Title

High quality draft genome sequence and analysis of Pontibacter roseus type strain SRC-1T (DSM 17521T) isolated from muddy waters of a drainage system in Chandigarh, India

Authors

Mukherjee, Supratim
Lapidus, Alla
Shapiro, Nicole
et al.

Publication Date

2015-06-01

High quality draft genome sequence and analysis of *Pontibacter roseus* type strain SRC-1^T (DSM 17521^T) isolated from muddy waters of a drainage system in Chandigarh, India

Supratim Mukherjee[1], Alla Lapidus[2,3], Nicole Shapiro[1], Jan-Fang Cheng[1], James Han[1], T.B.K. Reddy[1], Marcel Huntemann[1], Natalia Ivanova[1], Natalia Mikhailova[1], Amy Chen[5], Krishna Palaniappan[5], Stefan Spring[4], Markus Göker[4], Victor Markowitz[5], Tanja Woyke[1], Brian J. Tindall[4], Hans-Peter Klenk[4], Nikos C. Kyrpides[1,6] and Amrita Pati[1*]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] T. Dobzhansky Center for Genome Bionformatics, St. Petersburg State University, St. Petersburg, Russia
[3] Algorithmic Biology Lab, St. Petersburg Academic University, St. Petersburg, Russia
[4] Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
[5] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[6] King Abdulaziz University, Jeddah, Saudi Arabia

* Corresponding author: apati@lbl.gov (Amrita Pati)

November 26, 2014

ACKNOWLEDGMENTS:

This work was supported by the Director, Office of Science, Office of Biological and Environmental Research, Life Sciences Division, U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The authors gratefully acknowledge the assistance of Andrea Schutze for growing *P. roseus* DSM 17521T cultures and Evelyne-Marie Brambilla for DNA extraction and quality control (both at the DSMZ). A. L. was supported in part by Russian Ministry of Science Mega-grant no.11.G34.31.0068 (PI. Dr. Stephen J O'Brien).

High quality draft genome sequence and analysis of *Pontibacter roseus* type strain SRC-1^T (DSM 17521^T) isolated from muddy waters of a drainage system in Chandigarh, India

High quality draft genome sequence and analysis of *Pontibacter roseus* type strain SRC-1^T (DSM 17521^T) isolated from muddy waters of a drainage system in Chandigarh, India

Supratim Mukherjee[1], Alla Lapidus[2,3], Nicole Shapiro[1], Jan-Fang Cheng[1], James Han[1], T.B.K. Reddy[1], Marcel Huntemann[1], Natalia Ivanova[1], Natalia Mikhailova[1], Amy Chen[5], Krishna Palaniappan[5], Stefan Spring[4], Markus Göker[4], Victor Markowitz[5], Tanja Woyke[1], Brian J. Tindall[4], Hans-Peter Klenk[4], Nikos C. Kyrpides[1,6] and Amrita Pati[1*]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] T. Dobzhansky Center for Genome Bionformatics, St. Petersburg State University, St. Petersburg, Russia
[3] Algorithmic Biology Lab, St. Petersburg Academic University, St. Petersburg, Russia
[4] Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
[5] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[6] King Abdulaziz University, Jeddah, Saudi Arabia

* Corresponding author: apati@lbl.gov (Amrita Pati)

Keywords
Aerobic, Gram-negative, non-motile, obligate aerobe, halotolerant, menaquinone, GEBA, KMG-I

Abstract

Pontibacter roseus Suresh et al 2006 is a member of genus *Pontibacter* family *Cytophagaceae*, class *Cytophagia*. While the type species of the genus *Pontibacter actiniarum* was isolated in 2005 from a marine environment, subsequent species of the same genus have been found in different types of habitats ranging from seawater, sediment, desert soil, rhizosphere, contaminated sites, solar saltern and muddy water. Here we describe the features of *Pontibacter roseus* strain SRC-1^T along with its complete genome sequence and annotation from a culture of DSM 17521^T. The 4,581,480 bp long draft genome consists of 12 scaffolds with 4,003 protein-coding and 50 RNA genes and is a part of *Genomic Encyclopedia of Type Strains*, Phase I: the one thousand microbial genomes (KMG-I) project.

