Complete genome sequence of the orange-red pigmented, radioresistant *Deinococcus proteolyticus* type strain (MRP^T)

Alex Copeland[1], Ahmet Zeytun[1,2], Montri Yassawong[3], Matt Nolan[1], Susan Lucas[1], Nancy Hammon[1], Shweta Deshpande[1], Jan-Fang Cheng[1], Cliff Han[1,2], Roxanne Tapia[1,2], Lynne Goodwin[1,2], Sam Pitluck[1], Konstantinos Mavromatis[1], Konstantinos Liolios[1], Ioanna Pagani[1], Natalia Ivanova[1], Natalia Mikhailova[1], Amrita Pati[1], Amy Chen[3], Krishna Palaniappan[4], Miriam Land[1,5], Loren Hauser[1,5], Cynthia D. Jeffries[1,5], Evelyne-Marie Brambilla[6], Manfred Rohde[7], Johannes Sikorski[6], Rüdiger Pukall[6], Markus Göker[6], John C. Detter[1,2], Tanja Woyke[1], James Bristow[1], Jonathan A. Eisen[1,8], Victor Markowitz[4], Philip Hugenholtz[1,9], Nikos C. Kyrpides[1], Hans-Peter Klenk[6*], and Alla Lapidus[1]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
[3] Mahidol University, Faculty of Pharmacy, Department of Biochemistry, Bangkok, Thailand
[4] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[5] Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[6] Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
[7] HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany
[8] University of California Davis Genome Center, Davis, California, USA
[9] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

[*]Corresponding author: Hans-Peter Klenk

Keywords
strictly aerobic, non-motile, chemoorganotrophic, proteolytic, radioresistant, mesophile, carotenoid pigments, tetrad-forming cocci, Gram-positive, *Deinococcaceae*, GEBA

Abstract
Deinococcus proteolyticus (*ex* Kobatake *et al.* 1973) Brook and Murray 1981 is one of currently 47 species in the genus *Deinococcus* within the family *Deinococcaceae*. Strain MRPT^T was isolated from faeces of *Lama glama*; it shares with various other species of the genus the extreme radiation resistance, with *D. proteolyticus* being resistant up to 1.5 Mrad of gamma radiation. Strain MRPT^T is of further interest for its carotenoid pigment. The genome presented here is only the fifth completed genome sequence of a member of the genus *Deinococcus* (and the forth type strain) to be published, and will hopefully contribute to a better understanding of how members of this genus adapted to high gamma- or UV ionizing-radiation. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 2,886,836 bp long genome with its four large plasmids of 97 kbp, 132 kbp, 196 kbp and 315 kbp harbours 2,741 protein-coding and 58 RNA genes and is a part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project.

