Phylogenomics of Rhodobacteraceae Reveals Evolutionary Adaptation to Marine and Non-Marine Habitats

Meinhard Simon, Carmen Scheuner, Jan P. Meier-Kolthoff, Thorsten Brinkhoff, Irene Wagner-Döbler, Marcus Ulbrich, Hans-Peter Klenk, Dietmar Schomburg, Jörn Petersen, Markus Göker

Supplementary file 3

A. Detailed description of the habitat assignment (marine or saline vs. non-marine) based on the isolation sources

The habitat assignment was conducted based on the information on the isolation source (see Supplementary File 1). The reliability of these assignments was assessed by comparing the inferred habitat with the reported NaCl requirements, if any. Growth on NaCl was reported in the literature on almost all type strains included in our sampling (they comprise roughly 50% of all strains investigated). Additionally, 16S rNA gene sequences of all 106 strains were queried with blastn against the EMBL environmental database em_geo_rel_env, which contains geographic locations (Lopez et al. 2014). Only hits with a sequence identity of >= 99% were kept. The resulting hits for each query strain, if any, where subsequently plotted on a world map with the R package ggmap (Kahle and Wickham 2013).

Marine and saline habitats were distinguished from non-marine habitats. A marine/saline habitat was assigned when the strain was isolated from a marine source or a source with a salt concentration comparable to the marine environment. Marine isolation sources were mainly water samples, but also biofilms, sea ice, sediment or other organisms (e.g. algae, dinoflagellates, sponges, clams). Examples for habitats equivalent to the marine one include *Nesiotobacter exalbescens* DSM 16456^{T}, which was collected from water of a hypersaline lake on the uninhabited Laysan Atoll in the Northwestern Hawaiian Islands (Donachie et al. 2006), and *Sediminimonas qiaohouensis* DSM 21189^{T}, which was collected from an ancient salt sediment at the Qiaohou salt mine in Yunnan, south-west China (Wang et al. 2009). Both strains require NaCl for optimal growth (Donachie et al. 2006; Wang et al. 2009). Similar results were reported for *Ruegeria lacuscaerulensis* ITI-1157^{T}, which was isolated from a silica-rich geothermal lake, the Blue Lagoon in Iceland (Petursdottir and Kristjansson 1997). Although it was not isolated from a marine habitat, this strain does not grow without NaCl, and the optimal salinity for its growth is 3.5% - 4% NaCl (Petursdottir and Kristjansson 1997). *Maritimibacter* sp. HL-12 and *Oceanicola* sp. HL-35 were isolated from a microbial mat of Hot Lake, Washington (Lindemann et al. 2013). Hot Lake is a heliothermal hypersaline lake subject to great variation in salinity over the annual cycle. However, the salinity is dominated by magnesium sulfate rather than sodium chloride (Lindemann et al. 2013). The prevailing mixolimnion salinity at collection was 117.9 g/L for HL-12 and 134.7 g/L for HL-35 [S. Lindemann, personal communication].

Strains with non-marine habitats originated mainly from soil and compost samples. Strains for which the natural habitat was unknown were considered regarding their salt requirements. For example, *Ketogulonicigenium vulgare* WSH-001 and *Ketogulonicigenium vulgare* Y25 are industrial strains which can produce 2-keto-L-gulonic acid from L-sorbose, a key intermediate in the synthesis of vitamin C. The natural habitat of these two strains could not be traced back. However, other strains of the genus *Ketogulonicigenium* were isolated from soil samples and are reported to grow optimally without NaCl (Urbance et al. 2001). Similarly, the natural habitat of *Rubellimicrobium thermophilum* DSM 16684^T, which was isolated from coloured deposits (biofilm) in a pulp dryer in Finland (Denner et al. 2006), is unknown so far (Fiebig et al. 2013). However, other *Rubellimicrobium* species are reported to have low salt tolerance, that is, < 1% NaCl (Weon et al. 2009). Another strain with unknown origin is *Rhodobacter sphaeroides* 2.4.1^T, for which an optimal growth is reported in the absence of NaCl (Arunasri et al. 2008; Ramana et al. 2008).

The strains *Rhodobacter sphaeroides* ATCC 17025 and *Rhodobacter sphaeroides* ATCC 17029 were already isolated by C. van Niel. Neither their source of isolation (Choudhary et al. 2007) nor their salt requirements could be identified. However, since optimal growth was reported in the absence of NaCl for the type strain (see above), these strains probably also prefer low salt concentration.

The comparison of the inferred habitat assignments with the reported NaCl requirements uniformly confirms the assignments, as all marine/saline strains which have been tested on NaCl show either a dependency on NaCl or at least an optimum that is close to fully marine (3%) or brackish (1%) salt concentrations, and vice versa. Likewise, the BLAST search against the em_geo_rel_env database (see the appendix at the end of this document) either yielded no hits at all (and thus no reason to presume the respective strain occurs in other habitats than the one it had been isolated from) or hits that corresponded to the inferred habitat. In some cases hits of a marine organism to a non-marine environment were reported but closer examination revealed those as saline. For example, the hit of *Loktanella vestfoldensis* DSM 16212^T was located in a polysaline lake in central China. The marine query *Roseovarius* sp. TM1035 hit the Spanish Tabla de Daimiel National Park, a wetland on the La Mancha plain. Here, two rivers – salt water and fresh water – cause recurrent seasonal inundations of the area, thus explaining the high similarity to a marine organism.

Whereas its isolation source needs not in general correspond to the preferred habitat of a strain, these comparisons indicate that, when simplified to just two options, the habitat assignments found in the literature on roseobacters are highly reliable. This conclusion is re-enforced by the results from our phylogenetic analyses, which indicate that the habitat is phylogenetically conserved (see the main text and below).

B. Results of the maximum likelihood (ML) and maximum-parsimony (MP) analyses for the different supermatrices described in the text

General characteristics of the character matrices constructed with all strains included and the included and trees inferred from them. The score is the log likelihood for the ML analyses, the number of steps for MP analysis (not counting uninformative characters). For each matrix and inference method, two analyses were conducted, one without and one with removal of the outgroup strains. For ExaML and TNT, both normal (above) and partition bootstrapping (below) were conducted. Abbreviations used: MCG, most conserved genes; BS, bootstrap support; X, omitted for reasons of running time.

Matrix	# Genes	# Characters	Analysis	Model	Score	Average BS
50 MCG	50	18,383	RAxML	PROTCATLGF	-600,680.35	82.58
			ExaML	-	-572,682.46	77.82
						78.52
			TNT	-	99,581	88.99
						79.64
	[Ingroup]		RAxML	PROTCATLGF	-483,244.54	88.56
			ExaML	-	-457,862.95	85.77
						79.43
			TNT	-	77,374	89.41
						81.23
100 MCG	100	38,887	RAxML	PROTCATLGF	-1,469,712.63	93.31
			ExaML	-	-1,408,969.93	85.56
						85.03
			TNT	-	253,282	96.42
						88.81
	[Ingroup]		RAxML	PROTCATLGF	-1,179,891.21	95.67
			ExaML	-	-1,123,079.07	91.22
						87.30
			TNT	-	197,077	97.92
						90.94

150 MCG	150	59,265	RAxML	PROTCATLGF	-2,538,255.03	95.39
			ExaML	-	-2,448,015.94	89.23
						89.70
			TNT	-	449,917	94.75
						89.37
	[Ingroup]		RAxML	PROTCATLGF	-2,037,279.55	95.66
			ExaML	-	-1,948,280.55	91.72
						88.20
			TNT	-	351,302	91.80
						89.29
200 MCG	200	77,986	RAxML	PROTCATLGF	-3,770,138.50	97.55
			ExaML	-	-3,657,857.77	94.22
						94.04
			TNT	-	687,849	97.17
						94.61
	[Ingroup]		RAxML	PROTCATLGF	-3,038,324.97	98.18
			ExaML	-	-2,924,875.90	97.21
						94.31
			TNT	-	540,852	97.58
						94.47
Core Genes	208	80,578	RAxML	PROTCATLGF	-4,021,184.16	97.98
			ExaML	-	-3,906,397.28	96.50
						93.91
			TNT	-	747,377	97.05
						94.70
	[Ingroup]		RAxML	PROTCATLGF	-3,251,380.02	98.09
			ExaML	-	-3,133,168.42	94.21
						94.38
			TNT	-	582,938	97.38
						95.18
MARE Matrix	2,116	614,117	RAxML	PROTCATLGF	-32,085,776.20	99.18

			ExaML	-	-31,300,593.65	96.77	
						92.64	
			TNT	-	6,084,484	97.88	
						94.17	
	[Ingroup]		RAxML	PROTCATLGF	-26,699,340.11	98.13	
			ExaML	-	-25,860,253.33	X	
						X	
			TNT	-	4,898,106	97.36	
						93.39	
Full Matrix	14,042	3,855,635	RAxML	PROTGAMMALGF	-86,508,642.04	X	
			ExaML	-	-82,437,287.39	X	
						X	
			TNT	-	14,000,901	98.95	
						95.76	
	[Ingroup]		RAxML	PROTGAMMALGF	-71,799,644.20	X	
			ExaML	-	X	X	
						X	
			TNT	-	11,144,755	93.76	
						92.56	

General characteristics of the character matrices constructed with only the ingroup strains included and trees inferred from them. The score is the log likelihood for the ML analyses, the number of steps for MP analysis (not counting uninformative characters). For ExaML and TNT, both normal (above) and partition bootstrapping (below) were conducted. Abbreviations used: MCG, most conserved genes; BS, bootstrap support; X, omitted for reasons of running time.

Matrix	**# Genes**	**# Characters**	**Analysis**	**Model**	**Score**	**Average BS**
50 MCG	50	17,293	RAxML	PROTCATLGF	-408,140.96	92.67
			ExaML	-	-384,965.07	81.04
						80.79

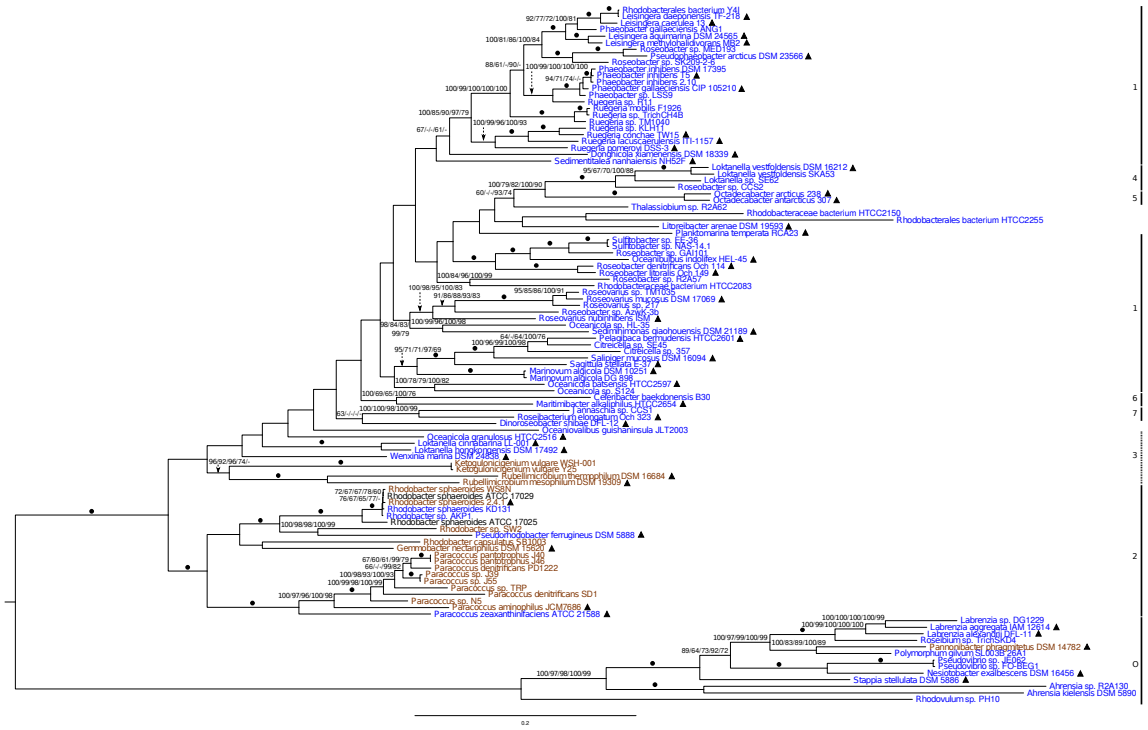
			TNT	-	64,353	89.78
						80.94
100 MCG	100	35,891	RAxML	PROTCATLGF	-976,238.27	91.77
			ExaML	-	-925,625.22	86.87
						86.80
			TNT	-	158,333	97.26
						87.70
150 MCG	150	55,234	RAxML	PROTCATLGF	-1,666,802.24	96.93
			ExaML	-	-1,585,188.16	87.87
						87.80
			TNT	-	278,031	94.62
						89.17
200 MCG	200	75,696	RAxML	PROTCATLGF	-2,544,613.25	98.31
			ExaML	-	-2,429,202.17	91.49
						92.14
			TNT	-	434,745	98.34
						91.97
250 MCG	250	94,224	RAxML	PROTCATLGF	-3,495,301.53	97.02
			ExaML	-	-3,355,092.27	90.68
						91.03
			TNT	-	611,849	98.37
						93.47
Core Genes	297	110,074	RAxML	PROTCATLGF	-4,534,876.05	97.47
			ExaML	-	-4,376,418.38	93.96
						94.14
			TNT	-	814,645	98.31
						95.99
MARE Matrix	2106	597,144	RAxML	PROTCATLGF	-26,318,037.63	97.44
			ExaML	-	-25,508,730.2	X
						X

			TNT	-	4,833,411	96.98 92.51
Full Matrix	12,389	3,383,815	RAxML	PROTGAMMALGF	-70,443,547.27	X
			ExaML	-	-66,869,731.93	X X
			TNT	-	11,144,755	94.14 92.46

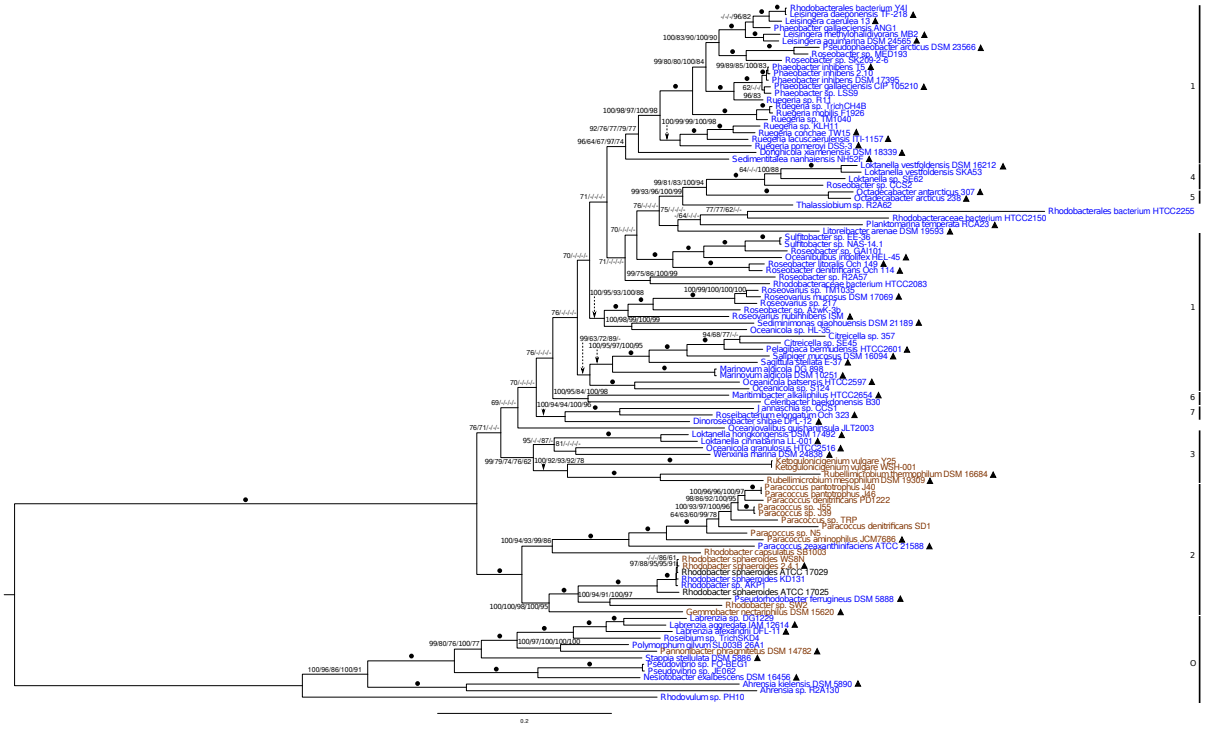
C. Phylogenetic analysis of alternative supermatrices with all strains

This chapter lists the trees inferred from alternative matrices with all strains but which are not shown in the main manuscript.

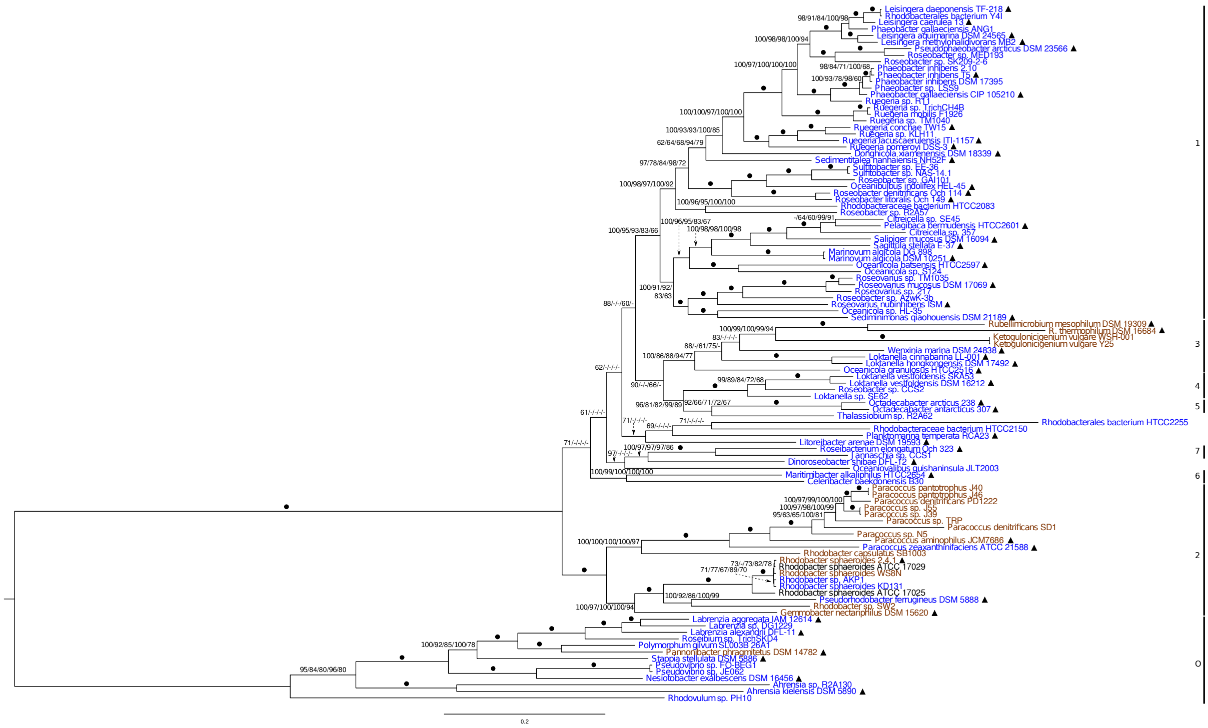
3. ML tree inferred from the supermatrix including the 50 most conserved genes under a single overall model of amino acid evolution and rooted with the included outgroup strains. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



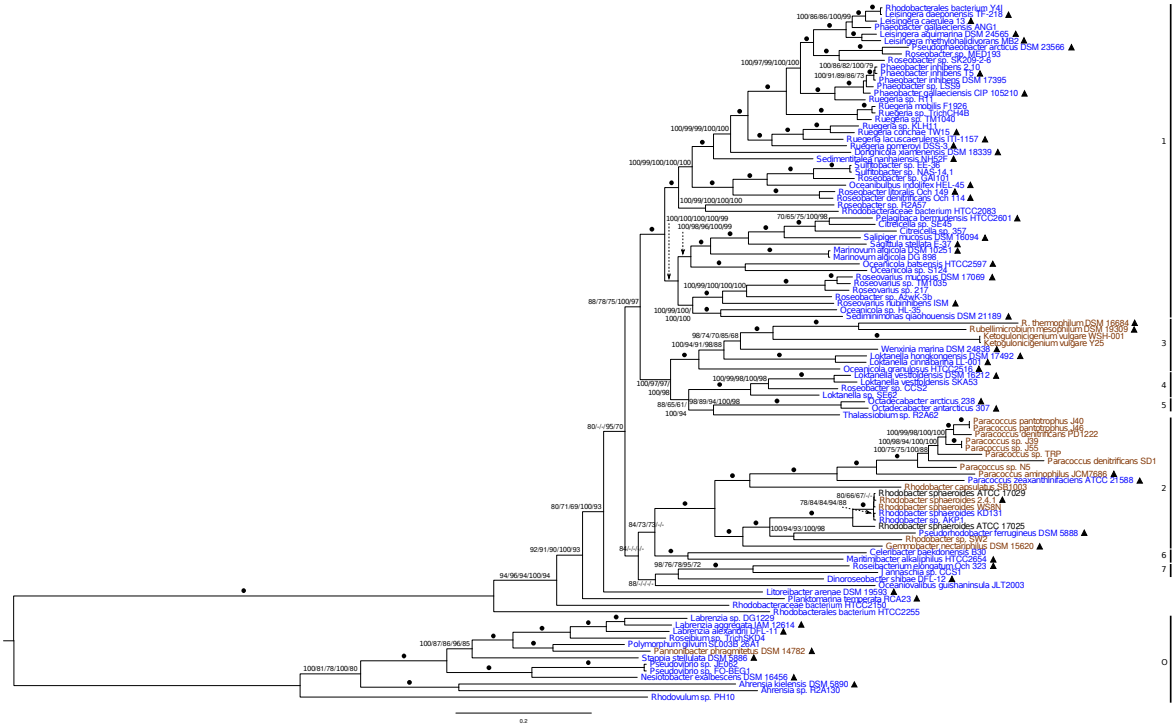
4. ML tree inferred from the supermatrix including the 100 most conserved genes under a single overall model of amino acid evolution and rooted with the included outgroup strains. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



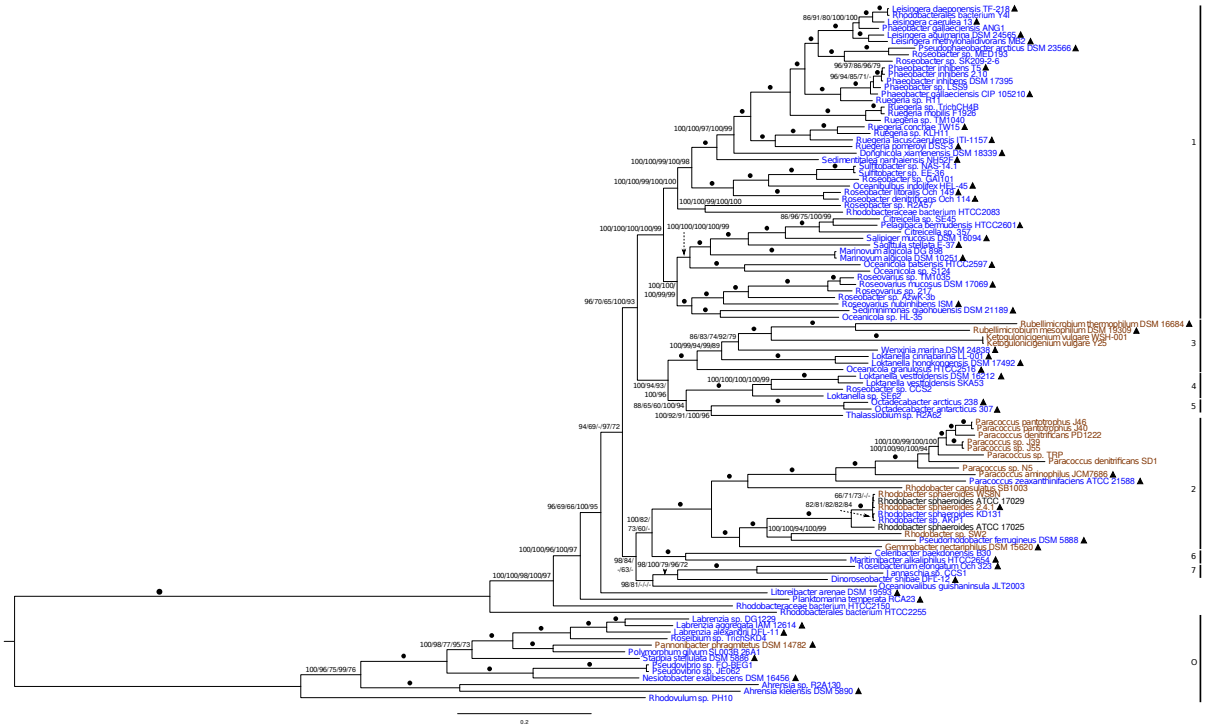
5. ML tree inferred from the supermatrix including the 150 most conserved genes under a single overall model of amino acid evolution and rooted with the included outgroup strains. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



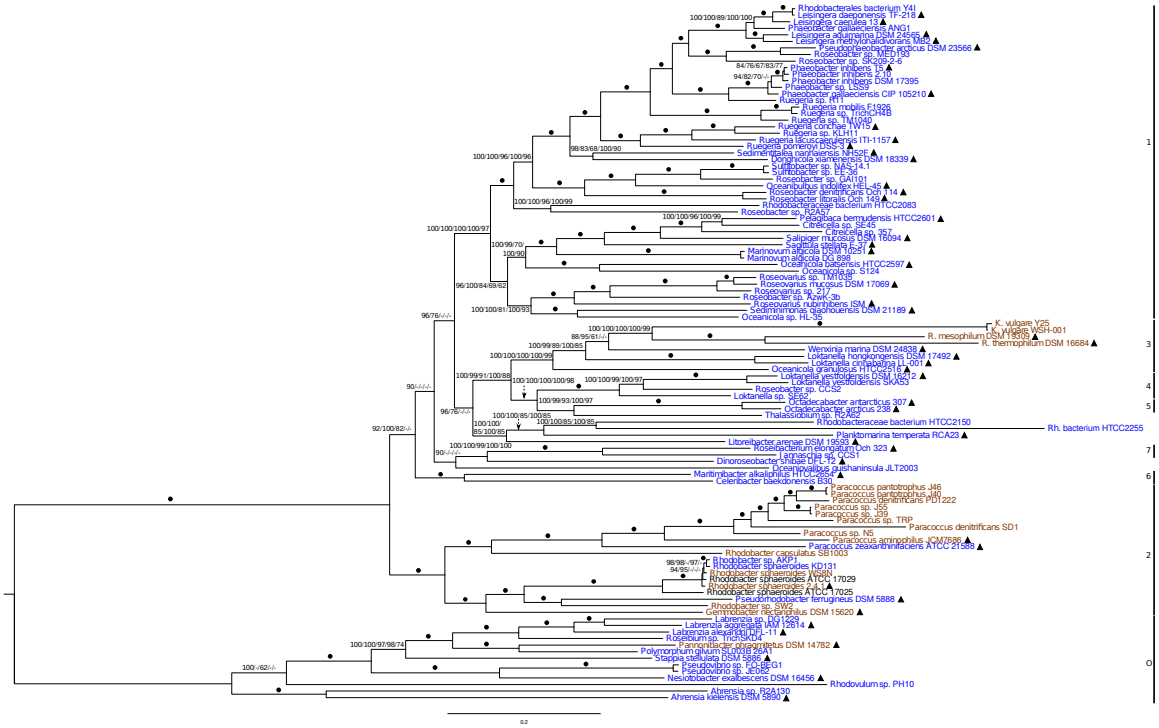
6. ML tree inferred from the supermatrix including the 200 most conserved genes under a single overall model of amino acid evolution and rooted with the included outgroup strains. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



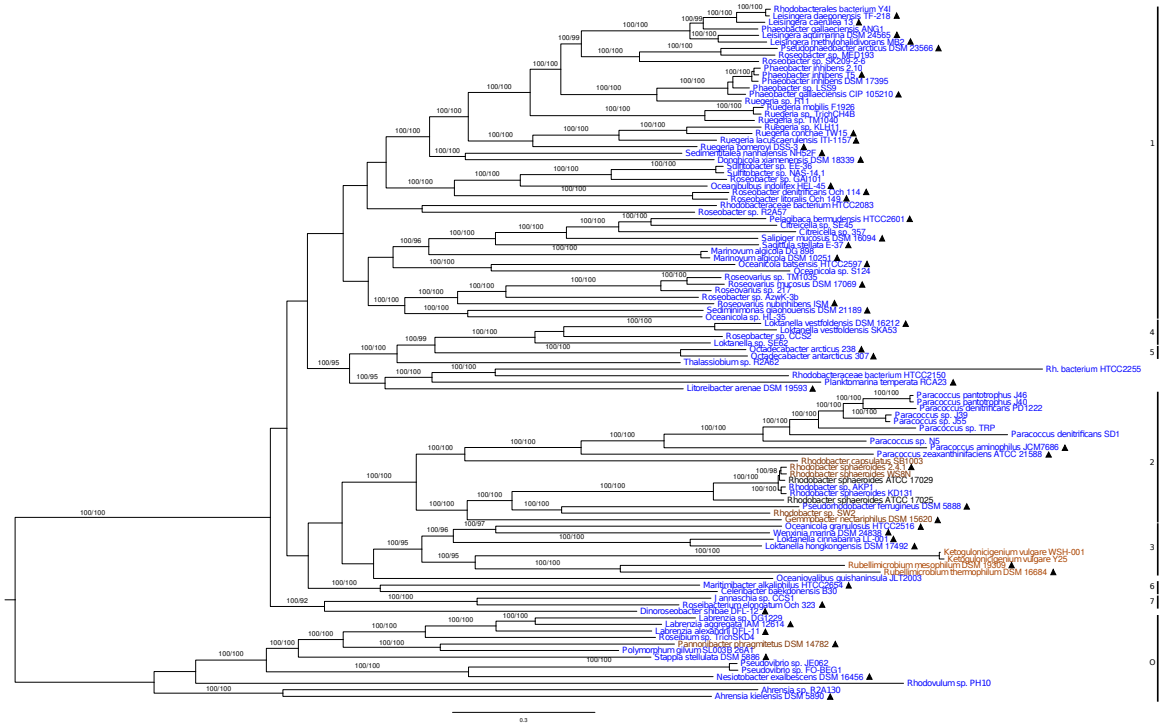
7. ML tree inferred from the core-gene matrix under a single overall model of amino acid evolution and rooted with the included outgroup strains. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



8. ML tree inferred from the MARE-filtered supermatrix under a single overall model of amino acid evolution and rooted with the included outgroup strains. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



9. ML tree inferred from the “full” supermatrix under a single overall model of amino acid evolution and rooted with the included outgroup strains. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under MP; (ii) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



D. Constrained analyses and paired-site tests

Results of the maximum likelihood (ML) and maximum-parsimony (MP) constraint analyses and paired-site tests for a selection of the supermatrices described in the text. (AU-test, approximately unbiased test; *after summing up the scores per gene).

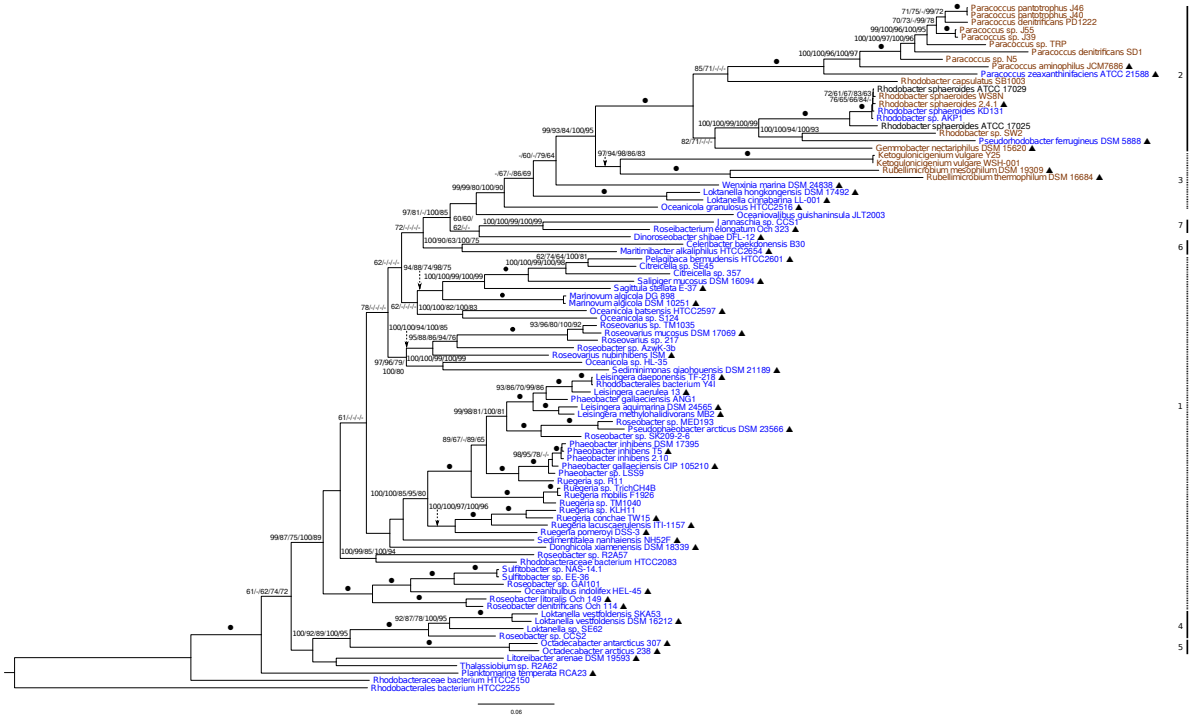
Analysis	**Log likelihood of the ML tree**		**Length of the best MP tree (steps)**		**P value**
	unconstrained	**constrained**	**uncon-strained**	**constrained**	
core genes (roseobacter constraint)	RAxML: -4,021,184.16 ExaML: -4,015,298.23	RAxML: -4,021,395.94 ExaML: -4,015,472.80	737,783	738,388	RAxML/AU-test: 0.029000 ExaML/Wilcox: 0.000000 ExaML/T-test: 0.062282 ExaML/Wilcox*: 0.148199 ExaML/T-test*: 0.167543 TNT/Wilcox: 0.000000 TNT/T-test: 0.000000 TNT/Wilcox*: 0.000031 TNT/T-test*: 0.000057
“full” (roseobacter constraint)	RAxML: -86,508,642.04 ExaML: -86,232,358.08	RAxML: -86,506,978.69 ExaML: -86,234,156.07	14,000,901	14,002,523	RAxML/AU-test: 0,000200 ExaML/Wilcox: 0.000070 ExaML/T-test: 0.007039 ExaML/Wilcox*: 0.000000

					ExaML/T-test*: 0.293506 TNT/Wilcox: 0.000000 TNT/T-test: 0.000000 TNT/Wilcox*: 0.003236 TNT/T-test*: 0.029498
MARE-filtered (roseobacter constraint)	ML analyses (RAxML, ExaML): roseobacters already monophyletic		6,007,432	6,008,833	ML (RAxML, ExaML) analyses: roseobacters already monophyletic TNT/Wilcox: 0.000000 TNT/T-test: 0.000000 TNT/Wilcox*: 0.000000 TNT/T-test*: 0.000000
MARE-filtered (core-genes constraint)	RAxML: -32,085,776.20 ExaML: -32,023,681.27	RAxML: -32,090,325.13 ExaML: -32,025,887.74	6,007,432	6,008,387	RAxML/AU-test: 0.000000 ExaML/Wilcox: 0.000000 ExaML/T-test: 0.000013 ExaML/Wilcox*: 0.772978 ExaML/T-test*: 0.097032 TNT/Wilcox:

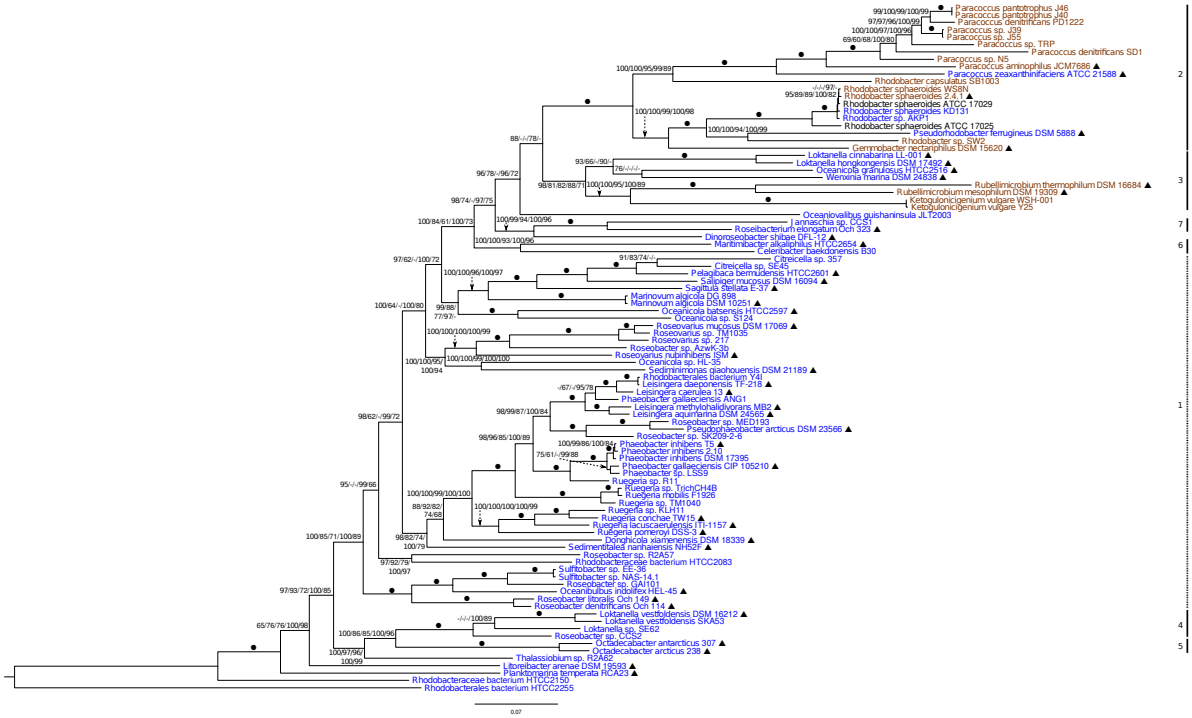
					0.000173 TNT/T-test: 0.000173 TNT/Wilcox*: 0.893966 TNT/T-test*: 0.109582
MARE-filtered (“full” constraint)	RAxML: -32,085,776.20 ExaML: -32,023,537.85	RAxML: -32,087,072.01 ExaML: -32,025,086.16	6,007,432	6,010,002	RAxML/AU-test: 8e-06 ExaML/Wilcox: 0.024465 ExaML/T-test: 0.000000 ExaML/Wilcox*: 0.000111 ExaML/T-test*: 0.042567 TNT/Wilcox: 0.000000 TNT/T-test: 0.000000 TNT/Wilcox*: 0.000000 TNT/T-test*: 0.003742