Introduction

The genus *Pontibacter* was first reported by Nedashkovskaya *et al.* [1] where they identified and described a menaquinone producing strain isolated from sea anemones. Several new species of the same genus have been reported in the literature since then. In addition to *Pontibacter roseus*, there are eighteen species with validly published names belonging to *Pontibacter* genus as of writing this manuscript. Members of genus *Pontibacter* including *P. roseus,* is of interest for genomic research due to their ability to synthesize and use menaquinone-7 (MK-7) as the primary respiratory quinone as well as to facilitate functional genomic studies within the group. Strain SRC-1^T (= DSM 17521 = CCTCC AB 207222 = CIP 109903 = MTCC 7260) is the type strain of *Pontibacter roseus,* which was isolated from muddy water from an occasional drainage system of a residential area in Chandigarh, India [2]. *P. roseus* SRC-1^T was initially reported to be *Effluviibacter roseus* SRC-1^T primarily due to its non-motile nature and fatty acid composition [2]. However, subsequent analysis of its fatty acid profile was shown to be more '*Pontibacter*-like' and gliding motility was observed to be variable in other *Pontibacter* species [3]. Further, its DNA G+C content, which was originally reported as 59 mol% [2], was also emended to be 52.0–52.3 mol% [3], a value more representative of members of the genus *Pontibacter*. As such, it was reclassified as *Pontibacter roseus* SRC-1^T [3]. Here we present a summary classification and features for *Pontibacter roseus* SRC-1^T, along with the genome sequence and annotation of DSM 17521^T.

Organism information

Classification and features

P. roseus SRC-1^T cells are non-motile, stain Gram-negative, do not form spores and are rod-shaped approximately 1.0–3.0 μm in length and 0.3–0.5 μm in width [2]. It is an obligate aerobe which can grow at a wide temperature range of 4-37 °C with the optimum being 30 °C (Table 1 and [2]). *P. roseus* SRC-1^T is a halotolerant microbe, can tolerate up to 8% NaCl and can utilize a wide range of sugars such as D-fructose, D-galactose, D-glucose, lactose, raffinose and sucrose as the sole source of carbon (Table 1 and [2]).

A representative genomic 16S rRNA sequence of *Pontibacter roseus* SRC-1^T was compared with the May 2013; release 13_5 of Greengenes database [4] using NCBI BLAST under default values. The top 250 hits with an alignment length cut-off of 1000 bp were retained among which genomes belonging to genus *Pontibacter* were the most abundant (45.6%) followed by *Adhaeribacter* (35.6%), those assigned to the family *Cytophagaceae* but without a defined genus name (16.4%) and *Hymenobacter* (2.4%). Among samples with available metadata, approximately 61% of the above hits were from a soil environment, 11% were isolated from skin and aapproximately 9% from aquatic samples. This distribution reflects the wide range of habitats commonly observed among members of the genus *Pontibacter* and its phylogenetic neighbors, ranging from forest soil to desert, contaminated aquatic and

soil environments, sediments and seawater among others [5–8]. Figure 1 shows the phylogenetic neighborhood of *Pontibacter roseus* SRC-1^T in a 16S rRNA based tree.

The predominant respiratory quinone for strain SRC-1^T is menaquinone 7 (MK-7), consistent with other members of the *Pontibacter* genus. Short chain menaquinones with six or seven isoprene units are characteristic of the different genera within the aerobic members of the phylum *Bacteroidetes.* The primary whole-cell fatty acids are branched chain iso-$C_{15:0}$ (14%), iso-$C_{17:0}$ 3-OH (14.7%) and summed feature 4 (34.9%, comprising of anteiso-$C_{17:1}$ B and/or iso-$C_{17:1}$ I, a pair of fatty acids that are grouped together for the purpose of evaluation by the Microbial Identification System(MIDI) as described earlier [24]) [2,3]. 2-OH fatty acids are absent. The original paper describing *P. roseus* SRC-1^T (as *Effluviibacter roseus*) [2] lists the polar lipids in strain SRC-1^T being phosphatidylglycerol, diphosphatidylglycerol and an unknown phospholipid. This is in stark contrast to the known lipid profile of this evolutionary group where phosphatidylethanolamine is usually the sole major digylceride based phospholipid and other non-phosphate based lipids make up a significant proportion of the polar lipids. Accordingly, while genes for phosphatidylserine synthase and a decarboxylase to convert the serine to phosphatidylethanolamine could be detected, we did not find any evidence in *P. roseus* DSM 17521^T genome to indicate that it produces the corresponding enzymes involved in the synthesis of phosphatidylglycerol or diphosphatidylglycerol. We therefore conclude that the original report on the lipid composition of strain SRC-1^T is probably in error. It should be noted that the original publication did not provide images of the TLC plates allowing others to examine these data set [2].

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [25,26]. It is a part of the *Genomic Encyclopedia of Type Strains*, Phase I: the one thousand microbial genomes (KMG-I) project [27], a follow-up of the *Genomic Encyclopedia of Bacteria and Archaea* (GEBA) project [28] which aims to increase the sequencing coverage of key reference microbial genomes and to generate a large genomic basis for the discovery of genes encoding novel enzymes [29]. KMG-I is a Genomic Standards Consortium project [30]. The genome project is deposited in the Genomes OnLine Database [11], the annotated genome is publicly available from the IMG Database [31] under the accession 2515154084, and the permanent draft genome sequence has been deposited at GenBank under accession number ARDO00000000. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI) using state of the art technology [32]. The project information is briefly summarized in Table 2.

