Draft genome sequence of the cellulolytic endophyte *Chitinophaga costaii* A37T2<sup>T</sup>

Diogo N. Proença<sup>1</sup>, William B. Whitman<sup>2</sup>, Nicole Shapiro<sup>3</sup>, Tanja Woyke<sup>3</sup>, Nikos C. Kyrpides<sup>3</sup> and Paula V. Morais<sup>1,4*</sup>

**Abstract**

Here we report the draft genome sequence of *Chitinophaga costaii* A37T2<sup>T</sup> (=CIP 110584<sup>T</sup> =LMG 27458<sup>T</sup>), which was isolated from the endophytic community of *Pinus pinaster* tree. The total genome size of *C. costaii* A37T2<sup>T</sup> is 5.07 Mbp, containing 4204 coding sequences. Strain A37T2<sup>T</sup> encoded multiple genes likely involved in cellulolytic, chitinolytic and lipolytic activities. This genome showed 1145 unique genes assigned into 109 Cluster of Orthologous Groups in comparison with the complete genome of *C. pinensis* DSM 2588<sup>T</sup>. The genomic information suggests the potential of the strain A37T2<sup>T</sup> to interact with the plant metabolism. As there are only a few bacterial genomes related to Pine Wilt Disease, this work provides a contribution to the field.

**Keywords:** *Chitinophaga costaii* A37T2, Cellulase, Chitinase, Genome sequence

**Introduction**

The genus *Chitinophaga* belongs to the family *Chitinophagaceae* (phylum *Bacteroidetes*) alongside with the genera *Arachidicoccus*, *Asinibacterium*, *Baheola*, *Cnueula*, *Crenotalea*, *Ferruginibacter*, *Filimonas*, *Flaviaesturaribacter*, *Flavimumibacter*, *Flavisolibacter*, *Flavitalae*, *Gracilimonas*, *Heliimonas*, *Hydrotalea*, *Lacibacter*, *Niabella*, *Niastella*, *Parasediminibacterium*, *Parasegetibacter*, *Sediminibacterium*, *Segetibacter*, *Taibaiella*, *Terrimonas*, *Thermoflavifillum* and *Vibrioronas*. The genus *Chitinophaga* is widely distributed in the environment and strains of this genus have been isolated from pine trees, soil, rhizosphere soil, roots, vermicompost and weathered rock [1]. Twenty-four species belonging to the genus *Chitinophaga* have been described [2], and only the type species of the genus *C. pinensis* has the complete genome sequenced [3].

*Pinus pinaster* trees from Central Portugal present a diverse endophytic microbial community. Strain A37T2<sup>T</sup> was isolated as part of the endophytic microbiome of pine trees affected by Pine Wilt Disease (PWD) which is a world devastating disease, consequence of *Bursaphelenchus xylophilus* colonization in pine trees [4]. Here, we show the second genome of the genus *Chitinophaga*, a draft genome of *Chitinophaga costaii* A37T2<sup>T</sup>, previously isolated as endophyte of *Pinus pinaster* affected by PWD [1].

