Genomics Data 2 (2014) 123–126

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Genome sequencing and annotation of *Afipia septicemium* strain OHSU_II

Philip Yang [1], Guo-Chiuan Hung [1], Haiyan Lei, Tianwei Li, Bingjie Li, Shien Tsai, Shyh-Ching Lo *

Tissue Microbiology Laboratory, Division of Cellular and Gene Therapies, Office of Cellular, Tissue and Gene Therapies, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD, USA

ARTICLE INFO

Article history:
Received 13 May 2014
Received in revised form 2 June 2014
Accepted 2 June 2014
Available online 10 June 2014

Keywords:
Afipia septicemium
Whole genome
Illumina-MiSeq
CLC bio Genomics Workbench version 6.0
Rapid Annotations using Subsystems Technology (RAST)

ABSTRACT

We report the 5.1 Mb noncontiguous draft genome of *Afipia septicemium* strain OHSU_II, isolated from blood of a female patient. The genome consists of 5,087,893 bp circular chromosome with no identifiable autonomous plasmid with a G + C content of 61.09% and contains 4898 protein-coding genes and 49 RNA genes including 3 rRNA genes and 46 tRNA genes.

Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Specifications	
Organism/cell line/tissue	*Afipia septicemium* OHSU_II isolated from blood of a female patient
Sex	N/A
Sample source location	Oregon Health and Science University, Portland, Oregon
Sequencer or array type	Sequencer; Illumina Miseq
Data format	Processed
Experimental factors	Microbial strain
Experimental features	draft genome sequencing of *A. septicemium*, assembly and annotation
Consent	N/A

Direct link to deposited data

http://www.ncbi.nlm.nih.gov/nuccore/APJI00000000.

Experimental design, materials and methods

The genus *Afipia* was first proposed by Brenner et al. in 1991 [1], containing *Afipia felis* sp. nov., *Afipia clevelandensis* sp. nov., *Afipia broomeae* sp. nov., and three unnamed *Afipia* genospecies. This genus comprises of gram-negative, motile, oxidase-positive bacteria belonging to the alpha-2 subgroup of the class *Proteobacteria*. Subsequently, the isolation of *Afipia birgiae*, *Afipia massiliensis* and *A. felis* genospecies A from hospital water supply using amoeba co-cultivation was described in 2002 [2]. More recently, our lab isolated a novel species *A. septicemium* OHSU_I and OHSU_II from the blood cultures of 2 patients with poorly defined clinical illnesses [3]. The bacteria were originally living in an inactive state in SP4 broth medium designed to culture fastidious mycoplasmas [4]. The bacteria gradually adapted to grow in modified SP4 medium supplemented with 5% fetal bovine serum, vitamin B_{12}, CKM ($CaCl_2$/KNO_3/$MgSO_4$) and NAD/NADP (β-nicotinamide adenine dinucleotide/β-nicotinamide adenine dinucleotide phosphate, 10 mg/L) and could form microscopic colonies on 1.5% Noble agar plates prepared using modified SP4 medium kept in room temperature [3]. The subsequent single-colony derived microbes could grow in YM (yeast malt) and BHI (brain heart infusion) broth without serum supplement.

Whole-genome sequencing was performed on DNA extracted from single-colony clones of *A. septicemium* OHSU_II grown in SP4 broth cultures. Briefly, 50 ng of purified DNA of a bacterial culture derived from a single colony were subjected to DNA library construction using the Nextera DNA Sample Prep Kit (Illumina). A mixture of 4 to 6 separate DNA libraries was usually analyzed for each run involving 2 × 250 bp pair-end sequencing. The sequencing data, sufficient to cover bacterial genomes of ~5 million bps 378-fold, was obtained by Illumina MiSeq platform. The raw reads generated from genomic DNA sequencing were first assembled into 42 contigs using CLC Genomics Workbench (CLC Bio version 6.0) using the de novo assembly after filtering out and trimming low-quality reads.

