RESEARCH NOTE

REVISED Draft genomes of two Australian strains of the plant pathogen, *Phytophthora cinnamomi* [version 2; referees: 3 approved]

Amy L. Longmuir[1], Peter L. Beech[1], Mark F. Richardson [2,3]

[1]Centre for Cellular and Molecular Biology, School of Life and Environmental Sciences, Deakin University, Burwood, VIC, 3125, Australia
[2]Bioinformatics Core Research Group, School of Life and Environmental Sciences, Deakin University, Geelong, VIC, 3220, Australia
[3]Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, VIC, 3220, Australia

v2 **First published:** 08 Nov 2017, **6**:1972 (doi: 10.12688/f1000research.12867.1)
Latest published: 28 Feb 2018, **6**:1972 (doi: 10.12688/f1000research.12867.2)

Abstract

Background: The oomycete plant pathogen, *Phytophthora cinnamomi*, is responsible for the destruction of thousands of species of native Australian plants, as well as several crops, such as avocado and macadamia, and has one of the widest host-plant ranges of the *Phytophthora* genus. The current reference genome of *P. cinnamomi* is based on an atypical strain and has large gaps in its assembly. To further studies of the pathogenicity of this species, especially in Australia, robust genome assemblies of more typical strains are required. Here we report the genome sequencing, draft assembly, and preliminary annotation of two geographically separated Australian strains of *P. cinnamomi*.

Findings: Some 308 million raw reads were generated for the two strains, DU054 and WA94.26. Independent genome assembly produced final genome sequences of 62.8 Mb (in 14,268 scaffolds) and 68.1 Mb (in 10,084 scaffolds), which are comparable in size and contiguity to other *Phytophthora* genomes. Gene prediction yielded > 22,000 predicted protein-encoding genes within each genome, while BUSCO assessment showed 94.4% and 91.5% of the stramenopile single-copy orthologs to be present in the assembled genomes, respectively.

Conclusions: The assembled genomes of two geographically distant isolates of *Phytophthora cinnamomi* will provide a valuable resource for further comparative analyses and evolutionary studies of this destructive pathogen, and further annotation of the presented genomes may yield possible targets for novel pathogen control methods.

GODAN Global Open Data for Agriculture & Nutrition

This article is included in the Global Open Data for Agriculture and Nutrition gateway.

Open Peer Review

Referee Status: ✔ ✔ ✔

	Invited Referees 1	2	3
REVISED version 2 published 28 Feb 2018			✔ report
version 1 published 08 Nov 2017	✔ report	✔ report	? report

1 **Nicolás Daniel Ayub**, National Scientific and Technical Research Council, Argentina

2 **Erik Andreasson**, Swedish University of Agricultural Sciences, Sweden
Laura Grenville Briggs, Swedish University of Agricultural Sciences, Sweden

3 **David J. Studholme**, University of Exeter, UK

Discuss this article

Comments (0)

Corresponding author: Mark F. Richardson (m.richardson@deakin.edu.au)

Author roles: Longmuir AL: Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation; **Beech PL**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Richardson MF**: Conceptualization, Data Curation, Formal Analysis, Methodology, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Longmuir AL, Beech PL and Richardson MF. **Draft genomes of two Australian strains of the plant pathogen, *Phytophthora cinnamomi* [version 2; referees: 3 approved]** *F1000Research* 2018, **6**:1972 (doi: 10.12688/f1000research.12867.2)

Grant information: The author(s) declared that no grants were involved in supporting this work.

First published: 08 Nov 2017, **6**:1972 (doi: 10.12688/f1000research.12867.1)

REVISED **Amendments from Version 1**

Updated corresponding author details; added an additional statement and reference highlighting the availability of two other *Phytophthora cinnamomi* genome assemblies; added sequencing library and paired-end read information; updated BUSCO analysis and included comparisons to previous *P. cinnamomi* assemblies; made preliminary gene predictions publically available; made all minor additional changes requested by reviewers.

