

# UC Davis

## UC Davis Previously Published Works

### Title

Complete genome sequence of Nocardioopsis dassonvillei type strain (IMRU 509T)

### Permalink

<https://escholarship.org/uc/item/5z0262q5>

### Journal

Environmental Microbiome, 3(3)

### ISSN

2524-6372

### Authors

Sun, Hui

Lapidus, Alla

Nolan, Matt

et al.

### Publication Date

2010-11-01

### DOI

10.4056/sigs.1363462

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

## Complete genome sequence of *Nocardiosis dassonvillei* type strain (IMRU 509<sup>T</sup>)

Hui Sun<sup>1</sup>, Alla Lapidus<sup>1</sup>, Matt Nolan<sup>1</sup>, Susan Lucas<sup>1</sup>, Tijana Glavina Del Rio<sup>1</sup>, Hope Tice<sup>1</sup>, Jan-Fang Cheng<sup>1</sup>, Roxane Tapia<sup>1,2</sup>, Cliff Han<sup>1,2</sup>, Lynne Goodwin<sup>1,2</sup>, Sam Pitluck<sup>1</sup>, Ioanna Pagani<sup>1</sup>, Natalia Ivanova<sup>1</sup>, Konstantinos Mavromatis<sup>1</sup>, Natalia Mikhailova<sup>1</sup>, Amrita Pati<sup>1</sup>, Amy Chen<sup>3</sup>, Krishna Palaniappan<sup>3</sup>, Miriam Land<sup>1,4</sup>, Loren Hauser<sup>1,4</sup>, Yun-Juan Chang<sup>1,4</sup>, Cynthia D. Jeffries<sup>1,4</sup>, Olivier Duplex Ngatchou Djao<sup>5</sup>, Manfred Rohde<sup>5</sup>, Johannes Sikorski<sup>6</sup>, Markus Göker<sup>6</sup>, Tanja Woyke<sup>1</sup>, James Bristow<sup>1</sup>, Jonathan A. Eisen<sup>1,7</sup>, Victor Markowitz<sup>3</sup>, Philip Hugenholtz<sup>1</sup>, Nikos C. Kyrpides<sup>1</sup>, and Hans-Peter Klenk<sup>6\*</sup>

<sup>1</sup> DOE Joint Genome Institute, Walnut Creek, California, USA

<sup>2</sup> Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

<sup>3</sup> Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>4</sup> Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

<sup>5</sup> HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany

<sup>6</sup> DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

<sup>7</sup> University of California Davis Genome Center, Davis, California, USA

\*Corresponding author: Hans-Peter Klenk

**Keywords:** Gram-positive, aerobic, pathogen, mesophile, non alkaliphilic, zig-zag-shaped mycelium, actinomycetoma, conjunctivitis, cholangitis, *Nocardiosaceae*, GEBA

---

*Nocardiosis dassonvillei* (Brocq-Rousseau 1904) Meyer 1976 is the type species of the genus *Nocardiosis*, which in turn is the type genus of the family *Nocardiosaceae*. This species is of interest because of its ecological versatility. Members of *N. dassonvillei* have been isolated from a large variety of natural habitats such as soil and marine sediments, from different plant and animal materials as well as from human patients. Moreover, representatives of the genus *Nocardiosis* participate actively in biopolymer degradation. This is the first complete genome sequence in the family *Nocardiosaceae*. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 6,543,312 bp long genome consist of a 5.77 Mbp chromosome and a 0.78 Mbp plasmid and with its 5,570 protein-coding and 77 RNA genes is a part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

---

### Introduction

Strain IMRU 509<sup>T</sup> (= DSM 43111 = ATCC 23218 = JCM 7437) is the type strain of *Nocardiosis dassonvillei*, which in turn is the type species of the genus *Nocardiosis*. Currently, *N. dassonvillei* is one of 40 validly published species belonging to the genus. The genus name derives from the Greek name *opsis*, appearance, and from Edmond Nocard, who first described in 1888 the type species of the genus *Nocardia*, *N. farcinica* [1,2]. *Nocardiosis* means “that which has the appearance

of *Nocardia*”. The species epithet is chosen in honor of Charles Dassonville, a contemporary French veterinarian [3]. The genus *Nocardiosis* was first described by Meyer in 1976 [4] for bacteria that were previously classified as either *Streptothrix dassonvillei* (Brocq-Rousseau 1904) [3], *Nocardia dassonvillei* [5], or *Actinomadura dassonvillei* [6] on the basis of their morphological characteristics and cell wall type [4]. The strain IMRU 509<sup>T</sup> is the neotype of the species *N.*