* Corresponding author. Tel.: +1 301 827 3170; fax: +1 301 827 0449.
E-mail address: shyhching.lo@fda.hhs.gov (S.-C. Lo).
[1] Both are first authors, equally contributed.

http://dx.doi.org/10.1016/j.gdata.2014.06.001
2213-5960/Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

The 16S rRNA gene sequence of *A. septicemium* was extracted from genome data and compared using NCBI BLAST under default settings (e.g. considering only the high-scoring segment pairs (HSP) from the best 250 hits) with the most recent database release in GenBank. The analysis showed that among all the matches for the 16S rRNA gene sequences, the species with the highest bit score of 2700 was uncultured *Afipia* sp. clone C-12 (AY568503), corresponding to an identity of 99.7%, followed by 2 uncultured *Afipia* clones and 2 uncultured bacterial clones, with bit scores ranging from 2676 to 2641. The highest scoring bacteria of currently recognized known species were *Afipia lausannensis* (bit score 2638) and *A. broomeae* (bit score 2630).

Fig. 1 shows the phylogenetic relatedness of *A. septicemium* with other closely related bacteria species based on the 16S rRNA sequences. The phylogenetic relatedness of difficult to culture *A. septicemium* with several other uncultured *Afipia* species using both Neighbor-Joining [5] and Minimal Evolution [6] criteria inferred by MEGA4 program [7] strongly supports the assigning of the newly identified microbes to the genus *Afipia*. It is important to note that the microbes in this genus, or even in this family, are known to have very high levels of homology in their 16S rRNA gene sequences [3,8]. Thus, what presently classified as *Afipia* bacteria based on 16S rRNA sequences are likely to be a heterogeneous group of *Bradyrhizobiaceae*. Genomic content differences between *A. septicemium* and other established *Afipia* species of bacteria including *A. broomeae* are more than 22% [3], which strongly supports that *A. septicemium* is a distinct *Afipia* species.

A working draft genome of the *A. septicemium* OHSU_II-C1 was constructed using the complete genome of *Bradyrhizobiaceae* bacterium strain SG-6C (GenBank accession number CM001195) and the 3 supercontigs (1.1, 1.2 and 1.3 without including supercontig 1.4 of 128 Kb) of *A. broomeae* (GenBank accession numbers NZ_KB375282.1, NZ_KB375283.1 and NZ_KB375284) [9] as reference genomes using CONTIGutor [10]. Sequence of supercontig 1.4 of *A. broomeae* could not be aligned with *Bradyrhizobiaceae* bacterium SG-6C genome sequence and revealed significant homology with several plasmids of alph-2 *Proteobacteria*, and therefore was not used as a reference for draft genome construction. The draft genome of *A. septicemium* OHSU_II-C1 consists of a 5,087,893 bp circular chromosome with no identifiable autonomous plasmid with a G + C content of 61.09% and contains 49 RNA genes including 3 rRNA genes and 46 tRNA genes predicted by RNAmmer 1.2 [11] and tRNAscan-SE-1.23 [12].

The functional annotation was carried out by RAST (Rapid Annotation using Subsystem Technology) hosted by Fellowship for Interpretation