**Organism information**

**Classification and features**

The type strain A37T2<sup>T</sup> (=CIP 110584<sup>T</sup> =LMG 27458<sup>T</sup>), was isolated from tree trunk of a *Pinus pinaster* tree affected by PWD and it described as *Chitinophaga costaii* (family *Chitinophagaceae*, phylum *Bacteroidetes*) [1]. It was Gram-stain-negative, facultative anaerobic, non-motile, formed rod-shaped cells, 0.5-1 μm in diameter and 1.8 μm in length after 48 h on R2A agar media (Fig. 1). Showed capacity to grow on R2A agar medium at 15-45 °C (optimum, 26-30 °C), at pH 5.5-8.0 (optimum, pH 7) and supplemented with up to 1% (w/v) NaCl (optimum without NaCl). The major fatty acids (>25%) showed by the strain A37T2<sup>T</sup> are saturated iso-C<sub>15:0</sub> and unsaturated C<sub>16:1 ω5c</sub>. The major polar lipids were identified as phosphatidylethanolamine, two unidentified aminophospholipids and one unidentified lipid. No glycolipid was detected. The menaquinone 7 (MK-7) was shown as the major respiratory lipoquinone. The determined DNA G + C content of the *C. costaii*...
A37T2\textsuperscript{T} was 46.6 mol%. Key features of this microorganism are summarized in Table 1. A phylogenetic tree based on the 16S rRNA gene sequence of this strain and its closest relative members are given in Fig. 2. The sequences were aligned by SINA (v1.2.9) using the SILVA SEED as reference alignment [5]. Sequences were included in 16S rRNA-based Living Tree Project (LTP) release 115 database [6] by parsimony implemented in the ARB software package version 5.5 [7]. Evolutionary distances were calculated [8] and phylogenetic dendrograms were constructed using the neighbor-joining [9] and Randomized Accelerated Maximum Likelihood (RAxML) method with GTRGAMMA model [10] included in the ARB software [7]. Trees topologies were evaluated by performing bootstrap analysis [11] of 1000 data sets by using ARB software package.

![Scanning electron micrograph of C. costaii A37T2\textsuperscript{T} after 48 h of growth on R2A agar plates at 30 °C.](image)

**Table 1** Classification and general features of *Chitinophaga costaii* A37T2\textsuperscript{T} according to the MIGS recommendations [26]

| MIGS ID | Property | Term | Evidence code\textsuperscript{a} |
|---------|----------|------|----------------------------------|
|         | **Classification** | | |
|         | Domain | Bacteria | TAS [27] |
|         | Phylum | Bacteroidetes | TAS [28, 29] |
|         | Class | Sphingobacteria | TAS [28, 30] |
|         | Order | Sphingobacteriales | TAS [28, 31] |
|         | Family | Chitinophagaceae | TAS [32] |
|         | Genus | Chitinophaga | TAS [33] |
|         | Species | *Chitinophaga costaii* | TAS [1] |
|         | Type strain: | A37T2\textsuperscript{T} (\text{=CIP 110584\textsuperscript{T}, =LMG 27458\textsuperscript{T}}) | |
|         | Gram stain | Negative | TAS [1] |
|         | Cell shape | Rod | TAS [1] |
|         | Motility | Non-motile | TAS [1] |
|         | Sporulation | Not reported | NAS |
|         | Temperature range | 15-45 °C | TAS [1] |
|         | Optimum temperature | 26-30 °C | TAS [1] |
|         | pH range; Optimum | 5.5-8.0, 7 | TAS [1] |
|         | Carbon source | Glucose | TAS [1] |
| MIGS-6 | Habitat | Endophyte of *Pinus pinaster* tree | TAS [1] |
| MIGS-6.3 | Salinity | 1.0% NaCl (w/v) | TAS [1] |
| MIGS-22 | Oxygen requirement | Facultative anaerobic | TAS [1] |
| MIGS-15 | Biotic relationship | Free-living | TAS [1] |
| MIGS-14 | Pathogenicity | Non-pathogen | NAS |
| MIGS-4 | Geographic location | Portugal | TAS [1] |
| MIGS-5 | Sample collection | July, 2009 | NAS |
| MIGS-4.1 | Latitude | 40.2962266 | NAS |
| MIGS-4.2 | Longitude | −7.9207357 | NAS |
| MIGS-4.4 | Altitude | 217 m | NAS |

\textsuperscript{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 [34]
Genome sequencing information

Genome project history

This Whole Genome Shotgun project has been deposited at ENA under the accession numbers FMAR01000001-FMAR01000056 and in the Integrated Microbial Genomes database (IMG) with Biosample ID SAMN05216457 [12]. The genome sequencing of this organism is part of the Genomic Encyclopedia of Bacteria and Archaea [13], 1000 Microbial Genomes project, phase III (KMG-III) [14], at the U.S. Department of Energy, Joint Genome Institute (JGI). The project information and its association with the MIGS is summarized in Table 2.

