Title

Draft Genome of Burkholderia cenocepacia TAtl-371, a Strain from the Burkholderia cepacia Complex Retains Antagonism in Different Carbon and Nitrogen Sources

Journal

Current Microbiology, 76(5)

ISSN

0343-8651

Authors

Rojas-Rojas, Fernando Uriel
Sánchez-López, David
Tapia-García, Erika Yanet
et al.

Publication Date

2019-05-01

DOI

10.1007/s00284-019-01657-0

Peer reviewed

Short Genome Reports

Improved, high-quality permanent draft genome of *Burkholderia cenocepacia* TAtl-371, a strain from the *Burkholderia cepacia* complex with a broad antagonistic spectrum

Authors

Fernando Uriel Rojas-Rojas[1], David López-Sánchez[1], Erika Yanet Tapia García[1], Maskit Maymon[2], Ethan Humm[2], Marcel Huntemann[3], Alicia Clum[3], Manoj Pillay[3], Krishnaveni Palaniappan[3], Neha Varghese[3], Natalia Mikhailova[3], Dimitrios Stamatis[3], T.B.K. Reddy[3], Victor Markowitz[3], Natalia Ivanova[3], Nikos Kyrpides[3], Tanja Woyke[3], Nicole Shapiro[3], Ann M. Hirsch[2,4], J. Antonio Ibarra[1] and Paulina Estrada-de los Santos[1]

Institutional Affiliation

[1] Instituto Politécnico Nacional, Escuela Nacional de Ciencias Biológicas, Prol. Carpio y Plan de Ayala s/n. Col. Santo Tomás, Del. Miguel Hidalgo. C.P. 11340. México, Cd. de México.

[2] Dept. of Molecular, Cell and Developmental Biology and [4]Molecular Biology Institute, University of California-Los Angeles, Los Angeles, California, United States of America 90095.

[3] DOE Joint Genome Institute 2800 Mitchell Drive Walnut Creek CA 94598.

Corresponding author

pestradadelossantos@gmail.com[1], Instituto Politécnico Nacional, Escuela Nacional de Ciencias Biológicas. Prol. Carpio y Plan de Ayala s/n. Col. Santo Tomás, Del. Miguel Hidalgo. C.P. 11340. México, Cd. de México.

Abstract

Burkholderia cenocepacia TAtl-371 was isolated from the rhizosphere of a tomato plant growing in Atlatlahuacan, Morelos, Mexico. This strain exhibited a broad antimicrobial spectrum against bacteria, yeast, and fungi. Here we report and describe the improved, high-quality permanent draft genome of *B. cenocepacia* TAtl-371, which was sequenced using a combination of PacBio RS and PacBio RS II sequencing methods. The 7,496,106 bp genome of the TAtl-371 strain is arranged in 3 scaffolds, contains 6722 protein coding genes, and 99 RNA only-encoding genes. Genome analysis revealed genes related to biosynthesis of antimicrobials such as non-ribosomal peptides, siderophores, chitinases, and bacteriocins.

Keywords

Betaproteobacteria –- *Burkholderia cepacia* complex –- *Burkholderia cenocepacia* – rhizosphere -antimicrobial - antagonistic.

Abbreviations

DOE, Department of Energy. JGI, Joint Genome Institute. ENA, European Nucleotide Archive. NRPS, Non-ribosomal peptide synthase. LlpA, lectin-like protein.

Introduction

The *Burkholderia cepacia* complex (Bcc) is a group of 22 closely related species [1] that share high similarity (>97.5%) in the 16S rRNA region of the *rrs* gene and low DNA-DNA hybridization values (30-60%) [2]. Bcc strains are versatile organisms, which are known as opportunistic pathogens in patients with cystic fibrosis (CF), but are also widely distributed in different environments including the rhizospheres of different crops [3]. Bcc members produce multiple antimicrobial compounds, which were recently reviewed by Depoorter *et al.* in 2016 [4].