Introduction

Strain MRP^T, also known as Kobatake strain MRP, (= DSM 20540 = ATCC 35074 = JCM 6276) is the type strain of *Deinococcus proteolyticus* [1], one of currently 47 validly named species in the genus *Deinococcus* [2]. The genus name is derived from the latinized Greek word *deinos* meaning 'strange or unusual' and the Neo-Latin word *coccus* meaning 'a grain or berry', yielding the Neo-Latin 'Deinococcus', meaning the 'unusual coccus' [1]. The species epithet is derived from the Neo-Latin word *proteolyticus*, meaning proteolytic [1]. Strain MRP^T was isolated in the early 1970s from faeces of *Lama glama* by Kobatake *et al*., and became known under its synonym "*Micrococcus radioproteolyticus*" [30], which according to Rule 12a of the Bacteriological Code was an illegitimate species epithet because it expressed more than one single concept [1]. The genus name "*Micrococcus*" was not considered for the Approved Lists of Bacterial Names published by Skerman *et al*. in 1980 [50]. In 1981 Brooks and Murray posited the family *Deinococcaceae* and the genus *Deinococcus*, with *D. radiodurans* as the type species of the type genus and *D. proteolyticus* as one out of three other members of the novel genus [1]. Many strains of the family *Deinococcaceae* resist to high levels of gamma and ultraviolet radiation [1]. Cells of deinococci are spherical or rod shaped [3]. Several distinct cell wall layers are visible in thin section and the cell wall contains lipoprotein [1]. The natural habitat of the members of genus *Deinococcus* was unknown for a long time, largely because of the recognition was not easy [4]. Plasmids of strain MRP^T were previously analysed by Mackay *et al*. [51], survival of repeated lyophilisation was studies by Rýznar and Drásil [52], hsp70 [53], hps40 [54], and SSB genes were sequenced [55], primarily for phylogenetic analyses. The members of the genus *Deinococcus* have been isolated from diverse environments [5-8], usually selected and characterized by survival after high-dose irradiation [4]. To date no further isolates of *D. proteolyticus* have been reported. Here we present a summary classification and a set of features for *D. proteolyticus* MRP^T, together with the description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of *D. proteolyticus* MRP^T was compared using NCBI BLAST [9,10] under default settings (e.g., considering only the high-scoring segment pairs (HSPs) from the best 250 hits) with the most recent release of the Greengenes database [11] and the relative frequencies of taxa and keywords (reduced to their stem [12]) were determined, weighted by BLAST scores. The most frequently occurring genus was *Deinococcus* (100.0%) (85 hits in total). Regarding the two hits to sequences from members of the species, the average identity within HSPs was 99.8%, whereas the average coverage by HSPs was 98.4%. Regarding the 52 hits to sequences from other members of the genus, the average identity within HSPs was 92.0%, whereas the average coverage by HSPs was 95.1%. Among all other species, the one yielding the highest score was *Deinococcus piscis* (DQ683348), which corresponded to an identity of 98.0% and an HSP coverage of 98.5%. (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 JF171367 ('skin antecubital fossa clone ncd1964b12c1'), which showed an identity of 95.1% and an HSP coverage of 89.1%. The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'skin' (20.6%), 'fossa' (10.2%), 'forearm' (9.2%), 'volar' (8.8%) and 'antecubit' (6.7%) (165 hits in total). Environmental samples which yielded hits of a higher score than the highest scoring species were not found.

Figure 1 shows the phylogenetic neighborhood of *D. proteolyticus* in a 16S rRNA based tree. The sequences of the three identical 16S rRNA gene copies in the genome differ by three nucleotides from the previously published 16S rRNA sequence (Y11331).

