Genome sequence of the *Wenxinia marina* type strain (DSM 24838T), a representative of the *Roseobacter* group isolated from oilfield sediments

Thomas Riedel[1,2], Anne Fiebig[3], James Han[4], Marcel Huntemann[4], Stefan Spring[3], Jörn Petersen[3], Natalia N. Ivanova[4], Victor Markowitz[5], Tanja Woyke[4], Markus Göker[3*], Nikos C. Kyrpides[4], Hans-Peter Klenk[3]

[1] **Sorbonne Universités, UPMC Univ Paris 06, USR 3579, LBBM, Observatoire Océanologique, Banyuls/Mer, France**
[2] **CNRS, USR 3579, LBBM, Observatoire Océanologique, Banyuls-sur-Mer, France**
[3] **Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany**
[4] **DOE Joint Genome Institute, Walnut Creek, California, USA**
[5] **Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA**

*Correspondence: Markus Göker

Wenxinia marina Ying *et al.* 2007 is the type species of the genus *Wenxinia*, a representative of the *Roseobacter* group within the alphaproteobacterial family *Rhodobacteraceae*, isolated from oilfield sediments of the South China Sea. This family was shown to harbor the most abundant bacteria especially from coastal and polar waters, but was also found in microbial mats, sediments and attached to different kind of surfaces.

Here we describe the features of *W. marina* strain HY34T together with the genome sequence and annotation of strain DSM 24838T and novel aspects of its phenotype. The 4,181,754 bp containing genome sequence encodes 4,047 protein-coding genes and 59 RNA genes. The genome of *W. marina* DSM 24838T was sequenced as part of the activities of the Genomic Encyclopedia of Type Strains, Phase I: the one thousand microbial genomes (KMG) project funded by the DoE and the Transregional Collaborative Research Centre 51 (TRR51) funded by the German Research Foundation (DFG).

Introduction

Strain HY34T (= DSM 24838T = CGMCC 1.6105T = JCM 14017T) is the type strain of *Wenxinia marina* in the monospecific genus *Wenxinia* [1,2], which belongs to the widely distributed marine *Roseobacter* group [3]. The strain was isolated from sediments of the Xijiang oilfield located in the South China Sea (China) [1]. The genus *Wenxinia* was named after Professor Wen-Xin Chen, a Chinese pioneer in soil microbiology. The species epithet *marina* refers to the Latin adjective *marina* ('of or belonging to the sea') [1,2]. Current PubMed records do not indicate any follow-up research with strain HY34T after the initial description of *W. marina* [1].

In this study we analyzed the genome sequence of *W. marina* DSM 24838T. We present a description of the genome sequencing and annotation and present a summary classification together with a set of features for strain HY34T, including novel aspects of its phenotype.

Classifications and features
16S rRNA gene analysis
Figure 1 shows the phylogenetic neighborhood of *W. marina* in a 16S rRNA gene based tree. The sequences of the two identical 16S rRNA gene copies in the genome, differ by three nucleotides from

the previously published 16S rRNA gene sequence (DQ640643).

A representative genomic 16S rRNA gene sequence of *W. marina* DSM 24838ᵀ was compared with the Greengenes database for determining the weighted relative frequencies of taxa and (truncated) keywords as previously described [4]. The most frequently occurring genera were *Ruegeria* (41.6%), *Paracoccus* (31.0%), *Oceanicola* (14.0%), *Silicibacter* (5.0%) and *Loktanella* (3.3%) (60 hits in total). Among all other species, the one yielding the highest score was *Oceanicola granulosus* (AAOT01000021), which corresponded to an identity of 94.7% and an HSP coverage of 99.6%. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification.) The highest-scoring environmental sequence was DQ640643 (Greengenes short name '*Rhodobacteraceae* South China Sea oil field sediment isolate HY34 *Rhodobacteraceae* str. HY34'), which showed an identity of 99.8% and an HSP coverage of 100.0%. The most frequently occurring keywords within the labels of all environmental samples that yielded hits were 'microbi' (4.3%), 'coral' (3.6%), 'sea' (2.6%), 'diseas' (2.5%) and 'china' (2.4%) (190 hits in total). The most frequently occurring keywords within the labels of those environmental samples which yielded hits of a higher score than the highest scoring species were 'antecubit, fossa, skin' (13.9%) and 'china, field, oil, rhodobacteracea, sea, sediment, south' (8.3%) (3 hits in total). Some of these keywords fit well to the isolation site of strain HY34ᵀ [1].
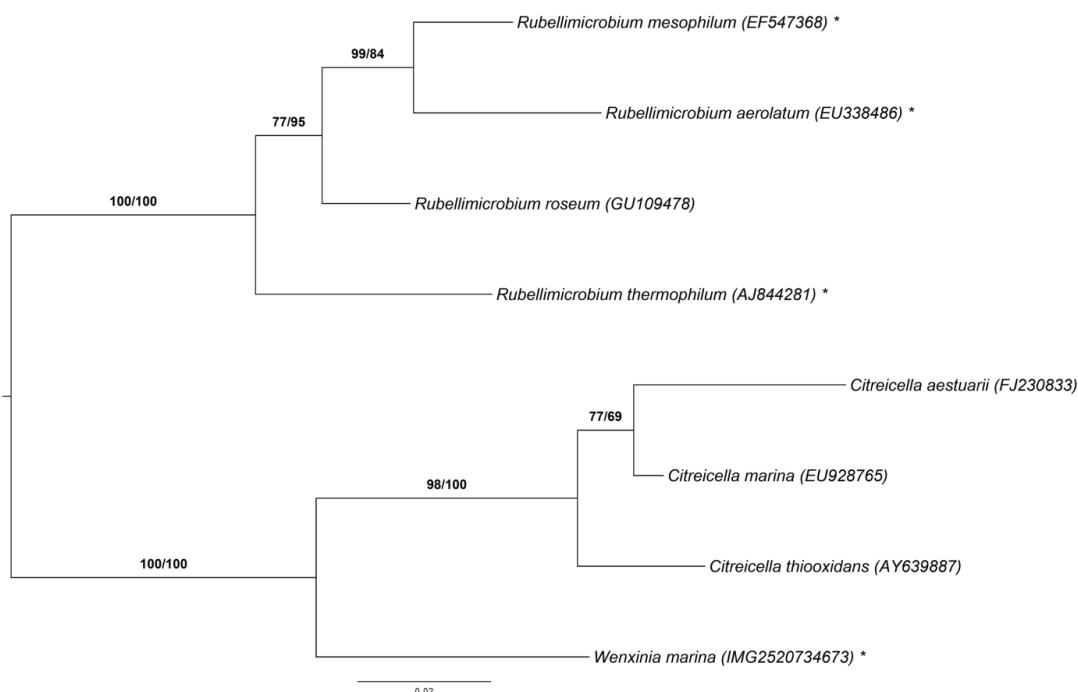