B. cenocepacia TAtl-371 was isolated from the rhizosphere of tomato growing in Atlatlahuacan, State of Mexico, and originally identified as *Burkholderia* (*Paraburkholderia*) *unamae* [5]. However, resequencing of the 16S region from the

rrs gene, as well as the *recA* gene showed that the strain belonged to the Bcc (Fig. 1). This species exhibits a broad-antimicrobial spectrum, which targets soil and nodulating bacteria, human and plant pathogenic bacteria, and also human opportunistic bacteria, yeast, and phytopathogenic fungi (unpublished results). The aim of sequencing the *B. cenocepacia* TAtl-371 genome was to identify genes related to the biosynthesis of antimicrobial compounds in order to achieve a deeper understanding of the broad antimicrobial spectrum of this strain. The *B. cenocepacia* TAtl-371 strain could be considered a source of antimicrobials for different biotechnological applications, including the biocontrol of phytopathogens, although more studies are needed because of its similarity to CF strains. The genome sequence was obtained in cooperation with Joint Genome Institute (JGI) of the US Department of Energy (DOE).

Organism Information

Classification and features

B. cenocepacia strain TAtl-371 is a motile, Gram-negative, strictly aerobic, nonsporulating rod in the order *Burkholderiales* of the class *Betaproteobacteria* (Table 1). The rod-shaped form varies in size with dimensions of 0.5-0.7 μm in width and 1-2 μm in length (Fig. 2). Cells grow fast in potato dextrose agar (PDA), Luria-Bertani agar (LB) and Trypticase soy agar (TSA) at 30ºC, and grow slowly in mineral media such as BSE, BAc and Az [6]. This bacterium can grow using phenol as a unique carbon source. Colonies on LB and TSA are uniform 1-3 mm diameter, circular, convex, with entire smooth edges, shiny, non-pigmented (cream) and opaque. On PDA, colonies are irregular, with undulate margins, smooth, shiny, non-pigmented (white), opaque and moderately mucoid. The optimal growth temperature is 30ºC; however, it can grow at 37º and 42ºC. The general features of this strain are shown in Table 1. The bacterium has natural resistance to tetracycline (100 μg/ml), chloramphenicol (60 μg/ml), kanamycin (50 μg/ml) and gentamicin (60 μg/ml). The strain produces siderophores, but phosphate solubilization is minimal and it lacks the ability to fix nitrogen. *B. cenocepacia* TAtl-371 is closely related to *B. cenocepacia* MC0-3, HI2424, and AU1054 (Fig 1).

Genome sequencing information

Genome project history

B. cenocepacia TAtl-371 was sequenced at the Department of Energy Joint Genome Institute (DOE-JGI) as a part of the project “Root nodule microbial communities of legume samples collected from USA, Mexico and Botswana” directed by Dr. Ann M. Hirsch. The goal of this project was to identify the microbial community housed within nodules of native legumes living in three arid or semi-arid, nutrient-poor environments in Mexico, Botswana, and the United States.

The complete sequence was finished in May 2015 and some features are presented in Table 2.

Growth conditions and genomic DNA preparation

B. cenocepacia TAtl-371 cells were grown in 5 ml of LB minus NaCl at 30ºC for 18 h at 120 rpm. The DNA extraction was done using the Invitrogen Purelink™ Genomic DNA Mini Kit. The purified DNA was monitored for integrity by gel electrophoresis, and then sent to the JGI for sequencing.

Genome sequencing and assembly

The draft genome of *B. cenocepacia* TAtl-371 was generated at the DOE Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology [7]. A Pacbio SMRTbellTM library was constructed and sequenced on the PacBio RS platform, which generated 213,715 filtered sub-reads totaling 943.3 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. The raw reads were assembled using HGAP (version: 2.3.0 p5,protocol version=2.3.0 method=RS HGAP Assembly.3,smrtpipe.py v1.87.139483,) [8]. The final draft assembly contained 3 contigs in 3 scaffolds, totalling 7.496 Mbp in size. The input read coverage was 91.6X.

Genome annotation

Genes were identified using Prodigal [9], followed by a round of manual curation using GenePRIMP [10] for finished genomes and draft genomes in fewer than 10 scaffolds. The predicted CDSs were translated and used to search the National

Center for Biotechnology Information (NCBI) no redundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool [11] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [12]. Other non–coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL [13]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG) platform [14] developed by the Joint Genome Institute, Walnut Creek, CA, USA [15].