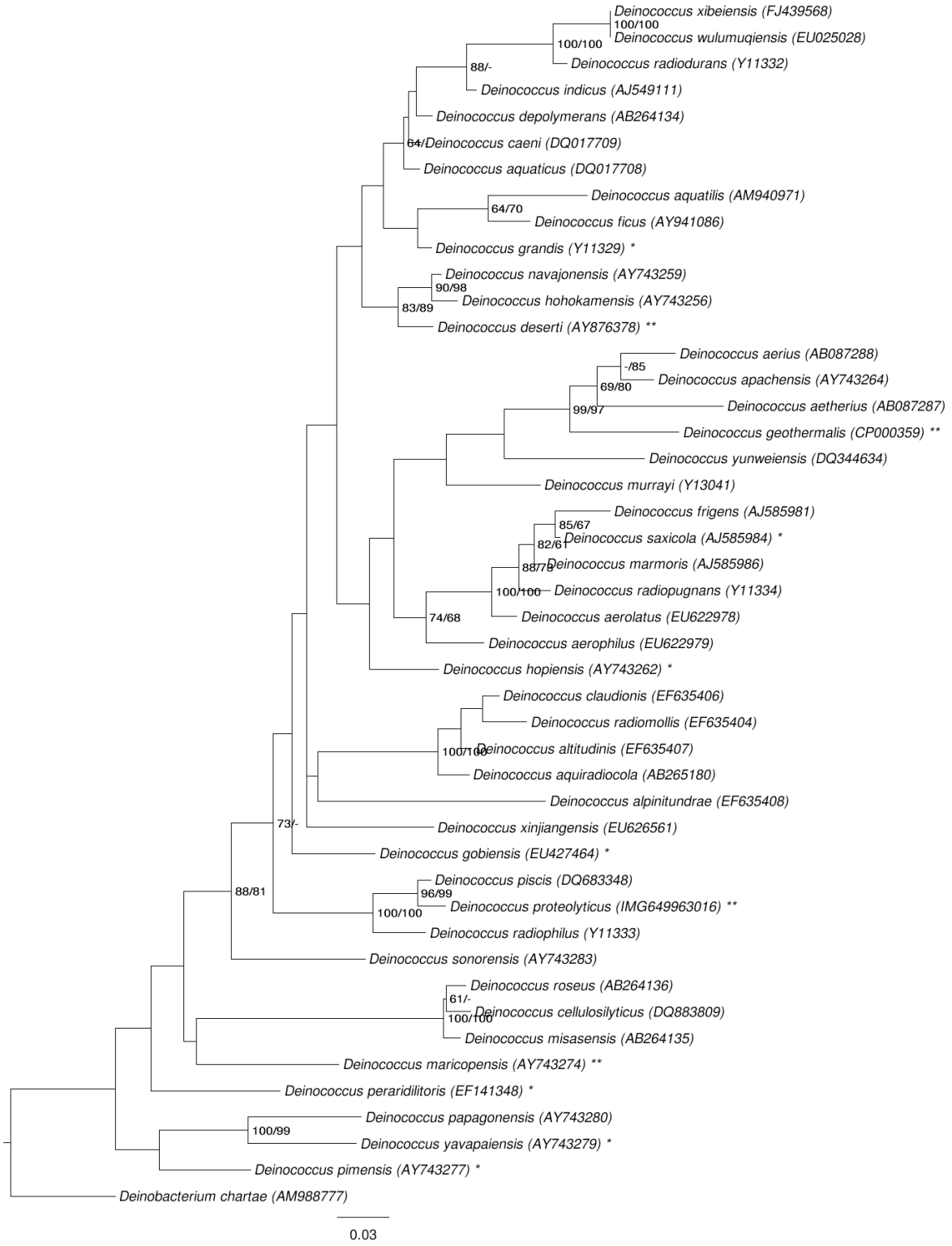

Figure 1. Phylogenetic tree highlighting the position of *D. proteolyticus* relative to the type strains of the other species within the family *Deinococcaceae*. The tree was inferred from 1,377 aligned characters [13,14] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [15]. Rooting was done initially using the midpoint method [16] and then checked for its agreement with the current classification (Table 1). The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 750 ML bootstrap replicates [17] (left) and from 1,000 Maximum-Parsimony bootstrap replicates [18] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [19] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks [46-48]. The genome of *D. radiodurans* published by White *et al.* in 1999 [49] later turned out not to be from the type strain [35].

Strain MRP^T is strictly aerobic, Gram-positive and non-motile [1]. Cells are spheres (Figure 2), 1.0 to 2.0 µm in diameter, occurring singly and in pairs [1]. Cells are dividing in two planes to form tetrads or tablets of cells, and the cell wall consists of at least three distinct layers [1]. Resting stages of cells are not known [1]. Colonies are orange-red, smooth and convex with a regular edge [1]. Multiple carotenoids are present in the cells [1]. The organism reveals the presence of polyphosphate granules which have a delicate granular structure [20]. Optimal growth temperature is 30ºC [1], but the organism is also able to grow at 37 ºC [21]. Growth was observed in media that contained 1% of NaCl [1], but not when the media contained 5% of NaCl [21]. Strain MRP^T is chemoorganotrophic with respiratory metabolism [1]. The organism produces catalase, but not *β*-galactosidase [21], and does not reduce nitrate to nitrite [21]. The reaction was negative for methyl red, Voges-Proskauer, indole and citrate tests [21]. Strain MRP^T does not produce acid from arabinose, galactose, lactose, maltose, manitol, sorbitol, sucrose or xylose [21]. Acid with no gas was produced from glucose or fructose, when the organism was grown on peptone-water basal medium or the basal medium according to subcommittee on taxonomy of staphylococci and micrococci [1,21,22]. Esculin was hydrolyzed by strain MRP^T [21]. The organism was more active in digesting proteins (milk, soya and gelatin) than *D. radiodurans* [1]; milk is peptonised and gelatine is liquefied by strain MRP^T [1]. Strain MRP^T resists to 1.5 Mrad of gamma radiation [1].