Growth conditions and DNA isolation

Pontibacter roseus DSM 17521^T, was grown aerobically in DSMZ medium 948 (Oxoid nutrient broth) [33] at 30°C. Genomic DNA was isolated using a Jetflex Genomic DNA Purification Kit (GENOMED 600100) following the standard protocol provided by the manufacturer with the following modifications: an additional incubation (60 min, 37°C) with 50 µl proteinase K and finally adding 200 µl protein precipitation buffer (PPT). DNA is available through the DNA Bank Network [34].

Genome sequencing and assembly

The draft genome of *Pontibacter roseus* DSM 17521^T was generated at the DOE Joint Genome Institute (JGI) using the Illumina technology [35]. An Illumina Std shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 12,071,874 reads totaling 1,810.8 Mbp. All general aspects of library construction and sequencing performed at the JGI is publicly available [36]. All raw Illumina sequence data was passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artifacts [37]. Following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet (version 1.1.04) [38], (2) 1–3 Kbp simulated paired end reads were created from Velvet contigs using wgsim [39], (3) Illumina reads were assembled with simulated read pairs using Allpaths–LG (version r41043) [40]. Parameters for assembly steps were: 1) Velvet (velveth: 63 –shortPaired and velvetg: –very clean yes –export- Filtered yes –min_contig_lgth 500 –scaffolding no –cov_cutoff 10) 2) wgsim (–e 0 –1 100 –2 100 –r 0 –R 0 –X 0) 3) Allpaths–LG (PrepareAllpathsInputs: PHRED_64=1 PLOIDY=1 FRAG_COVERAGE=125 JUMP_COVERAGE=25 LONG_JUMP_COV=50, RunAllpathsLG: THREADS=8 RUN=std_shredpairs TARGETS=standard VAPI_WARN_ONLY=True OVERWRITE=True). The final draft assembly contained 12 scaffolds. The total size of the genome is 4.6 Mbp and the final assembly is based on 562.0 Mbp of Illumina data, which provides an average 122.8 × coverage of the genome.

Genome annotation

Genes were identified using Prodigal [41] as part of the JGI genome annotation pipeline [42], followed by a round of manual curation using the JGI GenePRIMP pipeline [43]. 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. These data sources were combined to assert a product description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE [44], RNAMMer [45], Rfam [46], TMHMM [47], SignalP [48] and CRT [49]. Additional gene functional annotation and comparative analysis were performed within the Integrated Microbial Genomes (IMG-ER) platform [50].

Genome properties

The assembly of the draft genome sequence consists of 12 scaffolds amounting to a 4,581,480 bp long chromosome with a GC content of approximately 53% (Table 3 and Figure 2). Of the 4,053 genes predicted, 4,003 were protein-coding genes along with 50 RNAs. The majority of protein-coding genes (69.4%) were assigned with a putative function while the remaining ones were annotated as hypothetical proteins.

The functional distribution of genes assigned to clusters of orthologous groups of proteins (COGs) is shown in Table 4. A large percentage of the genes do not have an assigned COG category, are unknown or fall into general function prediction, which is typical for a newly sequenced organism that has not been well characterized yet.

Insights from the genome sequence

Menaquinone Biosynthesis.

Respiratory lipoquinones such as ubiquinone and menaquinone are essential components of the electron transfer pathway in bacteria and archaea. While ubiquinones are limited to members of *Alphaproteobacteria*, *Gammaproteobacteria* and *Betaproteobacteria* [51], menaquinones have been found to be more widespread among prokaryotes [52,53], occurring in both aerobes and anaerobes. Menaquinone is a non-protein lipid-soluble redox component of the electron transport chain, which plays an important role in mediating electron transfer between membrane-bound protein complexes. The classical menaquinone biosynthesis pathway was studied primarily in *Escherichia coli*; more recently, an alternate pathway was identified in *Streptomyces coelicolor* A3(2) as well as in pathogens such as *Helicobacter pylori* and *Campylobacter jejuni* [54,55], aspects of which remain to be fully elucidated.

All identified species of the genus *Pontibacter* are known to possess menaquinone – 7 (MK-7) [6] which is the primary respiratory quinone in *Pontibacter roseus* SRC-1^T [2]. Biosynthesis of menaquinone in this organism appears to occur via the classical pathway. Using comparative genomics we identified the genes possibly involved in menaquinone biosynthesis in *P. roseus* DSM 17521^T (Table 5). Menaquinone biosynthesis genes have been extensively studied in *E. coli* where they are organized in an operon and in *B. subtilis* where gene neighborhood was helpful in identifying *men*C and *men*H genes [56]. However, the *P. roseus* genes seem to be spread across its chromosome. It is well known that conservation of gene order in bacteria can be disrupted during the course of evolution [57]. For example, isolated genes belonging to the menaquinone biosynthesis pathway leading to phylloquinione biosynthesis were identified in *Synechocystis* sp. PCC 6803 through sequence similarity with *E. coli* followed by transposon mutagenesis [58,59]. As more genomes become available, these aspects can be investigated in greater detail.

