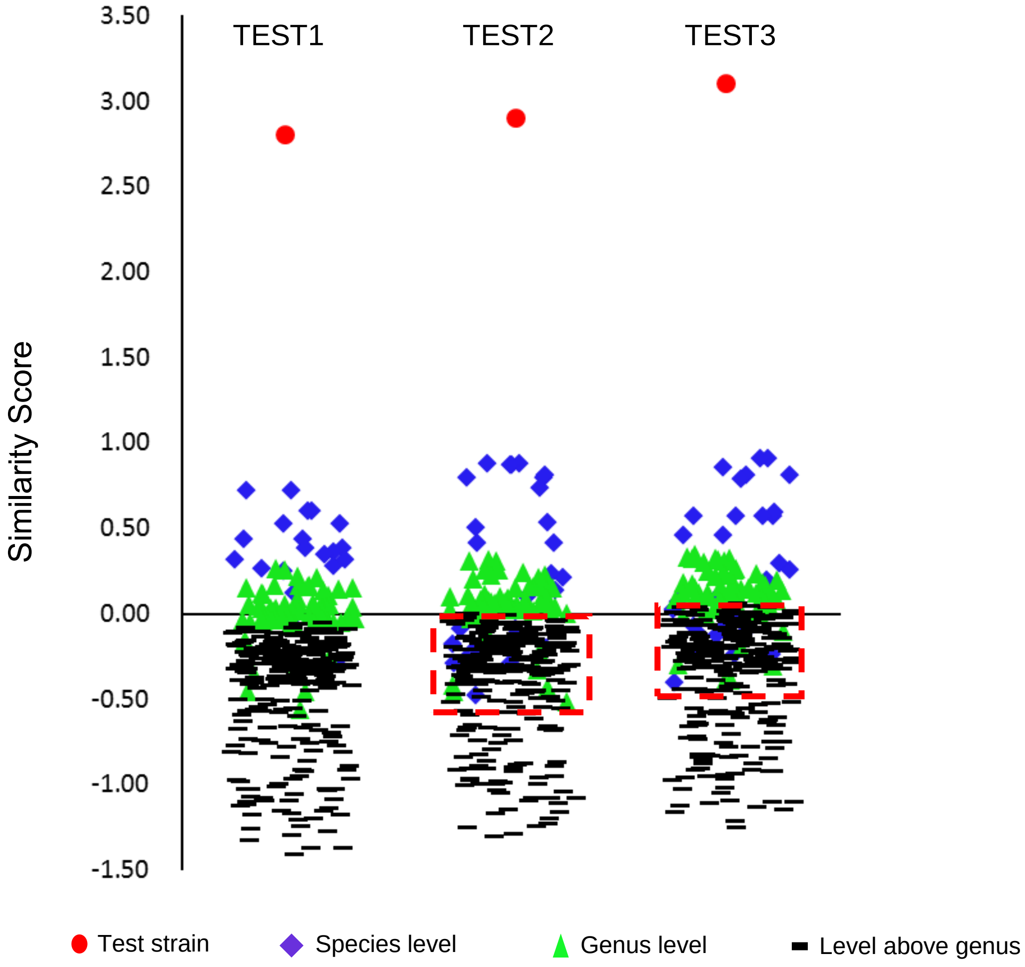
Supplementary Figure 1 | Phylogenetic relationships of known species in marine bacterial metabolite database by their 16s rRNA sequences.


Bacteria isolation
Pure bacteria recultured and checked by microscopy
Identification by 16s rRNA
RDP Classifier in the QIIME pipeline using the Silva108 database with a confidence level of 80%
Fermentation in a flask
Culture conditions: glucose, yeast extract, peptone, tryptone, sea salt, 37℃
Metabolite extraction
Ethyl acetate (2x volume) added to the bacterial broth and stirred for 10 min
Sample analysis by LC-MS
Method: 1.7 µm C_{18} column, 250 µl/min, 5%-95% CH3CN with 0.1% TFA, 30 min
Raw data conversion
CompassXport: from RAW to mzXML format
Feature map extraction
Noise filter
Feature
Consensus map generation
LC alignment and replicate merging by MBMSearcher
Chemical Fingerprint Database

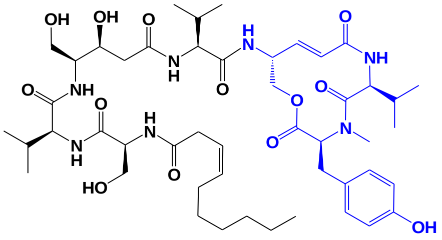
Supplementary Figure 2 | Flowchart of establishing a chemical fingerprint database.



Supplementary Figure 3 | Representative UPLC chromatograms of 5 biological replicates (*Thalassospira xiamenensis* strain) measured by UV detector at the wavelength of 210nm. The UPLC chromatograms overlapped very well, indicating good reproducibility. Most metabolite signals were found within the retention time range of 5 ~ 10 minutes (indicated by light yellow).