Chemotaxonomy

Cell wall of strain MRP^T possesses A3*β* type peptidoglycan [20], with L-ornithine in the peptide subunit and glycine in the interpeptide bridge [1]. The predominant fatty acid component is palmitoleate, whereas branched-chain fatty acids are not present [1]: $C_{16:1}$ (73.0%), $C_{18:1}$ (7.8% $C_{17:1}$ (6.9%), $C_{17:0}$ (4.8%), $C_{16:0}$ (3.7%), $C_{19:1}$ (2.4%), $C_{15:1}$ (0.9%), and trace amount of $C_{14:0,}$ $C_{14:1}$ and $C_{15:0}$ [21]. The fatty acid composition and the cell wall profiles of *D. proteolyticus* are similar to those of *D. radiodurans* and *D. radiophilus* [20,21].

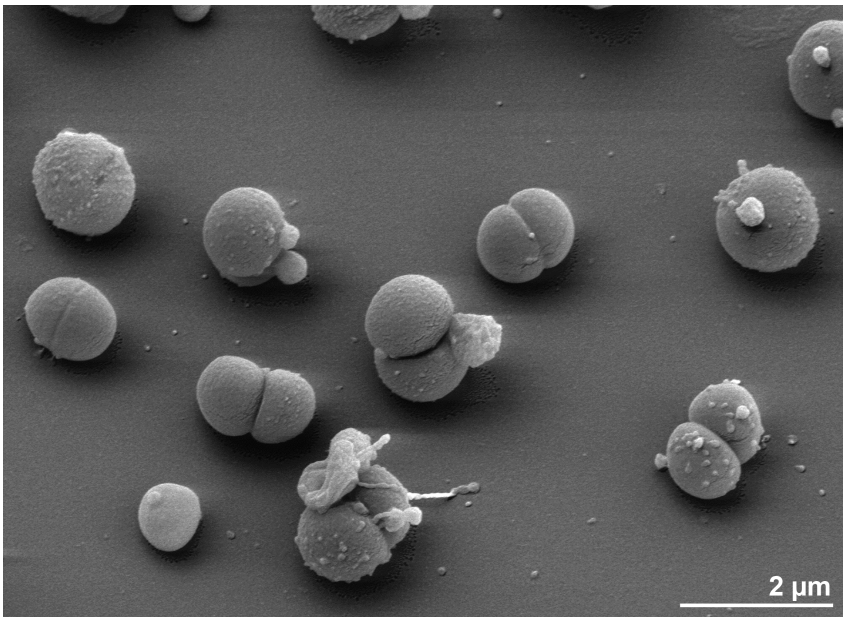

Figure 2. Scanning electron micrograph of *D. proteolyticus* MRP^T

Table 1. Classification and general features of *D. proteolyticus* MRP^T in accordance with the MIGS recommendations [23] and the NamesforLife database [24].

