UCLA

UCLA Previously Published Works

Title

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

Permalink

<https://escholarship.org/uc/item/00t1012p>

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

1 Short Genome Reports

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

5 Authors

6 Fernando Uriel Rojas-Rojas¹, David López-Sánchez¹, Erika Yanet Tapia García¹,
7 Maskit Maymon², Ethan Humm², Marcel Huntemann³, Alicia Clum³, Manoj Pillay³,
8 Krishnaveni Palaniappan³, Neha Varghese³, Natalia Mikhailova³, Dimitrios Stamatis³,
9 T.B.K. Reddy³, Victor Markowitz³, Natalia Ivanova³, Nikos Kyrpides³, Tanja Woyke³,
10 Nicole Shapiro³, Ann M. Hirsch^{2,4}, J. Antonio Ibarra¹ and Paulina Estrada-de los
11 Santos¹

12 Institutional Affiliation

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

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

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

20 Corresponding author

21 pestradadelossantos@gmail.com¹, Instituto Politécnico Nacional, Escuela Nacional
22 de Ciencias Biológicas. Prol. Carpio y Plan de Ayala s/n. Col. Santo Tomás, Del.
23 Miguel Hidalgo. C.P. 11340. México, Cd. de México.

24Abstract

25*Burkholderia cenocepacia* TAtl-371 was isolated from the rhizosphere of a tomato
26plant growing in Atlatlahuacan, Morelos, Mexico. This strain exhibited a broad
27antimicrobial spectrum against bacteria, yeast, and fungi. Here we report and
28describe the improved, high-quality permanent draft genome of *B. cenocepacia*
29TAtl-371, which was sequenced using a combination of PacBio RS and PacBio RS II
30sequencing methods. The 7,496,106 bp genome of the TAtl-371 strain is arranged
31in 3 scaffolds, contains 6722 protein coding genes, and 99 RNA only-encoding
32genes. Genome analysis revealed genes related to biosynthesis of antimicrobials
33such as non-ribosomal peptides, siderophores, chitinases, and bacteriocins.

34Keywords

35Betaproteobacteria -- *Burkholderia cepacia* complex -- *Burkholderia cenocepacia* -
36rhizosphere -antimicrobial - antagonistic.

37Abbreviations

38DOE, Department of Energy. JGI, Joint Genome Institute. ENA, European Nucleotide
39Archive. NRPS, Non-ribosomal peptide synthase. LlpA, lectin-like protein.

40Introduction

41The *Burkholderia cepacia* complex (Bcc) is a group of 22 closely related species [1]
42that share high similarity (>97.5%) in the 16S rRNA region of the *rrs* gene and low
43DNA-DNA hybridization values (30-60%) [2]. Bcc strains are versatile organisms,
44which are known as opportunistic pathogens in patients with cystic fibrosis (CF), but
45are also widely distributed in different environments including the rhizospheres of
46different crops [3]. Bcc members produce multiple antimicrobial compounds, which
47were recently reviewed by Depoorter *et al.* in 2016 [4].

48*B. cenocepacia* TAtl-371 was isolated from the rhizosphere of tomato growing in
49Atlatlahuacan, State of Mexico, and originally identified as *Burkholderia*
50(*Paraburkholderia*) *unamae* [5]. However, resequencing of the 16S region from the

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

63 Organism Information

64 Classification and features

65 *B. cenocepacia* strain TAtl-371 is a motile, Gram-negative, strictly aerobic,
66 nonsporulating rod in the order *Burkholderiales* of the class *Betaproteobacteria*
67 (Table 1). The rod-shaped form varies in size with dimensions of 0.5-0.7 μm in width
68 and 1-2 μm in length (Fig. 2). Cells grow fast in potato dextrose agar (PDA), Luria-
69 Bertani agar (LB) and Trypticase soy agar (TSA) at 30°C, and grow slowly in mineral
70 media such as BSE, BAc and Az [6]. This bacterium can grow using phenol as a
71 unique carbon source. Colonies on LB and TSA are uniform 1-3 mm diameter,
72 circular, convex, with entire smooth edges, shiny, non-pigmented (cream) and
73 opaque. On PDA, colonies are irregular, with undulate margins, smooth, shiny, non-
74 pigmented (white), opaque and moderately mucoid. The optimal growth
75 temperature is 30°C; however, it can grow at 37°C and 42°C. The general features of
76 this strain are shown in Table 1. The bacterium has natural resistance to
77 tetracycline (100 $\mu\text{g/ml}$), chloramphenicol (60 $\mu\text{g/ml}$), kanamycin (50 $\mu\text{g/ml}$) and
78 gentamicin (60 $\mu\text{g/ml}$). The strain produces siderophores, but phosphate
79 solubilization is minimal and it lacks the ability to fix nitrogen. *B. cenocepacia* TAtl-
80 371 is closely related to *B. cenocepacia* MC0-3, HI2424, and AU1054 (Fig 1).