See referee reports

Introduction

Phytophthora cinnamomi is a highly virulent plant pathogen that has a devastating impact on the Australian ecosystem, namely in the south-western areas of Western Australia and much of the south and east coasts of Victoria and New South Wales[1]. In the south west ecoregion of Western Australia, alone, over 40% of the 5710 plant species present have been shown to be susceptible to *P. cinnamomi*[2]. Significant genetic and phenotypic variation can occur within a signal clonal linage of *P. cinnamomi*[3] and susceptibility of a given host plant species has been shown to vary from site to site[4]. Furthermore, despite the general lack of crossing during sexual reproduction, *P. cinnamomi* excels at adapting to new environments and developing virulence to new host species through asexual growth, making it a deadly and difficult-to-control pathogen. Unravelling how *P. cinnamomi* is able to adapt so quickly, and remain virulent, to a wide range of hosts in Australia, is an important research goal.

Currently, three *P. cinnamomi* strains have genome assemblies (MP94.48 and NZFS375, see 5 and Joint Genome Institute (JGI); NCBI Accession no. PRJNA68241). However, only the genome of *P. cinnamomi* var. *cinnamomi* (JGI; NCBI Accession no. PRJNA68241) has a publically available annotation, serving as the species reference genome. The assembly is based on the Rands isolate from Sumatra in 1922, which has been in culture for many decades and may not be representative of the current pathogenic strains present in Australia. Here we report and make available two Australian *P. cinnamomi* genomes, isolated from geographically very separate areas with different available host species. After analyses of genetic differences between these two *P. cinnamomi* genomes, it may be that key genes or gene families under high evolutionary pressure can be identified; this may aid further studies on more effective control of this pathogen.

Sample collection and sequencing

Two isolates of *P. cinnamomi* were selected from areas of infection on either side of the Australian continent: one from the Brisbane Ranges in southeastern Australia (DU054, A2 mating type)[6] and the other from southwestern Western Australia (WA94.26, A2 mating type), both Deakin University culture collection. These isolates were selected to represent possible genetic diversity of *P. cinnamomi* in Australia arising from geographic isolation, and possible variation of selective pressures due to different host species. Isolates were maintained on V8 agar (V8A) [50 ml unclarified V8 'Original' Juice (Campbells, Australia), 0.5 g $CaCO_3$and 7.5 g biological agar per 500 mL of distilled water] at 25°C in darkness, as per 6. Genomic DNA was isolated from hyphae using a DNeasy Plant Mini Kit (Qiagen), following the manufacturers' protocol. Illumina TruSeq Nano library preparation (one per isolate) and sequencing on an Illumina HiSeq 2500 platform were performed by the Australian Genome Resources Facility (Walter and Eliza Hall Institute, Parkville, Australia) generating ~154 million paired-end (2 × 150 bp) raw reads per isolate. Raw reads are available in the NCBI Short Read Archive (SRA) under the Bioproject Accession: PRJNA413098.

Genome assembly

Raw sequencing data for each isolate were first pre-processed using Trimmomatic v0.33[7] with the following parameter values: *ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:4 AVGQUAL:30 MINLEN:36*, to remove Illumina adapters and filter reads based on quality scores (Phred score). Only reads with average Phred > 30 were retained. To ensure only the desired *P. cinnamomi* genomes were assembled, a second round of pre-processing was conducted to remove potential contaminants. MetaPhlAn v2[8], was run with default settings and identified the *Paenibacillus* genus as the only likely bacterial contaminate. Using BBMap v0.35 (BBMap - Bushnell. B), we mapped the Trimmomatic-filtered reads to the closest species match (*Paenibacillus* sp., JDR-2, GenBank accession: GCA_000023585.1, with 2.7% and 2.0% of DU054 and WA94.26 reads mapping, respectively; these *Paenibacillus* reads were subsequently removed. The remaining reads were then mapped using BBMap to the human genome (GRCh38; NCBI accession: GCA_000001405.15), with < 0.5% (~ 430,000 reads from DU054 and ~ 630,000 from WA94.26) being mapped and subsequently removed from the data set. Thus, the final set of reads (DU054, 149 million reads; WA94.26, 151 million reads) used for the assembly contained high-quality paired-end reads not belonging to either human or bacterial contaminants.