F. Phylogenetic analysis of normal matrices but without the outgroup

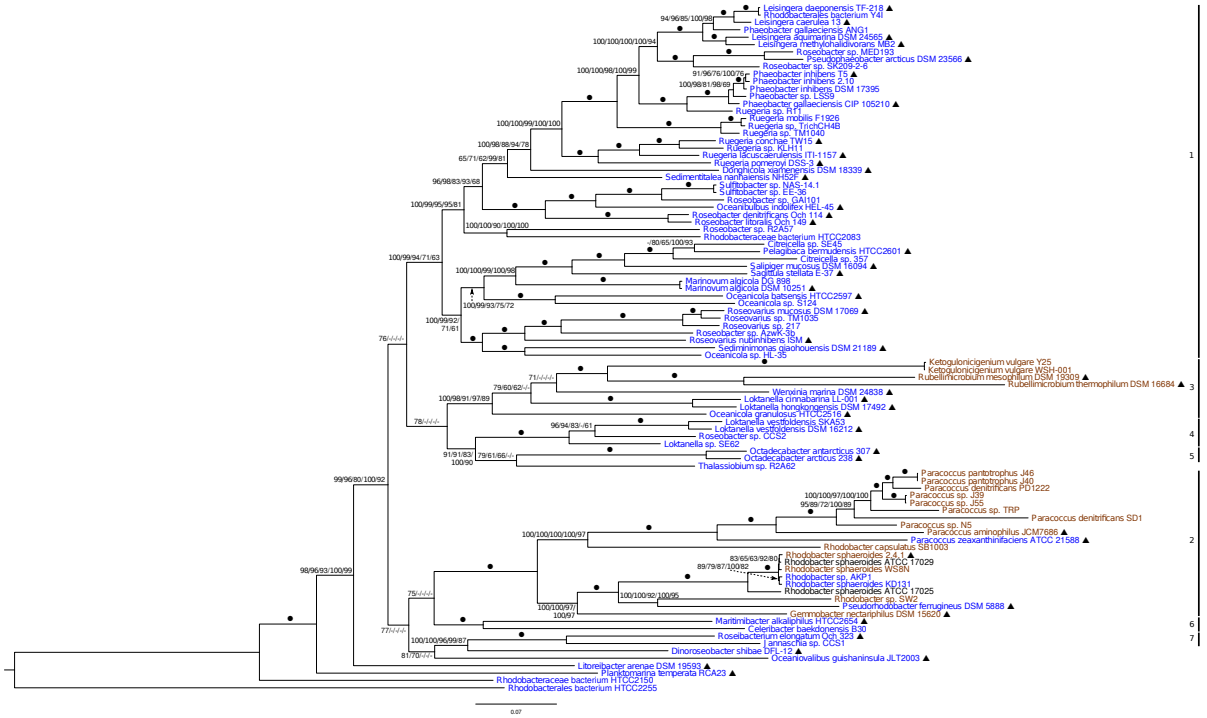
1. ML tree inferred from the supermatrix including the 50 most conserved genes after removal of the outgroup under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



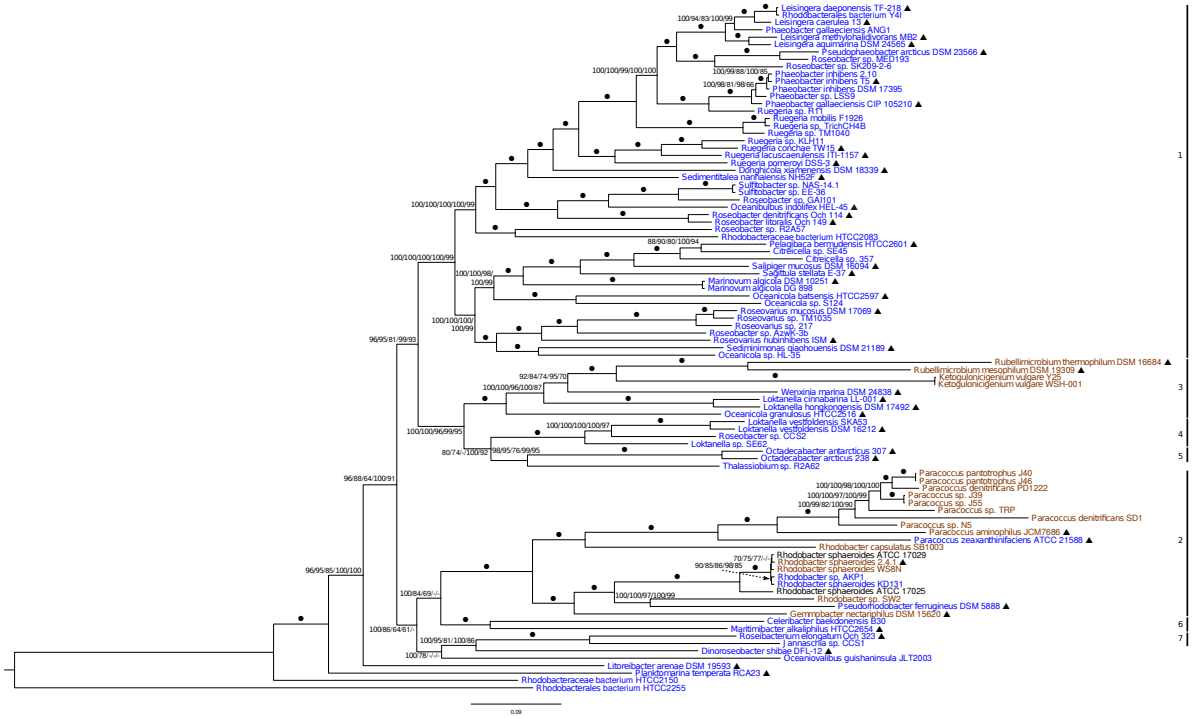
2. ML tree inferred from the supermatrix including the 100 most conserved genes after removal of the outgroup under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



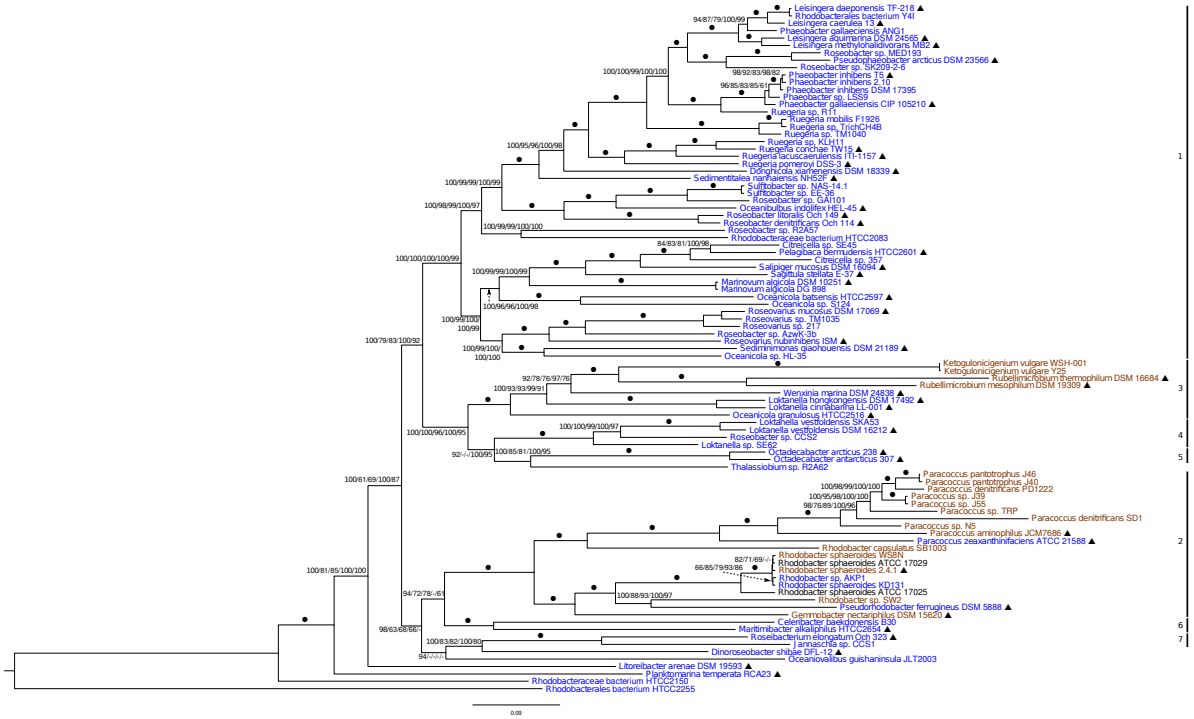
3. ML tree inferred from the supermatrix including the 150 most conserved genes after removal of the outgroup under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



4. ML tree inferred from the supermatrix including the 200 most conserved genes after removal of the outgroup under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



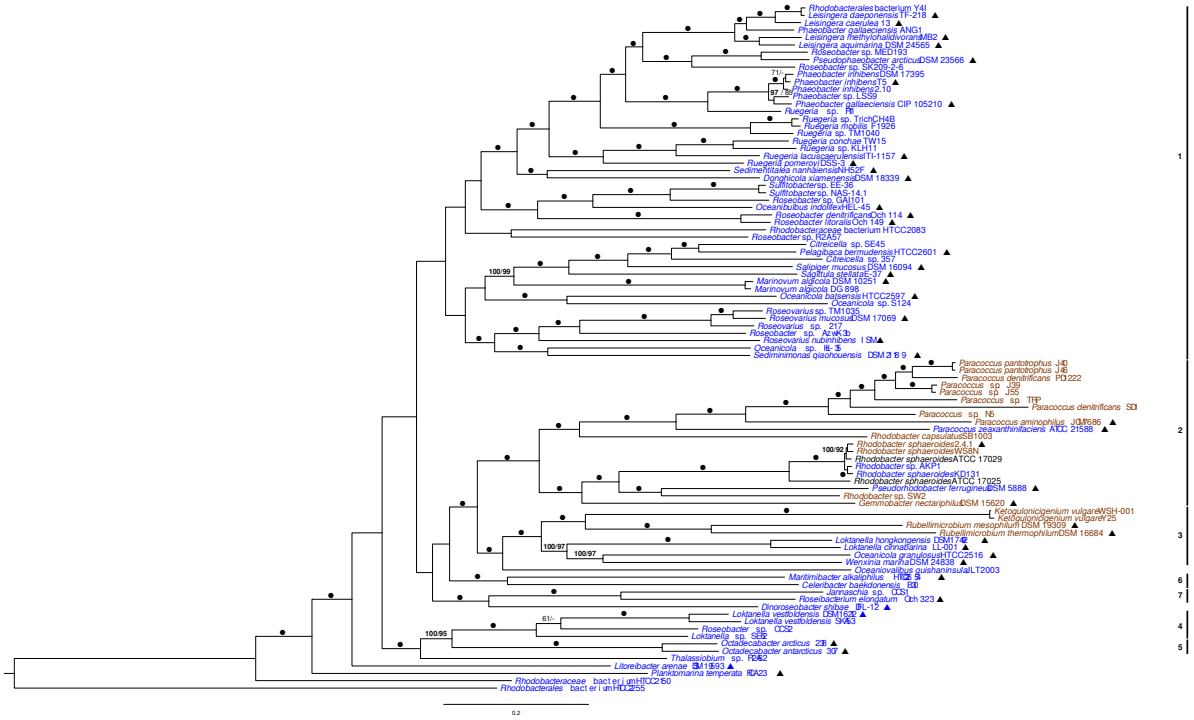
5. ML tree inferred from the core-gene matrix after removal of the outgroup under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



6. ML tree inferred from the MARE-filtered supermatrix after removal of the outgroup under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.

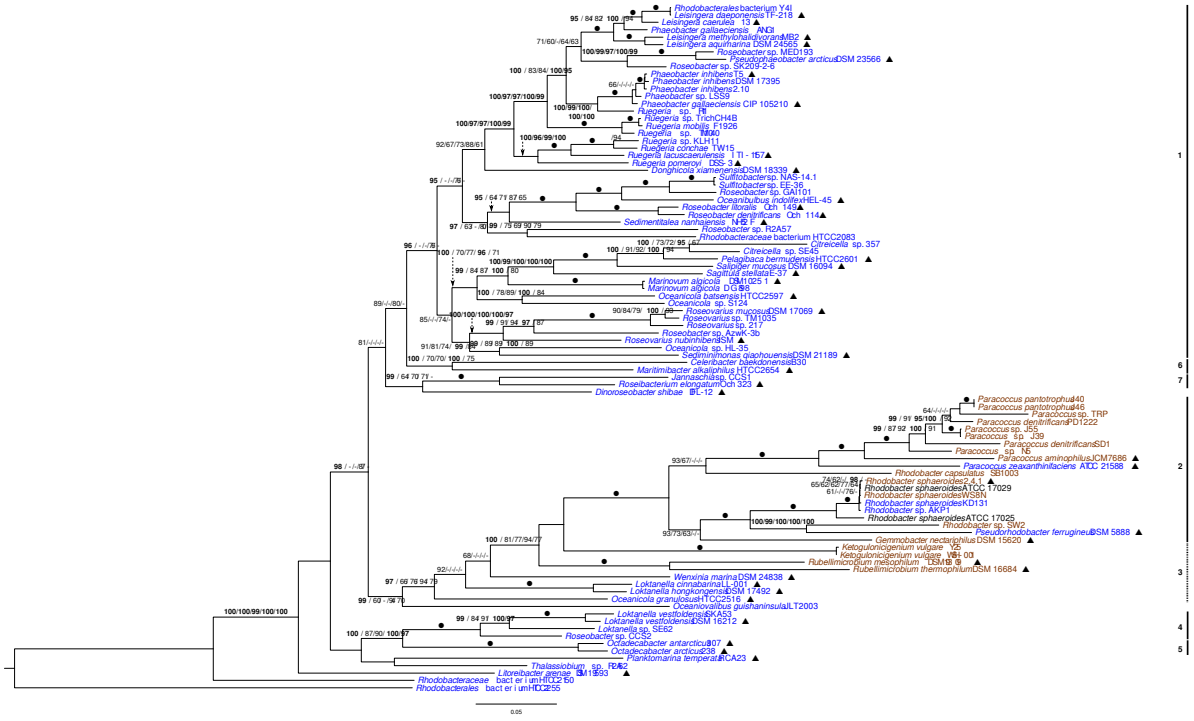
[Omitted for reasons of running time.]

7. ML tree inferred from the “full” supermatrix after removal of the outgroup under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under MP; (ii) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.

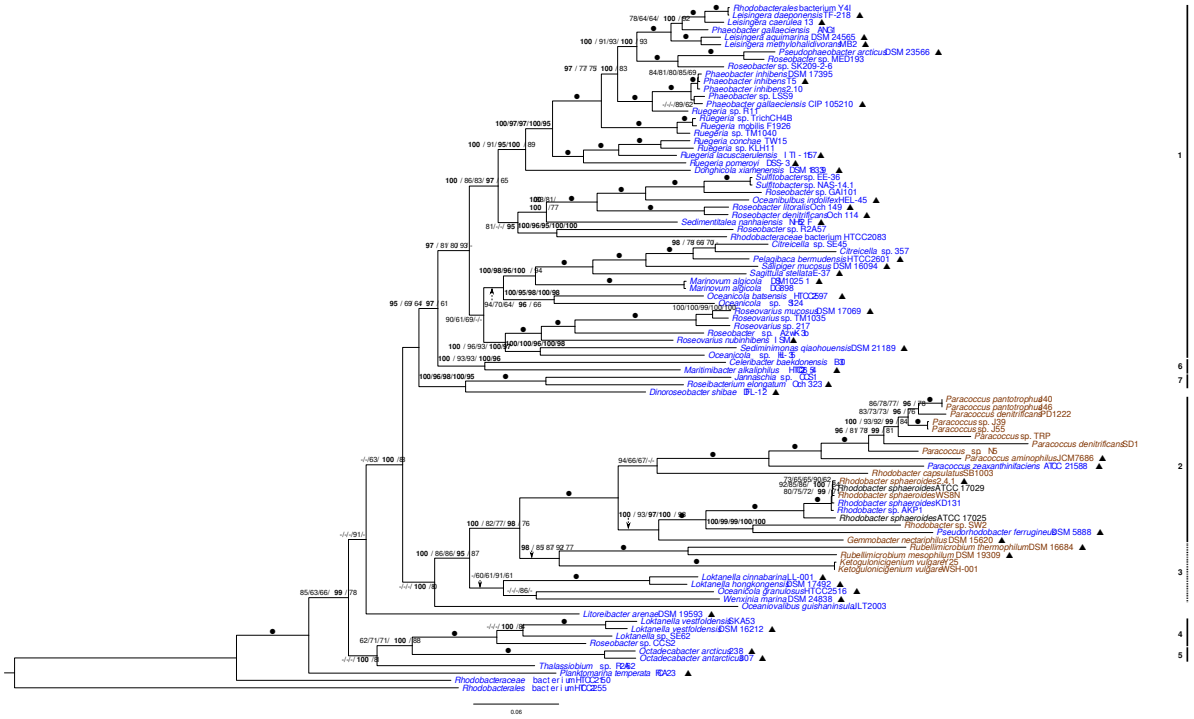


G. Phylogenetic analysis of matrices generated without the outgroup

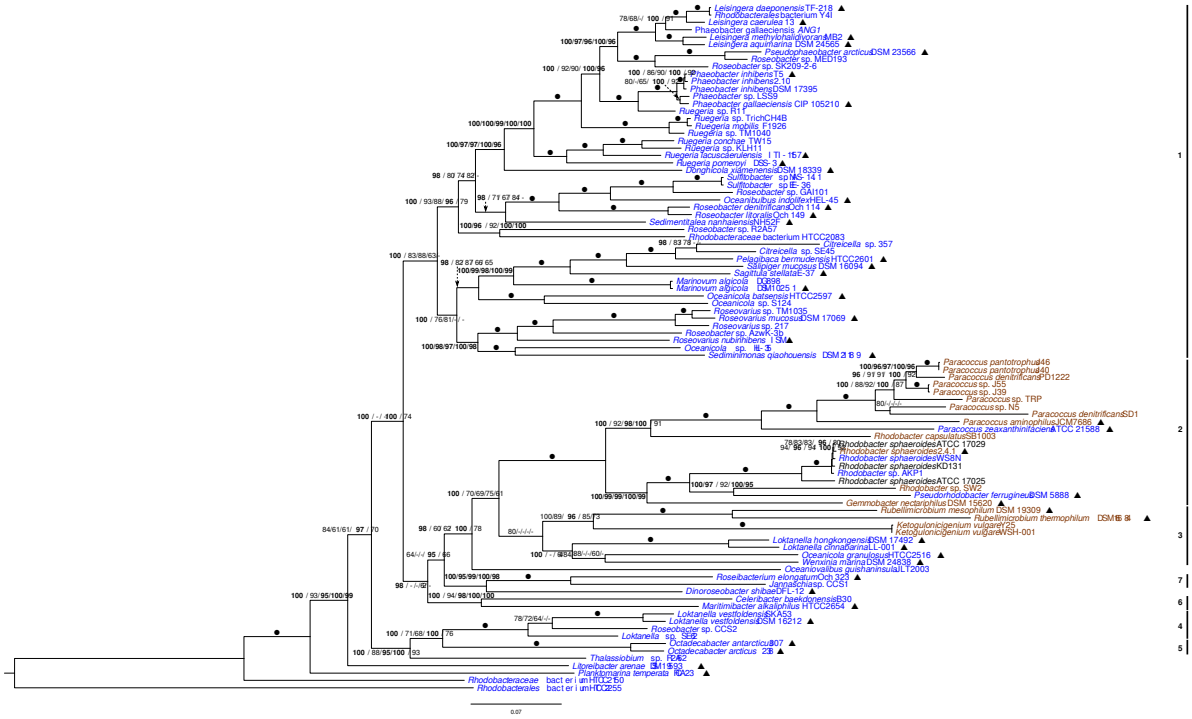
1. ML tree inferred from a second supermatrix including the 50 most conserved genes (generated anew after removal of the outgroup) under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



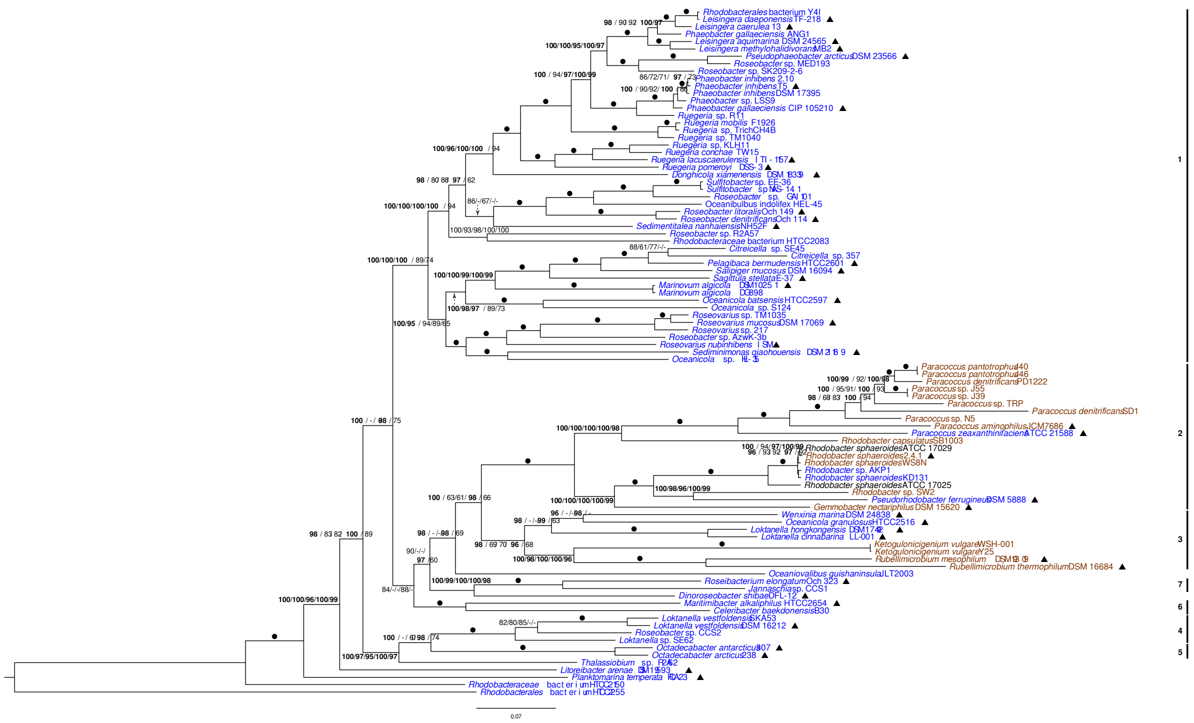
2. ML tree inferred from a second supermatrix including the 100 most conserved genes (generated anew after removal of the outgroup) under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



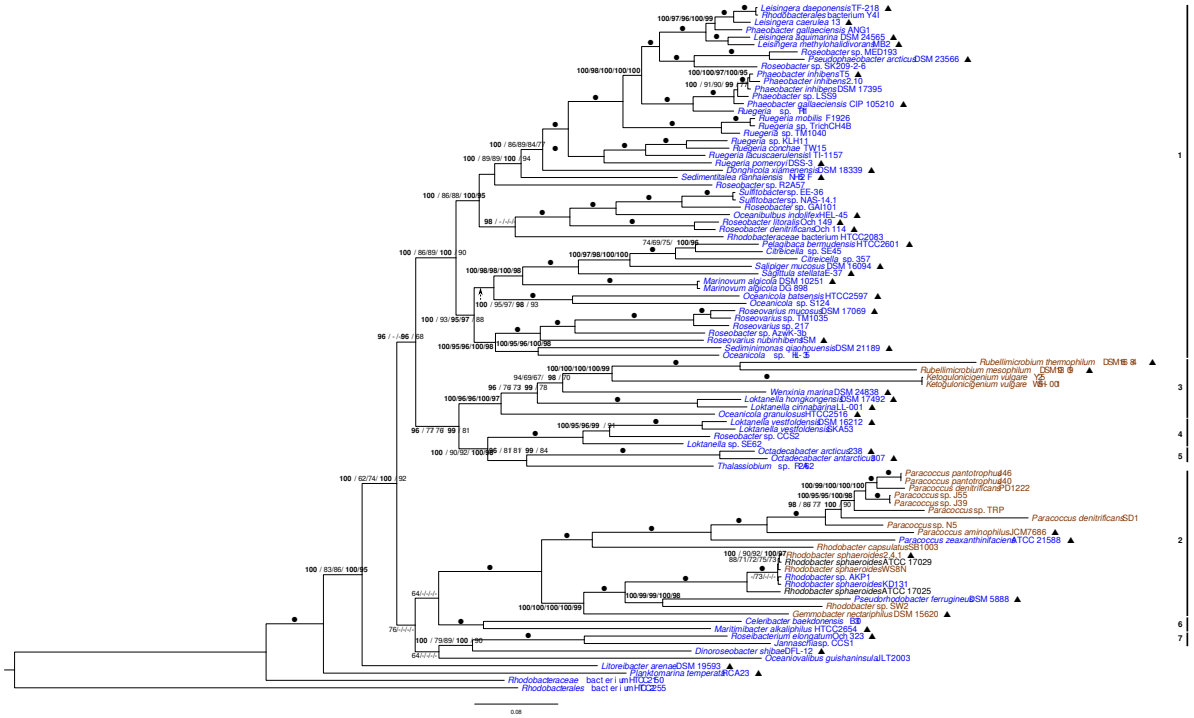
3. ML tree inferred from a second supermatrix including the 150 most conserved genes (generated anew after removal of the outgroup) under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



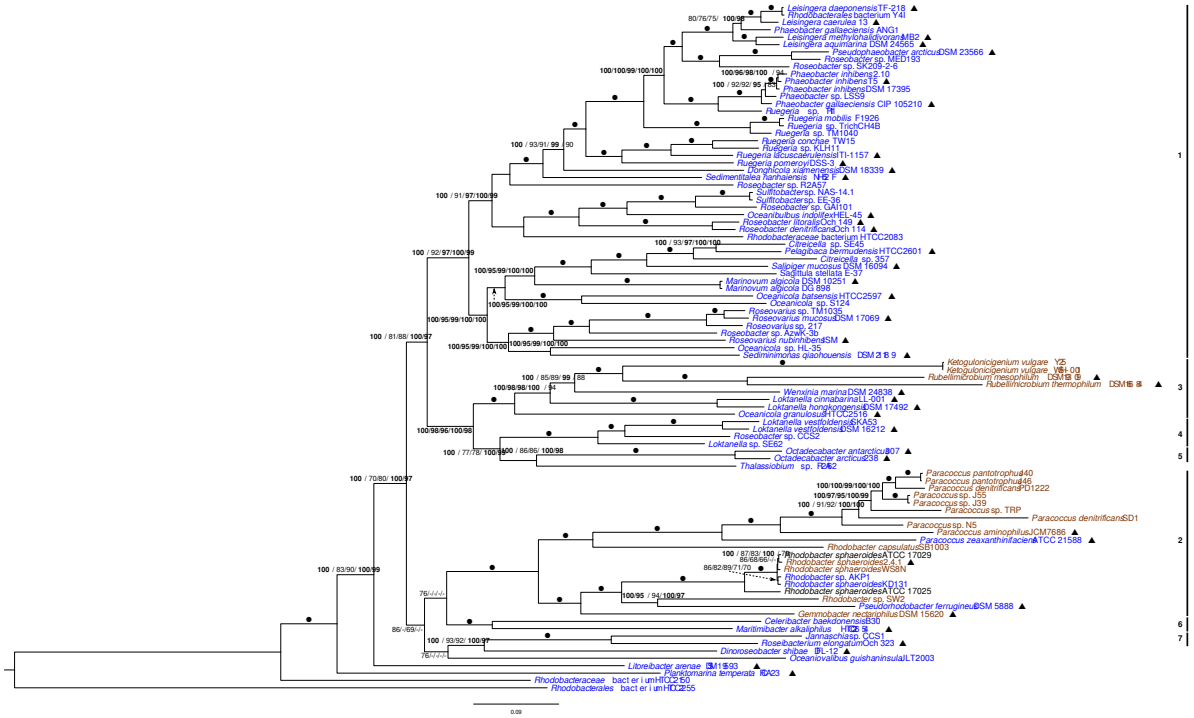
4. ML tree inferred from a second supermatrix including the 200 most conserved genes (generated anew after removal of the outgroup) under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



5. ML tree inferred from a second supermatrix including the 250 most conserved genes (generated anew after removal of the outgroup) under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



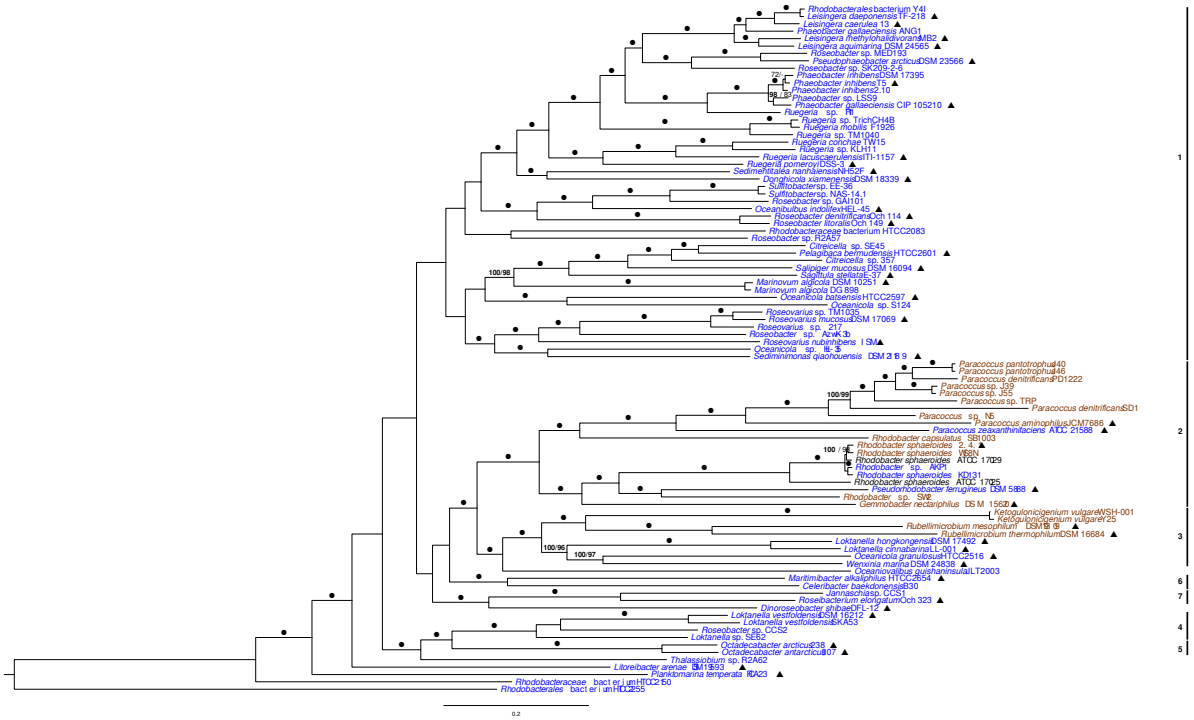
6. ML tree inferred from a second core-gene matrix (generated anew after removal of the outgroup) under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



7. ML tree inferred from a second MARE-filtered supermatrix (generated anew after removal of the outgroup) under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under ML with a single overall model of amino acid evolution; (ii) ordinary bootstrap under ML with one model per gene; (iii) partition bootstrap under ML; (iv) ordinary bootstrap under MP; (v) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.

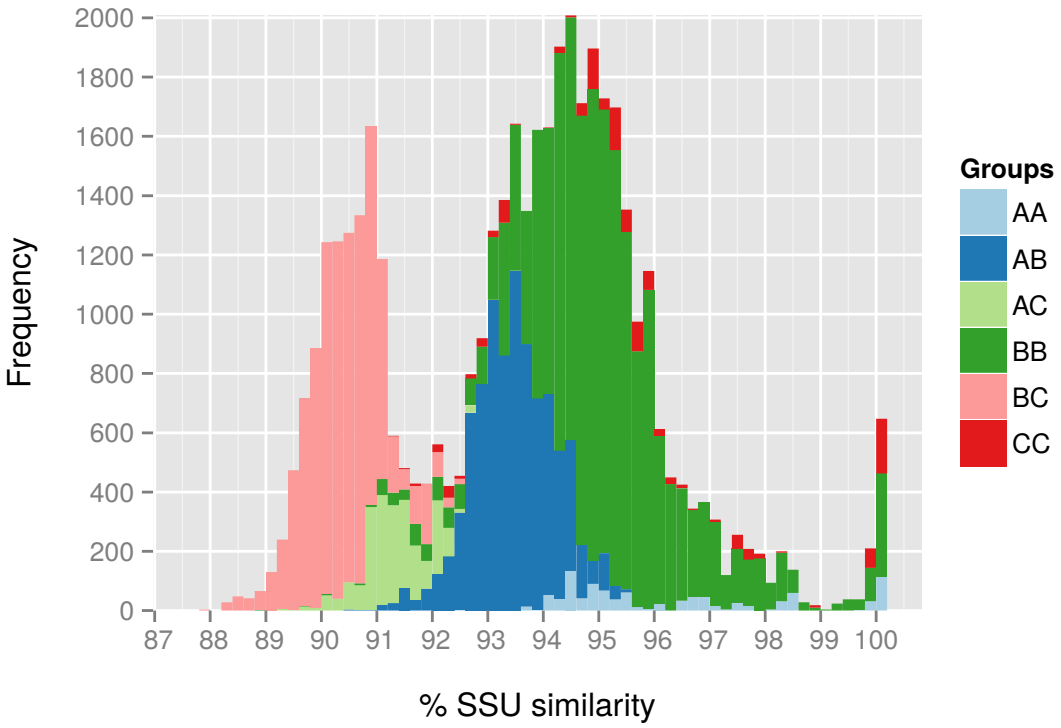
[Omitted for reasons of running time.]

8. ML tree inferred from a second “full” supermatrix (generated anew after removal of the outgroup) under a single overall model of amino acid evolution and rooted according to the LSD results. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches (from left to right) are bootstrapping support values if larger than 60% from (i) ordinary bootstrap under MP; (ii) partition bootstrap under MP. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. For the groups see the caption of Fig. 1 in the main manuscript. Triangles indicate type strains. The colours of the tip labels indicate the habitat: blue, marine; brown, non-marine; uncoloured, unknown.



H. 16S rRNA gene pairwise similarities and paired-site tests

Histograms of all possible pairwise 16S rRNA gene similarities between the following groups: *Rhodobacteraceae* except *Roseobacter* group (**A**), *Roseobacter* group (**B**) and outgroup (**C**). Labels '**AA**', '**BB**' and '**CC**' denote the respective within-group similarities.



The 16S rRNA gene sequences extracted from the genomes showed similarity values of >89% in all pairwise comparisons of *Rhodobacteraceae*, not only in the pairwise comparisons of roseobacters (supplementary file 2); similarities ≤89% were only observed between outgroup and *Rhodobacteraceae* sequences.

Results of the paired-site tests of the 16S rRNA gene sequences under MP. The constraints enforced the monophyly of all included strains except for: Constraint 1, *Rhodobacterales* bacterium HTCC2255 and outgroup; Constraint 2, *Rhodobacterales* bacterium HTCC2255, *Rhodobacteraceae* bacterium HTCC2150 and outgroup; Constraint 3, *Rhodobacterales* bacterium HTCC2255, *Rhodobacteraceae* bacterium HTCC2150, *Planktomarina temperata* RCA23^{T} and outgroup; Constraint 4, *Rhodobacterales* bacterium HTCC2255, *Rhodobacteraceae* bacterium HTCC2150, *Planktomarina temperata* RCA23^{T}, *Litoreibacter arenae* DSM 19593^{T} and outgroup. The best trees under each constraint were compared with the best trees from unconstrained search.

