

# Complete genome sequence of the *Phaeobacter gallaeciensis* type strain CIP 105210<sup>T</sup> (= DSM 26640<sup>T</sup> = BS107<sup>T</sup>)

Oliver Frank<sup>1</sup>, Silke Pradella<sup>1</sup>, Manfred Rohde<sup>2</sup>, Carmen Scheuner<sup>1</sup>, Hans-Peter Klenk<sup>1</sup>, Markus Göker<sup>1</sup>, Jörn Petersen<sup>1\*</sup>

<sup>1</sup> Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

<sup>2</sup> Helmholtz-Centre for Infection Research, Braunschweig, Germany

\*Correspondence: Jörn Petersen (joern.petersen@dsmz.de)

Keywords: *Alphaproteobacteria*, *Roseobacter* group, Plasmid wealth, Replication systems, Sister species, *Phaeobacter inhibens*.

---

*Phaeobacter gallaeciensis* CIP 105210<sup>T</sup> (= DSM 26640<sup>T</sup> = BS107<sup>T</sup>) is the type strain of the species *Phaeobacter gallaeciensis*. The genus *Phaeobacter* belongs to the marine *Roseobacter* group (*Rhodobacteraceae*, *Alphaproteobacteria*). *Phaeobacter* species are effective colonizers of marine surfaces, including frequent associations with eukaryotes. Strain BS107<sup>T</sup> was isolated from a rearing of the scallop *Pecten maximus*. Here we describe the features of this organism, together with the complete genome sequence, comprising eight circular replicons with a total of 4,448 genes. In addition to a high number of extrachromosomal replicons, the genome contains six genomic island and three putative prophage regions, as well as a hybrid between a plasmid and a circular phage. Phylogenomic analyses confirm previous results, which indicated that the originally reported *P. gallaeciensis* type-strain deposit DSM 17395 belongs to *P. inhibens* and that CIP 105210<sup>T</sup> (= DSM 26640<sup>T</sup>) is the sole genome-sequenced representative of *P. gallaeciensis*.

---

## Introduction

Strain CIP 105210<sup>T</sup> (= BS107<sup>T</sup> = DSM 26640<sup>T</sup>) is the type strain of *Phaeobacter gallaeciensis*, the type species of *Phaeobacter*, a genus of marine species of *Rhodobacteraceae* (*Rhodobacterales*, *Alphaproteobacteria*). BS107<sup>T</sup> was isolated from the scallop *Pecten maximus* and was initially described as the type strain of *Roseobacter gallaeciensis* [1]. After comprehensive reclassifications of *Rhodobacteraceae* genera, BS107<sup>T</sup> became the type strain of the species *P. gallaeciensis* [2], currently comprising the species *P. gallaeciensis*, *P. inhibens*, *P. caeruleus*, *P. daeponensis*, *P. leonis* and *P. arcticus*. A recent study [3] revealed the non-identity of the reported identical deposits DSM 17395 and CIP 105210<sup>T</sup> and confirmed that the strain CIP 105210<sup>T</sup> represents the original *P. gallaeciensis* isolate BS107<sup>T</sup>, which is now deposited in the DSMZ open collection as DSM 26640<sup>T</sup>. In contrast, strain DSM 17395 was reclassified as a representative of the sister species *P. inhibens*. Analysis of their similar, but distinct metabolic capacities allowed for a discrimination between the two strains, which were originally reported to

represent the same type strain [3]. Thus, in the absence of sequenced genomes, the assignment to species was essentially based on deviating plasmid profiles and molecular analyses (16S rDNA, ITS, DNA-DNA hybridization), which showed convergent results.

The genus *Phaeobacter* comprises effective surface colonizers. Comparative analyses of strains DSM 17395 and DSM 24588 (= 2.10) revealed a high level of adaptation to life on surfaces [4]. The production of the characteristic antibiotic tropodithietic acid (TDA) correlates with the formation of a brown pigment that is eponymous for *Phaeobacter* [1]. Current scientific interest in *Phaeobacter* is based on the role of its strains as probiotic agents in fish aquaculture [5] and as agents of bleaching diseases in marine red algae [6], as well as on their potential regulatory activity during phytoplankton blooms [7] via so-called roseobactinoids [8]. Here we present the complete genome sequence of *P. gallaeciensis* CIP 105210<sup>T</sup>, together with a summary classification and a set of features, including insights into genome architecture, genomic islands and phages.



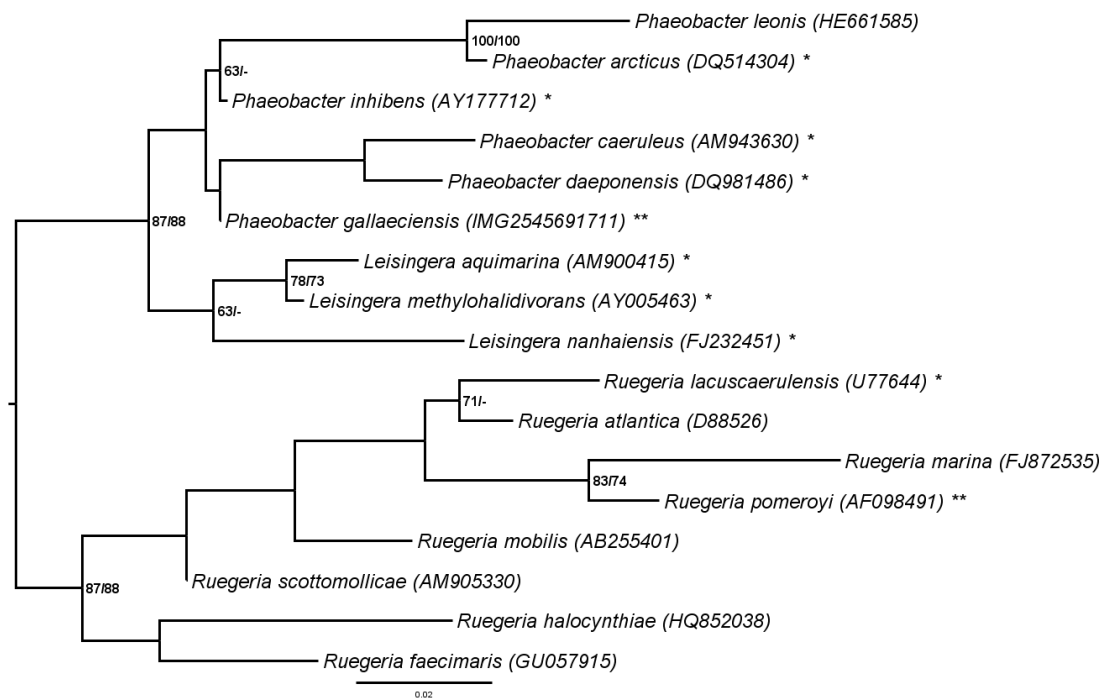
## Classification and features

### 16S rRNA gene analysis

Figure 1 shows the phylogenetic neighborhood of *P. gallaeciensis* CIP 105210<sup>T</sup> in a 16S rDNA gene sequence based tree. The sequences of the four 16S rRNA identical gene copies in the genome differ by five nucleotides from the previously published 16S rDNA gene sequence (Y13244 [1]).

A representative genomic 16S rDNA gene sequence of *P. gallaeciensis* CIP 105210<sup>T</sup> was compared with the Greengenes database for determining the weighted relative frequencies of taxa and (truncated) keywords as previously described [9], to infer the taxonomic and environmental affiliation of the strain. The most frequently occurring genera were *Ruegeria* (30.2%), *Phaeobacter* (29.4%), *Roseobacter* (13.9%), *Silicibacter* (13.7%) and *Nautella* (3.6%) (698 hits in total). Regarding the 30 hits to sequences from members of the species, the average identity within HSPs (high-scoring segment pairs) was 99.6%, whereas

the average coverage by HSPs was 18.7%. Regarding the 20 hits to sequences from other members of the genus, the average identity within HSPs was 98.0%, whereas the average coverage by HSPs was 18.7%. Among all other species, the one yielding the highest score was *P. inhibens* (AY177712), which corresponded to a 16S rDNA gene identity of 99.5% and an HSP coverage of 18.6%. (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 AJ296158 (Greengenes short name 'Spain:Galicia isolate str. PP-154'), which showed an identity of 99.8% and an HSP coverage of 18.7%. The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'microbi' (2.8%), 'marin' (2.5%), 'coral' (2.4%), 'sediment' (2.0%) and 'biofilm' (1.9%) (509 hits in total). Environmental samples which yielded hits of a higher score than the highest scoring species were not found.



**Figure 1.** Phylogenetic tree highlighting the position of *P. gallaeciensis* relative to the type strains of the other species within the genus *Phaeobacter* and the neighboring genus *Leisingera*. The tree was inferred from 1,381 aligned characters of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion as previously described [9]. *Ruegeria* spp. were included in the dataset for use as outgroup taxa. 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% [9]. Lineages with type strain genome sequencing projects registered in GOLD [10] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks. Genome sequences are available for *P. arcticus* (DQ514304) [11], *P. inhibens* (AY177712) [12], *P. caeruleus* (AM943630) [13], *P. daeponensis* (DQ981416) [14], *P. gallaeciensis* (IMG2545691711, this publication), *L. aquimarina* (AM900415) [15], *L. methylohalidivorans* (AY005463) [16] and *R. pomeroyi* (AF098491) [17].