MIGS ID	Property	Term	Evidence code
	Current classification	Domain *Bacteria*	TAS [25]
		Phylum "*Deinococcus-Thermus*"	TAS [26,27]
		Class *Deinococci*	TAS [28,29]
		Order *Deinococcales*	TAS [3]
		Family *Deinococcaceae*	TAS [1,3]
		Genus *Deinococcus*	TAS [1,3]
		Species *Deinococcus proteolyticus*	TAS [1]
		Type strain MRP	TAS [1,30]
	Gram stain	positive	TAS [1]
	Cell shape	spheres; singly, in pairs or tetrads	TAS [1]
	Motility	none	TAS [1]
	Sporulation	none	TAS [1]
	Temperature range	mesophile	TAS [1]
	Optimum temperature	30°C	TAS [1]
	Salinity	1% NaCl	TAS [1]
MIGS-22	Oxygen requirement	strictly aerobic	TAS [1]
	Carbon source	glucose	TAS [1]
	Energy source	chemoorganotroph	TAS [1]
MIGS-6	Habitat	soil, host	TAS [30]
MIGS-15	Biotic relationship	free-living	NAS
MIGS-14	Pathogenicity	none	NAS
	Biosafety level	1	TAS [31]
	Isolation	faeces of *Lama glama*	TAS [30]
MIGS-4	Geographic location	not reported	
MIGS-5	Sample collection time	1973 or before	TAS [30]
MIGS-4.1	Latitude	not reported	

MIGS-4.2	Longitude	not reported
MIGS-4.3	Depth	not reported
MIGS-4.4	Altitude	not reported

Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from of the Gene Ontology project [32].

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [33], and is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project [34]. The genome project is deposited in the Genome On Line Database [19] and the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Three genomic libraries: one 454 pyrosequence standard library, one 454 PE library (8 kb insert size), one Illumina library
MIGS-29	Sequencing platforms	454-GS-FLX-Titanium, Illumina GAii
MIGS-31.2	Sequencing coverage	249.0 × Illumina; 33.0 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3, VELVET version 0.7.63, phrap version SPS - 4.24
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002536
	Genbank Date of Release	October 7, 2011
	GOLD ID	Gc01666
	NCBI project ID	41911
	Database: IMG-GEBA	649633035
MIGS-13	Source material identifier	DSM 20540
	Project relevance	Tree of Life, GEBA

Growth conditions and DNA isolation

D. proteolyticus MRP^T, DSM 20540, was grown in DSMZ medium 53 (Corynebacterium Agar) [36] at 30°C. DNA was isolated from 0.5-1 g of cell paste using MasterPure Gram-positive DNA purification kit (Epicentre MGP04100) following the standard protocol as recommended by the manufacturer, with modification st/DL for cell lysis as described in Wu *et al.* [34]. DNA is available through the DNA Bank Network [45].

Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website [37].

Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 75 contigs in five scaffolds was converted into a phrap [38] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (721.9 Mb) was assembled with Velvet [39] and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. The 454 draft assembly was based on 146.0Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [38] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [37], Dupfinisher [40], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 169 additional reactions and two shatter library were necessary to close gaps and to raise the quality of the finished sequence. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [41]. The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 282.0 × coverage of the genome. The final assembly contained 149,969 pyrosequence and 20,053,100 Illumina reads.

Genome annotation

Genes were identified using Prodigal [42] as part of the Oak Ridge National Laboratory genome annotation pipeline, 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, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [44].

Genome properties

The genome consist of a 2,147,060 bp long chromosome and four large circular plasmids of 315,518 bp, 195,800 bp, 132,270 bp, and 97,188 bp length, respectively, and a G+C content of 65.6% (Table 3 and Figure 3). Of the 2,799 genes predicted, 2,741 were protein-coding genes, and 58 RNAs; 85 pseudogenes were also identified. The majority of the protein-coding genes (65.0%) were assigned a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	2,886,836	100.00
DNA coding region (bp)	2,524,665	87.45
DNA G+C content (bp)	1,894,892	65.64
Number of replicons	5	
Extrachromosomal elements	4	
Total genes	2,799	100.00
RNA genes	58	2.07

rRNA operons	3	
tRNA genes	47	1.68
Protein-coding genes	2,741	97.93
Pseudo genes	85	3.04
Genes with function prediction	1,818	64.95
Genes in paralog clusters	1,029	36.76
Genes assigned to COGs	2,042	72.95
Genes assigned Pfam domains	1,982	70.81
Genes with signal peptides	986	35.23
Genes with transmembrane helices	561	20.04
CRISPR repeats	3	

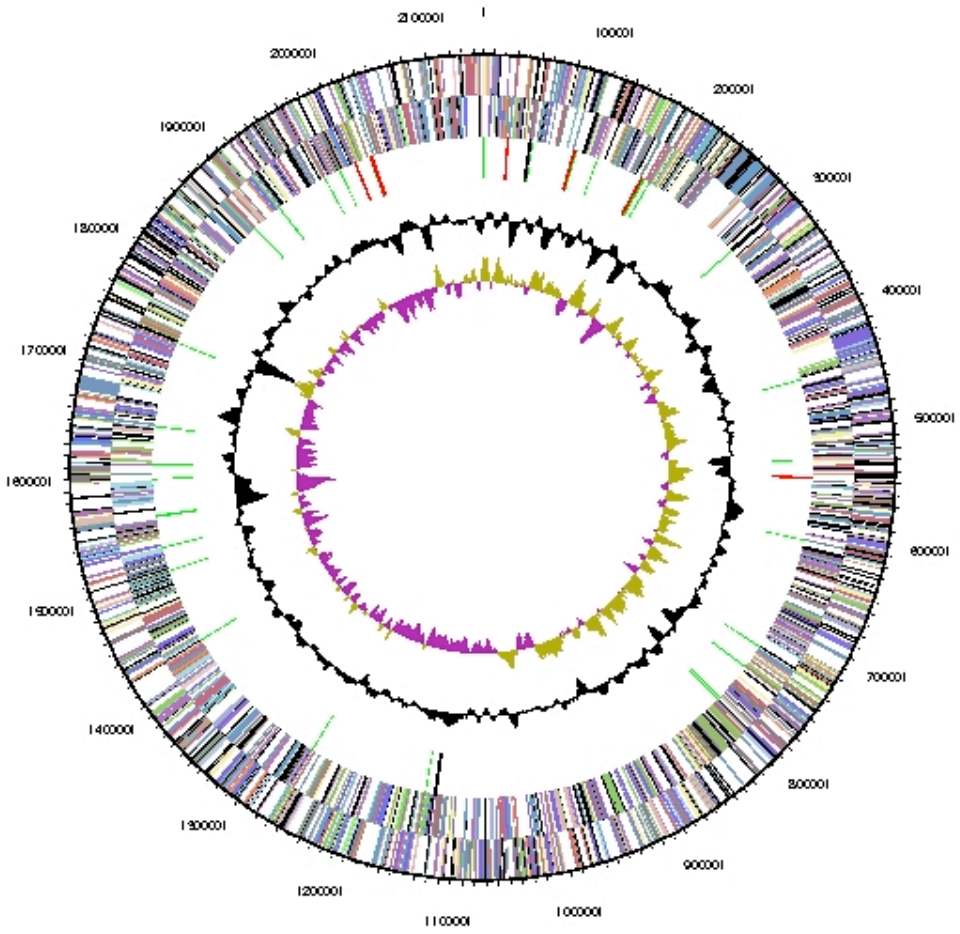

Figure 3. Graphical circular map of the chromosome (plasmids not shown); From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

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

Code	COG counts and percentage of protein-coding genes		Description
	Genome		
	value	% of total	
J	150	6.8	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	134	6.1	Transcription
L	158	7.1	Replication, recombination and repair
B	1	0.1	Chromatin structure and dynamics
D	32	1.5	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	44	2.0	Defense mechanisms
T	93	4.2	Signal transduction mechanisms
M	101	4.6	Cell wall/membrane/envelope biogenesis
N	30	1.4	Cell motility
Z	1	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	50	2.3	Intracellular trafficking, secretion, and vesicular transport
O	95	4.3	Posttranslational modification, protein turnover, chaperones
C	122	5.5	Energy production and conversion
G	105	4.8	Carbohydrate transport and metabolism
E	177	8.0	Amino acid transport and metabolism
F	75	3.4	Nucleotide transport and metabolism
H	109	4.9	Coenzyme transport and metabolism
I	80	3.6	Lipid transport and metabolism
P	119	5.4	Inorganic ion transport and metabolism
Q	38	1.7	Secondary metabolites biosynthesis, transport and catabolism
R	299	13.5	General function prediction only
S	199	9.0	Function unknown
-	757	27.1	Not in COGs