Best tree(s)	**MP score**	**Number of best MP trees**	**p-value Wilcox**	**P-value T-test**
Unconstrained	3262	9	-	-

Constraint 1	3264	18	0.41121-0.54026	0.44743-0.45270
Constraint 2	3270	13	0.26162-0.31018	0.27248-0.29260
Constraint 3	3270	9	0.28361-0.36064	0.27912-0.32615
Constraint 4	3275	15	0.14461-0.18642	0.16461-0.19657

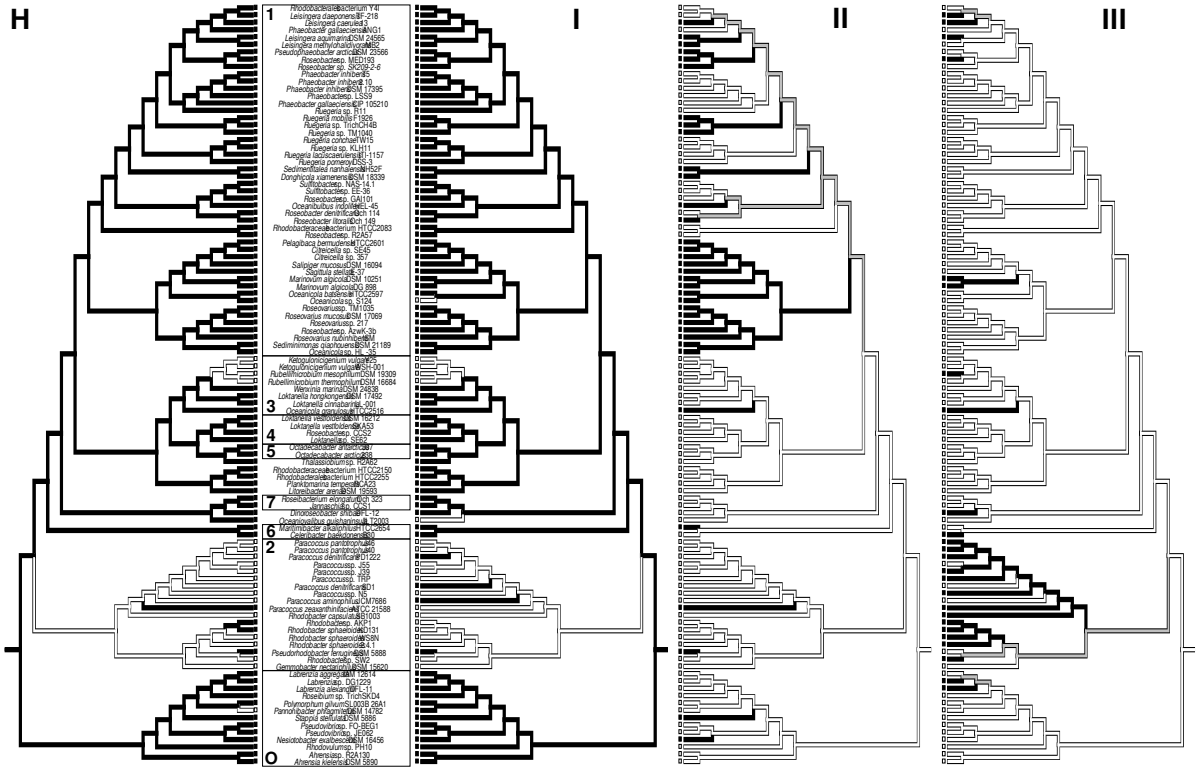
Results of the paired-site tests of the 16S rRNA gene sequences under ML. The constraints enforced the monophyly of all included strains except for: Constraint 1, *Rhodobacterales* bacterium HTCC2255 and outgroup; Constraint 2, *Rhodobacterales* bacterium HTCC2255, *Rhodobacteraceae* bacterium HTCC2150 and outgroup; Constraint 3, *Rhodobacterales* bacterium HTCC2255, *Rhodobacteraceae* bacterium HTCC2150, *Planktomarina temperata* RCA23^{T} and outgroup; Constraint 4, *Rhodobacterales* bacterium HTCC2255, *Rhodobacteraceae* bacterium HTCC2150, *Planktomarina temperata* RCA23^{T}, *Litoreibacter arenae* DSM 19593^{T} and outgroup. The best tree under each constraint was compared with the best tree from unconstrained search.

Best tree(s)	**Log likelihood**	**p-value AU test**	**p-value Wilcox**	**P-value T-test**
Unconstrained	-17315.899407	-	-	-
Constraint 1	-17344.731674	0.633	1	0.461546575404846
Constraint 2	-17339.357087	0.390	1	0.305064211535987
Constraint 3	-17356.114526	0.244	1	0.25290944692449
Constraint 4	-17330.499514	0.371	1	0.335270693544713

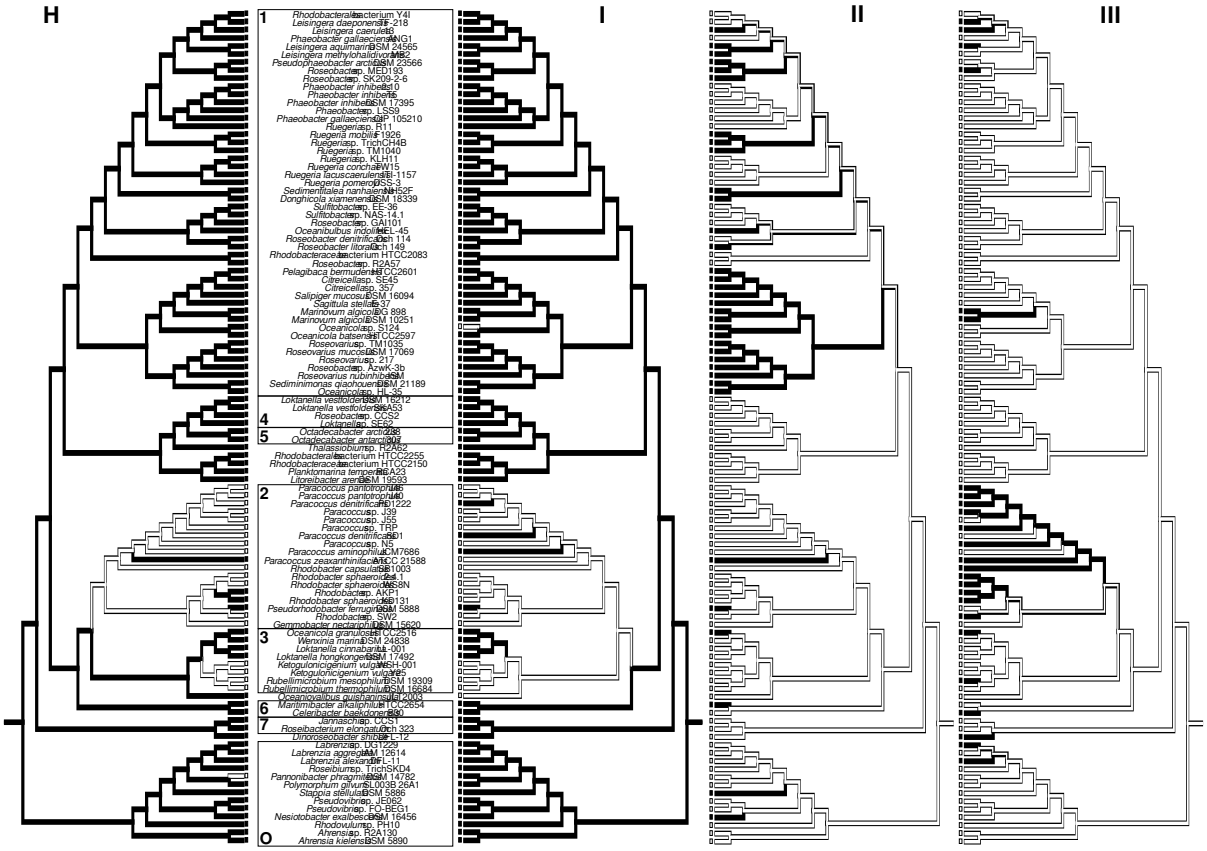
The best 16S rRNA gene trees from unconstrained search were not significantly better than the best trees obtained by enforcing the monophyly of the ingroup strains excluding "*Rhodobacterales* bacterium" HTCC2255, "*Rhodobacteraceae* bacterium" HTCC2150 or *Planktomarina temperata* RCA23^{T}, confirming the result from bootstrapping that the 16S rRNA gene does not significantly support the *Roseobacter* clade.

I. Ancestral character-state reconstruction on the MARE-filtered supermatrix tree

1. Ancestral character-state reconstruction under ordered MP for the presence (black) or absence (white) of H, marine or equivalent habitat; I, (S)-2-haloacid dehalogenase (EC 3.8.1.2); II, ectoine synthase (EC 4.2.1.108); III, 6-phosphofructokinase (EC 2.7.1.11). The tree topology is as in figure C/8. Grey shading indicates uncertainties in character-state assignment. The major types of phylogenetic distributions represented by the three genomic characters are: I, losses predominantly in non-marine strains; II, gains mainly in marine strains; III, gains predominantly in non-marine strains.



2. Ancestral character-state reconstruction under ordered MP for the presence (black) or absence (white) of H, marine or equivalent habitat; I, (S)-2-haloacid dehalogenase (EC 3.8.1.2); II, ectoine synthase (EC 4.2.1.108); III, 6-phosphofructokinase (EC 2.7.1.11). The tree topology is as in figure C/9. Grey shading indicates uncertainties in character-state assignment. The major types of phylogenetic distributions represented by the three genomic characters are: I, losses predominantly in non-marine strains; II, gains mainly in marine strains; III, gains predominantly in non-marine strains.



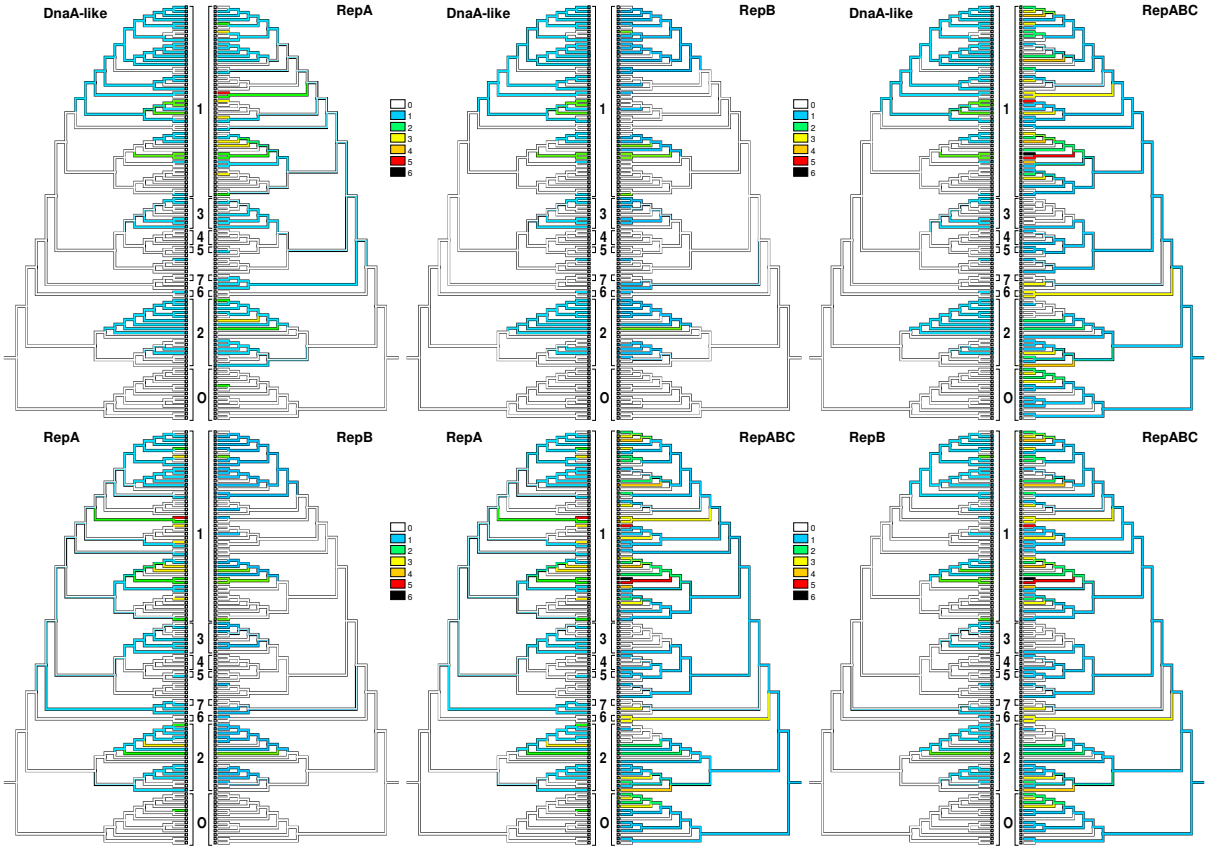
J. Ancestral character-state reconstructions of the number of extrachromosomal replicons per type of replication system

Overall we identified 325 plasmid replication modules in the 106 strains investigated in the current study and determined their compatibility groups (supplementary file 1). The ingroup genomes harboured between zero and 12 and on average 3.3 ECRs, and nearly all of their 307 replication modules fit into the existing classification scheme (Petersen et al. 2009 and 2011). Additional subtypes of the RepA-, RepB- and RepABC-type replication systems (supplementary file 1) might represent further yet unknown compatibility groups. We identified five new subtypes of replication systems with specific palindrome sequences (C-10, C-13, C-14, C-15, C-18) (Petersen et al. 2009), and the presence of two RepABC-9 subtype ECRs in *Leisingera daeponensis* TF-218T and *Marinovum algicola* DSM 10251T indicates that the existing classification scheme has to be extended. The current study represents the most comprehensive comparison of ECRs in *Rhodobacteraceae* and revealed the presence of up to twelve replicons (chromosome, chromids, plasmids; Harrison et al. 2010) in a single bacterium (supplementary file 1), which is in agreement with former results from the physical separation with the pulsed-field technology (Pradella et al. 2004 and 2010). Nevertheless, the identification of at least 23 distinct compatibility groups in *Rhodobacteraceae* indicates that individual roseobacter strains may harbour more than twelve replicons as recently reported for *Marinovum algicola* DG898 (Frank et al. 2015a). The abundance of the different plasmid types in *Rhodobacteraceae* is further reflected by 53 DnaA-like, 79 RepA, 52 RepB and 140 RepABC replication modules.

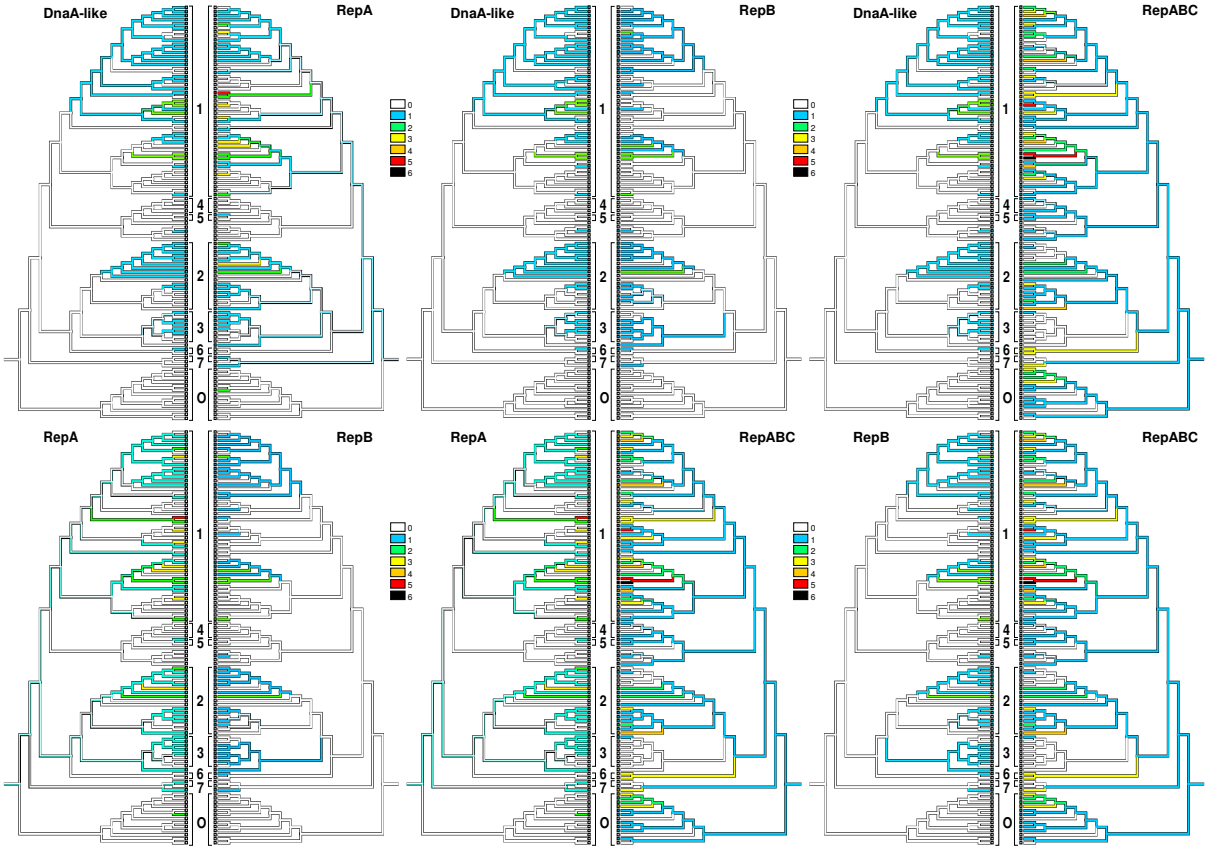
The replicases of the 18 ECRs in the outgroup genomes (2x RepA, 16x RepABC) grouped together with rhizobial sequences and were clearly separated from those of *Rhodobacteraceae* (Petersen et al. 2009). The sole exception is the replicase subtype RepABC-9 that branches as separate subtree within *Rhizobiales* as previously shown (Petersen et al. 2009). Our phylogeny-based classification of 325 plasmid replication systems thus documented that ECRs from *Rhodobacteraceae* belong to compatibility groups distinct from those found in the outgroup strains (supplementary file 1). In contrast, the BayesTraits analyses showed no correlation between the presence of ECRs and the habitat (supplementary file 4). Based on their ECR replication systems the marine *Rhodobacteraceae* are indistinguishable from non-marine ones (Petersen et al. 2009). This finding is in line with the fact that marine and non-marine *Rhodobacteraceae* are not phylogenetically separated (fig. 1).

In contrast, the presences and absences of ECR types DnaA-like, RepA and RepB significantly depended on each other, whereas RepABC showed no significant relationships to the others (fig. 3, supplementary file 4).

1. Ancestral character-state reconstruction under ordered MP for the number of ECR replicases of the distinct types DnaA, RepA, RepB and RepABC, and according pairwise phylogenetic cross-comparisons of their abundances in each genome. For the labels see figure V/1, where the same tree topology is depicted in exactly the same layout. The numbers between the trees refer to the clades as indicated in figure C/8 (O = outgroup). The colours indicate the number of replicases of each type as follows: white, 0; blue, 1; green, 2; yellow, 3; orange, 4; red, 5; black, 6. Presences and absences alone are correlated between DnaA, RepA and RepB but not between RepABC and the others.



2. Ancestral character-state reconstruction under ordered MP for the number of ECR replicases of the distinct types DnaA, RepA, RepB and RepABC, and according pairwise phylogenetic cross-comparisons of their abundances in each genome. For the labels see figure V/2, where the same tree topology is depicted in exactly the same layout. The numbers between the trees refer to the clades as indicated in figure C/9 (O = outgroup). The colours indicate the number of replicases of each type as follows: white, 0; blue, 1; green, 2; yellow, 3; orange, 4; red, 5; black, 6. Presences and absences alone are correlated between DnaA, RepA and RepB but not between RepABC and the others.



K. Results of the tip-permutation test for the phylogenetic conservation of number of extrachromosomal replicons per type of replication system

In the tip-permutation tests, the occurrences of all replication systems were phylogenetically conserved; after removal of the outgroup, only RepB and DnaA-like showed a significant conservation under all conditions. The significant positive correlations between the plasmid types DnaA-like, RepA and RepB, but not RepABC, is in agreement with their occurrence on evolutionarily stable chromids, which exhibit a codon usage comparable to that of each chromosome (Harrison et al. 2010). In contrast, RepABC-type ECRs frequently show a deviating codon usage and thus represent genuine plasmids. A prime example is the genus *Phaeobacter* with conserved RepA-I, RepB-I and DnaA-like I type chromids as well as a mobile gene pool essentially represented by RepABC-type plasmids (Frank et al. 2014; Dogs et al. 2013). Many of these plasmids contain type-IV secretion systems indicating that they are subjected to horizontal transfer between *Rhodobacteraceae* via conjugation (Petersen et al. 2013). This prediction is in line with the results of the permutation tests (supplementary file 1), suggesting a higher phylogenetic conservation in the other replication systems, and the higher number of phylogenetic correlations of the occurrences of RepABC and RepA with the presence of COGs associated with type-IV secretion.

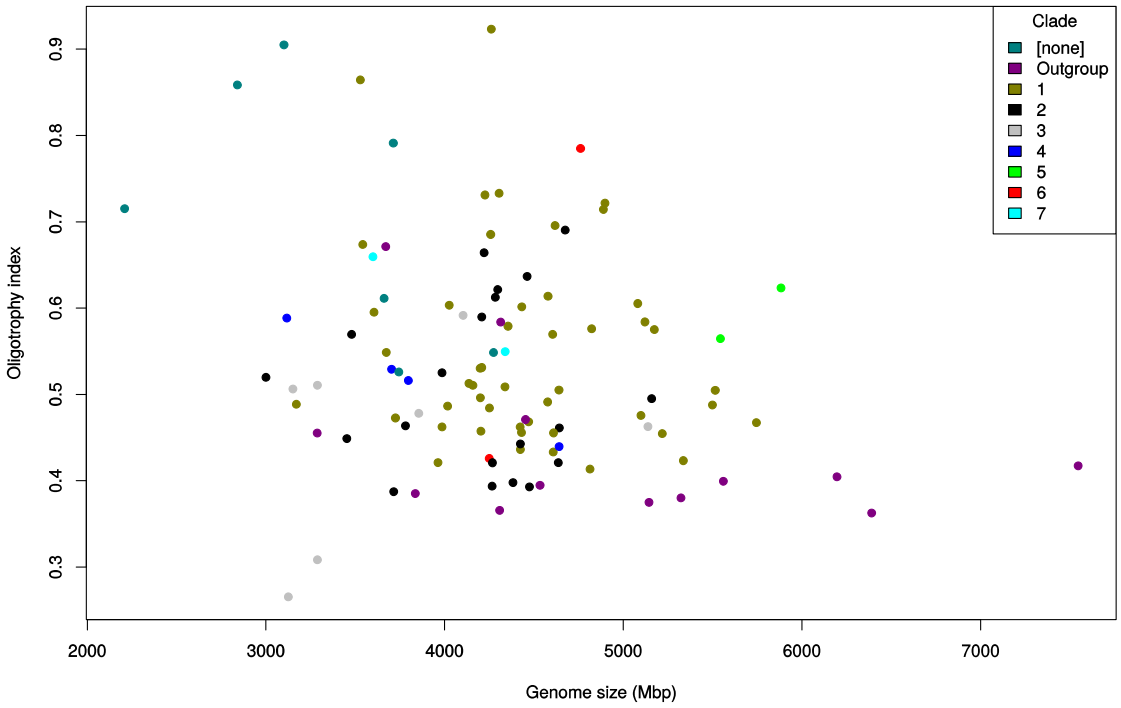
The trees used were inferred using ML from the indicated matrices. Permutation tests were conducted either with or without removing the outgroup strains from the tree. Length, number of steps (MP score) of the unpermuted tree; CI, consistency index; RI, retention index.

Tree	Dataset	Number	Character	Length	P-value	CI	RI
core genes	all	1	DnaA-like	22	0.001	0.091	0.608
core genes	all	2	RepA	52	0.008	0.096	0.365
core genes	all	3	RepB	28	0.001	0.071	0.480
core genes	all	4	RepABC	80	0.005	0.075	0.253
core genes	ingroup	1	DnaA-like	22	0.001	0.091	0.574
core genes	ingroup	2	RepA	50	0.061	0.100	0.274
core genes	ingroup	3	RepB	28	0.002	0.071	0.480
core genes	ingroup	4	RepABC	74	0.012	0.081	0.244
MARE	all	1	DnaA-like	22	0.001	0.091	0.608
MARE	all	2	RepA	52	0.004	0.096	0.365
MARE	all	3	RepB	28	0.001	0.071	0.480
MARE	all	4	RepABC	80	0.007	0.075	0.253

MARE	ingroup	1 DnaA-like	22	0.001	0.091	0.574
MARE	ingroup	2 RepA	50	0.063	0.100	0.274
MARE	ingroup	3 RepB	28	0.002	0.071	0.480
MARE	ingroup	4 RepABC	74	0.012	0.081	0.244
full	all	1 DnaA-like	22	0.001	0.091	0.608
full	all	2 RepA	50	0.001	0.100	0.392
full	all	3 RepB	27	0.001	0.074	0.500
full	all	4 RepABC	77	0.001	0.078	0.283
full	ingroup	1 DnaA-like	22	0.001	0.091	0.574
full	ingroup	2 RepA	48	0.009	0.104	0.306
full	ingroup	3 RepB	27	0.002	0.074	0.500
full	ingroup	4 RepABC	71	0.002	0.085	0.278

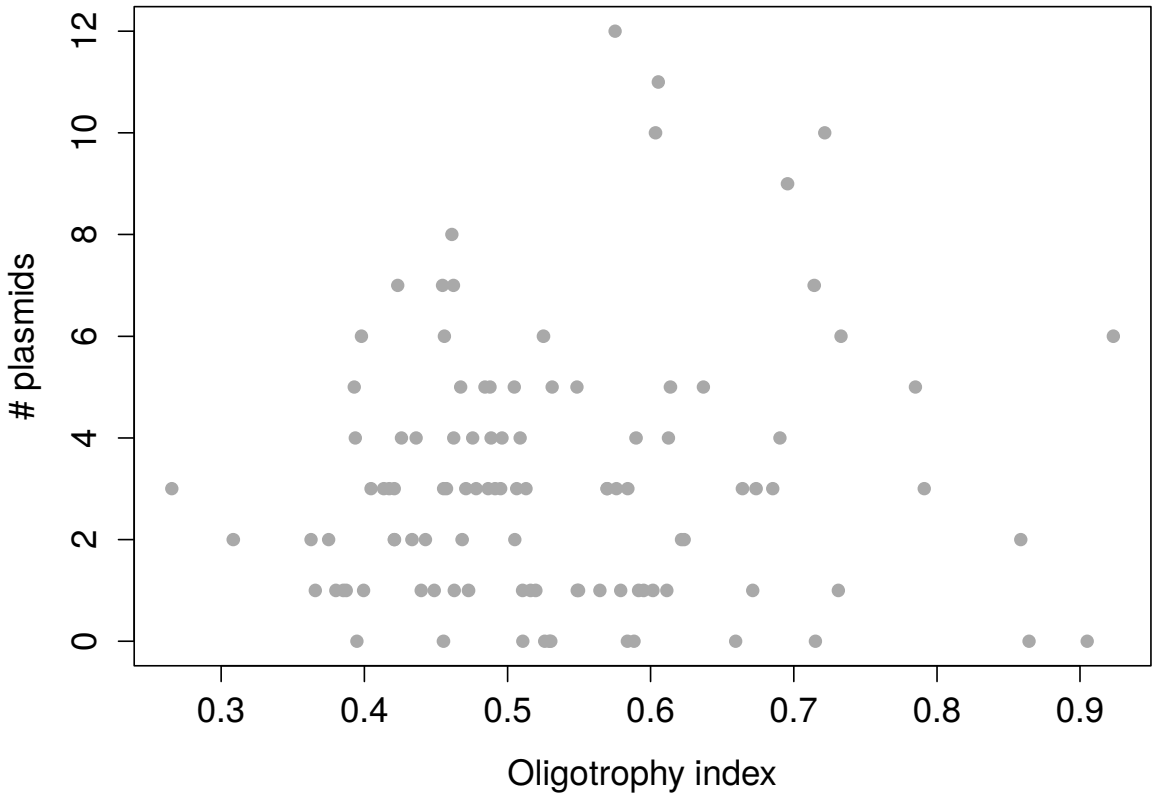
L. Relationship between genome size and oligotrophy

Oligotrophy index plotted against the genome size. Colours indicate major *Rhodobacteraceae* clades as defined in the main manuscript. The correlation, as determined using the Kendall coefficient, was not significant ($\alpha = 0.05$), but small genomes show a huge variety of oligotrophy indexes, whereas larger genomes are usually copiotrophic.



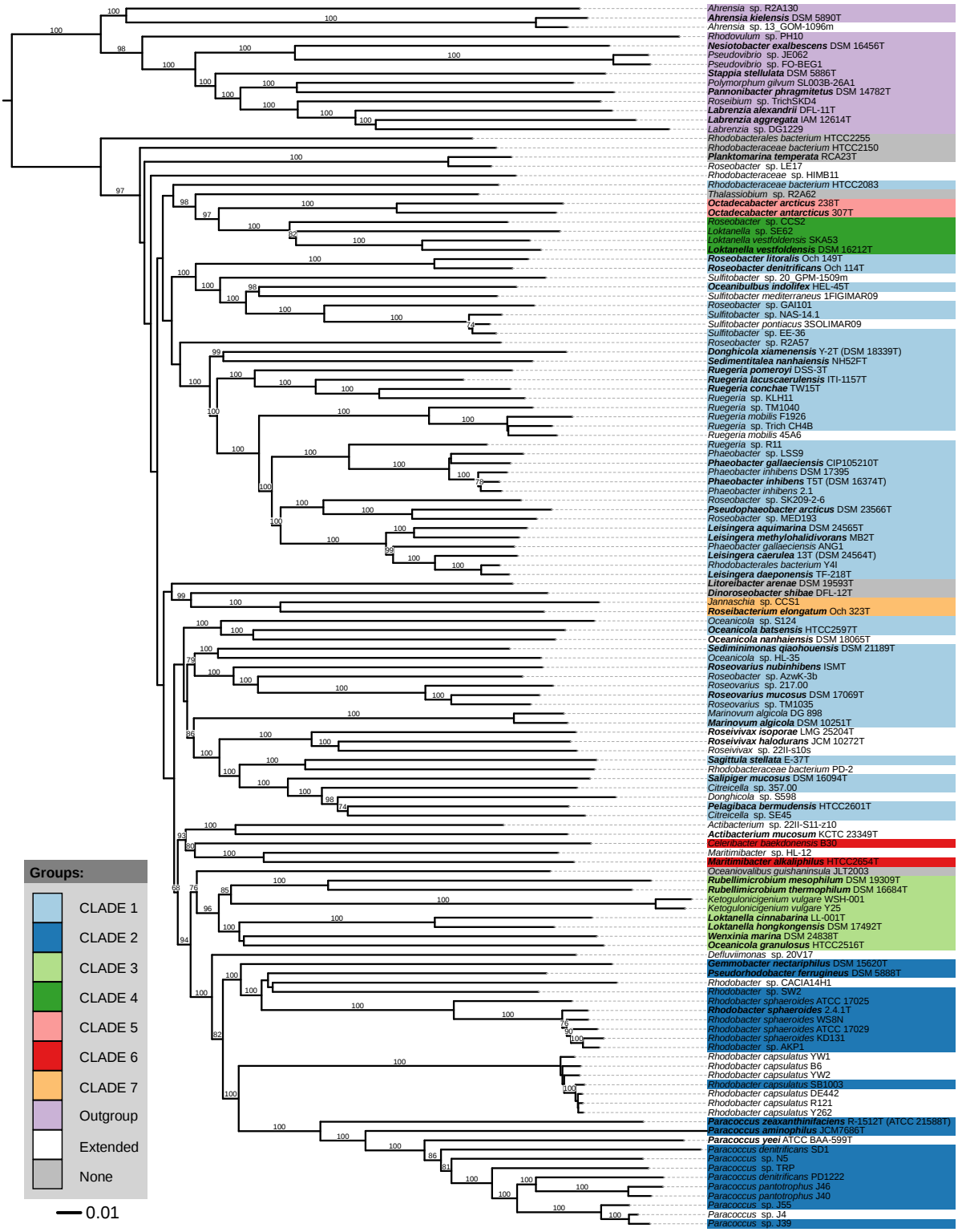
M. Relationship between genome size and number of ECRs

Number of extrachromosomal replicons (ECRs) plotted against the oligotrophy index. The correlation, as determined using the Kendall coefficient, was not significant (α = 0.05).



N. Whole proteome-based phylogenomic tree using the GBDP approach

Phylogeny is based on all pairwise intergenomic distances between the proteomes of 132 strains as calculated by GBDP's latest version (Meier-Kolthoff et al. 2013) and inferred using FastME v2.07 with TBR postprocessing (Lefort et al. 2015). The following GBDP settings were used: trimming algorithm, e-value threshold of 10^{-8} and formula d_5. Numbers above branches are greedy-with-trimming pseudo-bootstrap (Meier-Kolthoff et al. 2014) support values from 100 replicates and only branch support values above 60 % are shown. The tree was visualized using the web service 'Interactive Tree of Life' (Letunic and Bork 2011). Leaf labels are annotated according to their affiliation to previously established clades, with the only exception of the novel 'CLADE 8', which emerged from the addition of strain *Roseobacter* sp. LE17 to this increased dataset.


Groups:
CLADE 1
CLADE 2
CLADE 3
CLADE 4
CLADE 5
CLADE 6
CLADE 7
Outgroup
Extended
None
0.01
Ahrensia sp. R2A130
Ahrensia kielensis DSM 5890T
Ahrensia sp. 13_GOM-1096m
Rhodovulum sp. PH10
Nesiotobacter exalbescens DSM 16456T
Pseudovibrio sp. JE062
Pseudovibrio sp. FO-BEG1
Stappia stellulata DSM 5886T
Polymorphum gilvum SL003B-26A1
Pannonibacter phragmitetus DSM 14782T
Roseibium sp. TrichSKD4
Labrenzia alexandrii DFL-11T
Labrenzia aggregata IAM 12614T
Labrenzia sp. DG1229
Rhodobacterales bacterium HTCC2255
Rhodobacteraceae bacterium HTCC2150
Planktomarina temperata RCA23T
Roseobacter sp. LE17
Rhodobacteraceae sp. HIMB11
Rhodobacteraceae bacterium HTCC2083
Thalassiobium sp. R2A62
Octadecabacter arcticus 238T
Octadecabacter antarcticus 307T
Roseobacter sp. CCS2
Loktanella sp. SE62
Loktanella vestfoldensis SKA53
Loktanella vestfoldensis DSM 16212T
Roseobacter litoralis Och 149T
Roseobacter denitrificans Och 114T
Sulfitobacter sp. 20_GPM-1509m
Oceanibulbus indolifex HEL-45T
Sulfitobacter mediterraneus 1FIGIMAR09
Roseobacter sp. GAI101
Sulfitobacter sp. NAS-14.1
Sulfitobacter pontiacus 3SOLIMAR09
Sulfitobacter sp. EE-36
Roseobacter sp. R2A57
Donghicola xiamenensis Y-2T (DSM 18339T)
Sedimentitalea nanhaiensis NH52FT
Ruegeria pomeroyi DSS-3T
Ruegeria lacuscaerulensis ITI-1157T
Ruegeria conchae TW15T
Ruegeria sp. KLH11
Ruegeria sp. TM1040
Ruegeria mobilis F1926
Ruegeria sp. Trich CH4B
Ruegeria mobilis 45A6
Ruegeria sp. R11
Phaeobacter sp. LSS9
Phaeobacter gallaeciensis CIP105210T
Phaeobacter inhibens DSM 17395
Phaeobacter inhibens T5T (DSM 16374T)
Phaeobacter inhibens 2.1
Roseobacter sp. SK209-2-6
Pseudophaeobacter arcticus DSM 23566T
Roseobacter sp. MED193
Leisingera aquimarina DSM 24565T
Leisingera methylohalidivorans MB2T
Phaeobacter gallaeciensis ANG1
Leisingera caerulea 13T (DSM 24564T)
Rhodobacterales bacterium Y4I
Leisingera daeponensis TF-218T
Litoreibacter arenae DSM 19593T
Dinoroseobacter shibae DFL-12T
Jannaschia sp. CCS1
Roseibacterium elongatum Och 323T
Oceanicola sp. S124
Oceanicola batsensis HTCC2597T
Oceanicola nanhaiensis DSM 18065T
Sediminimonas qiaohouensis DSM 21189T
Oceanicola sp. HL-35
Roseovarius nubinhibens ISMT
Roseobacter sp. AzwK-3b
Roseovarius sp. 217.00
Roseovarius mucosus DSM 17069T
Roseovarius sp. TM1035
Marinovum algicola DG 898
Marinovum algicola DSM 10251T
Roseivivax isoporae LMG 25204T
Roseivivax halodurans JCM 10272T
Roseivivax sp. 22II-s10s
Sagittula stellata E-37T
Rhodobacteraceae bacterium PD-2
Salipiger mucosus DSM 16094T
Citreicella sp. 357.00
Donghicola sp. S598
Pelagibaca bermudensis HTCC2601T
Citreicella sp. SE45
Actibacterium sp. 22II-S11-z10
Actibacterium mucosum KCTC 23349T
Paracoccus haeundaensis B30
Maritimibacter sp. HL-12
Maritimibacter alkaliphilus HTCC2654T
Oceaniovalibus guishaninsula JLT2003
Rubellimicrobium mesophilum DSM 19309T
Rubellimicrobium thermophilum DSM 16684T
Ketogulonicigenium vulgare WSH-001
Ketogulonicigenium vulgare Y25
Loktanella cinnabarina LL-001T
Loktanella hongkongensis DSM 17492T
Wenxinia marina DSM 24838T
Oceanicola granulosus HTCC2516T
Defluviimonas sp. 20V17
Defluviimonas denitrificans DSM 15620T
Pseudorhodobacter ferrugineus DSM 5888T
Rhodobacter sp. CACIA14H1
Rhodobacter sp. SW2
Rhodobacter sphaeroides ATCC 17025
Rhodobacter sphaeroides 2.4.1T
Rhodobacter sphaeroides WS8N
Rhodobacter sphaeroides ATCC 17029
Rhodobacter sphaeroides KD131
Rhodobacter sp. AKP1
Rhodobacter capsulatus YW1
Rhodobacter capsulatus B6
Rhodobacter capsulatus YW2
Rhodobacter capsulatus SB1003
Rhodobacter capsulatus DE442
Rhodobacter capsulatus R121
Rhodobacter capsulatus Y262
Paracoccus zeaxanthinifaciens R-1512T (ATCC 21588T)
Paracoccus aminophilus JCM7686T
Paracoccus yeei ATCC BAA-599T
Paracoccus denitrificans SD1
Paracoccus sp. N5
Paracoccus sp. TRP
Paracoccus denitrificans PD1222
Paracoccus pantotrophus J46
Paracoccus pantotrophus J40
Paracoccus sp. J55
Paracoccus sp. J4
Paracoccus sp. J39

O. Corrected affiliations of strains to *Rhodobacteraceae* species and genera as inferred from their genomes

Our and previous analyses (Newton et al. 2010; Luo and Moran 2014) show that several genera of the Roseobacter group are not monophyletic, such as *Oceanicola*, which is spread over clades 1 and 3, and is not supported by 16S rRNA gene analysis either (Breider et al. 2014). This finding calls for the reclassification of the affected species. Suggestion for revised names of several strains are given here. Species affiliations have also to be corrected in some instances such as *Rhodobacterales* bacterium Y4I, for which digital DNA:DNA hybridization clearly shows that it is a *Leisingera daeponensis* strain.

