Title
Complete genome sequence of Cryptobacterium curtum type strain (12-3T)

Permalink
https://escholarship.org/uc/item/1nz955tm

Author
Mavromatis, Konstantinos

Publication Date
2009-09-20

Peer reviewed
Complete genome sequence of *Cryptobacterium curtum* type strain (12-3^T^)

Konstantinos Mavromatis^1^, Rüdiger Pukall^2^, Christine Rohde^2^, David Sims^1,3^, Thomas Brettin^1,3^, Cheryl Kuske^1,3^, John C. Detter^1,3^, Cliff Han^1,3^, Alla Lapidus^1^, Alex Copeland^1^, Tijana Glavina Del Rio^1^, Matt Nolan^1^, Susan Lucas^1^, Hope Tice^1^, Jan-Fang Cheng^1^, David Bruce^1,3^, Lynne Goodwin^1,3^, Sam Pitluck^1^, Galina Ovchinnikova^1^, Amrita Pati^1^, Natalia Ivanova^1^, Amy Chen^1^, Krishna Palaniappan^4^, Patrick Chain^1,5^, Patrik D'haeseleer^1,5^, Jim Bristow^1^, Jonathan A. Eisen^1,6^, Victor Markowitz^4^, Philip Hugenholtz^1^, Manfred Rohde^7^, Hans-Peter Klenk^2^, and Nikos C. Kyrpides^1^*

1^ DOE Joint Genome Institute, Walnut Creek, California, USA  
2^ DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany  
3^ Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico USA  
4^ Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA  
5^ Lawrence Livermore National Laboratory, Livermore, California 94550, USA  
6^ University of California Davis Genome Center, Davis, California, USA  
7^ HZI - Helmholtz Centre for Infection Research, Braunschweig, Germany

*Corresponding author: Nikos Kyrpides

Keywords  
Oral infections, opportunistic pathogenic, periodontitis, non-spore-former, anaerobic, asaccharolytic, *Coriobacteriaceae*

Abstract

*Cryptobacterium curtum* Nakazawa et al. 1999 is the type species of the genus, and is of phylogenetic interest because of its very distant and isolated position within the family *Coriobacteriaceae*. *C. curtum* is an asaccharolytic, opportunistic pathogen with a typical occurrence in the oral cavity, involved in dental and oral infections like periodontitis, inflammations and abscesses. Here we describe the features of this organism, together with the complete genome sequence, and annotation. This is the first complete genome sequence of the actinobacterial family *Coriobacteriaceae*, and this 1,617,804 bp long single replicon genome with its 1364 protein-coding and 58 RNA genes is part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

Introduction

*Cryptobacterium curtum* strain 12-3^T^ (DSM 15641 = ATCC 700683 = CCUG 43107) is the type strain of the species, representing the sole species of the genus *Cryptobacterium* [1]. *C. curtum* was described by Nakazawa et al. in 1999 [1]. The organism is of significant interest because of its position in the tree of life where it was initially wrongly placed close to *Eubacterium* (Firmicutes) to be then relocated in the phylum *Actinobacteria*, close to the *Coriobacteriaceae* [1, Fig. 1].
The type strain 12-3T and a second strain of the species, KV43-B, both classified in *Cryptobacterium curtum* gen. nov., sp. nov., were isolated from a periodontal pocket sample of an adult patient and from necrotic dental pulp, respectively [1]. *C. curtum* can also be isolated from human oral and dental infections like pulpal inflammations, advanced caries [1], dental abscesses or periodontitis [2]. 16S rRNA gene sequence analysis revealed that the two isolates represent a distinct lineage within the family *Coriobacteriaceae*, between the neighbouring genera *Eggerthella* and *Slackia*. No significant matches with any 16S rRNA sequences from environmental genomic samples and surveys are reported at the NCBI BLAST server (February 2009).

The very short and non-motile rods form tiny translucent colonies of less than 1 mm in diameter on BHI-blood agar without haemolysis after prolonged incubation under strictly anaerobic conditions. Transmission electron micrographs of ultrathin sections of *C. curtum* 12-3T showed a single-layered Gram-positive cell wall of approximately 10 nm thickness [1]. Carbohydrates are not metabolised, the species is asaccharolytic [1]. *C. curtum* is unreactive in most biochemical tests. The human oral cavity contains arginine and other amino acids and oligopeptides due to proteinase and peptidase activities. *C. curtum* degrades arginine through arginine deiminase pathway [3]. Like *Slackia exigua*, a closely related species, these bacteria are very difficult to cultivate. Optimal doubling time is 12 hours [3]. There are no chemotaxonomic data available to *C. curtium* strain 12-3T.