81Genome sequencing information

82Genome project history

83*B. cenocepacia* TAtl-371 was sequenced at the Department of Energy Joint Genome
84Institute (DOE-JGI) as a part of the project “Root nodule microbial communities of
85legume samples collected from USA, Mexico and Botswana” directed by Dr. Ann M.
86Hirsch. The goal of this project was to identify the microbial community housed
87within nodules of native legumes living in three arid or semi-arid, nutrient-poor
88environments in Mexico, Botswana, and the United States.

89The complete sequence was finished in May 2015 and some features are presented
90in Table 2.

91Growth conditions and genomic DNA preparation

92*B. cenocepacia* TAtl-371 cells were grown in 5 ml of LB minus NaCl at 30°C for 18 h
93at 120 rpm. The DNA extraction was done using the Invitrogen Purelink™ Genomic
94DNA Mini Kit. The purified DNA was monitored for integrity by gel electrophoresis,
95and then sent to the JGI for sequencing.

96Genome sequencing and assembly

97The draft genome of *B. cenocepacia* TAtl-371 was generated at the DOE Joint
98Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology
99[7]. A Pacbio SMRTbell™ library was constructed and sequenced on the PacBio RS
100platform, which generated 213,715 filtered sub-reads totaling 943.3 Mbp. All
101general aspects of library construction and sequencing performed at the JGI can be
102found at <http://www.jgi.doe.gov>. The raw reads were assembled using HGAP
103(version: 2.3.0 p5, protocol version=2.3.0 method=RS HGAP Assembly.3, smrtpipe.py
104v1.87.139483,) [8]. The final draft assembly contained 3 contigs in 3 scaffolds,
105totalling 7.496 Mbp in size. The input read coverage was 91.6X.

106Genome annotation

107Genes were identified using Prodigal [9], followed by a round of manual curation
108using GenePRIMP [10] for finished genomes and draft genomes in fewer than 10
109scaffolds. The predicted CDSs were translated and used to search the National

Center for Biotechnology Information (NCBI) non-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 17,496,106 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*,

140and *B. metallica* [19]. *B. cenocepacia* TAtl-371's antibacterial activity against other
141Bcc species (*B. multivorans*, *B. dolosa*, *B. stabilis* and *B. pyrrocinia*) was determined
142by our research group (data not shown). We are continuing to study whether other
143LlpA sequences in the TAtl-371 genome encode gene products with antibacterial
144activity.

145Chitinases are glycosyl hydrolases that catalyze the hydrolytic degradation of chitin
146which is one of the major constituent of cell walls of a variety of fungi [20]. The TAtl-
147371 genome contains a gene that encodes a chitinase belonging to the family 18 of
148the glycosyl hydrolases as well as a gene that encodes a predicted chitinase. Using
149BLAST tools [21], we identified widely conserved homologues of these genes in
150several *B. cenocepacia* strains, but to the best of our knowledge, no studies about
151antifungal activity brought about by chitinases in this species have been published.
152The chitinolytic activity of *B. cenocepacia* TAtl-371 and the role of these enzymes in
153antifungal activity are currently under study in our lab.

154The *B. cenocepacia* TAtl-371 genome also contains genes that encode for possible
155resistance to the metalloids arsenate and arsenite and several metal ions, including
156tellurium, silver, cobalt, zinc, cadmium, copper, and chromate. Genes potentially
157involved in copper and chromate resistance are common in the genome. Finally,
158gene clusters for the biosynthesis cellulose and vanillin were identified; both
159molecules have potential as TAtl-371 metabolites with biotechnological applications.

