Complete genome sequence of *Kytococcus sedentarius* type strain (541<sup>T</sup>)

David Sims<sup>1</sup>, Thomas Brettin<sup>1,2</sup>, John C. Detter<sup>1</sup>, Cliff Han<sup>1</sup>, Alla Lapidus<sup>2</sup>, Alex Copeland<sup>2</sup>, Tijana Glavina Del Rio<sup>2</sup>, Matt Nolan<sup>2</sup>, Feng Chen<sup>1</sup>, Susan Lucas<sup>2</sup>, Hope Tice<sup>2</sup>, Jan-Fang Cheng<sup>2</sup>, David Bruce<sup>1</sup>, Lynne Goodwin<sup>1</sup>, Sam Pitluck<sup>2</sup>, Galina Ovchinnikova<sup>2</sup>, Amrita Pati<sup>2</sup>, Natalia Ivanova<sup>2</sup>, Konstantinos Mavromatis<sup>2</sup>, Amy Chen<sup>1</sup>, Krishna Palaniappan<sup>2</sup>, Patrik D'haeseleer<sup>2,4</sup>, Patrick Chain<sup>2,4</sup>, Jim Bristow<sup>2</sup>, Jonathan A. Eisen<sup>2,5</sup>, Victor Markowitz<sup>3</sup>, Philip Hugenholtz<sup>2</sup>, Susanne Schneider<sup>6</sup>, Markus Göker<sup>6</sup>, Rüdiger Pukall<sup>6</sup>, Nikos C. Kyrpides<sup>2</sup>, and Hans-Peter Klenk*<sup>6</sup>

<sup>1</sup> Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA  
<sup>2</sup> DOE Joint Genome Institute, Walnut Creek, California, USA  
<sup>3</sup> Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA  
<sup>4</sup> Lawrence Livermore National Laboratory, Livermore, California, USA  
<sup>5</sup> University of California Davis Genome Center, Davis, California, USA  
<sup>6</sup> DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

*Corresponding author: Hans-Peter Klenk

**Keywords:** mesophile, free-living, marine, aerobic, opportunistic pathogenic, Dermaciococcaceae

*Kytococcus sedentarius* (ZoBell and Upham 1944) Stackebrandt et al. 1995 is the type strain of the species, and is of phylogenetic interest because of its location in the Dermaciococcaceae, a poorly studied family within the actinobacterial suborder Micrococcineae. *K. sedentarius* is known for the production of oligoketide antibiotics as well as for its role as an opportunistic pathogen causing valve endocarditis, hemorrhagic pneumonia, and pitted keratolysis. It is strictly aerobic and can only grow when several amino acids are provided in the medium. The strain described in this report is a free-living, nonmotile, Gram-positive bacterium, originally isolated from a marine environment. Here we describe the features of this organism, together with the complete genome sequence, and annotation. This is the first complete genome sequence of a member of the family Dermaciococcaceae and the 2,785,024 bp long single replicon genome with its 2639 protein-coding and 64 RNA genes is part of the Genomic Encyclopedia of Bacteria and Archaea project.

**Introduction**

Strain 541<sup>T</sup> (DSM 20547 = ATCC 14392 = JCM 11482 = CCM 314 and other culture collections) is the type strain of the species *Kytococcus sedentarius*, which is the type species of the genus *Kytococcus* [1]. Strain 541<sup>T</sup> was first described as *Micrococcus sedentarius* (ZoBell and Upham 1944) [2] and later emended as *K. sedentarius* in a taxonomic dissection of the genus *Micrococcus* [1]. The organism is of interest for its biotechnological potential as source of natural antibiotics (oligoketides), for its role as an opportunistic pathogen, and for its position in the tree of life, where it represents the scarcely populated genus *Kytococcus* (2 species) within in the actinobacterial family Dermaciococcaceae [1] (Figure 1). *K. sedentarius* 541<sup>T</sup> was first isolated around 1944 from a marine environment [2], but strains of the species were also frequently isolated from human skin [3]. More recently, closely related strains were also isolated from culture-dependant environmental screenings.
of a non-saline alkaline groundwater environment in Cabeco de Vide in southern Portugal [4], screening for pelagic bacteria in South Korea [5], tropical marine sediments from the intertidal zone off the coast of the Republic Palau [6], from the ciliate Collinia sp.), an endoparasite of euphausiids from the Gulf of California (unpublished literature, GenBank record EU090136), and in a culture-independent analysis of the microbial burden and diversity in commercial airline cabins [7]. Screening of environmental genomic samples and surveys reported at the NCBI BLAST server indicated no closely related phylotypes that can be linked to the species or genus. Here we present a summary classification and a set of features for K. sedentarius strain 541T (Table 1), together with the description of the complete genomic sequencing and annotation.

