Complete genome sequence of *Paenibacillus* sp. strain JDR-2

Virginia Chow¹, Guang Nong¹, Franz J. St. John², John D. Rice¹, Ellen Dickstein¹, Olga Cherikov⁴, David Bruce⁴, Chris Detter³, Thomas Brettin⁵, James Han¹, Tanja Woyke⁵, Sam Pitluck⁶, Matt Nolan⁶, Amrita Pati⁶, Joel Martin⁶, Alex Copeland⁶, Miriam L. Land⁵, Lynne Goodwin⁴, Jeffrey B. Jones¹, Lonnie O. Ingram¹, Keelnathan T. Shanmugam¹, and James F. Preston¹*

¹Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida, USA
²US Forest Service, Forest Products Laboratory, Madison, Wisconsin, USA
³Department of Plant Pathology, University of Florida, Gainesville, Florida, USA
⁴Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico USA
⁵Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
⁶DOE Joint Genome Institute, Walnut Creek, California, USA

*Corresponding author: James F. Preston (jpreston@ufl.edu)

Keywords: aerobic, mesophile, Gram-positive, Paenibacillus, xylanolytic, xylan

*Paenibacillus* sp. strain JDR-2, an aggressively xylanolytic bacterium isolated from sweetgum (*Liquidambar styraciflua*) wood, is able to efficiently depolymerize, assimilate and metabolize 4-O-methylglucuronoxylan, the predominant structural component of hardwood hemicelluloses. A basis for this capability was first supported by the identification of genes and characterization of encoded enzymes and has been further defined by the sequencing and annotation of the complete genome, which we describe. In addition to genes implicated in the utilization of β-1,4-xylan, genes have also been identified for the utilization of other hemicellulosic polysaccharides. The genome of *Paenibacillus* sp. JDR-2 contains 7,184,930 bp in a single replicon with 6,288 protein-coding and 122 RNA genes. Uniquely prominent are 874 genes encoding proteins involved in carbohydrate transport and metabolism. The prevalence and organization of these genes support a metabolic potential for bioprocessing of hemicellulose fractions derived from lignocellulosic resources.

Introduction

*Paenibacillus* sp. strain JDR-2 (Pjdr2) was isolated from wafers cut from live stems of sweet gum (*Liquidambar styraciflua*) placed in soil in an area populated predominantly by this tree species. The ability of this isolate to grow on 4-O-methylglucuronoxylane (MeGX) as the sole carbon source identified a metabolic potential not previously described. MeGX is released along with fermentable xylose during dilute acid pretreatment of lignocellulosic biomass. Since MeGX may represent 5 to 20% of the hemicellulose components from hardwoods and agricultural residues, this ability was of interest for increasing bioconversion yields of fermentable sugars from these resources [1,2].

Growth rates and yields of Pjdr2 with polymeric 4-O-methylglucuronoxylane (MeGXn) as substrate were much greater than with monosaccharides and oligosaccharides derived from MeGXn. These increases are presumably the result of a cell-associated multimodular GH10 endoxylanase that generates xylobiose, xylotriose, and the aldouronate, 4-O-methylglucuronoxylotriose (MeGX3), for direct assimilation and metabolism [2]. A cluster of genes was cloned and sequenced from Pjdr2 genomic DNA which contained two genes encoding transcriptional regulators, three genes encoding ABC transporters, and three sequential structural genes lacking secretion sequences encoding a GH67 α-glucuronidase, a
GH10 endoxylanase catalytic domain and a putative GH43 β-xylosidase. The expression of these genes, as well as a distal gene encoding a secreted cell-associated multimodal GH10 endoxylanase, was coordinately responsive to inducers and repressors, leading to their collective designation as a xylan-utilization regulon [3]. Physiological studies defining the preferential utilization of MeGXn compared to MeGX and MeGX3 support a process in which extracellular depolymerization, assimilation and intracellular metabolism are coupled, allowing the rapid and complete utilization of MeGXn [4].