An o-succinylbenzoate synthase (OSBS) that is part of the menaquinone biosynthetic pathway encoded by the *men*C gene in *E. coli* and *B. subtilis* is missing from the *Pontibacter roseus* genome. A gene annotated as muconate cycloisomerase in *Pontibacter roseus* DSM 17521^T (IMG geneID 2515480441) may perform this function. It contains conserved domains belonging to Muconate Lactonizing Enzyme (MLE) subgroup of the enolase superfamily. Sequence similarity between different members of the enolase superfamily is typically less than 25% [66]. Even though they possess similar structural scaffolds, they are known to have evolved significantly such that their functional role cannot be easily assigned through sequence similarity alone [67]. For example, *B. subtilis men*C was initially annotated as 'similar to muconate cycloisomerase of *Pseudomonas putida*' and 'N-acylamino acid racemase' but was later corrected to be OSBS [56]. The *P. roseus* gene shares protein level identity of 48% with muconate cycloisomerase 1 of *Pseudomonas putida* [65] and approximately 23% and 17% with *E. coli* and *B. subtilis* MenC respectively [56,68]. Multiple sequence alignment (Figure 3) of the above three genes reveal conservation of Asp^{161}, Glu^{190}, Asp^{213} and Lys^{235} (boxes in Figure 3) which have been predicted to be essential for OSBS in *E. coli* and other members of the enzyme family [66]. We thereby propose that IMG 2515480441 performs the function of MenC in *P. roseus* DSM 17521^T.

Multidrug Resistance (MDR) Efflux pump.

Resistance to antibiotic drugs is one of the major public health concerns of today as highlighted in the recent report by the Centers for Disease Control and Prevention (CDC) [69]. Among several other mechanisms, multidrug (MDR) resistance efflux pumps play a very important role in conferring decreased susceptibility to antibiotics in bacteria by transporting drugs across the bacterial membrane and preventing intracellular accumulation [70]. AcrAB-TolC is one of the most studied MDR efflux systems in Gram-negative bacteria. It is comprised of an inner membrane efflux transporter (AcrB), a linker protein (AcrA) and an outer membrane protein (TolC) which interacts with AcrA and AcrB and forms a multifunctional channel that is essential to pump cellular products out of the cell [70,71]. Previous reports have identified gene clusters predicted to confer antibiotic resistance in members of *Pontibacter* [6]. Applying comparative analysis with characterized proteins, we identified a set of genes (IMG ID 2515478940–43) that may function as a multi-drug resistance efflux pump in *P. roseus* DSM 17521^T (Figure 4).

P. roseus 2515478940 is 37% identical to *E. coli* multidrug efflux pump subunit AcrB [72]; 2515478941 shares 28% identity to *E. coli* AcrA [73] while 2515478942 is 20% identical to *E. coli* outer membrane protein TolC [74]. Additionally, there is a transcriptional repressor (2515478943) upstream of TolC which shares 25% protein level identity to HTH-type transcriptional repressor Bm3R1 [75] from

Bacillus megaterium and may act as a regulator of the MDR transport system in *P. roseus* DSM 17521^T.

Discussion

Members of the genus *Pontibacter* occupy a unique phylogenetic niche within the phylum *Bacteroidetes*. As of writing, this genome report is only the second for the entire genus. In addition to a detailed analysis of the *P. roseus* genome we highlight some of the key functional characteristics of the organism and summarizes the genes encoding enzymes leading to the biosynthesis of menaquinone, the primary respiratory quinone for majority of the species of the genus.

Acknowledgements

The authors gratefully acknowledge the assistance of Andrea Schütze for growing *P. roseus* DSM 17521^T cultures and Evelyne-Marie Brambilla for DNA extraction and quality control (both at the DSMZ). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. A. L. was supported in part by Russian Ministry of Science Mega-grant no.11.G34.31.0068 (PI. Dr. Stephen J O'Brien).

References

1. Nedashkovskaya OI, Kim SB, Suzuki M, *et al. Pontibacter actiniarum* gen. nov., sp. nov., a novel member of the phylum "*Bacteroidetes*", and proposal of *Reichenbachiella* gen. nov. as a replacement for the illegitimate prokaryotic generic name *Reichenbachia* Nedashkovskaya et al. 2003. *Int. J. Syst. Evol. Microbiol.* 2005;**55**:2583–2588.

