

DATA NOTE

Open Access



# First draft genome of *Thecaphora frezii*, causal agent of peanut smut disease

Renee S. Arias<sup>1\*</sup>, Cinthia Conforto<sup>2,3</sup>, Valerie A. Orner<sup>1</sup>, Edgardo J. Carloni<sup>4</sup>, Juan H. Soave<sup>5</sup>, Alicia N. Massa<sup>1</sup>, Marshall C. Lamb<sup>1</sup>, Nelson Bernardi-Lima<sup>2,3</sup> and Alejandro M. Rago<sup>3,6</sup>

## Abstract

**Objectives** The fungal pathogen *Thecaphora frezii* Carranza & Lindquist causes peanut smut, a severe disease currently endemic in Argentina. To study the ecology of *T. frezii* and to understand the mechanisms of smut resistance in peanut plants, it is crucial to know the genetics of this pathogen. The objective of this work was to isolate the pathogen and generate the first draft genome of *T. frezii* that will be the basis for analyzing its potential genetic diversity and its interaction with peanut cultivars. Our research group is working to identify peanut germplasm with smut resistance and to understand the genetics of the pathogen. Knowing the genome of *T. frezii* will help analyze potential variants of this pathogen and contribute to develop enhanced peanut germplasm with broader and long-lasting resistance.

**Data description** *Thecaphora frezii* isolate IPAVE 0401 (here referred as T.f.B7) was obtained from a single hyphal-tip culture, its DNA was sequenced using Pacific Biosciences Sequel II (PacBio) and Illumina NovaSeq6000 (Nova). Data from both sequencing platforms were combined and the de novo assembling estimated a 29.3 Mb genome size. Completeness of the genome examined using Benchmarking Universal Single-Copy Orthologs (BUSCO) showed the assembly had 84.6% of the 758 genes in fungi\_odb10.

**Keywords** PacBio, Genome, Pathogen, Groundnut, Smut disease, Fungi

## Objective

Peanut smut disease converts the peanut seed into a brown/black powder of fungal teliospores, example shown in Data File 1, Table 1 [1]. The disease is currently endemic to Argentina [2], where the pathogen *Thecaphora frezii* [1] has been reported to cause up to 51% disease incidence [3] and crop losses up to 35% [2, 4]. Global trade could potentially spread the disease to other areas. For example, United States is a major exporter of peanuts [5], but it also imports peanuts from other countries [6] including Argentina [7]. Scientists from the U.S. and Argentina are working in collaboration to better understand the disease and to identify resistant peanut germplasm [8–11]. Part of that effort is to know the genetics of the pathogen. Previously, we have sequenced the 123 kb mitochondrial genome of *T. frezii* based on DNA obtained from teliospores [12]. Other than that, the

\*Correspondence:

Renee S. Arias  
renee.arias@usda.gov

<sup>1</sup> USDA-ARS National Peanut Research Laboratory (NPRL), Dawson, GA, USA

<sup>2</sup> Instituto de Patología Vegetal, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria, IPAVE-CIAP-INTA, X5020 Córdoba, Argentina

<sup>3</sup> Unidad de Fitopatología y Modelización Agrícola, Consejo Nacional de Investigaciones Científicas y Técnicas, UFVMA-CONICET, X5020 Córdoba, Argentina

<sup>4</sup> Instituto de Fisiología y Recursos Genéticos Vegetales, Centro de Investigaciones Agropecuarias, Instituto Nacional de Tecnología Agropecuaria, IFRG-CIAP-INTA, X5020 Córdoba, Argentina