Pjdr2 was the first member of this genus to have its genome completely sequenced and made available for detailed analysis. The sequences of genomes of 2 strains of Paenibacillus polymyxa [5,6], “Paenibacillus vortex” [7], and Paenibacillus sp. Y412MC10 (NCBI NC_013406.1, unpublished results) have since been completed. The incomplete genome sequence Paenibacillus larvae subsp. larvae, the causative agent of American Foulbrood disease of honey bees, has also been analyzed [8].

Classification and features
A phylogenetic tree was constructed using the Neighbor-Joining method [9] for complete sequences of genes encoding 16S rRNA derived from sequenced genomes of Paenibacillus spp., along with the sequences of some members of the Bacillus spp., Microbacterium spp. and Clostridium spp., is presented in Figure 1. The sequence of the gene encoding 16S rRNA (AF355462) from Paenibacillus polymyxa PKB1 is included as representative of the type species of the genus [10].

The unrooted phylogenetic tree shows Pjdr2 in a branch that includes other Paenibacillus spp. in this comparison, supporting a lineage distinct from other Gram positive endospore-forming bacteria. Pjdr2 groups more closely with Paenibacillus lentimorbus and other Paenibacillus species that are insect pathogens than it does with another group that includes type species Paenibacillus polymyxa. From the standpoint of genome size and imputed metabolic potential based on sequence, it is surprising, based on 16S sequence, that it is not more closely related to Paenibacillus sp. Y412MC10. Despite a close similarity of Paenibacillus JDR-2 to Microbacterium species with respect to membrane fatty acids (see discussion below), it is clear that it is not related to members of the genus Microbacterium on the basis of 16S rRNA sequence. When grown on oat spelt xylan agar plates [2], colonies of strain Pjdr2 are white with smooth edges, surrounded by clearing zones resulting from the depolymerization of the xylan. This property was routinely used to monitor the purity of Pjdr2 cultures. As shown in Figure 2, cells of Pjdr2 are rod shaped, with swellings suggestive of sporulation. The properties evaluated for classification allows assignment as an endospore-forming bacterium in the phylum Firmicutes and genus Paenibacillus as noted in Table 1.

Chemotaxonomy
The fatty acid methyl esters analysis (FAME) of Pjdr2 provided an alternative approach for determination of relatedness to other bacteria. Cultures were grown to exponential phase (24 hrs) on Trypticase soy agars. Bacterial cells were harvested and extracted according to the standard MIDI protocol [26]. FAME analysis was conducted using the Sherlock Microbial Identification System 4.5 [27]. Analyses showed that the predominant fatty acid in Pjdr2 is anteiso-C15:0 (46.93%), which in addition to iso-C16:0 (23.02%) and C16:0 (13.48%), constituted >80% of the fatty acid composition of this strain. Minor fatty acids included iso-C14:0 (3.92%), C14:0 (2.35%), and iso-C15:0 (5.29%).

Strains with a similarity index (SI) value of 0.5 or higher indicate a good library comparison (MIDI 2002). The two strains that most closely match the profile of Pjdr2 are Microbacterium laevaniformans (SI = 0.75) and Cellulobacterium cellulans (SI = 0.51). We have included these two species in our phylogenetic analysis based upon their 16S rRNA sequences (Figure 1). The FAME analysis provided a rapid assignment of the species by comparing the fatty acid profile(s) with 60 strains (42 species) of Bacillus, 2 strains (1 species) of Cellulobacterium, 20 strains (19 species) of Microbacterium and 20 strains (18 species) of Paenibacillus, as well as other aerobic bacteria. Sequence analysis of 16S rRNA provides the acceptable basis for considering phylogenetic relationships. Nevertheless the FAME analysis provides a convenient method with which to confirm the identity of the organism as it is maintained and studied over time.
Figure 1. Phylogenetic analysis of *Paenibacillus* sp. JDR-2 was performed using MEGA4 [9] with the Neighbor-Joining method (bootstrap: 2,000 replicates). The species and GenBank accession numbers are: *Paenibacillus larvae* subsp. *pulvifaciens* DSM 3615 (AB073204); *Bacillus halodurans* C-125 (BA000004); *Bacillus subtilis* subsp. *subtilis* str. 168 (AL009126); *Bacillus clausii* KSM-K16 (AP006627); *Bacillus licheniformis* DSM 13 (AE017333); *Bacillus megaterium* str. KL-197 (AY030338); *Bacillus stearothermophilus* (AB021196); *Paenibacillus lentimorbus* (AB110988); *Paenibacillus popilliae* str. ATCC14706(T), (AF071859); *Paenibacillus thiaminolyticus* (D78475); *Paenibacillus nematophilus* str. NEM1b (AF480937); *Paenibacillus polymyxa* SC2 (CP002213); *Paenibacillus sp.* JDR-2 (CP001656); *Paenibacillus vortex* str. V453 (HQ005270); *Clostridium difficile* 630 (AM180355); *Clostridium polysaccharolyticum* DSM 1801 (X71858); *Clostridium acetobutylicum* DSM 1731 (X78071); *Clostridium pasteurianum* (M23930); *Microbacterium testaceum* StLB037 (AP012052); *Microbacterium laevaniformans* str. C820 (NR_036839); *Microbacterium luteolum* DSM 20143 (Y17235); *Cellulosimicrobium cellulans* str. ZFJ-17 (EU931556).
### Table 1. Classification and general features of *Paenibacillus* sp. JDR-2 according to the MIGS recommendations [11].