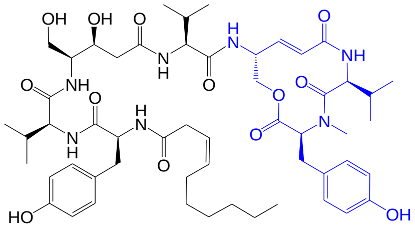


Supplementary Figure 4 | Similarity of secondary metabolite profiles of 3 test *Bacillus subtilis* strains and all other strains in the chemical fingerprint library. Three *Bacillus subtilis* strains isolated from marine environment were randomly selected to evaluate the similarity scoring function of our software against the entire database. The 3 test queries showed that the similarity scores against strains from the same species (blue diamonds) are generally higher than those against strains from the same genus but not the same species (green triangles), which in turn are higher than those from bacteria in different genera (black rectangles). However, some exceptions can be found (indicated in the red box). This result further supported our hypothesis that although similarity of secondary metabolite profiles are roughly correlated with taxonomical similarity (based on 16s rRNA), species identification based on 16s rRNA does not always adequately predict the secondary metabolite repertoire of a given bacterial strain.



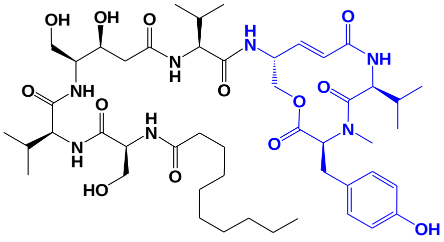
thalassospiramide A

MW: 957.6, RT: 13.5 min



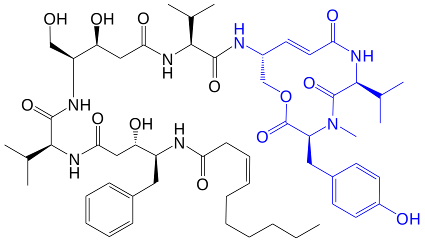
thalassospiramide A1

MW: 1033.6, RT: 14.7 min



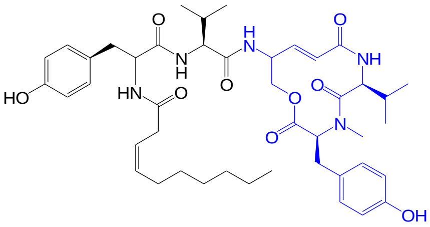
thalassospiramide A2

MW: 959.6, RT: 13.6 min



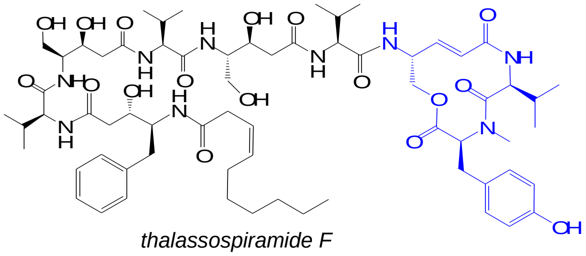
