

Data in Brief

The draft genome sequence of *Hymenobacter* sp. CRA2 isolated from Nama Karoo shrub land soils from South Africa



Mubanga H. Kabwe^{a,b}, Nerissa Govender^{a,c}, Surendra Vikram^{a,c}, Oliver M. Bezuidt^d, Thulani P. Makhalanyane^{a,c,*}

^a Centre for Microbial Ecology and Genomics, University of Pretoria, Pretoria 0028, South Africa

^b Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0028, South Africa

^c Department of Genetics, University of Pretoria, Pretoria 0028, South Africa

^d Biotechnology Platform, Agricultural Research Council, Pretoria 0028, South Africa

A B S T R A C T

Here we report the draft genome sequence of *Hymenobacter* sp. CRA2 isolated from the Nama Karoo shrub land soils of the Northern Cape, South Africa. This genome is approximately 5.88 Mb long and the assembly comprised 45 contigs. The draft genome sequence has been deposited in DDBJ/EMBL/GenBank under the accession number NZ_MVBC00000000 and is available for download at: https://www.ncbi.nlm.nih.gov/nucleotide/NZ_MVBC00000000.1.

Specifications

Organism/cell line/tissue	<i>Hymenobacter</i> sp. CRA2
Sex	Not applicable
Sequencer or array type	Illumina HiSeq
Data format	Analyzed
Experimental factors	<i>Hymenobacter</i> sp. CRA2 isolated from an environmental soil sample
Experimental features	Isolation of Bacteria, genomic DNA extraction, genome sequencing, de novo assembly, annotation, and analysis
Consent	Not applicable
Sample source location	Carnavon, South Africa, Latitude & Longitude – 31.13132 & 21.60795

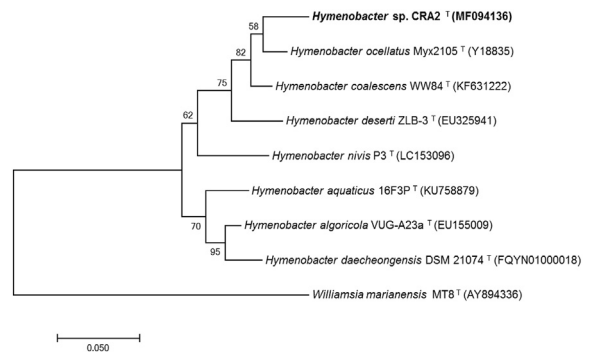


Fig. 1. Molecular phylogenetic analysis of *Hymenobacter* sp. CRA2 (shown in bold). The evolutionary history of the strain was inferred by applying the Maximum Likelihood method based on the Tamura-Nei model and 500 bootstraps following a ClustalW alignment [8]. Initial trees were obtained automatically by applying Neighbor-Join and BioNJ algorithms using the Maximum Composite Likelihood (MCL) approach. Topology with superior log likelihood value was then selected. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The 16S rRNA gene sequence of *Williamsia marianensis* was selected as an outgroup. Phylogenetic analyses were conducted using MEGA7 version 7.0.18 [9].

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/nucleotide/NZ_MVBC00000000.1

2. Introduction

The genus *Hymenobacter* was first described by Hirsch et al. (1998) from sandstone and soil samples retrieved from the Antarctic Dry

Valleys [1]. Members of this genus, which belong to the family Cytophagaceae (phylum Bacteroidetes) have a considerably high G + C content (55–70 mol%) compared to other genera of the same family [2]. Representatives in this phylum are Gram-negative, rod shaped,

* Corresponding author at: Room 3-14, Centre for Microbial Ecology and Genomics, Natural Sciences 2, University of Pretoria, Lynnwood Road, Pretoria 0028, South Africa.
E-mail address: thulani.makhalanyane@up.ac.za (T.P. Makhalanyane).

Table 1
List of tRNA annotated in the draft genome of *Hymenobacter* sp. CRA2 generated from ARAGORN v1.2.38.

tRNA	Occurrence
tRNA-Gln	2
tRNA-Thr	4
tRNA-Asn	1
tRNA-Cys	1
tRNA-Val	3
tRNA-His	1
tRNA-Asp	2
tRNA-Pro	3
tRNA-Tyr	1
tRNA-Met	3
tRNA-Leu	5
tRNA-Ile	1
tRNA-Arg	4
tRNA-Ala	2
tRNA-Glu	2
tRNA-Trp	1
tRNA-Ser	4
tRNA-Phe	1
tRNA-Lys	2
tRNA-Gly	5
Total	48

non-spore forming bacteria, and are found in both terrestrial and aquatic environments. Currently 44 species of this genus have been described and are validly published in the [LPSNbacterio.net](http://www.bacterio.net/hymenobacter.html#r) database (<http://www.bacterio.net/hymenobacter.html#r>). Some identified species of the genus *Hymenobacter* have been isolated from cold environments [3–5], which suggests that members may have capacity to rapidly adapt and survive in extreme ecosystems. The analysis of *Hymenobacter* genomes may provide crucial insights on the molecular adaptations to environmental extremes. Here we present the draft genome sequence of *Hymenobacter* sp. CRA2 isolated from Nama Karoo soils, along an aridity gradient.