All pairwise intergenomic distances between the genome sequences of the 132 strains were calculated using GBDP's latest version (Meier-Kolthoff et al. 2013). Under the established (Meier-Kolthoff et al. 2013) species delimitation thresholds, corresponding to 70 % DDH, the strains affiliation to known type strains was assessed. Recommendations for the renaming of incorrectly identified strains are given in the below table. Note that the maximum subtree height (Scheuner et al. 2014) was observed for the genus *Rubellimicrobum* (*R. thermophilum* DSM 16684^{T} and *R. mesophilum* DSM 19309^{T}).

Current name	**Recommended renaming**	**Comment**
Citreicella sp. SE45	*Pelagibaca* sp. SE45	height of the subtree is smaller than the maximum subtree height of established genera
Citreicella sp. 357	*Pelagibaca* sp. 357	height of the subtree is smaller than the maximum subtree height of established genera
Loktanella strains	[None, reclassification necessary]	*Loktanellas* is found to be non-monophyletic; *L. salsilacus* is type species, but not included in the analyses; thus either *L. cinnabarina* LL-001^{T} *and L. hongkongensis* DSM 17492^{T} or *L. vestfoldensis* DSM 16212 ^{T}, *L. vestfoldensis* SKA53 and *L.* sp. SE62 should be put in a new genus
Oceanicola batsensis HTCC2597^{T}	[None, reclassification necessary]	*Oceanicola* is found to be non-monophyletic; *O. granulosus* is type species;

		thus *O. batsensis* should be put in a new genus
Oceanicola sp. HL-35	*Sediminimonas* sp. HL-35	height of the subtree is smaller than the maximum subtree height of established genera
Oceanicola sp. S124	[None]	same genus as *O. batsensis*
Phaeobacter gallaeciensis ANG1	*Leisingera* sp. ANG1	< 70 % DDH, but clusters within strains of the genus *Leisingera*
Rhodobacter capsulatus SB1003	*Rhodobacteraceae* gen. sp. SB1003	height of the subtree is larger than the maximum subtree height of established genera
Rhodobacter sp. AKP1	*Rhodobacter sphaeroides* AKP1	≥ 70 % DDH to *Rhodobacter sphaeroides* 2.4.1^{T}
Rhodobacter sp. SW2	*Pseudorhodobacter* sp. SW2	height of the subtree is smaller than the maximum subtree height of established genera
Rhodobacterales bacterium Y4I	*Leisingera daeponensis* Y4I	≥ 70 % DDH to *Leisingera daeponensis* TF-218^{T}
Roseobacter sp. AzwK-3b	*Roseovarius* sp. AzwK-3b	< 70 % DDH, but clusters within strains of the genus *Roseovarius*
Roseobacter sp. CCS2	*Loktanella* sp. CCS 2	< 70 % DDH, but clusters within strains of the genus *Loktanella*; see comment for the *Loktanella* strains
Roseobacter sp. GAI101	*Rhodobacteraceae* gen. sp. GAI101	
Roseobacter sp. LE17	*Planktomarina tempera* LE17	≥ 70 % DDH to *Planktomarina tempera* RCA23^{T}
Roseobacter sp. MED193	*Pseudophaeobacter* sp. MED193	

Roseobacter sp. R2A57	*Rhodobacteraceae* gen. sp. R2A57	
Roseobacter sp. SK209-2-6	*Rhodobacteraceae* gen. sp. SK209-2-6	
Ruegeria mobilis F1926	*Rhodobacteraceae* gen. sp. F1926	
Ruegeria sp. R11	*Nautella italica* R11	
Ruegeria sp. TM1040	*Rhodobacteraceae* gen. sp. TM1040	
Ruegeria sp. TrichCH4B	*Rhodobacteraceae* gen. sp. TrichCH4B	

References

Arunasri K, Venkata Ramana V, Spröer C, Sasikala C, Ramana C V. 2008. *Rhodobacter megalophilus* sp. nov., a phototroph from the Indian Himalayas possessing a wide temperature range for growth. Int J Syst Evol Microbiol 58: 1792–1796. doi:10.1099/ijs.0.65642-0.

Breider S, Scheuner C, Schumann P, Fiebig A, Petersen J, Pradella S, Klenk HP, Brinkhoff T, Göker M. 2014. Genome-scale data suggest reclassifications in the Leisingera-Phaeobacter cluster including proposals for *Sedimentitalea* gen. nov. and *Pseudophaeobacter* gen. nov. *Front Microbiol*. 5:416.

Choudhary M, Zanhua X, Fu YX, Kaplan S. 2007. Genome Analyses of Three Strains of *Rhodobacter sphaeroides*: Evidence of Rapid Evolution of Chromosome II. J Bacteriol 189: 1914–1921. doi:10.1128/JB.01498-06.

Denner EBM, Kolari M, Hoornstra D, Tsitko I, Kämpfer P, et al. 2006. *Rubellimicrobium thermophilum* gen. nov., sp. nov., a red-pigmented, moderately thermophilic bacterium isolated from coloured slime deposits in paper machines. Int J Syst Evol Microbiol 56: 1355–1362. doi:10.1099/ijs.0.63751-0.

Dogs M, Voget S, Teshima H, Petersen J, Davenport K, Dalingault H, Chen A, Pati A, Ivanova N, Goodwin LA et al. 2013. Genome sequence of *Phaeobacter inhibens* type strain (T5T), a secondary metabolite producing representative of the marine *Roseobacter* clade, and emendation of the species description of *Phaeobacter inhibens*. *Stand Genomic Sci*. 9:334–350.

Donachie SP, Bowman JP, Alam M. 2006. *Nesiotobacter exalbescens* gen. nov., sp. nov., a moderately thermophilic alphaproteobacterium from an Hawaiian hypersaline lake. Int J Syst Evol Microbiol 56: 563–567. doi:10.1099/ijs.0.63440-0.

Fiebig A, Riedel T, Gronow S, Petersen J, Klenk H-P, et al. 2013. Genome sequence of the reddish-pigmented *Rubellimicrobium thermophilum* type strain (DSM 16684T), a member of the *Rhodobacterales* clade. Stand Genomic Sci 8: 480–490. doi:10.4056/sigs.4247911.

Frank O, Pradella S, Rohde M, Scheuner C, Klenk H-P, Göker M, Petersen J. 2014. Complete genome sequence of the *Phaeobacter gallaeciensis* type strain CIP 105210T (= DSM 26640T = BS107T). *Stand Genomic Sci*. 9: 914–932.

Frank O, Göker M, Pradella S, Petersen J. 2015a. Ocean's twelve: Flagellar and biofilm chromids in the multipartite genome of *Marinovum algicola* DG898 exemplify functional compartmentalization in Proteobacteria. Environ Microbiol. 17: 4019–4034.

Harrison PW, Lower RPJ, Kim NKD, Young JPW. 2010. Introducing the bacterial "chromid": Not a chromosome, not a plasmid. *Trends Microbiol*. 18:141–148.

Kahle D, Wickham H. 2013. ggmap: Spatial Visualization with ggplot2. The R Journal 5:144–161.

Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol 2015 32: 2798–2800.

Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39: W475–W478. doi:10.1093/nar/gkr201.

Lindemann SR, Moran JJ, Stegen JC, Renslow RS, Hutchison JR, et al. 2013. The epsomitic phototrophic microbial mat of Hot Lake, Washington: community structural responses to seasonal cycling. Front Microbiol 4: 323. doi:10.3389/fmicb.2013.00323.

Lopez, R., Cowley, A., Li, W. and McWilliam, H. 2014. Using EMBL-EBI Services via Web Interface and Programmatically via Web Services. Curr Protoc Bioinform 48: 3.12.1-3.12.50.

doi:10.1002/0471250953.bi0312s48

Luo H, Moran MA. 2014. Evolutionary ecology of the marine *Roseobacter* clade. Microbiol Mol Biol Rev. 78:573–587.

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14: 60. doi:10.1186/1471-2105-14-60.

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2014. Highly parallelized inference of large genome-based phylogenies. Concurr Comput Pract Exp 26: 1715–1729. doi:10.1002/cpe.3112.

Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE, Howard EC, King E, Oakley CA, Reisch CR, et al. 2010. Genome characteristics of a generalist marine bacterial lineage. ISME J. 4:784–798.

Petersen J, Brinkmann H, Pradella S. 2009. Diversity and evolution of repABC type plasmids in Rhodobacterales. *Environ Microbiol*. 11:2627–2638.

Petersen J, Frank O, Göker M, Pradella S. 2013. Extrachromosomal, extraordinary and essential - The plasmids of the Roseobacter clade. *Appl Microbiol Biotechnol*. 97:2805–2815.

Pétursdóttir SK, Kristjánsson JK. 1997. *Silicibacter lacuscaerulensis* gen. nov., sp. nov., a mesophilic moderately halophilic bacterium characteristic of the Blue Lagoon geothermal lake in Iceland. Extremophiles 1: 94–99.

Scheuner C, Tindall BJ, Lu M, Nolan M, Lapidus A, et al. 2014. Complete genome sequence of *Planctomyces brasiliensis* type strain (DSM 5305^T), phylogenomic analysis and reclassification of *Planctomycetes* including the descriptions of *Gimesia* gen. nov., *Planctopirus* gen. nov. and *Rubinisphaera* gen. nov. and emended descriptions of the order *Planctomycetales* and the family *Planctomycetaceae*. Stand Genomic Sc 9: 10. doi:10.1186/1944-3277-9-10.

Urbance JW, Bratina BJ, Stoddard SF, Schmidt TM. 2001. Taxonomic characterization of *Ketogulonigenium vulgare* gen. nov., sp. nov. and *Ketogulonigenium robustum* sp. nov., which oxidize L-sorbose to 2-keto-L-gulonic acid. Int J Syst Evol Microbiol 51: 1059–1070.

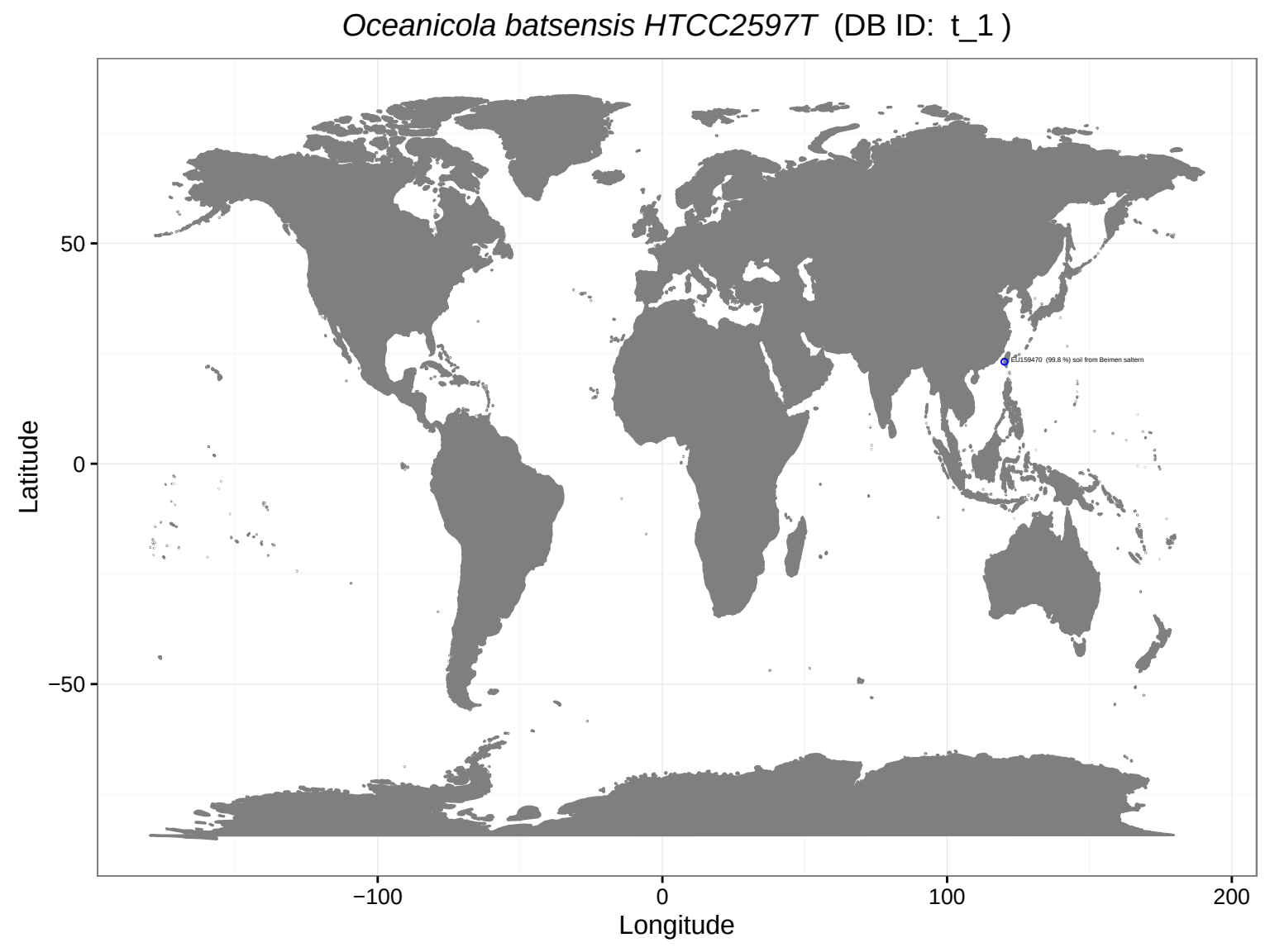
Venkata Ramana V, Sasikala C, Ramana C V. 2008. *Rhodobacter maris* sp. nov., a phototrophic alphaproteobacterium isolated from a marine habitat of India. Int J Syst Evol Microbiol 58: 1719–1722. doi:10.1099/ijs.0.65638-0.

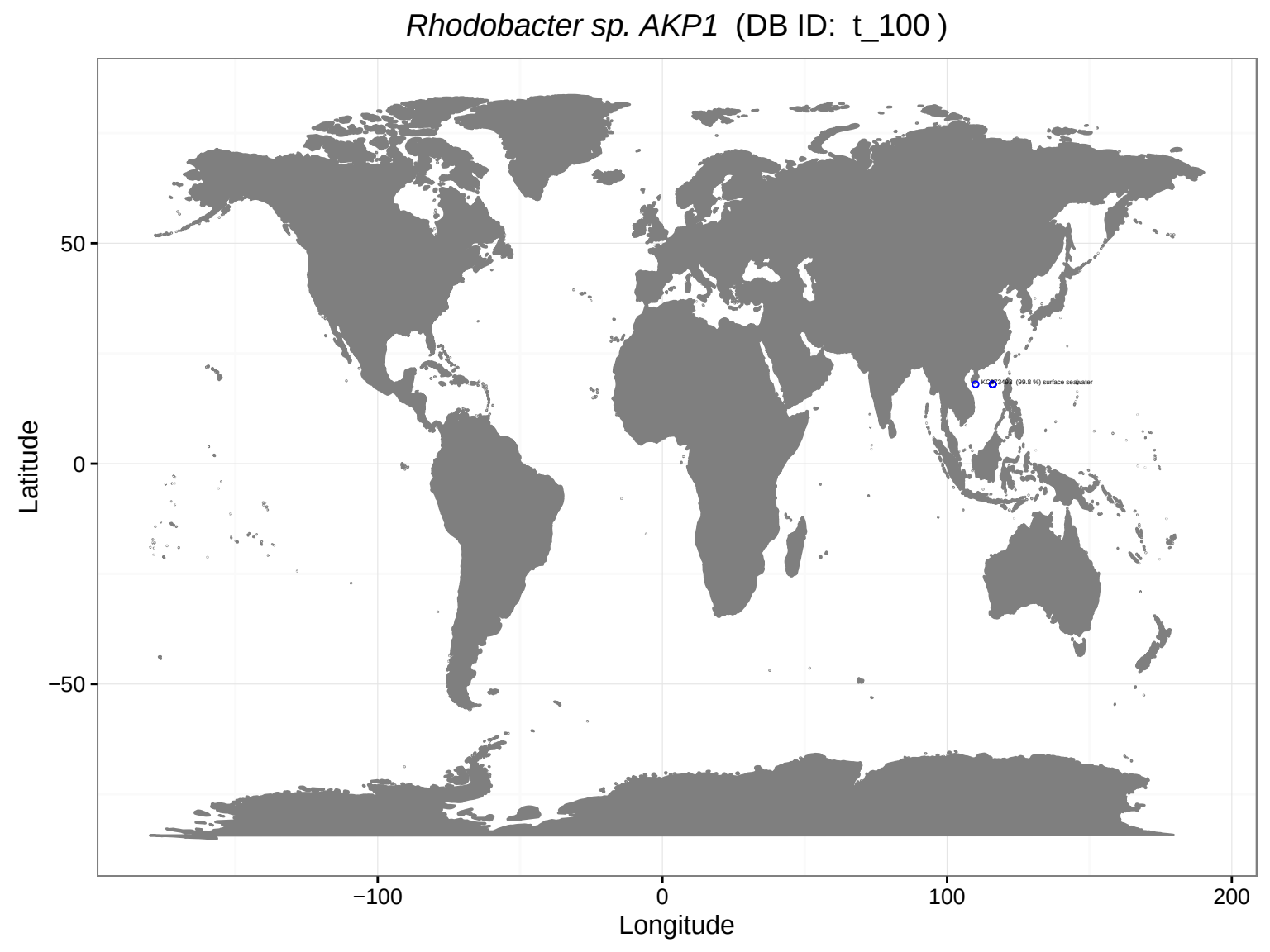
Wang Y-X, Wang Z-G, Liu J-H, Chen Y-G, Zhang X-X, et al. 2009. *Sediminimonas qiaohouensis* gen. nov., sp. nov., a member of the *Rhodobacterales* clade in the order *Rhodobacterales*. Int J Syst Evol Microbiol 59: 1561–1567. doi:10.1099/ijs.0.006965-0.

Weon H-Y, Son J-A, Yoo S-H, Hong S-B, Jeon Y-A, et al. 2009. *Rubellimicrobium aerolatum* sp. nov., isolated from an air sample in Korea. Int J Syst Evol Microbiol 59: 406–410. doi:10.1099/ijs.0.65856-0.

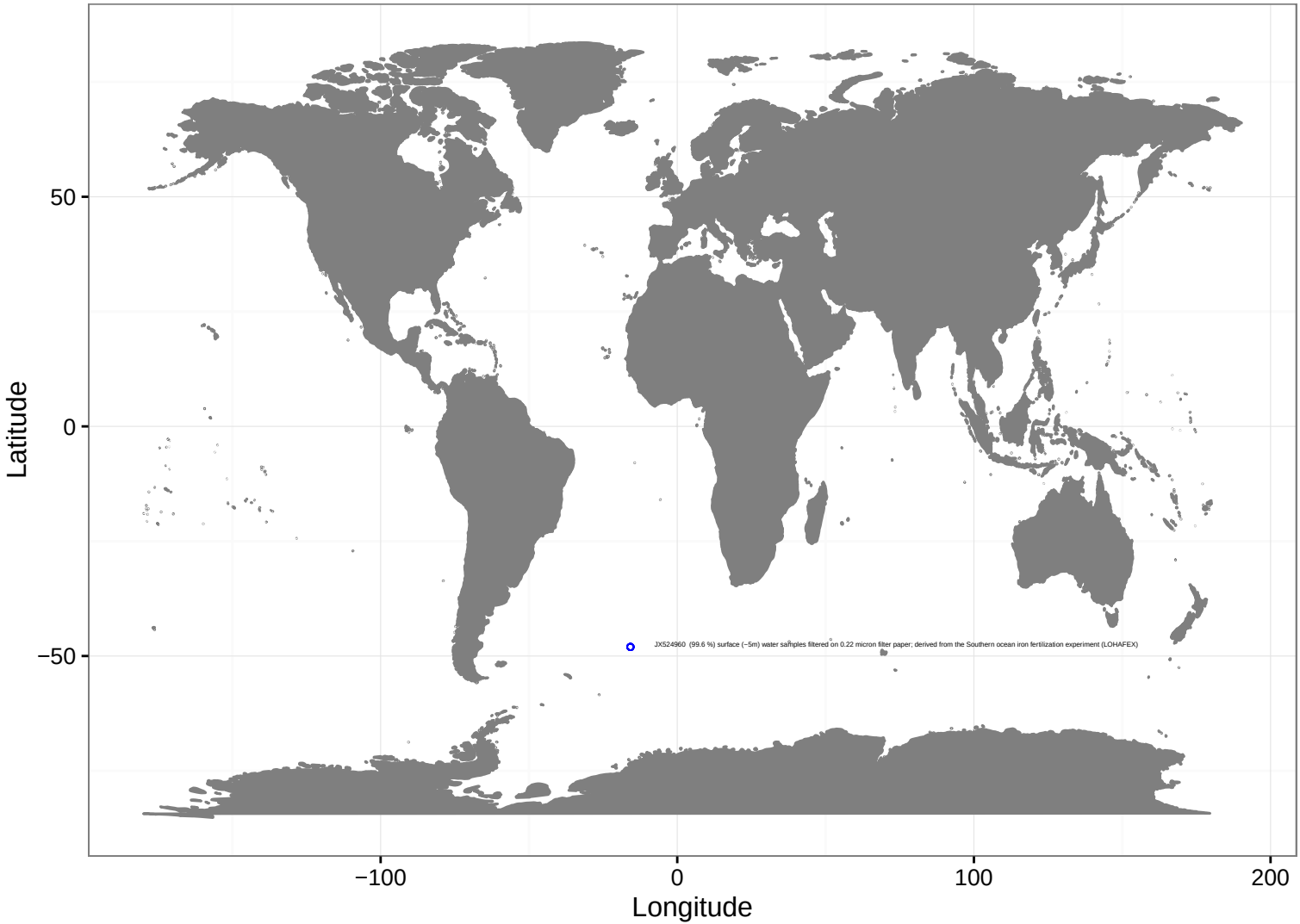
Appendix

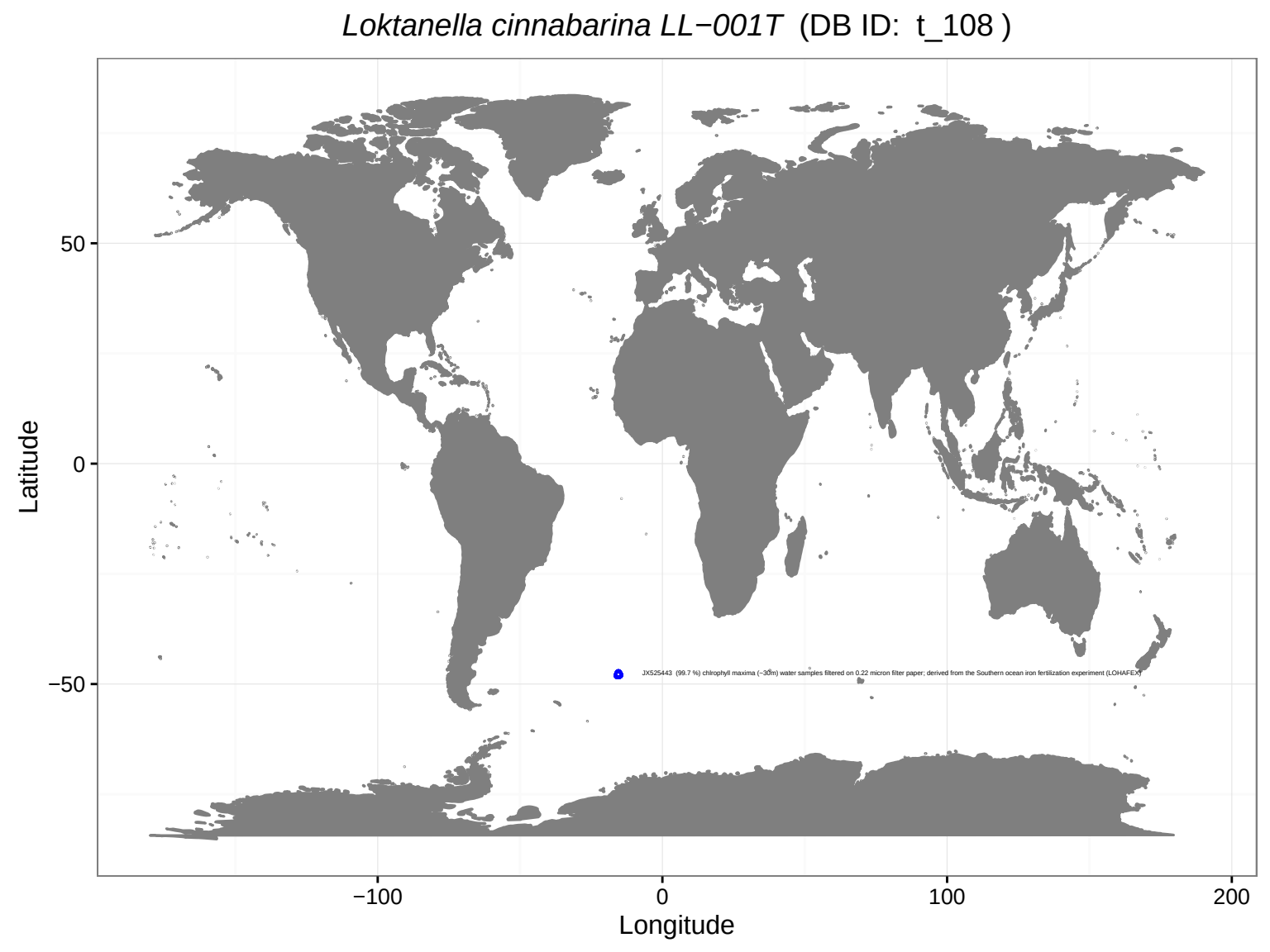
Results of BLASTN searches against the EMBL em_geo_rel_env environmental database. Each circle represents a the geographic location of a hit and is coloured according to the habitat assigned to the query as used in this study (blue: marine; red: non-marine/saline; grey: undefined). The labels associated with each hit provide information on the sampling site as included in the EMBL database.

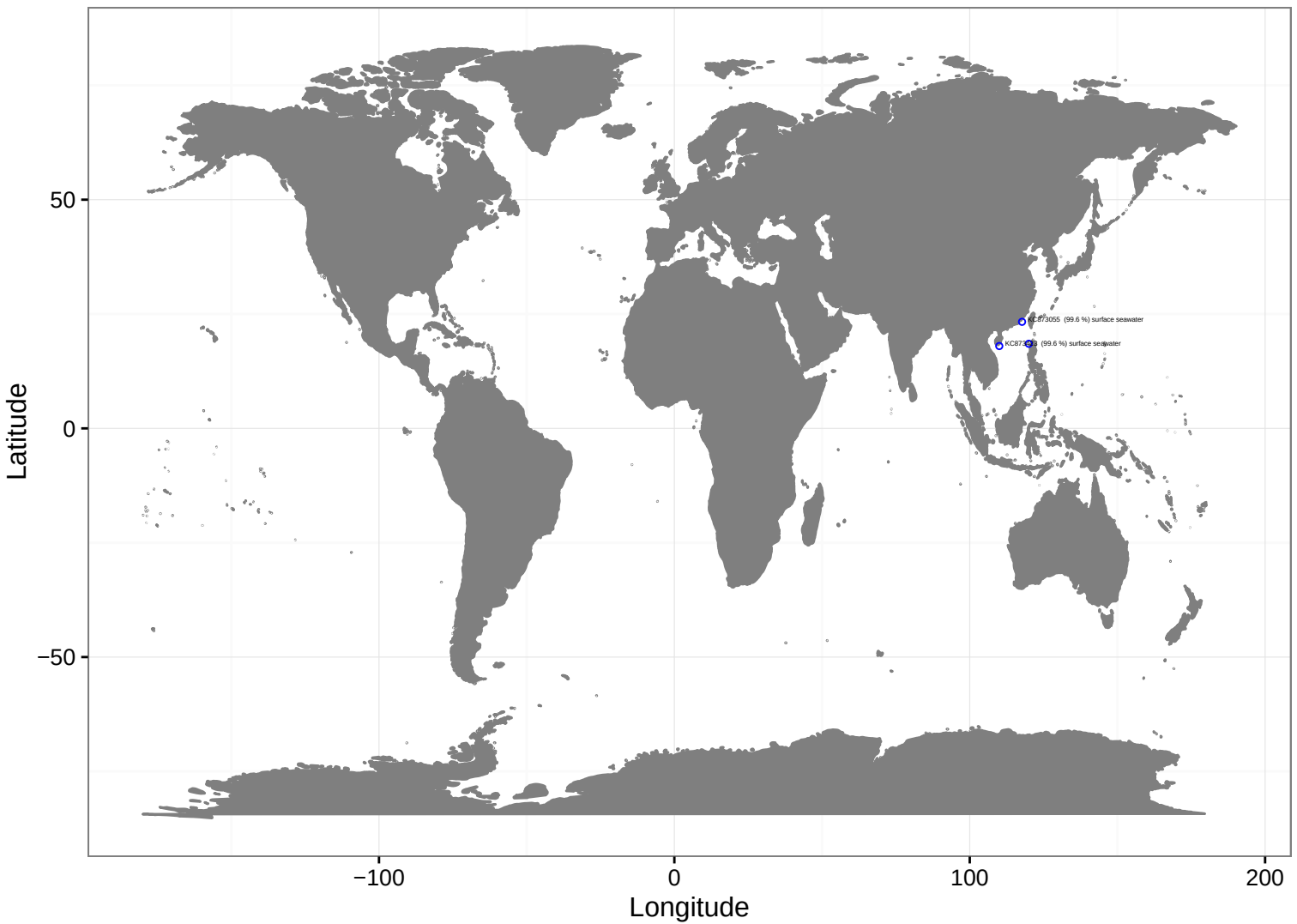

Oceanicola batsensis HTCC2597T (DB ID: t_1)
EU159470 (99.8 %) soil from Beimen saltern
50
0
−50
Latitude
−100
0
100
200
Longitude


Rhodobacter sp. AKP1 (DB ID: t_100)
50
0
−50
Latitude
−100
0
100
200
Longitude

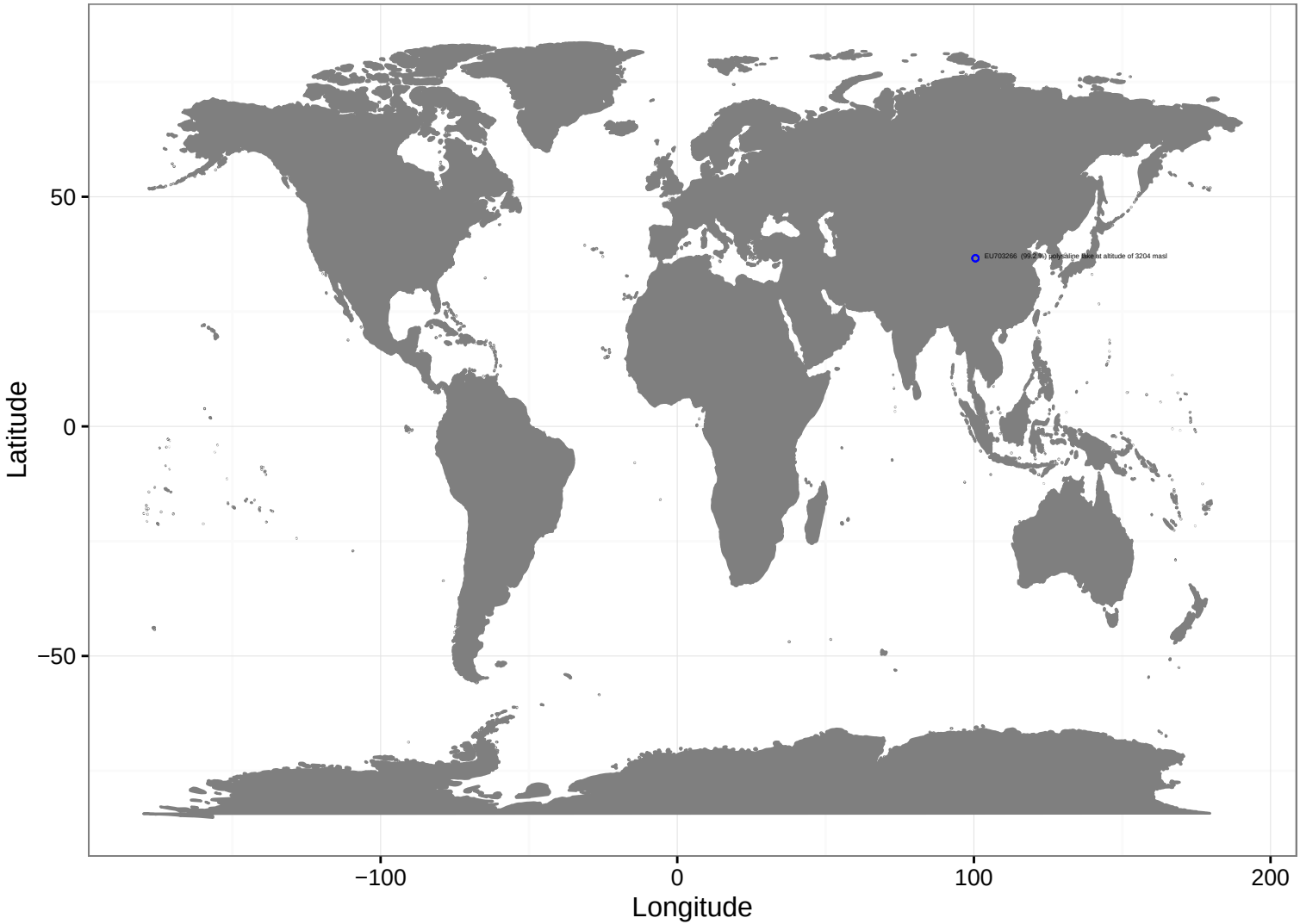
Celeribacter baekdonensis B30 (DB ID: t_101)