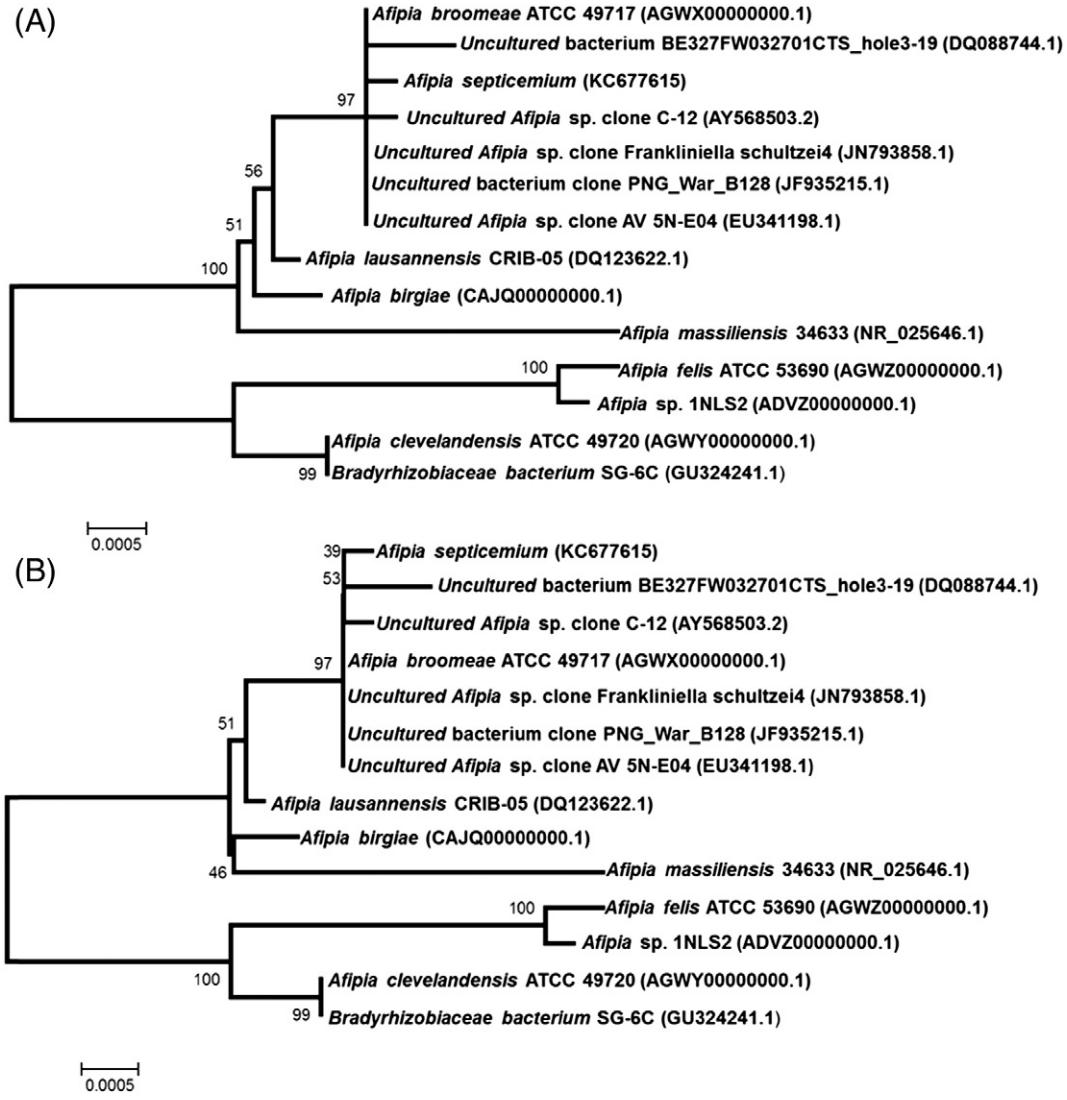

Fig. 1. Phylogenetic tree highlighting the position of *Afipia septicemium* relative to other members of *Afipia* and *Bradyrhizobiaceae* bacterium SG-6C. This tree is inferred from 1487 aligned positions of the 16S rRNA gene sequence under the Neighbor-Joining criteria (A) and Minimum Evolution criteria (B) using MEGA4. GenBank Accession numbers of sequences used in the analyses are shown in parentheses. The branches are scaled in terms of the expected number of substitution per site. Numerals indicate bootstrap percentages over 50 after 500 replications.