De novo contig assembly of the two genomes was conducted independently, using IDBA-UD v1.1.0[9]. IDBA-UD was run using the following parameter values: *--mink 20 --maxk 100 --step 20 --min_contig 500 --min_support 2 --min_count 3*. Briefly, these conducted a multiple K-mer assembly from k = 20 up to k = 100; only assembled contigs above 500 bp and those with a minimum depth coverage ≥ 3 were kept. As heterogeneous data can increase redundancy in genome assemblies (through heterozygous regions being assembled as separate contigs that results in highly fragmented assemblies[10]), the IDBA-UD assembled contigs were run through the Redundans pipeline v0.12c[10] with the following parameter values: *-threads 4 -min_length 500*. Redundans uses paired-end mapping data to reduce assembled sequence redundancy and scaffold contigs into longer less fragmented sequences. The final assembled genome sequence of DU054 was 62.80 Mb in 14,269 scaffolds with an N50 of 9,951 bp; the longest scaffold was 1.54 Mb in length (Table 1). For WA94.26, the final genome sequence was 68.07 Mb in length, in 10,085 scaffolds with the largest being 1.54 Mb and an N50 of 20,813 bp. GC content remained consistent, at ~ 53%, between both isolate genomes across both assemblies and before and after processing with Redundans. The quality, as

measured by the above metrics, of the presented genomes is comparable to the previously available *P. cinnamomi* var. *cinnamomi* Rands isolate genome (JGI). The final genome assemblies are available under the NCBI Bioproject Accession: PRJNA413098.

We used the BUSCO (benchmarking universal single-copy orthologs) pipeline v3.02[11] in genome mode, with the default e-value cutoff of 0.01, to assess the completeness of the assembled genomes and compared the results to the previously available Rands isolate and the *P. cinnamomi* assemblies from Studholme *et al.*[5]. Utilizing the set of 234 conserved stramenopile single-copy orthologs (hereafter BUSCOs), the analysis indicated 94.4% and 91.5% BUSCO completeness for the DU054 and WA94.26 genomes, respectively. For DU054, 221 complete BUSCOs (all single-copy with no duplicated BUSCOs) and 3 fragmented BUSCOs were identified, and 214 complete and 2 fragmented BUSCOs in WA94.26 (Table 2). Overall, we find a higher level of BUSCO completeness compared with the Rands isolate, and comparable (albeit it slightly lower) completeness compared to the two *P. cinnamomi* assemblies from Studholme *et al.*[5] (Table 2). This suggests our two Australian isolate assemblies are as complete references as those currently available.

Preliminary genome annotation

We conducted a preliminary protein-coding sequence prediction using GeneMark-ES v4.32[12], which utilises a self-training algorithm to identify exon, intron and intergenic regions as well as initiation and termination sites. GeneMark-ES was run using the default settings and a database of predicted gene models (i.e., predicted polypeptides) was constructed for DU054 and WA94.26 genomes (available in the associated data repository[13]). An initial 23,414 gene models were identified in DU054 and 22,573 in WA94.26. Of these, 14,735 pairs of predicted gene models appear to be orthologous between the two genomes (reciprocal best-hit Blastp, e value ≤ 1e-5). As a preliminary verification of these gene model builds, we identified orthologous counterparts to eight available *Phytophthora* genomes with annotations [*P. infestans*[14], *P. kernoviae*[15], *P. lateralis*[16], *P. nicotianae*[17], *P. parasitica* (P1569_v1; Broad Institute), *P. ramorum*[18], *P. sojae*[18] and *P. cinnamomi* var. *cinnamomi*]. Accordingly, we used OrthoFinder v1.1.10[19] with default parameter values, except we used DIAMOND[20] as the alignment program with the *diamond_more_sensitive* flag. OrthoFinder first identifies 'orthogroups' (an extension of orthologues to include groups of genes descended from a single gene in the last common ancestor of a group of species[19]) and then orthologues between each pair of species in the comparison. OrthoFinder assigned 88.5% (170,769) of the genes found in all the species to 19,089 orthogroups, and of these 50% of all the genes were contained in orthogroups, which had 10 or more genes within them. We found 2,931 orthogroups that contained genes for each of the species, and of these 1,309 orthogroups consisted entirely of single-copy genes; see associated data repository[13]. Using these single-copy orthogroups, gene trees were first constructed, then the species tree was inferred using the distance-based method implemented by fastme[21]. The resultant species tree (see associated data repository[13]) exhibits strong

Table 1. Summary of genomic features of assembled genomes comparing IDBA-UD output to scaffolded genome after Redundans processing and the *P. cinnamomi* Rands isolate genome.