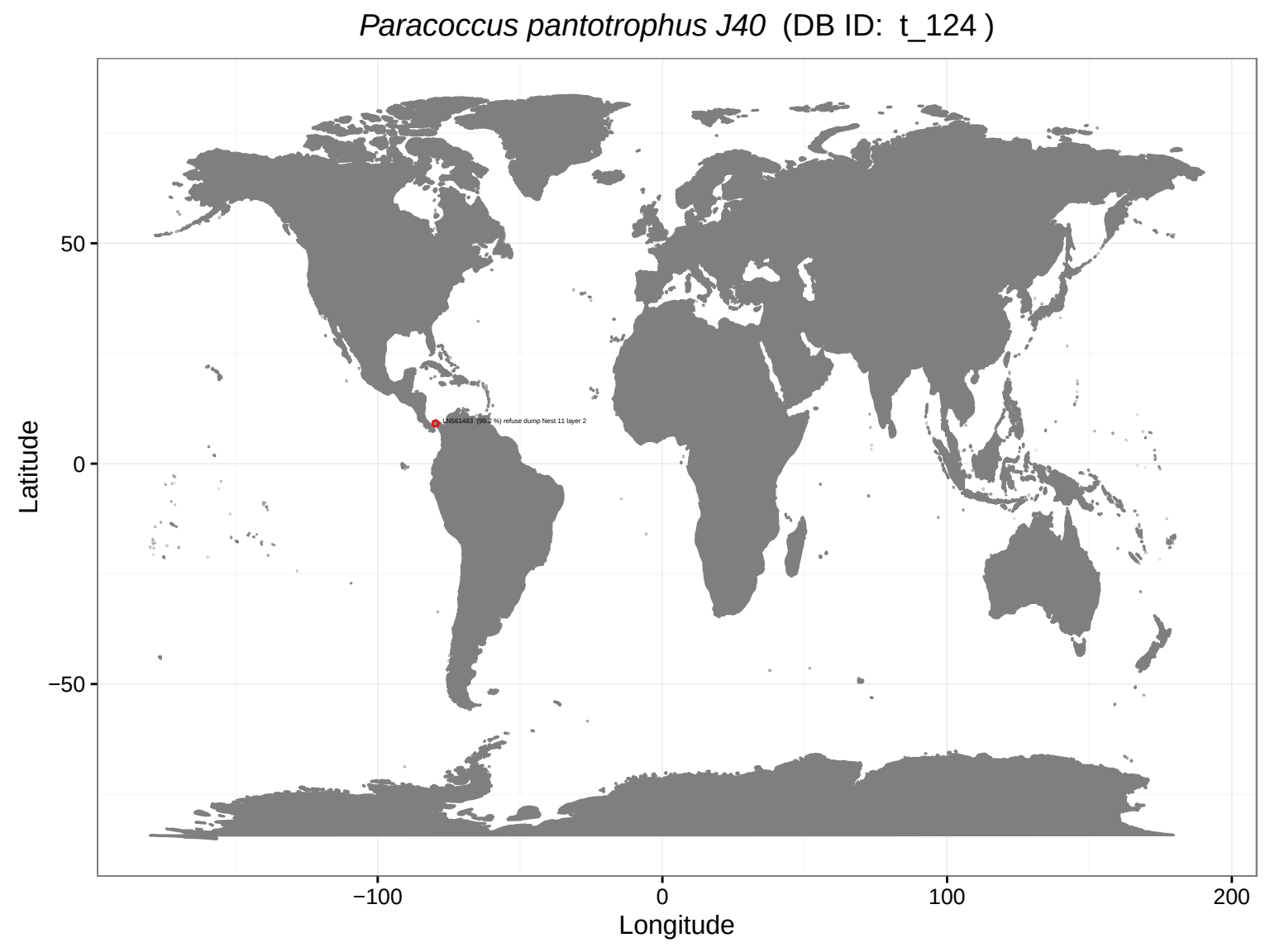


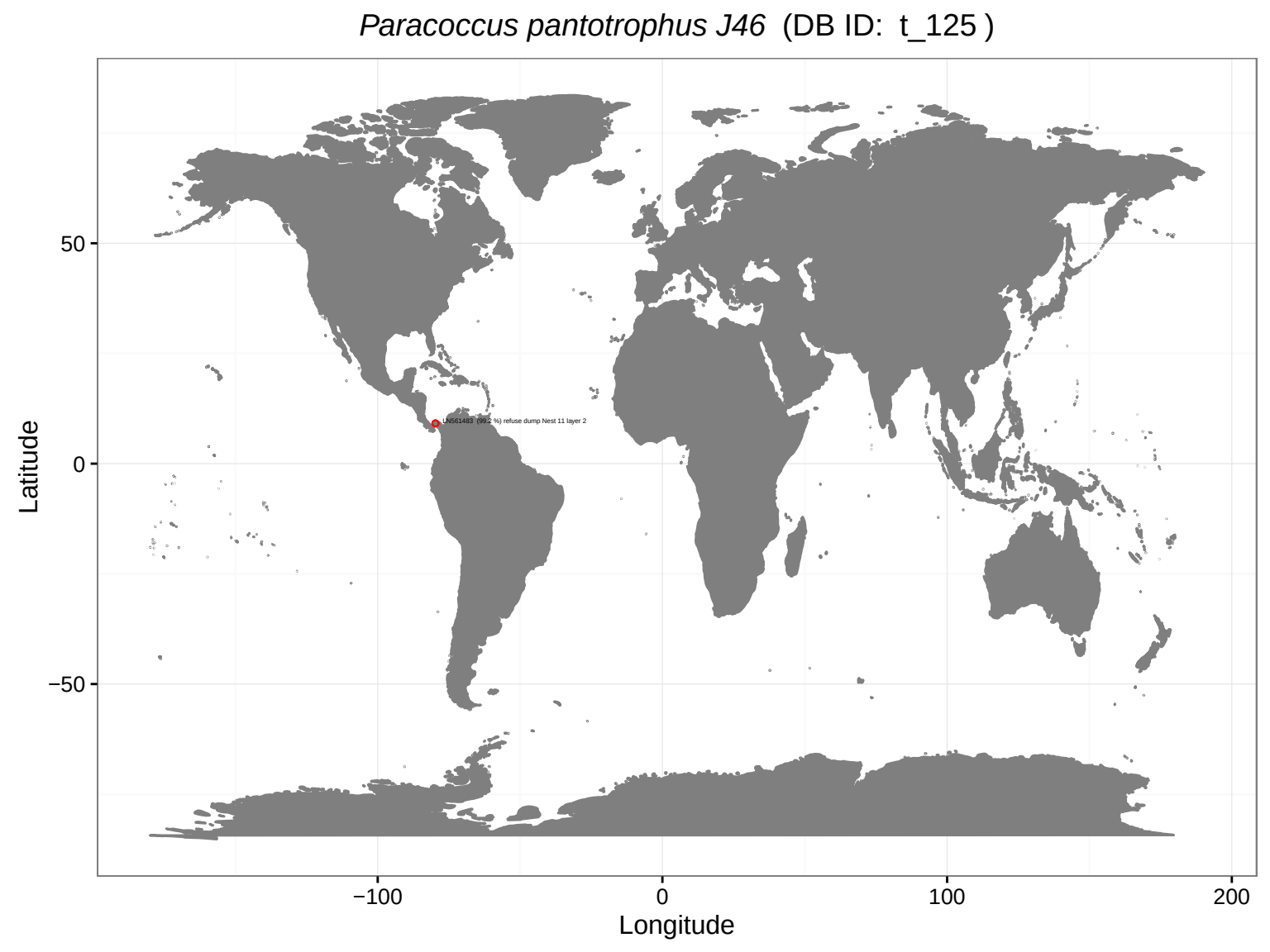

Loktanella cinnabarina LL−001T (DB ID: t_108)
JX525443 (99.7 %) chlrophyll maxima (~30m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
50
0
−50
Latitude
−100
0
100
200
Longitude

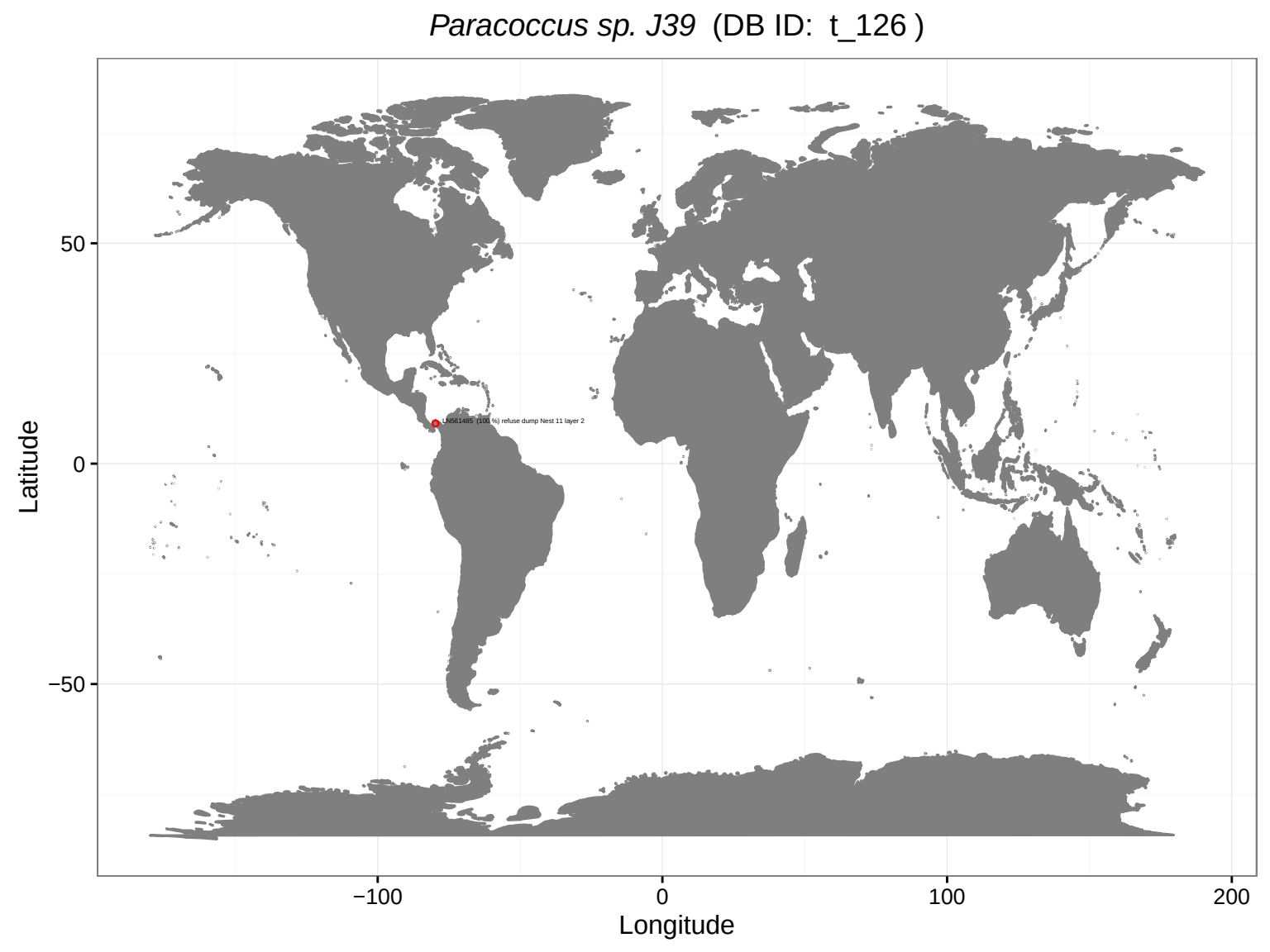

Ruegeria mobilis F1926 (DB ID: t_110)
50
0
−50
Latitude
−100
0
100
200
Longitude
KC873055 (99.6 %) surface seawater
KC873 (99.6 %) surface seawater

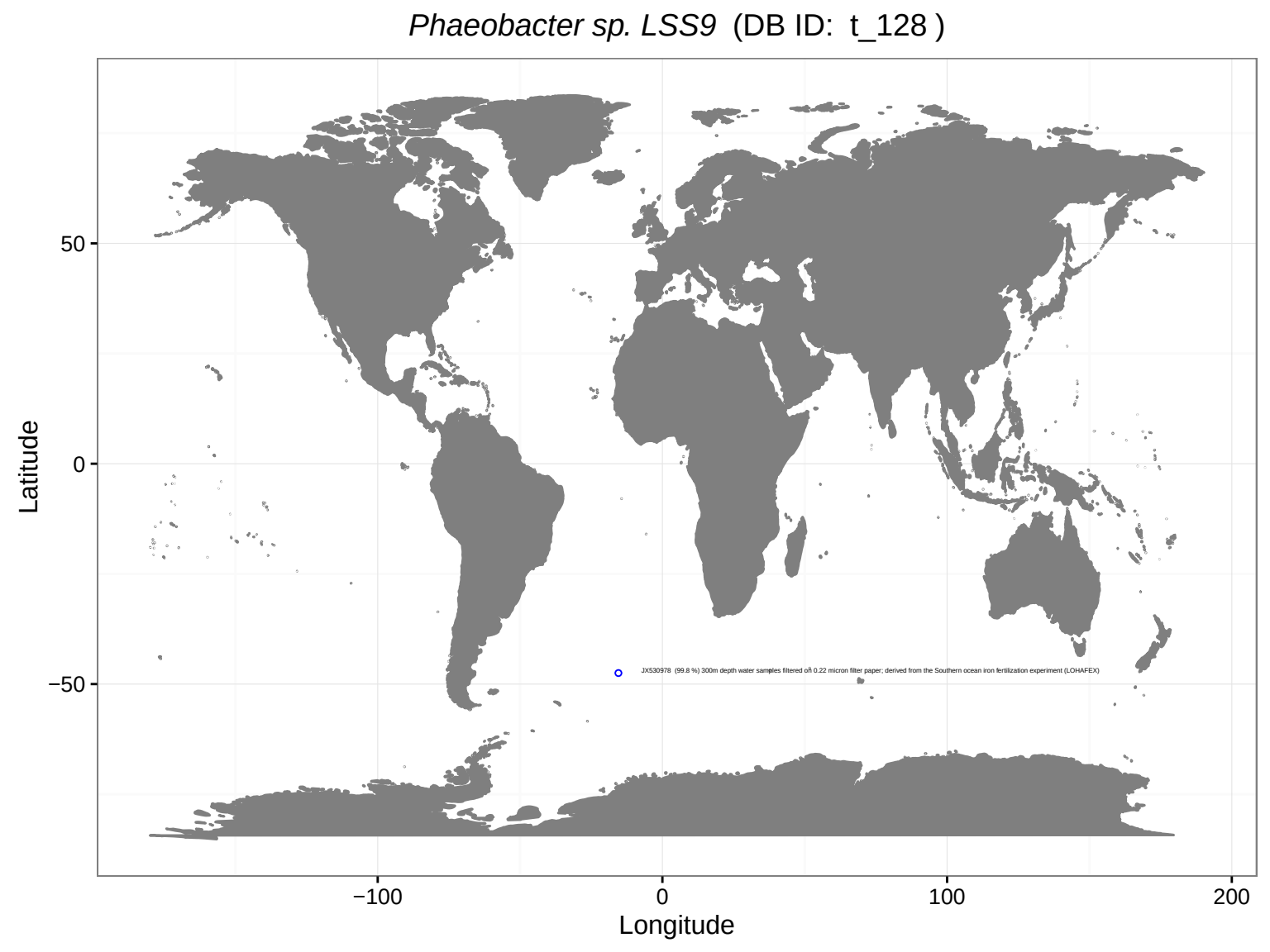
Loktanella vestfoldensis DSM 16212T (DB ID: t_111)

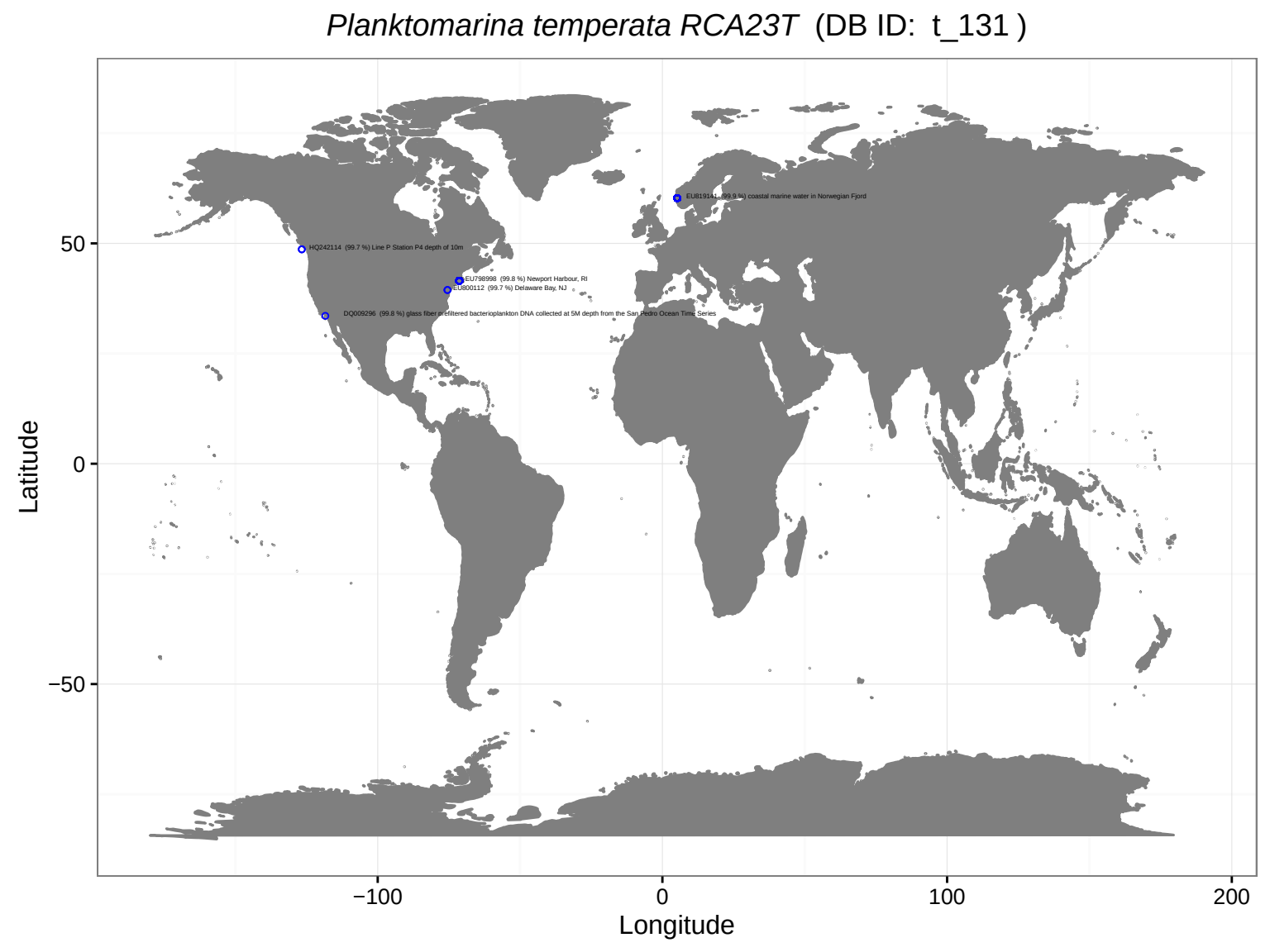


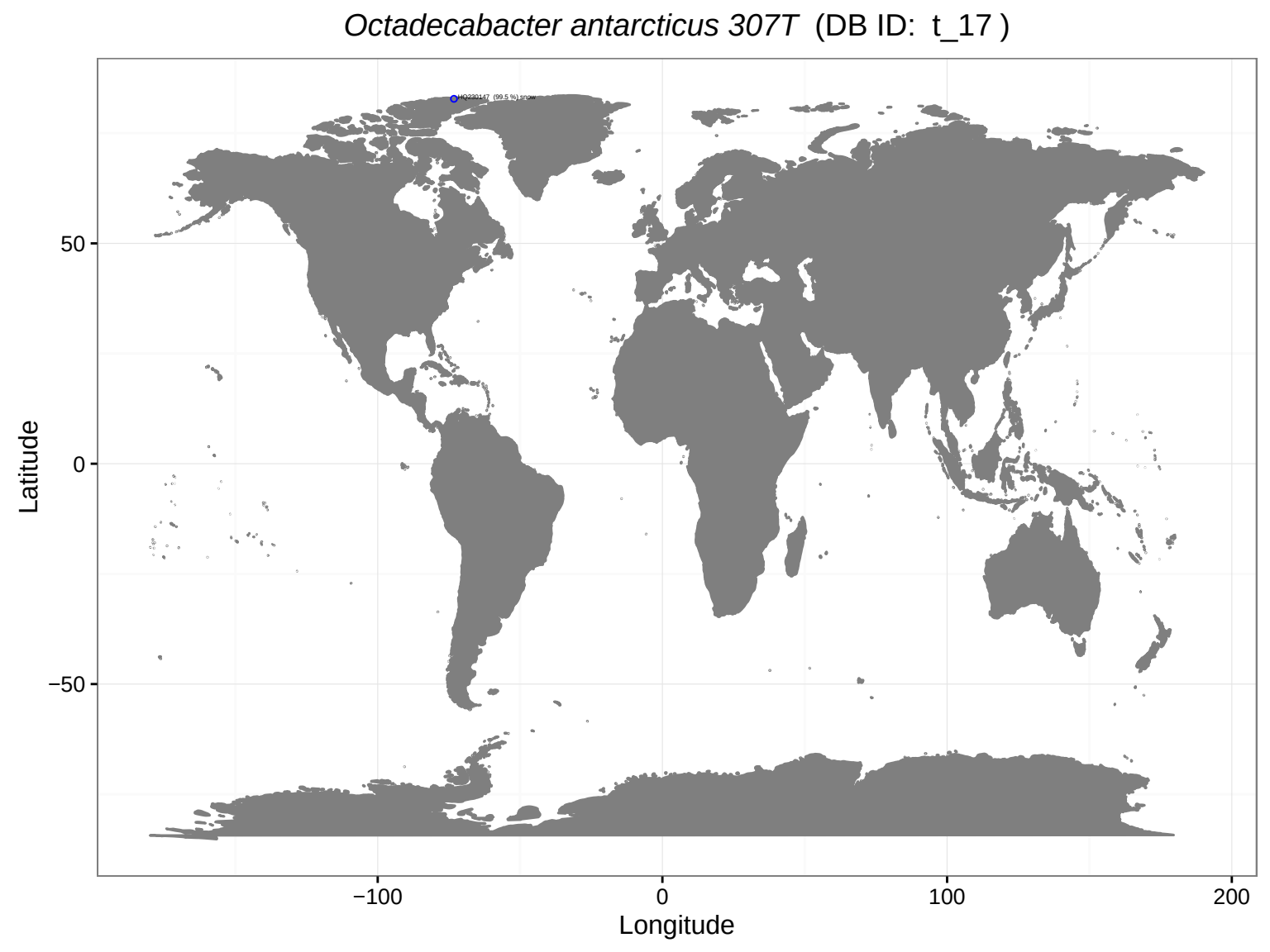

Paracoccus pantotrophus J40 (DB ID: t_124)
LN561483 (99.2 %) refuse dump Nest 11 layer 2
50
0
−50
Latitude
−100
0
100
200
Longitude

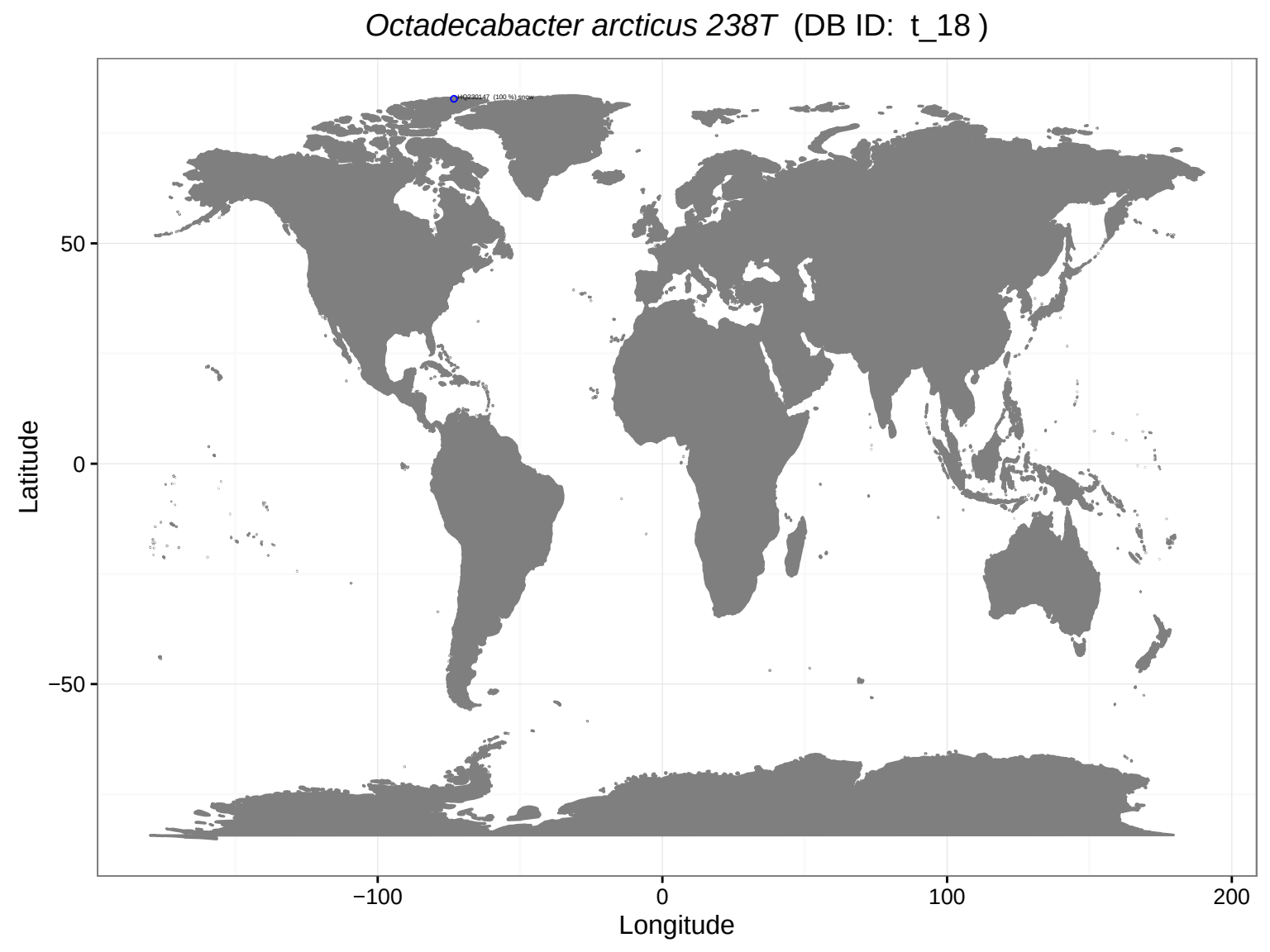

Paracoccus pantotrophus J46 (DB ID: t_125)
LN561483 (99.2 %) refuse dump Nest 11 layer 2
50
0
−50
Latitude
−100
0
100
200
Longitude

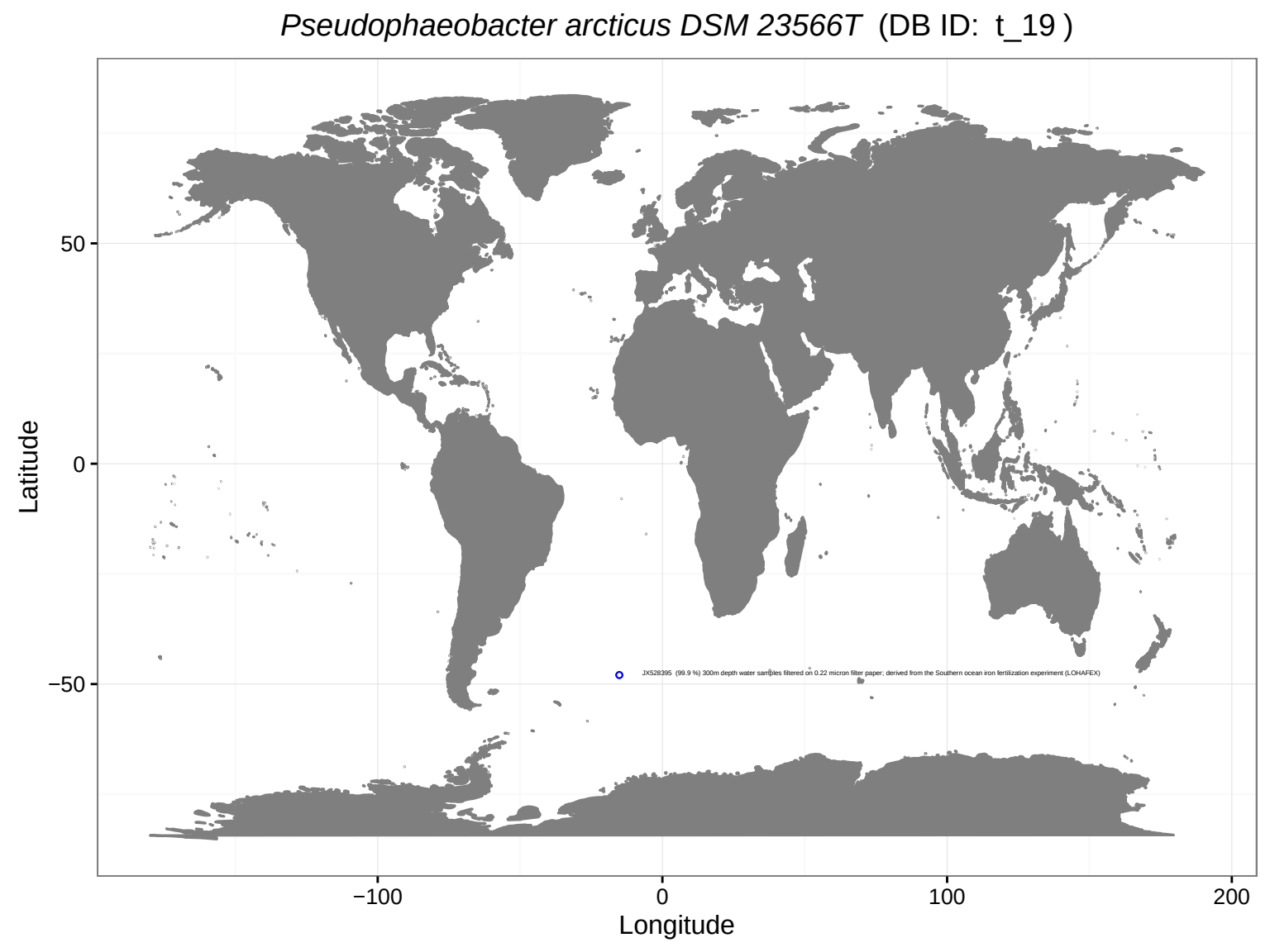

Paracoccus sp. J39 (DB ID: t_126)
LN561485 (100 %) refuse dump Nest 11 layer 2
50
0
−50
Latitude
−100
0
100
200
Longitude

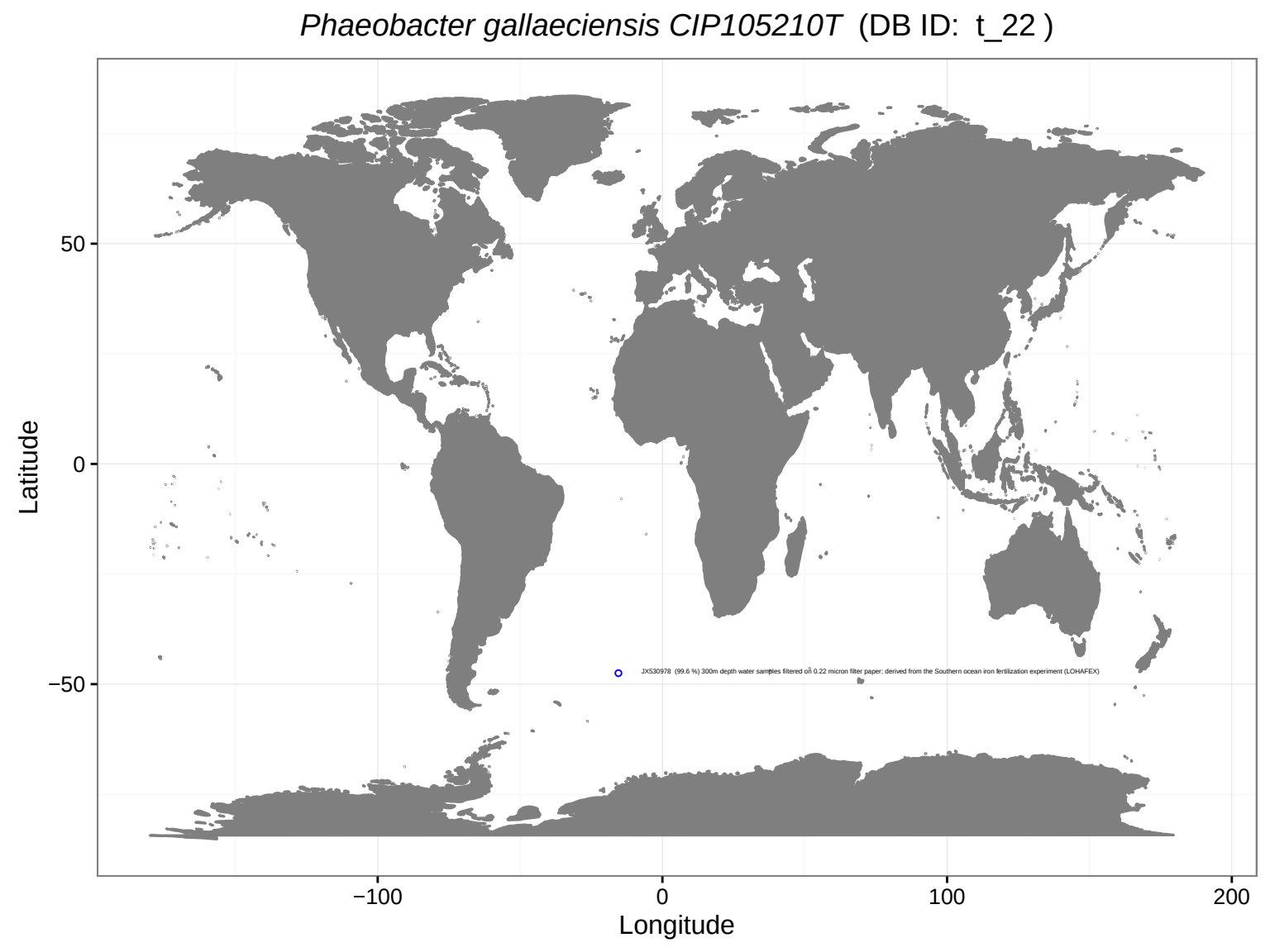

Phaeobacter sp. LSS9 (DB ID: t_128)
50
0
−50
Latitude
JX530978 (99.8 %) 300m depth water samples filtered oñ 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
−100
0
100
200
Longitude

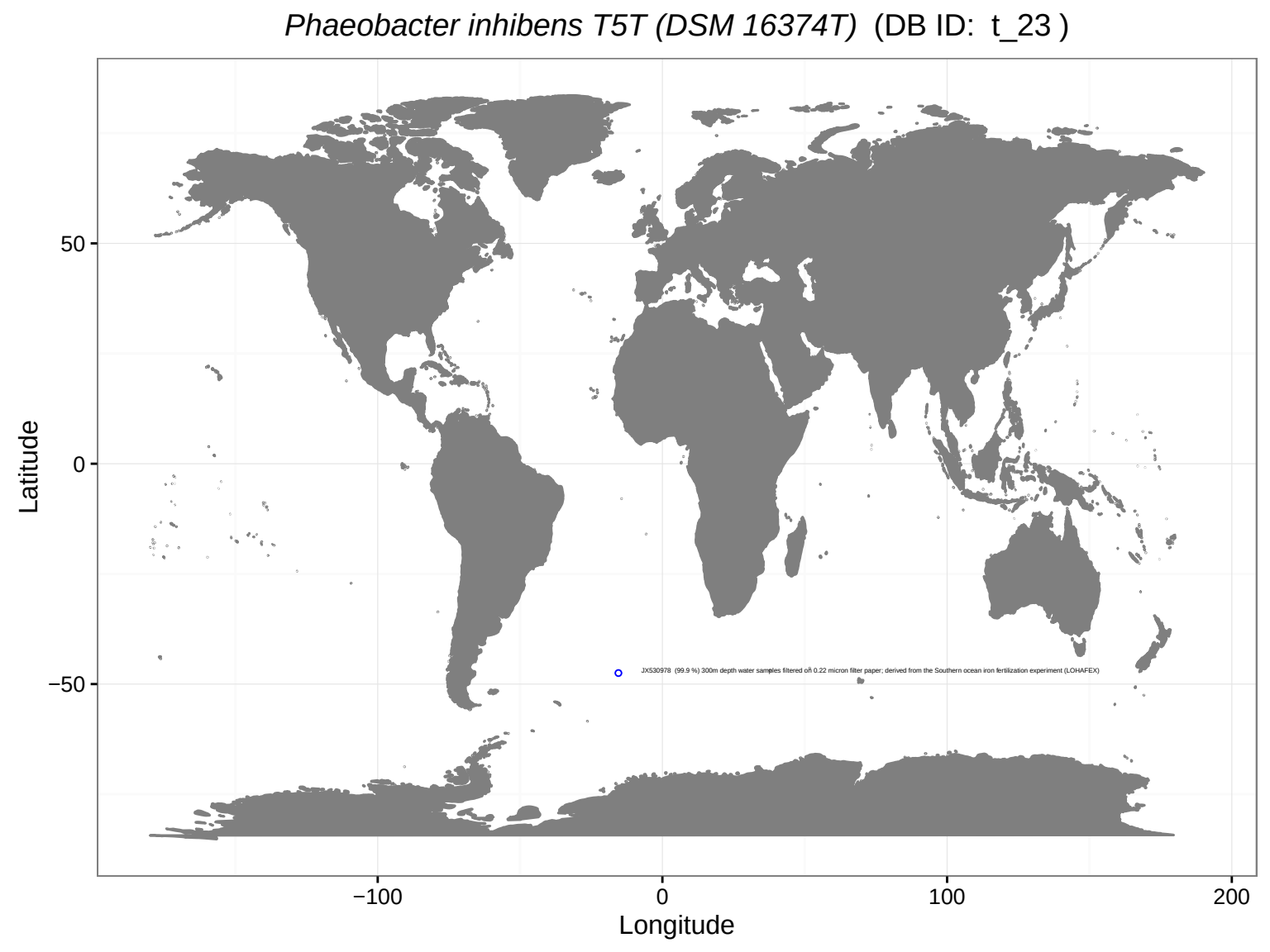

Planktomarina temperata RCA23T (DB ID: t_131)
EU819141 (99.9 %) coastal marine water in Norwegian Fjord
HQ242114 (99.7 %) Line P Station P4 depth of 10m
EU798998 (99.8 %) Newport Harbour, RI
EU800112 (99.7 %) Delaware Bay, NJ
DQ009296 (99.8 %) glass fiber prefiltered bacterioplankton DNA collected at 5M depth from the San Pedro Ocean Time Series
50
0
−50
Latitude
−100
0
100
200
Longitude

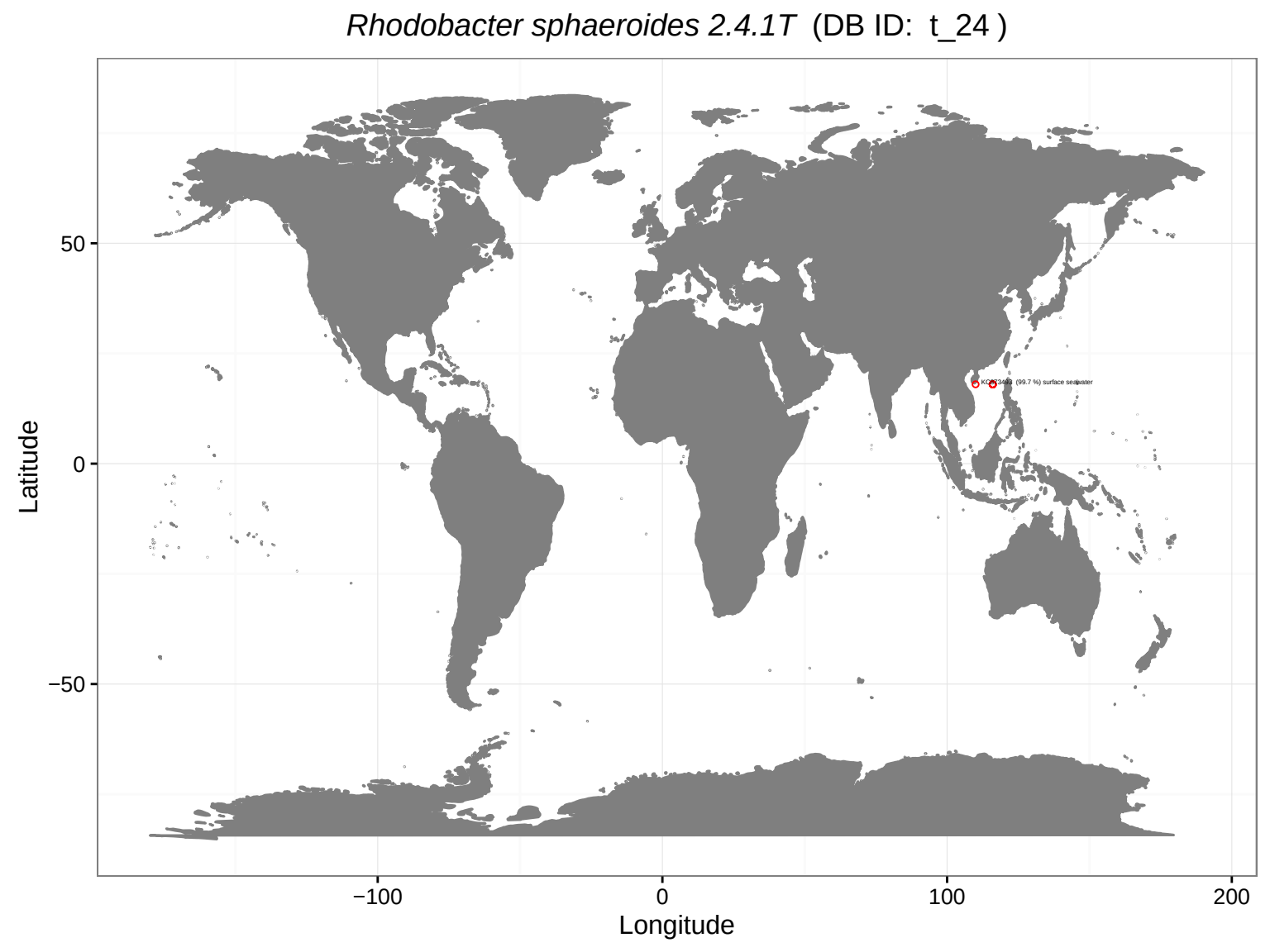

Octadecabacter antarcticus 307T (DB ID: t_17)
HQ220147 (99.5 %) snow
50
0
−50
Latitude
−100
0
100
200
Longitude