Genome properties

The *B. cenocepacia* TAtl-371 genome comprises three scaffolds with a total size of 7,496,106 bp and a GC content of 67.01% (Table 3, Figure 3). A total of 6821 genes were predicted of which 6722 (98.55%) are protein-coding genes, 62 (0.91%) are pseudogenes, 99 (1.45%) RNA genes (18 rRNA genes and 67 tRNA genes). The majority of the protein-coding genes (5554) (81.43%) were assigned to a putative function and 1168 (17.12%) were annotated as hypothetical proteins.

Insights from the genome sequence

B. cenocepacia TAtl-371 is able to kill soil and nodulating bacteria, human and plant pathogenic bacteria, human opportunistic bacteria and yeast, and phytopathogenic fungi. Genome analysis showed several genes involved in the biosynthesis of numerous antimicrobial compounds (Table 5), including the siderophores ornibactin and pyochelin, which have been identified in other *B. cenocepacia* genomes as the most common siderophores produced by this species [16]. The biological activity described for pyochelin is for iron uptake and antibacterial activity [17].

Bacteriocins are ribosomally synthesized proteinaceous compounds which inhibit growth of bacteria that are closely related to the producer strain [18]. The *B. cenocepacia* TAtl-371 genome contains two genes encoding a lectin-like bacteriocin, namely LlpA. These tandemly organized genes are homologues to *llpA*-like genes found in other *B. cenocepacia* strains [19]. The antimicrobial activity of *B. cenocepacia* LlpAs has been reported for *B. ambifaria, B. anthina, B. cenocepacia*,

and *B. metallica* [19]. *B. cenocepacia* TAtl-371's antibacterial activity against other Bcc species (*B. multivorans, B. dolosa, B. stabilis* and *B. pyrrocinia*) was determined by our research group (data not shown). We are continuing to study whether other LlpA sequences in the TAtl-371 genome encode gene products with antibacterial activity.

Chitinases are glycosyl hydrolases that catalyze the hydrolytic degradation of chitin which is one of the major constituent of cell walls of a variety of fungi [20]. The TAtl-371 genome contains a gene that encodes a chitinase belonging to the family 18 of the glycosyl hydrolases as well as a gene that encodes a predicted chitinase. Using BLAST tools [21], we identified widely conserved homologues of these genes in several *B. cenocepacia* strains, but to the best of our knowledge, no studies about antifungal activity brought about by chitinases in this species have been published. The chitinolytic activity of *B. cenocepacia* TAtl-371 and the role of these enzymes in antifungal activity are currently under study in our lab.

The *B. cenocepacia* TAtl-371 genome also contains genes that encode for possible resistance to the metalloids arsenate and arsenite and several metal ions, including tellurium, silver, cobalt, zinc, cadmium, copper, and chromate. Genes potentially involved in copper and chromate resistance are common in the genome. Finally, gene clusters for the biosynthesis cellulose and vanillin were identified; both molecules have potential as TAtl-371 metabolites with biotechnological applications.

An Average Nucleotide Identity (ANI) calculation was used (Table 6) to compare the genome of *B. cenocepacia* TAtl-371 with the other sequenced Bcc genomes. The accepted ANI cut-off for species is 95-96% [22]. An ANI value >95.0% confirmed that strain TAtl-371 belongs to *B. cenocepacia.* Our phylogenetic analysis using the sequence of *recA* gene as well as the ANI values confirms that *B. cenocepacia* TAtl-371 belongs to PHDC lineage of the *B. cenocepacia* species [23], which is formed by strains isolated not only isolated from CF patients but also from natural environments around the world [24-26].

Conclusions

B. cenocepacia TAtl-371 was isolated from the tomato rhizosphere in Mexico and characterized as an antagonistic bacterium with a broad antimicrobial spectrum. The genome analysis of this strain allowed the identification of gene clusters encoding putative antimicrobial compounds such as siderophores, bacteriocins, and various enzymes that could mediate the antagonistic activity of TAtl-371. Our research group is currently studying the role of these compounds in the broad-spectrum antagonistic activity of this strain. The combination of genes involved in antagonism against other microbes, as well as the presence of genes involved in cellulose and vanillin biosynthesis, suggests that *B. cenocepacia* could be a source of compounds with yet to be explored biotechnological applications. Nevertheless, its close relationship with bacteria isolated CF patients requires additional research before such applications can be accepted.