Figure 1. Phylogenetic tree highlighting the position of *W. marina* relative to the type strains of the neighboring genera *Citreicella* and *Rubellimicrobium*. The tree was inferred from 1,381 aligned characters of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion as previously described [4]. The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 1,000 ML bootstrap replicates (left) and from 1,000 maximum-parsimony bootstrap replicates (right) if larger than 60% [4]. Lineages with type strain genome sequencing projects registered in GOLD [5] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks [6].

Morphology and physiology

Cells of strain HY34ᵀ form Gram-negative, ovoid or short rods (0.7-0.8 μm in width and 1.3 μm in length) [Figure 2]. Motility and sporulation were not observed. Cells are strictly aerobic and display a heterotrophic lifestyle. When cultured on Marine Agar 2216 colonies with a weak pink color became visible, but bacteriochlorophyll *a* was not detected. The strain grows in a temperature range of 15-42°C with an optimum at 34-38°C. NaCl is required for growth (0.5-9%) with an optimum salt concentration at 1-4%. Further, the strain grows in a range of pH 6.5-8.5 with an optimum pH of 7.5-8.0. The strain is oxidase- and catalase-

positive. Nitrate is reduced to nitrite. Indole and H_2S are not produced. Cells hydrolyze urea and Tween 20, and a weak hydrolysis of Tween 40 and Tween 80 was also detected. The strain does not hydrolyze agar, casein, starch, DNA or CM-cellulose. Strain HY34[T] accumulates polyhydroxyalkanoates in its cells. Tests for arginine dehydrolase and lecithinase were negative. Further, cells utilize sucrose, lactose, galactose, maltose, melezitose, L-rhamnose, L-fucose, trehalose, cellobiose, gluconate, lactic acid, malate, L-glutamic acid. The strain utilizes D-melibiose, inulin, methyl α-D-glucoside, glycerol, sorbitol, butanol, pyruvate, formic acid, L-alanine and L-proline weakly. Utilization of D-raffinose, mannitol, L-sorbose, dulcitol, adonitol, *myo*-inositol, methanol, ethanol, citrate, malonate, butyric acid and caprate acid was not detected. Cells produce acid from D-xylose, cellobiose, lactose, L-rhamnose, L-arabinose, D-raffinose, and weakly from sucrose, maltose, mannose, trehalose, and ribose. Strong activities for esterase (C8) and α- and β-glucosidases were detected, as well as weak activities for alkaline phosphatase, leucine arylamidase, valine arylamidase and naphthol-AS-BI-phosphohydrolase. No activity was found for acid phosphatase, *N*-acetyl-β-cysteine arylamidase, glucosamidase, α- and β-galactosidase, α-mannosidase, α-chrymotrypsin, β-glucuronidase, α-fucosidase and lipase (C_{14}). Cells of strain HY34[T] are found to be resistant to norfloxacin, tetracycline and gentamicin as well as sensitive to neomycin, polymyxin B, streptomycin, ampicillin, carbenicillin, vancomycin, ciprofloxacin, rifampicin, chloramphenicol, benzylpenicillin, kanamycin, and erythromycin (all data from [1] and presented in Table 1).

The utilization of carbon compounds by *W. marina* DSM 24838[T] grown at 28°C was also determined for this study using Generation-III microplates in an OmniLog phenotyping device (BIOLOG Inc., Hayward, CA, USA) [7]. The microplates were inoculated with a cell suspension at a cell density of 95-96% turbidity and dye IF-A. Further additives were vitamin, micronutrient and sea-salt solutions, which had to be added for dealing with such marine bacteria [8]. The plates were sealed with parafilm to avoid a loss of fluid.

The exported measurement data were further analyzed with the opm package for R [9,10], using its facilities for statistically estimating parameters from the respiration curves such as the maximum height, and automatically translating these values

into negative, ambiguous, and positive reactions. The reactions were recorded in three biological replicates.