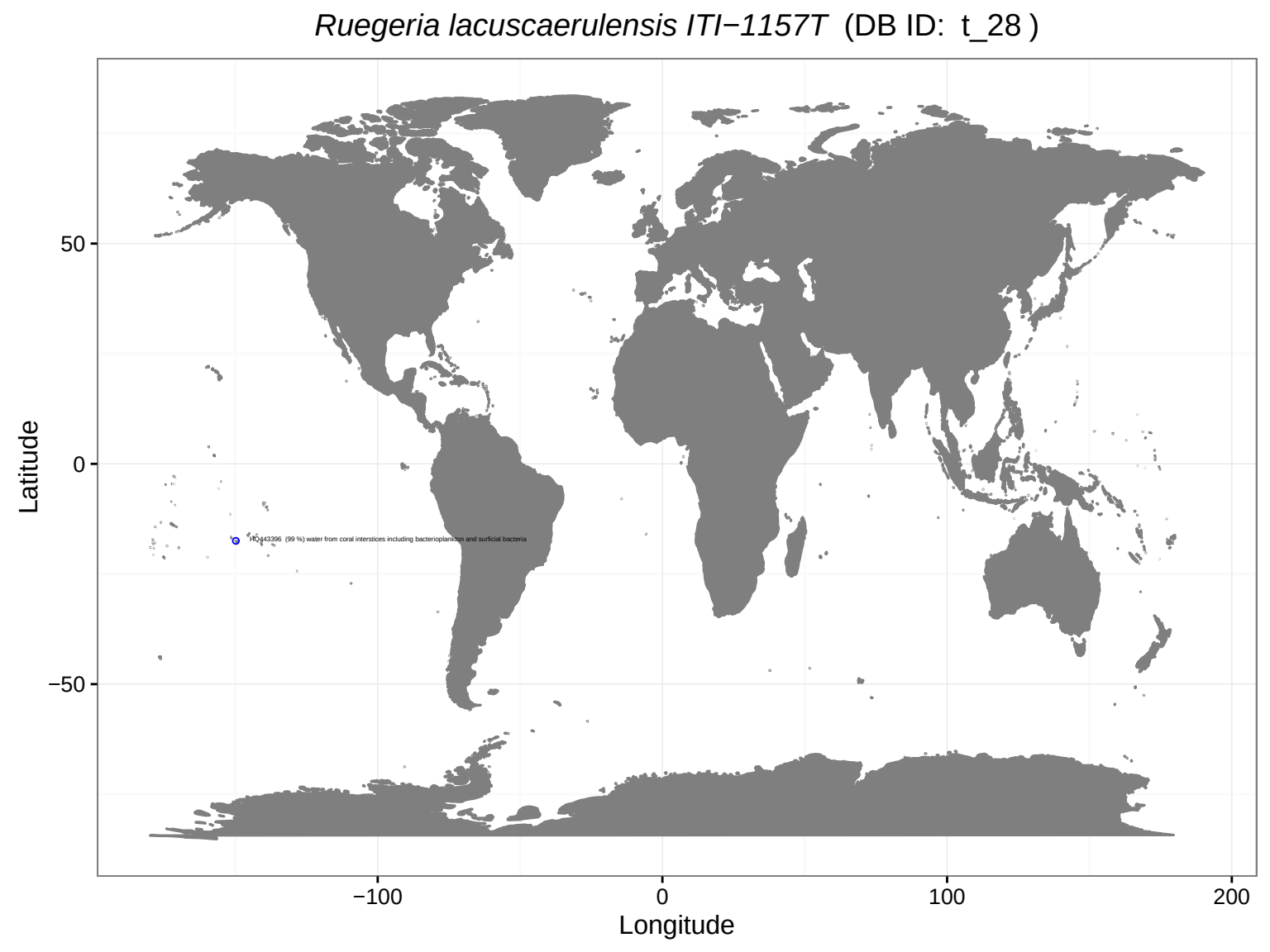

Octadecabacter arcticus 238T (DB ID: t_18)
HQ220147 (100 %) snow
50
0
−50
Latitude
−100
0
100
200
Longitude

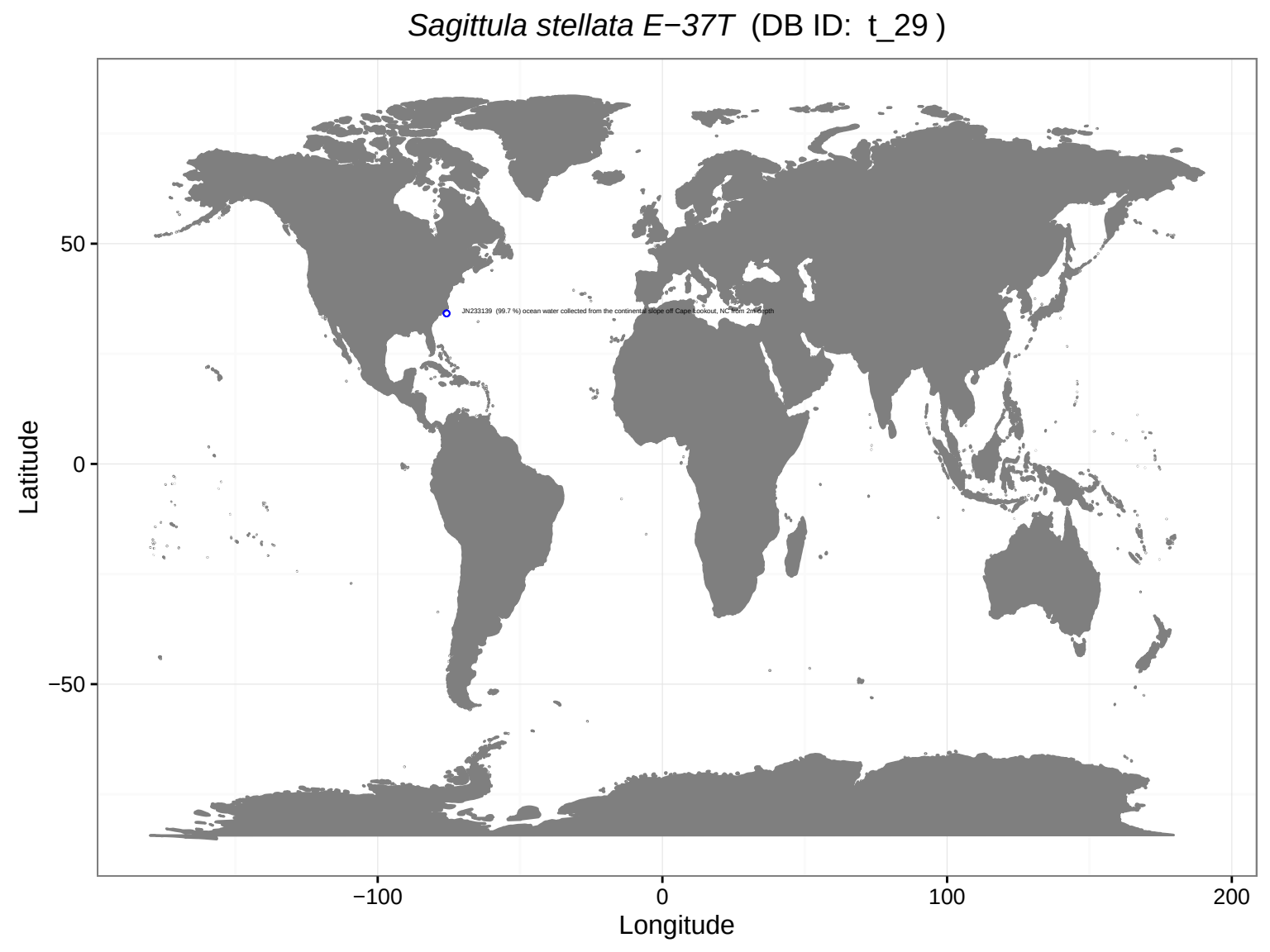

Pseudophaeobacter arcticus DSM 23566T (DB ID: t_19)
50
0
−50
Latitude
JX528395 (99.9 %) 300m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
−100
0
100
200
Longitude

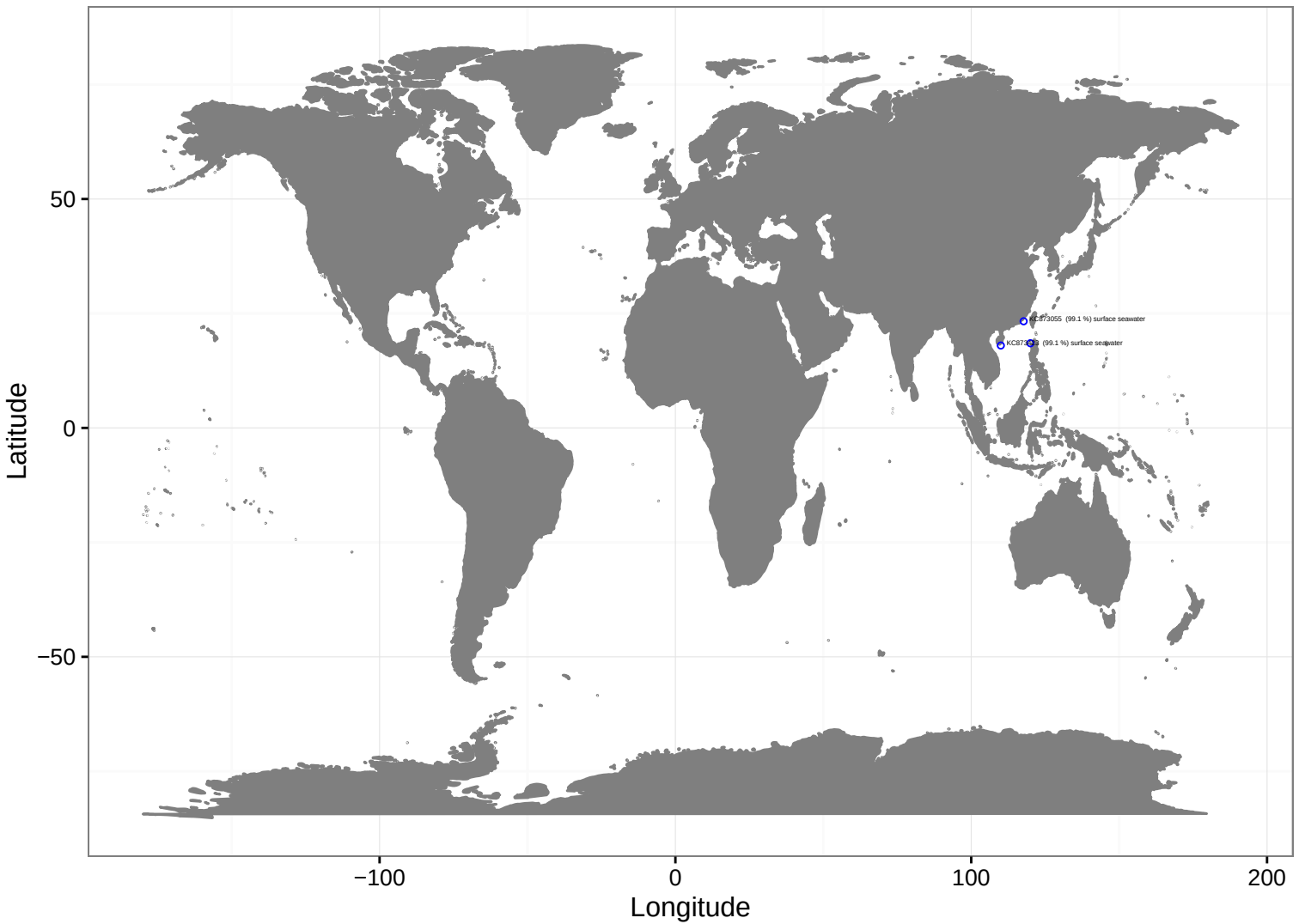

Phaeobacter gallaeciensis CIP105210T (DB ID: t_22)
50
0
−50
Latitude
JX530978 (99.6 %) 300m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
−100
0
100
200
Longitude


Phaeobacter inhibens T5T (DSM 16374T) (DB ID: t_23)
50
0
−50
Latitude
JX530978 (99.9 %) 300m depth water samples filtered oñ 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
−100
0
100
200
Longitude

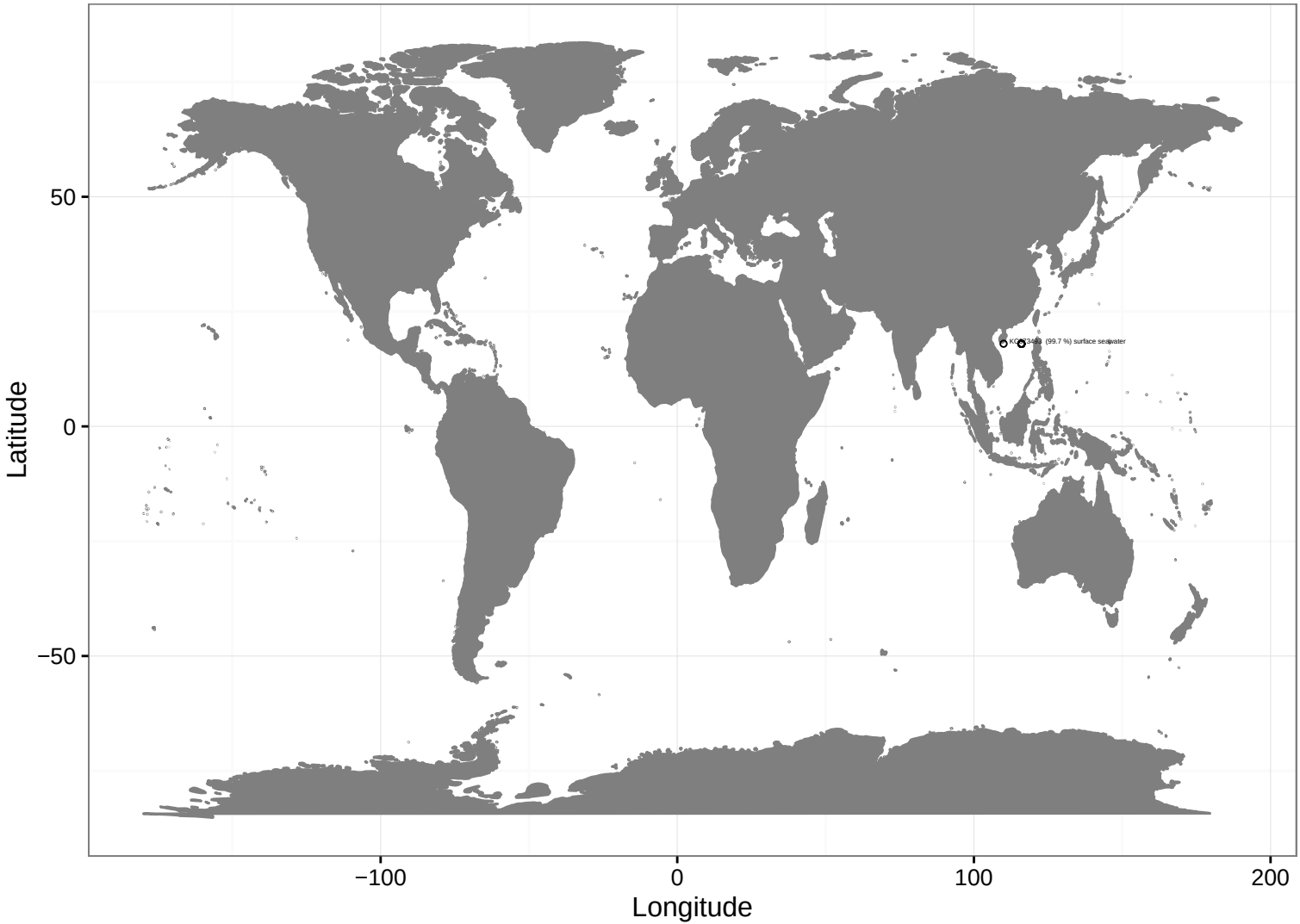

Rhodobacter sphaeroides 2.4.1T (DB ID: t_24)
50
0
−50
Latitude
−100
0
100
200
Longitude

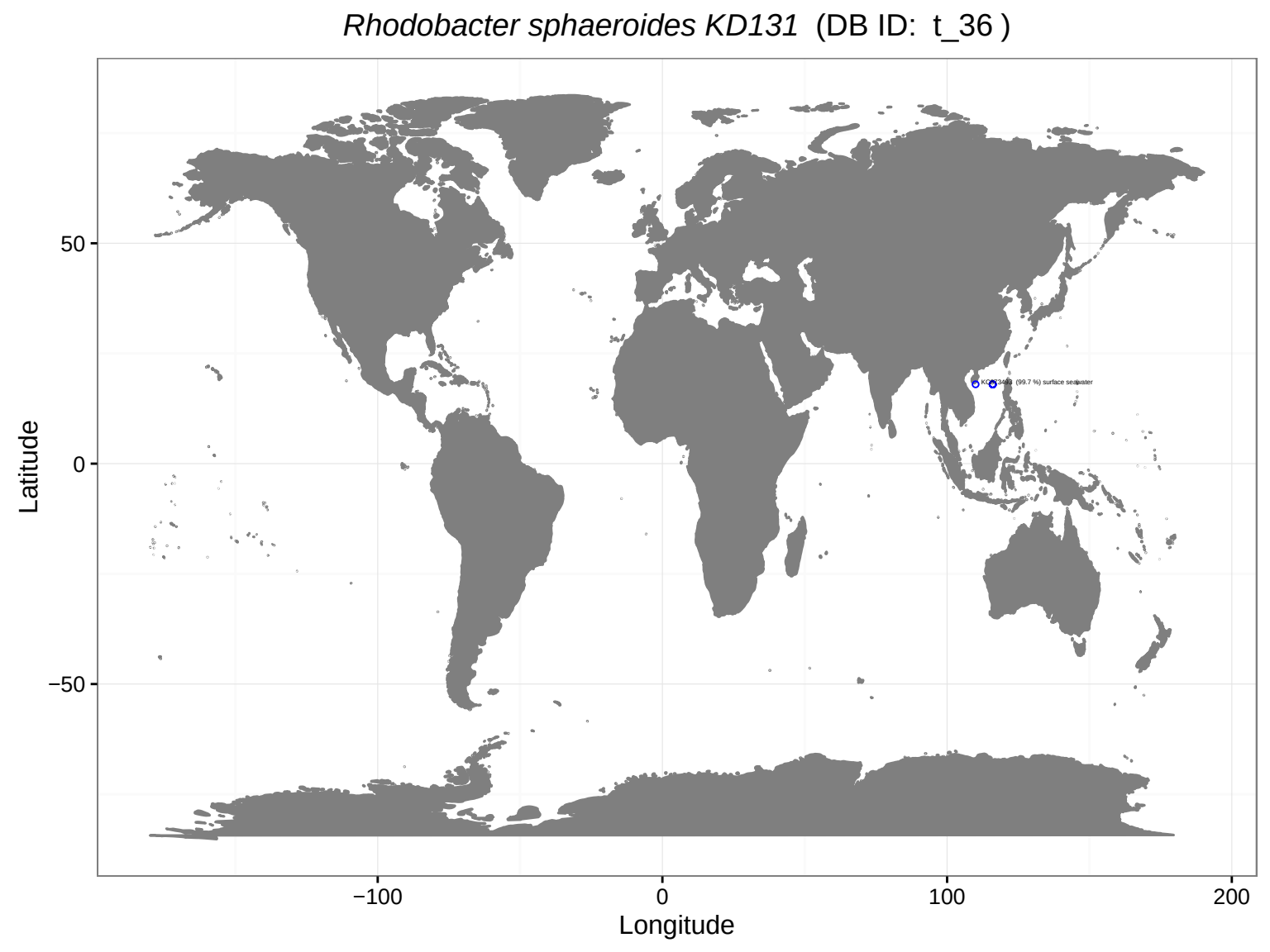

Ruegeria lacuscaerulensis ITI−1157T (DB ID: t_28)
HQ443396 (99 %) water from coral interstices including bacterioplankton and surficial bacteria
50
0
−50
Latitude
−100
0
100
200
Longitude

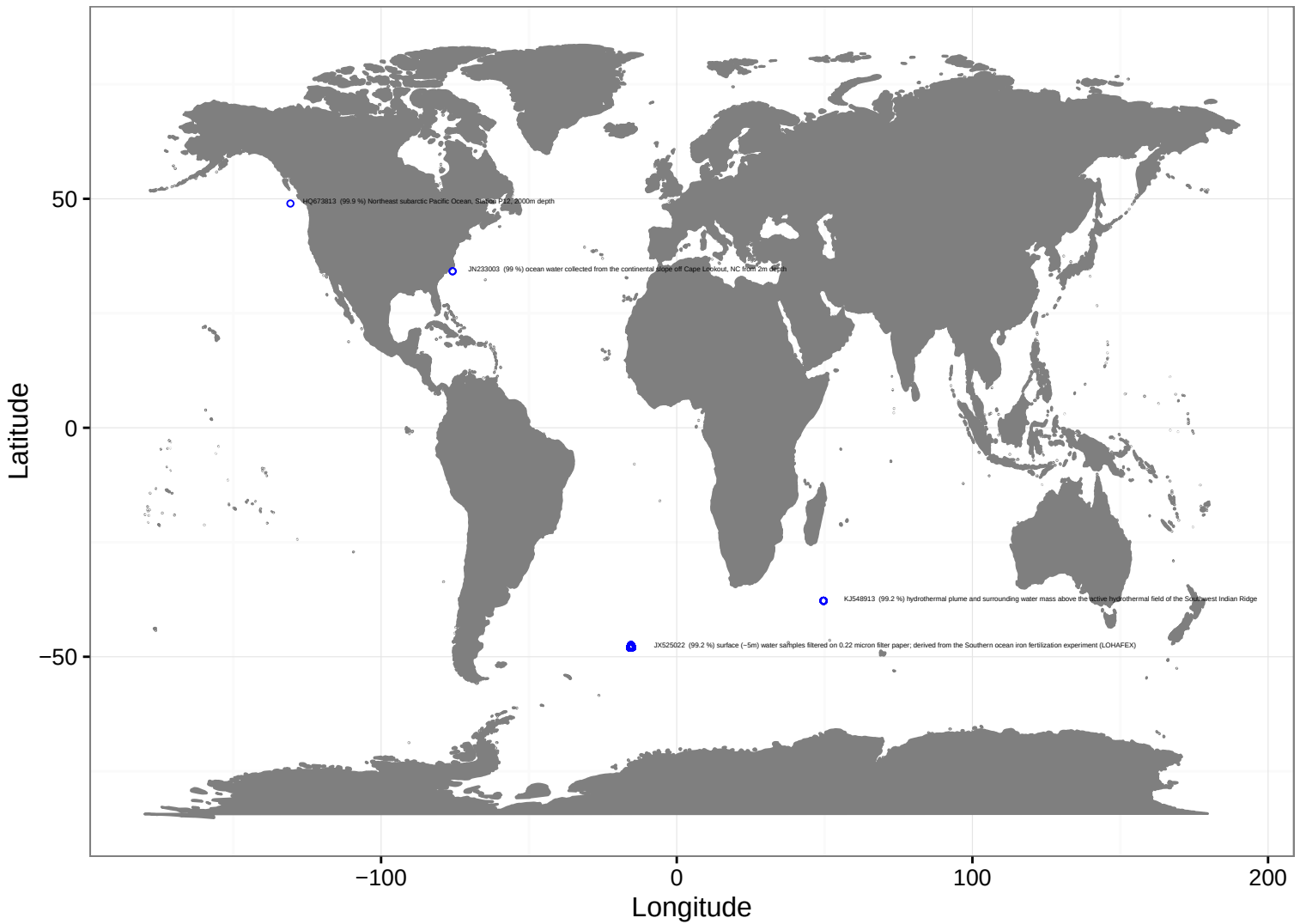

Sagittula stellata E−37T (DB ID: t_29)
JN233139 (99.7 %) ocean water collected from the continental slope off Cape Lookout, NC from 2m depth
50
0
−50
Latitude
−100
0
100
200
Longitude


Ruegeria sp. TM1040 (DB ID: t_32)
50
0
−50
Latitude
−100
0
100
200
Longitude
KC873055 (99.1 %) surface seawater
KC873058 (99.1 %) surface seawater

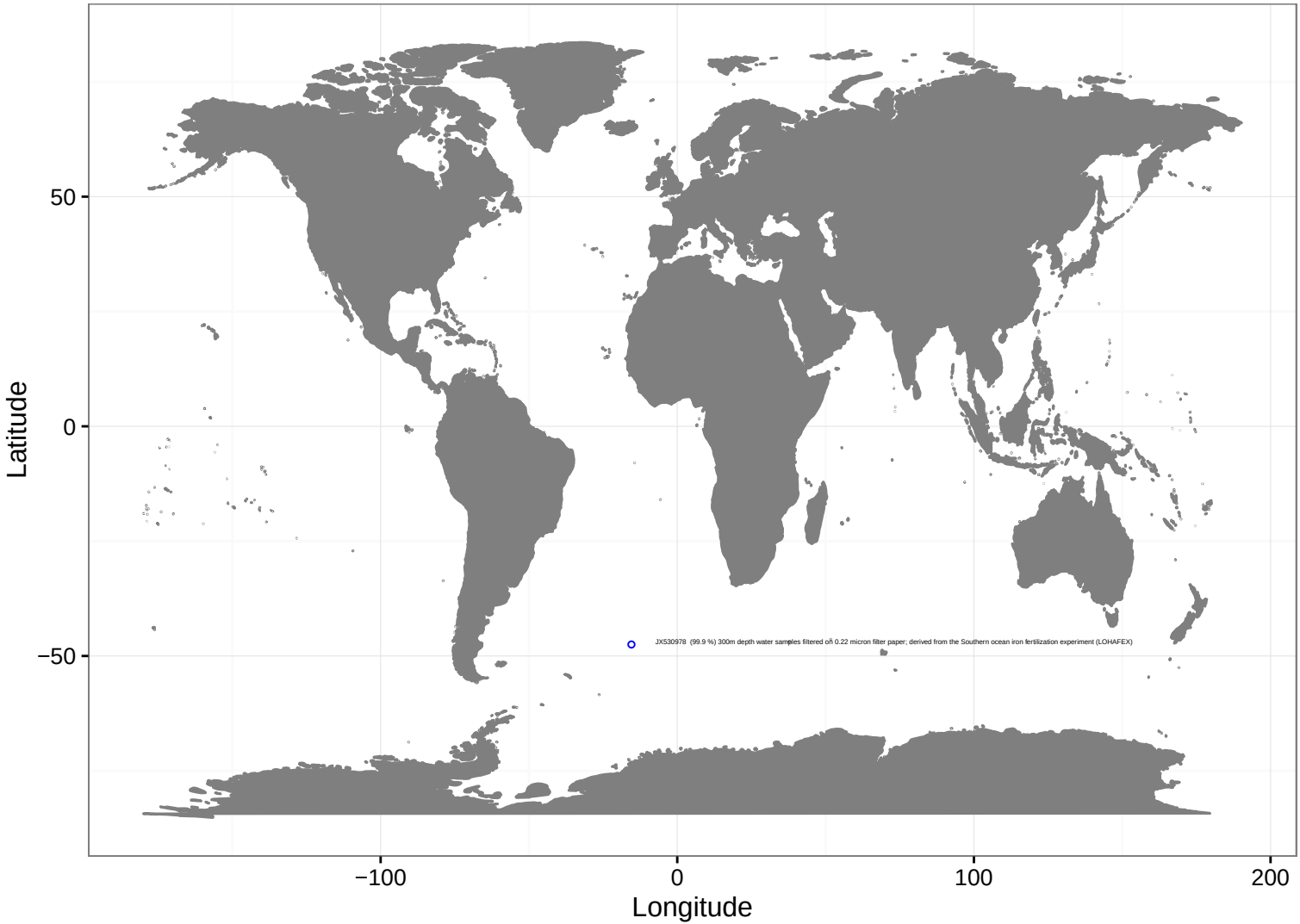
Rhodobacter sphaeroides ATCC 17029 (DB ID: t_34)

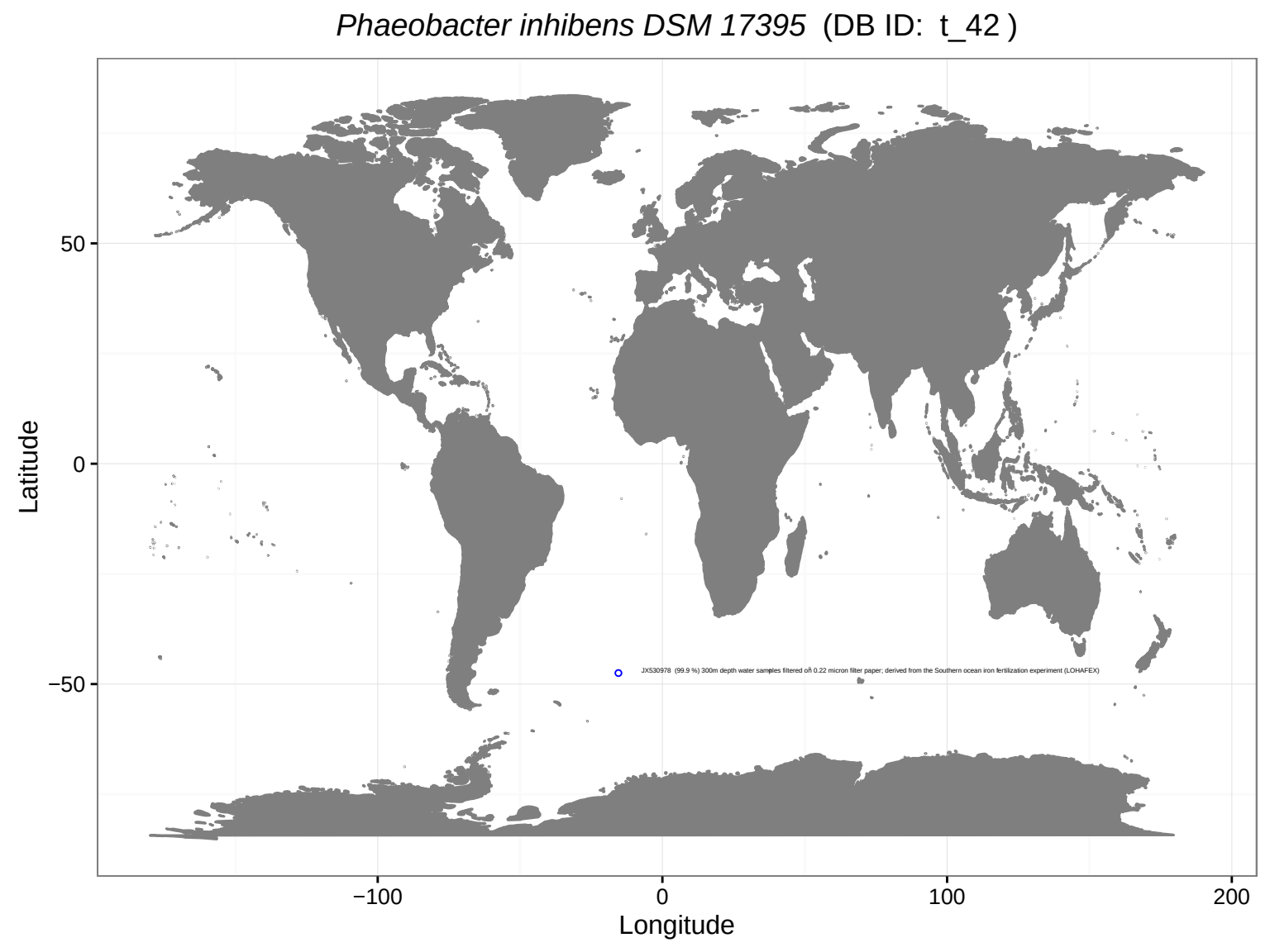


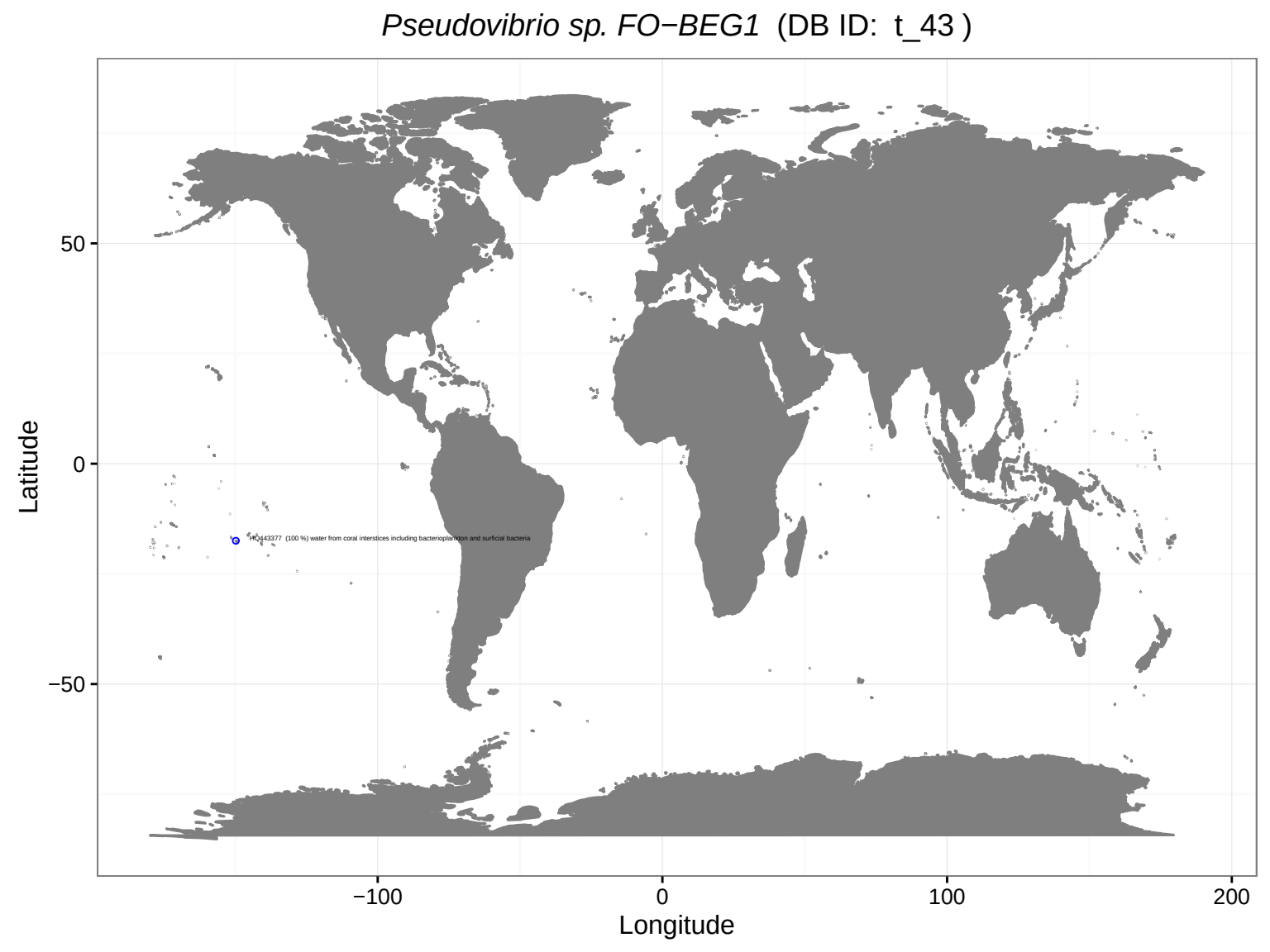

Rhodobacter sphaeroides KD131 (DB ID: t_36)
50
0
−50
Latitude
−100
0
100
200
Longitude