thalassospiramide B

MW: 1061.6, RT: 15.1 min



thalassospiramide C

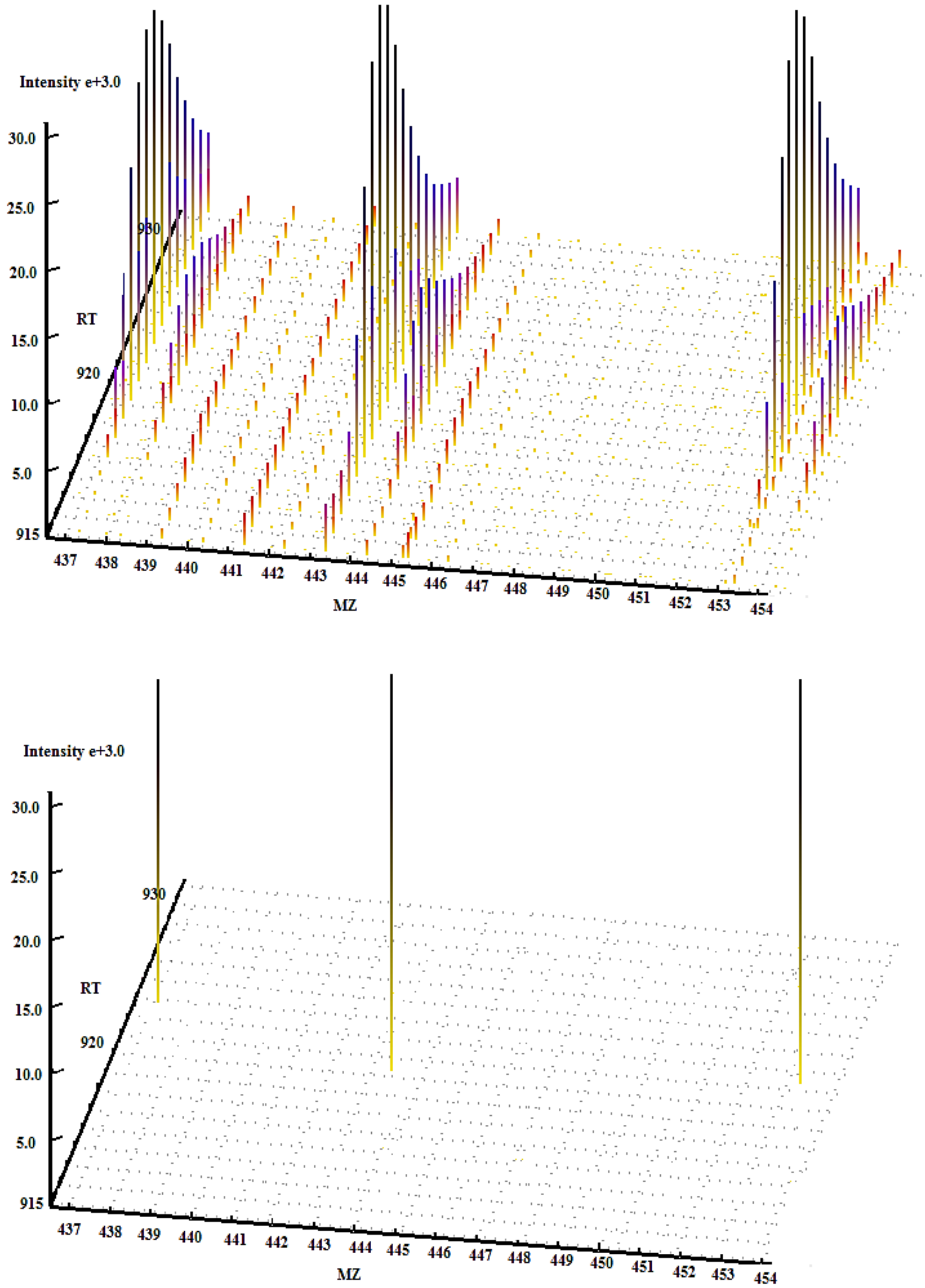
MW: 803.4, RT: 16.2 min



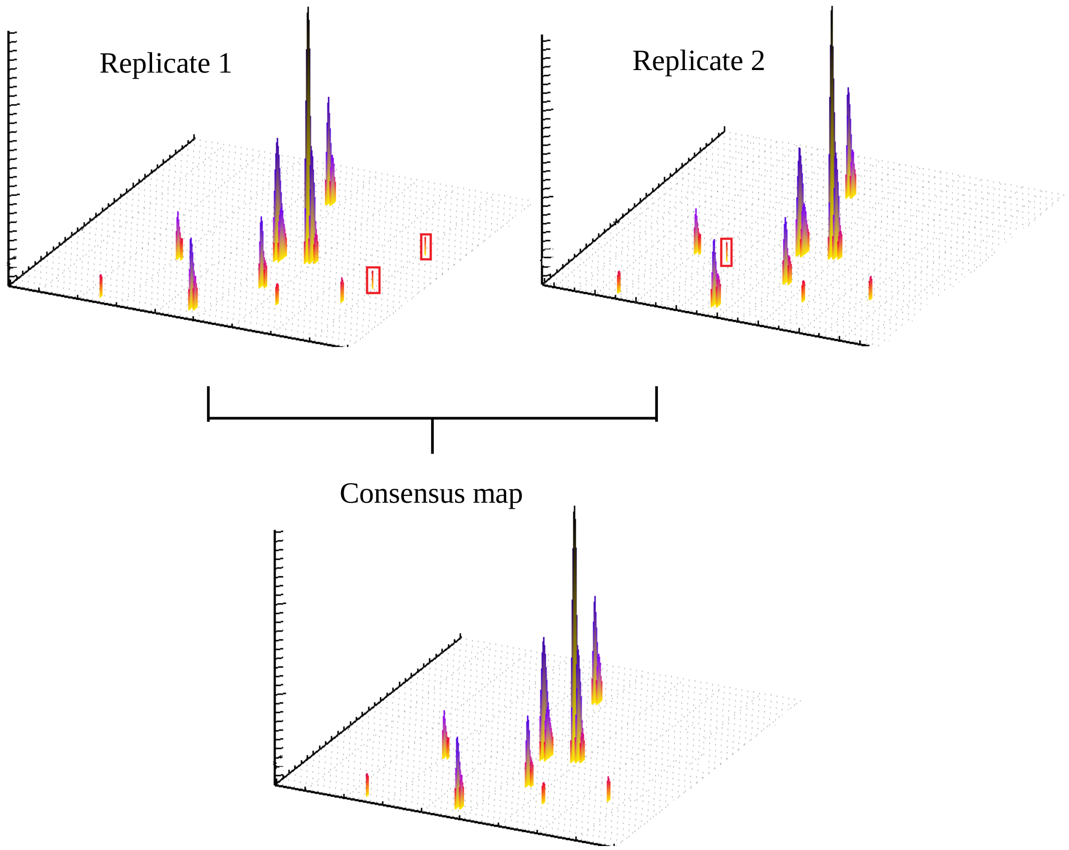
thalassospiramide F

MW: 1291.7, RT: 14.0 min

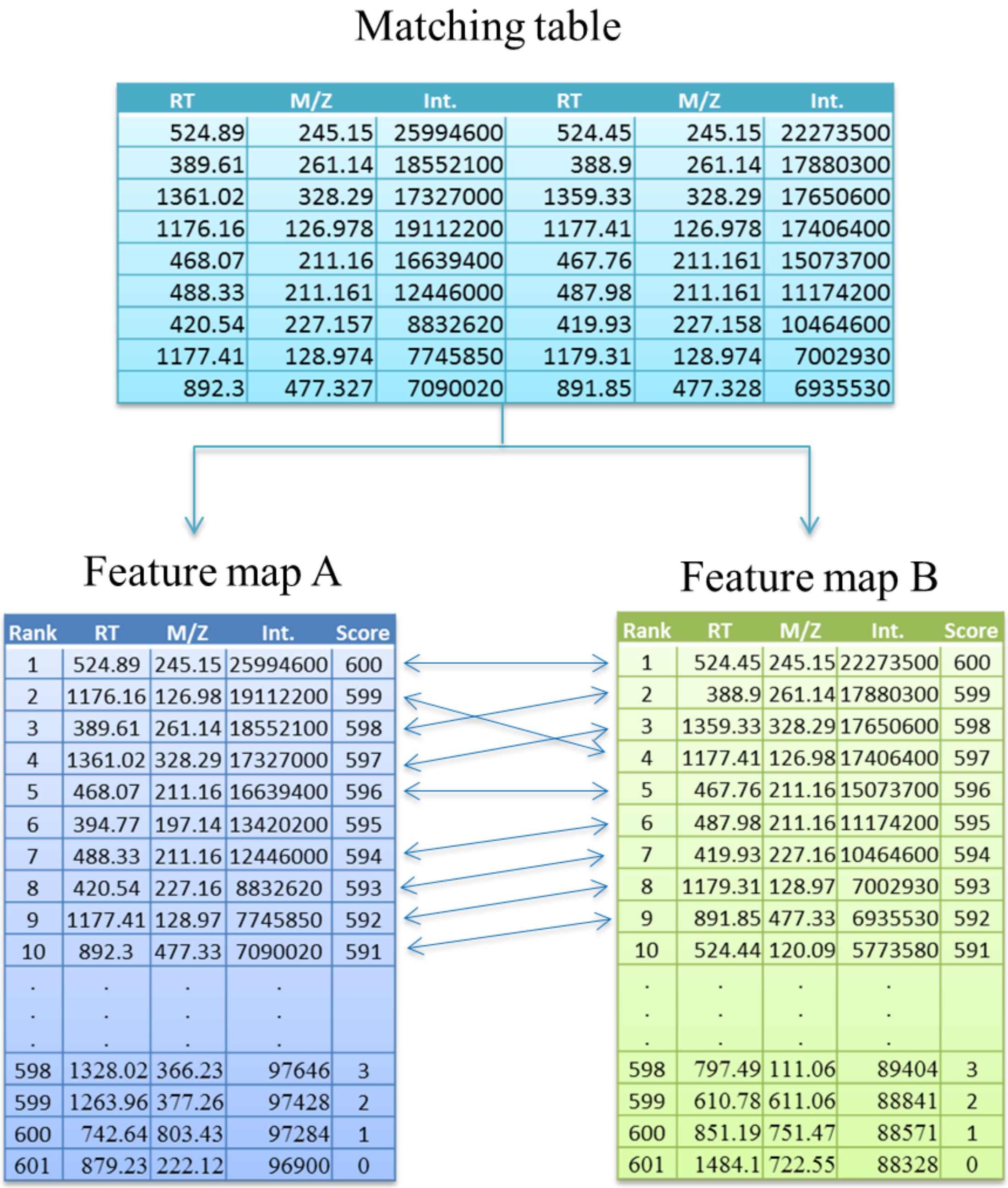
Supplementary Figure 5 | Chemical structure and retention time of *thalassospiramide A, A1, A2, B, C and F*. The common 12-membered ring structure indicated in blue.



Supplementary Figure 6 | Example of feature detection from LC-MS profile data. Three features (bottom panel) located at (437.1, 924s), (443.3, 922s) and (453.2, 923s) were detected from the LC-MS profile data (top panel) by the FeatureFinder function in *OpenMS*.