Growth conditions and genomic DNA preparation

The strain A37T2 was grown on R2A agar media at 30°C during 48 h and its genomic DNA was extracted using the E.Z.N.A. Bacterial DNA Kit (Omega Bio-Tek, Norcross, GA, USA) and purified using the Wizard Pure DNA Purification kit (Promega, Madison, WI, USA).

Table 2: Project information

| MIGS ID | Property                  | Term                                      |
|---------|---------------------------|-------------------------------------------|
| MIGS 31 | Finishing quality         | Draft                                     |
| MIGS-28 | Libraries used            | Illumina Regular Fragment, 300 bp, Tubes  |
| MIGS 29 | Sequencing platforms      | Illumina HiSeq 2500-1 TB                  |
| MIGS 31.2 | Fold coverage            | 297.2                                     |
| MIGS 30 | Assemblers                | SPAdes                                    |
| MIGS 32 | Gene calling method       | NCBI Prokaryotic Genome Annotation Pipeline|
|         | Locus Tag                 | GAD116948                                 |
|         | Genbank ID                | FMAR00000000                              |
|         | GenBank Date of Release   | August 3, 2016                            |
|         | GOLD ID                   | Gp0139259                                 |
|         | BIOPROJECT                | PRJNA322901                               |
| MIGS 13 | Source Material Identifier| A37T2                                      |
|         | Project relevance         | GEBA-KMG                                  |
Genome sequencing and assembly

The draft genome of *C. costaii A37T2<sup>T</sup>* was generated at the DOE Joint Genome Institute (JGI) using the Illumina technology [15]. An Illumina 300 bp insert standard shotgun library was constructed and sequenced using the Illumina HiSeq–2500 1 TB platform, generating 9,965,394 reads totaling 1494.8 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found at [16]. All raw Illumina sequence data was filtered using BBDuk [17], which removes known Illumina artifacts and PhiX. Reads with more than one “N” or with quality scores (before trimming) averaging less than 8 or reads shorter than 51 bp (after trimming) were discarded. Remaining reads were mapped to masked versions of human, cat and dog references using BBMAP [17] and discarded if identity exceeded 95%. Sequence masking was performed with BBMask [17]. Following steps were then performed for assembly: (1) artifact filtered Illumina reads were assembled using SPAdes (version 3.6.2) [18]; (2) assembled contigs were discarded if length was <1 kbp. Parameters for the
SPAdes assembly were --cov-cutoff auto --phred-offset 33 -t 8 -m 40 --careful -k 25,55,95 --12.

**Genome annotation**
Protein-coding genes were identified using Prodigal [19], as part of the DOE-JGI genome annotation pipeline [20]. Additional gene prediction analysis and manual functional annotation were performed within the Integrated Microbial Genomes Expert Review system (IMG-ER), which provides tools for analyzing and reviewing the structural and functional annotations of genomes in a comparative context [12, 21]. Genome annotation procedures are detailed in Markowitz et al. [12] and references therein. Briefly, the predicted CDSs were translated and used to search the NCBI nonredundant database, UniProt, TIGRFam, Pfam, KEGG, COG and InterPro databases. Transfer RNA genes were identified using the tRNAscan-SE tool and other non-coding RNAs were found using INFERNAL. Ribosomal RNA genes were predicted using hmmssearch against the custom models generated for each type of rRNA.

**Genome properties**
The draft genome sequence of *C. costaii* strain A37T2\textsuperscript{T} comprised 5,074,440 bp, based on 1494.8 Mbp of Illumina data with a mapped coverage of 297.2-fold of the genome. The final draft assembly contained 56 contigs in 56 scaffolds with more than 1052 bp. The G + C content was 47.6%. The genome encoded 4204 putative coding sequences (CDSs) (Table 3). Fifty four % of the CDSs, corresponding to 2284 proteins, could be assigned to Cluster of Orthologous Groups (COG) families [22] (Table 4). The draft genome sequence contained four ribosomal RNAs and 50 tRNAs loci (Table 3).