| MIGS ID | Property | Term | Evidence code |
|---------|----------|------|---------------|
|         | Domain   | *Bacteria* | TAS [12]     |
|         | Phylum    | *Firmicutes* | TAS [13,14] |
|         | Class     | *Bacilli* | TAS [15,16] |
|         | Current classification | Order *Bacillales* | TAS [17,18] |
|         |           | Family *Paenibacillaceae* | TAS [16,19] |
|         |           | Genus *Paenibacillus* | TAS [20-24] |
|         |           | Species *Paenibacillus* sp. Strain JDR-2 | TAS [2] |
|         | Gram stain | Positive | NAS     |
|         | Cell shape | Rod-shaped | NAS     |
|         | Sporulation | Spore-forming | NAS     |
|         | Temperature range | Mesophile, | TAS [2] |
|         | Optimum temperature | 30°C | TAS [2] |
|         | Salinity | Aerobic | IDA     |
|         | Carbon source | Glucose, xylose, β,4-xylan, β,4-1,3-glucan, 4-O-methyl-glucuronoxyllose | TAS [2] |
|         | Energy source | chemoorganotrophic |     |
|         | MIGS-6 | Habitat | Sweet Gum stem wood | TAS [2] |
|         | MIGS-15 | Biotic relationship | Free living | TAS [2] |
|         | MIGS-14 | Pathogenicity | Non pathogenic | NAS |
|         | Biosafety level | 1 | NAS |
|         | Isolation | Sweet Gum stem wood in soil | TAS [2] |
|         | MIGS-4 | Geographic location | Florida | TAS [2] |
|         | MIGS-5 | Sample collection time | 2000 | TAS [2] |
|         | MIGS-4.1 | Latitude | 29.4° | TAS |
|         | MIGS-4.2 | Longitude | 82.3° | TAS |
|         | MIGS-4.3 | Depth | 1 inch | TAS [2] |
|         | MIGS-4.4 | Altitude | 180 feet above msl | 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). Evidence codes are from the Gene Ontology project [25].
Figure 2. Scanning electron micrographs of *Paenibacillus sp.* JDR-2. Panel (a) is representative of the bacilli harvested in the vegetative state and panel (b) indicates individuals with expanded midsections which are entering the sporulation phase. Pjdr2 cells were grown in Luria Broth and harvested by centrifugation at the exponential growth phase (a) and post exponential phase (b), the pellets washed with water 3 times and prepared for scanning electron microscopy by the Electron Microscopy and Bio-Imaging laboratory, ICBR of the University of Florida.

