

Complete genome sequence of *Arthrobacter sp.* strain FB24

Cindy H. Nakatsu^{1*}, Ravi Barabote², Sue Thompson², David Bruce², Chris Detter², Thomas Brettin², Cliff Han², Federico Beasley¹, Weimin Chen¹, Allan Konopka³, and Gary Xie²

¹Department of Agronomy, Purdue University, West Lafayette, IN

²Los Alamos National Laboratories, Los Alamos, NM

³Pacific Northwest National Laboratory, Richland, WA

*Correspondence: Cindy H. Nakatsu (cnakatsu@purdue.edu)

Arthrobacter sp. strain FB24 is a species in the genus *Arthrobacter* Conn and Dimmick 1947, in the family *Micrococcaceae* and class *Actinobacteriaceae*. A number of *Arthrobacter* genome sequences have been completed because of their important role in soil, especially bioremediation. This isolate is of special interest because it is tolerant to multiple metals and it is extremely resistant to elevated concentrations of chromate. The genome consists of a 4,698,945 bp circular chromosome and three plasmids (96,488, 115,507, and 159,536 bp, a total of 5,070,478 bp), coding 4,536 proteins of which 1,257 are without known function. This genome was sequenced as part of the DOE Joint Genome Institute Program.

Abbreviations-IMG- Integrated microbial genomes, DOE-JGI- Department of Energy Joint Genome Institution, NCBI- National Center for Biotechnology Information (Bethesda, MD, USA), RDP- Ribosomal Database Project (East Lansing, MI, USA)

Introduction

Arthrobacter sp. strain FB24 was isolated from a microcosm made from soil collected at an Indiana Department of Transport facility in Seymour, Indiana. This site was of particular interest because the soils were contaminated by mixed waste, both petroleum hydrocarbons and extreme metal (chromium and lead) levels [1]. Details of microcosm enrichment and isolation procedures used to obtain the *Arthrobacter* strain have been described previously [2]. This isolate was of particular interest because of its extreme resistance to chromate [3,4]. This work is a part of a larger study determining the compositional and functional diversity of bacterial communities in soils exposed to long-term contamination with metals [5-7].

Classification and features

Arthrobacter sp. strain FB24 is a high G+C Gram-positive member of the *Micrococcaceae* (Figure 1, Table 1). The strain is a facultative, non-motile aerobe with characteristic morphology of rod-shaped cells (Figure 2) that become coccoid in stationary phase. Strain FB24 is able to use a number carbon sources for growth, including glucose, fructose,

lactate, succinate, malate, xylose and aromatic hydrocarbons (hydroxybenzoates, phthalate). Additionally, this *Arthrobacter sp.* strain is resistant to multiple metals: arsenate, arsenite, chromate, cadmium, lead, nickel, and zinc.

Genome sequencing information

Genome project history

Arthrobacter sp. strain FB24 was chosen for sequencing by DOE-JGI because of its extreme resistance to chromate. Table 2 presents the project information and its association with MIGS version 2.0 compliance [25].

Growth conditions and DNA isolation

The FB24 culture used for DNA extraction was started from the glycerol stock (stored at -80°C) that was made from the original isolate. Cells were streaked onto a 0.1× nutrient agar plate, incubated at 30°C, then a single colony was used to grow a culture in 0.25× nutrient broth (NB) (Difco, USA). Total genomic DNA was extracted from cells grown in liquid culture using the standard CTAB procedure [26].