**Classification and features**

*K. sedentarius* cells are spherical/coccoid and occur predominantly in tetrads which can be arranged in cubical packets [1] (Figure 2). Cells are described as Gram-positive, nonmotile, non-encapsulated, and not endospore-forming [1]. *K. sedentarius* 541T is strictly aerobic and chemooorganotrophic, requires methionine and other amino acids for growth, and grows well in NaCl at concentrations up to 10% (w/v) [1].

*K. sedentarius* (strain NK0508) is capable of degrading diphenylarsenic acid [8], but not starch [1], and does not produce acids from most carbohydrates and alcohols [1]. Its optimal growth temperature is 28-36°C. Nitrate is reduced to nitrite by some *K. sedentarius* strains [1]. *K. sedentarius* is not only described as the source of the oligoketide antibiotics monensin A and B [9], but has also been associated with pitted keratolysis [10], opportunistic infections, and fatal hemorrhagic pneumonia [11].

Figure 1 shows the phylogenetic neighborhood of *K. sedentarius* strain 541T in a 16S rRNA based tree. Analysis of the 16S rRNA gene copies in the genome of strain 541T differed by one nucleotide from each other, and by up to two nucleotides from the previously published 16S rRNA sequence generated from DSM 20547 (X87755).

**Chemotaxonomy**

The murein of *K. sedentarius* strain 541T contains L-Lys-Glu, a variation of cell wall type A4α [1]. Mycolic acids and teichonic acids were not reported [1]. Strain 541T contains only completely unsaturated menaquinones with 8-11 isoprene subunits (MK8 to MK11), with MK8 dominating [1]. The major cellular fatty acids are methyl-branched chain iso-C₁₇:₀ and anteiso-C₁₇:₀, as well as the straight chain saturated C₁₅:₀ and C₁₇:₀ [1]. Phosphatidylglycerol, diphosphatidylglycerol, and
phosphatidylinositol were identified as dominating polar lipids [1]. Reported cytochromes include \( a_{a3}, c_{626}, c_{550}, b_{557}, b_{561}, \) and \( b_{564} \) [1].

Figure 2. Scanning electron micrograph of \( K. \text{ sedentarius} \) strain 541\(^T \) (Manfred Rohde, Helmholtz Centre for Infection Biology, Braunschweig)

Table 1. Classification and general features of \( K. \text{ sedentarius} \) strain 541\(^T \) based on MIGS recommendations [16]