160An Average Nucleotide Identity (ANI) calculation was used (Table 6) to compare the
161genome of *B. cenocepacia* TAtl-371 with the other sequenced Bcc genomes. The
162accepted ANI cut-off for species is 95-96% [22]. An ANI value >95.0% confirmed
163that strain TAtl-371 belongs to *B. cenocepacia*. Our phylogenetic analysis using the
164sequence of *recA* gene as well as the ANI values confirms that *B. cenocepacia* TAtl-
165371 belongs to PHDC lineage of the *B. cenocepacia* species [23], which is formed by
166strains isolated not only isolated from CF patients but also from natural
167environments around the world [24-26].

168**Conclusions**

169*B. cenocepacia* TAtl-371 was isolated from the tomato rhizosphere in Mexico and
170characterized as an antagonistic bacterium with a broad antimicrobial spectrum.
171The genome analysis of this strain allowed the identification of gene clusters
172encoding putative antimicrobial compounds such as siderophores, bacteriocins, and
173various enzymes that could mediate the antagonistic activity of TAtl-371. Our
174research group is currently studying the role of these compounds in the broad-
175spectrum antagonistic activity of this strain. The combination of genes involved in
176antagonism against other microbes, as well as the presence of genes involved in
177cellulose and vanillin biosynthesis, suggests that *B. cenocepacia* could be a source
178of compounds with yet to be explored biotechnological applications. Nevertheless,
179its close relationship with bacteria isolated CF patients requires additional research
180before such applications can be accepted.

181**Competing interests**

182The authors declare that they have no competing interests.

183**Funding**

184The genome sequence was conducted by the U.S. Department of Energy, Joint
185Genome Institute, a DOE Office of Science User Facility, is supported by the Office of
186Sciences of the U.S. Department of Energy under the proposal 1572 and contract
187No. DE-AC02-05CH11231.

188**Authors' contributions**

189FURR performed some phenotypic characterization of the strain, prepared bacteria
190for shipping to UCLA Laboratory, analyzed data and drafted the manuscript. DLS
191and EYTG performed some phenotypic characterization of the strain. MM and EH
192prepared the DNA for sequencing. MH, AC, MP, KP, NV, NM, DS, TBKR, VM, NI, NK,
193TW and NS performed the technical work for sequencing, assembly, and annotation
194of the genome. AMH led the manual annotation group at UCLA and wrote and
195reviewed the final manuscript. JAI and PES analyzed data and drafted the
196manuscript. All authors read and approved the final manuscript.

197 Acknowledgements

198

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

203

204

205 References

206

2071. Weber CF, King GM: **Volcanic Soils as Sources of Novel CO-Oxidizing**
208 **Paraburkholderia and Burkholderia: Paraburkholderia hiiakae sp.**
209 **nov., Paraburkholderia metrosideri sp. nov., Paraburkholderia**
210 **paradisi sp. nov., Paraburkholderia peleae sp. nov., and**
211 **Burkholderia alpina sp. nov. a Member of the Burkholderia cepacia**
212 **Complex.** *Front Microbiol* 2017, **8**:207.
2132. Coenye T, Vandamme P, Govan JRW, LiPuma JJ: **Taxonomy and**
214 **identification of the Burkholderia cepacia complex.** *Journal of Clinical*
215 *Microbiology* 2001, **39**(10):3427-3436.
2163. Compant S, Nowak J, Coenye T, Clement C, Ait Barka E: **Diversity and**
217 **occurrence of Burkholderia spp. in the natural environment.** *FEMS*
218 *Microbiol Rev* 2008, **32**(4):607-626.
2194. Depoorter E, Bull MJ, Peeters C, Coenye T, Vandamme P, Mahenthiralingam E:
220 **Burkholderia: an update on taxonomy and biotechnological potential**
221 **as antibiotic producers.** *Appl Microbiol Biotechnol* 2016, **100**(12):5215-
222 5229.
2235. Caballero-Mellado J, Onofre-Lemus J, Estrada-De Los Santos P, Martínez-
224 Aguilar L: **The tomato rhizosphere, an environment rich in nitrogen-**
225 **fixing Burkholderia species with capabilities of interest for**
226 **agriculture and bioremediation.** *Applied and environmental microbiology*
227 2007, **73**(16):5308-5319.
2286. Estrada-De Los Santos P, Bustillos-Cristales Ro, Caballero-Mellado J:
229 **Burkholderia, a genus rich in plant-associated nitrogen fixers with**
230 **wide environmental and geographic distribution.** *Applied and*
231 *environmental microbiology* 2001, **67**(6):2790-2798.
2327. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P,
233 Bettman B et al: **Real-time DNA sequencing from single polymerase**
234 **molecules.** *Science* 2009, **323**(5910):133-138.