The Average Nucleotide Identity between *C. costaii* A37T2\textsuperscript{T} and *C. pinensis* DSM 2588\textsuperscript{T} was 70.9 based on 1593 of total Bidirectional Best Hits, using MiSI [23]. Figure 3 shows the circular graph of the genome of *C. costaii* A37T2\textsuperscript{T} query to the only available complete genome of the genus *Chitinophaga*, *C. pinensis* DSM 2588\textsuperscript{T} [2].

The comparison between the draft genome of *C. costaii* A37T2\textsuperscript{T} and the complete genome of *C. pinensis*
DSM 2588<sup>T</sup> showed 1145 unique genes only present in the genome of <i>C. costaii</i> A37T<sup>2T</sup> and 3493 unique genes only present in the genome of <i>C. pinensis</i> DSM 2588<sup>T</sup>. Focused on the unique genes present on the genome of strain A37T<sup>2T</sup> it was possible to assigned 109 COG, summarized in Table 5.

### Insights from the genome sequence

The draft genome sequence of <i>C. costaii</i> A37T<sup>2T</sup> carries multiple genes involved in cellulolytic activity, including one gene encoding the enzyme cellulase (SCC15587) and six genes encoding for β-glucosidase (SCB82491, SCB92249, SCB95191, SCC15475, SCC57293, SCC61957), which might be involved in cellulose degradation in the environment and in biotechnological processes [24]. As expected for this genus, four genes encoding chitinases (SCC19468, SCC19522, SCC23114, SCC34676) were found. Six genes encoded lysophospholipase L1, including representatives of both of size groups, i.e. less than 300aa (SCB77875, SCC28514, SCC37316, SCC54197) and less than 500aa (SCB98645, SCC50813). Moreover, the genome of strain A37T<sup>2T</sup> encoded 1-aminocyclopropane-1-carboxylate deaminase (SCB80758), a hydrolase that might be involved in lowering ethylene levels in the plant [25]. In summary, the genome sequence suggested multiple potentials for the strain to interact with the plant metabolism.

### Conclusions

This work contributed to the knowledge of the genome sequence of the type species of <i>C. costaii</i> A37T<sup>2T</sup> (=CIP 110584<sup>T</sup>, =LMG 27458<sup>T</sup>), an endophyte of <i>P. pinaster</i> affected by PWD. The genome encoded multiple genes involved in cellulolytic activity and the sequence provided insights into the role of bacteria in PWD. As there are only a few bacterial genomes related to PWD, this work provides a contribution to this field.

### Abbreviations

PWD: Pine wilt disease; PWN: Pinewood nematode

### Funding

We thank to Ana Paula Piedade for SEM analysis. This work was supported by CEMWPRE and by Fundação para a Ciência e a Tecnologia (FCT) under the project UID/EMS/00285/2013. D.N.P. was supported by FCT, postdoctoral fellowship SFRH/BPD/100721/2014. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