Here we present a summary classification and a set of features for *C. curtum* 12-3T (Tab. 1), together with the description of the complete genomic sequencing and annotation.

**Classification and features of organism**

Fig. 1 shows the phylogenetic neighborhood of *C. curtum* strain 12-3T in a 16S rRNA based tree. Analysis of the thee 16S rRNA gene sequences in the genome of strain 12-3T indicated that the genes differ by at most one nucleotide from eachother, but differ by 15 nucleotides and eight ambiguities (1.1%) from the previously published 16S rRNA sequence (AB019260) generated from DSM 15641. The significant differences between the genome data and the reported 16S rRNA gene sequence is most likely due to sequencing errors in the previously reported sequence data.

![Figure 1. Phylogenetic tree of *C. curtum* 12-3T and most type strains of the family *Coriobacteriaceae*, inferred from 1422 aligned 16S rRNA characters [4, 5] under the maximum likelihood criterion [6]. The tree was rooted with type strains of the genera](image-url)
Collinsella and Coriobacterium. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 1000 bootstrap replicates if larger than 60%. Strains with a genome sequencing project registered in GOLD [7] are printed in blue; published genomes in bold.

**Figure 2.** Scanning electron micrograph of *C. curtum* 12-3\(^T\)

![Scanning electron micrograph of *C. curtum* 12-3\(^T\)](image)

**Table 1.** Classification and general features of *C. curtum* 12-3\(^T\) in accordance to the MIGS recommendations [8]
MIGS ID | Property | Term | Evidence code\(^{ab}\)
--- | --- | --- | ---

| Current classification | Domain | Bacteria |
| | Phylum | Actinobacteria |
| | Class | Actinobacteria |
| | Order | Coriobacteriales |
| | Family | Coriobacteriaceae |
| | Genus | Cryptobacterium |
| | Species | Cryptobacterium curtum |
| Type strain | 12-3 |
| Gram stain | positive |
| Cell shape | very short rods |
| Motility | nonmotile |
| Sporulation | non-sporulating |
| Temperature range | mesophile |
| Optimum temperature | 37°C |
| Salinity | normal |

MIGS-22 | Oxygen requirement | obligate anaerobic |
| | Carbon source | asaccharolytic |
| | Energy source | arginine, lysine |

MIGS-6 | Habitat | human oral microflora |

MIGS-15 | Biotic relationship | growth on enzymatic degradation products of inflamed tissues |

MIGS-14 | Pathogenicity | periodontal infections |
| Biosafety level | 1 (+) |
| Isolation | infected human oral cavity |

MIGS-4 | Geographic location | not reported |

MIGS-5 | Sample collection time | about 1995 |

| MIGS-4.1 | Latitude – Longitude | not reported |
| MIGS-4.2 | Depth | not reported |
| MIGS-4.4 | Altitude | not reported |

a) Evidence code types – (R)eported for the purpose of this specific publication, directly observed by one of the authors or acknowledged person or institution for the living isolated sample, (C)ited: a direct report exists in the literature, or (I)nferred: not directly observed for the living, isolated sample, but based on a personally accepted property for this species, or anecdotal communication.

b) A general mapping of these evidence codes to those evidence codes ([http://www.geneontology.org/GO.evidence.shtml](http://www.geneontology.org/GO.evidence.shtml)) used by the Gene Ontology project [8] is: R= IDA; C=TAS; and I= NAS.

Genome sequencing and annotation information

Genome project history
This organism was selected for sequencing on the basis of each phylogenetic position, and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project. The genome project is deposited in the Genome OnLine Database [7] and the complete genome sequence in GenBank. Sequencing, finishing and annotation was performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.
Table 2. Genome sequencing project information

| MIGS ID | Property                        | Term                                                                 |
|---------|---------------------------------|----------------------------------------------------------------------|
| MIGS-31 | Finishing quality               | Finished                                                             |
| MIGS-28 | Libraries used                  | Three genomic libraries: two Sanger libraries - 8 kb pMCL200 and fosmid pcc1Fos - and one 454 pyrosequence standard library |
| MIGS-29 | Sequencing platforms            | ABI3730, 454 GS FLX                                                 |
| MIGS-31.2| Sequencing coverage            | 12.9 x Sanger; 20 x pyrosequence                                    |
| MIGS-30 | Assemblers                      | Newbler version 1.1.02.15, phrap                                      |
| MIGS-32 | Gene calling method             | Genemark 4.6b, tRNAscan-SE-1.23, infernal 0.81                      |
|         | INSDC / Genbank ID              | N/A                                                                 |
|         | Genbank Date of Release         | N/A                                                                 |
|         | GOLD ID                         | Gi02234                                                             |
|         | Database: IMG-GEBA              | 2500901758                                                          |
|         | Project relevance               | Tree of Life, GEBA                                                  |