2358. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A,
236 Copeland A, Huddleston J, Eichler EE *et al*: **Nonhybrid, finished microbial**
237 **genome assemblies from long-read SMRT sequencing data**. *Nat*
238 *Methods* 2013, **10**(6):563-569.
2399. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal:**
240 **prokaryotic gene recognition and translation initiation site**
241 **identification**. *BMC Bioinformatics* 2010, **11**:119.
24210. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A,
243 Kyrpides NC: **GenePRIMP: a gene prediction improvement pipeline for**
244 **prokaryotic genomes**. *Nat Methods* 2010, **7**(6):455-457.
24511. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of**
246 **transfer RNA genes in genomic sequence**. *Nucleic Acids Res* 1997,
247 **25**(5):955-964.
24812. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO:
249 **SILVA: a comprehensive online resource for quality checked and**
250 **aligned ribosomal RNA sequence data compatible with ARB**. *Nucleic*
251 *Acids Res* 2007, **35**(21):7188-7196.
25213. **INFERNAL**. <http://infernal.janelia.org/>. In.
25314. **IMG. The Integrated Microbial Genomes Plataform**
254 [<http://img.jgi.doe.gov/>]
25515. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC: **IMG**
256 **ER: a system for microbial genome annotation expert review and**
257 **curation**. *Bioinformatics* 2009, **25**(17):2271-2278.
25816. Esmaeel Q, Pupin M, Kieu NP, Chataigne G, Bechet M, Deravel J, Krier F, Hofte
259 M, Jacques P, Leclere V: **Burkholderia genome mining for nonribosomal**
260 **peptide synthetases reveals a great potential for novel siderophores**
261 **and lipopeptides synthesis**. *Microbiologyopen* 2016, **5**(3):512-526.
26217. Esmaeel Q, Pupin M, Jacques P, Leclère V: **Nonribosomal peptides and**
263 **polyketides of Burkholderia: new compounds potentially implicated**
264 **in biocontrol and pharmaceuticals**. *Environmental science and pollution*
265 *research international* 2017.
26618. Snyder AB, Worobo RW: **Chemical and genetic characterization of**
267 **bacteriocins: antimicrobial peptides for food safety**. *Journal of the*
268 *Science of Food and Agriculture* 2014, **94**(1):28-44.
26919. Ghequire MG, De Canck E, Wattiau P, Van Winge I, Loris R, Coenye T, De Mot
270 R: **Antibacterial activity of a lectin-like Burkholderia cenocepacia**
271 **protein**. *Microbiologyopen* 2013, **2**(4):566-575.
27220. Bhattacharya D, Nagpure A, Gupta RK: **Bacterial chitinases: properties**
273 **and potential**. *Critical reviews in biotechnology* 2007, **27**(1):21-28.
27421. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment**
275 **search tool**. *J Mol Biol* 1990, **215**(3):403-410.
27622. Richter M, Rosselló-Móra R: **Shifting the genomic gold standard for the**
277 **prokaryotic species definition**. *Proceedings of the National Academy of*
278 *Sciences* 2009, **106**(45):19126-19131.
27923. LiPuma JJ, Spilker T, Coenye T, Gonzalez CF: **An epidemic Burkholderia**
280 **cepacia complex strain identified in soil**. *The Lancet* 2002, **359**(9322).
28124. Coenye T, Spilker T, Van Schoor A, LiPuma J, Vandamme P: **Recovery of**
282 **Burkholderia cenocepacia strain PHDC from cystic fibrosis patients in**
283 **Europe**. *Thorax* 2004, **59**(11):952-954.

