

Reports

Chromium sequencing: the doors open for genomics of obligate plant pathogens

Alistair R McTaggart^{†,1,2}, Tuan A Duong^{†,1}, Vang Quy Le³, Louise S Shuey¹, Werner Smidt¹, Sanushka Naidoo¹, Michael J Wingfield¹ & Brenda D Wingfield^{*,1}

¹Department of Biochemistry, Genetics & Microbiology, Tree Protection Co-operative Programme (TCP), Forestry & Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa; ²Queensland Alliance for Agriculture & Food Innovation, The University of Queensland, Ecosciences Precinct, Brisbane, Queensland, Australia; ³Section of Molecular Diagnostics, Aalborg University Hospital, Denmark

BioTechniques 65: 253-257 (November 2018) 10.2144/btn-2018-0019

Keywords: chromium 10X • fungal genomes • myrtle rust • rust fungi

For reprint orders, please contact: reprints@futuremedicine.com

It is challenging to sequence and assemble genomes of obligate plant pathogens and microorganisms because of limited amounts of DNA, comparatively large genomes and high numbers of repeat regions. We sequenced the 1.2 gigabase genome of an obligate rust fungus, *Austropuccinia psidii*, the cause of rust on *Myrtaceae*, with a Chromium 10X library. This technology has mostly been applied for single-cell sequencing in immunological studies of mammals. We compared scaffolds of a genome assembled from the Chromium library with one assembled from combined paired-end and mate-pair libraries, sequenced with Illumina HiSeq. Chromium 10X provided a superior assembly, in terms of number of scaffolds, N50 and number of genes recovered. It required less DNA than other methods and was sequenced and assembled at a lower cost. Chromium sequencing could provide a solution to sequence and assemble genomes of obligate plant pathogens where the amount of available DNA is a limiting factor.

Obligate pathogens cannot be cultured. Consequently, there is often insufficient DNA available for whole-genome sequencing. This problem is reflected by the small number of genomes sequenced for obligate fungi in the 1K fungal genomes project [1,2]. Limited DNA quantity is an obstacle to study genomics of fungi and plant pathogens such as those from marine environments, obligate lichen associates, insect associates, downy mildews, powdery mildews, rust fungi and some obligate smut fungi [3,4].

Most genome-sequencing technologies require large amounts of DNA for preparation of a library. For example, 2–7 µg of DNA is recommended for preparation of paired-end and mate-pair libraries for sequencing on an Illumina platform. Long-read platforms, such as PacBio and Nanopore, require 1–15 µg per cell for sequencing; however, many cells are needed for significant sequencing depth

of large genomes. Larger amounts of DNA are difficult to obtain for microorganisms that cannot be isolated and cultured. Genomes that have been sequenced for obligate microorganisms are often those of well-studied models, for which material can be cultivated, for example poplar and wheat rust [5].

Rust fungi are one of the most speciose groups of obligate fungal plant pathogens, with genomes that are repeat rich and amongst the largest in the Fungal Kingdom [6]. Despite their diversity and significance as plant pathogens, relatively few genomes of rust fungi have been sequenced. This is partly because of the challenges posed by their obligate nature [3].

There are five near-complete, publicly available genomes of rust fungi from cereal grasses and poplar, assembled from 454 and Illumina sequencing [5,7–11]. However, five other projects have not yet published

their assembled genomes of rust fungi because of difficulties encountered from size, repeat content and different haplotypes in the genome [12–16].

Long-read sequencing platforms, such as PacBio and NanoPore, may offer solutions to sequence and assemble repetitive genomes of rust fungi [3]. For example, PacBio sequencing was recently tested to assemble and phase dikaryotic genomes of two rust fungi [17,18]. This research showed that long-read sequencing could resolve challenges for genome assembly of dikaryotic taxa. However, if DNA quantity is a limiting factor for nonmodel plant pathogens, these technologies cannot provide sequencing coverage depth, especially for organisms with large genomes.

Chromium sequencing, developed by 10X Genomics (www.10xgenomics.com), has provided a low-cost approach to assemble human genomes from limited

METHOD SUMMARY

We sequenced and assembled the genome of an obligate fungus to compare Chromium sequencing technology with paired-end and mate-pair libraries. This was the first application of Chromium sequencing for an obligate fungal pathogen, in which DNA quantity is a limiting factor for genomics.

Table 1. Statistics of the genomes assembled from different sequencing libraries.