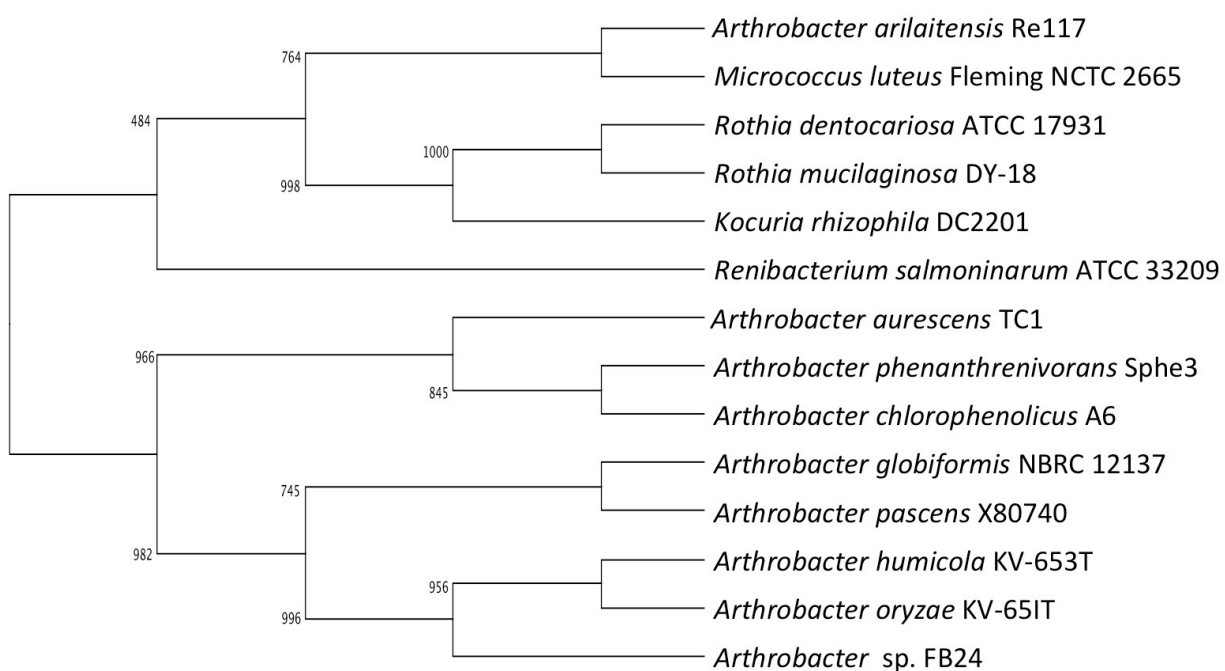


Figure 1. Phylogenetic tree of *Arthrobacter* strain FB24 relative to nearest neighboring *Arthrobacter* type strains and *Micrococcaceae* strains with finished genome sequences: *A. arilaitensis* re117 (FQ311476) [8], *A. aurescens* TC1 (NC_008709) [9], *A. chlorophenicus* A6 (NC_011886), *A. phenanthrenivorans* Sphe3 (CP002379 [10], *Kocuria rhizophila* DC2201, *Micrococcus luteus* Fleming NCTC 2665, *Renibacterium salmoninarum* ATCC 33209, *Rothia dentocariosa* ATCC 17931, and *Rothia mucilaginosa* DY-18. The sequences were aligned in ClustalX and a consensus tree was generated using a 1,000× repeated bootstrapping process [11,12].



Figure 2. Transmission electron micrograph of *Arthrobacter* sp. strain FB24. Cells were grown in nutrient broth for 15 h (~early stationary phase), fixed in 3% glutaraldehyde in 0.1 M cacodylate buffer, then fixed in reduced osmium, followed by a series of ethanol dehydration steps. Cells are then embedded in Spurr resin, stained with uranyl acetate and Reynold's lead citrate. Image was captured on Kodak SO-163 film at 33,000× magnification.

Table 1. Classification and general features of *Arthrobacter* strain FB24

MIGS ID	Property	Term	Evidence code ^a
		Domain <i>Bacteria</i>	TAS [13]
		Phylum <i>Actinobacteria</i>	TAS [14]
		Class <i>Actinobacteria</i>	TAS [15]
	Current classification	Order <i>Actinomycetales</i>	TAS [15-18]
		Family <i>Micrococcaceae</i>	TAS [15-17,19]
		Genus <i>Arthrobacter</i>	TAS [17,20-23]
		Species <i>Arthrobacter sp.</i>	TAS [14]
		Type strain	TAS [15]
	Gram stain	Positive	IDA
	Cell shape	Polymorphic: Coccus to rod shape	IDA
	Motility	Non-motile	IDA
	Sporulation	Non-sporulating	IDA
	Temperature range	4-37°C	IDA
	Optimum temperature	30°C	IDA
	Carbon source	Yeast extract, glucose, fructose, lactate, succinate, malate, xylose, hydroxybenzoates, phthalate	
	Energy source	Yeast extract, glucose, fructose, lactate, succinate, malate, xylose, hydroxybenzoates, phthalate	
	Terminal electron receptor	Oxygen or nitrate	IDA
MIGS-6	Habitat	Soil	TAS [1]
	Isolation	Chromate and xylene enriched micro-cosm composed of anthropogenically disturbed soils	TAS [2]
MIGS-6.3	Salinity		
MIGS-22	Oxygen	Facultative aerobe	IDA
MIGS-15	Biotic relationship	Free-living	IDA
MIGS-14	Pathogenicity	Non-pathogenic	NAS
MIGS-4	Geographic location	Seymour, Indiana, USA	TAS [1,2]
MIGS-5	Sample collection time	June 27, 2001	IDA
MIGS-4.1	Latitude	38.9591667	NAS
MIGS-4.2	Longitude	-85.8902778	NAS
MIGS-4.3	Depth	40-90 cm	NAS
MIGS-4.4	Altitude	583 feet	NAS

^aEvidence codes - IDA: Inferred from Direct Assay; 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 [24].