| MIGS ID | Property         | Term                                      | Evidence code |
|---------|------------------|-------------------------------------------|---------------|
|         | Domain           | \textit{Bacteria}                         |               |
|         | Phylum           | \textit{Actinobacteria}                    |               |
|         | Class            | \textit{Actinobacteria}                    | TAS [17]      |
|         | Order            | \textit{Actinomycetales}                  | TAS [18]      |
|         | Suborder         | \textit{Micrococccinea}                   | TAS [17]      |
|         | Family           | \textit{Dermacoccacea}                    | TAS [19]      |
|         | Genus            | \textit{Kytococcus}                       | TAS [1]       |
|         | Species          | \textit{Kytococcus sedentarius}           | TAS [1]       |
|         | Type strain      | 541                                       |               |
|         | Gram stain       | positive                                  | TAS [1]       |
|         | Cell shape       | spherical, predominantly in tetrads       | TAS [1]       |
|         | Motility         | nonmotile                                 | TAS [1]       |
|         | Sporulation      | non-sporulating                           | TAS [1]       |
|         | Temperature range | mesophilic                                | TAS [1]       |
|         | Optimum temperature | 28-36°C                  | TAS [1]       |
|         | Salinity         | nonhalophilic, but growth in media        | TAS [1]       |
|         |                  | up to 10% (w/v) NaCl                      |               |
Table 1. Classification and general features of *K. sedentarius* strain 541\textsuperscript{T} based on MIGS recommendations \[16\] (cont.)

| MIGS ID | Property | Term | Evidence code |
|---------|----------|------|---------------|
| MIGS-22 | Oxygen requirement | mandatory aerobe | TAS [1] |
|         | Carbon source | not reported | |
|         | Energy source | unknown, not starch | NAS |
| MIGS-6  | Habitat | marine | TAS [2] |
| MIGS-15 | Biotic relationship | free-living | NAS |
| MIGS-14 | Pathogenicity | in rare cases | TAS [10,11] |
|         | Biosafety level | 1 | TAS [20] |
|         | Isolation | slide submerged in sea water | TAS [2] |
| MIGS-4  | Geographic location | probably San Diego | TAS [2] |
| MIGS-5  | Sample collection time | about or before 1944 | TAS [2] |
| MIGS-4.1 | Latitude – Longitude | not reported | |
| MIGS-4.2 | | | |
| 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 \[21\]. If the evidence code is IDA, then the property was directly observed, for a live isolate by one of the authors, or another expert mentioned in the acknowledgements.

**Genome sequencing and annotation**

**Genome project history**
This organism was selected for sequencing on the basis of its phylogenetic position, and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project. The genome project is deposited in the Genome OnLine Database \[15\] and 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                                                             |
|         |                           | Two genomic Sanger libraries: 8kb pMCL200 and fosmid pcc1Fos libraries. |
| MIGS-28 | Libraries used            | ABI3730                                                              |
| MIGS-29 | Sequencing platforms      | ABI3730                                                              |
| MIGS-31.2 | Sequencing coverage      | 17.3x Sanger                                                          |
|         | Assemblers                | phrap                                                                |
| MIGS-30 | Gene calling method       | Genemark 4.6b, tRNAscan-SE-1.23, Infernal 0.81                      |
| MIGS-32 | Genbank ID                | ABUD0000000000                                                       |
|         | Genbank Date of Release   | N/A                                                                  |
|         | NCBI project ID           | 21067                                                                |
|         | GOLD ID                   | Gc01042                                                              |
|         | Database: IMG-GEBA        | 2500901761                                                           |
| MIGS-13 | Source material identifier | DSM 20547                                                           |
|         | Project relevance         | Tree of Life, GEBA                                                  |

**Growth conditions and DNA isolation**
*K. sedentarius* strain 541\textsuperscript{T}, DSM20547, was grown in DSMZ medium \textbf{92} (3% trypticase soy broth, 0.3% yeast extract) at 30°C. DNA was isolated from 1-1.5 g of cell paste using Qiagen Genomic 500 DNA Kit (Qiagen, Hilden, Germany) with a modified protocol for cell lysis in first freezing for 20 min. (-70°C), then heating 5 min. (98°C), and

http://standardsingenomics.org
cooling 15 min to 37°C; adding 1.5 ml lysozyme (standard: 0.3 ml, only), 1.0 ml achromopeptidase, 0.12 ml lysostaphine, 0.12 ml mutanolysine, 1.5 ml proteinase K (standard: 0.5 ml, only), followed by overnight incubation at 35°C.

