

# High quality draft genome sequence of *Segniliparus rugosus* CDC 945<sup>T</sup> (= ATCC BAA-974<sup>T</sup>)

Ashlee M. Earl<sup>1\*</sup>, Christopher A. Desjardins<sup>1</sup>, Michael G. Fitzgerald<sup>1</sup>, Harindra M. Arachchi<sup>1</sup>, Qiandong Zeng<sup>1</sup>, Teena Mehta<sup>1</sup>, Allison Griggs<sup>1</sup>, Bruce W. Birren<sup>1</sup>, Nadege C. Toney<sup>3</sup>, Janice Carr<sup>3</sup>, James Posey<sup>2</sup> and W. Ray Butler<sup>2</sup>

<sup>1</sup>Genome Sequencing and Analysis Program, Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA, USA

<sup>2</sup>Division of Tuberculosis Elimination, Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

<sup>3</sup>Division of Healthcare Quality Promotion, Clinical and Environmental Microbiology Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

**Keywords:** *Segniliparaceae*, genome sequencing, Human Microbiome Project

\*Corresponding author Ashlee M. Earl

---

*Segniliparus rugosus* represents one of two species in the genus *Segniliparus*, the sole genus in the family *Segniliparaceae*. A unique and interesting feature of this family is the presence of extremely long carbon-chain length mycolic acids bound in the cell wall. *S. rugosus* is also a medically important species because it is an opportunistic pathogen associated with mammalian lung disease. This report represents the second species in the genus to have its genome sequenced. The 3,567,567 bp long genome with 3,516 protein-coding and 49 RNA genes is part of the NIH Roadmap for Medical Research, Human Microbiome Project.

---

## Introduction

Strain CDC 945<sup>T</sup> (= ATCC BAA-974<sup>T</sup> = CIP 10838<sup>T</sup> = DSM 45345 = CCUG 50838<sup>T</sup> = JCM 13579<sup>T</sup>) is the type strain of the species *Segniliparus rugosus* in the *Segniliparaceae* family [1]. The genus name was created to acknowledge the presence of novel long carbon-chain fatty acids (mycolic acids) detected using the *Mycobacterium* species identification method with high performance liquid chromatography (HPLC) [2]. The name was formed from the Latin adjective 'segnis', meaning 'slow' and combined with the Greek adjective 'liparos' for 'fatty', to indicate the 'one with slow fats'. The name relates to the late elution of the apolar, alpha-mycolic acids (fatty acids) during HPLC analysis [1]. The specific epithet for the taxon name is from the Latin adjective 'rugosus', referring to the formation of wrinkled, rough colony morphology [1]. The type strain of *S. rugosus*, CDC 945<sup>T</sup>, was isolated from a human sputum specimen collected in Alabama, USA [1]. *S. rugosus* has been isolated from multiple patients with cystic fibrosis in the U.S. and Australia and appears to be a respiratory opportunistic pathogen [3,4]. A recent isolation

from a ~1 year old sea lion showing third-stage malnutrition with a 30% loss of body weight, moderate bradycardia and severe hypothermia, suggests a possible aquatic or marine niche for the species [5]. The only other validly named species of the genus is *Segniliparus rotundus* (CDC 1076<sup>T</sup>), which is the type strain of this species. *S. rotundus*. CDC 1076 shares 98.9% 16S rRNA sequence identity with *S. rugosus* CDC 945<sup>T</sup>, although the DNA-DNA hybridization is less than 28% [1]. The complete genome of *S. rotundus* was recently reported and has 3,157,527 bp with 3,081 protein-coding and 52 RNA genes [6]. Here we present a summary classification and a set of features for *S. rugosus* CDC 945<sup>T</sup>, together with the description of the high quality draft genomic sequencing and annotation.