Genome properties

The 5,070,478- base pair genome of *Arthrobacter FB24* is composed of a single 4,698,945-base pair circular chromosome and three large circular plasmids (96,488, 115,507, and 159,536 bp) (Table 3) with GC content of 65.5, 64.7, 63.3 and 65.0%, respectively. Based on a summary of genomic features listed on the Integrated Microbial Genomes (IMG) [34] there are 4,536 protein coding sequences identified, of which 3,279 (70.94%, Table 4) have been assigned to a COG functional category (Table 5, Figure 3 and Figure 4). There are 1,257 (27.19%) predicted genes without an associated function.

Genome sequencing and assembly

The random shotgun method was used in Sanger sequencing the genome of *Arthrobacter sp.* strain FB24 at the DOE-Joint Genome Institution (DOE-JGI). Medium (8 kb) and small (3 kb) insert random libraries were partially sequenced with average success rate of 88% and average high-quality read lengths of 614 nucleotides. Sequences were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher [27] or by analysis of transposon insertions in bridge clones. Gaps between contigs were closed by editing, custom primer walk or PCR amplification. The completed genome sequence of *Arthrobacter sp.* FB24 contains 89530 reads, achieving an average of 15-fold sequence coverage per base with an error rate less than 1 in 100,000. The sequences of *Arthrobacter sp.* FB24 can be accessed using the GenBank accession number NC_008541 for the chromosome and NC_008537, NC_008538, NC_008539 for three plasmids.

Genome annotation

Automated gene prediction was performed by using the output of Critica [28], combined with the output of Generation and Glimmer [29]. The assignment of product descriptions was made by using search results of the following curated databases in this order: TIGRFam; PRIAM (e^{-30} cutoff); Pfam; Smart; COGs; Swissprot/TrEMBL (SPTR); and KEGG. If there was no significant similarity to any protein in another organism, it was described as "hypothetical protein." "Conserved hypothetical protein" was used if at least one match was found to a hypothetical protein in another organism. EC numbering was based on searches in PRIAM at an e^{-10} cutoff; COG and KEGG functional classifications were based on homology searches in the respective databases. Additionally, the tRNAScanSE tool [30] was used to find tRNA genes, whereas ribosomal RNAs were found by using BLASTn vs. the 16S and 23S ribosomal RNA databases. Other "standard" structural RNAs (e.g., 5S rRNA, rnpB, tmRNA, SRP RNA) were found by using covariance models with the Infernal search tool [31]. The HMMTOP program was used to predict the number of transmembrane segments (TMSs) in each protein. Those predicted to have two or more TMSs (about 918 proteins) were used to interrogate the transporter database (TCDB). Peter Karp's pathologic tool was used for pathway prediction [32]. This method largely relies on the keyword matching and other automatic methods to manually curate some of the pathways, such as aromatic compound degradation. Metabolic pathways were constructed using MetaCyc as a reference data set [33].

Table 2. Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Small and medium random shotgun clones
MIGS-29	Sequencing platforms	Sanger
MIGS-31.2	Fold coverage	~15-fold
MIGS-30	Assemblers	Parallel PHRAP
MIGS-32	Gene calling method	Critica, Generation, Glimmer
	Genome Database release	March 1, 2007
	Genbank ID	12640
	Genbank Date of Release	October 24, 2006
	GOLD ID	Gc00445
	Project relevance	Bioremediation, biotechnological, environmental

Table 3. Summary of genome

Label	Size (bp)	Topology	INSDC identifier	RefSeq ID
Chromosome 1	4,698,945	Circular	CP000454.1	NC_008541.1
Plasmid pFB104	96,488	Circular	CP000457.1	NC_008539.1
Plasmid pFB105	115,507	Circular	CP000456.1	NC_008538.1
Plasmid pFB136	159,536	Circular	CP000455.1	NC_008537.1