Supplementary Figure 7 | Example of a consensus LC-MS map merged from 2 technical replicates. The feature maps of each pair of technical replicates were first aligned using LWBMatch. All of the shared features were included in a consensus map (unshared peaks are labeled in red).



Matching table

RT	M/Z	Int.	RT	M/Z	Int.
524.89	245.15	25994600	524.45	245.15	22273500
389.61	261.14	18552100	388.9	261.14	17880300
1361.02	328.29	17327000	1359.33	328.29	17650600
1176.16	126.978	19112200	1177.41	126.978	17406400
468.07	211.16	16639400	467.76	211.161	15073700
488.33	211.161	12446000	487.98	211.161	11174200
420.54	227.157	8832620	419.93	227.158	10464600
1177.41	128.974	7745850	1179.31	128.974	7002930
892.3	477.327	7090020	891.85	477.328	6935530

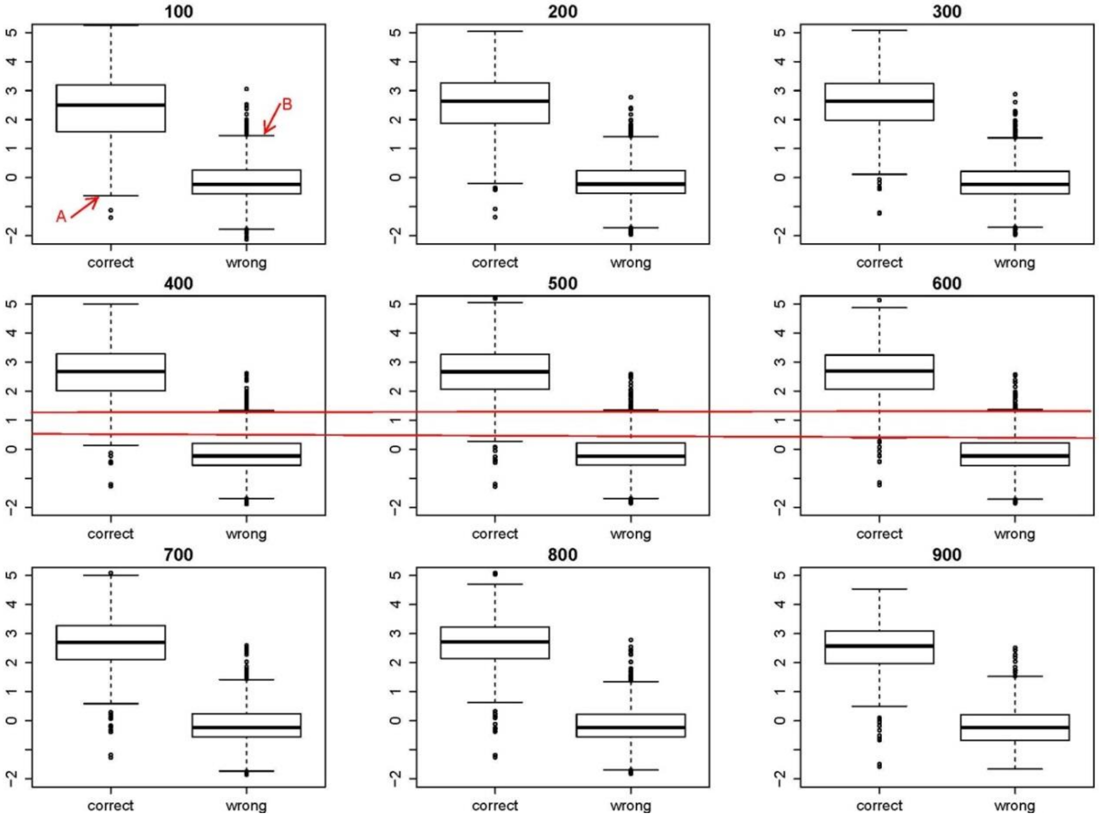
Feature map A

Rank	RT	M/Z	Int.	Score
1	524.89	245.15	25994600	600
2	1176.16	126.98	19112200	599
3	389.61	261.14	18552100	598
4	1361.02	328.29	17327000	597
5	468.07	211.16	16639400	596
6	394.77	197.14	13420200	595
7	488.33	211.16	12446000	594
8	420.54	227.16	8832620	593
9	1177.41	128.97	7745850	592
10	892.3	477.33	7090020	591
.	.	.	.	
.	.	.	.	
.	.	.	.	
598	1328.02	366.23	97646	3
599	1263.96	377.26	97428	2
600	742.64	803.43	97284	1
601	879.23	222.12	96900	0