3. Experimental design, materials and methods

Soils (0.5 g) were transferred into a sterile 2 mL tube, before homogenization in 1 mL of deionized water, and vortexed at maximum speed for 10 s. The solution was then centrifuged at 11000 rpm for 60 s. 100 µL of the supernatant was plated onto R2A agar plates, supplemented with cyclohexamine (100 mg/mL). After optimum growth at 22 °C for 72 h, genomic DNA was extracted using a method combining chemical lysis and bead-beating [6]. *Hymenobacter* sp. CRA2 was morphologically identified and further confirmed by 16S rRNA gene amplification and sequencing [7]. The partial 16S rRNA gene sequence of sp. CRA2 has been deposited at the NCBI GenBank (accession number MF094136). A phylogenetic tree was constructed from the near

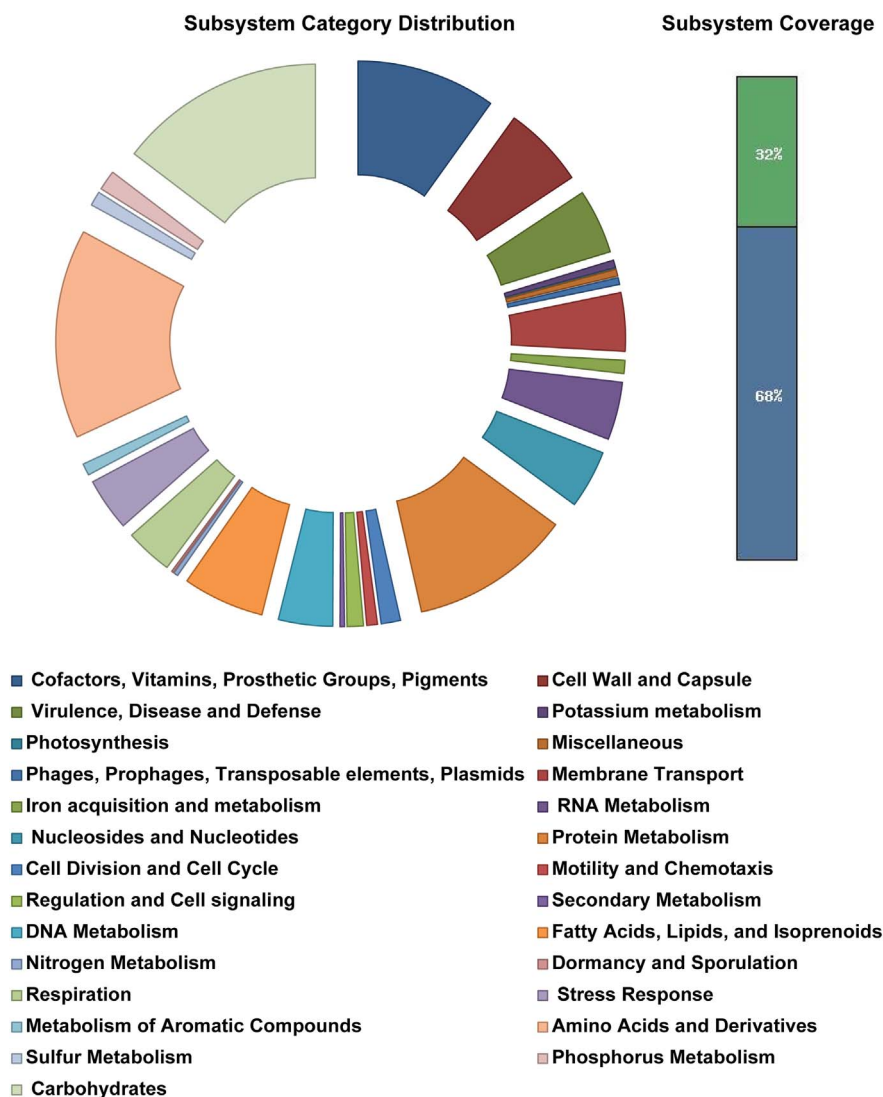


Fig. 2. Diagrammatic representation of the subsystem categories of the *Hymenobacter* sp. CRA2 genome annotated by RAST.

complete 16S rRNA gene sequence of CR2 (Fig. 1). The tree is drawn to scale and analysis involved 9 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 1365 positions in the final dataset.