Table 4. Nucleotide content and gene count levels of the genome

Attribute	Value	% of total ^a
Genome size (bp)	5,070,478	100.0
DNA coding region (bp)	4,552,065	89.78
DNA G+C content (bp)	3,315,507	65.39
Total genes ^b	4,622	100.00
RNA genes	86	1.86
Protein-coding genes with function prediction	3,279	70.94
Protein coding genes without function prediction	1,257	27.19
Genes in paralog clusters	965	20.88
Genes assigned to COGs	3,361	72.72
Genes with signal peptides	1,098	23.76
Genes with transmembrane helices	1,168	25.27
Paralogous groups	373	100.00

a) The total is based on either the size of the genome in base pairs or on the total number of protein coding genes in the annotated genome.

b) Also includes 54 pseudogenes and 5 other genes

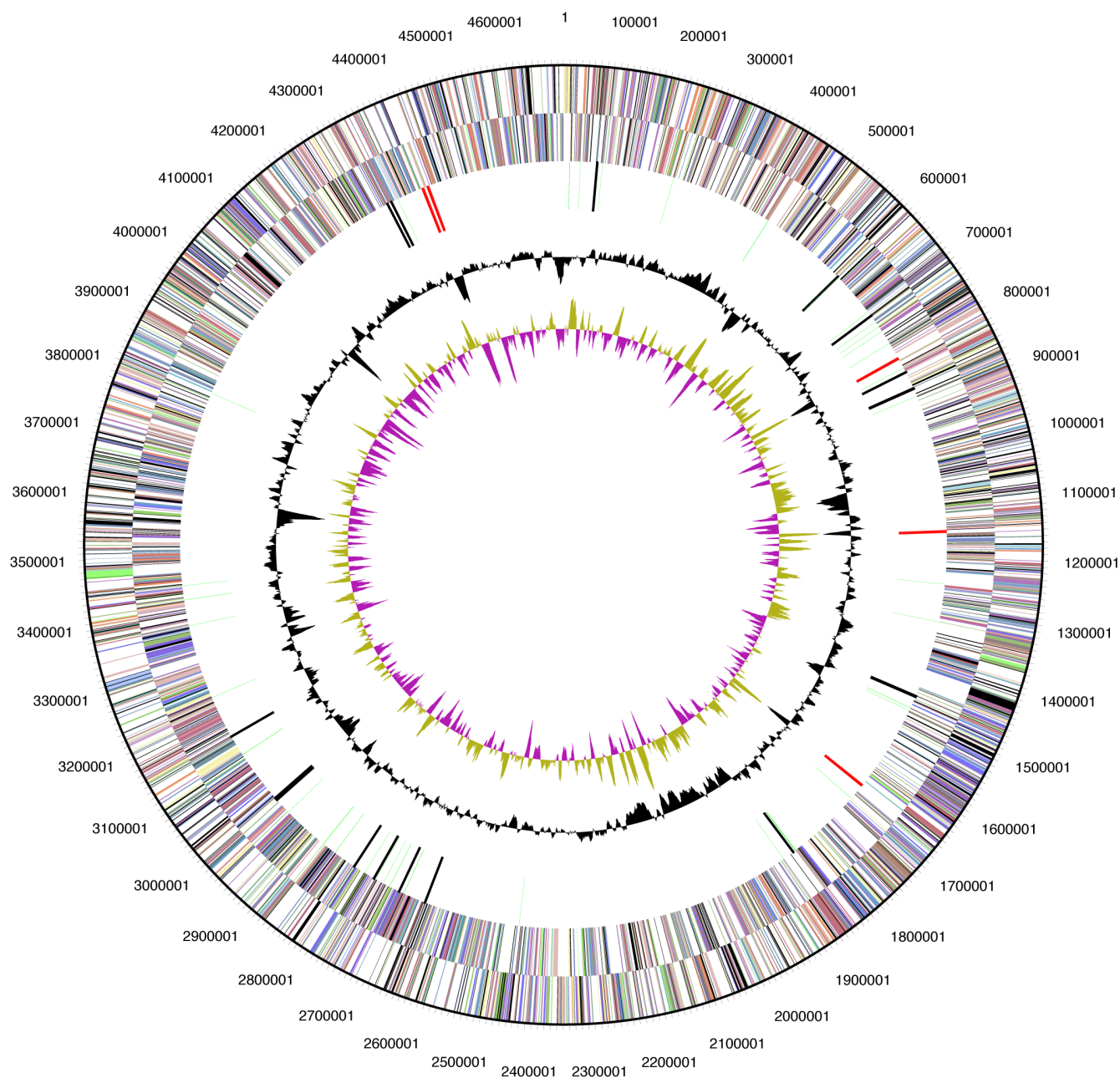


Figure 3. Circular map of FB24 chromosome, graphical depiction from outside to the center: genes on forward strand, genes on reverse strand (colored by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew. Chromosome is not to scale with plasmid maps.

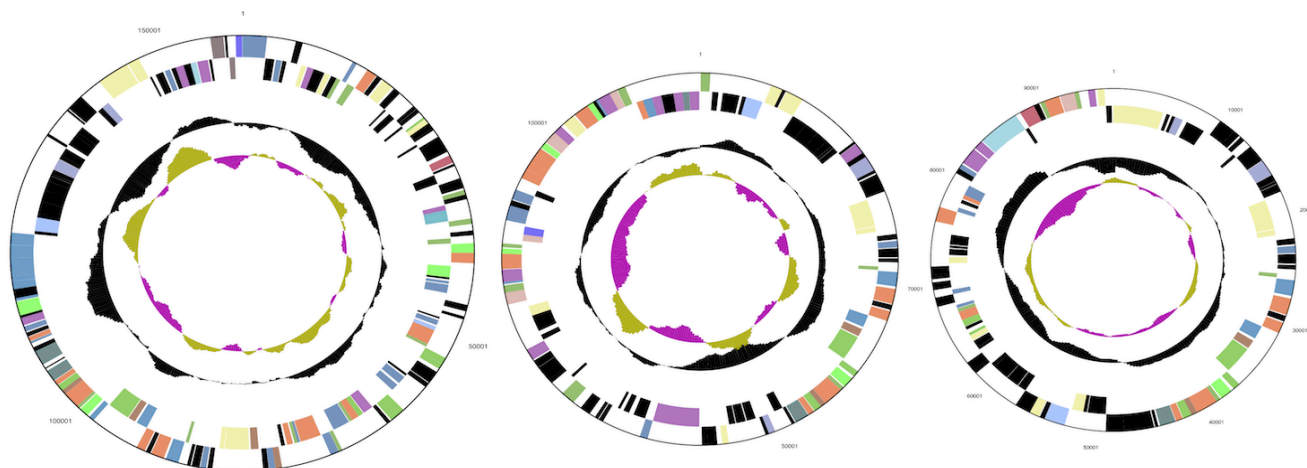


Figure 4. Circular map of three plasmids in FB24, graphical depiction from outside to the center: genes on forward strand, genes on reverse strand (colored by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew. Plasmid maps not to scale with each other or with chromosome map.

Table 5. Number of genes associated with general COG functional categories

Code	Value	%age ^a	Description
J	162	4.27	Translation, ribosomal structure and biogenesis
A	1	0.03	RNA processing and modification
K	363	9.57	Transcription
L	164	4.32	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	32	0.84	Cell cycle control, cell division, chromosome partitioning
Y	-	-	Nuclear structure
V	49	1.29	Defense mechanisms
T	162	4.27	Signal transduction mechanisms
M	171	4.51	Cell wall/membrane/envelope biogenesis
N	3	0.08	Cell motility
Z	1	0.03	Cytoskeleton
W	0	0.0	Extracellular structures
U	48	1.27	Intracellular trafficking, secretion, and vesicular transport
O	124	3.27	Posttranslational modification, protein turnover, chaperones
C	239	6.3	Energy production and conversion
G	436	11.49	Carbohydrate transport and metabolism
E	364	9.6	Amino acid transport and metabolism
H	98	2.58	Nucleotide transport and metabolism
I	155	4.09	Lipid transport and metabolism
P	207	5.46	Inorganic ion transport and metabolism
Q	112	2.95	Secondary metabolites biosynthesis, transport and catabolism
R	458	12.07	General function prediction only
S	286	7.54	Function unknown
-	1,261	27.28	Not in COG

a) The total is based on the total number of protein coding genes in the annotated genome.

Genome comparisons

A comparative analysis of genome sizes and protein coding genes in *Arthrobacter* sp. FB24 and other *Arthrobacter* species with finished sequences (Table 6) was made from data listed on the IMG website [34]. Included in the comparison is *A. arilaitensis* re117 (Gc01419, FQ311476) [8], *A. aurescens* TC1 (Gc00480, NC_008709) [9], *A. chlorophenolicus* A6 (Gc00930, NC_011886), *A. nitroguajacolicus* Rue61a (Gc0006272, CP003203), and *A. phenanthrenivorans* Sphe3 (Gc01621, CP002379) [10]. In addition, the draft genome of *A. globiformis* NBRC 12137 was included because its phylogenetic relatedness to FB24 based on the 16S rRNA gene sequence. Similarity between functional protein groups (based on COG,

clusters of orthologous groups) in the genomes of these strains were made and visualized using hierarchical clustering (Figure 5) with tools available on the Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) site. Also included in the tree were closely related species in the family *Micrococcaceae* with finished genomes *Kocuria rhizophila* DC2201 (Gc00769), *Micrococcus luteus* Fleming NCTC 2665 (Gc01033), *Renibacterium salmoninarum* ATCC 33209 (Gc00698), *Rothia dentocariosa* ATCC 17931 (Gc01662), and *Rothia mucilaginosa* DY-18 (Gc01162). Detailed information about the genome properties and genome annotation of these strains can be obtained from the JGI-IMG website at the JGI website [35].

Table 6. Comparison of genomes of the genus *Arthrobacter* with finished genome sequences

Genome Name	Genome size (bp)	Gene count	Protein coding	Protein with function	Without function	Plasmid number	rRNA operons
<i>Arthrobacter arilaitensis</i> re117, CIP108037	3,918,192	3,518	3,436	2,390	1,046	2	6
<i>Arthrobacter aurescens</i> TC1	5,226,648	4,793	4,699	3,419	1,280	2	6
<i>Arthrobacter chlorophenolicus</i> A6	4,980,870	4,744	4,641	3,125	1,516	2	5
<i>Arthrobacter nitroguajacolicus</i> Rue61a	5,081,038	4,655	4,584	3,800	784	2	6
<i>Arthrobacter phenanthrenivorans</i> Sphe3	4,535,320	4,273	4,209	3,101	1,108	2	4
<i>Arthrobacter</i> sp. FB24	5,070,478	4,622	4,536	3,279	1,257	3	5
<i>Arthrobacter globiformis</i> NBRC 12137*	4,954,410	4,582	4,529	2,784	1,745	?	1

*Sequence not fully assembled

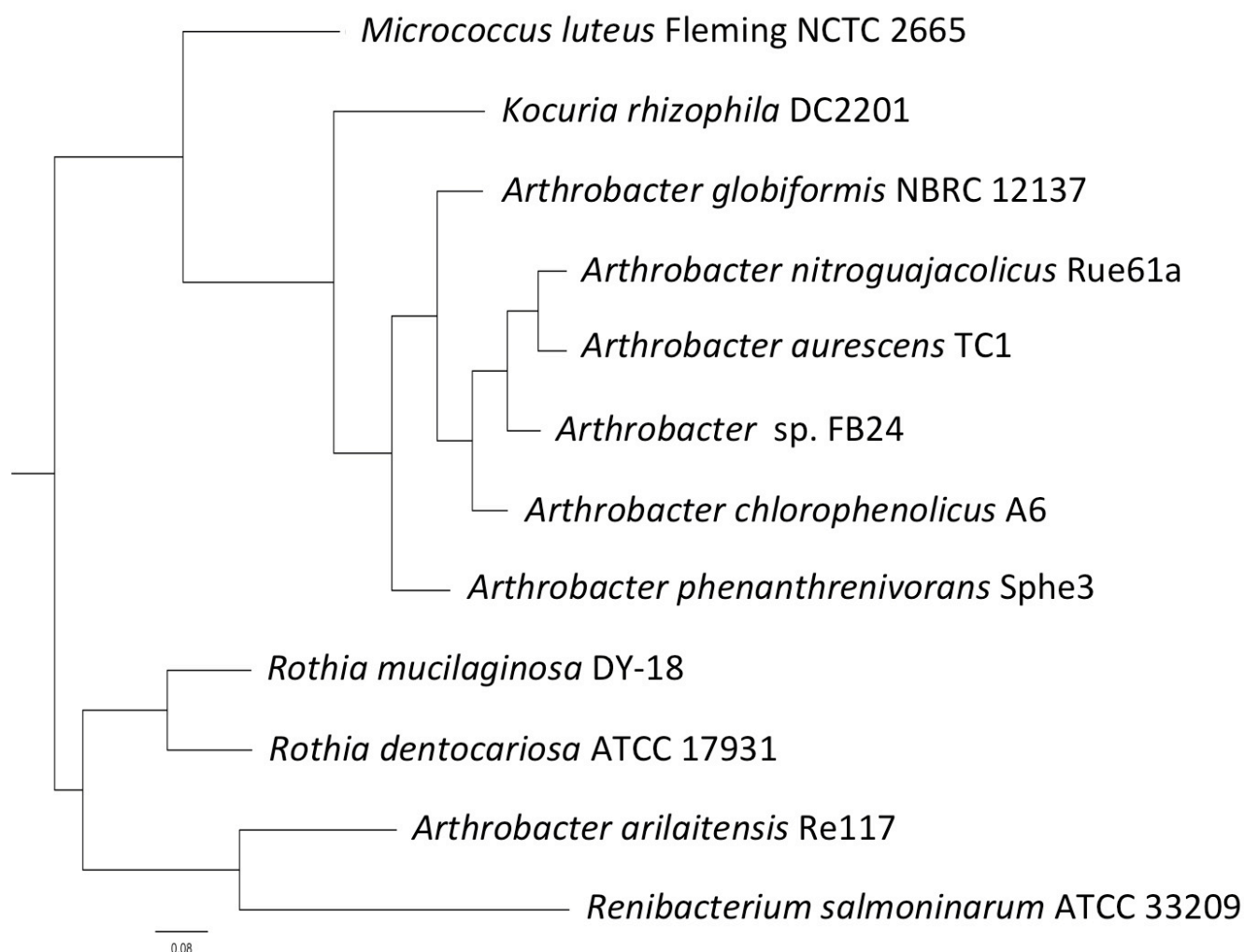


Figure 5. Hierarchical tree based on similarity of COG groups between genomes. Included are genomes of bacteria in the family *Micrococcaceae* with finished genome sequences.

References

- Joynt J, Bischoff M, Turco RF, Konopka A, Nakatsu CH. Microbial community analysis of soils contaminated with lead, chromium and organic solvents. *Microb Ecol* 2006; **51**:209-219. [PubMed](http://dx.doi.org/10.1007/s00248-005-0205-0) <http://dx.doi.org/10.1007/s00248-005-0205-0>