On the Generation-III plates, the strain was positive for pH 6, 1% NaCl, 4% NaCl, 8% NaCl, D-galactose, 3-O-methyl-D-glucose, D-fucose, L-fucose, L-rhamnose, 1% sodium lactate, myoinositol, rifamycin SV, L-aspartic acid, L-glutamic acid, L-histidine, L-serine, D-glucuronic acid, quinic acid, L-lactic acid, citric acid, α-keto-glutaric acid, D-malic acid, L-malic acid, nalidixic acid, and sodium formate.

W. marina HY34[T] was negative for the following tests*:* dextrin, D-maltose, D-trehalose, D-cellobiose, β-gentiobiose, sucrose, D-turanose, stachyose, pH 5, D-raffinose, α-D-lactose, D-melibiose, β-methyl-D-galactoside, D-salicin, *N*-acetyl-D-glucosamine, *N*-acetyl-β-D-mannosamine, N-acetyl-D-galactosamine, *N*-acetyl-neuraminic acid, D-glucose, D-mannose, D-fructose, inosine, fusidic acid, D-serine, D-sorbitol, D-mannitol, D-arabitol, glycerol, D-glucose-6-phosphate, D-fructose-6-phosphate, D-aspartic acid, D-serine, troleandomycin, minocycline, gelatin, glycyl-L-proline, L-alanine, L-arginine, L-pyroglutamic acid, lincomycin, guanidine hydrochloride, niaproof, pectin, D-galacturonic acid, L-galactonic acid-γ-lactone, D-gluconic acid, glucuronamide, mucic acid, D-saccharic acid, vancomycin, tetrazolium violet, tetrazolium blue, phydroxy-phenylacetic acid, methyl pyruvate, D-lactic acid methyl ester, bromo-succinic acid, lithium chloride, potassium tellurite, tween 40, γ-amino-n-butyric acid, α-hydroxy-butyric acid, β-hydroxy-butyric acid, α-keto-butyric acid, acetoacetic acid, propionic acid, acetic acid, aztreonam, butyric acid and sodium bromate.

The phenotype microarray results fit to those reported by Ying and colleagues [1] in large part. Only the utilization of lactose and D-trehalose could not be confirmed by respiration measurements under the given conditions. Interestingly, *W. marina* DSM 24838[T] showed a varying phenotype both in growth measurement [1] and in the respiration curves among replicates. Ying and colleagues reported eleven substrates yielding "weak" results, which complicates the exact comparison of substrate utilization [1]. In contrast to Ying and colleagues, the OmniLog measurements gave positive reactions for L-histidine and myoinositol. This may be due respiratory measurements being more sensitive than growth measurements [11].
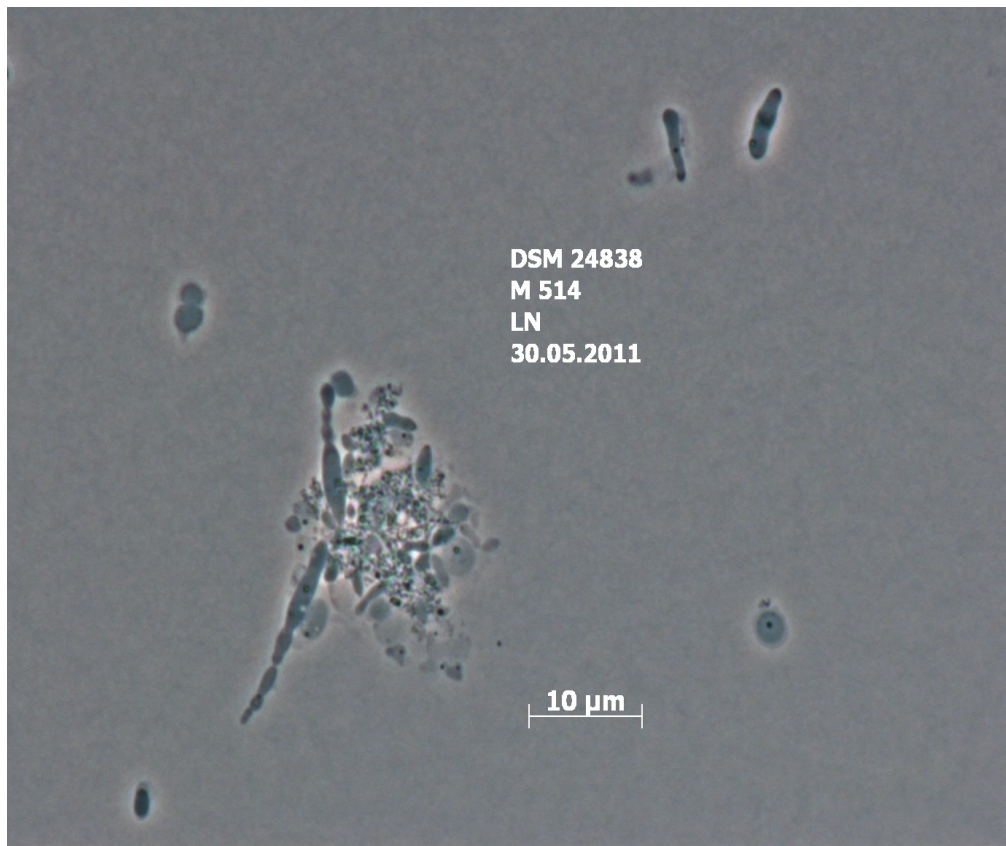
Figure 2. Phase contrast micrograph of *W. marina* DSM 24838^T.