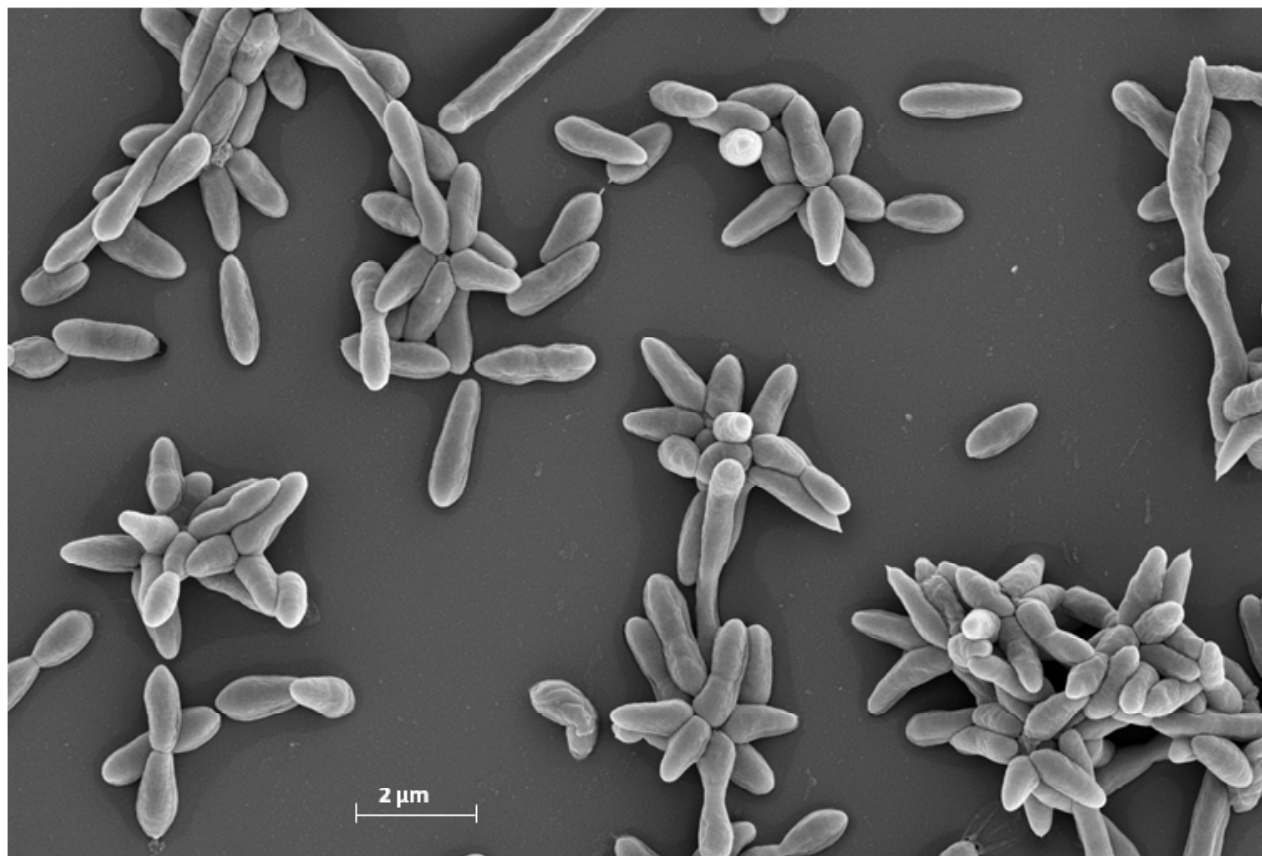
## Morphology and physiology

Cells of BS 107<sup>T</sup> stain Gram-negative and are ovoid-shaped rods ranging 0.7-1.0 µm in width and 1.7-2.5 µm in length. Motility is achieved by means of a polar flagellum (not visible in Figure 2). Young colonies grown on Marine Broth (MB) at 23°C are 0.5 mm in diameter, circular, smooth, convex and brownish with regular edges [1]. Colonies incubated for 7 days are 2 mm in diameter with irregular edges and produce a brown, diffusible pigment. Cells grow at temperatures between 15 and 37°C; optimal growth was observed in a range between 23 and 27°C. The optimal pH is 7.0, with growth occurring up to pH 10.0 but none below pH 4.0. Cells grow at salt concentrations ranging from 0.1 to 2.0 M NaCl, with 0.2 M being the optimal concentration. Additional thiamine (vitamin B<sub>2</sub>) is required for growth in minimal medium. Cells exhibit catalase and oxidase activity, but they do not exhibit amylase, gelatinase, β-galactosidase, tweenase, DNase, urease, arginine dihydrolase, lysine decarboxylase and ornithine decarboxylase activities [1].

BS107<sup>T</sup> is able to use the following substrates as sole carbon source and energy source: D- man-

nose, D-galactose, D-fructose, D-glucose, D-xylose, melibiose, trehalose, maltose, cellobiose, sucrose, meso-erythritol, D-mannitol, glycerol, D-sorbitol, meso-inositol, succinate, propionate, butyrate, γ-aminobutyrate, DL-hydroxybutyrate, 2-ketoglutarate, pyruvate, fumarate, glycine, L-*a*-alanine, p-alanine, L-glutamate, L-lysine, L-arginine, L-ornithine, L-proline, acetate and leucine. Bacteriochlorophyll *a* was not detected [1].

The metabolic properties of *Phaeobacter gallaeciensis* CIP 105210<sup>T</sup> and the *P. inhibens* strains DSM 17395, DSM 24588 (= 2.10) and DSM 16374<sup>T</sup> (= T5<sup>T</sup>) were compared using the more sensitive Phenotype MicroArray (PM) technology [3]. Here, using the statistical analysis (clustering and discretization) approaches as implemented in "opm" [18,19], the non-identity of strains CIP 105210<sup>T</sup> and DSM 17395 could be demonstrated despite an overall similar physiology. Differences could be found regarding the respiration of tyramine, which was positive in DSM 17395 and negative in CIP 105210<sup>T</sup>, and for butyrate, for which respiration was found to be negative in DSM 17395 and positive in CIP 105210<sup>T</sup> [3]. A summary of the classification and features of CIP 105210<sup>T</sup> is presented in Table 1.



**Figure 2.** Scanning electron micrograph of *P. gallaeciensis* CIP 105210T.

## Chemotaxonomy

The chemical composition of strain BS107<sup>T</sup> confirmed ubiquinones as the sole respiratory lipoquinones and revealed Q10 as predominant. Polar lipids consisted of an unidentified phospholipid, two uncharacterized lipids, aminolipids, phosphatidylenthanolamine,

phosphatidylglycerole and phosphatidylcholine [2].

The major fatty acids are the monounsaturated acids C<sub>18:1 ω7c</sub> (76.1%), and 11-methyl C<sub>18:1 ω7c</sub> (6.1%), followed by hydroxy fatty acid C<sub>16:0 2-OH</sub> (5.1%) as well as C<sub>16:0</sub> (4.0%), C<sub>14:1</sub> (3.1%), C<sub>18:0</sub> (2.6%), C<sub>10:0 3-OH</sub> (2.2%) and C<sub>18:1 ω9c</sub> (0.9%) [2].

**Table 1.** Classification and general features of *P. gallaeciensis* BS107<sup>T</sup> according to the MIGS recommendations [20] published by the Genome Standards Consortium [21].

MIGS ID	Property	Term	Evidence code <sup>a</sup>
		Domain <i>Bacteria</i>	TAS [22]
		Phylum <i>Proteobacteria</i>	TAS [23]
		Class <i>Alphaproteobacteria</i>	TAS [24,25]
	Current classification	Order <i>Rhodobacterales</i>	TAS [25,26]
		Family <i>Rhodobacteraceae</i>	TAS [25,27]
		Genus <i>Phaeobacter</i>	TAS [1,28]
		Species <i>Phaeobacter gallaeciensis</i>	TAS [1]
		Subspecific genetic lineage (strain) BS107 <sup>T</sup>	TAS [1]
MIGS-12	Reference for biomaterial	Ruiz-Ponte et al. 1998	TAS [1]
	Gram stain	Gram-negative	TAS [1]
	Cell shape	ovoid-rod-shaped	TAS [1]
	Motility	motile, via polar flagella	TAS [1]
	Sporulation	not reported	
	Temperature range	15-37°C, mesophile	TAS [1]
	Optimum temperature	23-27°C	TAS [1]
	Salinity	0.1-2.0 M NaCl	TAS [1]
MIGS-22	Relationship to oxygen	aerobe	TAS [1]
	Carbon source	complex substrates, butyrate, DL-hydroxybutyrate, D-xylose	TAS [1]
	Energy metabolism	chemoheterotrophic	TAS [1]
MIGS-6	Habitat	seawater, <i>Pecten maximus</i>	TAS [1]
MIGS-6.2	pH	4.0-10.0, optimum 7.0	TAS [1]
MIGS-15	Biotic relationship	free living, facultative symbiont	TAS [1]
MIGS-14	Known pathogenicity	-	IDA
MIGS-16	Specific host	<i>Pecten maximus</i>	
MIGS-18	Health status of host	not reported	
	Biosafety level	1	TAS [29]
MIGS-19	Trophic level	heterotroph	TAS [1]
MIGS-23.1	Isolation	seawater of larval cultures of the scallop <i>Pecten maximus</i>	TAS [1]
MIGS-4	Geographic location	A Coruna, Galicia, Spain	TAS [1]
MIGS-5	Time of sample collection	not reported	
MIGS-4.1	Latitude	43.3619	
MIGS-4.2	Longitude	-8.410	
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	about sea level	

<sup>a</sup>Evidence codes - IDA: Inferred from Direct Assay; 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 [30].

## Genome sequencing and annotation

**Table 2.** Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	One draft assembly of standard shotgun library, one 3 kbp paired-end library
MIGS-29	Sequencing platforms	Roche/454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	22 ×
MIGS-30	Assemblers	Newbler assembler version 2.6 (Software Release: 2.6 (20110517_1502)
MIGS-32	Gene calling method	Prodigal 1.4
	INSDC ID	CP002976.1
	GenBank Date of Release	1/31/2014
	GOLD ID	Gi24053
	NCBI project ID	188096
	Database: IMG	2531839720
MIGS-13	Source material identifier	CIP 105210 <sup>T</sup>
	Project relevance	Tree of Life, carbon cycle, scallop rearing, plasmid

### Growth conditions and DNA extraction

A culture of CIP 105210<sup>T</sup> was grown aerobically in 100 ml of DSMZ medium 514 [31] on a shaker at 28°C. Genomic DNA was isolated using the Qiagen Genomic DNA Kit, following the standard protocol for Bacteria 500G provided by the manufacturer. The extracted DNA had a concentration of 200 ng/μl. The quality of the DNA was checked with the NanoDrop.

### Genome sequencing and assembly

The genome of *P. gallaeciensis* CIP 105210<sup>T</sup> was sequenced using the Roche/454 GS FLX Titanium sequencing platform [Table 2]. A draft assembly based on 247,768 reads of a standard shotgun library and 204,863 reads of a 3 kbp paired-end library (LGC Genomics, Berlin, Germany) with a total of 138 Mb (22-fold coverage) was generated with Newbler assembler, Roche Diagnostics GmbH, Mannheim, Germany). This assembly consisted of 45 contigs 26 of which could be joined into 15 scaffolds. Gaps resulting from repetitive sequences were closed by PCR followed by Sanger sequencing, yielding a final genome size of 4,540,155 bp, that consists of one circular chro-

mosome of 3,776,653 bp and seven circular plasmids.