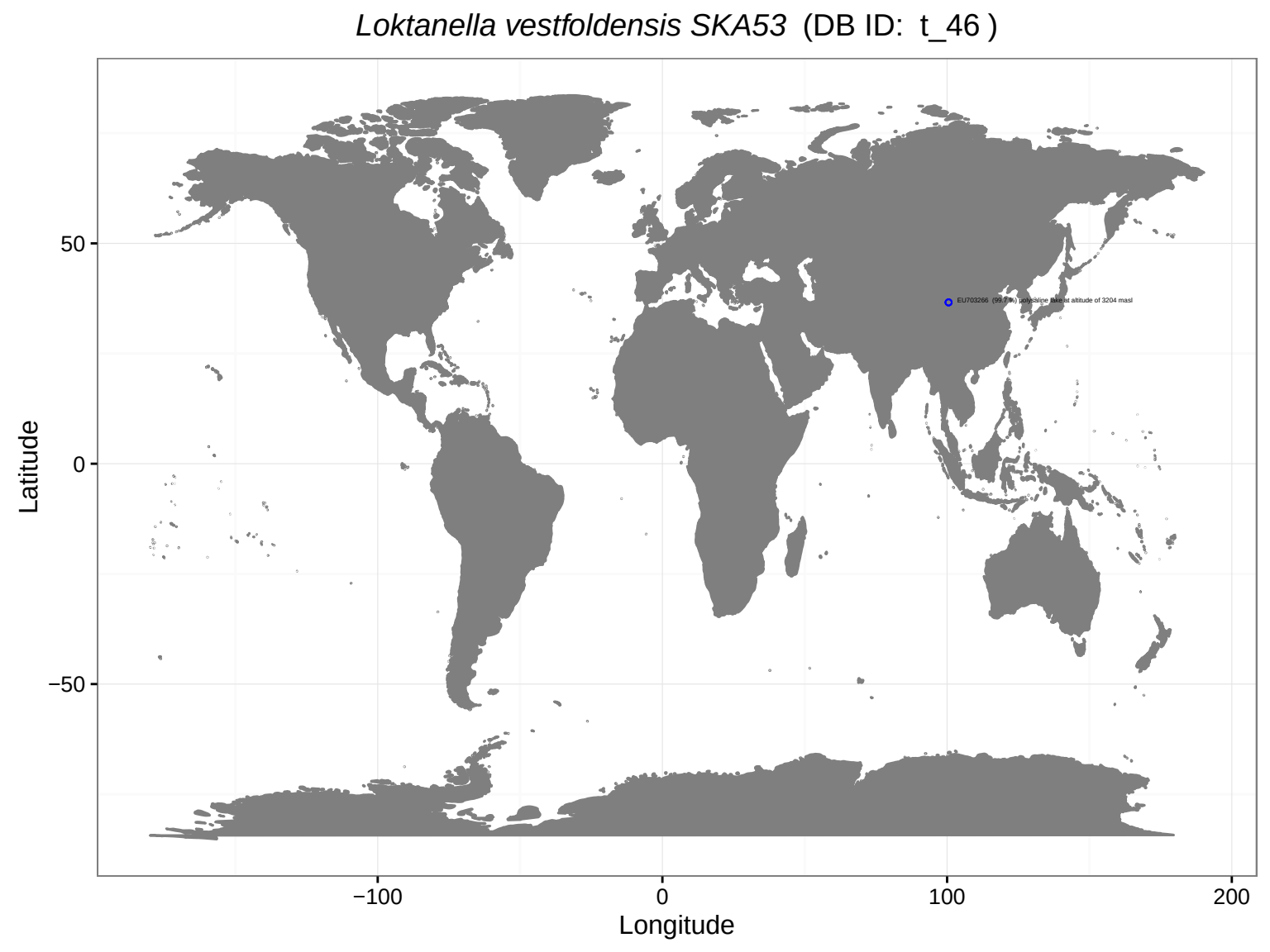

Oceanibulbus indolifex HEL−45T (DB ID: t_4)
HQ673813 (99.9 %) Northeast subarctic Pacific Ocean, Station P12, 2000m depth
JN233003 (99 %) ocean water collected from the continental slope off Cape Lookout, NC from 2m depth
KJ548913 (99.2 %) hydrothermal plume and surrounding water mass above the active hydrothermal field of the Southwest Indian Ridge
JX525022 (99.2 %) surface (~5m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
50
0
−50
Latitude
−100
0
100
200
Longitude

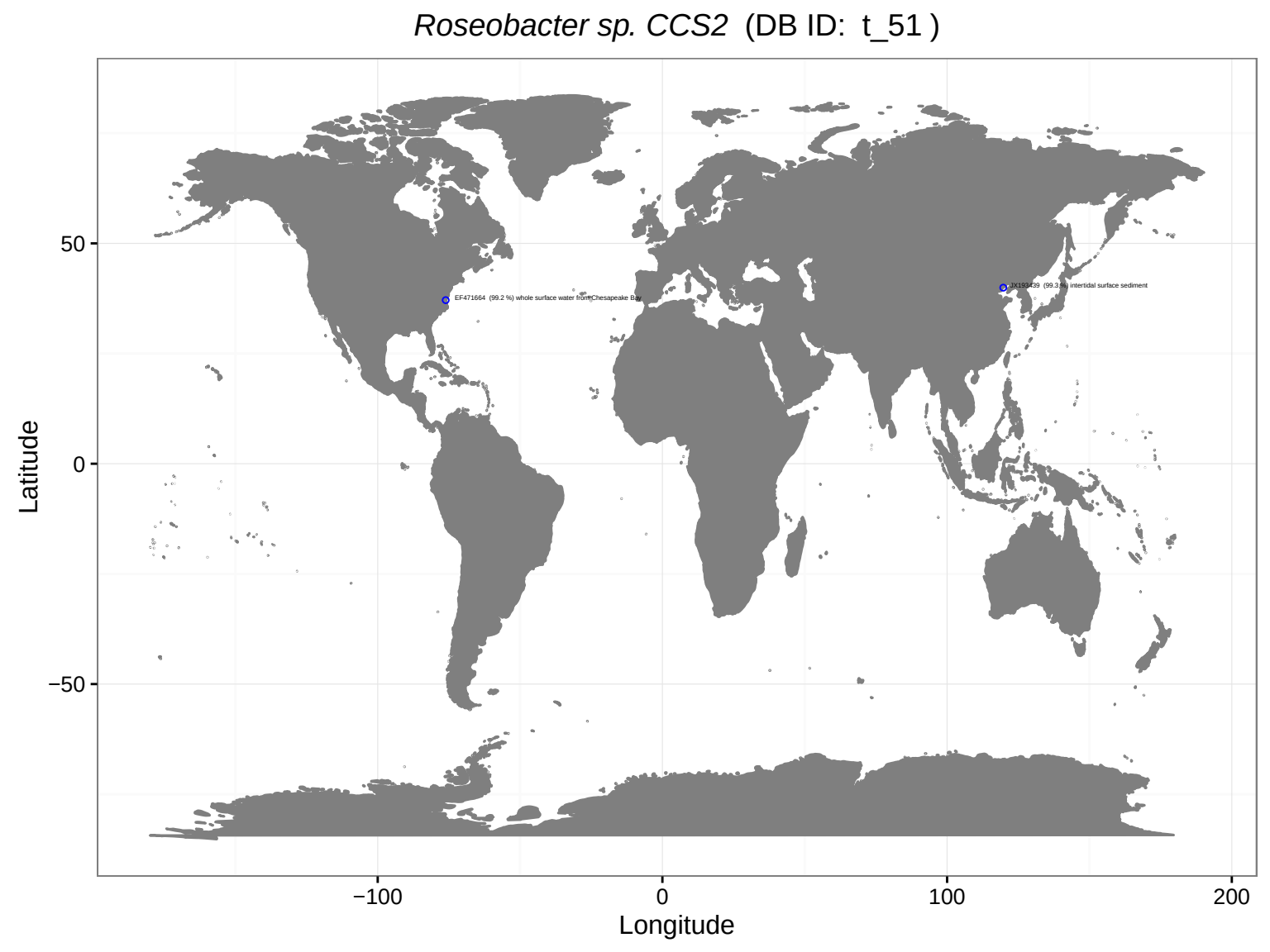
Phaeobacter inhibens 2.10 (DB ID: t_41)

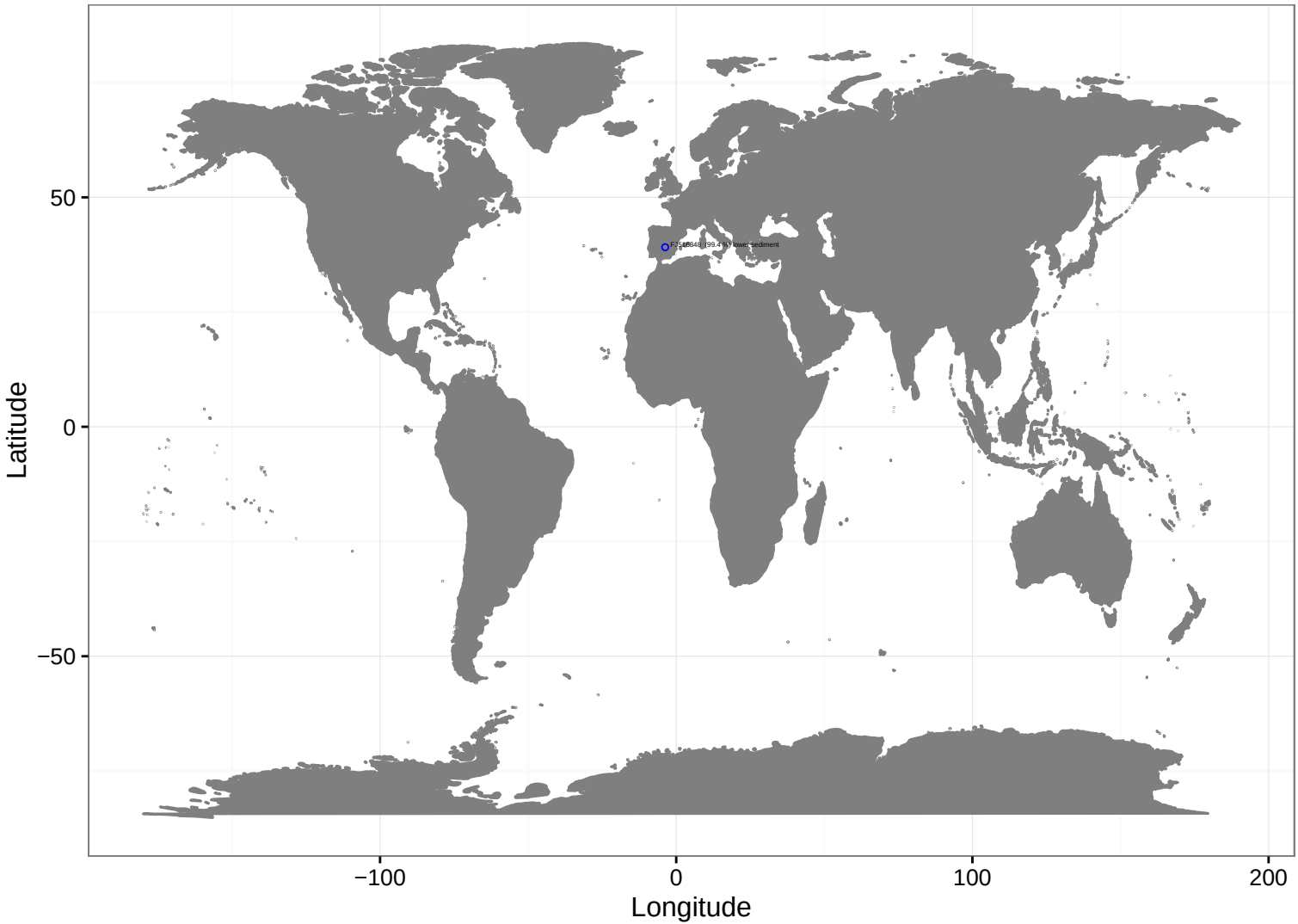


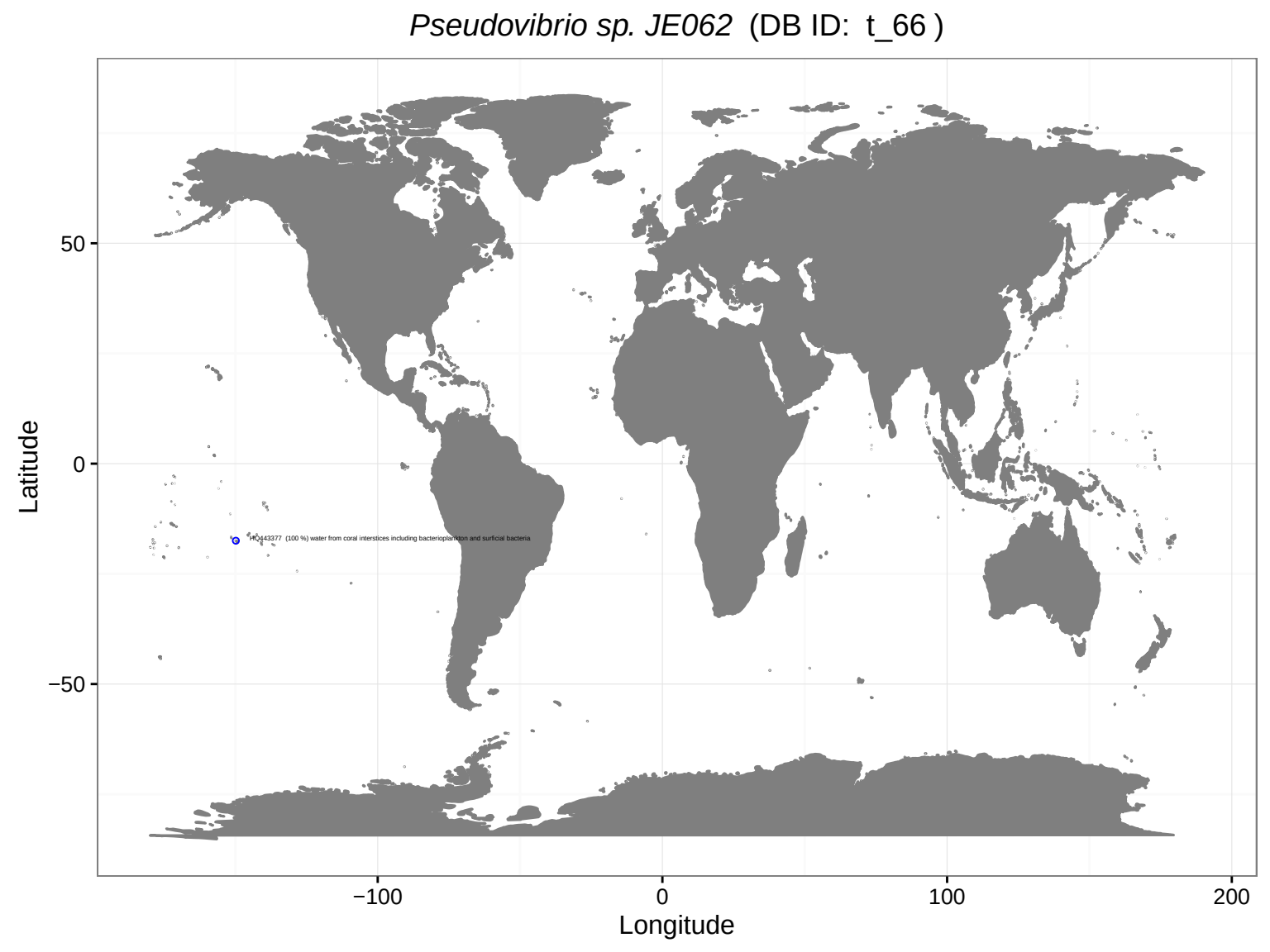

Phaeobacter inhibens DSM 17395 (DB ID: t_42)
50
0
−50
Latitude
JX530978 (99.9 %) 300m depth water samples filtered oñ 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
−100
0
100
200
Longitude


Pseudovibrio sp. FO−BEG1 (DB ID: t_43)
50
0
−50
Latitude
HQ443377 (100 %) water from coral interstices including bacterioplankton and surficial bacteria
−100
0
100
200
Longitude

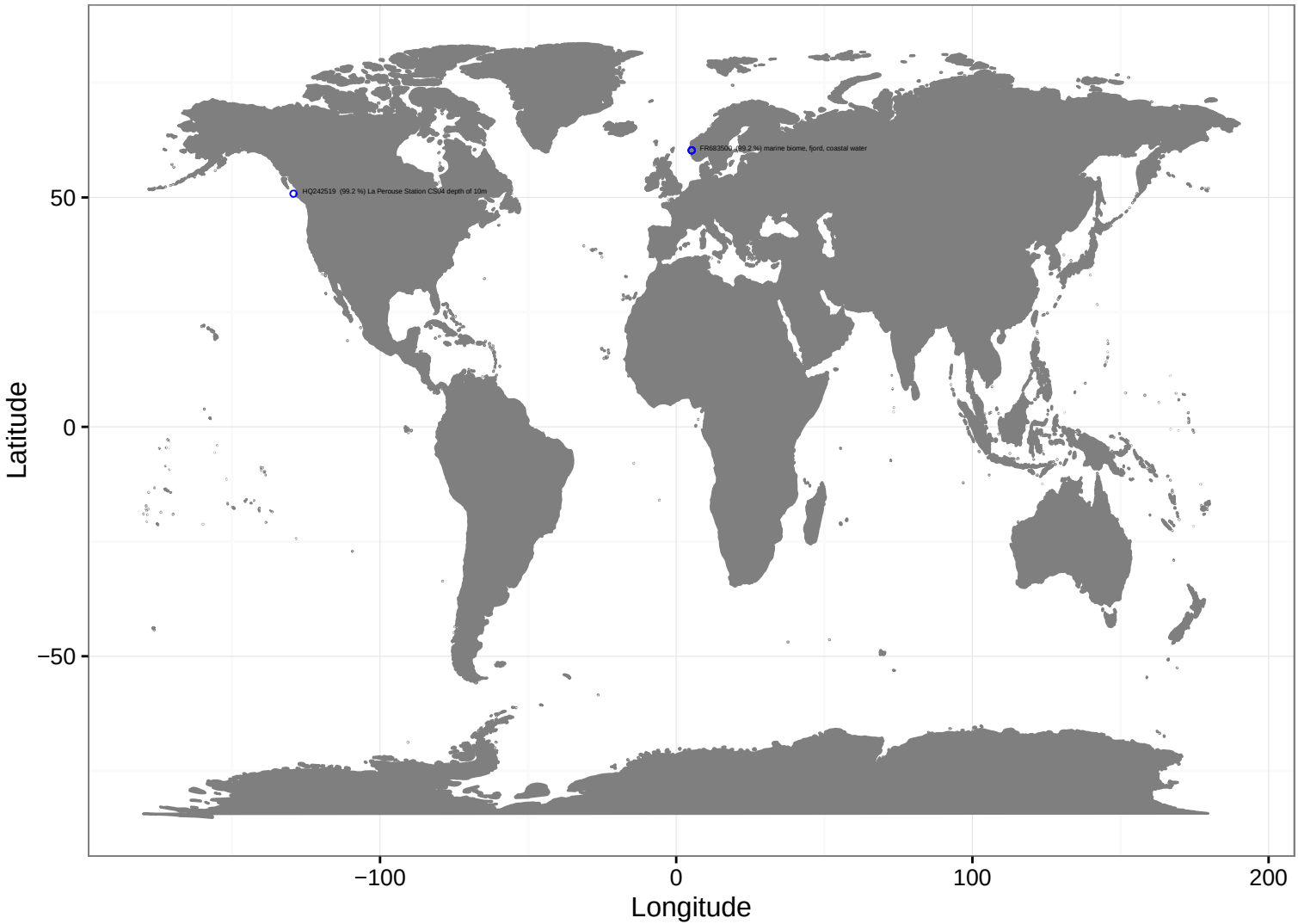

Loktanella vestfoldensis SKA53 (DB ID: t_46)
EU703266 (99.7%) | alkaline lake at altitude of 3204 masl
50
0
−50
Latitude
−100
0
100
200
Longitude

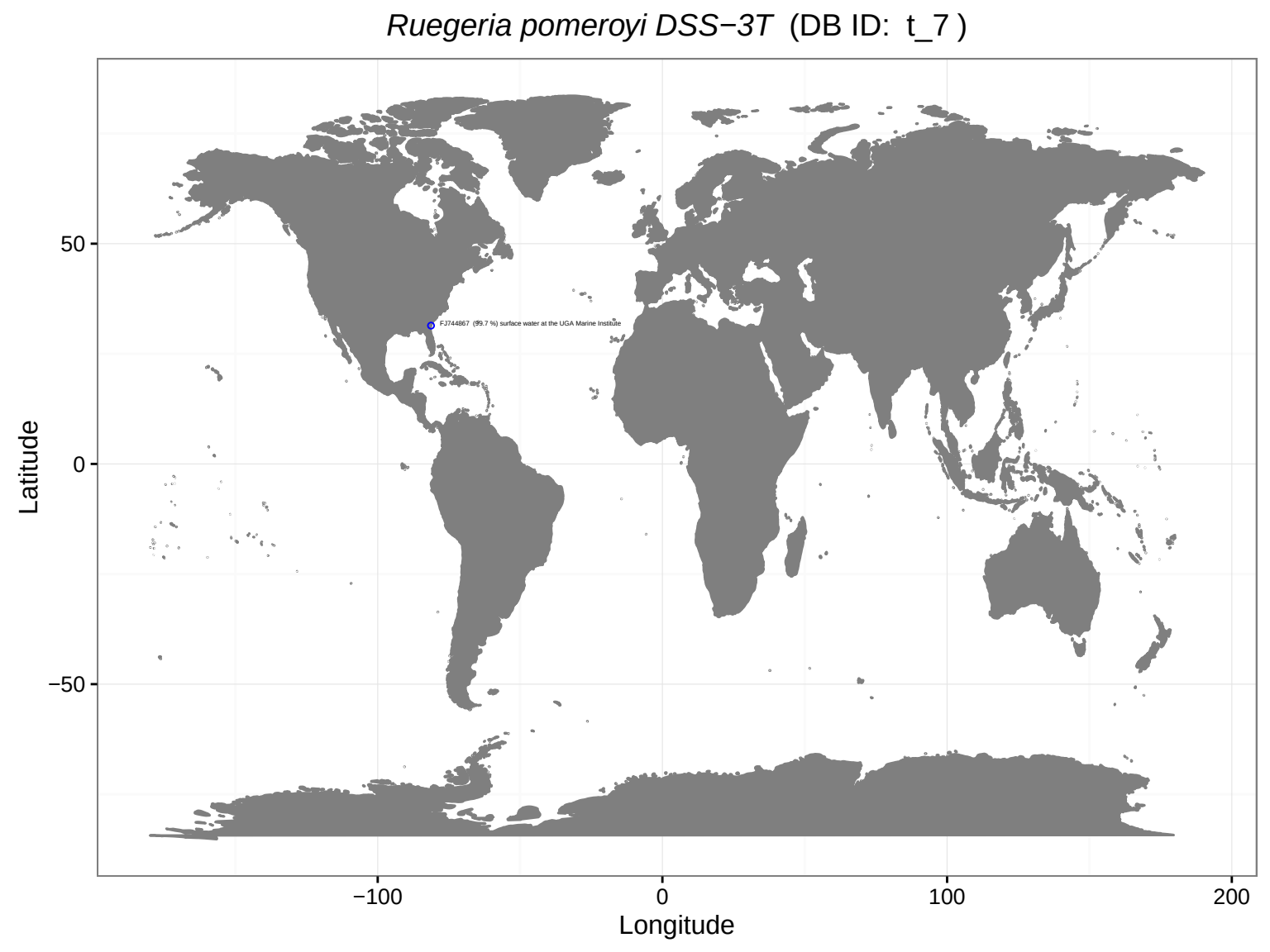

Roseobacter sp. CCS2 (DB ID: t_51)
EF471664 (99.2 %) whole surface water from Chesapeake Bay
JX193439 (99.3 %) intertidal surface sediment
50
0
−50
Latitude
−100
0
100
200
Longitude

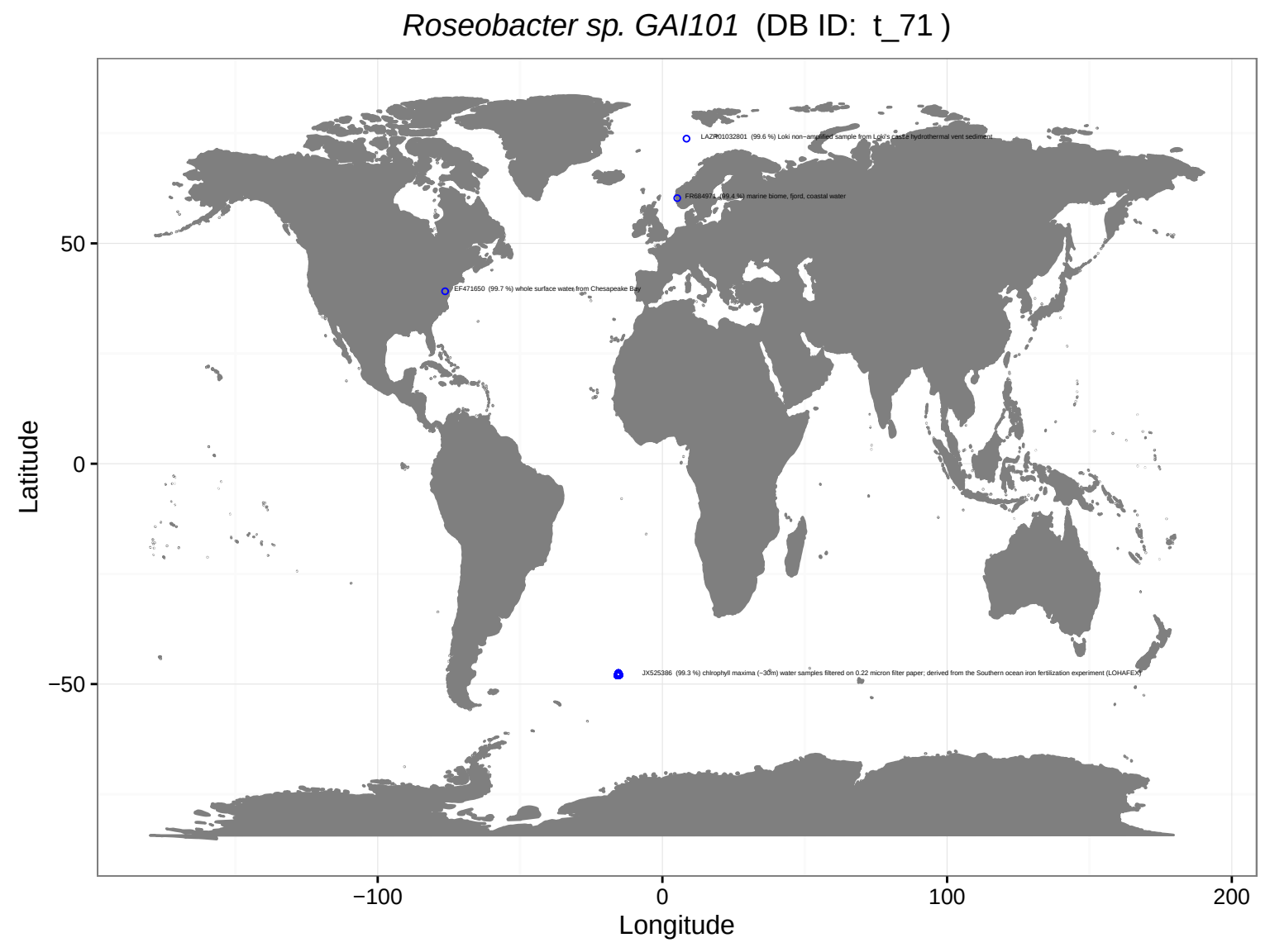

Roseovarius sp. TM1035 (DB ID: t_53)
50
0
−50
Latitude
−100
0
100
200
Longitude

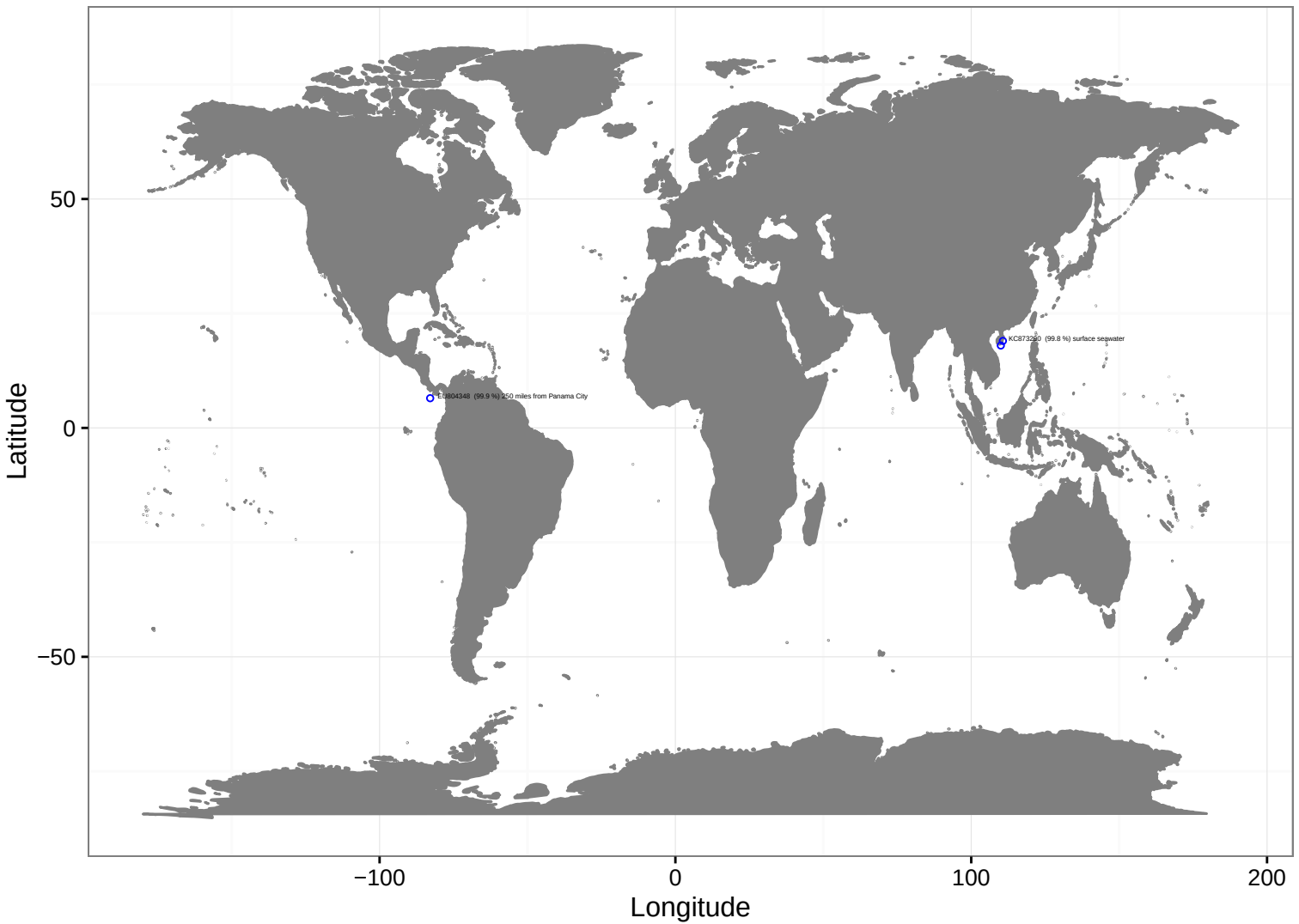

Pseudovibrio sp. JE062 (DB ID: t_66)
50
0
−50
Latitude
HQ443377 (100 %) water from coral interstices including bacterioplankton and surficial bacteria
−100
0
100
200
Longitude

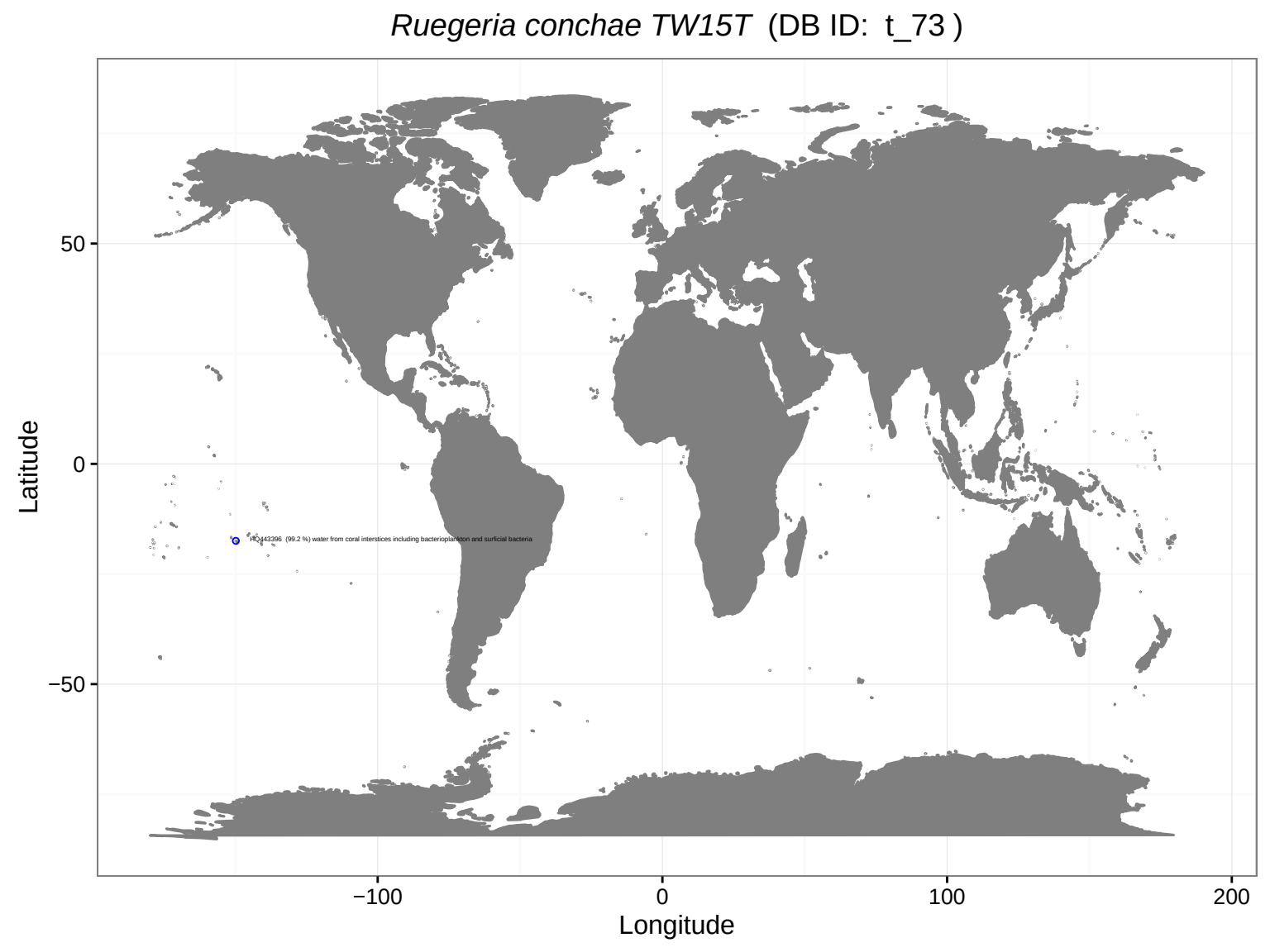
Rhodobacteraceae bacterium HTCC2083 (DB ID: t_67)

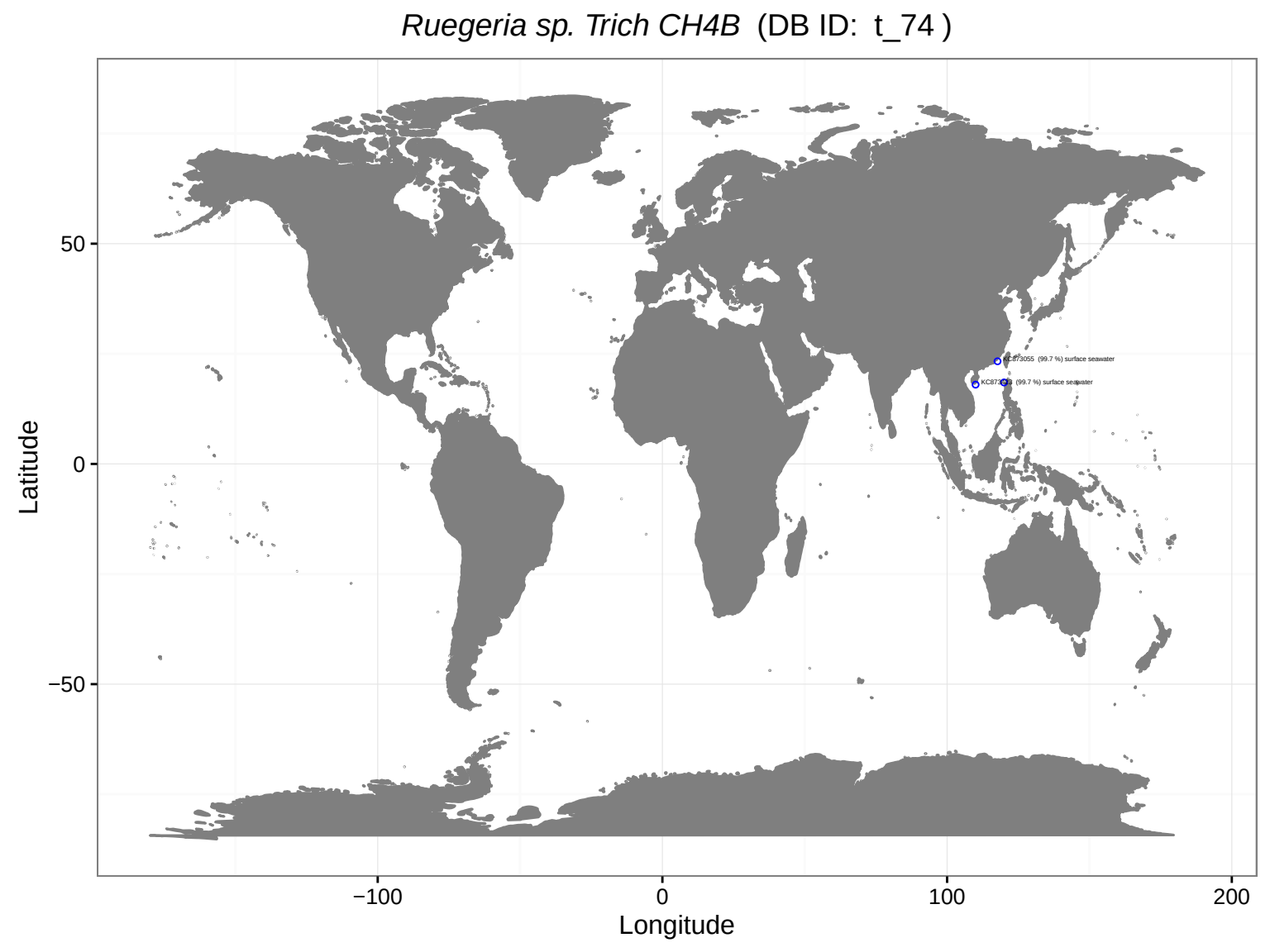



Ruegeria pomeroyi DSS−3T (DB ID: t_7)
FJ744867 (99.7 %) surface water at the UGA Marine Institute
50
0
−50
Latitude
−100
0
100
200
Longitude