**Growth conditions and DNA isolation**

For the preparation of genomic DNA, one of several colonies surrounded by a clear zone was picked from an agar plate (0.1% oat spelt xylan/0.1% yeast extract/ Zucker-Hankin medium [2], and grown in Zucker-Hankin/1% yeast extract at 30°C with shaking at 240 rpm. A culture (8 ml) at 0.6 OD<sub>600</sub> was inoculated into 48 ml of culture media (Zucker-Hankin, 1% yeast extract). The latter was grown to 0.6 OD<sub>600</sub> and cells were collected by centrifugation. High molecular weight DNA was prepared from these cells as per the protocol provided by JGI. Cells were suspended in TE buffer (10 mM Tris-HCl, 1.0 mM EDTA), pH 8.0 and treated with lysozyme to lyse the cell wall. SDS and Proteinase K were added to denature and degrade proteins. NaCl and CTAB were added to facilitate subsequent precipitation. Cell lysates were extracted with phenol and chloroform and the DNA was precipitated by addition of isopropanol. The nucleic acid pellet was washed with 70% ethanol, dissolved in water and then treated with RNase A.

**Genome sequencing and assembly**

The genome of Pjdr2 was sequenced at the JGI using a combination of 8 kb and 40 kb (fosmid) DNA libraries. In addition to Sanger sequencing, 454 pyrosequencing [28] was performed to a depth of 20× coverage. All general aspects of library construction and sequencing performed at the JGI can be found at the JGI website [29]. Draft assemblies were based on 39,689 total reads. All three libraries provided 5.1× coverage of the genome. The Phred/Phrap/Consed software package [30] was used for sequence assembly and quality assessment [31-33]. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher [34] or transposon bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom primer walk, or PCR amplification (Roche Applied Science, Indianapolis, IN). A total of 1,028 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The completed sequence analysis of Pjdr2 contained 45,057 reads, achieving an average of 5.5-fold sequence coverage per base, with an error rate less than 1 in 100,000. The complete nucleotide sequence of *Paenibacillus sp.* strain JDR-2 and its annotation can be found online at the IMG (Integrated Microbial Genome) portal of JGI [35], as well as at the genome resource site of NCBI [36].

http://standardsingenomics.org
**Genome annotation**

Genes were identified using Prodigal [37] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by manual curation using the JGI program GenePRIMP [38]. The predicted CDSs were translated and searched with the following databases to assign a product description for each predicted protein: the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE [39], RNAMMer [38], Rfam [40], TMHMM [41], and SignalP [42]. Genome statistics are provided in Table 2, and a full circular map in Figure 3 below.

---

**Figure 3.** Circular map of the genome of *Paenibacillus* sp JDR-2. Labeling from the outside circle towards the inside circles: circle 1. Nucleotide numbering system; circle 2 and 3. Predicted coding sequences on the forward strand and on the reverse strand with each gene colored by its assigned COG category; circle 4. RNA genes (tRNAs in green, rRNAs in red, other RNAs in black); circle 5. GC content; circle 6. GC skew.
Table 2. Genomic Statistics

| Attribute                        | Value   | % of Total |
|----------------------------------|---------|------------|
| Genome size (bp)                 | 7,184,930 | 100.00%   |
| DNA coding region (bp)           | 6,384,736 | 88.86%    |
| DNA G+C content (bp)             | 3,612,449 | 50.28%    |
| Number of replicons              | 1       |            |
| Extrachromosomal elements        | 0       |            |
| Total genes                      | 6,410   | 100.00%    |
| RNA genes                        | 122     | 1.90%      |
| rRNA genes                       | 35      | 0.55%      |
| Protein coding genes             | 6,288   | 98.10%     |
| Pseudo Genes                     | 75      | 1.17%      |
| Genes with function prediction   | 4,737   | 73.90%     |
| Protein coding genes with COGs   | 4,667   | 72.81%     |
| Protein coding genes with Pfam   | 5,128   | 80.00%     |
| Genes in paralog clusters        | 1,614   | 25.18%     |
| Protein coding genes coding signal peptides | 1,629 | 25.41% |
| Genes connected to transporter classification | 1,090 | 17.00% |