- Nakatsu CH, Carmosini N, Baldwin B, Beasley F, Kourtev P, Konopka A. Soil microbial community responses to additions of organic carbon substrates and heavy metals (Pb and Cr). *Appl Environ Microbiol* 2005; **71**:7679-7689. [PubMed](http://dx.doi.org/10.1128/AEM.71.12.7679-7689.2005) <http://dx.doi.org/10.1128/AEM.71.12.7679-7689.2005>
- Henne KL, Nakatsu CH, Thompson DK, Konopka AE. High-level chromate resistance in *Arthrobacter sp.* strain FB24 requires previously uncharacterized accessory genes. *BMC Microbiol* 2009; **9**:199. [PubMed](http://dx.doi.org/10.1186/1471-2180-9-199) <http://dx.doi.org/10.1186/1471-2180-9-199>
- Henne KL, Turse JE, Nicora CD, Lipton MS, Tollaksen SL, Lindberg C, Babnigg G, Giometti CS, Nakatsu CH, Thompson DK, et al. Global proteomic analysis of the chromate response in *Arthrobacter sp.* strain FB24. *J Proteome Res* 2009; **8**:1704-1716. [PubMed](http://dx.doi.org/10.1021/pr800705f) <http://dx.doi.org/10.1021/pr800705f>
- Becker JM, Parkin T, Nakatsu CH, Wilbur JD, Konopka A. Bacterial activity, community structure, and centimeter-scale spatial heterogeneity in contaminated soil. *Microb Ecol* 2006; **51**:220-231. [PubMed](http://dx.doi.org/10.1007/s00248-005-0002-9) <http://dx.doi.org/10.1007/s00248-005-0002-9>