Feature map B

Rank	RT	M/Z	Int.	Score
1	524.45	245.15	22273500	600
2	388.9	261.14	17880300	599
3	1359.33	328.29	17650600	598
4	1177.41	126.98	17406400	597
5	467.76	211.16	15073700	596
6	487.98	211.16	11174200	595
7	419.93	227.16	10464600	594
8	1179.31	128.97	7002930	593
9	891.85	477.33	6935530	592
10	524.44	120.09	5773580	591
.	.	.	.	
.	.	.	.	
.	.	.	.	
598	797.49	111.06	89404	3
599	610.78	611.06	88841	2
600	851.19	751.47	88571	1
601	1484.1	722.55	88328	0

Supplementary Figure 8 | Example of a feature matching table of two LC-MS maps, illustrating the calculation of the rank-transform dot product. The features in each feature map will first be sorted based on their signal intensities. We selected the top 600 features by intensity and assigned scores to them (from 600 to 1). Feature-to-feature mappings were built on the matching table. In this example, the rank-transform dot product between Feature map A and B will be calculated as: Rank-transform dot product = 600*600 + 599*597 + 598*599 + 597*598 + 596*596 + 594*595 + 593*594 + 592*593 + 591*592 = 2836425



Supplementary Figure 9 | Effect of the number of retained features on the score separation between correct and incorrect hits. To optimize the number of retained features for similarity scoring, 5 replicates each of 4 *Thalassospira xiamenensis* and 4 *Thalassospira profundimaris* strains were selected, and the pairwise similarity scores are calculated for all replicates. At different numbers of retained features (n = 100, 200, …, 900), the distribution of similarity scores are plotted as box-and-whisker diagrams separately for correct hits (pairs that belong to the same strains) and incorrect hits (pairs that belong to different strains). The bottom and top of the box represent the first and third quartiles, respectively, and the band inside the box is the second quartile. The ends of the whiskers below and above each box indicate the lowest and highest datum within 1.5 x interquartile range (IQR) of each quartile, respectively. Any data not included between whiskers are plotted as an outlier (small circle). We calculated the overlap between boundaries A and B for each diagram: Overlap (100) = 2.08, Overlap(200) = 1.68, Overlap(300) = 1.35, Overlap(400) = 1.16, Overlap(500) = 1.06, Overlap(600) = 0.82, Overlap(700) = 0.85, Overlap(800) = 0.95, Overlap(900) = 1.10. When *n* = 600, the best separation between correct and incorrect hits is achieved (marked by two red lines in the middle rows). Therefore, the default number of retained features is set to 600 in *MBMSearcher*.

Supplementary Tables

Supplementary Table 1 | Species name of identified marine bacteria in the database

Strain name	Strain name	Strain name	Strain name
Acremonium murorum	*Brevibacterium linens*	*Marinobacter salsuginis*	*Sagittula sp.*
Aerococcus viridans	*Cellulosimicrobium cellulans*	*Mesorhizobium sp.*	*Sagittula stellata*
Alcanivorax dieselolei	*Cellulosimicrobium funkei*	*Microbacterium aurum*	*Salegentibacter holothuriorum*
Algoriphagus hitonicola	*Chromohalobacter salexigens*	*Microbacterium hydrocarbonoxydans*	*Salinisphaera hydrothermalis*
Algoriphagus ornithinivorans	*Citreicella thiooxidans*	*Microbacterium lacus*	*Salinisphaera shabanensis*
Alteromonas addita	*Elizabethkingia miricola*	*Microbulbifer agarilyticus*	*Salinisphaera sp.*
Alteromonas genovensis	*Erythrobacter citreus*	*Microbulbifer variabilis*	*Staphylococcus epidermidis*
Alteromonas litorea	*Erythrobacter flavus*	*Micrococcus luteus*	*Stappia alba*
Alteromonas macleodii	*Erythrobacter vulgaris*	*Micrococcus yunnanensis*	*Stappia kahanamokuae*
Alteromonas marina	*Flexibacter tractuosus*	*Muricauda aquimarina*	*Sulfitobacter delicatus*
Aspergillus nidulans	*Gordonia lacunae*	*Muricauda lutimaris*	*Sulfitobacter dubius*
Bacillus aerius	*Gordonia terrae*	*Muricauda ruestringensis*	*Sulfitobacter litoralis*
Bacillus algicola	*Halomonas aquamarina*	*Nautella italica*	*Sulfitobacter pontiacus*
Bacillus alkalitelluris	*Halomonas denitrificans*	*Nocardioides basaltis*	*Tenacibaculum lutimaris*
Bacillus amyloliquefaciens	*Halomonas kenyensis*	*Nodulisporium sp.*	*Tenacibaculum mesophilum*
Bacillus aquimaris	*Halomonas meridian*	*Oceanibulbus indolifex*	*Thalassobius mediterraneus*
Bacillus badius	*Halomonas nitritophilus*	*Oceanicaulis alexandrii*	*Thalassococcus halodurans*
Bacillus barbaricus	*Halomonas sulfidaeris*	*Oceanicola marinus*	*Thalassospira lucentensis*
Bacillus boroniphilus	*Halomonas ventosae*	*Oceanicola nanhaiensis*	*Thalassospira permensis*
Bacillus circulans	*Henriciella litoralis*	*Oceanicola pacificus*	*Thalassospira profundimaris*
Bacillus firmus	*Henriciella marina*	*Oceanobacillus iheyensis*	*Thalassospira tepidiphila*