## Classification and features

The cells of CDC 945<sup>T</sup> are irregular rods ranging in length and width from 0.55-0.90 μm by 1.9-4.5 μm (Table 1 and Figure 1). Colonies are wrinkled,

rough and form in less than 7 days on 7H10 and 7H11 agar at an optimal temperature of 33°C [1]. CDC 945<sup>T</sup> is aerobic, non-motile, asporogenous, and stains bright red with acid alcohol stain [1]. It is mesophilic and demonstrates a temperature range for growth between 22 and 42 °C [1]. Colonies grown for < 4 weeks are non-pigmented, nonphotochromogenic and do not produce a diagnostic odor [1]. They do not produce aerial mycelium, spores or demonstrate true branching. Young colonies are creamy and smear easily when disturbed. Cell growth at ~4 weeks on Löwenstein-Jensen (LJ) medium produces a diffusible pink color in the agar at the leading edges of mature growth. Aged colonies on LJ develop a light buff pigment and demonstrate 'greening' from uptake of malachite green [1]. CDC 945<sup>T</sup> is weakly positive for arylsulfatase at 7 days but is strongly positive in 14 days. No growth on MacConkey agar not containing crystal violet. CDC 945<sup>T</sup> grows in the presence of 5% sodium chloride at 7 days, in lysozyme at 21 days. Positive for iron uptake, nitrate reduction, tellurite reduction and tween opacity. Negative for tween hydrolysis [1].

Results with the API CORYNE test kit shows CDC 945<sup>T</sup> is positive for β-glucosidase, and pyrazinamidase activities and negative for alkaline phosphatase, β-galactosidase, β-glucuronidase, α-glucosidase, *N*-acetyl-β-glucosaminidase and pyrrolidonyl arylamidase activity at 33°C [1]. It is susceptible to imipenem 4ug/ml, moxifloxacin 0.5 μg/ml, and trimethoprim-sulfamethoxazole ≤ 4.8 ug/ml, intermediate to cefoxitan 64 ug/ml and resistant to amikacin >128 ug/ml, clarithromycin 32 ug/ml, ciprofloxacin 16 ug/ml, ethambutol >16 ug/ml and tobramycin >64 ug/ml. [1,3]. Strain CDC 945<sup>T</sup> uses D-glucose, glycerol, maltose, mannitol, D-sorbitol and trehalose as sole carbon sources with the production of acid. No growth on adonitol, L-arabinose, cellobiose, citrate, dulcitol, i-erythriol, galactose, i-*myo*-inositol, lactose, mannose, melibiose, raffinose, L-rhamnose, salicin or sodium citrate [1]. The strain hydrolyzes urea but not acetamide adenine, casein, aesculin, hypoxanthine, tyrosine or xanthine [1].

### Chemotaxonomy

The cell wall of strain CDC 945<sup>T</sup> contains mycolic acids and meso-diaminopimelic acid [1]. The mycolic acid pattern developed with HPLC is a double cluster of peaks emerging at 7.24 min and the last

peak group is unresolved and elutes slightly before the 110 carbon chain length, high molecular weight internal standard [1,2]. Thin layer chromatography confirms 2 groups of apolar, α- and α'-alpha-mycolic acids lacking oxygen function, other than the hydroxyl group [1]. The HPLC and TLC results indicate that this strain produces a unique homologous subclass of long, alpha-mycolic acids with additional 90 to 110 carbons [1]. The fatty acid profile by gas-liquid chromatography is C<sub>10:0</sub> (8.65%), C<sub>12:0</sub> (1.33%), C<sub>14:0</sub> (8.49%), C<sub>16:0</sub> (18.34%), C<sub>18:1ω9c</sub> (8.93%), C<sub>18:0 10-methyl</sub> (tuberculos-tearic acid, 21.62%), and C<sub>20</sub> (28.51%) [1].

The phylogenetic association of *Segniliparus* species is shown in Figure 2 in a 16S rRNA based tree. The genus forms a distinct lineage relative to the other mycolic acid containing *Actinobacteria*. The positioning of *Segniliparus* in this phylogenetic analysis is consistent with its positioning in the "All-Species Living Tree Project" LTP release 106, August 2011, which is similarly based on 16S rRNA. [11].