<sup>5</sup> El Carmen S.A., General Cabrera, Córdoba, Argentina

<sup>6</sup> Centro de Investigaciones Agropecuarias - INTA, Córdoba, Argentina



This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023. **Open**

**Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Table 1** Overview of data files/data sets

Label	Name of data file/data set	File types (File extension)	Data repository and identifier (DOI or accession number)
Data file 1	Peanut smut in shell <i>Thecaphora frezii</i> culture	Tag Image File Format (.tiff)	<a href="https://doi.org/10.7910/DVN/EOYLUX">https://doi.org/10.7910/DVN/EOYLUX</a> [1] at Harvard Dataverse
Data file 2	QC_140bp	Excel (.xlsx)	<a href="https://doi.org/10.7910/DVN/I338KA">https://doi.org/10.7910/DVN/I338KA</a> [13] at Harvard Dataverse
Data file 3	Contig_Length_& Cumulative Length_TF	Portable Document Format (.PDF)	<a href="https://doi.org/10.7910/DVN/SOFHRR">https://doi.org/10.7910/DVN/SOFHRR</a> [14] at Harvard Dataverse
Data set 1	B7_S113_L004_R1_001 B7_S113_L004_R2_001	Illumina NovaSeq 6000 data; Fastq file (.fastq.gz)	SRR18840655 [15] at NCBI GenBank Database
Data set 2	M64069_211005_200549	PacBio, Binary alignment map (.BAM)	SRR18837637 [16] at NCBI GenBank Database
Data set 3	<i>T. frezii</i> Genome Assembly: 7166_Contigs_TF_B7_Genome	Fasta (.fasta)	JALNIF000000000 [17] at NCBI GeneBank Database

National Center for Biotechnology Information (NCBI) has only 170 entries for this species, those sequences are under 2,500 bp, and more than half correspond to microsatellites and ribosomal RNA. We have found that working with teliospores had its own constraints, thus, the current work aimed to sequence the genome of a hyphal tip culture of this fungus. The isolate referred as T.f.B7 was obtained from peanuts in Argentina and is now stored in the IPAVE culture collection as IPAVE 0401, both, the hyphal tip culture and the original teliospores are kept in the collection. Overall, BLAST analysis performed on assembled *T. frezii* contigs showed highest similarity (~78%) to *Anthracozygia flocculosa* (Syn. *Pseudozyma flocculosa*), followed by (~76%) similarity to *Sporisorium* spp. and lower similarity (~73%) to *Ustilago* spp. Given the limited information available about this pathogen, the data are very useful.

#### Data description

*Thecaphora frezii* isolate IPAVE 0401 (here referred as T.f.B7), was obtained as a hyphal culture from teliospores of a smut-affected peanut plant collected in 2018 from Hernando city, in Tercero Arriba, Cordoba (32° 24' 30.5028" S, 63° 42' 18.9468" W). Genomic DNA was extracted from T.f.B7, a culture expected to be haploid monokaryotic as shown by HCL-Giemsa and Propidium iodide nuclear staining, Data File 1, Table 1 [1], and was sequenced as paired end 150 base pairs (bp) using Illumina NovaSeq6000 which resulted in 114,541,382 reads, Data set 1, Table 1 [15]. After trimming for potential presence of adapters and removal of sequences shorter than 140 bp, 112,980,831 clean reads were available. A summary of read length and quality is listed in Data File 2, Table 1 [13]. A second DNA extraction of isolate T.f.B7 was processed following HiFi low DNA input library preparation with Single Molecule Real Time (SMRT) bell Express Template Prep Kit 2.0 and sequenced using

PacBio Sequel II (Pacific Biosciences, San Diego, CA) at the Genome Center, University of California Davis, CA. This generated 1,201,967 subreads, Data set 2, Table 1 [16], with average length 4,086 bases, N50=6,011, N90=2,669, and HiFi read count=17,988. Nova reads were mapped to the 123 kb mitochondrial DNA of *T. frezii* [12], the mapping had a 119,373 X coverage and the assembled consensus contig had two single nucleotide polymorphisms (SNPs) compared to the published mitogenome of *T. frezii*, contig\_7166\_TF\_mitochondrion Data set 3, Table 1 [17]. Nova reads that did not map to the mitochondrial DNA comprised 2.3 Gb that were de novo assembled using CLC Workbench. For the assembly, corrected PacBio Subreads were used as “guidance only”, application where guidance only reads are not used to create the *de Bruijn* graph but to resolve ambiguities in the graph. This resulted in 7,165 contigs with average coverage 80 X, and an estimated genome size of 29,333,160 bp when adding the mitochondrial DNA (contig number 7,166), Data set 3 [17] and Data file 3, Table 1 [14]. The assembly was performed with word size 23, bubble size 50, and resulted in N75=3,838, N50=9,027, N25=16,027, max length=60,190 bp, min length=377 bp, average=4,094 bp. As a comparison, a genome size of the related species, *Thecaphora thlaspeos*, is 20,591,600 bp [18]. The 9,012 bp Contig\_22 [17] contains the ribosomal RNA (rRNA) cistron; alignment of the 615 bp partial rRNA sequence of *T. frezii* (Sa-EM1) accession JX041638.1 [19] to Contig\_22 showed 100% identity. Further assessment of genome completeness was done using BUSCO [20] v.5.2.2, with the fungi\_odb10 database, which consists of 549 genomes ([www.orthodb.org](http://www.orthodb.org)). The results showed 641 of the 758 BUSCO genes (84.6%) were complete, with 632 complete single copy (83.4%), 9 complete duplicates (1.2%), 48 genes were fragmented (6.3%), and 69 BUSCO genes were missing (9.1%). Mapping and assembly were performed