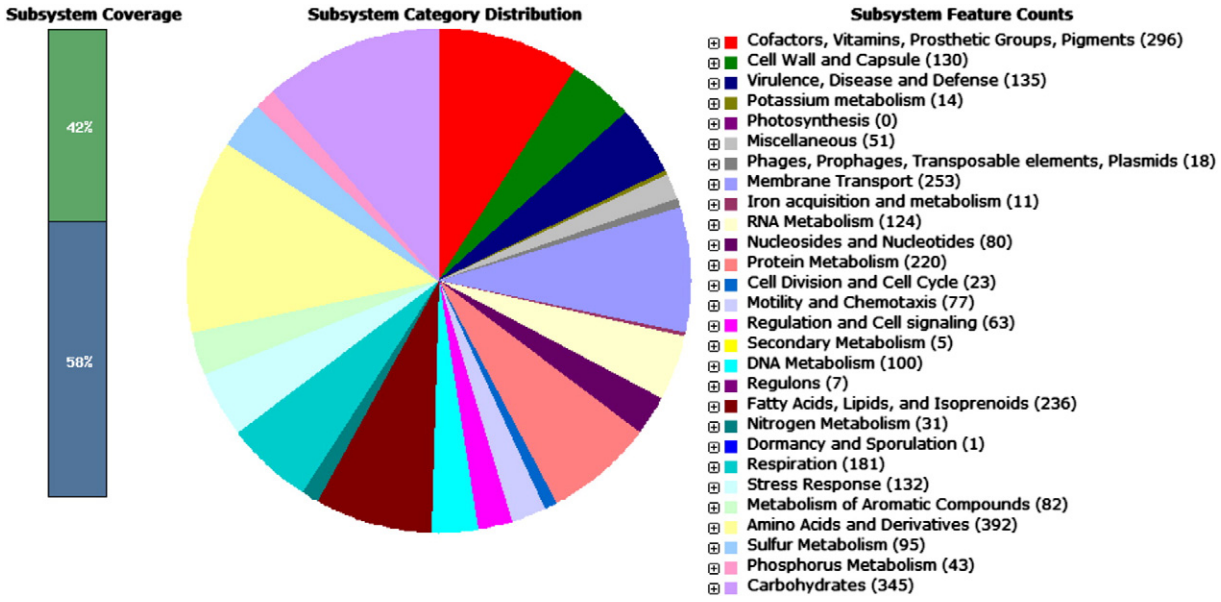

Fig. 2. Subsystem distribution of *Afipia septicemium* (based on RAST annotation server).

of Genomes [13]. The functional comparison of genome sequences available on the RAST server revealed the closest neighbors of *A. septicemium* as *Afipia* sp. 1NLS2 (score 537) followed by *Nitrobacter winogradskyi* Nb-255 (score 517), *A. clevelandensis* ATCC 49720 (score 496) and *Oligotropha carboxidovorans* OM4 (score 486). *A. broomeae* genome was not available on RAST server for functional comparison.

A total of 5061 coding regions (2337 from a positive strand and 2724 from a negative strand) were found in the genome of which 3487 (69%) could be functionally annotated. The genome coding density is 87% with an average gene length of 875 bp. Fig. 2 revealed that the annotated genome has 135 genes potentially involved in virulence, disease and defense including 109 genes for resistance to antibiotics and toxic compounds such as cobalt–zinc–cadmium, fluoroquinolones, arsenic, beta-lactamase and chromium compounds.

Additionally, the annotated genome has 18 genes involved in phages, prophages, transposable elements and plasmids (Fig. 2). Further examination of the 18 genes revealed that there is a cluster of 14 gene transfer agent (GTA) genes from the genomic position 3,629,808 to 3,652,259 (Fig. 3). These GTA genes have a similar organization of the first GTA described in the purple photosynthetic alpha-2 *Proteobacteria Rhodobacter capsulatus* (RcGTA) [14]. These GTA genes include terminase, the phage portal protein, the prohead protease, the major capsid protein, the head–tail adaptor protein, the major tail protein, the tail tape measure protein, and host specificity proteins (Fig. 3).