### Genome annotation

Genes were identified using Prodigal [32] as part of the Integrated Microbial Genomes Expert Review (IMG/ER) annotation pipeline [33]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases.

### Genome properties

The *Phaeobacter gallaeciensis* CIP 105210<sup>T</sup> genome statistics are provided in Table 3 and Figures 3a, 3b, 3c, 3d, 3e, 3f, 3g, 3h. The genome consists of eight circular replicons with a total length of 4,540,155 bp and a G+C content of 59.44%. The replicons correspond to a single chromosome (3,776,653 bp) and seven extrachromosomal elements ranging in size between 255,493 bp and 40,170 bp. From a total of 4,448 predicted genes, 4,369 were protein coding genes and 79 RNA genes. The distribution of genes into COGs functional categories is presented in Table 4.



**Table 3.** Genome statistics

Attribute	Value	% of total <sup>a</sup>
Genome size (bp)	4,540,155	100.00
DNA coding region (bp)	4,056,108	89.34
DNA G+C content (bp)	2,698,552	59.44
Number of replicons	8	
Extrachromosomal elements	7	
Total genes	4,448	100.00
RNA genes	79	1.78
rRNA operons	4	
tRNA genes	59	1.33
Protein-coding genes	4,369	98.22
Genes with function prediction	3,595	80.82
Genes in paralog clusters	3,475	78.13
Genes assigned to COGs	3,422	76.93
Genes assigned Pfam domains	3,657	82.22
Genes with signal peptides	457	10.27
Genes with transmembrane helices	975	21.92
CRISPR repeats	0	

<sup>a</sup>The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome

## Insights into the genome

### Unique genes

A search for specific genes in the genome of *P. gallaeciensis* CIP 105210<sup>T</sup> compared to the *P. inhibens* strains DSM 24588 (= 2.10), DSM 16374<sup>T</sup> (= T5<sup>T</sup>) and DSM 17395, based on an e-value of 1e-5 and a minimum identity of 30%, resulted in a total number of 551 specific genes. 296 (54%) of these genes were located on the chromosome and 255 (46%) on extrachromosomal replicons. In comparison with the other completely sequenced bacterial strains of the genus *Phaeobacter*, 8% of the chromosomal and 35% of the extrachromosomal *P. gallaeciensis* CIP 105210<sup>T</sup> genes were unique, thus reflecting the considerable contribution of extrachromosomal elements to unique gene content.

The observed distribution may be influenced by the presence of two chromosome-encoded bacterial MobC mobilization proteins (Gal\_00154, Gal\_01073). MobC, which is missing in all three completely sequenced *P. inhibens* strains, is part of the relaxosome at the origin of transfer and in-

creases the frequency of plasmid mobilization and therefore conjugal transfer of plasmids [34], which is also in agreement with the comparably large number of seven extrachromosomal replicons present in CIP 105210<sup>T</sup>.

The probable function of some of the unique genes is explained below. Genes Gal\_01405 and Gal\_01407 constitute methane monooxygenases (EC 1.14.13.25) facilitating the degradation of aromatic compounds and phenols [35]. Gal\_01397, a monoamine oxidase could provide an additional source of ammonium [36].

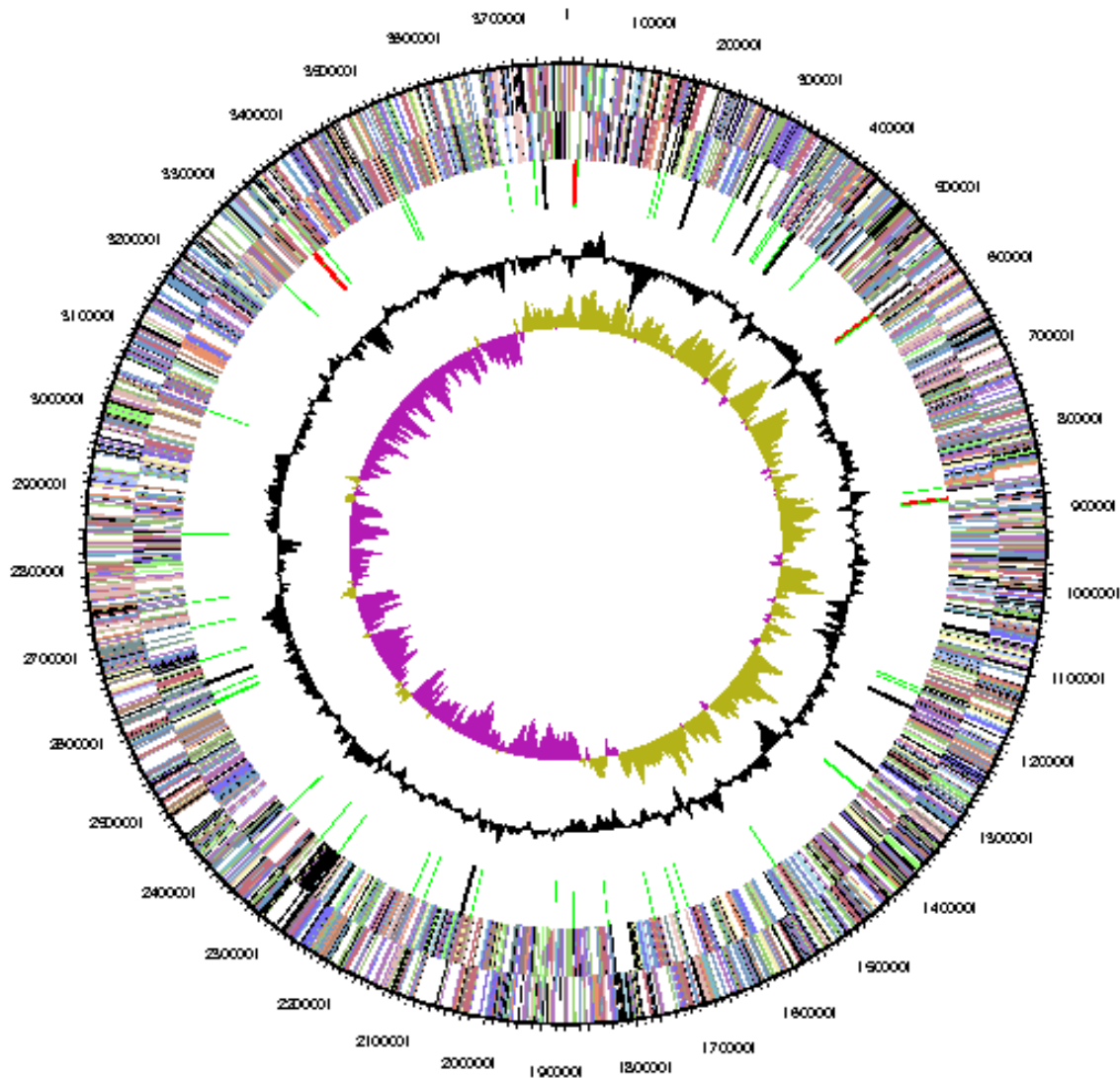
Unique genes are also provided by phage-like elements. In CIP 105210<sup>T</sup> these so-called “morons” (because they add “more on” the genome [37]) comprise, e.g., an ABC-2 family drug transporter (Gal\_01752) [38], and a negative regulator of beta-lactamase expression (Gal\_02239).

### Genomic islands

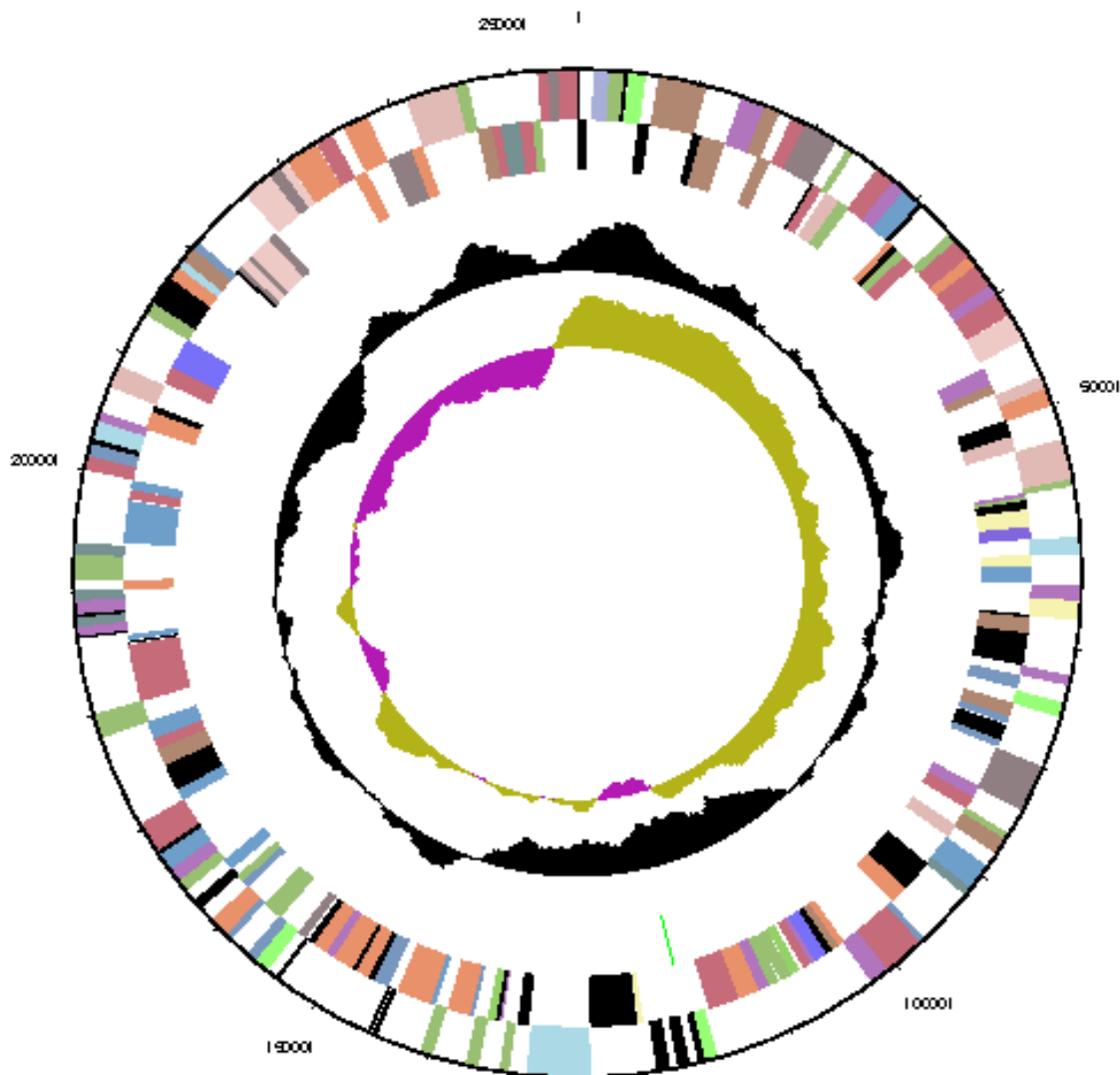
Six genomic islands could be identified on the chromosome with the web-based island-viewer system [39]. Island-viewer combines the methods IslandPick [40], which uses a comparative genomics approach, SIGI-HMM [41], which relies