in CLC\_Genomics Workbench 20.0.4 (Qiagen, Aarhus, Denmark), using CLC Genome Finishing Module for processing PacBio data. The data were deposited in NCBI, Bioproject PRJNA828173, Biosample SAMN27642199, Data sets 1, 2 and 3, Table 1 [15–17].

### Limitations

*Thecaphora frezii* isolation and DNA extraction proved very challenging since teliospores after germination form a very thin layer of budding-yeast phase, without aerial mycelium. A first whole genome sequencing of *T. frezii* as paired end 150 bases using NovaSeq 6000 and de novo assembly resulted in a relatively fragmented genome. Additional sequencing of a small amount of DNA of suboptimal quality was performed by PacBio resulting in 95% of reads shorter than 10 Kb, thus, the reads were corrected to keep only long sequences that were supported by multiple reads. A de novo assembly of data obtained from both platforms combined was still fragmented. Since the genetic information of *T. frezii* is very limited, the draft genome we obtained by combining both sequencing platforms and reported here, will allow mining for genes of interest, and perform studies on genetic diversity of this pathogen.

### Abbreviations

Nova	Illumina NovaSeq 6000
PacBio	Pacific Biosciences Sequel II
NCBI	National Center for Biotechnology Information
IPAVE	Instituto de Patología Vegetal
SMRT	Single Molecule Real Time
SNP	Single Nucleotide Polymorphism

### Acknowledgements

Not applicable.

### Authors' contributions

RSA, AMR, JHS, MCL: Conceptualization. JHS, MCL, BES, AMR: Funding acquisition and resources. RSA, VAO: Formal analysis, data curation and uploading. CEC, NBL: Fungal isolation, DNA extractions, photographs. RSA: Writing original draft. EJC: nuclear staining, microscopy. RSA, CEC, AMR, ANM, JHS, MCL: Review and editing. All authors reviewed and approved the manuscript.

### Authors' information

Not applicable.

### Funding

This work was funded by the United States Department of Agriculture, Agricultural Research Service, Project Number: 6604–21000-005-00D; and by Instituto Nacional de Tecnología Agropecuaria (INTA), Córdoba, Argentina. The funding bodies played no role in the design of the study, collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

The data described in this Data note can be freely and openly accessed on NCBI Bioproject PRJNA828173, Biosample SAMN27642199 (Data sets 1–3), Harvard Dataverse under <https://doi.org/10.7910/DVN/EOYLUX>; <https://doi.org/10.7910/DVN/I338KA> (Data files 1–3); and NCBI Accession numbers: SRR18840655, SRR18837637, JALNIF000000000. Please see Table 1 and references [1, 13, 14, 15, 16, 17] for details and links to the data.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare having no competing interest.