Assembler	Data used	Amount of starting material (μg)	Cost of data generation (€)	No. of scaffolds (>500 bp)	N50 (kb)	L50	Estimated genome size (bp)	Completeness of genome (BUSCO fungi; %)
Supernova	Chromium	1	3000	147,937	60.61	4702	1.20×10^9	80.30
ABYSS	Mate-pair and paired-end	7	6430	390,143	12.58	18,904	998.4×10^6	75.20

Comparison of genomes sequenced and assembled from Chromium with combined mate-pair and paired-end sequence data for *Austropuccinia psidii*. The genome of *A. psidii* assembled from the Chromium library in this study has been deposited at DDBJ/EMBL/GenBank under the accession QOUH00000000.

starting material [19]. Small amounts, as low as 1 ng, of high-quality DNA are fragmented into ~100-kb molecules, which are then barcoded and amplified within micelles to prepare a Chromium library. Homologous regions of DNA are unlikely to be present in the same micelle, which allows haplotypes to be phased using their micelle-specific barcodes. After amplification, the barcoded library is sequenced on an Illumina platform, and reads are assembled with Supernova Software from 10X Genomics.

The medical community has largely adopted Chromium sequencing for genomics and transcriptomics; however, at least three studies have used it to assemble, scaffold and phase genomes of plants [20–22]. Chromium sequencing has not yet been tested for assembly of fungal genomes; there would be few advantages to sequence haploid and culturable fungi with this approach because there is no need to phase haplotypes and quantity of DNA is not limiting. However, obligate, dikaryotic plant pathogens with large genome sizes are perfect candidates to sequence, assemble and phase genomes with Chromium sequencing.

We tested whether Chromium 10X provided a competitive platform for sequencing and *de novo* genome assembly of a rust pathogen, *Austropuccinia psidii*, here used as a model for other microorganisms. The genome assembly from a Chromium library was compared with a combined assembly from one paired-end and one mate-pair library sequenced on Illumina.

Materials & methods

Selection of isolate & DNA extraction

To obtain enough material from a single genotype of *A. psidii*, a single-pustule isolate was grown on *Syzygium jambos* to extract DNA from approximately 500 mg of urediniospores. This is a technically challenging and laborious process that is not always possible with other species of rust fungi. The urediniospores were ground in liquid N_2 and

DNA was extracted following the protocol of a Machery Nagel NucleoSpin® Plant II DNA extraction kit. The taxon selected, *A. psidii*, has an unreleased genome sequenced with paired-end reads on Illumina MiSeq [16], and a publicly available genome, approximately half-complete, sequenced with paired-end reads on Illumina HiSeq [23].

Genome sequencing

Genomic DNA was sent to Macrogen Korea for library preparation and sequencing. Three libraries were prepared, a Chromium 10X library, a mate-pair library with a 5-kb insert, and a paired-end library with a 350-bp insert. Macrogen prepared the paired-end and mate-pair libraries from 2 and 7 μg of DNA, respectively (~100 and 350 mg of spores), and the Chromium 10X library from 1 μg of DNA (~50 mg of spores) with fragments above 40 kb size-selected using BluePippin (Sage Science Inc.). The Chromium 10X and paired-end libraries were run on Illumina HiSeqX, and the mate-pair library was run on Illumina HiSeq4000.

Genome assembly Assembly of Chromium 10X sequencing data

A single run of a Chromium library on a HiSeq X Ten yielded 812.77 million 150-bp reads (~120 gigabytes of raw data). The Chromium data were processed (quality filtering and trimming) and assembled using Supernova v 1.2.2 from 10X Genomics [19]. The Supernova run on barcoded micelles resulted in a weighted mean molecule size of 24.47 kb. The assembly yielded 1443 long scaffolds (scaffolds ≥ 10 kb), with a total scaffold size of 74,872 Mb (only scaffolds ≥ 10 kb) and a N50 value of 81.25 kb. The final assembly was exported using the Supernova mkoutput function, with minimum size of scaffolds set to 500 bp, and this assembly has been deposited in GenBank (accession: QOUH00000000). The statistics for the final assembly from Supernova were reported in Table 1.