Chemotaxonomy

The principal cellular fatty acids of strain HY34^T are $C_{18:1\,\omega7c}$ (57.1%), $C_{16:0}$ (16.5%), 11-methyl $C_{18:1\,\omega7c}$ (5.4%), $C_{18:0}$ (3.9%), $C_{14:0}$ (3.7%), $C_{15:1\,iso\,G}$ and $C_{15:1\,iso\,I}$ (3.4%), summed feature 3 $C_{16:1\,\omega7c}$ and/or $C_{15:0\,2-OH}$ (1.9%), $C_{12:0}$ (1.6%) and $C_{13:0\,2-OH}$ (1.2%). The major respiratory lipoquinone was ubiquinone 10, which is a well-known characteristic of the *Alphaproteobacteria*. Phosphatidylglycerol and phosphatidylcholine were identified as the major polar lipids. In contrast to other representatives of the *Roseobacter* group such as *Marinovum algicola* FF3^T (DSM 10251^T) [12,13], strain HY34^T also contains an unidentified glycolipid called L1, which shows similarities to an unidentified phospholipid of *Ruegeria atlantica* DSM 5828^T (all data from [1]).

Genome sequencing and annotation
Genome project history

This strain was twice selected for genome sequencing on the basis of its phylogenetic position [14]. First as part of the DFG funded project "Ecology, Physiology and Molecular Biology of the *Roseobacter* clade: Towards a Systems Biology Understanding of a Globally Important Clade of Marine Bacteria" and later as part of the "Genomic Encyclopedia of Type Strains, Phase I: the one thousand microbial genomes (KMG) project" [15], a follow-up of the GEBA project [16], which aims in increasing the sequencing coverage of key reference microbial genomes. The strain was independently sequenced from the same source of DNA and produced draft sequences that were finally joined. The project information can found in the Genomes OnLine Database [5] and the Integrated Microbial Genomes database (IMG) [17]. A summary of the project information is shown in Table 2.

Growth conditions and DNA isolation

A culture of *W. marina* DSM 24838^T was grown aerobically in DSMZ medium 514 [18] at 30°C. Genomic DNA was isolated using Jetflex Genomic DNA Purification Kit (GENOMED 600100) following the standard protocol provided by the manufacturer but modified by an incubation time of 60 min, incubation on ice overnight on a shaker, the use of an additional 50 µl proteinase K, and the addition of 100 µl protein precipitation buffer. The DNA is available from the Leibniz-Institute DSMZ through the DNA Bank Network [19].

Table 1. Classification and general features of *W. marina* HY34[T] according to the MIGS recommendations [41] (published by the Genomic Standards Consortium [42]).

MIGS ID	Property	Term	Evidence code
		Domain *Bacteria*	TAS [43]
		Phylum *Proteobacteria*	TAS [44]
		Class *Alphaproteobacteria*	TAS [45,46]
	Current classification	Order *Rhodobacterales*	TAS [46,47]
		Family *Rhodobacteraceae*	TAS [48]
		Genus *Wenxinia*	TAS [1]
		Species *Wenxinia marina*	TAS [1]
		Strain HY34[T]	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	ovoid or short rods	TAS [1]
	Motility	non-motile	TAS [1]
	Sporulation	non-spore-forming	TAS [1]
	Temperature range	15-42°C	TAS [1]
	Optimum temperature	34-38°C	TAS [1]
	Salinity	0.5-9% (NaCl)	TAS [1]
MIGS-22	Oxygen requirement	aerobic	TAS [1]
	Carbon source	Yeast extract, peptone	TAS [1]
	Energy metabolism	heterotroph	TAS [1]
MIGS-6	Habitat	Oilfield sediment	TAS [1]
MIGS-15	Biotic relationship	Free living	TAS [1]
MIGS-14	Pathogenicity	None	NAS
	Biosafety level	1	TAS [49]
MIGS-23.1	Isolation	Oilfield sediment	TAS [1]
MIGS-4	Geographic location	Xijiang oilfield, South China Sea (China)	TAS [1]
MIGS-5	Sample collection time	before 2007	NAS
MIGS-4.1	Latitude	Not reported	
MIGS-4.2	Longitude	Not reported	
MIGS-4.3	Depth	100 m	TAS [1]
MIGS-4.4	Altitude	Not reported	

Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). Evidence codes are from of the Gene Ontology project [50].

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Non-contiguous finished
		Two genomic libraries: one Illumina PE library (539 bp insert size),
MIGS-28	Libraries used	one 454 PE library (3kb insert size)
MIGS-29	Sequencing platforms	Illumina GA IIx, Illumina MiSeq, 454 GS-FLX Titanium
MIGS-31.2	Sequencing coverage	356 ×
MIGS-30	Assemblers	velvet version 1.1.36, Newbler version 2.3, consed 20.0
MIGS-32	Gene calling method	Prodigal 1.4
	GenBank Date of Release	pending
	GOLD ID	Gi10895
	NCBI project ID	183669
		2519899719 (8 scaffold version)
	Database: IMG	and 2515154190 (41 scaffold version)
MIGS-13	Source material identifier	DSM 24838[T]
	Project relevance	Tree of Life, biodiversity

Genome sequencing and assembly

The genome sequencing under the DFG funded part of the project was perform as previously described for *Rubellimicrobium thermophilum* [6], with 3.3 million reads delivered by the first run on an Illumina GAII platform. To increase the sequencing depth, a second Ilumina run was performed, providing another 8.1 million reads. The first draft assembly from 9,139,639 filtered reads (median read length 122 nt) resulted in more than 300 contigs. To gain information on the contig arrangement an additional 454 run was performed. The paired-end pyrosequencing jumping library resulted in 158,608 reads, with an average read length of 450 bp. Both draft assemblies (Illumina and 454 sequences) were fractionated into artificial Sanger reads of 1,000 nt in length plus 75 bp overlap on each site. These artificial reads served as an input for the phred/phrap/consed package [20]. In combination the assembly resulted in 265 contigs in 26 scaffolds.