6. Kourtev PS, Nakatsu CH, Konopka A. Responses of the anaerobic bacterial community to addition of organic C in chromium(VI)- and iron(III)-amended microcosms. *Appl Environ Microbiol* 2006; **72**:628-637. [PubMed](#)
<http://dx.doi.org/10.1128/AEM.72.1.628-637.2006>
7. Kourtev PS, Nakatsu CH, Konopka A. Inhibition of nitrate reduction by chromium(VI) in anaerobic soil microcosms. *Appl Environ Microbiol* 2009; **75**:6249-6257. [PubMed](#)
<http://dx.doi.org/10.1128/AEM.00347-09>
8. Monnet C, Loux V, Gibrat J-Fo, Spinnler E, Barbe Vr, Vacherie B, Gavory F, Gourbeyre E, Siguier P, Chandler Ml and others. The *Arthrobacter arilaitensis* Re117 Genome Sequence Reveals Its Genetic Adaptation to the Surface of Cheese. *PLoS ONE* 2010; **5**:e15489. [PubMed](#)
<http://dx.doi.org/10.1371/journal.pone.0015489>
9. Mongodin EF, Shapir N, Daugherty SC, Deboy RT, Emerson JB, Shvartzbeyn A, Radune D, Vamathevan J, Riggs F, Grinberg V, *et al.* Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1. *PLoS Genet* 2006; **2**:e214. [PubMed](#)
<http://dx.doi.org/10.1371/journal.pgen.0020214>
10. Kallimanis A, LaButti K, Lapidus A, Clum A, Lykidis A, Mavromatis K, Pagani I, Liolios K, Ivanova N, Goodwin L, *et al.* Complete genome sequence of *Arthrobacter phenanthrenivorans* type strain (Sphe3T). *Stand Genomic Sci* 2011; **4**:123-130. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.1393494>
11. Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* 2007; **35**:D169-D172. [PubMed](#)
<http://dx.doi.org/10.1093/nar/gkl889>
12. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**:2947-2948. [PubMed](#)
<http://dx.doi.org/10.1093/bioinformatics/btm404>
13. 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>
14. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW, editors. *Bergey's Manual of Systematic Bacteriology*. Second Edition ed. Volume 1. New York: Springer; 2001. p 119-169.
15. 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.
<http://dx.doi.org/10.1099/00207713-47-2-479>
16. 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 Bacteriol* 2009; **59**:589-608. [PubMed](#)
17. 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>
18. Buchanan RE. Studies in the nomenclature and classification of bacteria. II. The primary subdivisions of the *Schizomycetes*. *J Bacteriol* 1917; **2**:155-164. [PubMed](#)
19. Pribram E. A contribution to the classification of microorganisms. *J Bacteriol* 1929; **18**:361-394. [PubMed](#)
20. Conn HJ, Dimmick I. Soil bacteria similar in morphology to *Mycobacterium* and *Corynebacterium*. *J Bacteriol* 1947; **54**:291-303. [PubMed](#)
21. Koch C, Schumann P, Stackebrandt E. Reclassification of *Micrococcus agilis* (Ali-Cohen 1889) to the genus *Arthrobacter* as *Arthrobacter agilis* comb. nov. and emendation of the genus *Arthrobacter*. *Int J Syst Bacteriol* 1995; **45**:837-839. [PubMed](#)
<http://dx.doi.org/10.1099/00207713-45-4-837>
22. Keddie RM. Genus II. *Arthrobacter* Conn and Dimmick 1947, 300. In: Buchanan RE, Gibbons NE (eds), *Bergey's Manual of Determinative Bacteriology*, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 618-625.
23. Judicial Commission. Opinion 24. Rejection of the Generic Name *Arthrobacter* Fischer 1895 and Conservation of the Generic Name *Arthrobacter* Conn and Dimmick 1947. *Int Bull Bacteriol Nomencl Taxon* 1958; **8**:171-172.
24. 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](http://dx.doi.org/10.1038/75556)
25. 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](http://dx.doi.org/10.1038/nbt1360)
26. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, eds. *Current Protocols in Molecular Biology*. Hoboken NJ: John Wiley & Sons; 2003.
27. Han C, Chain P. Finishing repetitive regions automatically with Dupfinisher. In: Arabnia HR, Valafar H, editors 2006; Las Vegas, Nevada, USA. CSREA Press. p 142-147.
28. Badger JH, Olsen GJ. CRITICA: Coding region identification tool invoking comparative analysis. *Mol Biol Evol* 1999; **16**:512-524. [PubMed http://dx.doi.org/10.1093/oxfordjournals.molbev.a026133](http://dx.doi.org/10.1093/oxfordjournals.molbev.a026133)
29. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 1999; **27**:4636-4641. [PubMed http://dx.doi.org/10.1093/nar/27.23.4636](http://dx.doi.org/10.1093/nar/27.23.4636)
30. Lowe TM, Eddy SR. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res* 1997; **25**:0955-964.
31. Eddy SR. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. *BMC Bioinform* 2002; **3**.
32. Karp PD, Paley S, Romero P. The Pathway Tools software. *Bioinformatics* 2002; **18**:S225-S232. [PubMed http://dx.doi.org/10.1093/bioinformatics/18.suppl_1.S225](http://dx.doi.org/10.1093/bioinformatics/18.suppl_1.S225)
33. Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, *et al.* MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* 2006; **34**:D511-D516. [PubMed http://dx.doi.org/10.1093/nar/gkj128](http://dx.doi.org/10.1093/nar/gkj128)
34. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Anderson I, Lykidis A, Mavromatis K, *et al.* The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Res* 2010; **38**:D382-D390. [PubMed http://dx.doi.org/10.1093/nar/gkp887](http://dx.doi.org/10.1093/nar/gkp887)
35. DOE Joint Genome Institute. <http://img.jgi.doe.gov/cgi-bin/w/main.cgi>