	DU054		WA94.26	
	IDBA-UD	Redundans	IDBA-UD	Redundans
Assemblysize (Mb)	71.29	62.80	76.95	68.07
No. scaffolds	33,475	14,268	36,333	10,084
N50 (bp)	4,085	9,951	4,075	20,813
No. predicted genes	NA	23,414	NA	22,573

Table 2. Summary of BUSCO assessment.

	DU054	WA94.26	*P. cinnamomi* var. *cinnamomi*	MP94.48[5]	NZFS375[5]
Total BUSCOs	234	234	234	234	234
Complete and single copy BUSCOs	221 (94.4%)	214 (91.5%)	202 (86.3%)	228 (97.4%)	228 (97.4%)
Complete and duplicate BUSCOs	0 (0%)	0 (0%)	4 (1.7%)	0 (0%)	0 (0%)
Fragmented BUSCOs	3 (1.3%)	2 (0.9%)	7 (3.0%)	2 (0.9%)	2 (0.9%)
Missing BUSCOs	10 (4.3%)	18 (7.6%)	21 (9.0%)	4 (1.7%)	4 (1.7%)

congruence to the *Phytophthora* phylogeny recently published by 22, providing more evidence that the genome assembly and preliminary annotation conducted here is valuable.

Conclusions

In summary, we present the genome assembly of two geographically separated isolates of *Phytophthora cinnamomi* from Australia. These high-quality genome assemblies may act as a valuable resource for comparative genomics and particularly for the further identification and analysis of protein-encoding genes expressed during plant infection, such as members of the avirulence gene families[23]. These gene families are of specific interest in the development of novel and effective pathogen control mechanisms.

Data availability

Raw reads are available in the NCBI SRA under the Bioproject Accession: PRJNA413098.

The final assemblies are available at DDBJ/EMBL/GenBank under the accessions, PDCY00000000 and PDCZ00000000 and under the Bioproject Accession: PRJNA413098.

Supporting data, including preliminary gene prediction, OrthoFinder analysis and BUSCO assessment results can be found in the associated data repository: doi, 10.4225/16/59d15a6917a5e[20].