Assembly of paired-end & mate-pair sequencing data

Runs on Illumina HiSeq X Ten and HiSeq4000 yielded 43 gigabytes of paired-end and 29 gigabytes of mate-pair data, both of which were 150-bp reads. Trimmomatic v 0.36 [24] and NxTrim v 0.4.2 [25] were used with the default options to trim paired-end and mate-pair reads. We used ABYSS 2.0 [26] to assemble (k-mer = 99) paired-end reads, and to scaffold with mate-pair reads. Only scaffolds ≥ 500 bp were retained. The statistics obtained from the assembly with ABYSS were reported in Table 1.

Comparison of Supernova & ABYSS assemblies

We used the NUCmer function in MUMmer [27] to compare the 18 largest scaffolds (>5 Mb) assembled by Supernova (from Chromium reads) to their homologous scaffolds (>5 Mb) assembled by ABYSS (from combined paired-end and mate-pair reads). We used Circos [28] to visualize the homology between 18 scaffolds from Supernova and 220 scaffolds from ABYSS (Figure 1).

Results & discussion

The first challenge to sequence genomes of obligate pathogens is the availability of DNA for preparation of a library. Chromium sequencing overcame this hurdle in the present study where a library was prepared from 1 μg of DNA. This amount of DNA was obtained from approximately 50 mg of spores. Long-read sequencing technologies and paired-end and mate-pair libraries are prepared from larger amounts of DNA; they are not practical options to obtain sequencing coverage depth when DNA quantity is a limiting factor.

Repeat-rich regions present another challenge to sequence and assemble genomes of obligate plant pathogens. Chromium 10X libraries outperformed combined paired-end and mate-pair libraries

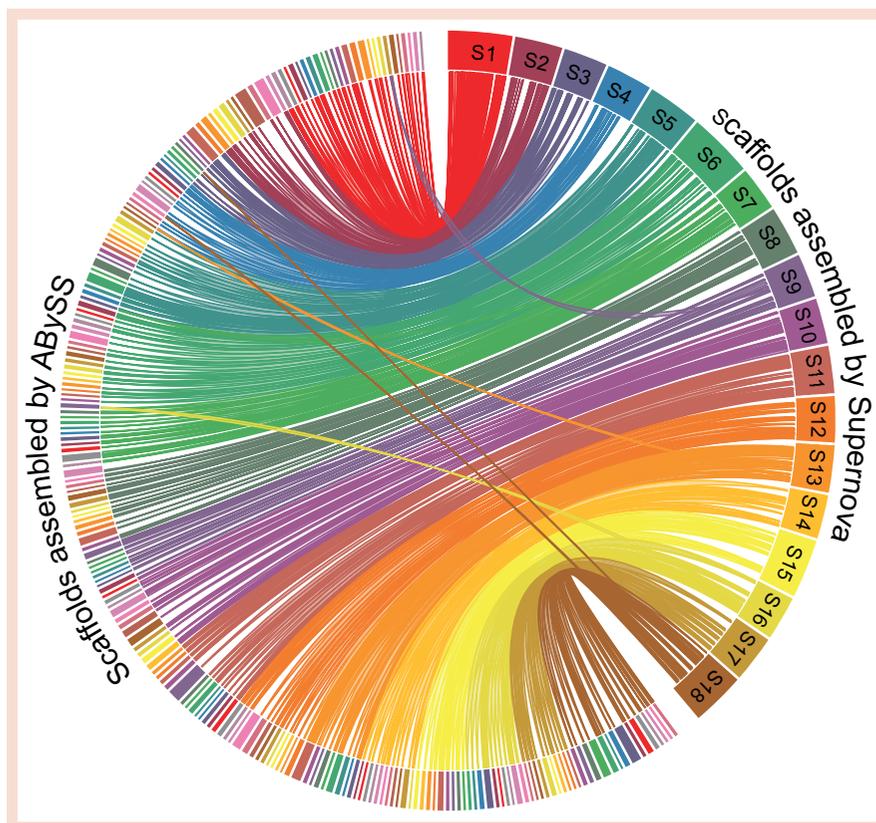


Figure 1. Circos plot showing 18 largest scaffolds (>0.5 Mb, indicated as S1 to S18) from the Supernova assembly (Chromium sequencing) and their 220 homologous scaffolds from the ABySS assembly (combined paired-end and mate-pair sequencing). The connected ribbons indicate the homologous relationship between scaffolds as determined using the NUCmer tool in MUMmer, which only included alignments larger than 5000 kb.

for assembly of repeat regions, and much larger contigs were assembled based on N50 and L50 (Table 1). Scaffolds obtained from Supernova and ABySS were mostly homologous; 220 scaffolds from ABySS were congruent with the largest 18 scaffolds from Supernova, with a few minor incongruences likely caused by repetitive regions (Figure 1). Supernova first assembles DNA molecules from barcoded reads in each micelle, which excludes homogenous repeat regions from other parts of the genome. However, certain types of repetitive regions, such as tandem repeats, still pose a challenge, and these may be better assembled with low-coverage long reads, such as those from PacBio or NanoPore, and combined in hybrid assemblies that provide more data for sequencing depth [29].