The genome sequencing under the DoE funded part of the project was performed as previously described for *Halomonas zhanjiangensis* [21] also using the Illumina technology [22]. An Illumina Standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform. All general aspects of library construction and sequencing performed at the JGI can be found at [23]. The final assembly for this part of the project resulted in 41 scaffolds covering 4,175,892 bp (ARAY00000000).

The draft sequence from the first (DFG-funded) part was mapped to the permanent draft version ARAY00000000 using minimus2 [24]. By manual editing the number of contigs was reduced to 22 in 8 scaffolds (AONG00000000). The combined sequences provided a 356 × coverage of the genome.

Genome annotation

Genes were identified using Prodigal [25] as part of the JGI genome annotation pipeline. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Identifications of RNA genes were carried out by using HMMER 3.0rc1 [26] (rRNAs) and tRNAscan-SE 1.23 (tRNAs) [27]. Other non-coding genes were predicted using INFERNAL 1.0.2 [28]. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [29] CRISPR elements were detected using CRT [30] and PILER-CR [51].

Genome properties

The genome statistics are provided in Table 3 and Figure 3. The genome of DSM 24838[T] has a total length of 4,181,754 bp and a G+C content of 70.5%. Of the 4,106 genes predicted, 4,047 were protein-coding genes, and 59 RNAs. The majority of the protein-coding genes (80.4%) were assigned a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3. Genome Statistics

| Attribute	Value	% of Total
Genome size (bp)	4,181,754	100.00
DNA coding region (bp)	3,740,397	89.45
DNA G+C content (bp)	2,948,333	70.50
Number of scaffolds	8	
Total genes	4,106	100.00
RNA genes	59	1.44
rRNA operons	2	
tRNA genes	45	1.10
Protein-coding genes	4,047	98.56
Genes with function prediction (proteins)	3,303	80.44
Genes in paralog clusters	3,408	83.00
Genes assigned to COGs	3,199	77.91
Genes assigned Pfam domains	3,379	82.29
Genes with signal peptides	430	10.47
Genes with transmembrane helices	904	22.02
CRISPR repeats	0	
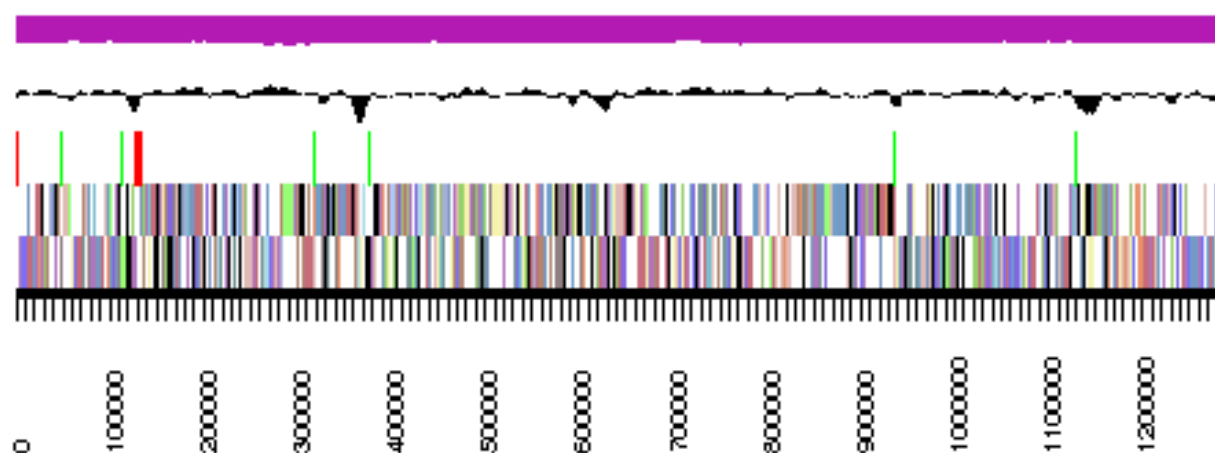
Figure 3. Graphical map of the largest scaffold. From bottom to the top: Genes on forward strand (colored by COG categories), Genes on reverse strand (colored by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content (black), GC skew (purple/olive).

Table 4. Number of genes associated with the general COG functional categories

Code	Value	%age	Description
J	170	4.9	Translation, ribosomal structure and biogenesis
A	3	0.1	RNA processing and modification
K	203	5.8	Transcription
L	153	4.4	Replication, recombination and repair
B	3	0.1	Chromatin structure and dynamics
D	27	0.8	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	40	1.1	Defense mechanisms
T	139	4.0	Signal transduction mechanisms
M	201	5.7	Cell wall/membrane/envelope biogenesis
N	24	0.7	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	54	1.5	Intracellular trafficking and secretion, and vesicular transport
O	123	3.5	Posttranslational modification, protein turnover, chaperones
C	214	6.1	Energy production and conversion
G	321	9.2	Carbohydrate transport and metabolism
E	372	10.6	Amino acid transport and metabolism
F	78	2.2	Nucleotide transport and metabolism
H	137	3.9	Coenzyme transport and metabolism
I	152	4.3	Lipid transport and metabolism
P	163	4.7	Inorganic ion transport and metabolism
Q	107	3.1	Secondary metabolites biosynthesis, transport and catabolism
R	439	12.5	General function prediction only
S	378	10.8	Function unknown
-	907	22.1	Not in COGs

Insights into the genome

Plasmids

Genome sequencing of *W. marina* DSM 24838ᵀ reveals the presence of one plasmid with a size of about 101 kb. The plasmid contains a characteristic replicase of the RepA-I type [31], but the typical module structure containing the replicase as well as a *par*AB partitioning operon was not found. A single *par*A gene (wenxma_04096) is located adjacent to the replicase and an additional *par*AB operon (wenxma_04090 to wenxma_04091) is located downstream of *rep*A-I. The plasmid harbors neither a plasmid stability module nor a type-IV secretion system.