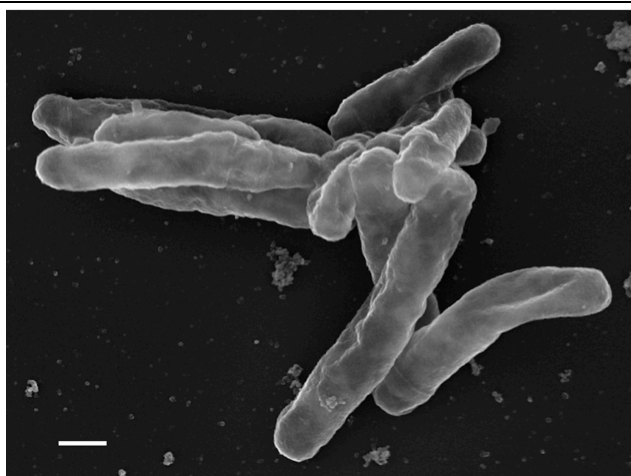
### Genome sequencing and annotation

#### Genome project history

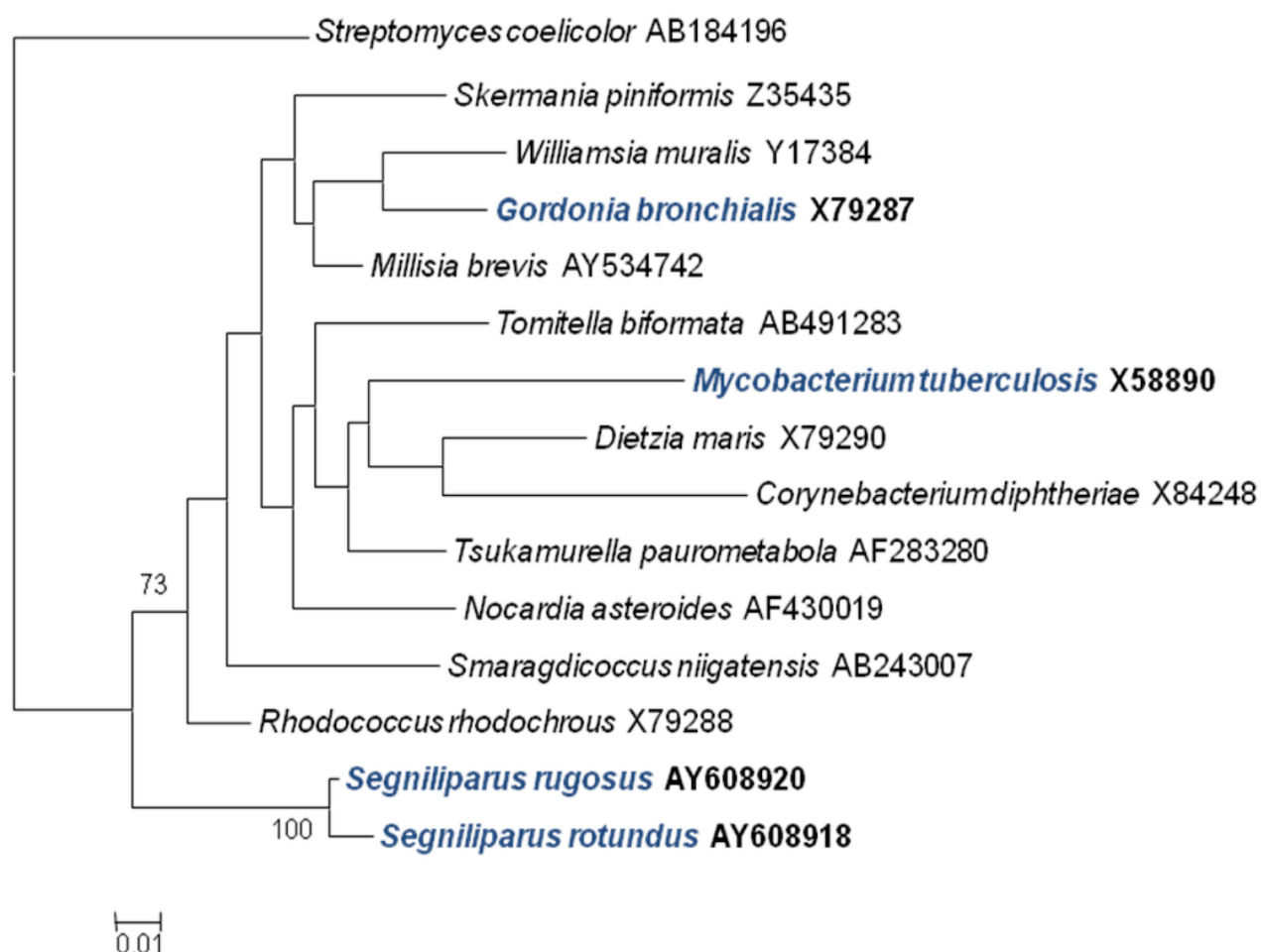
This organism was selected for sequencing on the basis of its association with respiratory lung disease and is part of the NIH Roadmap for Medical Research, Human Microbiome Project (HMP) [12]. The HMP presents reference genomes in Genome Online Database (GOLD) [10], the Human Microbiome Project Data Analysis and Coordination Center Project Catalog [13] and the complete high quality draft genome sequence is deposited in GenBank [14]. The Broad Institute performed the sequencing and annotation of this high quality draft genome [15]. A summary of the project is given in Table 2.