**Genome sequencing and assembly**
The genome was sequenced using a combination of 8 kb and fosmid DNA libraries. All general aspects of library construction and sequencing performed at the JGI website. Draft assemblies were based on 60,742 total reads. The Phred/Phrap/Consed software package was used for sequence assembly and quality assessment [22-24]. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher [25] or transposon bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom priming, or PCR amplification (Roche Applied Science, Indianapolis, IN). A total of 1,255 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The completed genome sequence of *K. sedentarius* 541T contains 61,582 reads. The error rate of the completed genome sequence is less than 1 in 100,000. Together all libraries provided > 17x coverage of the genome.

**Genome annotation**
Genes were identified using GeneMark [26] as part of the genome annotation pipeline in the Integrated Microbial Genomes Expert Review (IMG-ER) system [27], followed by a round of manual cura-

| Attribute                  | Value        | % of Total |
|----------------------------|--------------|------------|
| Genome size (bp)           | 2,785,024    |            |
| DNA Coding region (bp)     | 2,558,989    | 91.88%     |
| DNA G+C content (bp)       | 1,994,844    | 71.63%     |
| Number of replicons        | 1            |            |
| Extrachromosomal elements  | 0            |            |
| Total genes                | 2703         | 100.00%    |
| RNA genes                  | 64           | 2.37%      |
| rRNA operons               | 2            |            |
| Protein-coding genes       | 2639         | 97.63%     |
| Pseudo genes               | 84           | 3.11%      |
| Genes with function prediction | 1948   | 72.07%     |
| Genes in paralog clusters  | 288          | 10.65%     |
| Genes assigned to COGs     | 1851         | 68.48%     |
| Genes assigned Pfam domains| 1908         | 70.59%     |
| Genes with signal peptides | 539          | 19.94%     |
| Genes with transmembrane helices | 595     | 22.01%     |
| CRISPR repeats             | 0            | 0          |
**Figure 3. Graphical circular map of the genome.** 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.

The distribution of genes into COGs functional categories is presented in Table 4, and a cellular overview diagram is presented in Figure 4, followed by a summary of metabolic network statistics shown in Table 5.

| Code | Value | %  | Description                                 |
|------|-------|----|---------------------------------------------|
| J    | 151   | 5.7| Translation                                 |
| A    | 1     | 0.0| RNA processing and modification             |
| K    | 143   | 5.4| Transcription                               |
| L    | 160   | 6.1| Replication, recombination and repair       |
| B    | 2     | 0.1| Chromatin structure and dynamics            |
| D    | 22    | 0.8| Cell cycle control, mitosis and meiosis     |
| Y    | 0     | 0.0| Nuclear structure                           |
| V    | 56    | 2.1| Defense mechanisms                          |
| T    | 73    | 2.8| Signal transduction mechanisms              |
| M    | 111   | 4.2| Cell wall/membrane biogenesis              |
Table 4. Number of genes associated with the 21 general COG functional categories (cont.)

| Code | Value | %   | Description                                      |
|------|-------|-----|--------------------------------------------------|
| N    | 2     | 0.1 | Cell motility                                   |
| Z    | 1     | 0.0 | Cytoskeleton                                    |
| W    | 0     | 0.0 | Extracellular structures                        |
| U    | 27    | 1.0 | Intracellular trafficking and secretion         |
| O    | 64    | 2.4 | Posttranslational modification, protein turnover, chaperones |
| C    | 99    | 3.8 | Energy production and conversion                |
| G    | 116   | 4.4 | Carbohydrate transport and metabolism           |
| E    | 185   | 7.0 | Amino acid transport and metabolism             |
| F    | 75    | 2.8 | Nucleotide transport and metabolism             |
| H    | 101   | 3.8 | Coenzyme transport and metabolism               |
| I    | 86    | 3.3 | Lipid transport and metabolism                  |
| P    | 117   | 4.4 | Inorganic ion transport and metabolism          |
| Q    | 46    | 1.7 | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 229   | 8.7 | General function prediction only                |
| S    | 160   | 6.1 | Function unknown                                |
| -    | 788   | 29.9| Not in COGs                                     |