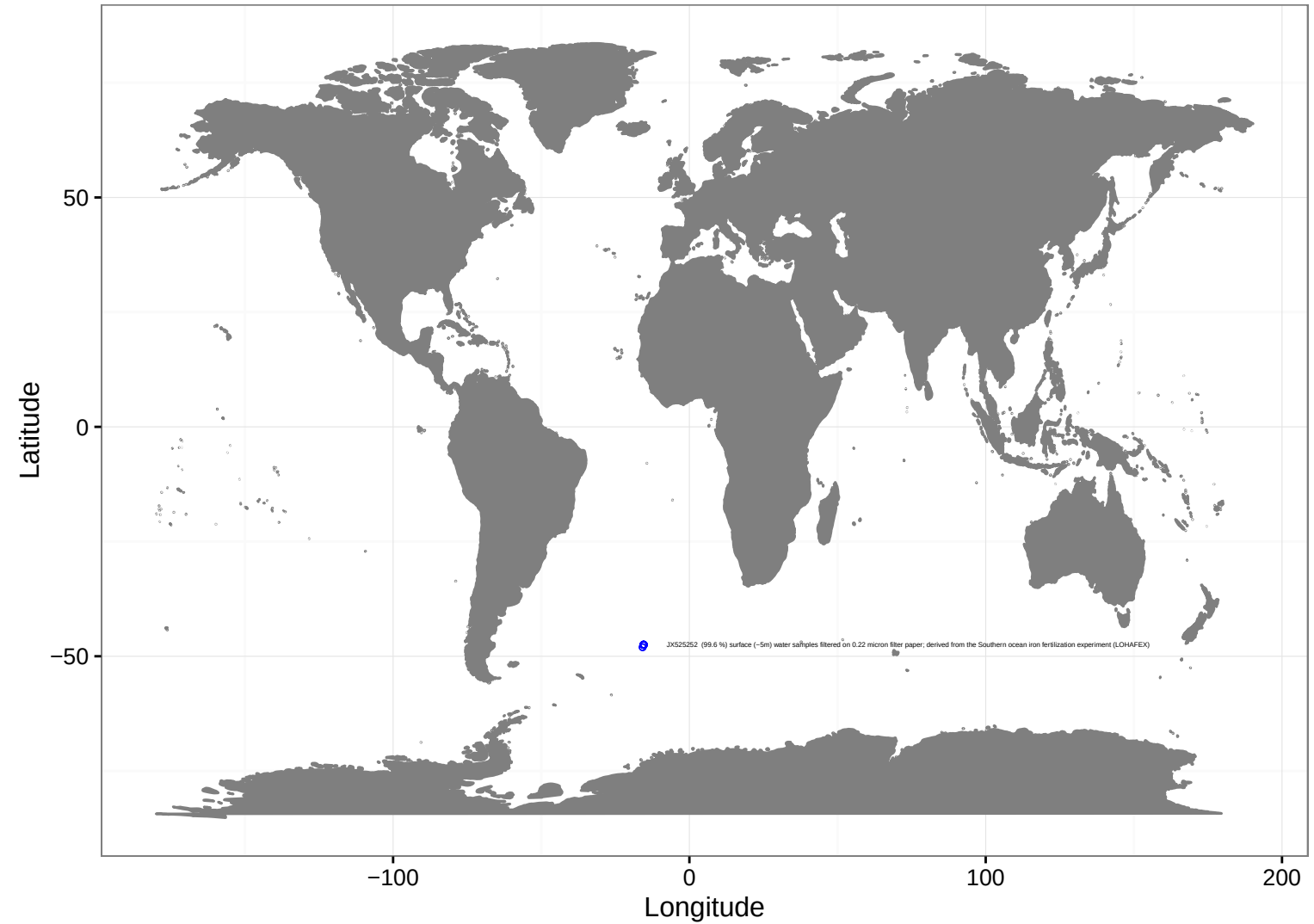

Roseobacter sp. GAI101 (DB ID: t_71)
LAZR01032801 (99.6 %) Loki non−amended sample from Loki's castle hydrothermal vent sediment
FR684974 (99.4 %) marine biome, fjord, coastal water
EF471650 (99.7 %) whole surface water from Chesapeake Bay
JX525386 (99.3 %) chlrophyll maxima (~30m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
Latitude
50
0
−50
−100
0
100
200
Longitude


Ruegeria sp. R11 (DB ID: t_72)
EU804348 (99.9 %) 250 miles from Panama City
KC873280 (99.8 %) surface seawater
50
0
−50
Latitude
−100
0
100
200
Longitude

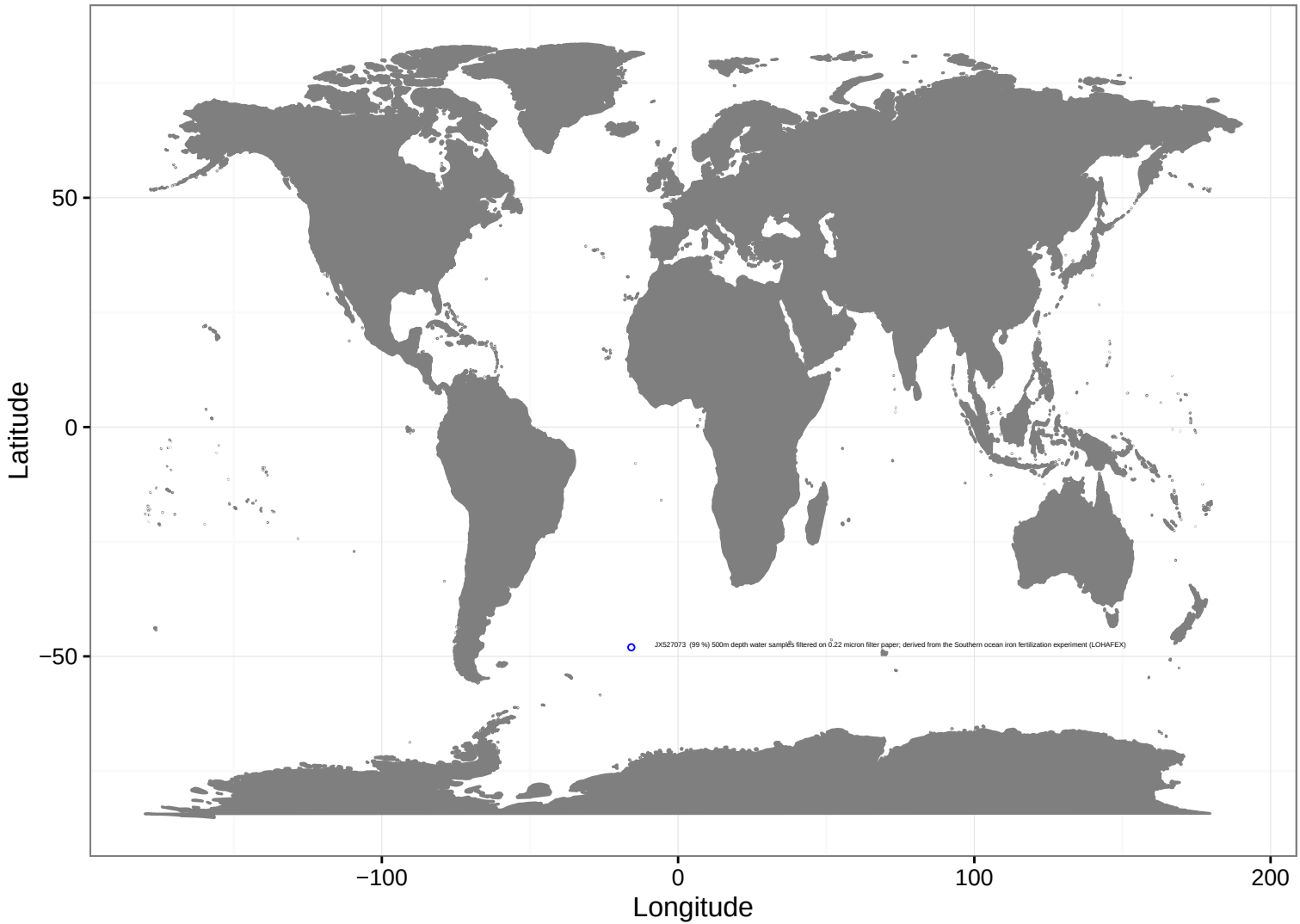

Ruegeria conchae TW15T (DB ID: t_73)
HQ443396 (99.2 %) water from coral interstices including bacterioplankton and surficial bacteria
50
0
−50
Latitude
−100
0
100
200
Longitude

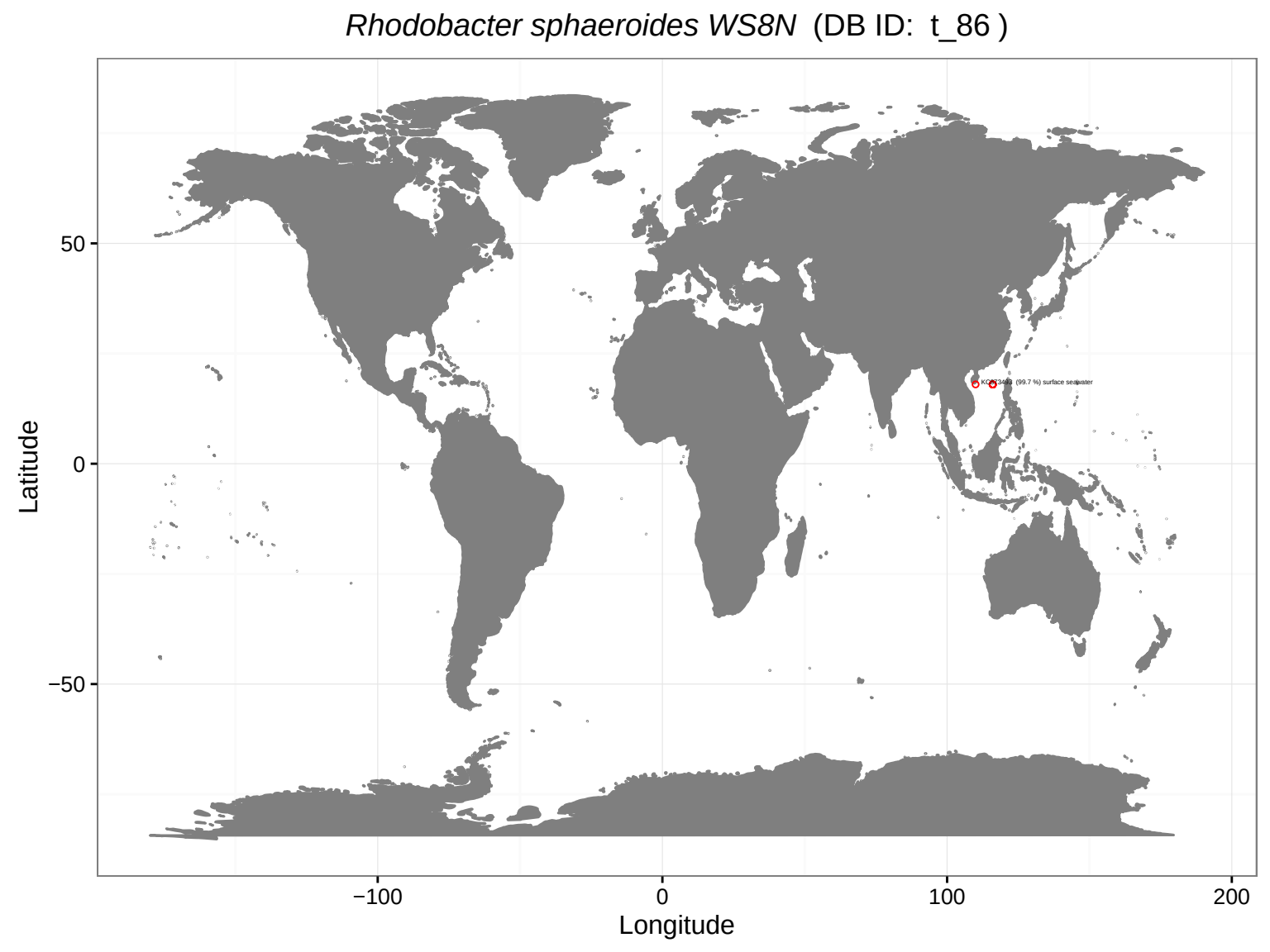

Ruegeria sp. Trich CH4B (DB ID: t_74)
50
0
−50
Latitude
−100
0
100
200
Longitude
KC873055 (99.7 %) surface seawater
KC873058 (99.7 %) surface seawater

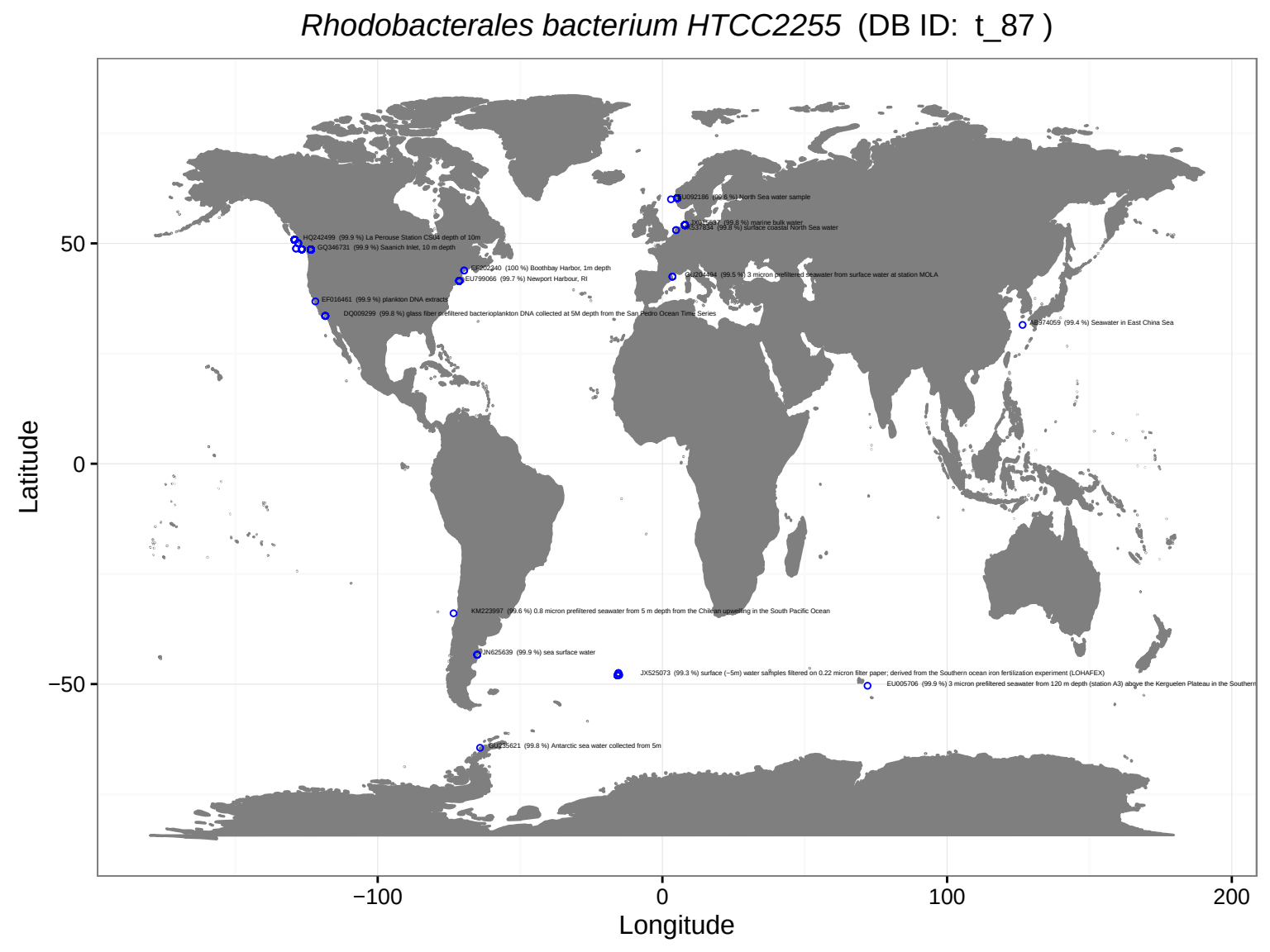
Ahrensia kielensis DSM 5890T (DB ID: t_76)



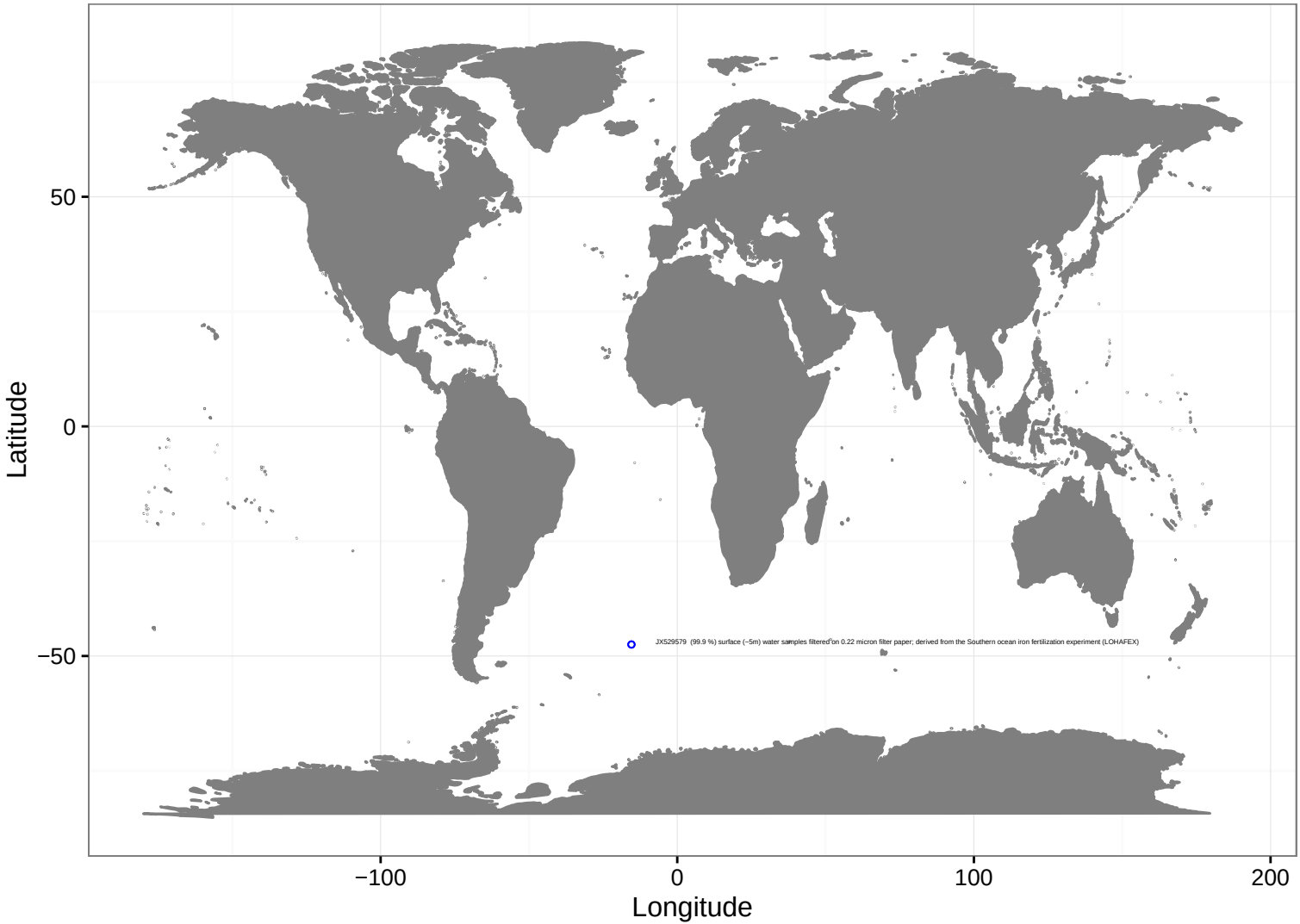
Loktanella hongkongensis DSM 17492T (DB ID: t_81)

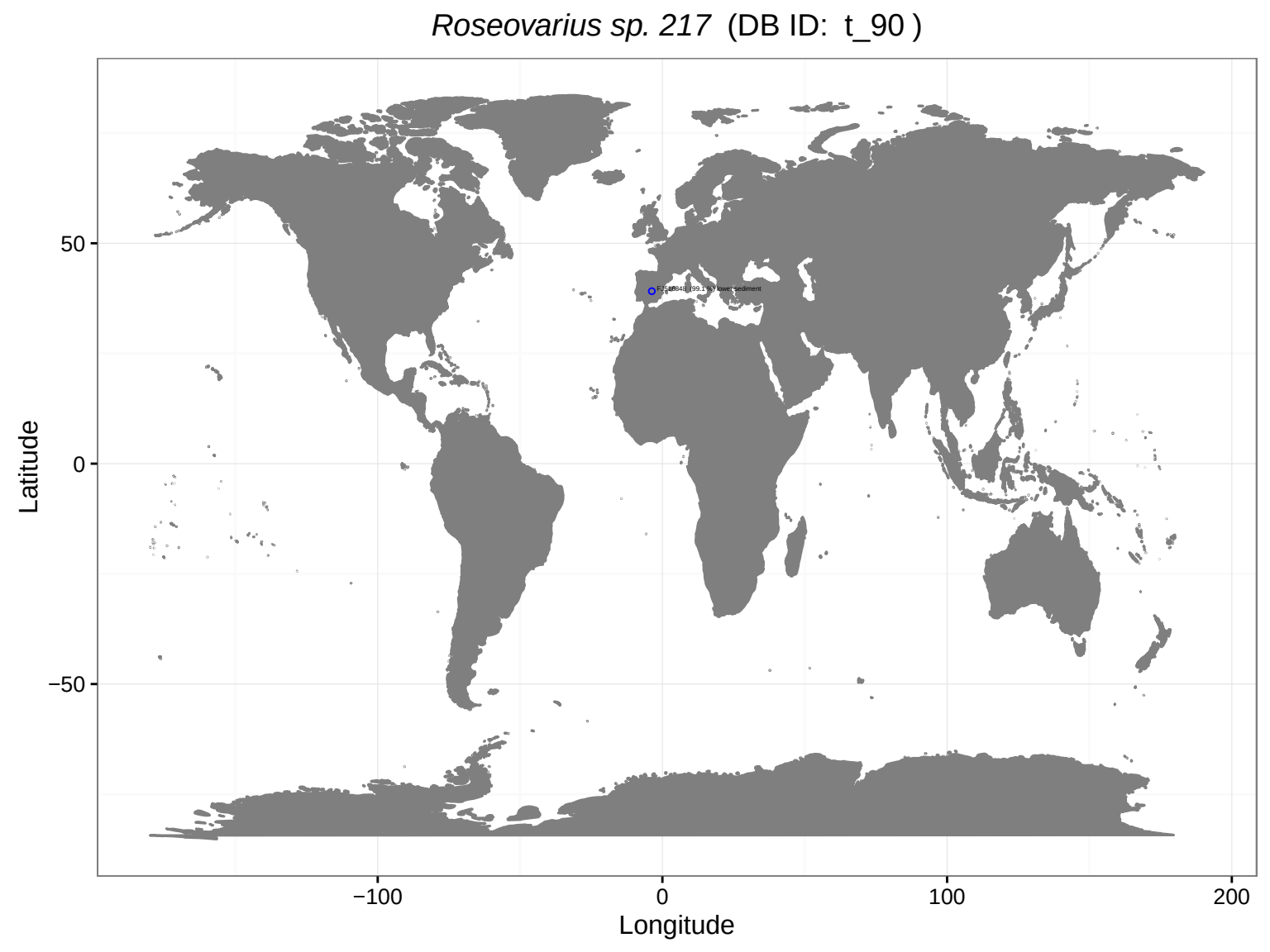


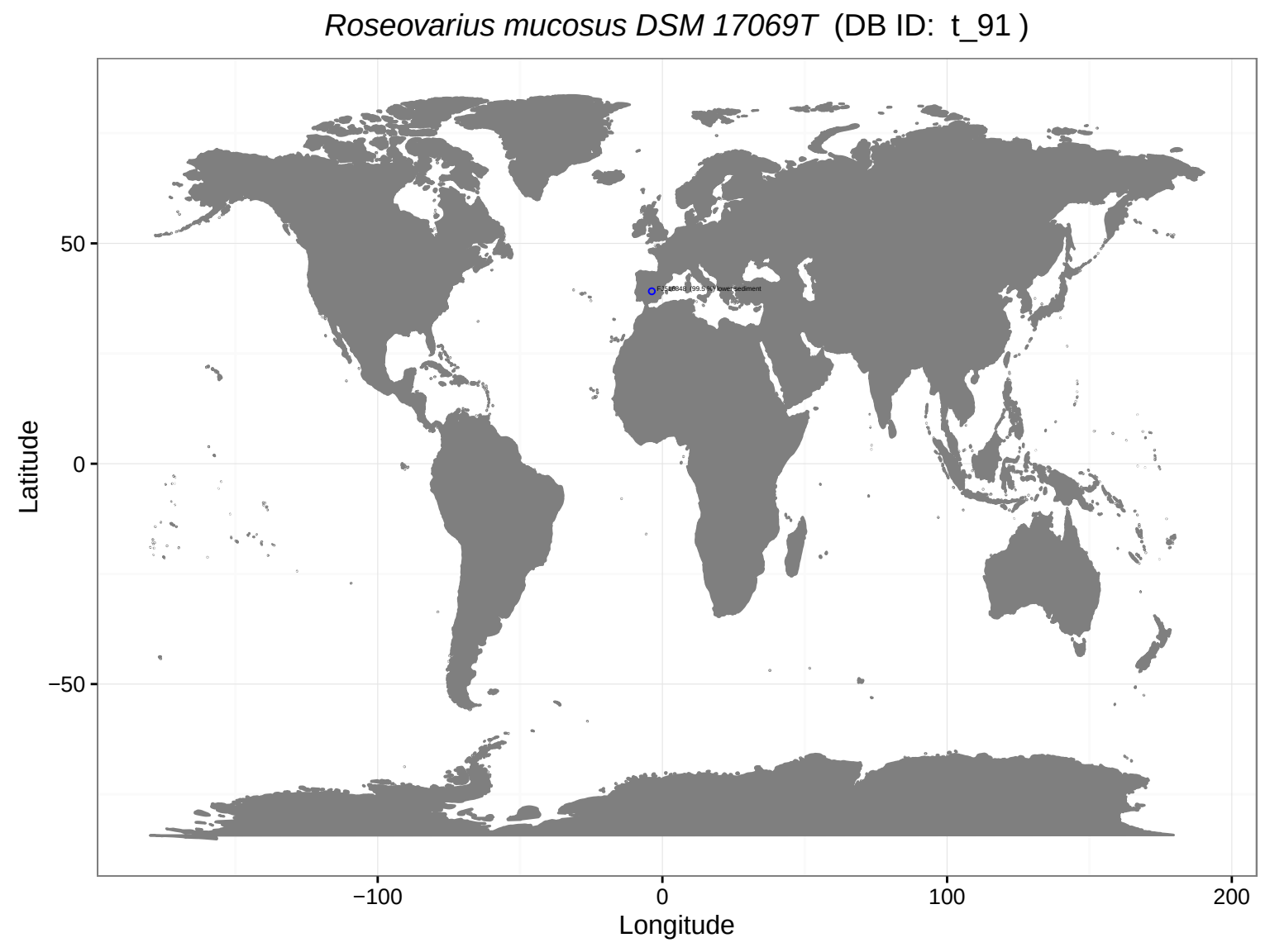

Rhodobacter sphaeroides WS8N (DB ID: t_86)
50
0
−50
Latitude
−100
0
100
200
Longitude

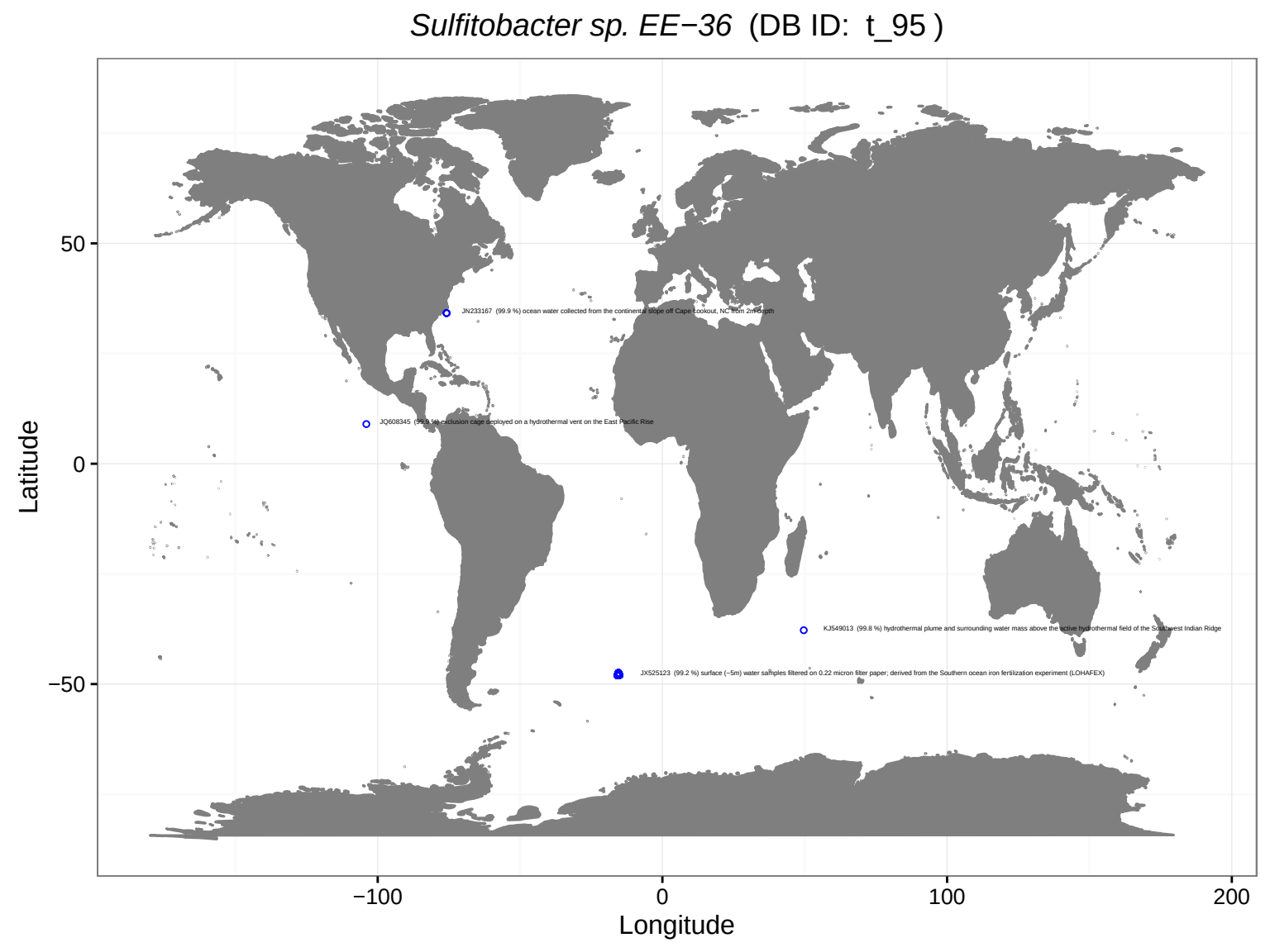

Rhodobacterales bacterium HTCC2255 (DB ID: t_87)
EU092186 (99.6 %) North Sea water sample
AM237834 (99.8 %) surface coastal North Sea water
HQ242499 (99.9 %) La Perouse Station CS04 depth of 10m
GQ346731 (99.9 %) Saanich Inlet, 10 m depth
EF202240 (100 %) Boothbay Harbor, 1m depth
EU799066 (99.7 %) Newport Harbour, RI
EF016461 (99.9 %) plankton DNA extracts
DQ009299 (99.8 %) glass fiber prefiltered bacterioplankton DNA collected at 5M depth from the San Pedro Ocean Time Series
AB374059 (99.4 %) Seawater in East China Sea
KM223997 (99.6 %) 0.8 micron prefiltered seawater from 5 m depth from the Chilean upwelling in the South Pacific Ocean
JN625639 (99.9 %) sea surface water
JX525073 (99.3 %) surface (~5m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
EU005706 (99.9 %) 3 micron prefiltered seawater from 120 m depth (station A3) above the Kerguelen Plateau in the Southern
JN625621 (99.8 %) Antarctic sea water collected from 5m
50
0
−50
Latitude
−100
0
100
200
Longitude

Roseovarius nubinhibens ISMT (DB ID: t_89)

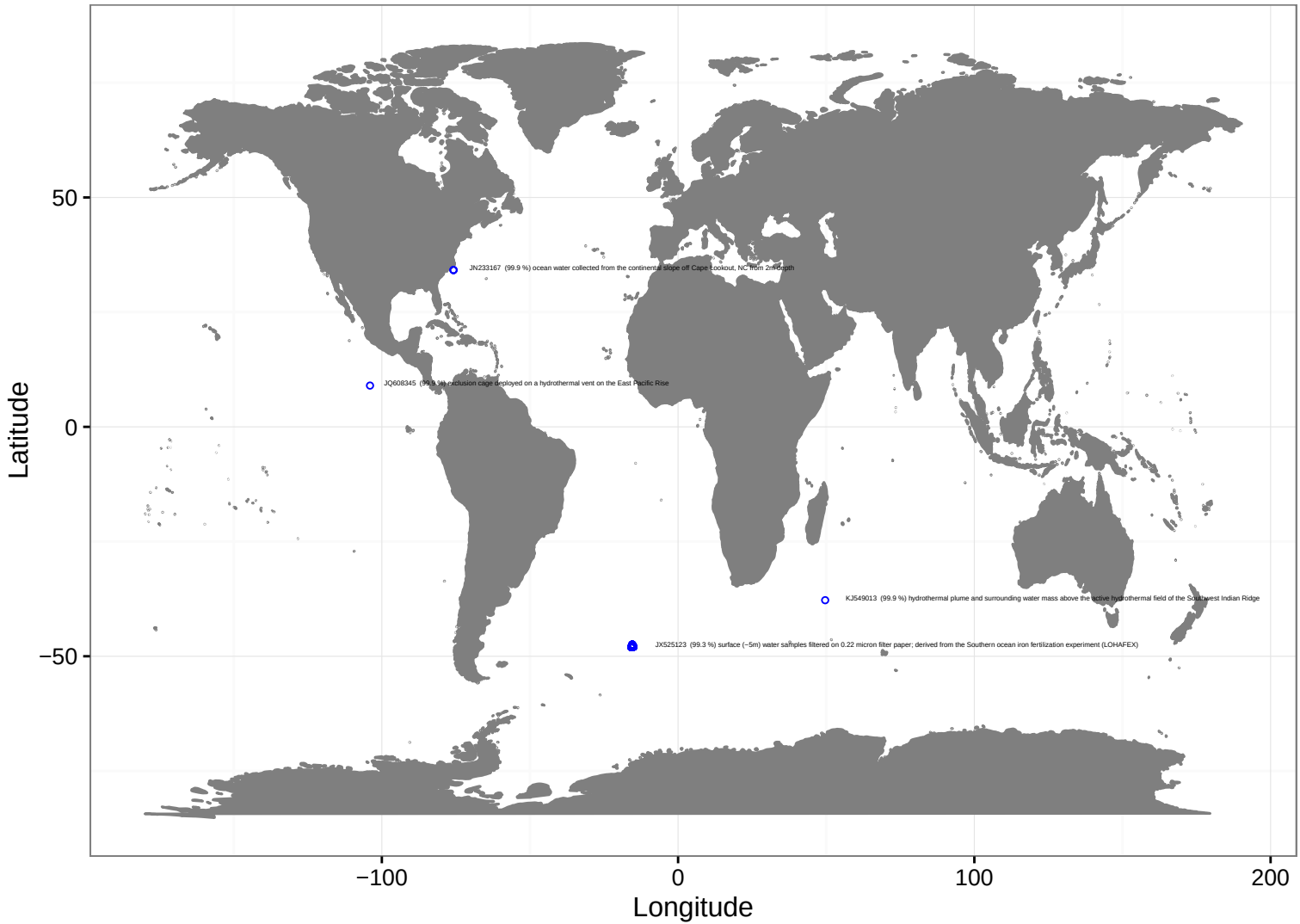



Roseovarius sp. 217 (DB ID: t_90)
50
0
−50
Latitude
−100
0
100
200
Longitude


Roseovarius mucosus DSM 17069T (DB ID: t_91)
50
0
−50
Latitude
−100
0
100
200
Longitude


Sulfitobacter sp. EE−36 (DB ID: t_95)
JN233167 (99.9 %) ocean water collected from the continental slope off Cape Lookout, NC from 2m depth
JQ608345 (99.9 %) exclusion cage deployed on a hydrothermal vent on the East Pacific Rise
KJ549013 (99.8 %) hydrothermal plume and surrounding water mass above the active hydrothermal field of the Southwest Indian Ridge
JX525123 (99.2 %) surface (~5m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)
50
0
−50
Latitude
−100
0
100
200
Longitude

Sulfitobacter sp. NAS−14.1 (DB ID: t_96)



Paracoccus sp. J55 (DB ID: t_99)