#### Growth conditions and DNA isolation

Strain CDC 945<sup>T</sup> was grown statically in Middlebrook 7H9 medium at 33°C until late log. DNA was isolated from whole cells after a chloroform/methanol wash with a disruption solution of guanidine thiocyanate, sarkosyl and mercaptoethanol as described in Mve-Obiang et al. [16]. The purity of DNA was assessed by The Broad Institute using the Quant-iT™ dsDNA Assay High Sensitivity Kit (Invitrogen, Carlsbad, CA) and according to the manufacturer's protocol.



**Figure 1.** Scanning electron micrograph of *S. rugosus* CDC 945<sup>T</sup>. The scale bar is 667nm.



**Figure 2.** Maximum likelihood phylogenetic tree was generated using PHYLML v2.2.4 [7] based on 16S rRNA sequences highlighting the position of *S. rugosus* CDC 945<sup>T</sup> relative to the other type strains of mycolic acid containing genera in the suborder *Corynebacterineae*. GenBank accession numbers are listed after the name. The tree was inferred from 1,468 bp positions aligned using Clustal W [8] in MEGA v4 [9]. Numbers at the branch nodes are support values from 1,000 bootstrap replicates if equal to or greater than 70%. The scale bar indicates substitutions per site. The tree was rooted with *Streptomyces coelicolor*. Lineages with type strain genome sequencing projects registered in GOLD [10] are shown in blue, published genomes in bold.

**Table 1.** Classification and general features of *S. rugosus* CDC 945<sup>T</sup> according to the MIGS recommendations [25].

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [26]
		Phylum <i>Actinobacteria</i>	TAS [27]
		Class <i>Actinobacteria</i>	TAS [28]
		Subclass <i>Actinobacteridae</i>	TAS [28]
	Current classification	Order <i>Actinomycetales</i>	TAS [28,29]
		Suborder <i>Corynebacterineae</i>	TAS [28,29]
		Family <i>Segniliparaceae</i>	TAS [1,29]
		Genus <i>Segniliparus</i>	TAS [1]
		Species <i>Segniliparus rugosus</i>	TAS [1]
		Type strain CDC 945	TAS [1]
	Gram stain	not reported	
	Cell shape	rods, irregular	TAS [1]
	Motility	nonmotile	TAS [1]
	Sporulation	non-sporulating	TAS [1]
	Temperature range	mesophile, 22-42 °C	TAS [1]
	Optimum temperature	33°C	TAS [1]
	Salinity	unknown	
MIGS-22	Oxygen requirement	aerobic	TAS [1]
	Carbon source	D-glucose, glycerol, maltose, mannitol, D-sorbitol and trehalose	TAS [1]
	Energy source	chemoorganotroph	TAS [1]
MIGS-6	Habitat	environmental water suggested	TAS [5]
MIGS-15	Biotic relationship	likely free-living	NAS [5]
MIGS-14	Pathogenicity	opportunistic pathogen	TAS [1,3]
	Biosafety level	2	TAS [1,3,30]
	Isolation	sputum, human	TAS [1,3]
MIGS-4	Geographic location	Alabama, USA	TAS [1]
MIGS-4.1	Latitude	not reported	
MIGS-4.2	Longitude	not reported	
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	not reported	
MIGS-5	Sample collection time	1998	TAS [1]

Evidence codes- 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 [31].

**Table 2.** Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High Quality Draft
MIGS-28	Libraries used	Two 454 pyrosequence libraries, one standard 0.6kb fragment library and one 2.5kb jump library
MIGS-29	Sequencing platforms	454 Titanium
MIGS-31.2	Sequencing coverage	13×
MIGS-30	Assemblers	Newbler Assembler version 2.3 PostRelease-11/19/2009
MIGS-32	Gene calling method	Glimmer; Metagene; PFAM; BLAST to non-redundant protein database; manual curation
	Genbank ID	ACZI01000000
	Genbank Date of Release	November 10, 2010
	GOLD ID	Gi05259
	NCBI project ID	40685
MIGS-13	Source material identifier	ATCC BAA-974 <sup>T</sup>
	Project relevance	Human Microbiome Project

### Genome sequencing and assembly

The genome of *Segniliparus rugosus* ATCC BAA-974 was sequenced using 454 pyrosequence fragment and jump libraries [17]. We assembled the 454 data, consisting of 135,510 fragment reads and 112,271 jump reads, using Newbler Assembler version 2.3 PostRelease-11/19/2009. The assembly is considered High-Quality Draft and consists of 262 contigs arranged in 30 scaffolds with a total size of 3,567,567 bases. The error rate of this draft genome sequence is less than 1 in 10,000 (accuracy of ~ Q40). Average sequence coverage is 13×. Assessment of coverage, GC content, contig BLAST and 16S contig classification were consistent with the species *Segniliparus*.