28425. Baldwin A, Mahenthiralingam E, Drevinek P, Vandamme P, Govan JR, Waine
285 DJ, LiPuma JJ, Chiarini L, Dalmastri C, Henry DA: **Environmental**
286 **Burkholderia cepacia complex isolates from human infections.**
287 *Emerging infectious diseases* 2007, **13**(3):458.
28826. Coenye T, LiPuma JJ: **Population structure analysis of Burkholderia**
289 **cepacia genomovar III: varying degrees of genetic recombination**
290 **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,
292 Thomson N, Allen MJ, Angiuoli SV *et al*: **The minimum information about a**
293 **genome sequence (MIGS) specification.** *Nat Biotechnol* 2008, **26**(5):541-
294 547.
29528. Woese CR, Kandler O, Wheelis ML: **Towards a natural system of**
296 **organisms: proposal for the domains Archaea, Bacteria, and Eucarya.**
297 *Proc Natl Acad Sci U S A* 1990, **87**(12):4576-4579.
29829. Garrity GM, Bell JA, Lilburn TE: **Phylum L. XIV. Proteobacteria phyl.**
299 **nov.** In: *Bergey's manual of systematic bacteriology*. Edited by Garrity GM,
300 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.**
302 In: *Bergey's Manual of systematic bacteriology*. Edited by Garrity GM,
303 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*
305 *manual of systematic bacteriology*. Edited by Garrity GM, Brenner DJ, Krieg
306 NR, Stanley JT, vol. 2, 2nd edn. New York: Springer; 2005.
30732. Garrity GM, Bell JA, Lisburn TE: **Family Burkholderiaceae.** In: *Bergey's*
308 *manual of systematic bacteriology*. Edited by Garrity GM, Brenner DJ, Krieg
309 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,
311 Arakawa M: **Proposal of Burkholderia gen. nov. and Transfer of Seven**
312 **Species of the Genus Pseudomonas Homology Group II to the New**
313 **Genus, with the Type Species Burkholderia cepacia (Palleroni and**
314 **Holmes 1981) comb. nov.** *Microbiology and Immunology* 1992,
315 **36**(12):1251-1275.
31634. Vandamme P, Holmes B, Coenye T, Goris J, Mahenthiralingam E, LiPuma JJ,
317 Govan JR: **Burkholderia cenocepacia sp. nov.--a new twist to an old**
318 **story.** *Res Microbiol* 2003, **154**(2):91-96.
31935. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP,
320 Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the**
321 **unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000,
322 **25**(1):25-29.
32336. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to**
324 **estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003,
325 **52**(5):696-704.
32637. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular**
327 **Evolutionary Genetics Analysis version 6.0.** *Mol Biol Evol* 2013,
328 **30**(12):2725-2729.

329

330

331

332

333

334

335

336

337

338

339

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

MIGS ID	Property	Term	Evidence code ^a
MIGS-6	Classification	Domain Bacteria	TAS [28]
		Phylum Proteobacteria	TAS [29]
		Class Betaproteobacteria	TAS [30]
		Order Burkholderiales	TAS [31]
		Family Burkholderiaceae	TAS [32]
		Genus <i>Burkholderia</i>	TAS [33]
		Species <i>Burkholderia cenocepacia</i>	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
MIGS-6.3	pH range; Optimum	6-8; 6	IDA
		Saccharose, succinic acid, malic acid,	
	Carbon source	glucose, mannitol, phenol, maltose	IDA
MIGS-6	Habitat	Rhizospheric soil	TAS [5]
MIGS-6.3	Salinity	Until 10% NaCl	IDA

	Oxygen		
MIGS-22	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 18° 56' 8.724''	IDA
MIGS-4.2	Longitude	O 98° 55' 2.588''	IDA
MIGS-4.4	Altitude	1656 M	IDA

342^a Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a
343direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly
344observed for the living, isolated sample, but based on a generally accepted property for the
345species, or anecdotal evidence). These evidence codes are from the Gene Ontology project
346[35]

347

348

349**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 13 th 2016
	GOLD ID	Ga0073272
	BIOPROJECT	PRNJA332745
MIGS 13	Source Material Identifier	
	Project relevance	Environmental

350

351

352**Table 3.** Genome statistics.

353Attribute	Value	% of Total
354	749610	100.0
355Genome size (bp)	6	0
356	659979	
357DNA coding (bp)	6	88.04
358DNA G+C (bp)	5	67.01
359		100.0
360DNA scaffolds	3	0
361		100.0
362Total genes	6821	0
363Protein 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 Cell cycle control, Cell division, chromosome
D	35	0.61	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 Posttranslational modification, protein turnover,
O	192	3.32	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 Secondary metabolites biosynthesis, transport and
Q	195	3.37	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.