The plasmid contains a large RTX-toxin (wenxma_04058) and is dominated by genes that are required for polysaccharide biosynthesis. It includes all four genes of the rhamnose pathway [32], but the *rmlA* gene for the glucose-1-phosphate thymidylyltransferase (EC 2.7.7.24; wenxma_04097) is separated from the remaining clustered genes (*rml*C, *rml*B, *rml*D; wenxma_04094 to wenxma_04092). The extrachromosomal replicon may be involved in surface attachment. Comparable RepA-I type plasmids with a similar genetic composition are also present in other *Rhodobacterales* including several *Phaeobacter* strains [33].

Phages

Many bacteria encode genome-inserted gene sequences, which are associated with prophages, one of the major reason for horizontal gene transfer and bacterial diversity [34,35]. The genome sequence of *W. marina* DSM 24838ᵀ was found to encode several prophage-associated gene sequences (e.g., wenxma_00641 to wenxma_00646, wenxma_00930 to wenxma_00936, wenxma_01496 to wenxma_01510).

Quorum Sensing

Analysis of the DSM 24838ᵀ genome sequence revealed the presence of gene sequences associated to quorum sensing (QS) [36-38]. QS is a bacterial communication system *via* chemical signal molecules called autoinducers, which are produced and released by QS bacteria to coordinate behaviors with respect to their population density [38]. Interestingly and surprisingly, QS induces also individual morphologies and cell division modes, which was recently shown for *D. shibae* DFL-12, another representative of the *Roseobacter* group

[39,40]. Regarding to QS the genome of DSM 24838ᵀ codes for, e.g., two N-acyl-L-homoserine-lactone synthetases (*Lux*I homologues, wenxma_01086 and wenxma_03269) and two genes possibly encoding QS-involved response and transcriptional regulators (*LuxR* homologues, wenxma_01085 and wenxma_03267).

Morphological traits

With regard to morphological traits, several genes associated with the putative production, biosynthesis and export of exopopolysaccharides (wenxma_00281, wenxma_02363 and wenxma_02364, wenxma_03720 and wenxma_03721) and capsule polysaccharides (wenxma_00822, wenxma_02023 to wenxma_02025, wenxma_02704 and wenxma_02705, wenxma_04069) were detected.

Interestingly, the genome of DSM 24838ᵀ was found to encode several gene sequences putatively involved in pili formation (e.g., wenxma_01776 to wenxma_01787, wenxma_03426 to wenxma_03435) and chemotaxis (e.g., wenxma_3823 to wenxma_03830), although the strain was described as non-motile [1]. Hence, it could be that the formed pili play a role for adhesion or switching-type motility on solid surfaces.

Further, according to its genome strain DSM 24838ᵀ accumulates polyhydroxyalkanoates as storage compounds (wenxma_02601 to wenxma_02604), which is in accordance with the findings of Ying and colleagues for strain HY34ᵀ [1].

References

1. Ying JY, Wang BJ, Dai X, Yang SS, Liu SJ, Liu ZP. *Wenxinia marina* gen. nov., a novel member of the *Roseobacter* clade isolated from oilfield sediments of the South China Sea. *Int J Syst Evol Microbiol* 2007; **57**:1711-1716. [Pub-Med](http://dx.doi.org/10.1099/ijs.0.64825-0)

2. Euzéby JP. List of prokaryotic names with standing in nomenclature. http://www.bacterio.cict.fr/s/wenxinia.html.

3. Buchan A, González JM, Moran MA. Overview of the marine *Roseobacter* lineage. *Appl Environ Microbiol* 2005; **71**:5665-5677. [Pub-Med](http://dx.doi.org/10.1128/AEM.71.10.5665-5677.2005)

4. Göker M, Cleland D, Saunders E, Lapidus A, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng JF, Tapia R, *et al*. Complete genome sequence of *Isosphaera pallida* type strain (IS1B^T). *Stand Genomic Sci* 2011; **4**:63-71. [Pub-Med](http://dx.doi.org/10.4056/sigs.1533840)

5. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2012; **40**:D571-D579. [Pub-Med](http://dx.doi.org/10.1093/nar/gkr1100)

6. Fiebig A, Riedel T, Gronow S, Petersen J, Klenk HP, Göker M. Genome sequence of the reddish-pigmented *Rubellimicrobium thermophilum* type strain (DSM 16684^T), a member of the *Roseobacter* clade. *Stand Genomic Sci* 2013; **8**:480-490. [Pub-Med](http://dx.doi.org/10.4056/sigs.4247911)

7. Bochner BR. Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* 2009; **33**:191-205. [PubMed](http://dx.doi.org/10.1111/j.1574-6976.2008.00149.x)