Competing interests

The authors declare that they have no competing interests.

Funding

The genome sequence was conducted by the U.S. Department of Energy, Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Sciences of the U.S. Department of Energy under the proposal 1572 and contract No. DE-AC02-05CH11231.

Authors' contributions

FURR performed some phenotypic characterization of the strain, prepared bacteria for shipping to UCLA Laboratory, analyzed data and drafted the manuscript. DLS and EYTG performed some phenotypic characterization of the strain. MM and EH prepared the DNA for sequencing. MH, AC, MP, KP, NV, NM, DS, TBKR, VM, NI, NK, TW and NS performed the technical work for sequencing, assembly, and annotation of the genome. AMH led the manual annotation group at UCLA and wrote and reviewed the final manuscript. JAI and PES analyzed data and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

FURR and EYTP are recipients of a fellowship from CONACyT. JAIG and PES are recipient of SNI, EDI and COFAA fellowships. We thank Dr. E.O. Lopez-Villegas (Escuela Nacional de Ciencias Biológicas, IPN) for the transmission electron microscopic analysis.

References

1. Weber CF, King GM: **Volcanic Soils as Sources of Novel CO-Oxidizing Paraburkholderia and Burkholderia: Paraburkholderia hiiakae sp. nov., Paraburkholderia metrosideri sp. nov., Paraburkholderia paradisi sp. nov., Paraburkholderia peleae sp. nov., and Burkholderia alpina sp. nov. a Member of the Burkholderia cepacia Complex**. *Front Microbiol* 2017, **8**:207.
2. Coenye T, Vandamme P, Govan JRW, LiPuma JJ: **Taxonomy and identification of the Burkholderia cepacia complex**. *Journal of Clinical Microbiology* 2001, **39**(10):3427-3436.
3. Compant S, Nowak J, Coenye T, Clement C, Ait Barka E: **Diversity and occurrence of Burkholderia spp. in the natural environment**. *FEMS Microbiol Rev* 2008, **32**(4):607-626.
4. Depoorter E, Bull MJ, Peeters C, Coenye T, Vandamme P, Mahenthiralingam E: **Burkholderia: an update on taxonomy and biotechnological potential as antibiotic producers**. *Appl Microbiol Biotechnol* 2016, **100**(12):5215-5229.
5. Caballero-Mellado J, Onofre-Lemus J, Estrada-De Los Santos P, Martínez-Aguilar L: **The tomato rhizosphere, an environment rich in nitrogen-fixing Burkholderia species with capabilities of interest for agriculture and bioremediation**. *Applied and environmental microbiology* 2007, **73**(16):5308-5319.
6. Estrada-De Los Santos P, Bustillos-Cristales Ro, Caballero-Mellado J: **Burkholderia, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution**. *Applied and environmental microbiology* 2001, **67**(6):2790-2798.
7. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al*: **Real-time DNA sequencing from single polymerase molecules**. *Science* 2009, **323**(5910):133-138.

23. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE *et al*: **Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data**. *Nat Methods* 2013, **10**(6):563-569.
24. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification**. *BMC Bioinformatics* 2010, **11**:119.
25. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC: **GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes**. *Nat Methods* 2010, **7**(6):455-457.
26. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence**. *Nucleic Acids Res* 1997, **25**(5):955-964.
27. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO: **SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB**. *Nucleic Acids Res* 2007, **35**(21):7188-7196.
28. **INFERNAL.** http://infernal.janelia.org/. In.
29. **IMG. The Integrated Microbial Genomes Plataform** [http://img.jgi.doe.gov/]
30. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC: **IMG ER: a system for microbial genome annotation expert review and curation**. *Bioinformatics* 2009, **25**(17):2271-2278.
31. Esmaeel Q, Pupin M, Kieu NP, Chataigne G, Bechet M, Deravel J, Krier F, Hofte M, Jacques P, Leclere V: **Burkholderia genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis**. *Microbiologyopen* 2016, **5**(3):512-526.
32. Esmaeel Q, Pupin M, Jacques P, Leclère V: **Nonribosomal peptides and polyketides of Burkholderia: new compounds potentially implicated in biocontrol and pharmaceuticals**. *Environmental science and pollution research international* 2017.
33. Snyder AB, Worobo RW: **Chemical and genetic characterization of bacteriocins: antimicrobial peptides for food safety**. *Journal of the Science of Food and Agriculture* 2014, **94**(1):28-44.
34. Ghequire MG, De Canck E, Wattiau P, Van Winge I, Loris R, Coenye T, De Mot R: **Antibacterial activity of a lectin-like Burkholderia cenocepacia protein**. *Microbiologyopen* 2013, **2**(4):566-575.
35. Bhattacharya D, Nagpure A, Gupta RK: **Bacterial chitinases: properties and potential**. *Critical reviews in biotechnology* 2007, **27**(1):21-28.
36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.
37. Richter M, Rosselló-Móra R: **Shifting the genomic gold standard for the prokaryotic species definition**. *Proceedings of the National Academy of Sciences* 2009, **106**(45):19126-19131.
38. LiPuma JJ, Spilker T, Coenye T, Gonzalez CF: **An epidemic Burkholderia cepacia complex strain identified in soil**. *The Lancet* 2002, **359**(9322).
39. Coenye T, Spilker T, Van Schoor A, LiPuma J, Vandamme P: **Recovery of Burkholderia cenocepacia strain PHDC from cystic fibrosis patients in Europe**. *Thorax* 2004, **59**(11):952-954.

28425. Baldwin A, Mahenthiralingam E, Drevinek P, Vandamme P, Govan JR, Waine DJ, LiPuma JJ, Chiarini L, Dalmastri C, Henry DA: **Environmental Burkholderia cepacia complex isolates from human infections**. *Emerging infectious diseases* 2007, **13**(3):458.

28826. Coenye T, LiPuma JJ: **Population structure analysis of Burkholderia cepacia genomovar III: varying degrees of genetic recombination characterize major clonal complexes**. *Microbiology* 2003, **149**(1):77-88.

29127. 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**(5):541-547.

29528. Woese CR, Kandler O, Wheelis ML: **Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya**. *Proc Natl Acad Sci U S A* 1990, **87**(12):4576-4579.

29829. Garrity GM, Bell JA, Lilburn TE: **Phylum L. XIV. Proteobacteria phyl. nov.** In: *Bergey's manual of systematic bacteriology.* Edited by Garrity GM, Brenner DJ, Krieg NR, Stanley JT, vol. 2, 2nd edn. New York: Springer; 2005.

30130. Garrity GM, Bell JA, Lilburn TE: **Class II. Betaproteobacteria class. nov**. In: *Bergey's Manual of systematic bacteriology.* Edited by Garrity GM, Brenner DJ, Krieg NR, Staley JT, vol. 2, 2nd edn. New York: Springer; 2005.

30431. Garrity GM, Bell JA, Lisburn TE: **Order I. Burkholderiales**. In: *Bergey's manual of systematic bacteriology.* Edited by Garrity GM, Brenner DJ, Krieg NR, Stanley JT, vol. 2, 2nd edn. New York: Springer; 2005.

30732. Garrity GM, Bell JA, Lisburn TE: **Family Burkholderiaceae**. In: *Bergey's manual of systematic bacteriology.* Edited by Garrity GM, Brenner DJ, Krieg NR, Staley JT, vol. 2, 2nd edn. New York: Springer; 2005

31033. Yabuuchi E, Kosako Y, Oyaizu H, Yano I, Hotta H, Hashimoto Y, Ezaki T, Arakawa M: **Proposal of Burkholderia gen. nov. and Transfer of Seven Species of the Genus Pseudomonas Homology Group II to the New Genus, with the Type Species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov**. *Microbiology and Immunology* 1992, **36**(12):1251-1275.

31634. Vandamme P, Holmes B, Coenye T, Goris J, Mahenthiralingam E, LiPuma JJ, Govan JR: **Burkholderia cenocepacia sp. nov.--a new twist to an old story**. *Res Microbiol* 2003, **154**(2):91-96.

31935. 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**(1):25-29.

32336. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood**. *Syst Biol* 2003, **52**(5):696-704.

32637. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular Evolutionary Genetics Analysis version 6.0**. *Mol Biol Evol* 2013, **30**(12):2725-2729.