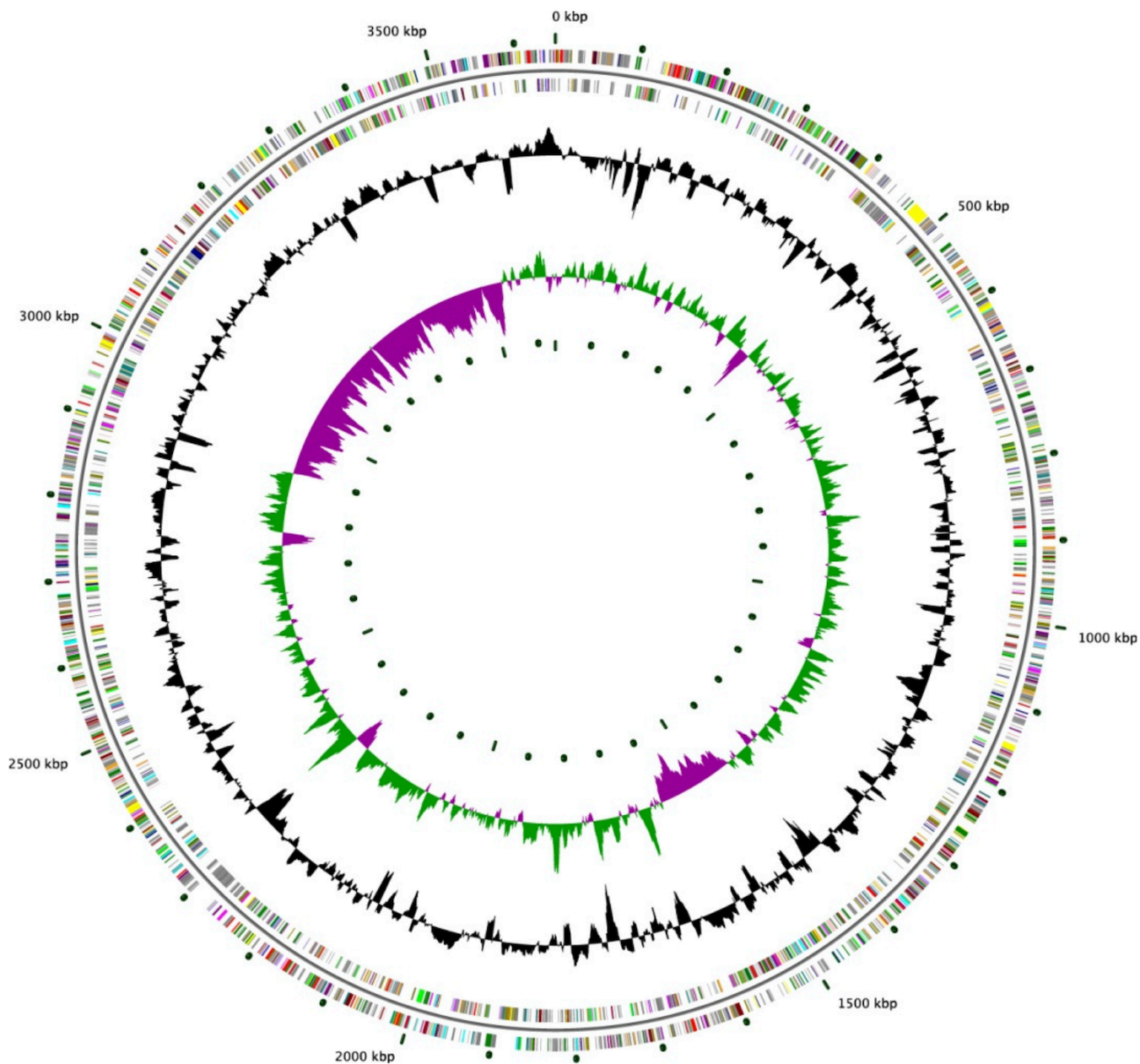
### Genome annotation

Protein-coding genes were predicted using four ORF-finding tools: GeneMark [18], Glimmer3 [19], Metagene [20], and findBlastOrfs (unpublished). This latter tool builds genes by extending whole-genome blast alignments, in-frame, to include start and stop codons. The final set of non-overlapping ORFs was selected from the output of these tools using an in-house gene-caller, which uses dynamic programming to score candidate gene models based on strength of similarity to entries in UniRef90, then

selects non-overlapping genes that, combined, have the highest overall score. In cases where predictions overlapped non-coding RNA features (see below), the genes were manually inspected and removed when necessary. Finally, the gene set was reviewed using both the NCBI discrepancy report and the internal Broad annotation metrics. Ribosomal RNAs (rRNAs) were identified with RNAmmer [21]. The tRNA features were identified using tRNAScan [22]. Other non-coding features were identified with Rfam [23]. The gene product names were assigned based on Hmmer equivalogs from TIGRfam and Pfam, and blast hits to KEGG and SwissProt protein sequence databases. This was done using the naming tool “Pidgin” [24].

### Genome properties

This 3,567,567 bp draft genome has high G+C content (Table 3 and Figure 3) and is predicted to encode 3,571 genes, 98% of which are protein coding. Nearly 70% of predicted proteins have a functional prediction and COG functional categories have been assigned to 53% of predicted proteins (Table 4).



**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), GC content, GC skew.

**Table 3.** Genome Statistics

<b>Attribute</b>	<b>Value</b>	<b>% of Total</b>
Genome size (bp)	3,647,826	100.00%
DNA coding region (bp)	3,156,492	86.35%
DNA G+C content (bp)	2,484,899	68.12%
Number of replicons	unknown	
Extrachromosomal elements	unknown	
Total genes	3,571	100.0%
tRNA genes	46	1.28%
rRNA genes	3	0.08%
rRNA operons	1	0.03%
CRISPR repeats	0	
Protein-coding genes	3,522	98.62%
Pseudo genes (partial genes)	6 (233)	0.17% (6.52%)
Genes with function prediction	2,486	69.62%
Genes in paralog clusters	194	5.43%
Genes assigned to COGS	1,897	53.12%
Genes assigned Pfam domains	2,451	68.64%
Genes with signal peptides	412	11.54%
Genes with transmembrane helices	590	16.52%