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

References

1. Cahill DM, Rookes JE, Wilson BA, *et al.*: ***Phytophthora cinnamomi* and Australia's biodiversity: impacts, predictions and progress towards control.** *Aust J Bot.* 2008; **56**(4): 279–310.
Publisher Full Text
2. Shearer BL, Crane CE, Barrett S, *et al.*: ***Phytophthora cinnamomi* invasion, a major threatening process to conservation of flora diversity in the South-west Botanical Province of Western Australia.** *Aust J Bot.* 2007; **55**(3): 225–238.
Publisher Full Text
3. Hüberli D, Tommerup IC, Dobrowolski MP, *et al.*: **Phenotypic variation in a clonal lineage of two *Phytophthora cinnamomi* populations from Western Australia.** *Mycol Res.* 2001; **105**(9): 1053–1064.
Publisher Full Text
4. Shearer BL, Dillon M: **Susceptibility of plant species in *Banksia* woodlands on the Swan Coastal Plain, Western Australia, to infection by *Phytophthora cinnamomi.*** *Aust J Bot.* 1996; **44**(4): 433–445.
Publisher Full Text
5. Studholme DL, McDougal RL, Sambles C, *et al.*: **Genome sequences of *six Phytophthora* species associated with forests in New Zealand.** *Genom Data.* 2016; **7**: 54–56.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Rookes JE, Wright ML, Cahill DM: **Elucidation of defence responses and signalling pathways induced in *Arabidopsis thaliana* following challenge with *Phytophthora cinnamomi*.** *Physiol Mol Plant P.* 2008; **72**(4–6): 151–161.
Publisher Full Text
7. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics.* 2014; **30**(15): 2114–20.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Segata N, Waldron L, Ballarini A, *et al.*: **Metagenomic microbial community profiling using unique clade-specific marker genes.** *Nat Methods.* 2012; **9**(8): 811–814.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Peng Y, Leung HC, Yiu SM, *et al.*: **IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth.** *Bioinformatics.* 2012; **28**(11): 1420–1428.
PubMed Abstract | Publisher Full Text
10. Pryszcz LP, Gabaldón T: **Redundans: an assembly pipeline for highly heterozygous genomes.** *Nucleic Acids Res.* 2016; **44**(12): e113.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212.
PubMed Abstract | Publisher Full Text
12. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, *et al.*: **Gene identification in novel eukaryotic genomes by self-training algorithm.** *Nucleic Acids Res.* 2005; **33**(20): 6494–6506.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Longmuir A, Beech P, Richardson M: **Supporting data for "Draft Genomes of two Australian strains of the plant pathogen, *Phytophthora cinnamomi*"**. [data collection], 2017.
Publisher Full Text
14. Haas BJ, Kamoun S, Zody MC, *et al.*: **Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*.** *Nature.* 2009; **461**(7262): 393–398.
PubMed Abstract | Publisher Full Text
15. Sambles C, Schlenzig A, O'Neill P, *et al.*: **Draft genome sequences of *Phytophthora kernoviae* and *Phytophthora ramorum* lineage EU2 from Scotland.** *Genom Data.* 2015; **6**: 193–194.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Quinn L, O'Neill PA, Harrison J, *et al.*: **Genome-wide sequencing of *Phytophthora lateralis* reveals genetic variation among isolates from Lawson cypress (*Chamaecyparis lawsoniana*) in Northern Ireland.** *FEMS Microbiol Lett.* 2013; **344**(2): 179–185.
PubMed Abstract | Publisher Full Text
17. Liu H, Ma X, Yu H, *et al.*: **Genomes and virulence difference between two physiological races of *Phytophthora nicotianae*.** *Gigascience.* 2016; **5**(1): 3.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Tyler BM, Tripathy S, Zhang X, *et al.*: ***Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis.** *Science.* 2006; **313**(5791): 1261–1266.
PubMed Abstract | Publisher Full Text
19. Emms DM, Kelly S: **OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy.** *Genome Biol.* 2015; **16**(1): 157.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using DIAMOND.** *Nat Methods.* 2015; **12**(1): 59–60.
PubMed Abstract | Publisher Full Text
21. Lefort V, Desper R, Gascuel O: **FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program.** *Mol Biol Evol.* 2015; **32**(10): 2798–800.
PubMed Abstract | Publisher Full Text | Free Full Text
22. McCarthy CGP, Fitzpatrick DA: **Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes.** *mSphere.* 2017; **2**(2): pii: e00095-17.
PubMed Abstract | Publisher Full Text | Free Full Text
23. Bos JI, Kanneganti TD, Young C: **The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*.** *Plant J.* 2006; **48**(2): 165–176.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Referee Status: ✔ ✔ ✔

Version 2

Referee Report 08 March 2018

doi:10.5256/f1000research.14404.r31293

✔ **David J. Studholme**
Biosciences, University of Exeter, Exeter, UK

The authors have satisfactorily dealt with all the issues raised in my previous report.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 28 November 2017

doi:10.5256/f1000research.13945.r28065

? **David J. Studholme**
Biosciences, University of Exeter, Exeter, UK

This manuscript announces the availability of genomic sequence data from Phytophthora cinnammomi strains DU054 and WA94.26. This is a useful resource for researchers interested in this important pathogen. The authors have deposited and made available the raw sequence data in SRA and their assemblies in GenBank, which is commendable. The genome annotation and protein sequences do not appear to be deposited in GenBank, however. This does not preclude publication, but when discussing predicted genes in the manuscript, the authors should be up-front about this or provide full details of the annotations in supplementary data or deposit them in some public repository.