329

330

Table 1. Classification and general features of *Burkholderia cenocepacia* strain TAtl-371 [27]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain Bacteria	TAS [28]
		Phylum Proteobacteria	TAS [29]
		Class Betaproteobacteria	TAS [30]
		Order Burkholderiales	TAS [31]
		Family Burkholderiaceae	TAS [32]
		Genus *Burkholderia*	TAS [33]
		Species *Burkholderia cenocepacia*	TAS [34]
		Strain: TAtl-371	
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	NAS
	Sporulation	Non-sporuling	NAS
	Temperature range	30-42°C	IDA
	Optimum temperature	30°C	IDA
	pH range; Optimum	6-8; 6	IDA
	Carbon source	Saccharose, succinic acid, malic acid, glucose, mannitol, phenol, maltose	IDA
MIGS-6	Habitat	Rhizospheric soil	TAS [5]
MIGS-6.3	Salinity	Until 10% NaCl	IDA

MIGS-22	Oxygen requirement	Aerobic	IDA
MIGS-15	Biotic relationship	Free-living	TAS [5]
MIGS-14	Pathogenicity	Pathogenic	NAS
MIGS-4	Geographic location	Morelos, Mexico	TAS [5]
MIGS-5	Sample collection	Date	TAS [5]
MIGS-4.1	Latitude	N 18º 56’ 8.724’’	IDA
MIGS-4.2	Longitude	O 98º 55’ 2.588’’	IDA
MIGS-4.4	Altitude	1656 M	IDA

[a] Evidence 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 [35]

Table 2. Project information.

MIGS ID	Property	Term
MIGS 31	Finishing quality	Level 3. Improved-High-Quality Draft
MIGS-28	Libraries used	PAcbio SMRTbell™
MIGS 29	Sequencing platforms	PacBio RS, PacBio RS II
MIGS 31.2	Fold coverage	91.6X
MIGS 30	Assemblers	HGAP version 2.3.0_p5
MIGS 32	Gene calling method	Prodigal
	Locus Tag	BLS50
	ENA accession number	SAMN05443026
	ENA Date of Release	October 13th 2016
	GOLD ID	Ga0073272
	BIOPROJECT	PRNJA332745
MIGS 13	Source Material Identifier	
	Project relevance	Environmental

Table 3. Genome statistics.

Attribute	Value	% of Total
	749610	100.0
Genome size (bp)	6	0
	659979	
DNA coding (bp)	6	88.04
	502308	
DNA G+C (bp)	5	67.01
		100.0
DNA scaffolds	3	0
		100.0
Total genes	6821	0
Protein coding genes	6722	98.55
RNA genes	99	1.45
Pseudo genes	62	0.91
Genes in internal clusters	2741	40.18
Genes with function prediction	5554	81.43
Genes assigned to COGs	5045	73.96
Genes with Pfam domains	5852	85.79
Genes with signal peptides	765	11.22
Genes with transmembrane helices	1599	23.44
CRISPR repeats	0	

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

Code	Value	%age	Description
J	247	4.27	Translation, ribosomal structure and biogenesis
A	1	0.02	RNA processing and modification
K	654	11.32	Transcription
L	115	1.99	Replication, recombination and repair
B	4	0.07	Chromatin structure and dynamics
D	35	0.61	Cell cycle control, Cell division, chromosome partitioning
V	126	2.18	Defense mechanisms
T	271	4.69	Signal transduction mechanisms
M	389	6.73	Cell wall/membrane biogenesis
N	118	2.04	Cell motility
U	123	2.13	Intracellular trafficking and secretion
O	192	3.32	Posttranslational modification, protein turnover, chaperones
C	358	6.2	Energy production and conversion
G	405	7.01	Carbohydrate transport and metabolism
E	612	10.59	Amino acid transport and metabolism
F	117	2.02	Nucleotide transport and metabolism
H	294	5.09	Coenzyme transport and metabolism
I	295	5.11	Lipid transport and metabolism
P	302	5.23	Inorganic ion transport and metabolism
Q	195	3.37	Secondary metabolites biosynthesis, transport and catabolism
R	567	9.81	General function prediction only
S	259	4.48	Function unknown
-	1776	26.04	Not in COGs