Growth conditions and DNA isolation

*C. curtum* strain 12-3°T, DSM 15641, was grown in DSMZ medium 78 (Chopped Meat Medium), supplemented with 1 g/l arginine, at 37°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 containing more lysozyme (5x) and proteinase K (3x), and overnight incubation at 35°C on a shaker.

Genome sequencing and assembly

The genome was sequenced using a combination of Sanger and 454 sequencing platforms. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. 454 pyrosequencing reads were assembled using the Newbler assembler version 1.1.02.15 (Roche). Large Newbler contigs were broken into 1,799 overlapping fragments of 1000bp 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 to adjust inflated q-scores. A hybrid 454/Sanger assembly was made using the parallel phrap assembler (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher [11] 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. 47 Sanger finishing reads were produced to close gaps, to resolve repetitive regions, and to raise the quality of the finished sequence. The error rate of the completed genome sequence is less than 1 in 100,000. Together all sequence types provided 32.9 x coverage of the genome.

Genome annotation *(to be revised by Nikos)*

Genes were identified using GeneMark [12] as part of the genome annotation pipeline in the Integrated Microbial Genomes Expert Review (IMG-ER) system (http://img.jgi.doe.gov/er) [13], followed by a round of manual curation using JGI’s GenePRIMP pipeline (http://geneprimp.jgi-psf.org) [14]. 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. The tRNAscanSE tool [15] was used to find tRNA genes, whereas ribosomal RNAs were found by using the tool RNAmmer [16]. Other non-coding RNAs were identified by searching the genome for the Rfam profiles using INFERNAL (v0.81) [17]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG) platform (http://img.jgi.doe.gov/) [18].

**Metabolic network analysis**

The metabolic Pathway/Genome Database (PGDB) was computationally generated using Pathway Tools software version 12.5 [19] and MetaCyc version 12.5 [20], based on annotated EC numbers and a customized enzyme name mapping file. It has undergone no subsequent manual curation and may contain errors, similar to a Tier 3 BioCyc PGDB [21].

**Genome properties**

The genome is 1,617,804 bp long and comprises one main circular chromosome with a 50.9% GC content. Of the 1422 genes predicted, 1364 were protein coding genes, and 58 RNAs. 7 pseudogenes were also identified. 78.5% of the genes were assigned with a putative function while the remaining are annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Table 5.

| Attribute                          | Value      | % of Total |
|-----------------------------------|------------|------------|
| Genome size (bp)                  | 1,617,804  | 100%       |
| DNA Coding region (bp)            | 1,438,957  | 88.95%     |
| DNA G+C content (bp)              | 823,649    | 50.91%     |
| Number of replicons               | 1          |            |
| Extrachromosomal elements         | 0          |            |
| Total genes                       | 1422       | 100.00%    |
| RNA genes                         | 58         | 2.37%      |
| rRNA operons                      | 3          |            |
| Protein-coding genes              | 1364       | 95.92%     |
| Pseudo genes                      | 7          | 0.49%      |
| Genes with function prediction    | 1117       | 78.55%     |
| Genes in paralog clusters         | 77         | 5.41%      |
| Genes assigned to COGs            | 1103       | 77.57%     |
| Genes assigned Pfam domains       | 1104       | 77.64%     |
| Genes with signal peptides        | 278        | 19.77%     |
| Genes with transmembrane helices  | 361        | 25.39%     |
| CRISPR repeats                    | 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, sRNAs red, other RNAs black), GC content, GC skew.
Table 4. Number of genes associated with the 21 general COG functional categories

| Code | COG counts and percentage of protein-coding genes | Description |
|------|---------------------------------------------------|-------------|
|      | Genome value % of total                           |             |
| J    | 128 9.38                                          | Translation, ribosomal structure and biogenesis |
| A    | 1 0.07                                            | RNA processing and modification               |
| K    | 94 6.89                                           | Transcription                                  |
| L    | 74 5.45                                           | Replication, recombination and repair          |
| B    | 1 0.07                                            | Chromatin structure and dynamics               |
| D    | 15 1.09                                           | Cell cycle control, mitosis and meiosis       |
| Y    | 0 0.00                                            | Nuclear structure                              |
| V    | 20 1.46                                           | Defense mechanisms                             |
| T    | 64 4.69                                           | Signal transduction mechanisms                |
| M    | 70 5.13                                           | Cell wall/membrane biogenesis                 |
| N    | 1 0.07                                            | Cell motility                                  |
| Z    | 1 0.07                                            | Cytoskeleton                                   |
| W    | 0 0.00                                            | Extracellular structures                       |
| U    | 20 1.46                                           | Intracellular trafficking and secretion       |
| O    | 55 4.03                                           | Posttranslational modification, protein turnover, chaperones |
| C    | 100 7.33                                          | Energy production and conversion              |
| G    | 41 3.00                                           | Carbohydrate transport and metabolism         |
| E    | 96 7.03                                           | Amino acid transport and metabolism           |
| F    | 47 3.44                                           | Nucleotide transport and metabolism           |
| Attribute                      | Value  |
|-------------------------------|--------|
| Total genes                   | 1422   |
| Enzymes                       | 316    |
| Enzymatic reactions           | 606    |
| Metabolic pathways            | 115    |
| Metabolites                   | 506    |

Acknowledgements
We would like to gratefully acknowledge the help of G. Gehrich-Schröter for growing *C. curtum* cultures 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’s Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, as well as German Research Foundation (DFG) INST 599/1-1.

References
1. Nakazawa F, Poco SE, Ikeda T, Sato M, Kalfas S, Sundqvist G, Hoshino E. Cryptobacterium curtum gen. nov., sp. nov., a new genus of Gram-positive anaerobic rod isolated from human oral cavities. *Int J Syst Bacteriol* 1999, **49**:1193-200.

2. Kumar PS, Griffen AL, Barton JA, Paster BJ, Moeschberger ML, Leys EJ. New bacterial species associated with chronic periodontitis. *J Dent Res* 2003, **82**:338-44.

3. Uematsu H, Sato N, Djais A, Hoshino E. Degradation of arginine by *Slackia exigua* ATCC 700122 and *Cryptobacterium curtum* ATCC 700683. *Oral Microbiol Immunol* 2006, **21**: 381-4.

4. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002, **18**:452-64

5. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000, **17**: 540-52.

6. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008, **57**:758-71.

7. Liolios, K., Mavromatis, K., Tavernarakis, N., and Kyrpides, N.C. The Genomes OnLine Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata *Nucleic Acids Res* 2008, **36**:D475-9.

8. Field D, Garrity G, Gray T, Morrison N, Selengut J, et al. Towards a richer description of our complete collection of genomes and metagenomes: the “Minimum Information about a Genome Sequence” (MIGS) specification. *Nature Biotechnology* 2008, **26**:541-7.

9. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol* 1997, **47**:479-91.

10. Biological Agents: Technical rules for biological agents [www.baua.de](http://www.baua.de) TRBA 466.

11. Han CS, Chain P. Finishing repeat regions automatically with Dupfinisher. In: Proceeding of the 2006 international conference on bioinformatics & computational biology. Hamid R. Arabnia & Homayoun Valafar (eds), CSREA Press. June 26-29, 2006:141-6.

12. 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**:2607-18.

13. Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC. Expert Review of Functional Annotations for Microbial Genomes. *Bioinformatics (in press)* 2009.

14. Pati et al. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes. *in preparation* 2009
15. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997, **25**:955-64.

16. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007, **35**:3100-8.

17. 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**:D121-4.

18. Markowitz VM, Szeto E, Palaniappan K, Grechkin Y, Chu K, Chen I-MA, Dubchak I, Anderson I, Lykidis A, Mavromatis K, Ivanova NN, Kyrpides NC. The Integrated Microbial Genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res* 2008, **36**:D528-33.

19. Karp PD, Paley SM, Romero P. The Pathway Tools Software. *Bioinformatics* 2000, **18**:S225-32.

20. Karp P, Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley SM, Rhee SY, Shearer A, Tissier C, Walk TC, Zhang P. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/Genome Databases. *Nucleic Acids Res* 2008, **36**:D623-31.

21. 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, **33**:6083-9.