One oversight that the authors should be aware of is the previous publication of two genome sequences of this species, one of which (MP94-48) is from Australia. See Studholme *et al*. (2015)[1]. So, the authors' assertion (in their Conclusions section) that this is the first genome assembly from an Australian strain should be revised. The authors should also include those two assemblies in their comparisons of assembly quality metrics. And also please revise the several other mentions of previously sequenced genome throughout the text in the light of the additional two previously sequenced genomes. Also, it would be interesting to assess how similar or different all these four available P.c. genome sequences are to each other, e.g. by calculating pairwise ANIs.

Some specific points that should be addressed around the methodology:

1. Why were reads mapped against the human genome? Why should contamination from human DNA be more prevalent or likely than from other organisms?

2. The authors make good efforts to remove contaminating Paenibacillus sequence reads. Interestingly, we also observed contamination of Phytophthora genomic DNA with this bacterial genus. However, the authors go on to claim that the data contained "highly quality reads not belonging to ... bacterial contaminants". Their approach does not remove non-Paenibacillus bacterial contaminants.

3. Please cite a reference to support the claim that "heterogeneous data can increase redundancy in genome assemblies". It is not entirely clear what this statement means, precisely, and in any case it is not self-evident and needs to be supported by peer-reviewed publication.

4. The use of BUSCO version 1.22 is questionable, given that versions 2 and 3 are now available. Furthermore, rather than using the general Eukaryote set of BUSCOs, the authors should use the Stramenopile set.

5. The completeness of the genome assemblies is rather poor (only < 65% of expected genes are present intact in a single copy). It would be useful to compare/benchmark this against other available Phytophthora genome sequences. For example, our recent sequencing of P. ramorum genomes, we found around 81- 85% of Stramenopile BUSCOs were intact and single-copy in each genome (See PubMed ID 28243575).

6. Towards the end of page 4, the authors claim that the "preliminary annotation ... is valuable". I agree and would go further to say that not just the annotation but the genome sequencing per se is valuable. I would also suggest including a brief explanation of how/why the presented data is valuable.

7. The authors say that their annotation is valuable, but the annotation has not apparently been deposited in a public repository. Therefore, please either make this valuable resource available, or remove the claim that it is valuable.

Some very minor points:

1. In the Introduction, it was not obvious to me what is meant by a "Botanical Province". Please consider explaining this term.

2. Please add an apostrophe after "manufacturers".

3. At several places in the text, the authors write "parameters" when they really mean "parameter values" or "options" or "switches". Please check and revise.

4. Please write "high-quality" not "highly quality".

5. On page 3, the authors say that no gene expression data are available for this species. This is untrue, since EST data (i.e. expressed sequence tags) are available. Furthermore, in the SRA, there are several RNAseq datasets available:

Illumina HiSeq 2000 sequencing; RNAseq analysis of germinating cysts of Phytophthora cinnamomi 1 ILLUMINA (Illumina HiSeq 2000) run: 46,420 spots, 4.2M bases, 3.5Mb downloads Accession: ERX709652 Select item 14623972.

Illumina HiSeq 2000 paired end sequencing 1 ILLUMINA (Illumina HiSeq 2000) run: 9.9M spots, 1.8G bases, 1.1Gb downloads Accession: ERX943317 Select item 1426113.

Phytophthora cinnamomi library 1 ILLUMINA (Illumina HiSeq 2000) run: 88.1M spots, 17.6G bases, 10.3Gb downloads Accession: SRX124562 Select item 1426104.

Phytophthora cinnamomi library 1 ILLUMINA (Illumina HiSeq 2000) run: 30,453 spots, 6.1M bases, 2.6Mb downloads Accession: SRX124561 Select item 1426095.

Phytophthora cinnamomi library1 ILLUMINA (Illumina HiSeq 2000) run: 50.6M spots, 10.1G bases, 5.9Gb downloads Accession: SRX124560 Select item 1426086.

Phytophthora cinnamomi library 2 ILLUMINA (Illumina HiSeq 2000) runs: 38.5M spots, 7.7G bases, 4.5Gb downloads Accession: SRX124559

6. When quoting N50 values, please include the units. For example, the N50 for DU054 was 9,951 bp or nt.

7. The authors refer to (on page 4) "more complete annotations" of several species. Among these examples is P. lateralis and a citation of our paper (PubMed 23678994) about the sequencing of this species' genome; however, I would not agree that its annotation is "more complete".