Bacillus foraminis	*Hypocrea jecorina*	*Paenibacillus barengoltzii*	*Thalassospira TrichSKD10*
Bacillus herbersteinensis	*Idiomarina baltica*	*Paracoccus chinensis*	*Thalassospira xiamenensis*
Bacillus horikoshii	*Idiomarina loihiensis*	*Paracoccus marcusii*	*Thanatephorus cucumeris*
Bacillus infantis	*Idiomarina seosinensis*	*Paracoccus niistensis*	*Tistrella bauzanensis*
Bacillus isabeliae	*Janibacter melonis*	*Paracoccus zeaxanthinifaciens*	*Tistrella mobilis*
Bacillus krulwichiae	*Kangiella aquimarina*	*Phaeobacter caeruleus*	*Tsukamurella tyrosinosolvens*
Bacillus licheniformis	*Kangiella japonica*	*Phaeobacter daeponensis*	*Vibrio atypicus*
Bacillus massiliensis	*Kocuria turfanensis*	*Pigmentiphaga daeguensis*	*Vibrio azureus*
Bacillus megaterium	*Kytococcus schroeteri*	*Pseudoalteromonas flavipulchra*	*Vibrio brasiliensis*
Bacillus muralis	*Labrenzia aggregata*	*Pseudoalteromonas nigrifaciens*	*Vibrio communis*
Bacillus mycoides	*Labrenzia alba*	*Pseudoalteromonas rubra*	*Vibrio fortis*
Bacillus niabensis	*Loktanella hongkongensis*	*Pseudomonas pseudoalcaligenes*	*Vibrio harveyi*
Bacillus pumilus	*Lysobacter sp.*	*Pseudomonas xanthomarina*	*Vibrio hepatarius*
Bacillus selenatarsenatis	*Maribacter goseongensis*	*Pseudovibrio denitrificans*	*Vibrio maritimus*
Bacillus simplex	*Maribaculum marinum*	*Rhizobium galegae*	*Vibrio mediterranei*
Bacillus soli	*Marinibacillus marinus*	*Rhodobacteraceae bacterium*	*Vibrio penaeicida*
Bacillus sonorensis	*Marinobacter algicola*	*Roseomonas mucosa*	*Vibrio rotiferianus*
Bacillus subtilis	*Marinobacter flavimaris*	*Ruegeria atlantica*	*Vibrio shilonii*
Bacillus tequilensis	*Marinobacter hydrocarbonoclasticus*	*Ruegeria lacuscaerulensis*	*VWinogradskyella poriferorum*
Bacillus thuringiensis	*Marinobacter koreensis*	*Ruegeria mobilis*	*Williamsia marianensis*
Bacillus vietnamensis	*Marinobacter lutaoensis*	*Ruegeria pelagia*	*Winogradskyella poriferorum*

Supplementary Table 2 | Common signals extraction for 3 bacterial species.

Bacillus subtilis

Signal ID	Rt (min)	m/z
1	8.74	505.34
2	12.58	382.27
3	13.74	328.22
4	15.52	356.39
5	16.57	317.36
6	16.57	658.43
7	16.57	663.46
8	16.59	299.32
9	16.59	440.34
10	18.26	299.31
11	18.42	493.35
12	21.98	614.53
13	24.92	803.59

Thalassospira xiamenesis

Signal ID	Rt (min)	m/z	Signal ID	Rt (min)	m/z
1	3.33	251.15	33	8.68	213.16
2	4.06	169.10	34	9.19	231.11
3	6.03	231.12	35	9.20	286.16
4	6.29	361.05	36	9.30	243.09
5	6.30	393.06	37	9.73	227.18
6	6.30	233.13	38	10.50	385.15
7	6.31	521.24	39	16.13	808.49
8	6.31	283.11	40	17.58	247.17
9	6.32	648.11	41	17.94	399.36
10	6.36	375.07	42	17.99	796.55
11	6.36	342.05	43	18.08	752.52
12	6.57	227.14	44	18.08	757.48
13	6.70	316.21	45	18.16	708.50
14	6.79	359.08	46	18.26	669.42
15	6.83	263.11	47	18.35	625.40
16	7.25	277.16	48	18.58	287.27
17	7.36	211.15	49	19.49	219.99
18	7.56	211.15	50	19.50	279.16
19	7.56	389.09	51	19.51	235.03