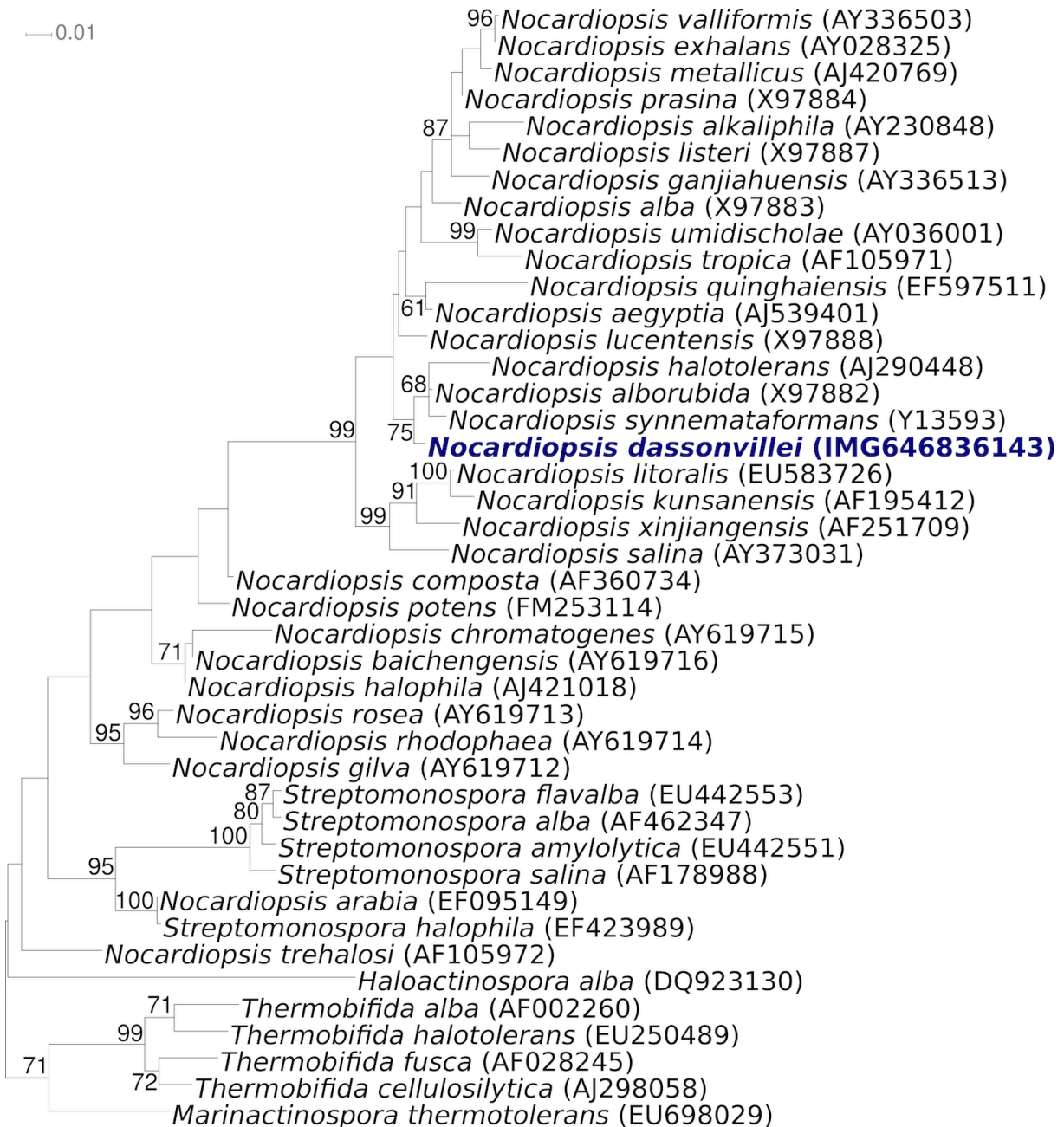
*dassonvillei* (Brocq-Rousseau 1904). Databases provide contradictory speculations on the ecological and geographical origin of strain IMRU 509<sup>T</sup> (e.g., soil from Paris, France; mildewed grain of unspecified geographical origin), however, solid information could not be extracted from the original literature [4,5,7-9]. Members of this species can be isolated from a variety of different habitats, including mildewed grain and fodder [3], different soils [10-13], antarctic glacier [14], marine sediments [10,15], actinoryzal plant rhizosphere [16], gut tract of animals [17], active stalactites [18], cotton waste and occasionally in hay [19], air of a cattle barn [20], atmosphere of a composting facility [21], salterns [22] and from patients suffering from conjunctivitis [23] or cholangitis [8]. *N. dassonvillei* strains were also isolated from nodules and draining sinuses associated with an actinomycetoma of the anterior aspect of the right leg below the knee of a 39-year-old man [24]. A microorganism identical to *Streptothrix dassonvillei* was isolated two years later, but was placed in the genus *Nocardia* and designated *N. dassonvillei* [23]. Subsequently, the genus *Actinomadura* was described to harbor, among other species, also *N. dassonvillei* (Brocq-Rousseau) Liegard and Landrieu [4,8]. Further analysis supplied evidence that *A. dassonvillei* is not related to nocardiae [7]. Therefore, a new genus was created for *A. dassonvillei* on the basis of the characteristic development of spores, including the specific zig-zag formation of aerial hyphae before spore dispersal and the lack of madurose [4]. In 1976, *A. dassonvillei* was transferred to this new genus and was designated *Nocardioopsis dassonvillei* [4]. Also, *N. dassonvillei* is an earlier heterotypic synonym of *N. alborubida* [25]. The species epithet *alborubida* was considered as orthographically incorrect and corrected by Evtushenko to *albirubida* [10]. Subsequently, the species *N. dassonvillei* has been divided into three subspecies, namely subsp. *prasina* [26], subsp. *albirubida* (Grund and Kroppensstedt 1990) [10] and subsp. *dassonvillei* (Brocq-Rousseau 1904) [4,27], which is an earlier heterotypic synonym of *Streptomyces flavidofuscus* Preobrazhenskaya 1986 [28]. DNA-DNA hybridization data, as well as the results of biochemical tests, indicated that *N. alborubida* DSM 40465, *N. antarctica* DSM 43884, and *N. dassonvillei* DSM 43111 represent a single species designated *N. dassonvillei* [25]. Here we present a summary

classification and a set of features for *N. dassonvillei* strain IMRU 509<sup>T</sup>, together with the description of the complete genomic sequencing and annotation.

## Classification and features

The 16S rRNA gene sequences of the strain IMRU 509<sup>T</sup> share 95.9 to 99.5% sequence similarity with the 16S rRNA gene sequences of the type strains from the other members of the genus *Nocardioopsis* [29]. The 16S rRNA gene of the strain IMRU 509<sup>T</sup> also shares 99% similarity with an uncultured 16S rRNA gene sequence of the clone AKIW919 from urban aerosol in USA [30], but none of the sequences in metagenomic libraries (env\_nt) shares more than 89% sequence identity, indicating that members of the species, genus and even family are poorly represented in the habitats screened thus far (as of November 2010). A representative genomic 16S rRNA sequence of *N. dassonvillei* was compared with the most recent release of the Greengenes database [31] using NCBI BLAST under default values and the relative frequencies of taxa and keywords, weighted by BLAST scores, were determined. The three most frequent genera were *Nocardioopsis* (91.1%), *Streptomyces* (7.1%) and *Prauseria* (1.8%). The species yielding the highest score was *N. dassonvillei* (including hits to *N. dassonvillei* subsp. *dassonvillei*, formerly also known as *Streptomyces flavidofuscus* [9,28]). The five most frequent keywords within the labels of environmental samples which yielded hits were 'soil(s)' (15.4%), 'algeria, nocardioopsis, saccharothrix, saharan' (5.7%), 'source' (2.0%) and 'alkaline' (2.0%). These keywords fit to the morphology of the type strain as well as to the ecology of habitats from which the type strain and also other members of the species were isolated. The single most frequent keyword within the labels of environmental samples which yielded hits of a higher score than the highest scoring species was 'desert/soil' (50.0%).

Figure 1 shows the phylogenetic neighborhood of *N. dassonvillei* strain IMRU 509<sup>T</sup> in a 16S rRNA based tree. The sequences of the five 16S rRNA gene copies in the genome differ from each other by up to ten nucleotides, and differ by up to eight nucleotides from the previously published 16S rRNA sequence (X97886).



**Figure 1.** Phylogenetic tree highlighting the position of *N. dassonvillei* strain IMRU 509<sup>T</sup> relative to the type strains of the other species within the genus and to the type strains of the other genera within the family *Nocardiothermobifidaceae*. The trees were inferred from 1,442 aligned characters [32,33] of the 16S rRNA gene sequence under the maximum likelihood criterion [34] and rooted in accordance with the current taxonomy [35]. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 750 bootstrap replicates [36] if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [37] are shown in blue, published genomes in bold [38]. Note that the tree is more in accordance with the view of Grund and Kroppenstedt (1990) [39] to treat *N. alborubida* as a species of its own, rather than with the view of Yassin *et al.* (1997) [25] and Evtushenko *et al.* 2000 [10] to regard it as a subspecies of *N. dassonvillei* based on a 71% DDH value [10].

**Table 1.** Classification and general features of *N. dassonvillei* strain IMRU 509<sup>T</sup> according to the MIGS recommendations [40]

| MIGS ID  | Property               | Term   | Evidence code           |         |
|----------|------------------------|--|-------------------------|---------|
|          |                        | Domain <i>Bacteria</i>   | TAS [41]                |         |
|          |                        | Phylum <i>Actinobacteria</i>   | TAS [42]                |         |
|          |                        | Class <i>Actinobacteria</i>  | TAS [43]                |         |
|          | Current classification | Order <i>Actinomycetales</i>   | TAS [43-46]             |         |
|          |                        | Family <i>Nocardiopsaceae</i>  | TAS [9,46,47]           |         |
|          |                        | Genus <i>Nocardiopsis</i>  | TAS [4,48]              |         |
|          |                        | Species <i>Nocardiopsis dassonvillei</i>   | TAS [4,48]              |         |
|          |                        | Type strain IMRU 509   | TAS [4]                 |         |
|          |                        | Gram stain   | positive                | TAS [4] |
|          |                        | Cell shape   | zig-zag shaped mycelium | NAS [4] |
|          | Motility               | none   | NAS                     |         |
|          | Sporulation            | yes  | TAS [4]                 |         |
|          | Temperature range      | mesophile, up to 42°C, but not 45°C  | TAS [9]                 |         |
|          | Optimum temperature    | about 28°C   | IDA                     |         |
|          | Salinity               | probably 0% NaCl, no growth at 20% NaCl  | NAS                     |         |
| MIGS-22  | Oxygen requirement     | aerobic  | TAS [4]                 |         |
|          | Carbon source          | carbohydrates  | TAS [4]                 |         |
|          | Energy source          | L-arabinose, D-xylose, D-mannose, D-glucose, L-rhamnose, maltose, lactose, adonitol, dulcitol, D-mannitol, i-inositol, D-fructose, sucrose, raffinose, and glycerol. | TAS [4]                 |         |
| MIGS-6   | Habitat                | soil, mildewed grain, and clinical materials of animal and human origin  | TAS [4]                 |         |
| MIGS-15  | Biotic relationship    | free-living  | NAS                     |         |
| MIGS-14  | Pathogenicity          | actinomycetoma, conjunctivitis, cholangitis  | TAS [8,23,24]           |         |
|          | Biosafety level        | 2  | TAS [49]                |         |
|          | Isolation              | not clearly reported   | NAS                     |         |
| MIGS-4   | Geographic location    | probably Paris, France   | NAS                     |         |
| MIGS-5   | Sample collection time | not reported   | TAS                     |         |
| MIGS-4.1 | Latitude               | 48.85  | NAS                     |         |
| MIGS-4.2 | Longitude              | 2.35 or other (see text)   | NAS                     |         |
| MIGS-4.3 | Depth                  | not reported   | NAS                     |         |
| MIGS-4.4 | Altitude               | not reported   | NAS                     |         |

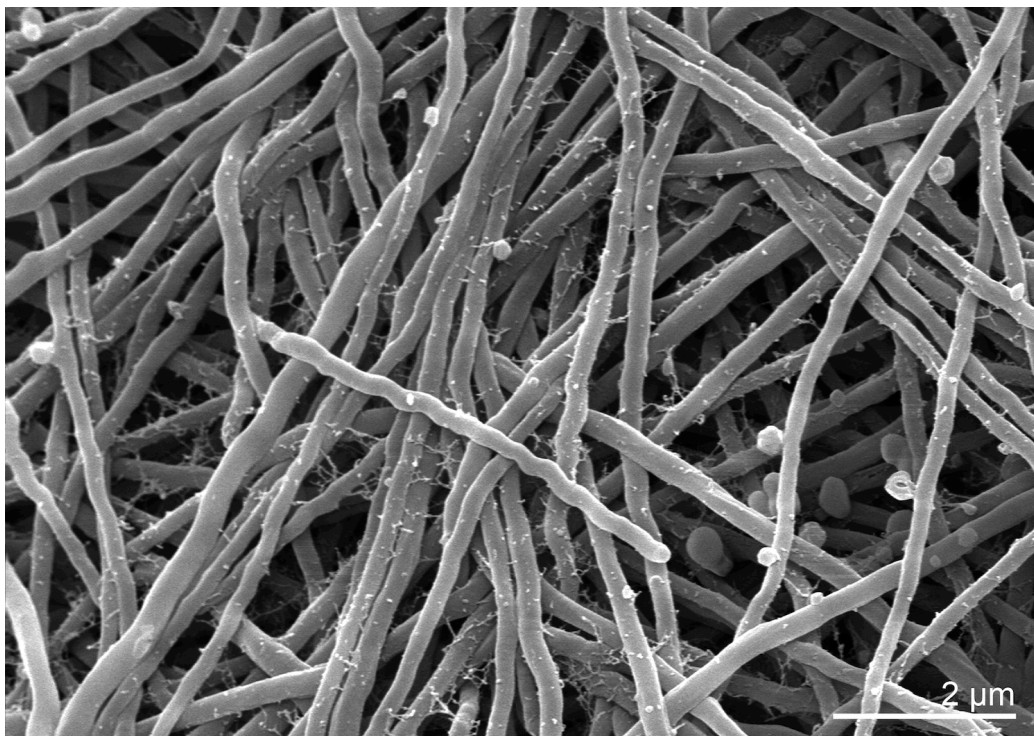
Evidence codes - IDA: Inferred from Direct Assay (first time in publication); TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from of the Gene Ontology project [50]. If the evidence code is IDA, then the property was directly observed by one of the authors or an expert mentioned in the acknowledgements.

The cells of strain IMRU 509<sup>T</sup> are aerobic and Gram-positive [4]. (Table 1). Aerial mycelia are long, moderately branched, and, at the beginning of sporulation, more or less zig-zag-shaped (Figure 2). Later, the hyphae are straight or somewhat coiled [4]. They then divide into long segments which subsequently subdivide into smaller spores of irregular size [4]. Spores are elongated and smooth.

Depending on the medium used, the color of the substrate mycelium is either yellowish-brown or olive to dark brown [4]. The aerial mycelium varies from a sparse coating to a thick, farinaceous to woolly cover of the colonies on oatmeal agar, oatmeal-nitrate agar, Bennett agar, Czapek-sucrose agar, inorganic salt-starch agar, yeast extract-malt extract agar, and complex organic medium 79 [4] of

Prauser and Falta [51]. The color of the aerial mycelium is white or yellowish to grayish [4]. Colonies of substrate mycelia have dense filamentous margins [4]. Hyphae of the substrate mycelium fragment into coccoid elements after 3 to 4 weeks, depending on the medium used [4]. Soluble pigment is not produced [4]. Melanoid pigments are not produced on ISP 6 or tyrosine agar [4]. Growth of strain IMRU 509<sup>T</sup> was tested on basal medium with and without carbohydrates. No growth was detected in the absence of carbohydrates. Strain IMRU 509<sup>T</sup> was able to use N-acetyl-D-glucosamine, *p*-arbutin, D-galactose, gluconate, D-maltose, D-ribose, salicin, D-threulose, maltitol, putrescine, 4-aminobutyrate, azelate, citrate, fumarate, DL-lactate, L-alanine,  $\beta$ -alanine, L-aspartate, L-leucine and phenylacetate as sole carbon sources, but not  $\alpha$ -D-melibiose, acetate, propionate, glutarate, L-malate, mesaconate, oxoglutarate, pyruvate, suberate, L-histidine, L-phenylalanine, L-proline, L-serine, L-tryptophan and 4-hydroxybenzoate [52]. However, L-arabinose, D-xylose, D-mannose, D-glucose, L-rhamnose, maltose, D-mannitol, D-fructose, sucrose and glycerol are the main carbohydrates used [4]. Acid is produced from L-arabinose, galactose, mannitol, sucrose and D-

xylose [8]. Moreover, adonitol, dulcitol, *i*-inositol are not utilized [4]. L-alanine, proline and serine are also used as sole carbon as well as nitrogen sources, although proline and serine are weakly utilized [25]. Strain IMRU 509<sup>T</sup> was found to hydrolyze *p*-nitrophenyl  $\alpha$ -D-glucopyranoside, *p*-nitrophenyl  $\beta$ -D-glucopyranoside, *p*-nitrophenyl phenylphosphonate and L-alanine *p*-nitroanilide, but not aesculin, bis-*p*-nitrophenyl phosphate, *p*-nitrophenyl phosphorylcholine, L-glutamate- $\gamma$ -3-carboxy-*p*-nitroanilide and L-proline *p*-nitroanilide [52]. Meyer (1976) reported that IMRU 509<sup>T</sup> was not able to liquefy gelatin [4], while Yassin *et al.* reported in 1997 the opposite [25]. Strain IMRU 509<sup>T</sup> is able to hydrolyze starch, to peptonize milk, to decompose esculin and to reduce nitrate to nitrite [4]. Strains of *N. dasonvillei* show positive tests of the decarboxylation of lactate, oxalate and propionate [8]. They also decompose casein, tyrosine and Tween 85. They show optimal growth at mildly alkaline conditions of pH 8, and at a salinity of 0% NaCl [8]. No growth is observed at 20% NaCl or at 45°C [8]. The catalase test is positive [4]. Strain IMRU 509<sup>T</sup> hydrolyses adenine, xanthine and hypoxanthine [25].



**Figure 2.** Scanning electron micrograph of *N. dasonvillei* strain IMRU 509<sup>T</sup>

## Chemotaxonomy

The cell wall of the strain IMRU 509<sup>T</sup> belongs to the chemotype III, which corresponds to the peptidoglycan type A1 $\gamma$  [53], i.e., N-acetyl-muramic acid, N-acetyl-glucosamine, alanine, glutamic acid, and *meso*-2, 6-diaminopimelic acid [4,8]. The products of the degradation of the cell wall are glycerol and glucose [8]. Strain IMRU 509<sup>T</sup> is susceptible to lysozyme [4]. The polar lipids found in strain IMRU 509<sup>T</sup> are phosphatidylinositol mannosides (PIM), phosphatidylinositol (PI), phosphatidylcholine (PC), monomannosyl diglyceride (MDG), phosphatidylglycerol (PG), phosphatidylmethylethanolamine (PME), monoacetylated glucose (AG), diphosphatidyl-glycerol (DPG), unknown phospholipids specific for *Nocardiosis*,  $\beta$ -lipids of unknown structure (PL) [8]. The menaquinone type 4C2 was detected [8]. The menaquinone patterns of the strain IMRU 509<sup>T</sup> contain menaquinones from MK-10 to MK-10(H<sub>8</sub>) and sugar type C [8]. Small amounts of the MK-9 and /or MK-12 series are also found [8].

The main fatty acids detected in the strain IMRU509<sup>T</sup> were, *iso*-C<sub>16:0</sub> (26.7%), *anteiso*-C<sub>17:0</sub> (19.8%) and C<sub>18:1</sub> (18.3%). Minor fatty acids detected included C<sub>18:0</sub> (5.8%), C<sub>17:1</sub> (5.2%), *anteiso*-C<sub>15:0</sub> (3.2%), C<sub>16:0</sub> (2.2%), *iso*-C<sub>17:0</sub> (2.1%), C<sub>16:1</sub> (1.2%) and *iso*-C<sub>15:0</sub> (0.8%) [10].

## Genome sequencing and annotation

### Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [54], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [55]. The genome project is deposited in the Genome OnLine Database [37] 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             | Genomic libraries: one Sanger 8 kb pMCL200 library, one fosmid (pcc1Fos) library, one 454 pyrosequence standard library and one Illumina standard library |
| MIGS-29   | Sequencing platforms       | ABI3730, 454 GS FLX Titanium, Illumina GAii   |
| MIGS-31.2 | Sequencing coverage        | 9.0 × Sanger; 19.8 × pyrosequence; 22.3 × Illumina  |
| MIGS-30   | Assemblers                 | Newbler version 1.1.03.24, phrap  |
| MIGS-32   | Gene calling method        | Prodigal 1.4, GenePRIMP   |
|           | INSDC ID                   | CPO02040 chromosome<br>CPO02041 plasmid   |
|           | Genbank Date of Release    | June 4, 2010  |
|           | GOLD ID                    | Gc01339   |
|           | NCBI project ID            | 19709   |
|           | Database: IMG-GEBA         | 646564557   |
| MIGS-13   | Source material identifier | DSM 43111   |
|           | Project relevance          | Tree of Life, GEBA  |

## Growth conditions and DNA isolation

*N. dassonvillei* strain IMRU 509<sup>T</sup>, DSM 43111, was grown in DSMZ medium 65 (GYM *Streptomyces* medium) [56] at 28°C. DNA was isolated from 0.5–1 g of cell paste using Qiagen Genomic 500 DNA Kit (Qiagen, Hilden, Germany) following the standard protocol as recommended by the manufac-

turer, with modification st/DALM for cell lysis as described in Wu *et al.* [55].

## Genome sequencing and assembly

The genome was sequenced using a combination of Sanger and 454 sequencing platforms. All general aspects of library construction and sequenc-

ing can be found at the JGI website [57]. Pyrosequencing reads were assembled using the Newbler assembler version 2.1-PreRelease (Roche). Large Newbler contigs were broken into 6,356 overlap ping fragments of 1,000 bp and entered into assembly as pseudo-reads. The sequences were assigned quality scores based on Newbler consensus q-scores with modifications to account for overlap redundancy and adjust inflated q-scores. A hybrid 454/Sanger assembly was made using the PGA assembler. Possible mis-assemblies were corrected and gaps between contigs were closed by editing in Consed, by custom primer walks from sub-clones or PCR products. A total of 462 Sanger finishing reads were produced to close gaps, to resolve repetitive regions, and to raise the quality of the finished sequence. Illumina reads were used to improve the final consensus quality using an in-house developed tool (the Polisher ) [58]. The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Sanger and 454 sequencing platforms provided 28.77 × coverage of the genome. The final assembly contains 68,385 Sanger reads and 1,376,163 pyrosequencing reads.

### Genome annotation

Genes were identified using Prodigal [59] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [60]. 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. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [61].

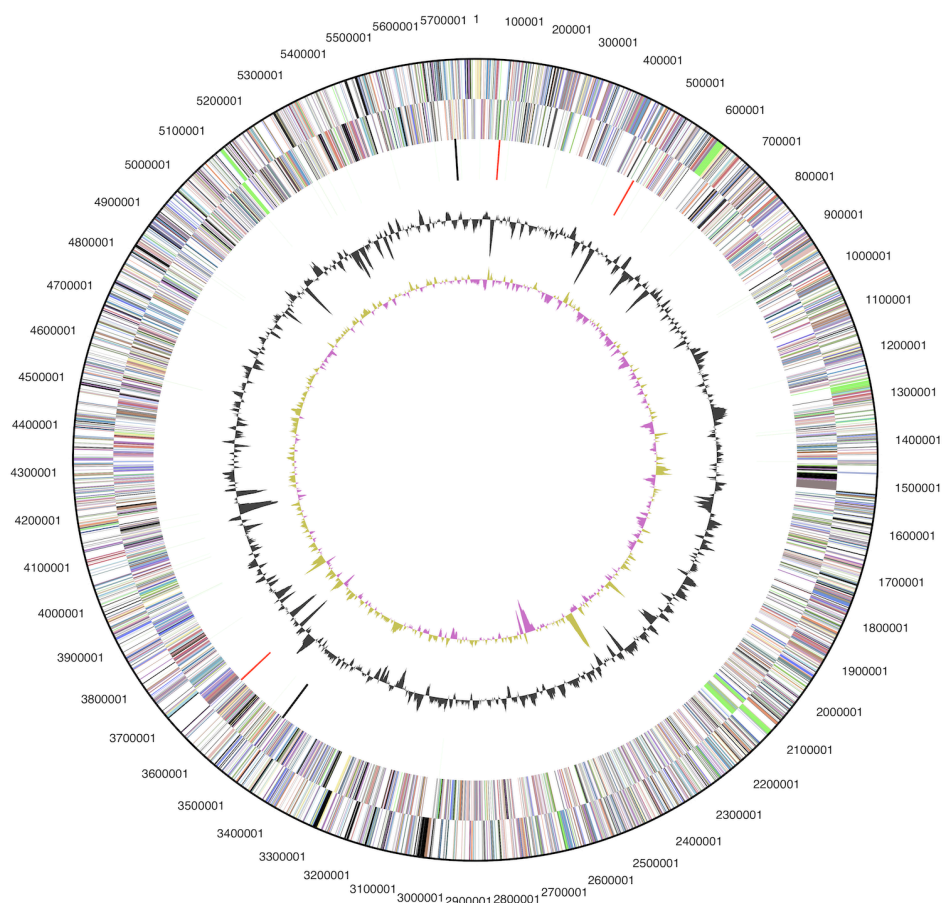
### Genome properties

The genome consists of a 5,767,958 bp long chromosome with a 73% GC content, and a 775,354 bp long plasmid a 72% GC content (Table 3 and Figure 3a and Figure 3b). Of the 5,647 genes predicted, 5,570 were protein-coding genes, and 77 RNAs; 73 pseudogenes were also identified. The majority of the protein-coding genes (69.6%) were assigned with 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.

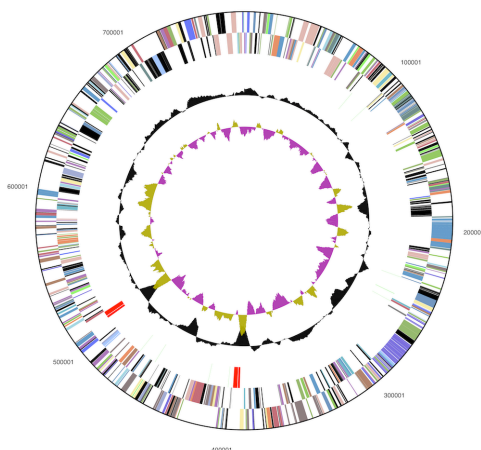
**Table 3.** Genome Statistics

| Attribute                        | Value     | % of Total |
|----------------------------------|-----------|------------|
| Genome size (bp)                 | 6,543,312 | 100.00%    |
| DNA coding region (bp)           | 5,543,886 | 84.73%     |
| DNA G+C content (bp)             | 4,758,935 | 72.73%     |
| Number of replicons              | 2         |            |
| Extrachromosomal elements        | 1         |            |
| Total genes                      | 5,647     | 100.00%    |
| RNA genes                        | 77        | 1.36%      |
| rRNA operons                     | 5         |            |
| Protein-coding genes             | 5,570     | 98.64%     |
| Pseudo genes                     | 73        | 1.29%2     |
| Genes with function prediction   | 3,930     | 69.59%     |
| Genes in paralog clusters        | 1,055     | 18.68%     |
| Genes assigned to COGs           | 3,793     | 67.17%     |
| Genes assigned Pfam domains      | 4,204     | 74.45%     |
| Genes with signal peptides       | 1,686     | 29.86%     |
| Genes with transmembrane helices | 1,337     | 23.68%     |
| CRISPR repeats                   | 8         |            |





**Figure 3a.** Graphical circular map of the chromosome (not drawn to scale with plasmid). 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.



**Figure 3b.** Graphical circular map of the plasmid (not drawn to scale with chromosome). 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 | value | %age | Description  |
|------|-------|------|--|
| J    | 180   | 4.1  | Translation, ribosomal structure and biogenesis                  |
| A    | 1     | 0.0  | RNA processing and modification                                  |
| K    | 518   | 11.9 | Transcription  |
| L    | 173   | 4.0  | Replication, recombination and repair                            |
| B    | 1     | 0.0  | Chromatin structure and dynamics                                 |
| D    | 38    | 0.9  | Cell cycle control, cell division, chromosome partitioning       |
| Y    | 0     | 0.0  | Nuclear structure  |
| V    | 103   | 2.4  | Defense mechanisms   |
| T    | 279   | 6.4  | Signal transduction mechanisms                                   |
| M    | 190   | 4.4  | Cell wall/membrane/envelope biogenesis                           |
| N    | 3     | 0.1  | Cell motility  |
| Z    | 2     | 0.1  | Cytoskeleton   |
| W    | 0     | 0.0  | Extracellular structures   |
| U    | 35    | 0.8  | Intracellular trafficking and secretion, and vesicular transport |
| O    | 115   | 2.6  | Posttranslational modification, protein turnover, chaperones     |
| C    | 266   | 6.1  | Energy production and conversion                                 |
| G    | 371   | 8.5  | Carbohydrate transport and metabolism                            |
| E    | 354   | 8.1  | Amino acid transport and metabolism                              |
| F    | 102   | 2.3  | Nucleotide transport and metabolism                              |
| H    | 210   | 4.8  | Coenzyme transport and metabolism                                |
| I    | 184   | 4.2  | Lipid transport and metabolism                                   |
| P    | 212   | 4.9  | Inorganic ion transport and metabolism                           |
| Q    | 144   | 3.3  | Secondary metabolites biosynthesis, transport and catabolism     |
| R    | 578   | 13.3 | General function prediction only                                 |
| S    | 295   | 6.8  | Function unknown   |
| -    | 1,854 | 32.8 | Not in COGs  |

## Acknowledgements

We would like to gratefully acknowledge the help of Marlen Jando for growing cultures of *N. dasonvillei* and Susanne Schneider for DNA extraction and quality analysis (both at DSMZ). 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 599/1-1.

## References

1. Buchanan RE, Gibbons NE. 1974. *Bergey's Manual of Determinative Bacteriology*, 8th Edition. 1,246 p. The Williams and Wilkins Company. Baltimore, Maryland.
2. Lechevalier MP. 1976. The Biology of *Nocardiae*, p. 517. In M. Goodfellow, G. H. Brownell, and J. A. Serrano (ed.). Academic Press, London.

3. Brocq-Rousseau D. Sur un *Streptothrix*. *Rev Gen Bot* 1904; **16**:219-230.
4. Meyer J. *Nocardiopsis*, a new genus of the order *Actinomycetales*. *Int J Syst Bacteriol* 1976; **26**:487-493. [doi:10.1099/00207713-26-4-487](https://doi.org/10.1099/00207713-26-4-487)
5. Gordon RE, Horan AC. *Nocardia dassonvillei*, a macroscopic replica of *Streptomyces griseus*. *J Gen Microbiol* 1968; **50**:235-240. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/)
6. Lechevalier HA, Lechevalier MP. 1970. A critical evaluation of the genera of aerobic actinomycetes. In H. Prauser (ed.), *The Actinomycetales*. VEB Gustav Fischer Verlag, Jena p. 393-405.
7. Goodfellow M. Numerical taxonomy of some nocardioform bacteria. *J Gen Microbiol* 1971; **69**:33-80. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/)
8. Kroppenstedt RM, Evtushenko LI. 2006. The family *Nocardiopsaceae*. In: M Dworkin, S Falkow, E Rosenberg, KH Schleifer E Stackebrandt (eds), *The Prokaryotes*, 3. ed, vol. 3. Springer, New York, p. 754-795.
9. Rainey FA, Ward-Rainey N, Kroppenstedt RM, Stackebrandt E. The genus *Nocardiopsis* represents a phylogenetically coherent taxon and a distinct actinomycete lineage: proposal of *Nocardiopsaceae* fam. nov. *Int J Syst Bacteriol* 1996; **46**:1088-1092. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/) [doi:10.1099/00207713-46-4-1088](https://doi.org/10.1099/00207713-46-4-1088)
10. Evtushenko LI, Taran VV, Akimov VN, Kroppenstedt RM, Tiedje JM, Stackebrandt E. *Nocardiopsis tropica* sp. nov., *Nocardiopsis trehalosi* sp. nov., nom. rev. and *Nocardiopsis dassonvillei* subsp. *albirubida* subsp. nov., comb. nov. *Int J Syst Evol Microbiol* 2000; **50**:73-81. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/)
11. Jiang C, Xu L. 1998. Actinomycete diversity in unusual habitats. In: C. Jiang and L. Xu (Eds.) *Actinomycetes Research*. Yunnan University Press. Yunnan, China. 259-270.
12. Wang Y, Zhang ZS, Ruan JS, Wang YM, Ali SM. Investigation of actinomycete diversity in the tropical rainforests of Singapore. *J Ind Microbiol Biotechnol* 1999; **23**:178-187. [doi:10.1038/sj.jim.2900723](https://doi.org/10.1038/sj.jim.2900723)
13. Xu LH, Tiang YQ, Zhang YF, Zhao LX, Jiang CL. *Streptomyces thermogriseus*, a new species of the genus *Streptomyces* from soil, lake and hot spring. *Int J Syst Bacteriol* 1998; **48**:1089-1093. [doi:10.1099/00207713-48-4-1089](https://doi.org/10.1099/00207713-48-4-1089)
14. Abyzov SS, Philipova SN, Kuznetsov VD. *Nocardiopsis antarcticus*, a new species of actinomycetes, isolated from the ice sheet of the central Antarctic glacier. *Izv Akad Nauk SSSR Ser Biol* 1983; **4**:559-568.
15. Dixit VS, Pant A. Hydrocarbon degradation and protease production by *Nocardiopsis* sp. NCIM 5124. *Lett Appl Microbiol* 2000; **30**:67-69. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/) [doi:10.1046/j.1472-765x.2000.00665.x](https://doi.org/10.1046/j.1472-765x.2000.00665.x)
16. Boivin-Jahns V, Bianchi A, Ruimy R, Garcin J, Daumas S, Christen R. Comparison of phenotypical and molecular methods for the identification of bacterial strains isolated from a deep subsurface environment. *Appl Environ Microbiol* 1995; **61**:3400-3406. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/)
17. Vasanthi V, Hoti SL. Microbial flora in gut of *Culex quinquefasciatus* breeding in cess pits. *SE Asian J Trop Med Pub Health* 1992; **23**:312-317.
18. Laiz L, Groth I, Schumann P, Zezza F, Felske A, Hermosin B, Saiz-Jimenez C. Microbiology of the stalactites from Grotta dei Cervi, Porto Badisco, Italy. *Int Microbiol* 2000; **3**:25-30. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/)
19. Lacey J. The ecology of actinomycetes in fodders and related substrates: Proceeding of the Warsaw symposium on *Streptomyces* and *Nocardia*. *Zentralbl Bakteriol Parasitenkd Infektionskr Hyg Abt 1 Supp* 1977; **6**:161-170.
20. Andersson MA, Mikkola R, Kroppenstedt RM, Rainey FA, Peltola J, Helin J, Sivonen K, Salkinonja-Salonen MS. The mitochondrial toxin produced by *Streptomyces griseus* strains isolated from indoor environment is valinomycin. *Appl Environ Microbiol* 1998; **64**:4767-4773. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/)
21. Kämpfer P, Busse HJ, Rainey F. *Nocardiopsis compostus* sp. nov., from the atmosphere of a composting facility. *Int J Syst Evol Microbiol* 2002; **52**:621-627. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/)
22. Chun J, Bae KS, Moon EY, Jung SO, Lee HK, Kim SJ. *Nocardiopsis kunsanensis* sp. nov., a moderately halophilic actinomycete isolated from a saltern. *Int J Syst Evol Microbiol* 2000; **50**:1909-1913. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/)
23. Liegard H, Landrieu M. Un cas de mycose conjunctivale. *Ann Ocul (Paris)* 1911; **146**:418-426.
24. Sindhuphak W, Macdonald E, Head E. Actinomycetoma caused by *Nocardiopsis dassonvillei*. *Arch Dermatol* 1985; **121**:1332-1334. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10169111/) [doi:10.1001/archderm.121.10.1332](https://doi.org/10.1001/archderm.121.10.1332)
25. Yassin AF, Rainey FA, Burghardt J, Gierth D, Ungerechts J, Lux I, Seifert P, Bal C, Schaal KP. Description of *Nocardiopsis synnematafomans* sp. nov., elevation of *Nocardiopsis alba* subsp. *prasina* to *Nocardiopsis prasina* comb. nov., and designation of *Nocardiopsis antarctica* and *No-*

- cardiopsis alborubida* as later subjective synonyms of *Nocardiopsis dassonvillei*. *Int J Syst Bacteriol* 1997; **47**:983-988. [PubMed](#) [doi:10.1099/00207713-47-4-983](https://doi.org/10.1099/00207713-47-4-983)
26. Miyashita K, Mikami Y, Arai T. Alkalophilic actinomycete, *Nocardiopsis dassonvillei* subsp. *prasinana* subsp. nov. isolated from soil. *Int J Syst Bacteriol* 1984; **34**:405-409. [doi:10.1099/00207713-34-4-405](https://doi.org/10.1099/00207713-34-4-405)
27. Hill LR, Skerman VBD, Sneath PHA. Corrigenda to the approved lists of bacterial names edited for the International Committee on Systematic Bacteriology. *Int J Syst Bacteriol* 1984; **34**:508-511. [doi:10.1099/00207713-34-4-508](https://doi.org/10.1099/00207713-34-4-508)
28. Tamura T, Ishida Y, Otoguro M, Hatano K, Suzuki K. Reclassification of *Streptomyces flavidofuscus* as a synonym of *Nocardiopsis dassonvillei* subsp. *dassonvillei*. *Int J Syst Evol Bacteriol* 2008; **58**:2321-2323.
29. Chun J, Lee JH, Jung Y, Kim M, Kim S, Kim BK, Lim YW. EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol Microbiol* 2007; **57**:2259-2261. [PubMed](#) [doi:10.1099/ijs.0.64915-0](https://doi.org/10.1099/ijs.0.64915-0)
30. Brodie EL, DeSantis TZ, Parker JP, Zubieta IX, Piceno YM, Andersen GL. Urban aerosols harbor diverse and dynamic bacterial populations. *Proc Natl Acad Sci USA* 2007; **104**:299-304. [PubMed](#) [doi:10.1073/pnas.0608255104](https://doi.org/10.1073/pnas.0608255104)
31. DeSantis TE, Hugenholtz P, Larsen N, Rojas M, Brodie E, Keller K, Huber T, Dalevi D, Hu P, Andersen G. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072. [PubMed](#) [doi:10.1128/AEM.03006-05](https://doi.org/10.1128/AEM.03006-05)
32. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. [PubMed](#)
33. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. [PubMed](#) [doi:10.1093/bioinformatics/18.3.452](https://doi.org/10.1093/bioinformatics/18.3.452)
34. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 2008; **57**:758-771. [PubMed](#) [doi:10.1080/10635150802429642](https://doi.org/10.1080/10635150802429642)
35. Yarza P, Richter M, Peplies J, Euzéby J, Amann R, Schleifer KH, Ludwig W, Glöckner FO, Rosselló-Móra R. The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* 2008; **31**:241-250. [PubMed](#) [doi:10.1016/j.syapm.2008.07.001](https://doi.org/10.1016/j.syapm.2008.07.001)
36. 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. [doi:10.1007/978-3-642-02008-7\\_13](https://doi.org/10.1007/978-3-642-02008-7_13)
37. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008; **36**:D475-D479. [PubMed](#) [doi:10.1093/nar/gkm884](https://doi.org/10.1093/nar/gkm884)
38. Lykidis A, Mavromatis K, Ivanova N, Anderson I, Land M, DiBartolo G, Martinez M, Lapidus A, Lucas S, Copeland A, et al. Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *J Bacteriol* 2007; **189**:2477-2486. [PubMed](#) [doi:10.1128/JB.01899-06](https://doi.org/10.1128/JB.01899-06)
39. Grund E, Kroppenstedt RM. Chemotaxonomy and numerical taxonomy of the genus *Nocardiopsis* Meyer 1976. *Int J Syst Bacteriol* 1990; **40**:5-11. [doi:10.1099/00207713-40-1-5](https://doi.org/10.1099/00207713-40-1-5)
40. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](#) [doi:10.1038/nbt1360](https://doi.org/10.1038/nbt1360)
41. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](#) [doi:10.1073/pnas.87.12.4576](https://doi.org/10.1073/pnas.87.12.4576)
42. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.
43. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol* 1997; **47**:479-491. [doi:10.1099/00207713-47-2-479](https://doi.org/10.1099/00207713-47-2-479)
44. Buchanan RE. Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the *Schizomycetes*. *J Bacteriol* 1917; **2**:155-164. [PubMed](#)

45. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980; **30**:225-420. [doi:10.1099/00207713-30-1-225](https://doi.org/10.1099/00207713-30-1-225)
46. Zhi XY, Li WJ, Stackebrandt E. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *Actinobacteria*, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int J Syst Evol Microbiol* 2009; **59**:589-608. [PubMed doi:10.1099/ijs.0.65780-0](https://pubmed.ncbi.nlm.nih.gov/19111111/)
47. Zhang Z, Wang Y, Ruan J. Reclassification of *Thermomonospora* and *Microtetraspora*. *Int J Syst Bacteriol* 1998; **48**:411-422. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10000000/)
48. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. *Int J Syst Bacteriol* 1980; **30**:225-420. [doi:10.1099/00207713-30-1-225](https://doi.org/10.1099/00207713-30-1-225)
49. Classification of bacteria and archaea in risk groups. [http://www.baua.de TRBA 466](http://www.baua.de/TRBA_466).
50. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. [PubMed doi:10.1038/75556](https://pubmed.ncbi.nlm.nih.gov/10000000/)
51. Prauser H, Falta R. Phagensensibilität, Zellwand-Zusammensetzung und Taxonomie von *Actinomyces*. *Z Allg Mikrobiol* 1968; **8**:39-46. [PubMed doi:10.1002/jobm.3630080106](https://pubmed.ncbi.nlm.nih.gov/10000000/)
52. Kämpfer P, Busse HJ, Rainey FA. *Nocardioopsis compostus* sp. nov., from the atmosphere of a composting facility. *Int J Syst Evol Microbiol* 2002; **52**:621-627. [PubMed](https://pubmed.ncbi.nlm.nih.gov/12000000/)
53. Schleifer KH, Kandler O. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* 1972; **36**:407-477. [PubMed](https://pubmed.ncbi.nlm.nih.gov/10000000/)
54. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; **33**:175-182. [PubMed doi:10.1016/j.syapm.2010.03.003](https://pubmed.ncbi.nlm.nih.gov/20000000/)
55. 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. [PubMed doi:10.1038/nature08656](https://pubmed.ncbi.nlm.nih.gov/18000000/)
56. List of growth media used at DSMZ: [http://www.dsmz.de/microorganisms/media\\_list.php](http://www.dsmz.de/microorganisms/media_list.php).
57. DOE Joint Genome Institute. <http://www.jgi.doe.gov>
58. 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. AGBT, Marco Island, FL, 2008.
59. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [PubMed doi:10.1186/1471-2105-11-119](https://pubmed.ncbi.nlm.nih.gov/20000000/)
60. Pati A, Ivanova N, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A gene prediction improvement pipeline for microbial genomes. *Nat Methods* 2010; **7**:455-457. [PubMed doi:10.1038/nmeth.1457](https://pubmed.ncbi.nlm.nih.gov/20000000/)
61. Markowitz VM, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed doi:10.1093/bioinformatics/btp393](https://pubmed.ncbi.nlm.nih.gov/19000000/)