Received: 17 May 2022 Accepted: 9 February 2023

Published online: 16 February 2023

## References

1. Data file 1: Photograph of peanut smut disease and *Thecaphora frezii* culture, by Conforto EC. Harvard Dataverse. (2022). <https://doi.org/10.7910/DVN/EOYLUX>.
2. Rago AM, Cazón LI, Paredes JA, Molina JPE, Conforto EC, Bisonard EM, Oddino C. Peanut smut: from an emerging disease to an actual threat to Argentine peanut production. *Plant Dis.* 2017;101(3):400–8.
3. Oddino C, Marinelli A, March G, Garcia JDJ, Tarditi L, D'Eramo L, Ferrari S. Relation between potential inoculum of *Thecaphora frezii*, smut disease intensity and crop yield. Córdoba: XXV National Peanut Congress; 2010. 1–2 (#6).
4. Paredes JA, Cazón LI, Osella A, Peralta V, Alcalde M, Kearney MI, Zuza MS, Rago AM, Oddino C. Regional survey of peanut smut and estimated losses caused by the disease. Córdoba: XXXI National Peanut Congress; 2016. 21–2 (#21).
5. American Peanut Council, United States Exports to World [[https://www.peanutsusa.com/phocadownload/Export\\_Statistics/USAnnualExports/Thru2020/U.S.%20Annual%20Exports-%20All%20Products.pdf](https://www.peanutsusa.com/phocadownload/Export_Statistics/USAnnualExports/Thru2020/U.S.%20Annual%20Exports-%20All%20Products.pdf)]
6. American Peanut Council, Global Imports from World [[https://www.peanutsusa.com/phocadownload/Export\\_Statistics/GlobalImports/World%20Peanut%20Imports%202015-2020.pdf](https://www.peanutsusa.com/phocadownload/Export_Statistics/GlobalImports/World%20Peanut%20Imports%202015-2020.pdf)]
7. Arias SL, Mary VS, Velez PA, Rodriguez MG, Otaiza-Gonzalez SN, Theumer MG. Where does the peanut smut pathogen, *Thecaphora frezii*, fit in the spectrum of smut diseases? *Plant Dis.* 2021;105(9):2268–80.
8. de Blas FJ, Bruno CI, Arias RS, Ballen-Taborda C, Mamani E, Oddino C, Rosso M, Costero BP, Bressano M, Soave JH, et al. Genetic mapping and QTL analysis for peanut smut resistance. *BMC Plant Biol.* 2021;21(1):312.
9. Massa AN, Bressano M, Soave JH, Buteler MI, Seijo G, Sobolev VS, Orner VA, Oddino C, Soave SJ, Faustinielli PC, et al. Genotyping tools and resources to assess peanut germplasm: smut-resistant landraces as a case study. *PeerJ.* 2021;9:e10581.
10. Bressano M, Massa AN, Arias RS, de Blas F, Oddino C, Faustinielli PC, Soave S, Soave JH, Perez MA, Sobolev VS, et al. Introgression of peanut smut resistance from landraces to elite peanut cultivars (*Arachis hypogaea* L.). *Plos One.* 2019;14(2):e0211920.
11. de Blas FJ, Bressano M, Teich I, Balzarini MG, Arias RS, Manifesto MM, Costero BP, Oddino C, Soave SJ, Soave JA, et al. Identification of smut resistance in wild *Arachis* species and its introgression into peanut elite lines. *Crop Sci.* 2019;59(4):1657–65.
12. Arias RS, Cazon LI, Massa AN, Scheffler BE, Sobolev VS, Lamb MC, Duke MV, Simpson SA, Conforto C, Paredes JA, et al. Mitogenome and nuclear-encoded fungicide-target genes of *Thecaphora frezii* - causal agent of peanut smut. *Fungal Genom Biol.* 2019;9(1):160–8. <https://www.longdom.org/archive/fgb-volume-9-issue-1-year-2019.html>
13. Data file 2: Length and quality of Illumina sequencing reads. Harvard Dataverse. (2022). <https://doi.org/10.7910/DVN/I338KA>.
14. Data file 3: Contig length and cumulative length of assembled *Thecaphora frezii* genome. Harvard Dataverse. (2022). <https://doi.org/10.7910/DVN/SOFHRR>.
15. National Center for Biotechnology Information. Illumina sequencing reads of *Thecaphora frezii*. Sequence Read Archive. SRR18840655. (2022). <https://www.ncbi.nlm.nih.gov/sra/SRR18840655>.

16. National Center for Biotechnology Information. PacBio sequencing reads of *Thecaphora frezii*. Sequence Read Archive: SRR18837637. (2022). <https://www.ncbi.nlm.nih.gov/sra/SRR18837637>.
17. National Center for Biotechnology Information. Genome Assembly Bio-project PRJNA828173. (2022). <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA828173>.
18. Courville KJ, Frantzeskakis L, Gul S, Haeger N, Kellner R, Hessler N, Day B, Usadel B, Gupta YK, van Esse HP, et al. Smut infection of perennial hosts: the genome and the transcriptome of the Brassicaceae smut fungus *Thecaphora thlaspeos* reveal functionally conserved and novel effectors. *New Phytol.* 2019;222(3):1474–92.
19. Conforto C, Cazón I, Fernández FD, Marinelli A, Oddino C, Rago AM. Molecular sequence data of *Thecaphora frezii* affecting peanut crops in Argentina. *Eur J Plant Pathol.* 2013;137(4):663–6.
20. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2.

### Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