369**Table 5.** *Burkholderia cenocepacia* TAtl-371 genes related to
370biotechnological applications

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

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

			lectin
<i>pchD</i> ²	Ga0073272_5545	1653	Chtinase family 18
<i>pchR</i> ²	Ga0073272_5546	942	Isochorismate synthase
<i>pchE</i> ²	Ga0073272_5547	4374	Isochorismate pyruvate lyase
----- ²	Ga0073272_5548	5478	Pyochelin biosynthetic protein PchC
<i>pchG</i> ²	Ga0073272_5549	1053	Pyochelin biosynthesis protein PchD
-----	Ga0073272_5789	2442	
Metal-ion resistance			Transcriptional regulator, AraC family
BCAL1231	Ga0073272_0413	756	
			Dihydroaeruginoic acid synthetase
	Ga0073272_0796	612	
-----			Pyochelin synthetase
	Ga0073272_0797	543	

	Ga0073272_1837	528	Pyochelin biosynthetic protein PchG
BCAL0581	Ga0073272_2310	1212	
BCAL2754			Predicted chitinase
	Ga0073272_2473	555	
BCAL2607			
	Ga0073272_2517	1206	
BCAL2450			Membrane protein
	Ga0073272_3568	348	

-----			TerC, possibly involved in tellurium resistance
BCAM0713	Ga0073272_4355	3213	
<i>czcB</i>	Ga0073272_4357	1470	Chromate transporter
<i>czcC</i>	Ga0073272_4358	1314	Chromate transporter
			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 CzcA
<i>cusA</i>	Ga0073272_4636	3207	
BCAM0434	Ga0073272_4637	1536	Membrane fusion protein, cobalt-zinc-cadmium efflux system
			outer membrane

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

BCAL1391 ³			(CbtB)
	Ga0073272_0287	1230	Probable cobalt transporter subunit (CbtA)
<i>bcsZ</i> ³	Ga0073272_0288	2346	
<i>bcsB</i> ³			Arsenite efflux membrane protein ArsB
	Ga0073272_4981	831	
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

371^a, NRPS gene cluster for siderophore ornibactin biosynthesis

372^b, gene cluster for siderophore pyochelin biosynthesis

373^c, genes related to cellulose biosynthesis

374^d, genes related to vanillin biosynthesis

375

376

377

378

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

381

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

382

383

384

385

386

387

388

389Figure legends

390

391Figure 1.

392Phylogenetic tree highlighting the position of *Burkholderia cenocepacia* TAtl-371 in
393relation to other *Burkholderia cepacia* complex strains. *B. mallei* ATCC 23344^T was
394used as outgroup. The tree is based on *recA* gene alignments and was performed by
395the maximum-likelihood (ML) method with the General Time Reversible model using
396PhyML [36]. Bootstrap values inferred from 100 replicates are shown at nodes. The
397tree was visualized with program MEGA 6 [37]. All strains used have a sequenced
398genome. The GenBank accession number of each sequence is in brackets.

399

400Figure 2.

401Transmission Electron Microscopy of negatively stained *Burkholderia*
402*cenocepacia* TAtl-371 cells. The strain was grown on LB medium, cells were cultured
403by centrifugation at 3000 rpm, were washed with PBS, exposed to glutaraldehyde
404for one hour and finally were washed and resuspended in PBS. A drop of the
405suspension was placed on a formvar-coated copper grid and air-dried for 20 min to
406allow the cells to adhere. The grid was then covered for 20 s with a solution of 0.5%
407uranyl acetate, the excess liquid was removed with a filter paper, and then air-
408dried. A JEOL JEM-1010 transmission electron microscope, operated at 60 kV, was
409used to observe and photograph negatively stained preparations. F, stands for
410flagella.

411

412

413Figure 3.

414Graphical map of the three scaffolds of the genome of *B. cenocepacia* TAtl-371.
415From the bottom to the top of each scaffold: Genes on forward strand (color by COG

416categories as denoted by the IMG platform), Genes on reverse strand (color by COG
417categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC
418skew

419