**Table 4.** Number of genes associated with the general COG functional categories

<b>Code</b>	<b>Value</b>	<b>%age</b>	<b>Description</b>
J	121	3.4	Translation, ribosomal structure and biogenesis
A	1	0.0	RNA processing and modification
K	89	2.5	Transcription
L	84	2.4	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	18	0.5	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	22	0.6	Defense mechanisms
T	51	1.4	Signal transduction mechanisms
M	85	2.4	Cell wall/membrane/envelope biogenesis
N	3	0.1	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	11	0.3	Intracellular trafficking and secretion, and vesicular transport
O	77	2.2	Posttranslational modification, protein turnover, chaperones
C	140	4.0	Energy production and conversion
G	100	2.8	Carbohydrate transport and metabolism
E	225	6.4	Amino acid transport and metabolism
F	74	2.1	Nucleotide transport and metabolism
H	102	2.9	Coenzyme transport and metabolism
I	120	3.4	Lipid transport and metabolism
P	110	3.1	Inorganic ion transport and metabolism
Q	92	2.6	Secondary metabolites biosynthesis, transport and catabolism
R	246	7.0	General function prediction only
S	126	3.6	Function unknown
-	1625	46.1	Not in COGs

## Acknowledgements

The authors gratefully acknowledge the Broad Genome Sequencing Platform, Lucia Alvarado-Balderrama for data submission and Susanna G. Hamilton for project

management. We also acknowledge NIH for funding this project with grants to the Broad Institute (grants HHSN272200900017C and U54-HG004969).

## References

1. Butler WR, Floyd MM, Brown JM, Toney SR, Daneshvar MI, Cooksey RC, Carr J, Steigerwalt AG, Charles N. Novel mycolic acid-containing bacteria in the family *Segniliparaceae* fam. nov., including the genus *Segniliparus* gen. nov., with descriptions of *Segniliparus rotundus* sp. nov. and *Segniliparus rugosus* sp. nov. *Int J Syst Evol Microbiol* 2005; **55**:1615-1624. [PubMed](#) [doi:10.1099/ijs.0.63465-0](https://doi.org/10.1099/ijs.0.63465-0)
2. Butler WR, Guthertz LS. Mycolic acid analysis by high-performance liquid chromatography for identification of *Mycobacterium* species. *Clin Microbiol Rev* 2001; **14**:704-726. [PubMed](#) [doi:10.1128/CMR.14.4.704-726.2001](https://doi.org/10.1128/CMR.14.4.704-726.2001)
3. Butler WR, Sheils CA, Brown-Elliott BA, Charles N, Colin AA, Gant MJ, Goodill J, Hindman D, Toney SR, Wallace RJ, Jr., et al. First isolations of *Segniliparus rugosus* from patients with cystic fibrosis. *J Clin Microbiol* 2007; **45**:3449-3452. [PubMed](#) [doi:10.1128/JCM.00765-07](https://doi.org/10.1128/JCM.00765-07)
4. Hansen T, Van-Kerckhof J, Jelfs P, Wainwright C, Ryan P, Coulter C. *Segniliparus rugosus* infection, Australia. *Emerg Infect Dis* 2009; **15**:611-613. [PubMed](#) [doi:10.3201/eid1504.081479](https://doi.org/10.3201/eid1504.081479)
5. Evans RH. *Segniliparus rugosus*-associated bronchiolitis in California sea lion. *Emerg Infect Dis* 2011; **17**:311-312. [PubMed](#)
6. Sikorski J, Lapidus A, Copeland A, Misra M, Glavina Del Rio T, Nolan M, Lucas S, Chen F, Tice H, Cheng JF, et al. Complete genome sequence of *Segniliparus rotundus* type strain (CDC 1076). *Stand Genomic Sci* 2010; **2**:203-211. [PubMed](#) [doi:10.4056/sigs.791633](https://doi.org/10.4056/sigs.791633)
7. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003; **52**:696-704. 54QH07WB5K5XCX4 [pii]
8. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**:4673-4680. [PubMed](#) [doi:10.1093/nar/22.22.4673](https://doi.org/10.1093/nar/22.22.4673)
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](#) [doi:10.1093/molbev/msm092](https://doi.org/10.1093/molbev/msm092)
10. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010; **38**:D346-D354.
11. Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer KH, Glockner FO, Rossello-Mora R. Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Syst Appl Microbiol* 2010; **33**:291-299. [PubMed](#) [doi:10.1016/j.syapm.2010.08.001](https://doi.org/10.1016/j.syapm.2010.08.001)
12. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, et al. The NIH Human Microbiome Project. *Genome Res* 2009; **19**:2317-2323. [PubMed](#) [doi:10.1101/gr.096651.109](https://doi.org/10.1101/gr.096651.109)
13. The National Institutes of Health. Data Analysis and Coordination Center for Human Microbiome Project. <http://www.hmpdacc.org>.
14. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2011; **39**:D32-D37. [PubMed](#) [doi:10.1093/nar/gkq1079](https://doi.org/10.1093/nar/gkq1079)
15. The Broad Institute. <http://www.broadinstitute.org>.
16. Mve-Obiang A, Mestdagh M, Portaels F. DNA isolation from chloroform/methanol-treated mycobacterial cells without lysozyme and proteinase K. *Biotechniques* 2001; **30**:272-274, 276. [PubMed](#)
17. Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W, Daza R, Erlich RL, Gianoukos G, Green L, et al. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biol* 2010; **11**:R15. [PubMed](#) [doi:10.1186/gb-2010-11-2-r15](https://doi.org/10.1186/gb-2010-11-2-r15)
18. Borodovsky M, McIninch J. GENMARK: Parallel gene recognition for both DNA strands. *Comput Chem* 1993; **17**:123-133. [doi:10.1016/0097-8485\(93\)85004-V](https://doi.org/10.1016/0097-8485(93)85004-V)
19. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification



- with GLIMMER. *Nucleic Acids Res* 1999; **27**:4636-4641. [PubMed](#) [doi:10.1093/nar/27.23.4636](https://doi.org/10.1093/nar/27.23.4636)
20. Noguchi H, Park J, Takagi T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* 2006; **34**:5623-5630. [PubMed](#) [doi:10.1093/nar/gkl723](https://doi.org/10.1093/nar/gkl723)
  21. 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. [PubMed](#) [doi:10.1093/nar/gkm160](https://doi.org/10.1093/nar/gkm160)
  22. 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](#) [doi:10.1093/nar/25.5.955](https://doi.org/10.1093/nar/25.5.955)
  23. 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-D124. [PubMed](#) [doi:10.1093/nar/gki081](https://doi.org/10.1093/nar/gki081)
  24. The Broad Institute. Automated gene naming tool. <http://genepidgin.sourceforge.net>.
  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](#) [doi:10.1038/nbt1360](https://doi.org/10.1038/nbt1360)
  26. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](#) [doi:10.1073/pnas.87.12.4576](https://doi.org/10.1073/pnas.87.12.4576)
  27. 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. New York: Springer; 2001. p119-169.
  28. Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a New Hierarchic Classification System, *Actinobacteria* classic nov. *Int J Syst Bacteriol* 1997; **47**:479-491. [doi:10.1099/00207713-47-2-479](https://doi.org/10.1099/00207713-47-2-479)
  29. Zhi XY, Li WJ, Stackebrandt E. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *Actinobacteria*, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int J Syst Evol Microbiol* 2009; **59**:589-608. [PubMed](#) [doi:10.1099/ijs.0.65780-0](https://doi.org/10.1099/ijs.0.65780-0)
  30. Chosewood L, Wilson D, eds. Biosafety in Microbiological and Biomedical Laboratories (BMBL). 5th ed. rev. Dec. 2009 ed. Washington D.C.: US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute of Health; 2009. pnas.050566797. 8. Ironside JW and JE Bell.
  31. 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. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](#) [doi:10.1038/75556](https://doi.org/10.1038/75556)