De novo genome assembly with Chromium was more economical than combined paired-end and mate-pair libraries, which is another advantage of this technology. A comparable amount of data obtained from PacBio would have required ~80 cells (at 1.2 gigabytes of data per cell), cost over 70,000€ (at 900€ per cell), and needed at least 100 µg of DNA (based on

a recommended minimum 5 µg per SMRT cell).

The genome size of *A. psidii* was first reported as 103–145 megabases [16], whereas it was estimated as >1.2 Gb in the present study. The difference between the two sequenced genomes may be that they are different strains, the pandemic strain sequenced by Tan *et al.* [16] and the South African strain [30] used here. However, intra-specific genomic diversity in *P. striiformis* f. sp. *tritici* and *P. triticinina* was attributed to differences in sequencing platforms used, rather than differences between strains of rust [7–10,31,32]. We hypothesize the size difference of ~1 Gb between the strains is a reflection of the sequencing technology used, a Chromium 10X library sequenced on HiSeqX10 in the present study compared with MiSeq in Tan *et al.* [16].

Chromium 10X sequencing provided an effective method to sequence and assemble the genome of an obligate plant pathogen from a small quantity of starting material. The Chromium library preparation of *A. psidii* used less DNA at a cheaper cost to obtain a better assembly, in terms of number of scaffolds (Figure 1) and BUSCO

genes recovered [33], than paired-end and mate-pair libraries (Table 1). Chromium sequencing has opened the door for *de novo* genome assembly, comparative genomics and phylogenomics of obligate plant pathogens and microorganisms, especially when DNA is a limiting factor.

Author contributions

AM, TD, VQ and BW designed the study. TD, VQ, WS and LS performed the experiments and analyzed the data. AM and TD wrote the manuscript. SN, MW and BW supervised the study and corrected the manuscript.

Acknowledgments

Dr Roger Shivas and three anonymous reviewers are thanked for their improvements to the manuscript.

Competing & financial interests disclosure

This work is based on research partially supported by the Tree Protection Co-operative Programme (TPCP) and the National Research Foundation of South Africa and the DST-NRF Centre of Excellence in Tree Health Biotechnology (CTHB). The grant holders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by NRF-supported research are that of the authors and that the NRF accepts no liability whatsoever in this regard. ARM acknowledges the University of Queensland Development Fellowships (UQFEL1718905) and support from the Department of the Environment and Energy under the Australian Biological Resources Study (grant number RG18-43). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