2. Suresh K, Mayilraj S, Chakrabarti T. *Effluviibacter roseus* gen. nov., sp. nov., isolated from muddy water, belonging to the family "*Flexibacteraceae*." *Int. J. Syst. Evol. Microbiol.* 2006;**56**:1703–1707.

3. Wang Y, Zhang K, Cai F, *et al. Pontibacter xinjiangensis* sp. nov., in the phylum "*Bacteroidetes*", and reclassification of [*Effluviibacter] roseus* as *Pontibacter roseus* comb. nov. *Int. J. Syst. Evol. Microbiol.* 2010;**60**:99–103.

4. DeSantis TZ, Hugenholtz P, Larsen N, *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 2006;**72**:5069–5072.

5. Singh AK, Garg N, Lata P, *et al.* *Pontibacter indicus* sp. nov., isolated from hexachlorocyclohexane-contaminated soil. *Int. J. Syst. Evol. Microbiol.* 2014;**64**:254–259.

6. Joshi MN, Sharma AC, Pandya RV, *et al.* Draft Genome Sequence of *Pontibacter sp*. nov. BAB1700, a Halotolerant, Industrially Important Bacterium. *J. Bacteriol.* 2012;**194**:6329–6330.

7. Zhou Y, Wang X, Liu H, *et al.* *Pontibacter akesuensis* sp. nov., isolated from a desert soil in China. *Int. J. Syst. Evol. Microbiol.* 2007;**57**:321–325.

8. Nedashkovskaya OI, Kim SB, Suzuki M, *et al.* *Pontibacter actiniarum* gen. nov., sp. nov., a novel member of the phylum "*Bacteroidetes*", and proposal of *Reichenbachiella* gen. nov. as a replacement for the illegitimate prokaryotic generic name *Reichenbachia* Nedashkovskaya et al. 2003. *Int. J. Syst. Evol. Microbiol.* 2005;**55**:2583–2588.

9. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 1987;**4**:406–425.

10. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 2011;**28**:2731–2739.

11. Pagani I, Liolios K, Jansson J, *et al.* The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 2012;**40**:D571–579.

12. Copeland A, Zhang X, Misra M, *et al.* Complete genome sequence of the aquatic bacterium *Runella slithyformis* type strain (LSU 4^{T}). *Stand. Genomic Sci.* 2012;**6**.

13. Mongodin EF, Nelson KE, Daugherty S, *et al.* The genome of *Salinibacter ruber*: Convergence and gene exchange among hyperhalophilic Bacteria and*Archaea*. *Proc. Natl. Acad. Sci. U. S. A.* 2005;**102**:18147–18152.

14. Field D, Garrity G, Gray T, *et al.* The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol.* 2008;**26**:541–547.

15. Field D, Amaral-Zettler L, Cochrane G, *et al.* The Genomic Standards Consortium. *PLoS Biol.* 2011;**9**:e1001088.

16. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* 1990;**87**:4576–4579.

17. Krieg N, Ludwig W, Euzéby J, Whitman W. Phylum XIV. *Bacteroidetes* phyl. nov. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W,

Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, second edition vol. 4 (The *Bacteroidetes*, *Spirochaetes*, *Tenericutes (Mollicutes)*, *Acidobacteria*, *Fibrobacteres*, *Fusobacteria*, *Dictyoglomi*, *Gemmatimonadetes*, *Lentisphaerae*, *Verrucomicrobia*, *Chlamydiae*, *and Planctomycetes*).Vol 4. Springer New York; 2010:25.

18. List of new names and new combinations previously effectively, but not validly, published. *Int. J. Syst. Evol. Microbiol.* 2012;**62**:1–4.

19. Nakagawa Y. Class IV. *Cytophagia* class. nov. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, second edition, vol. 4 (The *Bacteroidetes*, *Spirochaetes*, *Tenericutes (Mollicutes)*, *Acidobacteria*, *Fibrobacteres*, *Fusobacteria*, *Dictyoglomi*, *Gemmatimonadetes*, *Lentisphaerae*, *Verrucomicrobia*, *Chlamydiae*, *and Planctomycetes*.) Springer New York; 2010:370.

20. Skerman VBD, McGOWAN V, Sneath PHA. Approved Lists of Bacterial Names. *Int. J. Syst. Bacteriol.* 1980;**30**:225–420.

21. Leadbetter E. Order II. *Cytophagales* Nomen novum. In: Buchanan RE, Gibbons NE (eds), Bergey's Manual of Determinative Bacteriology. Eight Edition. The Williams and Wilkins Co., Baltimore; 1974:99.

