Rethinking the Niche of Upper-Atmosphere Bacteria: Draft Genome Sequences of *Bacillus aryabhattai* C765 and *Bacillus aerophilus* C772, Isolated from Rice Fields

Paula Ramos-Silva,[a] Patrícia H. Brito,[a] Mónica Serrano,[b] Adriano O. Henriques,[b] José B. Pereira-Leal[a]

Instituto Gulbenkian de Ciência, Oeiras, Portugal[a]; Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal[b]

Here, we report two genome sequences of endospore-forming bacteria isolated from the rice fields of Comporta, Portugal, identified as *Bacillus aryabhattai* C765 and *Bacillus aerophilus* C772. Both species were previously identified in air samples from the upper atmosphere, but our findings suggest their presence in a wider range of environmental niches.

Address correspondence to Adriano O. Henriques, aoh@itqb.unl.pt, or José B. Pereira-Leal, jleal@igc.gulbenkian.pt.

The strains *Bacillus aryabhattai* C765 and *Bacillus aerophilus* C772 were identified in samples collected from the rice fields of the Portuguese coastal area of Comporta, which is near the Sado River and the Atlantic Ocean. This finding confirms the presence of these species in a broader set of environments than was previously suggested (1, 2), supporting a freshwater ecological niche.

When grown in Difco sporulation medium (DSM), both organisms form subterminal ellipsoidal endospores that do not swell the mother cell. Prior to DNA extraction, the cells were grown at 37°C in LB medium until reaching an optical density at 600 nm (OD_{600}) of 1. The genomes were sequenced with the Illumina MiSeq system (300-bp paired-end library), generating >1.8 million reads, with average coverages of 123× and 146× for *B. aryabhattai* C765 and *B. aerophilus* C772, respectively. More than 94% of the reads were assembled using SPAdes 3.0.0 (3). A total of 49 contigs (5.5 Mbp, 38% G+C content) were obtained for *B. aryabhattai* C765, and 17 contigs were obtained for *B. aerophilus* (3.7 Mbp, 41% G+C content). Only the contigs of >1,000 bp and with a coverage of >10× were used. In the assembly of *B. aryabhattai*, one contig corresponds to a complete sequence of a plasmid of 10,154 bp, with a coverage of 1,339×. Another five contigs showed high identity (≥88%) and significant overlap (≥18%) with the plasmid sequences of *Bacillus megaterium* QMB1551; still, it was not possible to conclude whether these contigs belong to plasmid or chromosomal DNA.

A total of 5,683 gene predictions in *B. aryabhattai* C765 were generated with Prodigal 2.60 (4) and annotated using Prokka 1.9 (5). The same software predicted 3,809 genes in *B. aerophilus* C772.

Searches in the 16S rRNA database EzTaxon (6) indicated *B. aryabhattai* B8W22[T] and *B. aerophilus* 28K[T] as the closest species (>98.6% sequence identity).

Reciprocal Best-Hits BLAST analyses revealed that strain C765 has a surprising 896 strain-specific genes in relation to the closest genomes of *B. aryabhattai* GZ203 and *B. megaterium* WSH-002 (5), and these organisms share a core of 4,361 orthologues. Both strains of *B. aryabhattai* share about 4% more orthologous genes than they do with *B. megaterium*. Strain C772 has 226 strain-specific genes and shares 3,391 orthologous genes with *Bacillus* sp. M 2-6 strain KACC 16563, a possible representative of *B. aerophilus* (6), and *Bacillus pumilus* MTCC B6033 (7). Strain C772 shares 6% more orthologous genes with strain KACC 16563 than it does with *B. pumilus*. In both genomes, the top GO terms (7) of specific genes corresponded to DNA binding (GO:0003677) and ATP binding functions (GO:0005524).

The plasmid identified in *B. aryabhattai* C765, termed pBa1, contains 10 coding sequences (CDSs), including two putative relaxase/mobilization proteins, one 3-oxo-5-alpha-steroid 4-dehydrogenase, one plasmid replication protein (RepL), and one antibiotic resistance regulatory protein of the MarR family.

Nucleotide sequence accession numbers. The whole-genome shotgun projects of *B. aryabhattai* C765 and *B. aerophilus* C772 were deposited at DDBJ/EMBL/GenBank under the accession numbers JXRC00000000 and JXRO00000000.

REFERENCES

1. Shivaji S, Chaturvedi P, Suresh K, Reddy GS, Dutt CB, Wainwright M, Narlikar JV, Bhargava PM. 2006. *Bacillus aerius* sp. nov., *Bacillus aerophilus* sp. nov., *Bacillus stratosphericus* sp. nov. and *Bacillus altitudinis* sp. nov., isolated from cryogenic tubes used for collecting air samples from high altitudes. Int J Syst Evol Microbiol 56:1465–1473. http://dx.doi.org/10.1099/ijs.0.64029-0.
2. Shivaji S, Chaturvedi P, Begum Z, Pindi PK, Manorama R, Padmanaban DA, Shouche YS, Pawar S, Vaishampayan P, Dutt CB, Datta GN, Manchanda RK, Rao UR, Bhargava PM, Narlikar JV. 2009. *Janibacter hoylei* sp. nov., *Bacillus isronensis* sp. nov. and *Bacillus aryabhattai* sp. nov., isolated from cryotubes used for collecting air from the upper atmosphere. Int J Syst Evol Microbiol 59:2977–2986. http://dx.doi.org/10.1099/ijs.0.002527-0.

3. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

4. **Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.

5. **Seemann T**. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.

6. **Kim O-S, Cho Y-J, Lee K, Yoon S-H, Kim M, Na H, Park S-C, Jeon YS, Lee J-H, Yi H, Won S, Chun J.** 2012. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol **62:**716–721. http://dx.doi.org/10.1099/ijs.0.038075-0.

7. **Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G**. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet **25:**25–29. http://dx.doi.org/10.1038/75556.