Genome Sequence of *Afipia birgiae*, a Rare Bacterium Associated with Amoebae

Isabelle Pagnier, Olivier Croce, Catherine Robert, Didier Raoult, and Bernard La Scola

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, CNRS, UMR 6236 – IRD 198, Faculté de Médecine, Aix-Marseille Université, Marseille, France

***Afipia birgiae* is an alphaproteobacterium from the family *Bradyrhizobiaceae*, growing in amoebae, and a potential human pathogen. We sequenced the genome of type strain 34632^T. It is composed of 5,325,467 bp and contains 5,160 protein-coding genes and 53 RNA genes, including 3 rRNA genes.**

Afipia birgiae was first isolated from a hospital water supply using a coculture procedure with *Acanthamoeba polyphaga* amoebae (8, 9). This Gram-negative bacillus is classified within the alphaproteobacteria in the family *Bradyrhizobiaceae*. Like *Legionella* spp., which are also amoeba-associated microorganisms, it grows well on BCYE agar but weakly or not at all on Columbia sheep blood agar. The genus *Afipia* currently includes 5 species (*A. birgiae*, *A. broomeae*, *A. clevelandensis*, *A. felis*, and *A. massiliensis*) and several genospecies, and many other probable new species have been detected in water and soil (15). Members of the genus are common in water, including hospital water supplies, and are suspected to be agents of hospital-acquired pneumonia (7), but *A. birgiae* is rarely isolated.

The genome was pyrosequenced with a 454 GS FLX titanium platform (Roche, Branford, CT) (11) using paired-end and shotgun techniques. A total of 512,586 reads were assembled with Newbler software v2.6 (Roche). Scaffolding was improved using Opera software v1.1 (4), combined with GapFiller v1.10 (3) and manual curation.

The draft genome of *A. birgiae* 34632^T consists of 65 contigs containing 5,302,490 bp and an estimated size, including gaps, of 5,325,467 bp. The G+C content of this genome is 60.58%, which is similar to that of the other completed genome of *Afipia* (*Afipia* sp. strain 1NLS2, with 61% G+C). Using BLASTN, Aragorn (10), and RNAmmer (6), the genome was shown to encode 53 RNA genes, including 3 rRNAs in a single operon and 50 tRNAs. However, considering the large size of the genome, the gaps between contigs could have included other RNA operons.

Potential coding sequences were predicted using Prodigal software (5). Assignment of protein functions was performed by searching against GenBank, Clusters of Orthologous Groups (COG), and Pfam databases using BLASTP (1, 12, 14). A total of 5,160 genes were identified, representing a coding capacity of 4,598,637 bp (86.4% of the sequenced genome). Among these genes, 169 (3.2%) were assigned as putative proteins, 1,975 (38.3%) were assigned as hypothetical proteins, and 309 (6%) encode proteins of unknown function. Moreover, 5,124 genes matched at least one sequence in the COG database (13) with BLASTP default parameters.

Comparison with the *Afipia* sp. strain 1NLS2 genome (GenBank accession number [ADVZ00000000](#)) using RAST (2) shows that *A. birgiae* 32632^T exhibits at least 80% common genes involved in the same functioning parts of the metabolism. Among the differences, some genes are related to the metabolism of aromatic compounds, where *A. birgiae* presents many genes for benzoate transport and degradation. In addition, *A. birgiae* has more genes involved in resistance mechanisms, such as those for chromium resistance.

Nucleotide sequence accession numbers. The *A. birgiae* genome, consisting of 65 contigs arranged in 5 scaffold contigs, was deposited in GenBank under accession numbers [CAJQ02000001](#) to [CAJQ02000065](#).

ACKNOWLEDGMENT

We report no funding source.

REFERENCES

1. Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
2. Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi:10.1186/1471-2164-9-75.
3. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol.* 13:R56.
4. Gao S, Sung W, Nagarajan N. 2011. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-end sequences. *J. Comput. Biol.* 18:1681–1691.
5. Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi:10.1186/1471-2105-11-119.
6. Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108.
7. Lamoth F, Greub G. 2010. Amoebal pathogens as emerging causal agents of pneumonia. *FEMS Microbiol. Rev.* 34:260–280.
8. La Scola B, Barrassi L, Raoult D. 2000. Isolation of new fastidious alpha Proteobacteria and *Afipia felis* from hospital water supplies by direct plating and amoebal co-culture procedures. *FEMS Microbiol. Ecol.* 34:129–137.
9. La Scola B, Mallet M, Grimont PAD, Raoult D. 2002. Description of *Afipia birgiae* sp. nov. and *Afipia massiliensis* sp. nov. and recognition of *Afipia felis* genospecies A. *Int. J. Syst. Evol. Microbiol.* 52:1773–1782.
10. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11–16.
11. Margulies M, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
12. Punta M, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.
13. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28:33–36.
14. Tatusov RL, Koonin EV, Lipman DJ. 1977. A genomic perspective on protein families. *Science* 278:631–637.
15. Thomas V, Casson N, Greub G. 2007. New *Afipia* and *Bosea* strains isolated from various water sources by amoebal co-culture. *Syst. Appl. Microbiol.* 30:572–579.

Received 1 October 2012 Accepted 11 October 2012

Address correspondence to Bernard La Scola, bernard.la-scola@univ-amu.fr.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.01918-12