Table 5. Metabolic Network Statistics

| Attribute                  | Value |
|----------------------------|-------|
| Total genes                | 2703  |
| Enzymes                    | 531   |
| Enzymatic reactions        | 922   |
| Metabolic pathways         | 185   |
| Metabolites                | 662   |

Figure 4. Schematic cellular overview of all pathways of the *K. sedentarius* strain 541T metabolism. Nodes represent metabolites, with shape indicating class of metabolite. Lines represent reactions.
Acknowledgements

We would like to gratefully acknowledge the help of Katja Steenblock (DSMZ) for growing K. sedentarius 541T 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, as well as German Research Foundation (DFG) INST 599/1-1.

References

1. Stackebrandt E, Koch C, Gvozdiak O, Schumann P. Taxonomic dissection of the genus Micrococcus: Kocuria gen. nov., Nesterenkonia gen. nov., Kytococcus gen. nov., Dermacoccus gen. nov., and Micrococcus Cohn 1872 gen. emend. Int J Syst Bacteriol 1995; 45:682-692. PMID:7547287 doi:10.1099/00207713-45-4-682

2. ZoBell CE, Upham HC. A list of marine bacteria including descriptions of sixty new species. Bull Scripps Inst Oceanogr Calif 1944; 5:239-292

3. Kloss WE, Tornabebe TG, Schleifer KH. Isolation and characterization of Micrococci from human skin, including two species: Micrococcus lylae and Micrococcus kytococcus Cohn 1974; 24:79-101. PMID:9933918 doi:10.1099/00207713-24-1-79

4. Tiago I, Chung AP, Verissimo A. Bacterial diversity in a nonsaline alkaline environment: heterotrophic aerobic populations. Appl Environ Microbiol 2004; 70:7378-MID:15574939 doi:10.1128/AEM.70.12.7378-7387.2004

5. Bhattarai HD, Lee YK, Cho KH, Lee HK, Shin HE. The study of antagonistic interactions among pelagic bacteria: a promising way to coin environmental friendly anti fouling compounds. Hydrobiologica 2006; 568:417-423 doi:10.1007/s10750-006-0220-2

6. Gontang EA, Fenical W, Jensen PR. Phylogenetic diversity of gram-positive bacteria cultured from marine sediments. Appl Environ Microbiol 2007; 73:3272-MID:17400789 doi:10.1128/AEM.02811-06

7. Osman S, La Duc MT, Dekas A, Newcombe D, Venkateswaran K. Microbial burden and diversity of commercial airline cabin air during short and long durations of travel. ISME J 2008; 2:482-497 PMID:18256704 doi:10.1038/ismej.2008.11 MID:18256704 doi:10.1038/ismej.2008.11

8. Nakamiya K, Nakayama T, Ito H, Edmonds JS, Shibata Y, Morita M. Degradation of arylarsonic compounds by microorganisms. FEMS Microbiol Lett 2007; 274:184-MID:17697081 doi:10.1111/j.1574-6968.2007.00835.x

9. Pospisil S, Benada O, Kofronova O, Petricek M, Janda L, Havlicek V. Kytococcus sedentarius (formerly Micrococcus sedentarius) and Dermacoccus nishinomiyaensis (formerly Micrococcus nishinomiyaensis) produce monensins, typical Streptomyces cinnamomensis metabolites. Can J Microbiol 1998; 44:1007-MID:9933918 doi:10.1139/cjm-44-10-1007

10. Longshaw CM, Wright JD, Farrell AM, Holland KT. Kytococcus sedentarius, the organism associated with pitted keratolysis, produces two keratin-degrading enzymes. J Appl Microbiol 2002; 93:810-MID:12392527 doi:10.1046/j.1365-2672.2002.01742.x