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

Table 5. *Burkholderia cenocepacia* TAtl-371 genes related to biotechnological applications

Gene symbol	Locus Tag	Size (bp)	Gene product name
Antimicrobial compounds-related genes			
orbL[1]	Ga0073272_0013	1017	Protein N-acetyltransferase, RimJ/RimL family
orbF[1]	Ga0073272_0014	840	Formyl transferase
orbA[1]	Ga0073272_0015	2265	Iron complex outermembrane recepter protein
pvdA[1]	Ga0073272_0016	1377	L-ornithine N5-oxygenase
orbK[1]	Ga0073272_0017	1032	Acetyltransferase (GNAT) domain-containing protein
orbJ[1]	Ga0073272_0018	4986	Amino acid adenylation domain-containing protein
orbI[1]	Ga0073272_0019	9660	Non-ribosomal peptide synthase domain TIGR01720/amino acid adenylation domain-containing protein
orbE[1]	Ga0073272_0020	1746	

orbB[1]	Ga0073272_0021	1029	Putative ATP-binding cassette transporter
orbF[1]	Ga0073272_0022	801	Iron complex transport system substrate-binding protein
orbD[1]	Ga0073272_0023	2091	Ferric iron reductase protein FhuF
orbC[1]	Ga0073272_0024	849	Iron complex transport system permease protein
orbG[1]	Ga0073272_0025	1008	Iron complex transport system ATP-binding protein
orbH[1]	Ga0073272_0026	243	Taurine dioxygenase, alpha-ketoglutarate-dependent
orbS[1]	Ga0073272_0027	669	MbtH protein
-----	Ga0073272_0087	813	RNA polymerase, sigma-24 subunit, RpoE
-----	Ga0073272_0088	897	D-mannose binding lectin
BCAL1722	Ga0073272_3288	1356	D-mannose binding
pchA[2]	Ga_0073272_5542	1434	
pchB[2]	Ga0073272_5543	306	
pchC[2]	Ga0073272_5544	774	

			lectin
pchD[2]	Ga0073272_5545	1653	Chtinase family 18
pchR[2]	Ga0073272_5546	942	Isochorismate synthase
pchE[2]	Ga0073272_5547	4374	Isochorismate pyruvate lyase
------[2]	Ga0073272_5548	5478	Pyochelin biosynthetic protein PchC
pchG[2]	Ga0073272_5549	1053	
------	Ga0073272_5789	2442	Pyochelin biosynthesis protein PchD
Metal-ion resistance			Transcriptional regulator, AraC family
BCAL1231	Ga0073272_0413	756	
	Ga0073272_0796	612	Dihydroaeruginoic acid synthetase
--------	Ga0073272_0797	543	Pyochelin synthetase
--------	Ga0073272_1837	528	Pyochelin biosynthetic protein PchG
BCAL0581	Ga0073272_2310	1212	
BCAL2754	Ga0073272_2473	555	Predicted chitinase
BCAL2607	Ga0073272_2517	1206	
BCAL2450	Ga0073272_3568	348	Membrane protein

-------			TerC, possibly involved in
BCAM0713	Ga0073272_4355	3213	tellurium resistance
czcB	Ga0073272_4357	1470	Chromate transporter
czcC	Ga0073272_4358	1314	Chromate transporter
----	Ga0073272_4621	942	Chromate transporter
----	Ga0073272_4623	345	Heavy-metal resistance
BPSS0456	Ga0073272_4624	1296	Chromate transporter
			Cu and Ag efflux protein CusF
BCAM0436	Ga0073272_4635	369	Cobalt-zinc-cadmium resistance protein
cusA	Ga0073272_4636	3207	CzcA
BCAM0434	Ga0073272_4637	1536	Membrane fusion protein, cobalt-zinc-cadmium efflux system
			outer membrane