8. On page 3, second paragraph, the authors write "the available genome". It is not the "genome" that is available; rather it is the "genome sequence".

Once the authors have addressed all these issues, I would be very pleased to see this indexed.

References

1. Studholme DJ, McDougal RL, Sambles C, Hansen E, Hardy G, Grant M, Ganley RJ, Williams NM: Genome sequences of six Phytophthora species associated with forests in New Zealand.*Genom Data*. 2016; **7**: 54-6 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 28 Feb 2018

Mark Richardson, Deakin University, Australia

Thank you very much for providing a thorough review and pointing out several oversights we made. We have endeavoured to rectify these as you will see from our responses below. Importantly, we have included the preliminary gene prediction results in the associated public repository with the current supplementary materials. We have also revised the manuscript to include the additional genomes from Studholme *et al.* 2015 and included them in a comparative BUSCO completeness assessment. We feel a more comprehensive comparative analysis (including ANIs) is beyond the scope of this research note, but this will be part of a future paper.

For clarity, the below responses are separated into Major and Minor subheadings and numbered as per points in the review in order to avoid duplication of text.

Major:

1. Contamination for human DNA should not be more prevalent than any other. As this was one of the first times we cultured this species we carried out this pre-filtering to ascertain whether or not we had any inadvertent contamination. The results show this was not the case.2. We find it interesting that the reviewer has also detected Paenibacillus contamination during their work.

2. While removing contamination through mapping to the Paenibacillus genome alone would not warrant our statement, this is not what we did. We used MetaPHlAn to first screen our raw reads to identify which, if any, bacterial species might be present. Only Paenibacillus could be detected. Thereby, once removed, we are confident that no other bacterial contamination exists. If others had been identified with MetaPHlA then they could be removed in the same way.

3. We have added a citation to this extent and clarified what we mean in the text.

4. We have repeated this analysis with version 3.02 and used the suggested ortholog set.

5. With respect, we feel that the reviewer's statement that the completeness was poor is unfounded, especially if we consider that they are not making a 'like for like' comparison by comparing results from the eukaryotic set to those from the stramenopile set. Nevertheless, the updated BUSCO analysis using the stramenopile set reveals the genome assemblies presented here have BUSCO completeness of ~91 to 94 %, which falls within the range for the previous *P. cinnamomi* assembles (86 -97% completeness, see Table 2 for full comparison).

6. Thank you for this suggestion, we have done so.

7. This is a very valid point. We have now included the preliminary gene predictions with the supplementary data.

Minor:

1. We have changed this to the more commonly understood ‘ecoregion’
2. Done
3. Done
4. Done
5. We have removed this statement.
6. Done
7. We have removed this statement.
8. Addressed

Competing Interests: No competing interests were disclosed.

Referee Report 20 November 2017

doi:10.5256/f1000research.13945.r27740

Erik Andreasson , **Laura Grenville Briggs**
Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Alnarp, Sweden

This data adds information about this important organism in the standard format to report a draft genome these days so it looks fine. They used hiseq and sequence coverage (BUSCO) looks appropriate and expected although there are relatively large differences between the two isolates (i.e .different final genome sizes and busco completeness scores). One suggestion is to add information on how many libraries they sequenced, and if it was only paired end and not also mate paired.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 28 Feb 2018

Mark Richardson, Deakin University, Australia

Thank you for the positive review, we have added the additional information you have requested.

Competing Interests: No competing interests were disclosed.

Referee Report 17 November 2017

doi:10.5256/f1000research.13945.r27739

Nicolás Daniel Ayub
National Scientific and Technical Research Council, Buenos Aires, Argentina

The work was carried out professionally and resulted in good draft genomes of two pathogen strains belonging to Phytophthora genus. In my opinion, this article is an important contribution to future studies about the molecular mechanism involved in *Phytophthora*-plant interaction. Particularly, in the first steps of pathogen adhesion, where the virulence factors related to this are little known.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 28 Feb 2018

Mark Richardson, Deakin University, Australia

Thank you for the positive review.

Competing Interests: No competing interests were disclosed.