Insights from genome sequencing

Utilization of lignocellulosics

The nucleotide sequence of a cluster of genes which included the α-glucuronidase gene served as a marker for the sequenced genome. The sequence of this cluster was previously determined in a cosmid clone of the genomic DNA of Pjdr2. The presence of this unique contiguous sequence in a single copy without orthologs or paralogs supported the final genomic sequence as representative of a single genome from a pure culture. This aldouronate-utilization gene cluster, in conjunction with the distal gene encoding a multimodular cell-associated GH10 endoxylanase, constitutes a xylan-utilization regulon as previously defined [3]. The coordinate expression of the genes in this regulon supports a process in which assimilation of the aldouronate, 4-0-methylglucuronoxylotriose, generated by a cell-associated GH10 endoxylanase, is coupled to extracellular depolymerization, assimilation and metabolism as previously described [4]. The sequencing of the genome of Paenibacillus sp. strain JDR-2 has allowed further analysis of its xylan-utilization regulon and the identification of similar regulons involved in the depolymerization and utilization of soluble β-glucans.

A noteworthy feature of the genome of Pjdr2 is the large number (874) of genes involved in carbohydrate metabolism and transport constituting 17% of the genome (Table 3). This characteristic contrasted with 9% and 291 genes in Bacillus subtilis subtilis 168 and 11% and 481 genes in Paenibacillus polymyxa E861. The recently completed genome Paenibacillus sp. Y412MC10, however, is quite similar to Pjdr2 and contains 16% and 828 genes in this category.

Table 3. Number of genes associated with the general COG functional categories

| Code | value | %age | Description                                      |
|------|-------|------|--------------------------------------------------|
| J    | 199   | 3.89 | Translation, ribosomal structure and biogenesis   |
| A    | -     | -    | RNA processing and modification                   |
| K    | 580   | 11.34| Transcription                                     |
| L    | 149   | 2.91 | Replication, recombination and repair             |
| B    | 1     | 0.02 | Chromatin structure and dynamics                  |
| D    | 36    | 0.70 | Cell cycle control, cell division, chromosome partitioning |
| Y    | -     | -    | Nuclear structure                                 |
| V    | 104   | 2.03 | Defense mechanisms                                |
| T    | 426   | 8.33 | Signal transduction mechanisms                    |
| M    | 255   | 4.98 | Cell wall/membrane/envelope biogenesis            |
| N    | 70    | 1.37 | Cell motility                                    |
Table 3 (cont.) Number of genes associated with the general COG functional categories

| Code | value | %age | Description                                      |
|------|-------|------|--------------------------------------------------|
| Z    | 1     | 0.02 | Cytoskeleton                                     |
| W    | -     | -    | Extracellular structures                         |
| U    | 57    | 1.11 | Intracellular trafficking, secretion, and vesicular transport |
| O    | 116   | 2.27 | Posttranslational modification, protein turnover, chaperones |
| C    | 180   | 3.52 | Energy production and conversion                 |
| G    | 874   | 17.08| Carbohydrate transport and metabolism            |
| E    | 316   | 6.18 | Amino acid transport and metabolism              |
| F    | 115   | 2.25 | Nucleotide transport and metabolism              |
| H    | 151   | 2.95 | Coenzyme transport and metabolism                |
| I    | 120   | 2.35 | Lipid transport and metabolism                   |
| P    | 273   | 5.34 | Inorganic ion transport and metabolism           |
| Q    | 99    | 1.94 | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 613   | 11.98| General function prediction only                 |
| S    | 381   | 7.45 | Function unknown                                 |
| -    | 1,743 | 27.19| Not in COGs                                     |

Acknowledgements

We thank the Electron Microscopy and Bio-Imaging laboratory, Interdisciplinary Center for Biotechnology Research, University of Florida for their assistance in preparing the scanning electron micrographs of Strain Pjdr2. We also thank Len Pennacchio, Natalia Ivanova, Roxanne Tapia and Shunsheng Han for their contributions in genome sequencing and annotations of this organism. The work of genomic sequencing was conducted by the U.S. Department of Energy Joint Genome Institute and supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was supported by the funds from the Department of Energy via the Consortium for Plant Biotechnology Research and the Joint Genome Institute (Project ID 4043135).