High molecular weight genomic DNA, isolated from *Hymenobacter* sp. CRA 2, was sent for sequencing at the Molecular Research LP next generation sequencing service (www.mrdnalab.com, Shallowater, TX, USA). DNA libraries were constructed using the paired-end DNA sample prep kit v1 according Illumina specifications using Illumina paired-end adapters [10]. Sequencing was performed on the Illumina HiSeq 2500 in paired-end mode using Illumina data analysis pipeline version 1.8 [11].

The generated paired end reads were processed with the PRINSEQ lite version 0.20.4 [12] for sequence read quality assessment. Subsequently reads were assembled into a draft genome with careful option and coverage cut off value set at auto using SPAdes version 3.7.1 [13].

4. Data description

The draft genome of *Hymenobacter* sp. CRA2 constituted a total of 45 contigs (> 500 bp) amounting to 5,879,518 bp with a G + C content of 62.8% and a N50 contig size of 298, 233 bp. A total of 48 tRNA genes were also identified using ARAGORN v1.2.38 [14] Table 1.

The assembled draft genome was annotated through Rapid Annotation using Subsystems Technology (RAST) version 2.0 [15] and the NCBI prokaryotic genome annotation pipeline (PGAP) [16]. The RAST annotation revealed 376 subsystems (Fig. 2).

Subsystem coverage showed that 32% of the coverage was within the subsystem and constituted a total of 1465 non-hypothetical and 65 hypothetical proteins. Whereas, 68% coverage was not in the subsystem and constituted a total of 3359 proteins of which 1310 were non-hypothetical and 2049 hypothetical protein. Annotation also revealed an abundance of genes involved in the synthesis of amino acids and their derivatives.

References

[1] P. Hirsch, W. Ludwig, C. Hethke, M. Sittig, B. Hoffmann, C. Gallikowski,

- Hymenobacter roseosalivarius* gen. nov., sp. nov. from continental Antarctic soils and sandstone: bacteria of the Cytophaga/Flavobacterium/Bacteroides line of phylogenetic descent, *Syst. Appl. Microbiol.* 21 (1998) 374–383.
- [2] K. Tang, B. Yuan, Q. Lai, R. Wang, H. Bao, F.Y. Feng, *Hymenobacter terrenus* sp. nov., isolated from biological soil crusts, *Int. J. Syst. Evol. Microbiol.* 65 (2015) 4557–4562.
- [3] D.H. Ahn, S.-R. Han, T.-J. Oh, H. Park, Complete genome sequence of ionizing radiation-resistant *Hymenobacter* sp. strain PAMC26628 isolated from an Arctic lichen, *J. Biotechnol.* 223 (2016) 50–51.
- [4] J.L. Klassen, Carotenoid Diversity in Novel *Hymenobacter* Strains Isolated From Victoria Upper Glacier, Antarctica, and Implications for the Evolution of Microbial Carotenoid Biosynthesis, (2009).
- [5] H. Koo, T. Ptacek, M. Crowley, A.K. Swain, J.D. Osborne, A.K. Bej, D.T. Andersen, Draft genome sequence of *Hymenobacter* sp. strain IS2118, isolated from a freshwater lake in Schirmacher Oasis, Antarctica, reveals diverse genes for adaptation to cold ecosystems, *Genome Announc.* 2 (2014) (e00739-00714).
- [6] D. Miller, J. Bryant, E. Madsen, W. Ghiorse, Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples, *Appl. Environ. Microbiol.* 65 (1999) 4715–4724.
- [7] M. Drancourt, C. Bollet, A. Carlouz, R. Martelin, J.-P. Gayral, D. Raoult, 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates, *J. Clin. Microbiol.* 38 (2000) 3623–3630.
- [8] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [9] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, *Mol. Biol. Evol.* (2016) msw054.
- [10] D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, K.P. Hall, D.J. Evers, C.L. Barnes, H.R. Bignell, Accurate whole human genome sequencing using reversible terminator chemistry, *Nature* 456 (2008) 53–59.
- [11] D. Aird, M.G. Ross, W.-S. Chen, M. Danielsson, T. Fennell, C. Russ, D.B. Jaffe, C. Nusbaum, A. Gnirke, Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries, *Genome Biol.* 12 (2011) R18.
- [12] R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets, *Bioinformatics* 27 (2011) 863–864.
- [13] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477.
- [14] D. Laslett, B. Canback, ARAGORN, a program for the detection of transfer RNA and transfer-messenger RNA genes, *Nucleic Acids Res.* 32 (2004) 11–16.
- [15] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formisano, S. Gerdes, E.M. Glass, M. Kubal, The RAST server: rapid annotations using subsystems technology, *BMC Genomics* 9 (2008) 75.
- [16] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E.P. Nawrocki, L. Zaslavsky, A. Lomsadze, K.D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline, *Nucleic Acids Res.* (2016) gkw569.