---

### Table 5 Unique Cluster Orthologous Groups present in the genome of <i>C. costaii</i> A37T<sup>2T</sup>

| Category Code | Category | COG ID |
|---------------|----------|--------|
| C             | Energy production and conversion | COG0280, COG0374, COG0680, COG1740 |
| E             | Amino acid transport and metabolism | COG1027, COG1586, COG2355, COG3104 |
| F             | Nucleotide transport and metabolism | COG0027 |
| G             | Carbohydrate transport and metabolism | COG0021, COG0058, COG0588, COG0662, COG0837, COG1080, COG1803, COG1925, COG2079, COG2893, COG3444, COG3716, COG3994 |
| H             | Coenzyme transport and metabolism | COG0561, COG1056, COG2091, COG2227, COG2329 |
| I             | Lipid transport and metabolism | COG0671, COG0821, COG2246 |
| J             | Translation, ribosomal structure and biogenesis | COG0060, COG0255, COG0257, COG0267, COG0268, COG0333, COG4680 |
| K             | Transcription | COG1476, COG4933 |
| L             | Replication, recombination and repair | COG0863, COG1722 |
| M             | Cell wall/membrane/envelope biogenesis | COG1083, COG1922, COG2089, COG2829, COG2982, COG3511, COG3637 |
| O             | Posttranslational modification, protein turnover, chaperones | COG0068, COG0298, COG0309, COG0409 |
| P             | Inorganic ion transport and metabolism | COG0428, COG1218, COG1230, COG1416, COG4772 |
| Q             | Secondary metabolites biosynthesis, transport and catabolism | COG2130, COG2162, COG3733, COG4242 |
| R             | General function prediction only | COG0312, COG0375, COG0429, COG0457, COG1062, COG1373, COG2320, COG3153, COG3488, COG4674, COG0561, COG2130, COG4242 |
| S             | Function unknown | COG0393, COG1286, COG2442, COG2962, COG3219, COG3247, COG3310, COG3361, COG3461, COG3477, COG3487, COG3489, COG3528, COG3548, COG3918, COG3943, COG4487, COG4700, COG4859, COG4924 |
| T             | Signal transduction mechanisms | COG0517, COG2184, COG2203, COG3292, COG1925 |
| U             | Intracellular trafficking, secretion, and vesicular transport | COG1272, COG1826, COG3451 |
| V             | Defense mechanisms | COG0286, COG0610, COG0732, COG3512, COG3513, COG4823, COG5499 |
| X             | Mobilome: prophages, transposons | COG3385, COG3436, COG3600, COG3654 |
Competing interests

The authors have no competing interests to declare.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Proença DN, Nobre MF, Morais PV. Chitinophaga costaii sp. nov., an endophyte of Pinus pinaster, and emended description of Chitinophaga niabensis. Int J Syst Evol Microbiol. 2014;64:1237–43.
2. List of prokaryotic names with standing in nomenclature. http://www. bacterio.net. Accessed 2 Dec 2016.
3. Glavina Del Rio T, Abt S, Spring S, Lapidus A, Nolan M, Tice H, et al. Complete genome sequence of Chitinophaga pinensis type strain (UQM 2034T). Stand Genom Sci. 2010;2:87–95.
4. Proença DN, Grass G, Morais PV. Understanding pine wilt disease: roles of the pine endophytic bacteria and of the bacteria carried by the disease-causing pinewood nematode. Microbiology. 2016;0:1–9.
5. Ptues e, Pepljes J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.
6. “The All-Species Living Tree” Project. http://www.arb-silva.de/projects/living-tree. Accessed 15 Jan 2016.
7. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a software environment for sequence data. Nucleic Acids Res. 2004;32:1363–71.
8. Jukes TH, Cantor CR. Evolution of protein molecules. In: Munro HN, editor. Mamm. Protein Metab. New York: Academic Press; 1969. p. 21–132.
9. Saizou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.
10. Stanmatsa K, RAV-ML VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22:2688–90.
11. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution (N Y). 1985;39:783–91.
12. Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Pillay M, et al. IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res. 2014;42:568–73.
13. Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Göker M, Parker CT, et al. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. PLoS Biol. 2014;12:1–7.
14. Whitman WB, Woyke T, Klenk H-P, Zhou Y, Lilburn TG, Beck BJ, et al. Genomic encyclopedia of bacterial and Archaeal type strains, phase III: the genomes of soil and plant-associated and newly described type strains. Stand Genomic Sci. 2015;10:26.
15. Bennett S. Soleya Ltd Pharmacogenomics. 2004;4:433–8.
16. Joint Genome Institute. http://www.jgi.doe.gov. Accessed 2 Jan 2017.
17. BBMap short read aligner, and other bioinformatic tools. http://sourceforge.net/projects/bbmap. Accessed 15 Apr 2016.
18. Bankevich A, Nurk S, Antipov D, Gurevich A A, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:45–57.