References

1. Preston JF, Hurlbert JC, Rice JD, Ragunathan A, St. John FJ. Microbial Strategies for the Depolymerization of Glucuronoxylan: Leads to the Biotechnological Applications of Endoxylanases in “Application of Enzymes to Lignocellulosics”, eds S.D. Mansfield and J.N. Saddler. ACS Symposium Series No. 855. Ch 12. pp191-210. 2003.

2. StJohn FJ, Rice J, Preston J. Paenibacillus sp. strain JDR-2 and XynA1: a novel system for methylglucuronoxylan utilization. *Appl Environ Microbiol* 2006; 72:1496-1506. PubMed [http://dx.doi.org/10.1128/AEM.72.2.1496-1506.2006](http://dx.doi.org/10.1128/AEM.72.2.1496-1506.2006)

3. Chow V, Nong G, Preston J. Structure, function, and regulation of the aldouronate utilization gene cluster from Paenibacillus sp. strain JDR-2. *J Bacteriol* 2007; 189:8863-8870. PubMed [http://dx.doi.org/10.1128/JB.01141-07](http://dx.doi.org/10.1128/JB.01141-07)

4. Nong G, Rice J, Chow V, Preston J. Aldouronate utilization in Paenibacillus sp. strain JDR-2: Physiological and enzymatic evidence for coupling of extracellular depolymerization and intracellular metabolism. *Appl Environ Microbiol* 2009; 75:4410-4418. PubMed [http://dx.doi.org/10.1128/AEM.02354-08](http://dx.doi.org/10.1128/AEM.02354-08)

5. Ma M, Wang C, Ding Y, Li L, Shen D, Jiang X, Guan D, Cao F, Chen H, Feng R, et al. Complete genome sequence of Paenibacillus polymyxia SC2, a strain of plant growth-promoting Rhizobacterium with broad-spectrum antimicrobial activity. *J Bacteriol* 2011; 193:311-312. PubMed [http://dx.doi.org/10.1128/JB.00983-10](http://dx.doi.org/10.1128/JB.00983-10)

6. Kim JF, Jeong H, Park SY, Kim SB, Park YK, Choi SK, Ryu CM, Hur CG, Ghim SY, Oh TK, et al. Genome sequence of the polymyxin-producing plant-probiotic rhizobacterium Paenibacillus polymyxia E681. *J Bacteriol* 2010; 192:6103-6104. PubMed [http://dx.doi.org/10.1128/JB.00983-10](http://dx.doi.org/10.1128/JB.00983-10)

7. Sirota-Madi A, Olender T, Helman Y, Ingham C, Brainis I, Roth D, Hagi E, Brodsky L, Leshkowitz D, Galatenko V, et al. Genome sequence of the pattern forming Paenibacillus vortex bacterium reveals potential for thriving in complex environments. *BMC Genomics* 2010; 11:710. PubMed [http://dx.doi.org/10.1186/1471-2164-11-710](http://dx.doi.org/10.1186/1471-2164-11-710)
8. Chan QW, Melathopoulos AP, Pernal SF, Foster LJ. The innate immune and systemic response in honey bees to a bacterial pathogen, *Paenibacillus larvae*. BMC Genomics 2009; 10:387. PubMed http://dx.doi.org/10.1186/1471-2164-10-387

9. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24:1596-1599. PubMed http://dx.doi.org/10.1093/molbev/msm092

10. Li J, Beatty PK, Shah S, Jensen SE. Use of PCR-targeted mutagenesis to disrupt production of fusaricidin-type antifungal antibiotics in *Paenibacillus polymyxa*. Appl Environ Microbiol 2007; 73:3480-3489. PubMed http://dx.doi.org/10.1128/AEM.02662-06

11. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008; 26:541-547. PubMed http://dx.doi.org/10.1038/nbt1360

12. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 1990; 87:4576-4579. PubMed http://dx.doi.org/10.1073/pnas.87.12.4576

13. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), *Bergey's Manual of Systematic Bacteriology*, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.

14. 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.

15. Ludwig W, Schleifer KH, Whitman WB. Class I. Bacilli class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 19-20.