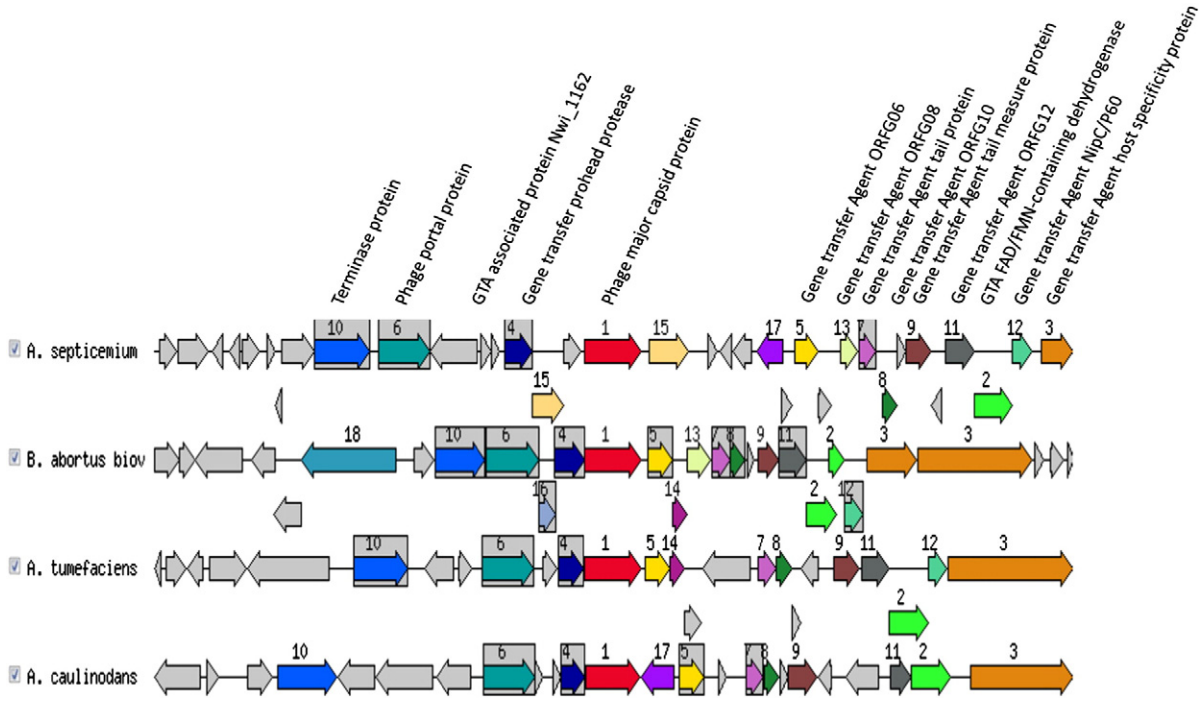

Fig. 3. Comparative analysis of genomic regions containing RcGTA like genes for *Afipia septicemium*, *Brucella abortus*, *Agrobacterium tumefaciens* and *Azorhizobium caulinodans*. Sets of homologous proteins located in the genomic regions are presented in the same color, with the phage major capsid protein in the middle of the region as number 1 while other numbers are randomly assigned by RAST. The arrow for each gene is pointing in the direction of transcription.

This particular region identified in *A. septicemium* genome revealed the highest similarity with that of *Brucella abortus*, a known pathogenic member of *Rhizobiales*, when compared using the RAST database (Fig. 3). Several potentially pathogenic *Afipia* bacteria such as *A. felis*, *Afipia* sp. 1NLS2 and *A. birgiae* are also found to possess similar numbers of RcGTA genes in their genomes. However, only the terminase gene is found in genomes of *A. clevelandensis* and *A. broomeae*.

Alphaproteobacteria are known to transfer their genomes into their eukaryotic hosts and to receive DNA from other organisms by bidirectional lateral sequence transfer through mechanisms such as transformation, transduction, conjugation, membrane vesicles, intracellular nanotube or GTAs [15]. Recently A novel putative GTA (BaGTA) was recently discovered in the mouse-infecting bacterium *Bartonella grahamii*, another species of alphaproteobacteria within the order of *Rhizobiales*. It was suggested that BaGTA contributed to the genomic plasticity of *Bartonella* to allow the microbe to evolve adaptively towards infecting a broader range of different hosts [16]. It is presently unclear if BaGTAs or any novel GTAs other than the RcGTA-like genes are present in *A. septicemium* genome.