11. Levaenga H, Donnelly P, Blijleven N, Verweij P, Shirango H, de Pauw B. Fatal hemorrhagic pneumonia caused by infection due to Kytococcus sedentarius--a pathogen or passenger? Ann Hematol 2004; 83:447-MID:14689234 doi:10.1007/s00277-003-0831-x

12. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. Bioinformatics 2002; 18:452-464 PMID:11934745 doi:10.1093/bioinformatics/18.3.452

13. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 1981; 17:368-MID:7288891 doi:10.1007/BF01734359

14. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol 2008; 57:758-MID:18853362 doi:10.1080/10635150802429642

15. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes On Line Database (GOLD) in 2007: status of genomic and metage-
Kytococcus sedentarius strain 541T

An anonymous paper discussing the Metagenomics projects and their associated metadata.

Nucleic Acids Res 2008; 36:D475-479. PMID:17981842 doi:10.1093/nar/gkm884

16. 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. PMID:18464787 doi:10.1038/nbt1360

17. Stackebrandt E, Rainey F, Ward-Rainey N. Proposal for a new hierarchic classification system, Actinobacteria classis nov. Int J Syst Bacteriol 1997; 47:479-491. doi:10.1109/00207713-47-2-479

18. Buchanan RE. Studies in the Nomenclature and Classification of the Bacteria: VIII. The Sub-groups and Genera of the Actinomycetales. J Bacteriol 1918; 3:403-406. PMID:1655883

19. Stackebrandt E, Schumann P. Description of Bogoriellaceae fam. nov., Dermacoccaceae fam. nov., Rarobacteraceae fam. nov. and Sanguibacteraceae fam. nov. and emendation of some families of the suborder Micrococccineae. Int J Syst Evol Microbiol 2000; 50:1279-1285. PMID:10843073

20. Anonymous. Biological Agents: Technical rules for biological agents. <www.baua.de>.

21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight DS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25:25-29. PMID:10802651 doi:10.1038/75556

22. Ewing B, Green P. Base-calling of automated sequence traces using phred. II. Error probabilities. Genome Res 1998; 8:186-196. PMID:9521922 doi:10.1101/gr.8.3.186

23. 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. PMID:9521921 doi:10.1101/gr.8.3.175

24. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. Genome Res 1998; 8:195-202. PMID:9521923 doi:10.1101/gr.8.3.195

25. Han CS, Chain P. Finishing repeat regions automatically with Dupfinisher CSREA Press. In: Arabia AR, Valafar H, editors. Proceedings of the 2006 international conference on bioinformatics & computational biology; 2006 June 26-29. CSREA Press. p 141-146.

26. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 2001; 29:2607-2618. PMID:11410670 doi:10.1093/nar/29.12.2607

27. Markowitz V, Mavromatis K, Ivanova N, Chen I-M, Chu K, Palaniappan K, Szeto E, Anderson I, Lykidis A, Kyrpides N. Expert Review of Functional Annotations for Microbial Genomes. (Submitted 2009).

28. Pati A. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes (In preparation 2009).

29. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997; 25:955-M. PMID:9023104 doi:10.1093/nar/25.5.955

30. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007; 35:3100-3108. PMID:17452365 doi:10.1093/nar/gkm160

31. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res 2005; 33:D121-D127. PMID:15608160 doi:10.1093/nar/gki081

32. Markowitz VM, Szeto E, Palaniappan K, Grechkin Y, Chu K, Chen IM, Dubchak I, Anderson I, Lykidis A, Mavromatis K, et al. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. Nucleic Acids Res 2008; 36:D528-D533. PMID:17933782 doi:10.1093/nar/gkm846

33. Karp P, Paley S, Romero P. The pathway tools software. Bioinformatics 2002; 18 Suppl1:S225-232. PMID:12169551

34. Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee SY, Shearer AG, Tissier C, et al. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 2008; 36:D528-D533. PMID:17933782 doi:10.1093/nar/gkm900

35. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res 2005;