16. Euzéby J. List of new names and new combinations previously effectively published outside the IJSB. Int J Syst Bacteriol 1994; 44:852. http://dx.doi.org/10.1099/00207713-44-4-852

17. Tindall BJ. The type species of the genus *Paenibacillus*? Request for an Opinion. Int J Syst Evol Microbiol 2000; 50:939-940. PubMed http://dx.doi.org/10.1099/00207713-50-2-939

18. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; 30:225-420. http://dx.doi.org/10.1099/00207713-30-1-225

19. De Vos P, Ludwig W, Schleifer KH, Whitman WB. Family IV. *Paenibacillaceae* fam. nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 269.

20. Ash C, Priest FG, Collins MD. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. Antonie van Leeuwenhoek 1993; 64:253-260. PubMed http://dx.doi.org/10.1007/BF00873085

21. Murray RGE, ed. Validation List no. 51. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. *Int J Syst Bacteriol* 1994; 44:852. http://dx.doi.org/10.1099/00207713-44-4-852

22. Euzéby JP. Taxonomic note: necessary correction of specific and subspecific epithets according to Rules 12c and 13b of the International Code of Nomenclature of Bacteria (1990 Revision). *Int J Syst Bacteriol* 1998; 48:1073-1075. http://dx.doi.org/10.1099/00207713-48-3-1073

23. Tindall BJ. What is the type species of the genus *Paenibacillus*? Request for an Opinion. *Int J Syst Evol Microbiol* 2000; 50:939-940. PubMed http://dx.doi.org/10.1099/00207713-50-2-939

24. Trüper HG. The type species of the genus *Paenibacillus* Ash et al. 1994 is *Paenibacillus polymyxa*. Opinion 77. Judicial Commission of the International Committee on Systematics of Prokaryotes. *Int J Syst Evol Microbiol* 2005; 55:513. PubMed

25. 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 http://dx.doi.org/10.1038/75556

26. Sasser M. Microbial Identification by gas chromatographic analysis of fatty acid methyl esters (GC_FAME). MIDI Technical Note 101. MIDI Inc. Newark, DE; 2009.

27. MIDI. MIS Operating Manual. MIDI, Inc., Newark, DE 19713; 2002.

http://standardsingenomics.org
28. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; 437:376-380. PubMed

29. DOE Joint Genome Institute. [http://www.jgi.doe.gov](http://www.jgi.doe.gov)

30. The Phred/Phrap/Consed software package. [http://www.phrap.com](http://www.phrap.com)

31. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; 8:186-194. PubMed

32. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; 8:175-185. PubMed

33. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; 8:195-202. PubMed

34. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. In: Arabnia H, Valafar, H, editor. Proceedings of the 2006 International Conference on Bioinformatics & Computational Biology. CSREA Press; 2006. p 141-146.

35. Integrated Microbial Genome portal of JGI. [http://img.jgi.doe.gov/cgi-bin/w/main.cgi?section=TaxonDetail&taxon_oid=644736396](http://img.jgi.doe.gov/cgi-bin/w/main.cgi?section=TaxonDetail&taxon_oid=644736396)

36. NCBI. [http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome&cmd=Retrieve&dopt=Overview&list_uids=1609](http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome&cmd=Retrieve&dopt=Overview&list_uids=1609)

37. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; 11:119. PubMed [http://dx.doi.org/10.1186/1471-2105-11-119](http://dx.doi.org/10.1186/1471-2105-11-119)

38. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. Rfam: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; 35:3100-3108. PubMed [http://dx.doi.org/10.1093/nar/gkm160](http://dx.doi.org/10.1093/nar/gkm160)

39. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; 25:955-964. PubMed [http://dx.doi.org/10.1093/nar/25.5.955](http://dx.doi.org/10.1093/nar/25.5.955)

40. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res* 2003; 31:439-441. PubMed [http://dx.doi.org/10.1093/nar/gkg006](http://dx.doi.org/10.1093/nar/gkg006)

41. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; 305:567-580. PubMed [http://dx.doi.org/10.1006/jmbi.2000.4315](http://dx.doi.org/10.1006/jmbi.2000.4315)

42. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; 340:783-795. PubMed [http://dx.doi.org/10.1016/j.jmb.2004.05.028](http://dx.doi.org/10.1016/j.jmb.2004.05.028)