8. Buddruhs N, Pradella S, Göker M, Päuker O, Michael V, Pukall R, Spröer C, Schumann P, Petersen J, Brinkhoff T. Molecular and phenotypic analyses reveal the non-identity of the *Phaeobacter gallaeciensis* type strain deposits CIP 105210^T and DSM 17395. *Int J Syst Evol Microbiol* 2013; **63**:4340-4349. [Pub-Med](http://dx.doi.org/10.1099/ijs.0.053900-0)

9. Vaas LA, Sikorski J, Hofner B, Fiebig A, Buddruhs N, Klenk HP, Göker M. opm: an R package for analyzing Omnilog(R) phenotype microarray data. *Bioinformatics* 2013; **29**:1823-1824. [Med](http://dx.doi.org/10.1093/bioinformatics/btt291)

10. Vaas LA, Sikorski J, Michael V, Göker M, Klenk HP. Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLoS ONE* 2012; **7**:e34846. [Pub-Med](http://dx.doi.org/10.1371/journal.pone.0034846)

11. Vaas LAI, Marheine M, Sikorski J, Göker M, Schumacher M. Impacts of pr-10a overexpression at the molecular and the phenotypic level. *Int J Mol Sci* 2013; **14**:15141-15166; http://dx.doi.org/10.3390/ijms140715141. [PubMed]

12. Lafay B, Ruimy R, de Traubenberg CR, Breittmayer V, Gauthier MJ, Christen R. *Roseobacter algicola* sp. nov., a new marine bacterium isolated from the phycosphere of the toxin-producing dinoflagellate *Prorocentrum lima*. *Int J Syst Bacteriol* 1995; **45**:290-296. [Pub-Med](http://dx.doi.org/10.1099/00207713-45-2-290)

13. Martens T, Heidorn T, Pukall R, Simon M, Tindall BJ, Brinkhoff T. Reclassification of *Roseobacter gallaeciensis* Ruiz-Ponte et al. 1998 as *Phaeobacter gallaeciensis* gen. nov., comb. nov., description of *Phaeobacter inhibens* sp. nov., antibiotic-producing members of the *Roseobacter* clade, reclassification of *Ruegeria algicola* (Lafay et al. 1995) Uchino et al. 1998 as *Marinovum algicola* gen. nov., comb. nov., and emended descriptions of the genera *Roseobacter*, *Ruegeria* and *Leisingera*. *Int J Syst Evol Microbiol* 2006; **56**:1293-1304. [Pub-Med](http://dx.doi.org/10.1099/ijs.0.63724-0)

14. Göker M, Klenk HP. Phylogeny-driven target selection for large-scale genome sequencing (and other) projects. *Stand Genomic Sci* 2013; **8**:360-374. [Pub-Med](http://dx.doi.org/10.4056/sigs.3446951)

15. Kyrpides NC, Woyke T, Eisen JA, Garrity G, Lilburn TG, Beck BJ, Whitman WB, Hugenholz P, Klenk HP. Genomic Encyclopedia of Type Strains, Phase I: the one thousand microbial genomes (KMG-I) project. *Stand Genomic Sci* 2013; **9**:628-634. http://dx.doi.org/10.4056/sigs.5068949

16. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*. A phylogeny-driven Genomic Encyclopaedia of Bacteria and *Archaea*. *Nature*

2009; **462**:1056-1060. [Pub-](#)[Med](#) [http://dx.doi.org/10.1038/nature08656](#)

17. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, *et al*. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 2012; **40**:D115-D122. [PubMed](#) [http://dx.doi.org/10.1093/nar/gkr1044](#)

18. List of growth media used at the DSMZ: http://www.dmsz.de/catalogues/cataloque-microorganisms/culture-technology/list-of-media-for-microorganisms.html.

19. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar G, Klenk HP, Güntsch AM, Berendsohn WG, Wägele JW. The DNA Bank Network: the start from a German initiative. *Biopreserv Biobank* 2011; **9**:51-55. [http://dx.doi.org/10.1089/bio.2010.0029](#)

20. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**:195-202. [Med](#) [http://dx.doi.org/10.1101/gr.8.3.195](#)

21. Zhou Y, Li R, Gao XY, Lapidus A, Han J, Haynes M, Lobos E, Huntemann M, Pati A, Ivanova NN, *et al*. High quality draft genome sequence of the slightly halophilic bacterium *Halomonas zhanjiangensis* type strain JSM 078169^T (DSM 21076^T) from a sea urchin in southern China. *Stand Genomic Sci* (under review).

22. Bennett S. Solexa Ltd. *Pharmacogenomics* 2004; **5**:433-438. [Med](#) [http://dx.doi.org/10.1517/14622416.5.4.433](#)

23. The DOE Joint Genome Institute. http://www.jgi.doe.gov

24. Sommer DD, Delcher AL, Salzberg SL, Pop M. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* 2007; **8**:64. [Pub-](#)[Med](#) [http://dx.doi.org/10.1186/1471-2105-8-64](#)

25. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [Pub-](#)[Med](#) [http://dx.doi.org/10.1186/1471-2105-11-119](#)

26. Finn DR, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. *Web Server Issue* 2011; **39**:W29-W37.

27. Lowe TM, Eddy SR. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res* 1997;

25:955-964. [Med](#) [http://dx.doi.org/10.1093/nar/25.5.0955](#)

28. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: Inference of RNA alignments. *Bioinformatics* 2009; **25**:1335-1337. [Pub-](#)[Med](#) [http://dx.doi.org/10.1093/bioinformatics/btp157](#)

29. Markowitz VM, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [Pub-](#)[Med](#) [http://dx.doi.org/10.1093/bioinformatics/btp393](#)

30. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 2007; **8**:209. [Med](#) [http://dx.doi.org/10.1186/1471-2105-8-209](#)

31. Petersen J. Phylogeny and compatibility: plasmid classification in the genomics era. *Arch Microbiol* 2011; **193**:313-321. [PubMed](#)

32. Giraud MF, Naismith JH. The rhamnose pathway. *Curr Opin Struct Biol* 2000; **10**:687-696. [Pub-](#)[Med](#) [http://dx.doi.org/10.1016/S0959-440X(00)00145-7](#)

33. Thole S, Kalhoefer D, Voget S, Berger M, Engelhardt T, Liesegang H, Wollherr A, Kjelleberg S, Daniel R, Simon M, *et al*. *Phaeobacter gallaeciensis* genomes from globally opposite locations reveal high similarity of adaption to surface life. *ISME J* 2012; **6**:2229-2244. [Pub-](#)[Med](#) [http://dx.doi.org/10.1038/ismej.2012.62](#)

34. Canchaya C, Proux C, Fournous G, Bruttin A. Prophage genomics. *Microbiol Mol Biol Rev* 2003; **67**:238-276. [Pub-](#)[Med](#) [http://dx.doi.org/10.1128/MMBR.67.2.238-276.2003](#)

35. Paul JH. Prophages in marine bacteria: dangerous molecular bombs or the key to survival in the seas? *ISME J* 2008; **2**:579-589. [Pub-](#)[Med](#) [http://dx.doi.org/10.1038/ismej.2008.35](#)

36. Bassler BL. How bacteria talk to each other: regulation of gene expression by quorum sensing. *Curr Opin Microbiol* 1999; **2**:582-587. [Pub-](#)[Med](#) [http://dx.doi.org/10.1016/S1369-5274(99)00025-9](#)

37. Henke JM, Bassler BL. Bacterial social engagements. *Trends Cell Biol* 2004; **14**:648-656. [Pub-](#)[Med](#) [http://dx.doi.org/10.1016/j.tcb.2004.09.012](#)

38. Waters CM, Bassler BL. Quorum Sensing: Cell-to-Cell Communication in Bacteria. *Annu Rev Cell Dev Biol* 2005; **21**:319-246. PubMed http://dx.doi.org/10.1146/annurev.cellbio.21.012704.131001

39. Wagner-Döbler I, Ballhausen B, Berger M, Brinkhoff H, Buchholz I, Bunk B, Cypionka H, Daniel R, Gerdts G, Hahnke S, *et al*. The complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: a hitchhiker's guide to life in the sea. *ISME J* 2010; **4**:61-77. PubMed http://dx.doi.org/10.1038/ismej.2009.94

40. Patzelt D, Wang H, Buchholz I, Rohde M, Gröbe L, Pradella S, Neumann A, Schulz S, Heyber S, Münch K, *et al*. You are what you talk: quorum sensing induces individual morphologies and cell division modes in *Dinoroseobacter shibae*. *ISME J* 2013; **7**:2274-2286. PubMed http://dx.doi.org/10.1038/ismej.2013.107

41. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

42. Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, Gilbert J, Glöckner FO, Hirschman L, Karsch-Mzrachi I, *et al*. The Genomic Standards Consortium. *PLoS Biol* 2011; **9**:e1001088. PubMed http://dx.doi.org/10.1371/journal.pbio.1001088

43. Woese CR, Kandler O, Weelis ML. Towards a natural system of organisms. Proposal for the domains *Archaea, Bacteria* and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. Med http://dx.doi.org/10.1073/pnas.87.12.4576

44. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl nov. *In*: Brenner DJ, Krieg NR, Stanley JT, Garrity GM (*eds*), Bergey's Manual of Sytematic Bacteriology, second edition. Vol. 2 (The *Proteobacteria*), part B (The *Gammaproteobacteria*), Springer, New York, 2005, p. 1.

45. Garrity GM, Bell JA, Lilburn T. Class I. *Alphapro teobacteria* class. nov. *In*: Brenner DJ, Krieg NR, Stanley JT, Garrity GM (*eds*), Bergey's Manual of Sytematic Bacteriology, second edition. Vol. 2 (The *Proteobacteria*), part C (The *Alpha-, Beta-, Delta-*, and *Epsilonproteobacteria*), Springer, New York, 2005, p. 1.

46. Validation List No. 107. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2006; **56**:1-6. Med http://dx.doi.org/10.1099/ijs.0.64188-0

47. Garrity GM, Bell JA, Lilburn T. Order III. *Rhodobacterales* ord. nov. *In*: Brenner DJ, Krieg NR, Staley JT, Garrity GM (*eds*), Bergey's Manual of Systematic Bacteriology, second edition. vol. 2 (The *Proteobacteria*), part C (The *Alpha-, Beta-, Delta-*, and *Epsilonproteobacteria*), Springer, New York, 2005, p. 161.

48. Garrity GM, Bell JA, Lilburn T. Family I. *Rhodobacteraceae* fam. nov. *In*: Brenner DJ, Krieg NR, Staley JT, Garrity GM (*eds*), Bergey's Manual of Systematic Bacteriology, second edition. vol. 2 (The *Proteobacteria*), part C (The *Alpha-, Beta-, Delta-*, and *Epsilonproteobacteria*), Springer, New York, 2005, p. 161.

49. BAuA. Classification of *Bacteria* and *Archaea* in risk groups. *TRBA* 2010; **466**:93.

50. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. PubMed http://dx.doi.org/10.1038/75556

51. PILER Genomic repeat analysis software. 2009.