upon deviating codon usage signatures, and IslandPath-DIMOB [42], which identifies genomic islands based on deviating GC content, dinucleotide bias in gene clusters and the presence of island specific genes like mobility genes and tRNAs. Island-I ranging from position 155,977 to 177,667 (21,690 bp) contains a tRNA gene (Phe GAA, Gal\_00137) next to a site-specific recombinase XerD (Gal\_00138) and the bacterial mobilization protein (MobC, Gal\_00154; see above). Further-

more, it contains a transcriptional regulator of the LysR family (Gal\_00160) and an adjacent ABC-type transport system for glycine/proline/betaine. Island-II (422,441 to 434,165; 11,725 bp) mainly consists of hypothetical proteins, but it also contains a large type II restriction enzyme (905aa, Gal\_00442) and another site specific XerD recombinase (Gal\_00444) next to a tRNA for proline (Gal\_00445). Island-III (1,085,143 to 1,096,105; 10,962 bp)



**Figure 3a.** Circular graphical map of the chromosome. From margin to center: 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, GC skew.



**Figure 3b.** Circular graphical map of the extrachromosomal replicon pGal\_A255. From margin to center: 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, GC skew.

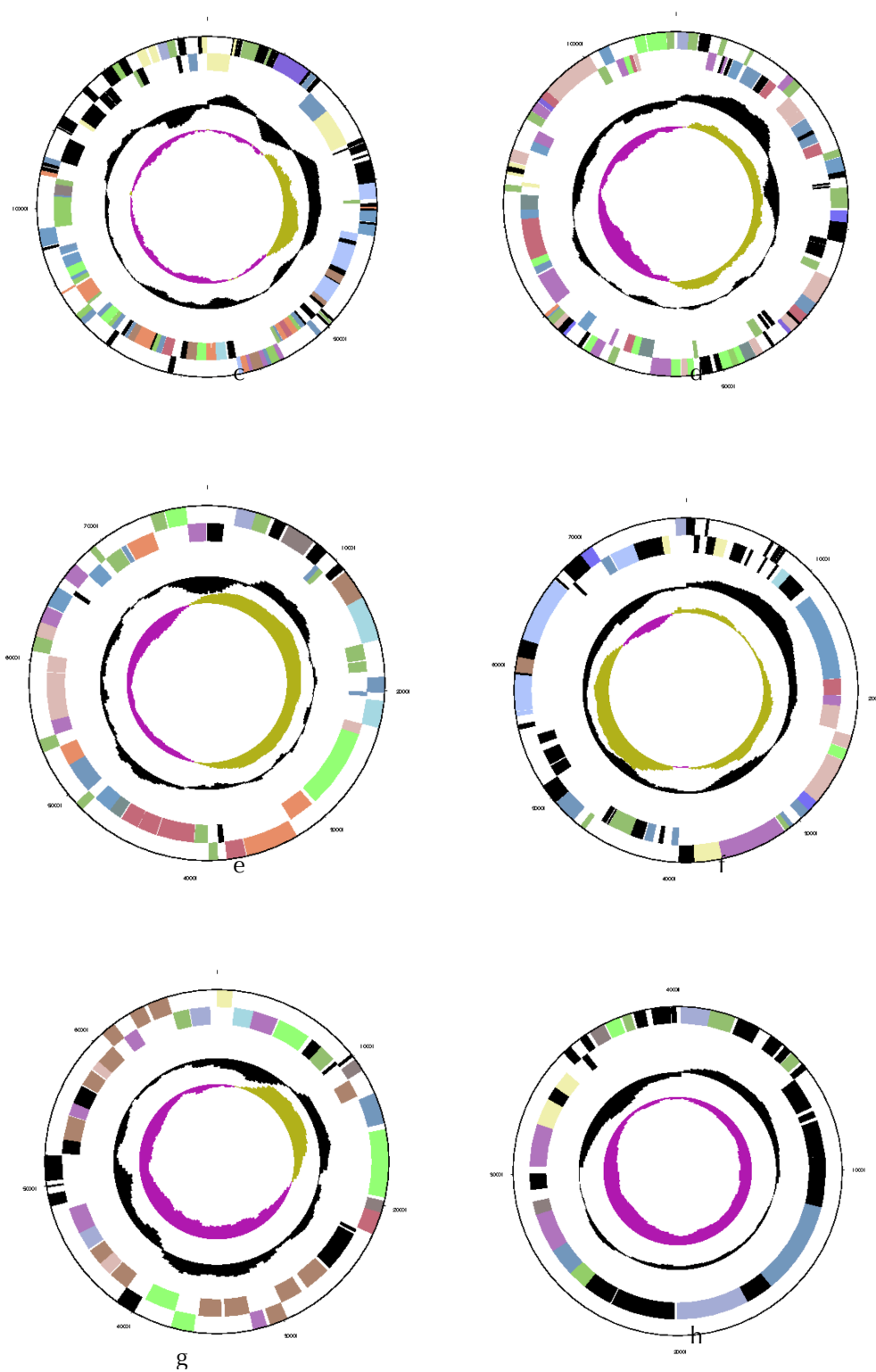
contains three XerD recombinases in row (Gal\_01065 to Gal\_01067, a MobC protein (Gal\_01073) and the typical VirD2 relaxase (Gal\_01074) as well as the VirD4 coupling protein (Gal\_01075) of type IV secretion systems [43] indicating a plasmid-derived origin of this island. Island-IV (1,626,663 to 1,641,677; 15,014 bp) contains an ABC-type cobalt transport system and a XerC recombinase (Gal\_01616). Island-V (2,821,359 to 2,848,860; 27,501 bp) consists mainly of regulated TRAP C4-dicarboxylate and ABC-type dipeptide/oligopeptide/nickel transport proteins and also the epsilon subunit of DNA pol-

ymerase III (Gal\_02817). Island-VI (3,328,870 to 3,344,910; 16,040 bp) lies adjacent to a ribosomal rRNA-operon and contains an ABC-type amino acid/amide transport system and an E1 component of the pyruvate dehydrogenase complex (Gal\_03286, E.C.: 1.2.4.1).

#### Phage-like elements

The presence of phage-like elements was analyzed with the online tool PHAST [44]. The program identified 16 genes representing a gene transfer agent (GTA [45];) and three incomplete clusters of phage-derived genes with sizes between 15 kb and 40 kb (Table 5).





**Figure 3c-h.** Circular graphical map of the extrachromosomal replicons: (c) pGal\_B134, (d) pGal\_C110, (e) pGal\_D78, (f) pGal\_E78, (g) pGal\_F69, and (h) pGal\_G40. From margin to center: 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, GC skew.

**Table 4.** Number of genes associated with the 25 general COG functional categories

Code	Value	%age <sup>a</sup>	Description
J	170	4.51	Translation, ribosomal structure and biogenesis
A	1	0.03	RNA processing and modification
K	310	8.23	Transcription
L	150	3.98	Replication, recombination and repair
B	3	0.08	Chromatin structure and dynamics
D	35	0.93	Cell cycle control, cell division, chromosome partitioning
Y	0	0	Nuclear structure
V	50	1.33	Defense mechanisms
T	163	4.33	Signal transduction mechanisms
M	204	5.41	Cell wall/membrane/envelope biogenesis
N	53	1.41	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0	Extracellular structures
U	87	2.31	Intracellular trafficking, secretion, and vesicular transport
O	146	3.87	Posttranslational modification, protein turnover, chaperones
C	235	6.24	Energy production and conversion
G	212	5.63	Carbohydrate transport and metabolism
E	436	11.57	Amino acid transport and metabolism
F	81	2.15	Nucleotide transport and metabolism
H	160	4.28	Coenzyme transport and metabolism
I	151	4.01	Lipid transport and metabolism
P	205	5.44	Inorganic ion transport and metabolism
Q	132	3.50	Secondary metabolites biosynthesis, transport and catabolism
R	451	12.00	General function prediction only
S	333	9.00	Function unknown
-	1,026	23.07	Not in COGs

<sup>a</sup>The total is based on the total number of protein coding genes in the annotated genome.

**Table 5.** Prophage regions in the chromosome of *P. gallaeciensis* CIP 105210<sup>Ta</sup>

Region	Length	Completeness	Score	CDS	Coordinates	Specific keyword	GC%
1	14.2 kb	Questionable	80	16	1,566,488-1,580,693	Gene transfer agent (GTA)	64.6
2	25.1 kb	Incomplete	30	32	1,781,279-1,806,383	integrase, region invertase, helicase	56.9
3	14.7 kb	Incomplete	40	18	1,800,767-1,815,474	Portal protein, head maturation protease	57.7
4	39.6 kb	Incomplete	60	45	2,265,763-2,305,412	Integrase, peptidoglycan hydrolase	58.5

<sup>a</sup>Completeness, a prediction of whether the region contains an intact or incomplete prophage based on the applied criteria of PHAST; Score, the score of the region based on the applied criteria of PHAST; CDS, the number of coding sequences; Coordinates, the start and end positions of the region on the bacterial chromosome; GC%, the percentage of GC nucleotides of the region.

### Extrachromosomal replicons

Complete genome sequencing of *Phaeobacter gallaeciensis* CIP 105210<sup>T</sup> resulted in eight replicons ranging from 40 kb to 3.8 MB in size. For the seven extrachromosomal replicons, ranging in size between 40 kb and 255 kb (Table 6), circular confirmation has been experimentally validated. The

extrachromosomal replicons were analyzed as described in [46] and [47]. They contain characteristic replication modules [43] of the RepABC-, DnaA-like, RepA- and RepB-type comprising a replicase and a parAB partitioning operon [48]. Plasmid pGal\_E78 also contains a replicase that is homologous to those of RepABC-type plasmids, but the partitioning genes repAB are missing. The solitary

replicase cannot be classified according to the established scheme [49] and is designated as RepC\_soli-1a (RepC' [50]). The respective replicases of the other extrachromosomal replicons that mediate the initiation of replication are

designated according to the established classification scheme [51]. The numbering of specific replicases corresponds to plasmid compatibility groups that are required for a stable coexistence of the replicons within the same cell [49].

**Table 6.** General genomic features of the chromosome and extrachromosomal elements from *P. gallaeciensis* strain CIP 105210<sup>Ta</sup>

Replicon	No.	Replicase	Length (bp)	GC (%)	Topology	No. Genes <sup>#</sup>
Chromosome	1	DnaA	3,776,653	60	circular	3,703
pGal_A255	2	DnaA-like-I	255,493	58	circular	237
pGal_B134	3	RepABC-5	133,631	60	circular	155
pGal_C110	4	RepABC-8	109,815	56	circular	115
pGal_D78	5	RepB-I	77,876	62	circular	62
pGal_E78	6	RepC_soli-1a	77,775	55	circular	81
pGal_F69	7	RepA-I	68,752	58	circular	56
pGal_G40	8	RepABC-4	40,170	56	circular	51

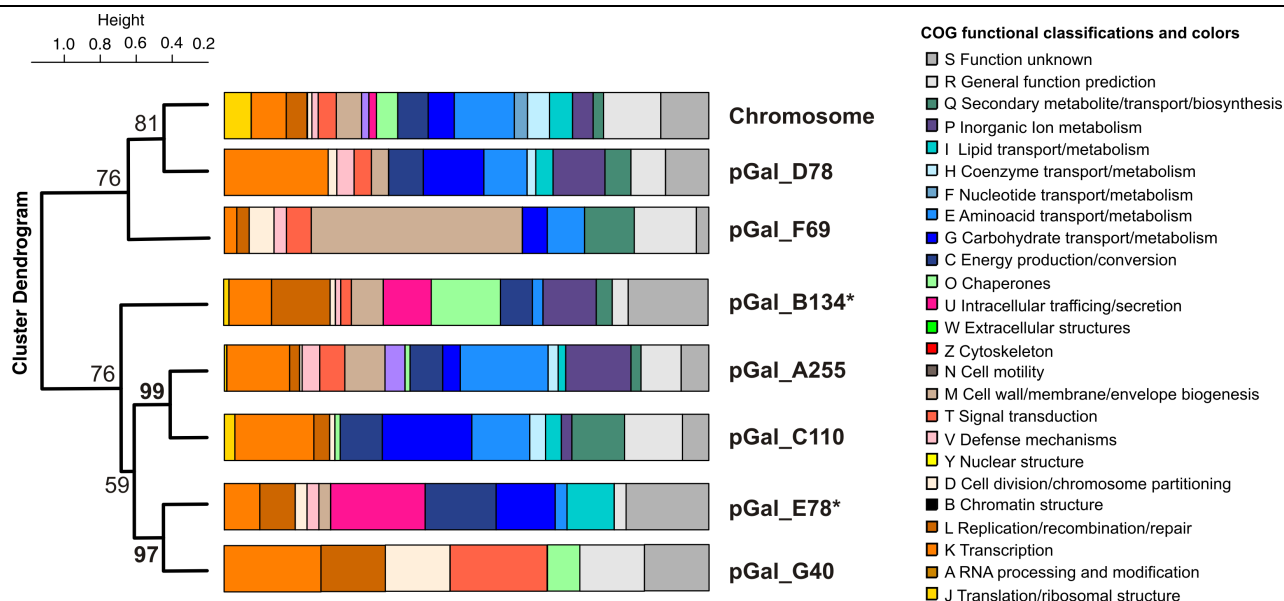
<sup>a</sup>deduced from automatic annotation.

The comparison of the extrachromosomal replicons from *P. gallaeciensis* CIP 105210<sup>T</sup> and *P. inhibens* DSM 17395 documents a strong conservation and long-range synteny of three replicons. The largest 255 kb DnaA-like-I replicon (pGal\_A255) is slightly smaller than the 262 kb equivalent (NC\_018291.1), sharing 89% identity on nucleotide level. The RepB-I type replicon pGal\_D78 exactly matches the size of the DSM 17395 replicon (NC\_018287.1, 91% identity), whereas the RepA-I type replicon pGal\_F69 is

slightly larger than its equivalent (65 kb; NC\_018288.1, 91% identity). On the contrary, RepABC-type replicons are not present in the DSM 17395 genome. However, only two of the four additional plasmids, the RepABC-5 type replicon pGal\_B134 and the RepC\_soli-1a-type replicon pGal\_E78 possess type IV secretion systems that are required for conjugative transfer [52]. Finally, the three replicons pGal\_A255, pGal\_B134, pGal\_C110 are equipped with stabilizing toxin/antitoxin modules [53] (Table 7).

**Table 7.** Integrated Microbial Genome (IMG) locus tags of *P. gallaeciensis* CIP 105210<sup>T</sup> genes for the initiation of replication, toxin/antitoxin modules and type IV secretion systems (T4SS) required for conjugation.

Replicon	Replication Initiation		Plasmid Stability		Type IV Secretion	
	Replicase	Locus Tag	Toxin	Antitoxin	VirB4	VirD4
Chromosome	DnaA	Gal_00001	-	-	-	-
pGal_A255	DnaA-like-I	Gal_03722	Gal_03770	Gal_03771	-	-
pGal_B134	RepABC-5	Gal_03960	Gal_03975	Gal_03974	Gal_04010	Gal_03992
pGal_C110	RepABC-8	Gal_04107	Gal_04110	Gal_04111	-	-
pGal_D78	RepB-I	Gal_04221	-	-	-	-
pGal_E78	RepC_soli-1a	Gal_04283	-	-	Gal_04360	Gal_04345
pGal_F69	RepA-I	Gal_04364	-	-	-	-
pGal_G40	RepABC-4	Gal_04417	-	-	-	-



**Figure 4.** Summary and representation of the assortment of replicons in CIP 105210<sup>T</sup> as previously described in [47]. Bars show the relative frequency of functional classes according to the database of clusters of orthologous groups of proteins (COGs). The cluster dendrogram arranges the replicons according to their overall codon usage. Codon-usage matrices were generated using yet unpublished scripts for the statistical analysis software R (version 2.15.0.) [54]. The hierarchical clustering analysis was conducted using the pvclust function [55] with “complete” as agglomeration method. Pvclust also returns the AU (Approximately Unbiased) values as statistical support for clusters in percent. Support values >95% are given in bold. Asterisks indicate the presence of genes for conjugation (T4SS) as listed in Table 7.

The 255 kb DnaA-like-I replicon pGal\_A255 is largely constituted by genes coding for proteins in COG E “amino-acid transport and metabolism” and COG P “inorganic ion metabolism” (Figure 4). The latter category comprises, for example, a Fe<sup>3+</sup> siderophore complex (Gal\_03846 to Gal\_03848), which contains ferric-iron chelating agents that facilitate enhanced uptake of this essential compound [56]. pGal\_A255 furthermore harbors six genes involved in chemotaxis, a tRNA (Gal\_03828) and a cluster for the biosynthesis of coenzyme PQQ, a redox factor (Gal\_03896). The genes for the synthesis of the antibiotic tropodithietic acid (TDA) [57] are consolidated in a cluster on pGal\_A255 and comprise *tdaA* (Gal\_03819), *tdaB* (Gal\_03818), *tdaC* (Gal\_03817), *tdaE* (Gal\_03815) and *tdaF* (Gal\_03802). The 134 kb RepABC-5 type plasmid pGal\_B134 harbors in comparison to the other seven replicons the most chaperones (COG O, Figure 4), owing to an elevated presence of cytochromes and disulfide bond formation proteins. pGal\_B134 also holds a dimethyladenosine-transferase (Gal\_03978) that facilitates RNA methylation and a T4S system (Table 7), thus combining on this plasmid genes for epigenetic modifications. The RepABC-8 type plasmid pGal\_C110 consists mainly of amino acid and car-

bohydrate transporters (COGs E and G) and biogenesis of secondary metabolites (COG Q). COG K, transcription is also elevated, due to the presence of 15 transcriptional regulators. On the RepB-I replicon pGal\_D78, COG K transcription is elevated, owing to the presence of twelve transcriptional regulators and a RNA-polymerase (Gal\_04277). This replicon also contains genes for siderophore synthetases (Gal\_04241 to Gal\_04247) and a catalase/peroxidase (Gal\_04279). On the RepC\_soli-1a plasmid pGal\_E78, proteins of COG C energy production and conversion are constituted by pyruvate dehydrogenase E1 and E2 components, which play a role in the citrate cycle and gluconeogenesis. The RepA-I replicon pGal\_F69 contains an RTX toxin [58] (Gal\_04412) and exhibits a strong accumulation of COG M, “cell-envelope biogenesis”. It harbors several polysaccharide export proteins including a type I secretion system ABC transporter (Gal\_04381, Gal\_04382), and a complete rhamnose operon [59]. *P. gallaeciensis* CIP 105210<sup>T</sup> (= DSM 26440<sup>T</sup>) forms strong biofilms (unpublished results) and the extrachromosomal 69 kb replicon seems to be responsible for the attached lifestyle as previously proposed for the *P. inhibens* strains DSM 17395 and DSM 24588 (2.10) [3]. pGal\_G40 represents a hybrid between

a plasmid and a circular phage, comparable to the coliphage N15 [60,61]. It contains an N-acyl-L-homoserine lactone synthetase (Gal\_04460) and a complete *repABC* operon. This interesting finding draws a direct connection between RepABC directed replication [49], horizontal gene transfer and AHL-mediated quorum sensing [62].

Genome sequencing of *P. inhibens* DSM 16374<sup>T</sup> (T5<sup>T</sup>) revealed the presence of the complete dissimilatory nitrate reduction pathway and anaerobic growth on nitrite has been validated experimentally [12]. The genes of the pathway are located on three different replicons, i.e. the chromosome, the DnaA-like I type plasmid pInhi\_A227 and the RepABC-8 type plasmid pInhi\_B88. The genome of the sister species *P. gallaeciensis* CIP 105210<sup>T</sup> exhibits a conspicuous synteny for the chromosome and three extrachromosomal replicons (DnaA-like I (pGal\_A255, pInhi\_A227), RepB-I (pGal\_D78, pInhi\_C78), RepA-I (pGal\_F69, pInhi\_D69)). However, the RepABC-8 type plasmid including the crucial nitrous oxide reductase (EC 1.7.2.4) is missing in *P. gallaeciensis* CIP 105210<sup>T</sup>, and this strain is accordingly unable to grow anaerobically.

### Phylogenomic analyses

The phylogenetic analysis of 16S rRNA gene type-strain sequences places *P. gallaeciensis* together with both *P. caeruleus* and *P. daeponensis*, whereas

*P. inhibens* forms a cluster with *P. leonis* and *P. arcticus*. Both clusters are set apart from each other, but the 16S rRNA gene tree is unresolved and does not allow one to infer the evolutionary inter-relationships in this group. Previous results [4] showed that the reported *P. gallaeciensis* type-strain deposit DSM 17395 belongs to *P. inhibens* and that CIP 105210<sup>T</sup> (= DSM 26640<sup>T</sup>) is the authentic type strain of *P. gallaeciensis*. Moreover, the genome sequenced strain ANG1 has been referred to as *P. gallaeciensis* based on 16S rRNA analyses [63], but our recent study revealed a well-supported association with *P. caeruleus* and *P. daeponensis* [4]. The relationships between these *Phaeobacter* strains have not been corroborated using genome sequences. Thus, we used the Genome-to-Genome Distance Calculator (GGDC) [64] to investigate the affiliation of strain ANG1 and the genomic similarities between *P. inhibens* and *P. gallaeciensis* strains from available genome sequences and conducted phylogenomic analyses to address the relationship between *P. gallaeciensis* and *P. inhibens*.

Table 8 shows the results of the calculated digital DNA-DNA hybridization (DDH) similarities of *P. gallaeciensis* CIP 105210<sup>T</sup> and *P. inhibens* DSM 16374<sup>T</sup> (T5<sup>T</sup>) to other *Phaeobacter* strains. For DDH values <math>\geq 70\%</math> the respective query strain would be considered as belonging to a different species than the strain used as a reference [65,66].

**Table 8.** DDH similarities with standard deviations between *P. gallaeciensis* CIP 105210<sup>T</sup>, *P. inhibens* DSM 16374<sup>T</sup> (T5<sup>T</sup>) and other *Phaeobacter* strains calculated in silico with the GGDC server version 2.0 [64]. The numbers in parentheses are IMG Taxon IDs identifying the genome sequence.

Formula reference species	identities/HSP length [%] <i>P. gallaeciensis</i> DSM 26640 <sup>T</sup> (= CIP 105210 <sup>T</sup> = BS107 <sup>T</sup> )	identities/HSP length [%] <i>P. inhibens</i> DSM 16374 <sup>T</sup> (T5 <sup>T</sup> )
<i>P. inhibens</i> DSM 24588 (2.10) (2501651220)	38.00% ± 2.49	79.50% ± 2.80
<i>P. inhibens</i> DSM 17395 (2510065029)	38.40% ± 2.50	78.70% ± 2.83
<i>P. gallaeciensis</i> ANG 1 (2526164696)	21.40% ± 2.34	21.10% ± 2.33
<i>P. inhibens</i> DSM 16374 <sup>T</sup> (T5 <sup>T</sup> ) (2516653078)	38.20% ± 2.50	100%
<i>P. gallaeciensis</i> DSM 26640 <sup>T</sup> (= CIP 105210 <sup>T</sup> ) (2545555837)	100%	38.20% ± 2.50

With the exception of *P. gallaeciensis* ANG 1, which neither belongs to *P. gallaeciensis* nor *P. inhibens* based on DDH values, the analysis supports the current classification. *P. inhibens* with the type strain DSM 16374<sup>T</sup> (T5<sup>T</sup>) includes the strains DSM

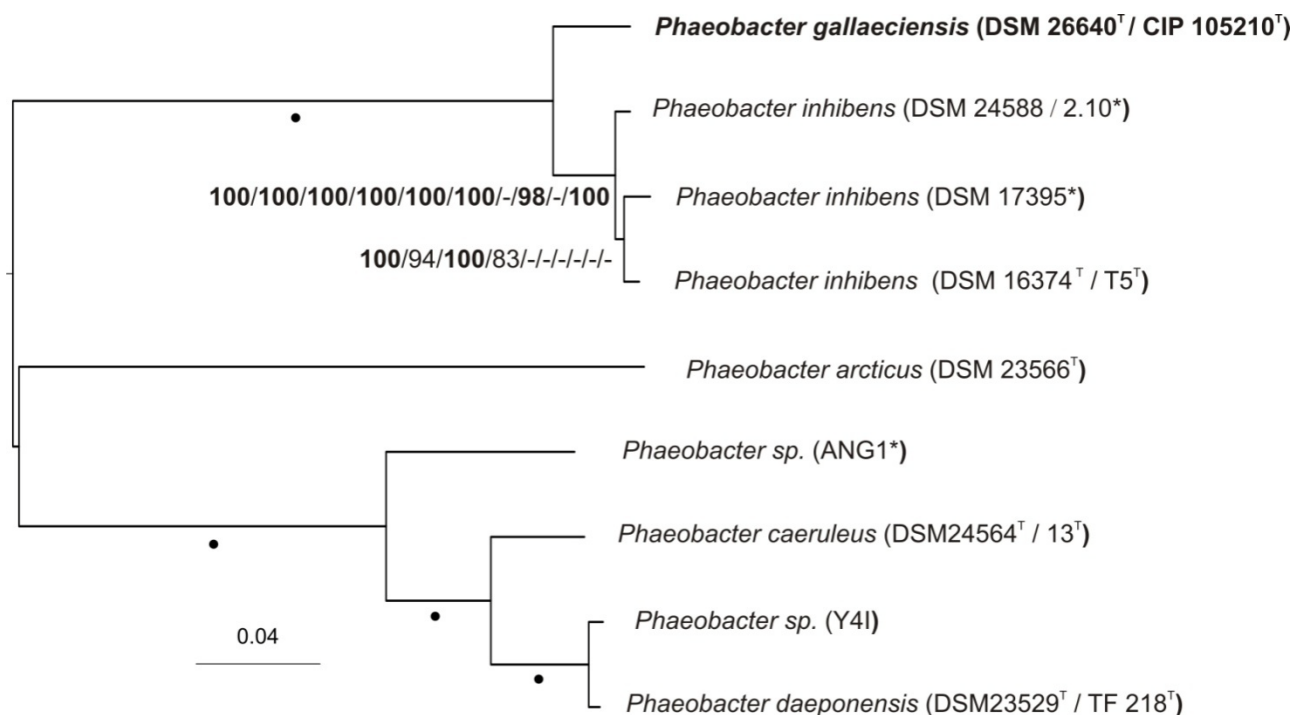
17395 and DSM 24588 (2.10), whereas the strain *P. gallaeciensis* CIP 105210<sup>T</sup> (= DSM 26640<sup>T</sup>) is the sole representative of *P. gallaeciensis* analyzed in the current study.



For the phylogenomic analysis, protein sequences from the available *Phaeobacter* genomes were retrieved from the IMG website (*P. arcticus* DSM 23566<sup>T</sup>; ID 2516653081; *P. caeruleus* DSM 24564<sup>T</sup> (13<sup>T</sup>), ID 2512047087; *P. daeponensis* DSM 23529<sup>T</sup> (TF-218<sup>T</sup>), ID 2516493020; *P. inhibens* DSM 16374<sup>T</sup> (T5<sup>T</sup>), ID 2516653078) or from NCBI (*P. inhibens* DSM 24588 (2.10), CP002972 – CP002975; *P. sp.* ANG1, AFCF00000000; *P. gallaeciensis* CIP 105210<sup>T</sup> (= DSM 26640<sup>T</sup>), AOQA00000000; *P. inhibens* DSM 17395, CP002976 – CP002979; *P. sp.* Y4I, ABXF00000000).

These sequences were investigated using the DSMZ phylogenomics pipeline as previously described [67-70] using NCBI BLAST [71], TribeMCL [72], OrthoMCL [73], MUSCLE [74], RASCAL [75], GBLOCKS [68] and MARE [76] to generate gene- and ortholog-content matrices as well as concatenated alignments of distinct selections of genes.

Maximum likelihood (ML) [77] and maximum-parsimony (MP) [78,79] trees were inferred from the data matrices with RAxML [80,81] and PAUP\* [82], respectively, as previously described [68,70,72,83].



**Figure 5.** Phylogenetic tree inferred from the MARE-filtered supermatrix under the maximum likelihood (ML) criterion [77] and rooted using the midpoint rooting approach [84]. The branches are scaled in terms of the expected number of substitutions per site. Numbers above the branches (from left to right) are bootstrap support values [85] (if greater than 60%) from ML/MP MARE-filtered supermatrix; ML/MP unfiltered (full) supermatrix; ML/MP core-genes supermatrix; ML/MP gene-content matrix; ML/MP ortholog-content matrix. Values larger than 95% are shown in bold; dots indicate branches with maximum support under all settings. Genomes marked with stars have been renamed according to this study and [3].

The results of the phylogenomic analyses are shown in Figure 5. The “full” and MARE-filtered supermatrix trees were topologically identical and the tree of the latter analysis is shown in Figure 5 together with ML and MP bootstrap support values from all analyses if larger than 60%. The tree inferred from the core-gene matrix showed a distinct grouping within *Phaeobacter inhibens*, i.e. *P. inhibens* DSM 17395 as sister of the clade comprising *P. inhibens* DSM 16374<sup>T</sup> (T5<sup>T</sup>) and *P. inhibens* DSM 24588 (2.10). The topologies of both MP and

ML “full” and MARE-filtered supermatrix trees were identical, whereas the MP core-genes tree was topologically identical to the ML core-genes tree. Both gene-content and ortholog-content MP trees were topologically identical and showed *P. inhibens* DSM 16374<sup>T</sup> (T5<sup>T</sup>) as a sister taxon of *P. inhibens* DSM 24588 (2.10) and *P. inhibens* DSM 17395. Only the ML gene-content and ortholog-content trees deviated regarding the species boundaries, showing a clade comprising *P. inhibens* DSM 16374<sup>T</sup> (T5<sup>T</sup>) and *P. gallaeciensis* CIP

105210<sup>T</sup> (= DSM 26640<sup>T</sup>) as well as a clade comprising *P. inhibens* DSM 24588 (2.10) and *P. inhibens* DSM 17395.

Thus, the analyses supported the earlier conclusion [3] that DSM 17395 belongs to *P. inhibens*. The analyses also confirmed that *P. "gallaeciensis"* ANG1 belongs neither to *P. gallaeciensis* nor to *P. inhibens* and might therefore represent a novel, not yet named seventh species in the genus *Phaeobacter*. Further, the analysis confirms *P. gallaeciensis* CIP 105210<sup>T</sup> (= DSM 26640<sup>T</sup>) as the sole representative of the species *Phaeobacter gallaeciensis*.

## Acknowledgments

The authors gratefully acknowledge the assistance of Victoria Michael for growing *P. gallaeciensis* cultures and DNA extraction and quality control. The work conducted by members of the *Roseobacter* consortium was supported by the German Research Foundation (DFG) Transregio-SFB 51.

## References

1. Ruiz-Ponte C, Cilia V, Lambert C, Nicolas JL. *Roseobacter gallaeciensis* sp. nov., a new marine bacterium isolated from rearings and collectors of the scallop *Pecten maximus*. *Int J Syst Bacteriol* 1998; **48**:537-542. [PubMed](#) <http://dx.doi.org/10.1099/00207713-48-2-537>
2. 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., reclassification of *Ruegeria algicola* (Lafay et al. 1995) Uchino et al. 1999 as *Marinovum algicola*. *Int J Syst Evol Microbiol* 2006; **56**:1293-1304. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.63724-0>
3. Buddruhs N, Pradella S, Göker M, Päufer O, 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<sup>T</sup> and DSM 17395. *Int J Syst Evol Microbiol* 2013; **63**:4340-4349. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.053900-0>
4. 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 adaptation to surface life. *ISME J* 2012; **6**:2229-2244. [PubMed](#) <http://dx.doi.org/10.1038/ismej.2012.62>
5. Porsby CH, Nielsen KF, Gram L. *Phaeobacter* and *Ruegeria* species of the *Roseobacter* clade colonize separate niches in a Danish Turbot (*Scophthalmus maximus*)-rearing farm and antagonize *Vibrio anguillarum* under different growth conditions. *Appl Environ Microbiol* 2008; **74**:7356-7364. [PubMed](#) <http://dx.doi.org/10.1128/AEM.01738-08>
6. Fernandes N, Case RJ, Longford SR, Seyedsayamdost MR, Steinberg PD, Kjelleberg S, Thomas T. Genomes and virulence factors of novel bacterial pathogens causing bleaching disease in the marine red alga *Delisea pulchra*. *PLoS ONE* 2011; **6**:e27387. [PubMed](#) <http://dx.doi.org/10.1371/journal.pone.0027387>
7. Seyedsayamdost MR, Case RJ, Kolter R, Clardy J. The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nat Chem* 2011; **3**:331-335. [PubMed](#) <http://dx.doi.org/10.1038/nchem.1002>
8. Seyedsayamdost MR, Carr G, Kolter R, Clardy J. Roseobactin: small molecule modulators of an algal-bacterial symbiosis. *J Am Chem Soc* 2011; **133**:18343-18349. [PubMed](#) <http://dx.doi.org/10.1021/ja207172s>
9. 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<sup>T</sup>). *Stand Genomic Sci* 2011; **4**:63-71. [PubMed](#) <http://dx.doi.org/10.4056/sigs.1533840>
10. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010; **38**:D346-D354. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkp848>
11. Freese HM, Dalingault H, Petersen J, Pradella S, Davenport K, Teshima H, Chen A, Pati A, Ivanova N, Goodwin LA, et al. Genome sequence of the phage-gene rich marine *Phaeobacter arcticus* type strain DSM 23566T. *Stand Genomic Sci* 2013; **8**:450-464. [PubMed](#) <http://dx.doi.org/10.4056/sigs.383362>
12. Dogs M, Teshima H, Petersen J, Fiebig A, Chertkov O, Dalingault H, Chen A, Pati A, Goodwin LA, Chain P, et al. Genome sequence of *Phaeobacter inhibens* type strain (T5<sup>T</sup>), a secondary metabolite producing member of the marine *Roseobacter* clade, and emendation of the species description of *Phaeobacter inhibens*. *Stand Genomic Sci* 2013; **9**:142-159. [PubMed](#) <http://dx.doi.org/10.4056/sigs.4287962>
13. Beyersmann PG, Chertkov O, Petersen J, Fiebig A, Chen A, Pati A, Ivanova N, Lapidus A, Goodwin LA, Chain P, et al. Genome sequence of *Phaeobacter caeruleus* type strain (DSM 24564<sup>T</sup>), a surface-associated member of the marine

- Roseobacter* clade. *Stand Genomic Sci* 2013; **8**:403-419. [PubMed](#)  
<http://dx.doi.org/10.4056/sigs.3927626>
14. Dogs M, Teshima H, Petersen J, Fiebig A, Chertkov O, Dalingault H, Chen A, Pati A, Goodwin LA, Chain P, et al. Genome sequence of *Phaeobacter daeponensis* type strain (DSM 23529T), a facultatively anaerobic bacterium isolated from marine sediment, and emendation of *Phaeobacter daeponensis*. *Stand Genomic Sci* 2013; **9**:142-159. [PubMed](#)  
<http://dx.doi.org/10.4056/sigs.4287962>
  15. Riedel T, Teshima H, Petersen J, Fiebig A, Davenport K, Daligault H, Erkkila T, Gu W, Munk C, Xu Y, et al. Genome sequence of the *Leisingera aquimarina* type strain DSM 24565<sup>T</sup>, a member of the marine *Roseobacter* clade rich in extrachromosomal elements. *Stand Genomic Sci* 2013; **8**:389-402. [PubMed](#)  
<http://dx.doi.org/10.4056/sigs.3858183>
  16. Buddruhs N, Chertkov O, Petersen J, Fiebig A, Chen A, Pati A, Ivanova N, Lapidus A, Goodwin LA, Chain P, et al. Complete genome sequence of the marine methyl-halide oxidizing *Leisingera methylhalidivorans* type strain (DSM 14336T), a member of the *Roseobacter* clade. *Stand Genomic Sci* 2013; **9**:128-141. [PubMed](#)  
<http://dx.doi.org/10.4056/sigs.4297965>
  17. Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, et al. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* 2004; **432**:910-913. [PubMed](#) <http://dx.doi.org/10.1038/nature03170>
  18. Vaas LAI, 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. [PubMed](#)  
<http://dx.doi.org/10.1371/journal.pone.0034846>
  19. Vaas LAI, Sikorski J, Hofner B, Buddruhs N, Fiebig A, Klenk HP, Göker M. opm: An R package for analysing OmniLog® Phenotype MicroArray Data. *Bioinformatics* 2013; **29**:1823-1824. [PubMed](#)  
<http://dx.doi.org/10.1093/bioinformatics/btt291>
  20. 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>
  21. Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, Gilbert J, Glöckner FO, Hirschman L, Karsch-Mzrachhi I, et al. Clarifying Concepts and Terms in Biodiversity Informatics. *PLoS Biol* 2011; **9**:e1001088. [PubMed](#)  
<http://dx.doi.org/10.1371/journal.pbio.1001088>
  22. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](#)  
<http://dx.doi.org/10.1073/pnas.87.12.4576>
  23. Garrity G, Bell J, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Brenner D, Krieg N, Staley J, Garrity G, eds. *Bergey's Manual of Systematic Bacteriology*, Vol. 2 Part B *The Gammaproteobacteria*. Second Edition. New York: Springer; 2005:1.
  24. Garrity G, Bell J, Lilburn T. Class I. *Alphaproteobacteria* class. nov. In: Garrity G, Brenner D, Krieg N, Staley J, eds. *Bergey's Manual of Systematic Bacteriology*, Volume 2, Part C. Second Edition. New York: Springer; 2005:1.
  25. 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. [PubMed](#)  
<http://dx.doi.org/10.1099/ijs.0.64188-0>
  26. Garrity G, Bellm J, Lilburn T. Order III. *Rhodobacterales* ord. nov. In: Garrity G, Brenner D, Krieg N, Staley J, eds. *Bergey's Manual of Systematic Bacteriology*, Volume 2, Part C. Second Edition. New York: Springer; 2005:161.
  27. Garrity GM, Bell JA, Lilburn T. Family III. *Rhodobacteraceae* fam. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, eds. *Bergey's Manual of Systematic Bacteriology*, Volume 2, Part C. Second Edition. New York: Springer; 2005:161.
  28. Yoon JH, Kang SJ, Lee SY, Oh TK. *Phaeobacter daeponensis* sp. nov., isolated from a tidal flat of the Yellow Sea in Korea. *Int J Syst Evol Microbiol* 2007; **57**:856-861. [PubMed](#)  
<http://dx.doi.org/10.1099/ijs.0.64779-0>
  29. BAuA. 2010, Classification of bacteria and archaea in risk groups. <http://www.baua.de/TRBA168>, p.
  30. 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>
  31. List of growth media at the DSMZ: <http://www.dsmz.de/catalogues/catalogue-microorganisms/culture-technology/list-of-media-for-microorganisms.html>
  32. 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.



- [PubMed http://dx.doi.org/10.1186/1471-2105-11-119](http://dx.doi.org/10.1186/1471-2105-11-119)
33. Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC. The DOE-JGI Standard operating procedure for the annotations of microbial genomes. *Stand Genomic Sci* 2009; **1**:63-67. [PubMed](#) <http://dx.doi.org/10.4056/sigs.632>
34. Zhang S, Meyer R. The relaxosome protein MobC promotes conjugal plasmid mobilization by extending DNA strand separation to the nick site at the origin of transfer. *Mol Microbiol* 1997; **25**:509-516. [PubMed](#) <http://dx.doi.org/10.1046/j.1365-2958.1997.4861849.x>
35. Colby J, Stirling DI, Dalton H. The soluble methane mono-oxygenase of *Methylococcus capsulatus* (Bath). Its ability to oxygenate n-alkanes, n-alkenes, ethers, and alicyclic, aromatic and heterocyclic compounds. *Biochem J* 1977; **165**:395-402. [PubMed](#)
36. Schilling B, Lerch K. Cloning, sequencing and heterologous expression of the monoamine oxidase gene from *Aspergillus niger*. *Mol Gen Genet* 1995; **247**:430-438. [PubMed](#) <http://dx.doi.org/10.1007/BF00293144>
37. Cumby N, Davidson AR, Maxwell KL. The moron comes of age. *Bacteriophage* 2012; **2**:225-228. [PubMed](#) <http://dx.doi.org/10.4161/bact.23146>
38. Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF, Kim SH. Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature* 1998; **396**:703-707. [PubMed](#) <http://dx.doi.org/10.1038/25393>
39. Langille MGI, Brinkman FSL. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 2009; **25**:664-665. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/btp030>
40. Langille MG, Hsiao WWL, Brinkman FSL. Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics* 2008; **9**:329. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-9-329>
41. Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovik K, Meinicke P, Merkl R. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 2006; **7**:142. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-7-142>
42. Hsiao W, Wan I, Jones SJ, Brinkman FSL. IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 2003; **19**:418-420. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/btg004>
43. del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M, Diaz-Orejes R. Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* 1998; **62**:434-464. [PubMed](#)
44. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A fast phage search tool. *Nucleic Acids Res* 2011; **39**:W347-W352. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkr485>
45. Biers EJ, Wang K, Pennington C, Belas R, Chen F, Moran MA. Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl Environ Microbiol* 2008; **74**:2933-2939. [PubMed](#) <http://dx.doi.org/10.1128/AEM.02129-07>
46. Harrison PW, Lower RPJ, Kim NKD, Young JPW. Introducing the bacterial "chromid": not a chromosome, not a plasmid. *Trends Microbiol* 2010; **18**:141-148. [PubMed](#) <http://dx.doi.org/10.1016/j.tim.2009.12.010>
47. Petersen J, Frank O, Göker M, Pradella S. Extrachromosomal, extraordinary and essential the plasmids of the *Roseobacter* clade. *Appl Microbiol Biotechnol* 2013; **97**:2805-2815. [PubMed](#) <http://dx.doi.org/10.1007/s00253-013-4746-8>
48. Petersen J, Brinkmann H, Berger M, Brinkhoff T, Päucker O, Pradella S. Origin and evolution of a novel DnaA-like plasmid replication type in *Rhodobacterales*. *Mol Biol Evol* 2011; **28**:1229-1240. [PubMed](#) <http://dx.doi.org/10.1093/molbev/msq310>
49. Petersen J, Brinkmann H, Pradella S. Diversity and evolution of repABC type plasmids in *Rhodobacterales*. *Environ Microbiol* 2009; **11**:2627-2638. [PubMed](#) <http://dx.doi.org/10.1111/j.1462-2920.2009.01987.x>
50. Bartosik D, Włodarczyk M, Thomas CM. Complete nucleotide sequence of the replicator region of *Paracoccus* (*Thiobacillus*) *versutus* pTAV1 plasmid and its correlation to several plasmids of *Agrobacterium* and *Rhizobium* species. *Plasmid* 1997; **38**:53-59. [PubMed](#) <http://dx.doi.org/10.1006/plas.1997.1295>
51. Petersen J. Phylogeny and compatibility: plasmid classification in the genomics era. *Arch Microbiol* 2011; **193**:313-321. [PubMed](#)
52. Cascales E, Christie PJ. The versatile bacterial type IV secretion systems. *Nat Rev Microbiol* 2003; **1**:137-149. [PubMed](#) <http://dx.doi.org/10.1038/nrmicro753>
53. Zielenkiewicz U, Ceglowski P. Mechanisms of plasmid stable maintenance with special focus on plasmid addiction systems. *Acta Biochim Pol* 2001; **48**:1003-1023. [PubMed](#)
54. R Development Core Team. R: A language and environment for statistical computing. R Founda-

- tion for Statistical Computing, Vienna, Austria 2008. ISBN 3-900051-07-0.
55. Suzuki R, Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 2006; **22**:1540-1542. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/btl117>
  56. Neilands JB. Siderophores: Structure and Function of Microbial Iron Transport Compounds. *J Biol Chem* 1995; **270**:26723-26726. [PubMed](#) <http://dx.doi.org/10.1074/jbc.270.45.26723>
  57. Geng H, Bruhn JB, Nielsen KF, Gram L, Belas R. Genetic dissection of tropodithietic acid biosynthesis by marine roseobacters. *Appl Environ Microbiol* 2008; **74**:1535-1545. [PubMed](#) <http://dx.doi.org/10.1128/AEM.02339-07>
  58. Lally ET, Hill RB, Kieba IR, Korostoff J. The interaction between RTX toxins and target cells. *Trends Microbiol* 1999; **7**:356-361. [PubMed](#) [http://dx.doi.org/10.1016/S0966-842X\(99\)01530-9](http://dx.doi.org/10.1016/S0966-842X(99)01530-9)
  59. Giraud MF, Naismith JH. The rhamnase pathway. *Curr Opin Struct Biol* 2000; **10**:687-696. [PubMed](#) [http://dx.doi.org/10.1016/S0959-440X\(00\)00145-Z](http://dx.doi.org/10.1016/S0959-440X(00)00145-Z)
  60. Ravin NV. N15: the linear phage-plasmid. *Plasmid* 2011; **65**:102-109. [PubMed](#) <http://dx.doi.org/10.1016/j.plasmid.2010.12.004>
  61. Rybchin VN, Svarchevsky AN. The plasmid prophage N15: a linear DNA with covalently closed ends. *Mol Microbiol* 1999; **33**:895-903. [PubMed](#) <http://dx.doi.org/10.1046/j.1365-2958.1999.01533.x>
  62. Kumari A, Pasini P, Deo SK, Flomenhoft D, Shashidhar S, Daunert S. Biosensing Systems for the Detection of Bacterial Quorum Signaling Molecules. *Anal Chem* 2006; **78**:7603-7609. [PubMed](#) <http://dx.doi.org/10.1021/ac061421n>
  63. Collins AJ, Nyholm SV. Draft genome of *Phaeobacter gallaeciensis* ANG1, a dominant member of the accessory nidamental gland of *Euprymna scolopes*. *J Bacteriol* 2011; **193**:3397-3398. [PubMed](#) <http://dx.doi.org/10.1128/JB.05139-11>
  64. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013; **14**:60. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-14-60>
  65. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E, et al. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int J Syst Bacteriol* 1987; **37**:463-464. <http://dx.doi.org/10.1099/00207713-37-4-463>
  66. Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 2010; **60**:249-266. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.016949-0>
  67. Spring S, Scheuner C, Lapidus A, Lucas S, Glavina Del Rio T, Tice H, Copeland A, Cheng JF, Chen F, et al. The genome sequence of *Methanohalophilus mahii* SLPT reveals differences in the energy metabolism among members of the *Methanosarcinaceae* inhabiting freshwater and saline environments. *Archaea* 2010; **2010**:690737. [PubMed](#) <http://dx.doi.org/10.1155/2010/690737>
  68. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. [PubMed](#) <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334>
  69. Göker M, Scheuner C, Klenk HP, Stielow JB, Menzel W. Codivergence of Mycoviruses with Their Hosts. *PLoS ONE* 2011; **6**:e22252. [PubMed](#) <http://dx.doi.org/10.1371/journal.pone.0022252>
  70. Abt B, Han C, Scheuner C, Lu M, Lapidus A, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng JF, et al. Complete genome sequence of the termite hindgut bacterium *Spirochaeta coccoides* type strain (SPN1<sup>T</sup>), reclassification in the genus *Sphaerochaeta* as *Sphaerochaeta coccoides* comb. nov. and emendations of the family *Spirochaetaceae* and the genus *Sphaerochaeta*. *Stand Genomic Sci* 2012; **6**:194-209. [PubMed](#) <http://dx.doi.org/10.4056/sigs.2796069>
  71. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**:3389-3402. [PubMed](#) <http://dx.doi.org/10.1093/nar/25.17.3389>
  72. Enright AJ, Van Dongen SM, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 2002; **30**:1575-1584. [PubMed](#) <http://dx.doi.org/10.1093/nar/30.7.1575>
  73. Li L, Stoeckert CJ, Jr., Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003; **13**:2178-2189. [PubMed](#) <http://dx.doi.org/10.1101/gr.1224503>
  74. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**:1792-1797. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkh340>



75. Thompson JD, Thierry JCC, Poch O. RASCAL: rapid scanning and correction of multiple sequence alignments. *Bioinformatics* 2003; **19**:1155-1161. [PubMed](#)  
<http://dx.doi.org/10.1093/bioinformatics/btg133>
76. Meusemann K, Von Reumont BM, Simon S, Roeding F, Strauss S, Kück P, Ebersberger I, Walz M, Pass G, Breuers S, et al. A phylogenomic approach to resolve the arthropod tree of life. *Mol Biol Evol* 2010; **27**:2451-2464. [PubMed](#)  
<http://dx.doi.org/10.1093/molbev/msq130>
77. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 1981; **17**:368-376. [PubMed](#)  
<http://dx.doi.org/10.1007/BF01734359>
78. Fitch WM. Toward defining the course of evolution: minimum change on a specified tree topology. *Syst Zool* 1971; **20**:406-416.  
<http://dx.doi.org/10.2307/2412116>
79. Goloboff PA. Parsimony, likelihood, and simplicity. *Cladistics* 2003; **19**:91-103.  
<http://dx.doi.org/10.1111/j.1096-0031.2003.tb00297.x>
80. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006; **22**:2688-2690. [PubMed](#)  
<http://dx.doi.org/10.1093/bioinformatics/btl446>
81. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 2008; **57**:758-771. [PubMed](#)  
<http://dx.doi.org/10.1080/10635150802429642>
82. Swofford DL. PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods), Version 4.0 b10. *Sinauer Association, MA: Sunderland*, 2002.
83. Anderson I, Scheuner C, Göker M, Mavromatis K, Hooper SD, Porat I, Klenk HP, Ivanova N, Kyrpides NC. Novel insights into the diversity of catabolic metabolism from ten haloarchaeal genomes. *PLoS ONE* 2011; **6**:e20237. [PubMed](#)  
<http://dx.doi.org/10.1371/journal.pone.0020237>
84. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc Lond* 2007; **92**:669-674.  
<http://dx.doi.org/10.1111/j.1095-8312.2007.00864.x>
85. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *J Comput Biol* 2010; **17**:337-354. [PubMed](#)  
<http://dx.doi.org/10.1089/cmb.2009.0179>