Nucleotide sequence accession number

A. septicemium draft genome shotgun (WGS) project has been deposited at DDBJ/EMBL/GenBank under the project accession APJI00000000. It has a bioproject ID of PRJNA191806 and a GOLD ID of Gi51905.

Conflict of interest

The authors declare that there is no conflict of interest on any work published in this paper.

Acknowledgments

The authors would like to thank Drs. Robert Aksamit and Jakob Reiser (CBER/FDA) for critically reading and carefully editing the manuscript. We would also like to thank Marco Galardini for his discussions regarding CONTIGutaor bioinformatics tools. This study was supported in part by the FDA Modernizing Science Grant.

References

[1] D.J. Brenner, D.G. Hollis, C.W. Moss, C.K. English, G.S. Hall, et al., Proposal of *Afipia* gen. nov. with *Afipia felis* sp. nov. (formerly the cat scratch disease bacillus), *Afipia clevelandensis* sp. nov. (formerly the Cleveland Clinic Foundation strain), *Afipia broomeae* sp. nov., and three unnamed genospecies. J. Clin. Microbiol. 29 (1991) 2450–2460.
[2] B. La Scola, M.N. Mallet, P.A. Grimont, D. Raoult, Description of *Afipia birgiae* sp. nov. and *Afipia massiliensis* sp. nov. and recognition of *Afipia felis* genospecies A. Int. J. Syst. Evol. Microbiol. 52 (2002) 1773–1782.
[3] S.C. Lo, G.C. Hung, B. Li, H.Y. Lei, T.W. Li, et al., Isolation of novel *Afipia septicemium* and identification of previously unknown bacteria *Bradyrhizobium* sp. OHSU_III from blood of patients with poorly defined illnesses. PLoS One 8 (2013) e76142.
[4] S.C. Lo, M.M. Hayes, R.Y. Wang, P.F. Pierce, H. Kotani, et al., Newly discovered mycoplasma isolated from patients infected with HIV. Lancet 338 (1991) 1415–1418.
[5] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4 (1987) 406–425.
[6] K. Takahashi, M. Nei, Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol. Biol. Evol. 17 (2000) 1251–1258.
[7] K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24 (2007) 1596–1599.
[8] G.E. Fox, J.D. Wisotzkey, P. Jurtshuk Jr., How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. 42 (1992) 166–170.
[9] *Afipia broomeae* ATCC 49717 RefSeq Genome, Available http://www.ncbi.nlm.nih.gov/bioproject/179398 (Accessed 2014 May 5).
[10] M. Galardini, E.G. Biondi, M. Bazzicalupo, A. Mengoni, CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol. Med. 6 (2011) 11, http://dx.doi.org/10.1186/1751-0473-6-11.
[11] K. Lagesen, P. Hallin, E.A. Rodland, H.H. Staerfeldt, T. Rognes, et al., RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35 (2007) 3100–3108.
[12] T.W. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25 (1997) 955–964.
[13] R.K. Aziz, D. Bartels, A.A. Best, M. Dejongh, T. Disz, et al., The RAST server: rapid annotation using subsystems technology. BMC Genomics 9 (2008) 75.
[14] B.L. Marrs, Genetic recombination in *Rhodopseudomonas capsulata*. Proc. Natl. Acad. Sci. U. S. A. 71 (1974) 971–973.
[15] P.T. Le, P. Pontarotti, D. Raoult, Alphaproteobacteria species as a source and target of lateral sequence transfers. Trends Microbiol. 22 (2014) 147–156.
[16] L. Guy, B. Nystedt, C. Toft, K. Zaremba-Niedzwiedzka, E.C. Berglund, et al., A gene transfer agent and a dynamic repertoire of secretion systems hold the keys to the explosive radiation of the emerging pathogen *Bartonella*. PLoS Genet. 9 (2013) e1003393.