22. Stanier RY. Studies on the *Cytophagas*. *J. Bacteriol.* 1940;**40**:619–635.

23. Ashburner M, Ball CA, Blake JA, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 2000;**25**:25–29.

24. Pandey KK, Mayilraj S, Chakrabarti T. *Pseudomonas indica* sp. nov., a novel butane-utilizing species. *Int. J. Syst. Evol. Microbiol.* 2002;**52**:1559–1567.

25. Göker M, Klenk H-P. Phylogeny-driven target selection for large-scale genome-sequencing (and other) projects. *Stand. Genomic Sci.* 2013;**8**:360–374.

26. Klenk H-P, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst. Appl. Microbiol.* 2010;**33**:175–182.

27. Kyrpides NC, Woyke T, Eisen JA, *et al.* Genomic Encyclopedia of Type Strains, Phase I: the one thousand microbial genomes (KMG-I) project. *Stand. Genomic Sci.* 2014;**9**.

28. Wu D, Hugenholtz P, Mavromatis K, *et al.* A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea. Nature* 2009;**462**:1056–1060.

29. Piao H, Froula J, Du C, *et al.* Identification of novel biomass-degrading enzymes from genomic dark matter: Populating genomic sequence space with functional annotation. *Biotechnol. Bioeng.* 2014;**111**:1550–1565.

30. Field D, Sterk P, Kottmann R, *et al.* Genomic Standards Consortium Projects. *Stand. Genomic Sci.* 2014;**9**:599–601.

31. Markowitz VM, Chen I-MA, Palaniappan K, *et al.* IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* 2014;**42**:D560–D567.

32. Mavromatis K, Land ML, Brettin TS, *et al.* The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PloS One* 2012;**7**:e48837.

33. List of growth media used at DSMZ. http://www.dsmz.de/catalogues/catalogue-microorganisms/culture-technology/list-of-media-for-microorganisms.html.

34. Gemeinholzer B, Dröge G, Zetzsche H, *et al.* The DNA Bank Network: The Start from a German Initiative. *Biopreservation Biobanking* 2011;**9**:51–55.

35. Bennett S. Solexa Ltd. *Pharmacogenomics* 2004;**5**:433–438.

36. DOE Joint Genome Institute: A DOE Office of Science User Facility of Lawrence Berkeley National Laboratory. *DOE Jt. Genome Inst.* http://jgi.doe.gov

37. Mingkun L, Copeland A, Han J. unpublished. 2011.

38. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;**18**:821–829.

39. Wgsim. Available at: https://github.com/lh3/wgsim.

40. Gnerre S, Maccallum I, Przybylski D, *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 2011;**108**:1513–1518.

41. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119.

42. Mavromatis K, Ivanova NN, Chen I-MA, 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.

43. Pati A, Ivanova NN, Mikhailova N, *et al.* GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat. Methods* 2010;**7**:455–457.

44. 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.

45. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;**35**:3100–3108.

46. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res.* 2003;**31**:439–441.

47. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 2001;**305**:567–580.

48. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 2004;**340**:783–795.

49. Bland C, Ramsey TL, Sabree F, *et al.* CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 2007;**8**:209.

50. Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinforma. Oxf. Engl.* 2009;**25**:2271–2278.

51. Tindall BJ. Section 4 update: Respiratory lipoquinones as biomarkers. In: Kowalchuk GA, Bruijn FJ de, Head IM, Akkermans AD, Elsas JD van, eds. Molecular Microbial Ecology Manual. Springer Netherlands; 2004:2907–2928.

52. Collins MD, Jones D. Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. *Microbiol. Rev.* 1981;**45**:316–354.

53. Zhi X-Y, Yao J-C, Tang S-K, Huang Y, Li H-W, Li W-J. The Futalosine Pathway Played an Important Role in Menaquinone Biosynthesis during Early Prokaryote Evolution. *Genome Biol. Evol.* 2014;**6**:149–160.

54. Hiratsuka T, Furihata K, Ishikawa J, *et al.* An Alternative Menaquinone Biosynthetic Pathway Operating in Microorganisms. *Science* 2008;**321**:1670–1673.

55. Dairi T. Menaquinone biosyntheses in microorganisms. *Methods Enzymol.* 2012;**515**:107–122.

56. Palmer DR, Garrett JB, Sharma V, Meganathan R, Babbitt PC, Gerlt JA. Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase. *Biochemistry (Mosc.)* 1999;**38**:4252–4258.

57. Itoh T, Takemoto K, Mori H, Gojobori T. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol. Biol. Evol.* 1999;**16**:332–346.

58. Johnson TW, Shen G, Zybailov B, *et al.* Recruitment of a Foreign Quinone into the A1 Site of Photosystem I I. GENETIC AND PHYSIOLOGICAL CHARACTERIZATION OF PHYLLOQUINONE BIOSYNTHETIC PATHWAY MUTANTS IN SYNECHOCYSTIS SP. PCC 6803. *J. Biol. Chem.* 2000;**275**:8523–8530.

59. Zhang S, Laborde SM, Frankel LK, Bricker TM. Four Novel Genes Required for Optimal Photoautotrophic Growth of the *Cyanobacterium Synechocystis* sp. Strain PCC 6803 Identified by In Vitro Transposon Mutagenesis. *J. Bacteriol.* 2004;**186**:875–879.

60. Daruwala R, Kwon O, Meganathan R, Hudspeth ME. A new isochorismate synthase specifically involved in menaquinone (vitamin K2) biosynthesis encoded by the menF gene. *FEMS Microbiol. Lett.* 1996;**140**:159–163.

61. Rowland B, Hill K, Miller P, Driscoll J, Taber H. Structural organization of a *Bacillus subtilis* operon encoding menaquinone biosynthetic enzymes. *Gene* 1995;**167**:105–109.

62. Sharma V, Hudspeth ME, Meganathan R. Menaquinone (vitamin K2) biosynthesis: localization and characterization of the menE gene from *Escherichia coli*. *Gene* 1996;**168**:43–48.

63. Sharma V, Suvarna K, Meganathan R, Hudspeth ME. Menaquinone (vitamin K2) biosynthesis: nucleotide sequence and expression of the menB gene from *Escherichia coli*. *J. Bacteriol.* 1992;**174**:5057–5062.

64. Suvarna K, Stevenson D, Meganathan R, Hudspeth ME. Menaquinone (vitamin K2) biosynthesis: localization and characterization of the menA gene from *Escherichia coli*. *J. Bacteriol.* 1998;**180**:2782–2787.

65. Aldrich TL, Frantz B, Gill JF, Kilbane JJ, Chakrabarty AM. Cloning and complete nucleotide sequence determination of the catB gene encoding cis,cis-muconate lactonizing enzyme. *Gene* 1987;**52**:185–195.

66. Thompson TB, Garrett JB, Taylor EA, Meganathan R, Gerlt JA, Rayment I. Evolution of Enzymatic Activity in the Enolase Superfamily: Structure of o-Succinylbenzoate Synthase from *Escherichia coli* in Complex with Mg2+ and o-Succinylbenzoate. *Biochemistry (Mosc.)* 2000;**39**:10662–10676.

67. Gerlt JA, Babbitt PC. Can sequence determine function? *Genome Biol.* 2000;**1**:REVIEWS0005.

68. Sharma V, Meganathan R, Hudspeth ME. Menaquinone (vitamin K2) biosynthesis: cloning, nucleotide sequence, and expression of the menC gene from *Escherichia coli*. *J. Bacteriol.* 1993;**175**:4917–4921.

69. CDC. ANTIBIOTIC RESISTANCE THREATS in the United States, 2013. Centers for Disease Control and Prevention; 2013. http://www.cdc.gov/drugresistance/threat-report-2013/.

70. Piddock LJV. Multidrug-resistance efflux pumps ? not just for resistance. *Nat. Rev. Microbiol.* 2006;**4**:629–636.

71. Viveiros M, Dupont M, Rodrigues L, *et al.* Antibiotic Stress, Genetic Response and Altered Permeability of *E. coli. PLoS ONE* 2007;**2**:e365.

72. Yu EW, Aires JR, Nikaido H. AcrB Multidrug Efflux Pump of *Escherichia coli:* Composite Substrate-Binding Cavity of Exceptional Flexibility Generates Its Extremely Wide Substrate Specificity. *J. Bacteriol.* 2003;**185**:5657–5664.

73. Ma D, Cook DN, Alberti M, Pon NG, Nikaido H, Hearst JE. Molecular cloning and characterization of acrA and acrE genes of *Escherichia coli. J. Bacteriol.* 1993;**175**:6299–6313.

74. Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* 2000;**405**:914–919.

75. Shaw GC, Fulco AJ. Barbiturate-mediated regulation of expression of the cytochrome P450BM-3 gene of *Bacillus megaterium* by Bm3R1 protein. *J. Biol. Chem.* 1992;**267**:5515–5526.

Tables

Table 1. Classification and general features of *Pontibacter roseus* SRC-1^T according to the MIGS recommendations [14], published by the Genomic Standards Consortium [15]

MIGS ID	Property	Term	Evidence code
MIGS-2		Domain *Bacteria*	TAS [16]
		Phylum *Bacteroidetes*	TAS [17,18]
		Class *Cytophagia*	TAS [18,19]
	Current classification	Order *Cytophagales*	TAS [20,21]
MIGS-3		Family *Cytophagaceae*	TAS [20,22]
		Genus *Pontibacter*	TAS [2,3,20]
		Species *Pontibacter roseus*	TAS [2,3,20]
MIGS-7	Strain	SRC-1^T	TAS [2,3,20]
	Gram stain	Gram-negative	TAS [2]
MIGS-37.1	Cell shape	Irregular rods	TAS [2]
MIGS-37.2	Motility	Non-motile	TAS [2]
MIGS-37.3	Sporulation	Non-sporulating	TAS [2]
MIGS-6.1	Temperature range	4-37 °C	TAS [2]
MIGS-6.1	Optimum temperature	30°C	TAS [2]
MIGS-6.3	Salinity	Halotolerant	TAS [2]
MIGS-22	Relationship to oxygen	Obligate aerobe	TAS [2]
	Carbon source	Sugars (Glucose, Galactose etc)	TAS [2]
MIGS-37.11	Energy metabolism	Chemoheterotroph	TAS [2]
MIGS-6	Habitat	Wastewater, aquatic	TAS [2]
MIGS-6.2	pH	pH 6.0–10.0	TAS [2]
MIGS-15	Biotic relationship	Free living	NAS
MIGS-14	Known pathogenicity	Not reported	
	Biosafety level	1	NAS
MIGS-23	Isolation	Muddy water	TAS [2]
MIGS-4	Geographic location	Chandigarh, India	TAS [2]
MIGS-5	Time of sample collection	Before 2006	TAS [2]
MIGS-4.1	Latitude	30.733	TAS [2]
MIGS-4.2	Longitude	76.779	TAS [2]

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 the Gene Ontology project [23].

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-Quality draft
MIGS-28	Libraries used	Illumina Std shotgun library
MIGS-29	Sequencing platforms	Illumina HiSeq 2000
MIGS-31.2	Sequencing coverage	122.8 × Illumina
MIGS-30	Assemblers	Velvet v. 1.1.04, ALLPATHS v. R41043
MIGS-11	Size (kb)	4581
	ORFS	4053
MIGS-35	GC Content	52.65%
MIGS-32	Gene calling method	Prodigal, GenePRIMP
	INSDC ID	ARDO01000000
	GOLD ID	Gi11777
	NCBI project ID	169723
	Release date	Please provide
	Database: IMG	2515154084
MIGS-13	Source material identifier	DSM 17521
	Project Relevance	GEBA-KMG, Tree of Life

Table 3. Genome Statistics

Attribute	Number	% of Total
Genome size (bp)	4,581,480	100.00%
DNA coding region (bp)	3,984,478	86.97%
DNA G+C content (bp)	2,411,942	52.65%
Total genes	4,053	100.00%
RNA genes	50	1.23%
rRNA operons	1	
tRNA genes	41	1.01%
Protein-coding genes	4,003	98.77%
Pseudo genes	-	0.00%
Genes with function prediction	2,813	69.41%
Genes in paralog clusters	1,373	33.88%
Genes assigned to COGs	2,790	68.84%
Genes assigned Pfam domains	3,062	75.55%
Genes with signal peptides	611	15.08%
Genes with transmembrane helices	997	24.60%

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

Code	Value	% of total[a]	Description
J	167	4.17	Translation
A	1	0.02	RNA processing and modification
K	134	3.35	Transcription
L	105	2.62	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	27	0.67	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	60	1.50	Defense mechanisms
T	94	2.35	Signal transduction mechanisms
M	229	5.72	Cell wall/membrane biogenesis
N	11	0.27	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	30	0.75	Intracellular trafficking and secretion
O	94	2.35	Posttranslational modification, protein turnover, chaperones
C	153	3.82	Energy production and conversion
G	126	3.15	Carbohydrate transport and metabolism
E	203	5.07	Amino acid transport and metabolism
F	65	1.62	Nucleotide transport and metabolism
H	99	2.47	Coenzyme transport and metabolism
I	96	2.40	Lipid transport and metabolism
P	149	3.72	Inorganic ion transport and metabolism
Q	69	1.72	Secondary metabolites biosynthesis, transport and catabolism
R	328	8.19	General function prediction only
S	229	5.72	Function unknown
-	1790	44.72	Not in COGs

a) The total is based on the total number of protein coding genes in the annotated genome.

Table 5. Predicted menaquinone biosynthesis genes in *Pontibacter roseus* DSM 17521^T

IMG GeneID	IMG Description	Identity to characterized proteins	Reference
2515478196	isochorismate synthases	28% identity to *E. coli* MenF	[60]
2515478195	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase	30% identity to *B. subtilis* MenD	[61]
2515478193	Acyl-CoA synthetases	26% identity to *E. coli* MenE	[62]
2515478204	naphthoate synthase	52% identity to *E. coli* MenB	[63]
2515479036	1,4-dihydroxy-2-naphthoate octaprenyltransferase	31% identity to *E. coli* MenA	[64]
2515480441	muconate and chloromuconate cycloisomerases	48% identity with muconate cycloisomerase 1 of *Pseudomonas putida*	[65]