	Ga0073272_4928	357	protein, cobalt-
BCAM0002			zinc-cadmium
	Ga0073272_4931	555	efflux system
BCAM2840			
	Ga0073272_5287	210	Putative copper
BCAM2487			resistance protein D
	Ga0073272_5288	723	Cu and Ag efflux
BCAM2486			protein CusF
	Ga0073272_6324	1254	Multicopper oxidase with three cupredoxin
------			domains (includes cell division protein FtsP and spore coat
Biosynthesis of molecules with biotechnological importance	Ga0073272_0280	198	protein CotA) Cu and Ag efflux protein CusF
BCAL1397[3]	Ga0073272_0281	1557	Cu(I)/Ag(I) efflux system membrane
BCAL1396[3]			protein CusA/SilA
	Ga0073272_0282	2538	
bcsA[3]			Membrane fusion protein, Cu(I)/Ag(I)
	Ga0073272_0283	780	efflux system
BCAL1394[3]			Arsenate reductase
	Ga0073272_0284	246	
--------[3]	Ga0073272_0285	1971	chromate reductase
BCAL1392[3]			Probable cobalt
	Ga0073272_0286	3885	transporter subunit

			(CbtB)
BCAL1391[3]			
			Probable cobalt
	Ga0073272_0287	1230	transporter subunit
			(CbtA)
bcsZ[3]	Ga0073272_0288	2346	
			Arsenite efflux
bcsB[3]			membrane protein
	Ga0073272_4981	831	ArsB
BCAM2793[4]			
	Ga0073272_4982	1449	
BCAM2792[4]			
			Hypothetical protein
			Cellulose synthase operon protein YhjU
			Cellulose synthase (UDP-forming)
			Cellulose synthase operon protein YhjQ
			Hypothetical protein
			Cellulose biosynthesis protein BcsE
			Tetratricopeptide

			repeat-containing protein
			Endoglucanase
			Cellulose synthase subunit
			Vanillin synthase /trans-feruloyl-CoA hydratase
			Vanillin dehydrogenase

[a], NRPS gene cluster for siderophore ornibactin biosynthesis
[b], gene cluster for siderophore pyochelin biosynthesis
[c], genes related to cellulose biosynthesis
[d], genes related to vanillin biosynthesis

Table 6. Average nucleotide identity (ANI) of strain TAtl-371 with other Bcc genomes.

Strain	Average Nucleotide Identity (%)
B. cenocepacia **C14**	98.48
B. cenocepacia **D2AES**	98.42
B. cenocepacia **AU1054**	98.38
B. cenocepacia **HI2424**	98.38
B. cenocepacia **MC0-3**	98.04
B. cenocepacia **J2315^{T}**	95.69
B. pyrrocinia **DSM 10685**	92.73
B. contaminans **LMG 23361^{T}**	92.37
B. vietnamiensis **AU5i**	92.37
B. cepacia **ATCC 25416**	92.36
B. lata **LK13**	92.35
B. ambifaria **IOP40-10**	91.33
B. dolosa **AU0158**	90.30
B. multivorans **ATCC 17616**	89.50
B. ubonensis **Bu^{T}**	89.16

Figure legends

Figure 1.

Phylogenetic tree highlighting the position of *Burkholderia cenocepacia* TAtl-371 in relation to other *Burkholderia cepacia* complex strains. *B. mallei* ATCC 23344^{T} was used as outgroup. The tree is based on *recA* gene alignments and was performed by the maximum-likelihood (ML) method with the General Time Reversible model using PhyML [36]. Bootstrap values inferred from 100 replicates are shown at nodes. The tree was visualized with program MEGA 6 [37]. All strains used have a sequenced genome. The GenBank accession number of each sequence is in brackets.

Figure 2.

Transmission Electron Microscopy of negatively stained *Burkholderia cenocepacia* TAtl-371 cells. The strain was grown on LB medium, cells were cultured by centrifugation at 3000 rpm, were washed with PBS, exposed to glutaraldehyde for one hour and finally were washed and resuspended in PBS. A drop of the suspension was placed on a formvar-coated copper grid and air-dried for 20 min to allow the cells to adhere. The grid was then covered for 20 s with a solution of 0.5% uranyl acetate, the excess liquid was removed with a filter paper, and then air-dried. A JEOL JEM-1010 transmission electron microscope, operated at 60 kV, was used to observe and photograph negatively stained preparations. F, stands for flagella.

Figure 3.

Graphical map of the three scaffolds of the genome of *B. cenocepacia* TAtl-371. From the bottom to the top of each scaffold: Genes on forward strand (color by COG

categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew