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

Alex Copeland¹, Ahmet Zeytun¹,₂, Montri Yassawong³, Matt Nolan¹, Susan Lucas¹, Nancy Hammon¹, Shweta Deshpande¹, Jan-Fang Cheng¹, Cliff Han¹,₂, Roxanne Tapia¹,₂, Lynne A. Goodwin¹,₂, Sam Pitluck¹, Konstantinos Mavromatis¹, Konstantinos Liolios¹, Jo安娜 Pagani¹, Natalia Ivanova¹, Natalia Mikhailova¹, Amrita Pati¹, Amy Chen¹, Krishna Palaniappan², Miriam Land¹,⁵, Loren Hauser¹,₂,⁵, Cynthia D. Jeffries¹,⁵, Evelyne-Marie Brambilla⁶, Manfred Rohde⁷, Johannes Sikorski⁷, Rüdiger Pukall⁷, Markus Göker⁷, John C. Detter¹,₂, Tanja Woyke¹, James Bristow¹, Jonathan A. Eisen¹,⁸, Victor Markowitz⁴, Philip Hugenholtz¹,⁹, Nikos C. Kyrpides¹, Hans-Peter Klenk⁶*, and Alla Lapidus¹

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*, GBEA

*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 MRPT¹ was isolated from feces of *Lama glama* and possesses extreme radiation resistance, a trait is shares with various other species of the genus *Deinococcus*, with *D. proteolyticus* being resistant up to 1.5 Mrad of gamma radiation. Strain MRPT³ 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 lengths 97 kbp, 132 kbp, 196 kbp and 315 kbp harbors 2,741 protein-coding and 58 RNA genes and is a part of the **Genomic Encyclopedia of Bacteria and Archaea** project.

---

**Introduction**

Strain MRPT, 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 epitet is derived from the Neo-Latin word *proteolyticus*, meaning proteolytic [1]. Strain MRPT was isolated in the early 1970s from feces of *Lama glama* by Kobatake et al., and became known under its synonym *Micrococcus radioproteolyticus* [3], 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 [4]. 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 are resistant to high levels of gamma and ultraviolet radiation [1]. Cells of deinococci are spherical or rod shaped [5]. Several distinct cell wall layers have been observed in thin sections 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 [6]. Plasmids of strain MRPT were previously analyzed by Mackay et al. [7], and survival of repeated lyophilisation was studied by Rýznar and Drásil [8]. The Genes hsp70 [9], hps40 [10], and SSB genes were sequenced [11], primarily for phylogenetic analyses. The members of the genus Deinococcus have been isolated from diverse environments [12-15], usually selected and characterized by survival after high-dose irradiation [6]. 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 MRPT, together with the description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of D. proteolyticus MRPT was compared using NCBI BLAST [16,17] 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 [18] and the relative frequencies of taxa and keywords (reduced to their stem [19]) 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).

Strain MRPT 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 divided into two planes to form tetrad 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 [41]. Optimal growth temperature is 30ºC [1], but the organism is also able to grow at 37 ºC [42]. Growth was observed in media that contained 1% of NaCl [1], but not when the media contained 5% of NaCl [42]. Strain MRPT is chemoorganotrophic with respiratory metabolism [1]. The organism produces catalase, but not β-galactosidase [42], and does not reduce nitrate to nitrite [42]. The reaction was negative for methyl red, Voges-Proskauer, indole and citrate tests [42]. Strain MRPT does not produce acid from arabinose, galactose, lactose, maltose, manitol, sorbitol, sucrose or xylose [42]. 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 [1,42,43]. Esculin was hydrolyzed by strain MRPT [42]. The organism was more active in digesting proteins (milk, soya and gelatin) than D. radiodurans [1]; milk is peptonized and gelatine is liquefied by strain MRPT [1]. Strain MRPT is resistant to to gamma radiation up to 1.5 Mrad. [1].
Figure 1. Phylogenetic tree highlighting the position of *D. proteolyticus* relative to the type strains of the other species within the family *Deinococccaceae*. The tree was inferred from 1,377 aligned characters [20,21] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [22]. Rooting was done initially using the midpoint method [23] 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 [24] (left) and from 1,000 maximum parsimony bootstrap replicates [25] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [26] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks [27-29]. The genome of *D. radiodurans* published by White et al. in 1999 [30] later turned out not to be from the type strain [31].
Chemotaxonomy
The cell wall of strain MRPT contains A3β type peptidoglycan [41], with L-ornithine in the peptide subunit and glycine in the interpeptide bridge [1]. The predominant fatty acid is palmmitoleate, whereas branched-chain fatty acids are present in minor amounts only [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 amounts of C_{14:0}, C_{14:1} and C_{15:0} [42]. The fatty acid composition and the cell wall profiles of 
\textit{D. proteolyticus} are similar to those of 
\textit{D. radiodurans} and 
\textit{D. radiophilus} [41,42].

Genome sequencing and annotation
Genome project history
This organism was selected for sequencing on the basis of its phylogenetic position [44], and is part of the \textit{Genomic Encyclopedia of Bacteria and Archaea} project [45]. The genome project is deposited in the Genome On Line Database [26] 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.

Growth conditions and DNA isolation
\textit{D. proteolyticus} MRPT, DSM 20540, was grown in DSMZ medium 53 (\textit{Corynebacterium Agar}) [46] 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. [45]. DNA is available through the DNA Bank Network [47].

Genome sequencing and assembly
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 [48]. 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 [49] 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 [50] 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 1350 -g

Figure 2. Scanning electron micrograph of \textit{D. proteolyticus} MRPT

http://standardsingenomics.org
Deinococcus proteolyticus type strain (MRPT)

-m -ml 20. The Phred/Phrap/Consed software package [49] 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 [48], Dupfinisher [51], 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 libraries 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 [52]. 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.

Table 1. Classification and general features of *D. proteolyticus* MRPT according to the MIGS recommendations [32] and the NamesforLife database [33].

| MIGS ID | Property                      | Term                                      | Evidence code |
|---------|-------------------------------|-------------------------------------------|---------------|
|         | **Current classification**    |                                           |               |
|         | Domain                        | *Bacteria*                                | TAS [34]      |
|         | Phylum                        | "*Deinococcus-Thermus*"                    | TAS [35,36]   |
|         | Class                         | *Deinococcus*                             | TAS [37,38]   |
|         | Order                         | *Deinococcales*                           | TAS [5]       |
|         | Family                        | *Deinococcaceae*                          | TAS [1,5]     |
|         | Genus                         | *Deinococcus*                             | TAS [1,5]     |
|         | Species                       | *Deinococcus proteolyticus*               | TAS [1]       |
|         | Type strain                   | MRP                                        | TAS [1]       |
|         | 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 [3]       |
| MIGS-15 | Biotic relationship           | free-living                                | NAS           |
| MIGS-14 | Pathogenicity                 | none                                       | NAS           |
|         | Biosafety level               | 1                                          | TAS [39]      |
|         | Isolation                     | feces of *Lama glama*                     | TAS [3]       |
| MIGS-4  | Geographic location           | not reported                               |               |
| MIGS-5  | Sample collection time        | 1973 or before                             | TAS [3]       |
| 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 – IDA: Inferred from Direct Assay (first time in publication); 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 the Gene Ontology project [40]. If the evidence code is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.
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 GAi                                    |
| 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 (chromosome)  CP002537-40 (pDEIRP01-04)
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

Genome annotation
Genes were identified using Prodigal [53] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [54]. 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 [55].

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, 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       |
Deinococcus proteolyticus type strain (MRPT)

Figure 3. Graphical circular map of the chromosome (plasmids not shown, but accessible through the JGI web pages [48]); 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 (black), GC skew (purple/olive).
Table 4. Number of genes associated with the general COG functional categories

| Code | Value | %age  | Description                                           |
|------|-------|-------|-------------------------------------------------------|
| 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 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. [http://dx.doi.org/10.1099/00207713-31-3-353](http://dx.doi.org/10.1099/00207713-31-3-353)

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. [http://dx.doi.org/10.1099/00207713-47-2-590](http://dx.doi.org/10.1099/00207713-47-2-590)

3. 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.

4. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980; 30:225-247.
Deinococcus proteolyticus type strain (MRPT)

420. http://dx.doi.org/10.1099/00207713-30-1-225

5. 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. http://dx.doi.org/10.1099/00207713-47-2-510

6. Counsell TJ, Murray RGE. Polar lipid profiles of the genus Deinococcus. Int J Syst Bacteriol 1986; 36:202-206. http://dx.doi.org/10.1099/00207713-36-2-202

7. 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. http://dx.doi.org/10.1007/BF00446746

8. 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. http://dx.doi.org/10.1007/BF02935825

9. Gupta RS, Bustard K, Falah M, Singh D. Sequencing of heat shock protein 70 (Dnak) homologs from Deinococcus proteolyticus and Thermococcus roseum and their integration in a protein-based phylogeny of prokaryotes. J Bacteriol 1997; 179:345-357.

10. Bustard K, Gupta RS. The sequences of heat shock protein 40 (Dnal) homologs provide evidence for a close evolutionary relationship between the Deinococcus-thermus group and cyanobacteria. J Mol Evol 1997; 45:193-205. http://dx.doi.org/10.1007/PL00006219

11. Filipkowski P, Kur J. Identification and properties of the Deinococcus grandis and Deinococcus proteolyticus single-stranded DNA binding proteins (SSB). Acta Biochim Pol 2007; 54:79-87.

12. Christensen EA, Kristensen H. Radiation-resistance of micro-organisms from air in clean premises. Acta Pathol Microbiol Scand [B] 1981; 89:293-301.

13. Ito H. Isolation of Micrococcus radiodurans occurring in radurized sawdust culture media of mushroom. Agric Biol Chem 1977; 41:35-41. http://dx.doi.org/10.1271/bbb1961.41.35

14. Kristensen H, Christensen EA. Radiation-resistant micro-organisms isolated from textiles. Acta Pathol Microbiol Scand [B] 1981; 89:303-309.

15. 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. http://dx.doi.org/10.1271/bbb1961.47.1239

16. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-410.

17. Korf I, Yandell M, Bedell J. BLAST. O’Reilly: Sebastopol, CA; 2003.

18. 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. http://dx.doi.org/10.1128/AEM.03006-05

19. Porter MF. An algorithm for suffix stripping. Program: electronic library and information systems 1980; 14:130-137.

20. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 2000; 17:540-552. http://dx.doi.org/10.1093/oxfordjournals.molev.a026334

21. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. Bioinformatics 2002; 18:452-464. http://dx.doi.org/10.1093/bioinformatics/18.3.452

22. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol 2006; 55:472-487. http://dx.doi.org/10.1080/10635150500379734

23. Christensen EA, Kristensen H. Radiation-resistance of micro-organisms from air in clean premises. Acta Pathol Microbiol Scand [B] 1981; 89:293-301.

24. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? Lect Notes Comput Sci 2009; 5541:184-200. http://dx.doi.org/10.1007/978-3-642-02008-7_13

25. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). 4.0 b10. Sunderland: Sinauer Associates; 2002.

26. 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. [http://dx.doi.org/10.1093/nar/gkr1100](http://dx.doi.org/10.1093/nar/gkr1100)

27. Makarova KS, Omelchenko MV, Gaidamakov EA, Matrosovsa YY, Vasilenko A, Zhai M, Lapidus A, Copeland A, Kim E, Land M, et al. *Deinococcus geothermals*: the pool of extreme radiation resistance genes shrinks. *PLoS ONE* 2007; 2:e955. [http://dx.doi.org/10.1371/journal.pone.0009055](http://dx.doi.org/10.1371/journal.pone.0009055)

28. de Groot A, Dulermo R, Blanchard L, Guérin P, Fernandez B, Vacherie B, Dossat C, Jolivet E, Siguiér P, Chandler M, et al. Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. *PLoS Genet* 2009; 5:e1000434. [http://dx.doi.org/10.1371/journal.pgen.1000434](http://dx.doi.org/10.1371/journal.pgen.1000434)

29. Pukall R, Zeytun A, Lucas S, Lapidus A, Hammon N, Deshpande S, Nolan M, Cheng JF, Pitluck S, Liliolos K, et al. Complete genome sequence of *Deinococcus maricopenisis* type strain (LB-344). *Stand Genomic Sci* 2011; 4:163-172. [http://dx.doi.org/10.4056/sigs.1633949](http://dx.doi.org/10.4056/sigs.1633949)

30. 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. [http://dx.doi.org/10.1126/science.286.5444.1571](http://dx.doi.org/10.1126/science.286.5444.1571)

31. Corrections and Clarifications [Erratum: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.] *Science* 2004; 303:766b. [http://dx.doi.org/10.1126/science.303.5659.766b](http://dx.doi.org/10.1126/science.303.5659.766b)

32. 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. [http://dx.doi.org/10.1038/nbt1360](http://dx.doi.org/10.1038/nbt1360)

33. Garrity G. *NamesforLife*. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; 37-9.

34. 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. [http://dx.doi.org/10.1073/pnas.87.12.4576](http://dx.doi.org/10.1073/pnas.87.12.4576)

35. 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.

36. 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. [http://dx.doi.org/10.1016/S0723-2674(09)80051-7](http://dx.doi.org/10.1016/S0723-2674(09)80051-7)

37. 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. [http://dx.doi.org/10.1099/ijs.0.02358-0](http://dx.doi.org/10.1099/ijs.0.02358-0)

38. 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.

39. BAuA. Classification of bacteria and archaea in risk groups. TRBA 466. p. 70. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Germany. 2010.

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

41. Sleytr UB, Silva MT, Kocur M, Lewis NF. The fine structure of *Micrococcus radiophilus* and *Micrococcus radioproteolyticus*. *Arch Microbiol* 1976; 107:313-320. [http://dx.doi.org/10.1007/BF00425346](http://dx.doi.org/10.1007/BF00425346)

42. Brooks BW, Murray RGE, Johnson JL, Stackebrandt E, Woese CR, Fox GE. Red-pigmented micrococci: a basis for taxonomy. *Int J Syst Bacteriol* 1980; 30:627-646. [http://dx.doi.org/10.1099/00207713-30-4-627](http://dx.doi.org/10.1099/00207713-30-4-627)

43. ICBN. Subcommittee on taxonomy of staphylococci and micrococci: Recommendations. *Int Bull Bacteriol Nomencl Taxon* 1965; 15:109-110. [http://dx.doi.org/10.1099/00207713-15-2-109](http://dx.doi.org/10.1099/00207713-15-2-109)

44. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*. *Syst Appl Microbiol* 2010; 33:175-182. [http://dx.doi.org/10.1016/j.syapm.2010.03.003](http://dx.doi.org/10.1016/j.syapm.2010.03.003)
Deinococcus proteolyticus type strain (MRPT)

45. 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. http://dx.doi.org/10.1038/nature08656

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

47. Gemeinholzer B, Dröge G, Zetsche 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. http://dx.doi.org/10.1089/bio.2010.0029

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

49. Phrap and Phred for Windows. MacOS, Linux and Unix. http://www.phrap.com.

50. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008; 18:821-829. http://dx.doi.org/10.1101/gr.074492.107

51. 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.

52. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltzman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. 2008; Marco Island, FL.

53. Hyatt D, Chen GL, LoCascio PF, Land ML, Larkin FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010; 11:119. http://dx.doi.org/10.1186/1471-2105-11-119

54. 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. http://dx.doi.org/10.1038/nmeth.1457

55. 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. http://dx.doi.org/10.1093/bioinformatics/btp393