20	7.57	511.18	52	19.55	396.37
21	7.57	356.07	53	19.67	384.37
22	7.88	566.21	54	19.95	554.45
23	7.88	511.18	55	21.56	876.67
24	7.88	513.18	56	22.15	512.51
25	7.88	311.08	57	22.49	673.54
26	7.89	356.07	58	22.64	953.69
27	7.89	548.15	59	23.38	793.57
28	8.09	245.13	60	23.63	810.57
29	8.10	255.17	61	23.74	267.16
30	8.51	335.03	62	24.71	605.43
31	8.51	579.15	63	24.73	649.46
32	8.51	581.15			

Tistrella mobilis

Signal ID	Rt (min)	m/z	Signal ID	Rt (min)	m/z
1	6.30	327.20	19	11.23	217.05
2	6.30	349.19	20	11.62	219.17
3	6.61	371.23	21	16.64	159.13
4	6.88	437.24	22	17.86	173.15
5	6.88	432.28	23	17.93	211.15
6	6.88	415.26	24	18.18	533.33
7	7.11	459.28	25	18.19	267.17
8	7.11	476.31	26	18.33	506.40
9	7.11	481.26	27	18.56	713.44
10	7.31	520.34	28	18.59	496.34
11	7.31	503.31	29	18.80	576.41
12	7.49	564.36	30	18.88	532.38
13	7.50	569.32	31	19.32	522.35
14	7.65	608.39	32	19.64	279.16
15	7.79	652.42	33	22.48	646.42
16	7.93	696.44	34	23.00	256.26
17	8.05	740.47	35	23.75	776.23
18	8.59	430.25	36	24.06	270.28

Supplementary Table 3 | Search result of *Thalassospira sp.TrichSKD10* against the database. The top 5 strains (most similar by secondary metabolites profiles) are highlighted in gray.

Rank	Strain ID	Strain name
1	MarineB0718	*Thalassospira xiamenensis*
2	MarineB0711	*Thalassospira lucentensis*
3	MarineB0701	*Thalassospira xiamenensis*
4	MarineB0717	*Thalassospira xiamenensis*
5	MarineB0685	*Thalassospira profundimaris*
6	MarineB0691	*Thalassospira lucentensis*
7	MarineB0694	*Thalassospira profundimaris*
8	MarineB0676	*Thalassospira profundimaris*
9	MarineB0703	*Thalassospira xiamenensis*
10	MarineB0678	*Thalassospira profundimaris*
11	MarineB0728	*Thalassospira xiamenensis*
12	MarineB0801	*Pseudovibrio denitrificans*
13	MarineB0462	*Thalassospira xiamenensis*
14	MarineB0463	*Thalassospira xiamenensis*
15	MarineB0729	*Tistrella bauzanensis*
16	MarineB0468	*Thalassospira xiamenensis*
17	MarineB0473	*Thalassospira sp.*
18	MarineB0468	*Thalassospira xiamenensis*
19	MarineB0802	*Tistrella Mobilis*
20	MarineB0492	*Thalassospira sp.*

Supplementary Method

Definition of the normalized rank-transform dot product as the similarity score between 2 LC-MS feature maps

MBMSearcher uses a rank-transform dot product to evaluate the similarity of two LC-MS feature maps after alignment. The features of each feature map were first sorted from highest to lowest based on their signal intensities to form a vector. We selected the top 600 dominant features and assigned scores. **Supplementary Fig. 9** shows that selection of the top 600 features provided a better separation between correct and incorrect hits.

The rank-transform dot product between 2 feature maps (*ref* and *sam*) was calculated as follows:

$$\mathrm{RankDot}(ref, sam) = \sum_{i=1}^{600} Score(ref_i) \times Score(Match(ref_i))$$

where ref_i refers to the i_{th} feature in ref_i and $Match(ref_i)$ is the corresponding feature in *sam*, which is aligned with ref_i in the matching table. An example is shown in **Supplementary Fig. 8**. To normalize the rank-transform dot product, we calculated the sample mean (Mean (*ref*)) and standard deviation (SD(*ref*)) as follows :

$$\mathrm{Mean}(ref) = \frac{\sum_{i=1}^{n} \mathrm{RankDot}(ref, b_i)}{n}$$

$$\mathrm{SD}(ref) = \sqrt{\frac{\sum_{i=1}^{n}\left(\mathrm{RankDot}(ref, db_i) - \mathrm{Mean}(ref)\right)^2}{n}}$$

where *n* is the number of all feature maps stored in the bacterial database *db,* and db_i refers to the *i*th feature map in the database. The normalized dot product is defined as follows:

$$\mathrm{NRankDot}(ref, sam) = \frac{\mathrm{RankDot}(ref,sam) - \mathrm{Mean}(ref)}{\mathrm{SD}(ref)}.$$

We used $\mathrm{NRankDot}(ref, sam)$ as a measure of the similarity between 2 feature maps.