- Grigoriev IV, Nikitin R, Haridas S *et al.* MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42, D699–D704 (2014).
- Aylward J, Steenkamp ET, Dreyer LL *et al.* A plant pathology perspective of fungal genome sequencing. *IMA Fungus* 8, 1–15 (2017).
- Aime MC, McTaggart AR, Mondo SJ, Duplessis S. Phylogenetics and phylogenomics of rust fungi. In: *Advanced Genetics*. Townsend JP, Wang Z (Eds.). Academic Press, 100, 267–307 (2017).
- Hacquard S. The genomics of powdery mildew fungi. *Adv. Bot. Res.* 70, 109–142 (2014).
- Duplessis S, Cuomo CA, Lin YC *et al.* Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl Acad. Sci.* 108, 9166–9171 (2011).
- Tavares S, Ramos AP, Pires AS *et al.* Genome size analyses of *Pucciniales* reveal the largest fungal genomes. *Front. Plant Sci.* 5, 422 (2014).
- Cantu D, Govindarajulu M, Kozik A *et al.* Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE* 6, e24230 (2011).
- Cantu D, Segovia V, MacLean D *et al.* Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* 14, 270 (2013).
- Cuomo CA, Bakkeren G, Khalil HB *et al.* Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. *G3 (Bethesda)* 7, 361–376 (2017).
- Kiran K, Rawal HC, Dubey H *et al.* Draft genome of the wheat rust pathogen (*Puccinia triticina*) unravels genome-wide structural variations during evolution. *Genome Biol. Evol.* 8, 2702–2721 (2016).
- Rochi L, Diéguez MJ, Burguener G *et al.* Characterization and comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize common rust. *Fungal Genet. Biol.* 112, 31–39 (2018).
- Cristancho MA, Botero Rozo DO, Giraldo W *et al.* Annotation of a hybrid partial genome of the Coffee Rust (*Hemileia vastatrix*) contributes to the gene repertoire catalogue of the Pucciniales. *Front. Plant Sci.* 5, 594 (2014).
- Link T, Seibel C, Voegelé RT. Early insights into the genome sequence of *Uromyces fabae*. *Front. Plant Sci.* 5, 587 (2014).
- Loehrer M, Vogel A, Huettel B *et al.* On the current status of *Phakopsora pachyrhizi* genome sequencing. *Front. Plant Sci.* 5, 377 (2014).
- Nemri A, Saunders D, Anderson C *et al.* The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. *Front. Plant Sci.* 5, 98 (2014).
- Tan MK, Collins D, Chen Z, Englezou A, Wilkins MR. A brief overview of the size and composition of the myrtle rust genome and its taxonomic status. *Mycology* 5, 52–63 (2014).
- Miller ME, Zhang Y, Omidvar V *et al.* *De novo* assembly and phasing of dikaryotic genomes from two isolates of *Puccinia coronata* f. sp. *avenae*, the causal agent of oat crown rust. *mBio* 9, e01650 (2018).
- Schwessinger B, Sperschneider J, Cuddy WS *et al.* A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high interhaplotype diversity. *mBio* 9, e02275 (2018).
- Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. *Genome Res.* 27, 757–767 (2017).
- Hulse-Kemp AM, Maheshwari S, Stoffel K *et al.* Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic. Res.* 5, 4 (2018).
- Zhang GQ, Liu KW, Li Z *et al.* The *Apostasia* genome and the evolution of orchids. *Nature* 549, 379 (2017).
- Crepeau MW, Langley CH, Stevens KA. From pine cones to read clouds: resc scaffolding the megagenome of sugar pine (*Pinus lambertiana*). *G3 (Bethesda)* 7, 1563–1568 (2017).
- Sandhu KS, Karaoglu H, Zhang P, Park RF. Simple sequence repeat markers support the presence of a single genotype of *Puccinia psidii* in Australia. *Plant Pathol.* 65, 1084–1094 (2016).
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).
- O'Connell J, Schulz-Trieglaff O, Carlson E *et al.* NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics* 31, 2035–2037 (2015).
- Jackman SD, Vandervalk BP, Mohamadi H *et al.* ABYSS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* 27, 768–777 (2017).
- Kurtz S, Phillippy A, Delcher AL *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12 (2004).
- Krzywinski M, Schein J, Birol I *et al.* Circos: sn information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645 (2009).
- Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Rev. Genet.* 15, 121 (2014).
- Roux J, Granados GM, Shuey L *et al.* A unique genotype of the rust pathogen, *Puccinia psidii*, on Myrtaceae in South Africa. *Australas. Plant Path.* 45, 645–652 (2016).
- Kiran K, Rawal HC, Dubey H *et al.* Dissection of genomic features and variations of three pathotypes of *Puccinia striiformis* through whole genome sequencing. *Sci. Rep.* 7, 42419 (2017).
- Zheng W, Huang L, Huang J *et al.* High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nat. Commun.* 4, 2673 (2013).
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212 (2015).

First draft submitted: 19 March 2018;
Accepted for publication: 31 July 2018

Address correspondence to: Brenda D Wingfield, Department of Biochemistry, Genetics & Microbiology, Tree Protection Co-operative Programme (TPCP), Forestry & Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa; Brenda.Wingfield@fabi.up.ac.za

†Authors contributed equally

To purchase reprints of this article contact:
s.cavana@future-science.com



Blot-to-blot consistency. Time after time.

To achieve meaningful comparison of expressed proteins from blot-to-blot, trustworthy data is essential. **BlotCycler™** is designed to eliminate slight variations in fluid delivery, timing and washing that may occur when protocols are performed manually. With rigorous precision, **BlotCycler** delivers an affordable solution to busy labs demanding reproducible results.



BlotCycler™
by PRECISION BIOSYSTEMS

www.blotcyclers.com