Acknowledgements

We would like to gratefully acknowledge the help of Katja Steenblock (DSMZ) for growing *D. proteolyticus* cultures. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

References

1. Brooks BW, Murray RG. Nomenclature for "*Micrococcus radiodurans*" and other radiation-resistant cocci: *Deinococcaceae* fam. nov. and *Deinococcus* gen. nov., including five species. *Int J Syst Bacteriol* 1981; **31**:353-360.
2. Euzeby JP. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* 1997; **47**:590-592.
3. Rainey FA, Nobre MF, Schumann P, Stackebrandt E, da Costa MS. Phylogenetic diversity of the deinococci as determined by 16S ribosomal DNA sequence comparison. *Int J Syst Bacteriol* 1997; **47**:510-514.
4. Counsell TJ, Murray RGE. Polar lipid profiles of the genus *Deinococcus*. *Int J Syst Bacteriol* 1986; **36**:202-206.
5. Christensen EA, Kristensen H. Radiation-resistance of micro-organisms from air in clean premises. *Acta Pathol Microbiol Scand B* 1981; **89**:293-301.
6. Ito H. Isolation of *Micrococcus radiodurans* occuring in radurized sawdust culture media of mushroom. *Agric Biol Chem* 1977; **41**:35-41.
7. Kristensen H, Christensen EA. Radiation-resistant micro-organisms isolated from textiles. *Acta Pathol Microbiol Scand B* 1981; **89**:303-309.
8. Ito H, Watanabe H, Takehisa M, Izuka H. Isolation and identification of radiation-resistant cocci belonging to the genus *Deinococcus* from sewage sludges and animal feeds. *Agric Biol Chem* 1983; **47**:1239-1247.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-410.
10. Korf I, Yandell M, Bedell J. BLAST. O'Reilly: Sebastopol, CA; 2003.
11. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072.
12. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems* 1980; **14**:130-137.
13. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552.
14. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464.
15. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 2008; **57**:758-771.
16. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc* 2007; **92**:669-674.
17. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How Many Bootstrap Replicates Are Necessary? *Lecture Notes in Computer Science* 2009; **5541**:184-200.
18. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). 4.0 b10. Sunderland: Sinauer Associates; 2002.
19. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2012; **40**:D571-D579.

20. Sleytr UB, Silva MT, Kocur M, Lewis NF. The fine structure of *Micrococcus radiophilus* and *Micrococcus radioproteolyticus*. *Arch Microbiol* 1976; **107**:313-320.
21. Brooks BW, Murray RG, Johnson JL, Stackebrandt E, Woese CR, Fox GE. Red-pigmented micrococci: a basis of taxonomy. *Int J Syst Bacteriol* 1980; **30**:627-646.
22. ICBN. Subcommittee on taxonomy of staphylococci and micrococci: Recommendations. *Int Bull Bacteriol Nomencl Taxon* 1965; **15**:109-110.
23. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV and others. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547.
24. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; **37**:9.
25. 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.
26. Garrity GM, Holt JG. Taxonomic outline of the *Archaea* and *Bacteria*. *In*: Garrity GM, Boone DR, Castenholz RW (*eds*), Bergey's manual of systematic bacteriology. Second Edition, Volume 1, Springer, New York, 2001, p. 155-166.
27. Weisburg WG, Giovannoni SJ, Woese CR. The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst Appl Microbiol* 1989; **11**:128-134.
28. List Editor. Validation List no. 85. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. *Int J Syst Evol Microbiol* 2002; **52**:685-690.
29. Garrity GM, Holt JG. Class I. *Deinococci* class. nov. *In*: Garrity GM, Boone DR, Castenholz RW (*eds*), Bergey's manual of systematic bacteriology. Second Edition, Volume 1, Springer, New York, 2001, p. 395.
30. Kobatake M, Tamabe S, Hasegawa S. Nouveau micrococcus radioresistant à pigment rouge, isolé de feces de *Lama glama*, et son utilisation comme indicateur microbiologique de la radiosterilisation *C R Soc Biol* 1973; **167**:1506-1510.
31. BAuA. Classification of bacteria and archaea in risk groups. TRBA 466. p. 70. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Germany. 2010.
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT and others. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29.
33. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; **33**:175-182.
34. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*. A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea*. *Nature* 2009; **462**:1056-1060.
35. Corrections and Clarifications [Erratum: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.]. Science 2004; **303**:766b.
36. List of growth media used at DSMZ: http//www.dsmz.de/catalogues/catalogue-microorganisms/culture-technology/list-of-media-for-microorganisms.html.
37. DOE Joint Genome Institute. http://www.jgi.doe.gov.
38. Phrap and Phred for Windows. MacOS, Linux and Unix. http://www.phrap.com.
39. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829.

40. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. *In:* Proceeding of the 2006 international conference on bioinformatics & computational biology. Arabnia HR, Valafar H (*eds*), CSREA Press. June 26-29, 2006: 141-146..
41. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltsman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. 2008; Marco Island, FL.
42. 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.
43. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457.
44. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278.
45. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar G, Klenk HP, Güntsch A, Berendsohn WG, Wägele JW. The DNA Bank Network: the start from a German initiative. *Biopreserv Biobank* 2011; **9**:51-55.
46. Makarova KS, Omelchenko MV, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Lapidus A, Copeland A, Kim E, Land M, *et al*. *Deinococcus geothermalis*: the pool of extreme radiation resistance genes shrinks. *PLoS One* 2007; **2**:e955.
47. de Groot A, Dulermo R, Blanchard L, Guérin P, Fernandez B, Vacherie B, Dossat C, Jolivet E, Siguier P, Chandler M, *et al*. Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. *PLoS Genet* 2009; **5**:e1000434.
48. Pukall R, Zeytun A, Lucas S, Lapidus A, Hammon N, Deshpande S, Nolan M, Cheng JF, Pitluck S, Liolios K, *et al*. Complete genome sequence of Deinococcus maricopensis type strain (LB-34T). *Stand Genomic Sci* 2011 **4**:163-172.
49. White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, Haft DH, Gwinn ML, Nelson WC, Richardson DL, *et al*. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science 1999; **286**:1571-1577.
50. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980; **30**:225-420.
51. Mackay MW, al-Bakri GH, Moseley BE. The plasmids of *Deinococcus* spp. and the cloning and restriction mapping of the *D. radiophilus* plasmid pUE1. *Arch Microbiol* 1985; **141**:91-94.
52. Rýznar L, Drásil V. Influence of repeated lyophylisation on the survival of *Deinococcus proteolyticus*, *Micrococcus luteus* and *Escherichia coli*. *Folia Microbiol (Praha)* 1991: **36**:71-74.
53. Gupta RS, Bustard K, Falah M, Singh D. Sequencing of heat shock protein 70 (DnaK) homologs from *Deinococcus proteolyticus* and *Thermomicrobium roseum* and their integration in a protein-based phylogeny of prokaryotes. *J Bacteriol* 1997; **179**:345-357.
54. Bustard K, Gupta RS. The sequences of heat shock protein 40 (DnaJ) homologs provide evidence for a close evolutionary relationship between the *Deinococcus*-thermus group and cyanobacteria. *J Mol Evol* 1997; **45**:193-205.
55. Filipkowski P, Kur J. Identification and properties of the *Deinococcus grandis* and *Deinococcus proteolyticus* single-stranded DNA binding proteins (SSB). *Acta Biochim Pol* 2007; **54**:89-87.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.