

# Reference Genome Assembly for Australian *Ascochyta rabiei* Isolate ArME14

Ramisah Mohd Shah,<sup>†,1</sup> Angela H. Williams,<sup>†,\*</sup> James K. Hane,<sup>\*,†</sup> Julie A. Lawrence,<sup>\*</sup>

Lina M. Farfan-Caceres,<sup>\*</sup> Johannes W. Debler,<sup>\*</sup> Richard P. Oliver,<sup>†,\*</sup> and Robert C. Lee<sup>\*,2</sup>

<sup>\*</sup>Centre for Crop and Disease Management, School of Molecular and Life Sciences, Curtin University, Bentley, WA, Australia, <sup>†</sup>Murdoch University, Murdoch, WA, Australia, and <sup>‡</sup>Department of Environment and Agriculture, Curtin University, Bentley, WA, Australia

ORCID IDs: 0000-0003-0196-0022 (A.H.W.); 0000-0002-7651-0977 (J.K.H.); 0000-0002-3604-051X (J.W.D.); 0000-0001-7290-4154 (R.P.O.); 0000-0002-4174-7042 (R.C.L.)

**ABSTRACT** *Ascochyta rabiei* is the causal organism of ascochyta blight of chickpea and is present in chickpea crops worldwide. Here we report the release of a high-quality PacBio genome assembly for the Australian *A. rabiei* isolate ArME14. We compare the ArME14 genome assembly with an Illumina assembly for Indian *A. rabiei* isolate, ArD2. The ArME14 assembly has gapless sequences for nine chromosomes with telomere sequences at both ends and 13 large contig sequences that extend to one telomere. The total length of the ArME14 assembly was 40,927,385 bp, which was 6.26 Mb longer than the ArD2 assembly. Division of the genome by OcculterCut into GC-balanced and AT-dominant segments reveals 21% of the genome contains gene-sparse, AT-rich isochores. Transposable elements and repetitive DNA sequences in the ArME14 assembly made up 15% of the genome. A total of 11,257 protein-coding genes were predicted compared with 10,596 for ArD2. Many of the predicted genes missing from the ArD2 assembly were in genomic regions adjacent to AT-rich sequence. We compared the complement of predicted transcription factors and secreted proteins for the two *A. rabiei* genome assemblies and found that the isolates contain almost the same set of proteins. The small number of differences could represent real differences in the gene complement between isolates or possibly result from the different sequencing methods used. Prediction pipelines were applied for carbohydrate-active enzymes, secondary metabolite clusters and putative protein effectors. We predict that ArME14 contains between 450 and 650 CAZymes, 39 putative protein effectors and 26 secondary metabolite clusters.

## KEYWORDS

PacBio  
*Pleosporales*  
*Dothideomycetes*  
plant pathogen  
chickpea

Ascochyta blight disease of chickpea (*Cicer arietinum* L.), is caused by *Ascochyta rabiei* (Pass.) Labrousse (teleomorph: *Didymella rabiei* (Kovachevski) von Arx, syn. *Mycosphaerella rabiei* (Kovachevski),

syn *Phoma rabiei*) (de Gruyter *et al.* 2009; Aveskamp *et al.* 2010) and is the major biotic constraint affecting production of chickpea worldwide. Chickpea is a legume crop species of global economic importance (Fondevilla *et al.* 2015). Worldwide production was 12 M tonne in 2016 with two-thirds grown in India. Australia is a major exporter and the combined control costs and losses attributed to ascochyta blight in Australia in 2012 were around AU\$40 M (Murray and Brennan 2012) which represents approximately 10% of crop value.

*A. rabiei* is a haploid, heterothallic, *Dothideomycetes* fungus (Order *Pleosporales*) and it causes disease by producing necrosis in leaf, stem and pod tissues (Trapero-Casas and Kaiser 1992; Peever *et al.* 2004). Little is known about how it causes disease. It produces solanapyrones, phytotoxic secondary metabolites in culture filtrates, and these have long been considered to play a role in pathogenicity (Alam *et al.* 1989; Hamid and Strange 2000). However deletion of the

Copyright © 2020 Mohd Shah *et al.*

doi: <https://doi.org/10.1534/g3.120.401265>

Manuscript received January 12, 2020; accepted for publication April 23, 2020; published Early Online April 28, 2020.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25387/g3.11589420>.

<sup>1</sup>Present address: Faculty of Fisheries and Food Science, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia.

<sup>2</sup>Corresponding author: Centre for Crop and Disease Management, School of Molecular and Life Sciences, Curtin University, Kent St, Bentley, WA, Australia, 6102. E-mail: Robert.Lee@curtin.edu.au

biosynthesis genes responsible for the synthesis of solanapyrones did not affect virulence (Kim *et al.* 2015a,b). Attention has since shifted to effectors, proteins that control the plant-pathogen interaction and in some cases induce necrosis (Lo Presti *et al.* 2015; Tan and Oliver 2017). In terms of pathogen life cycle and population structure, the two mating types are found in *A. rabiei* populations in Israel (Lichtenzweig *et al.* 2005), North America (Peever *et al.* 2004) and Canada (Armstrong *et al.* 2001), thus providing a mechanism for sexual recombination. In Australia, most reports suggest the presence of only mating-type MAT1-2 (Leo *et al.* 2015; Mehmood *et al.* 2017) and the absence of mating in the population. SSR genotyping has found the Australian *A. rabiei* population to be highly homogenous with around 70% of isolates tested being from a single dominant haplotype designated ARH01 (Leo *et al.* 2015; Mehmood *et al.* 2017).

In 2016, an Illumina short-read genome assembly of an Indian *A. rabiei* isolate, ArD2, was published (Verma *et al.* 2016). In their analysis Verma *et al.* (2016) predicted 758 secreted proteins from a total of 10,596 proteins encoded by the 34.6 Mb genome assembly (Verma *et al.* 2016). Of the 758 predicted secreted proteins, 201 proteins were annotated as Carbohydrate-Active Enzymes (CAZymes) (Lombard *et al.* 2014). These included proteins containing carbohydrate-binding modules and LysM domains that characterize chitin-binding effectors, which suppress the immune response of plants to fungal pathogens (de Jonge *et al.* 2010). There were 323 putative effectors with no known protein domain and 70 proteins of the predicted secretome showed sequence similarity to members of the Pathogen-Host Interaction (PHI) database (Winnenburg *et al.* 2006) with annotated involvement in virulence or pathogenicity (Verma *et al.* 2016). In addition to proposed virulence proteins, transcription factors that control gene expression during plant infection by *A. rabiei* have been predicted for ArD2 and assessed for their contribution to pathogenesis (Verma *et al.* 2017). Verma *et al.* (2017) suggest that for *A. rabiei*, Myb transcription factors play a role in regulating the expression of genes encoding secreted proteins, such as effectors. Gene regulation by transcription factors during plant infection may also vary depending on specific isolate and host interactions (Verma *et al.* 2017).

Here, we present a genome assembly for an Australian *A. rabiei* isolate, ArME14 that was produced by amalgamation of whole genome DNA sequence data from both Illumina and PacBio SMRT sequencing. The updated genome assembly features 12 full-length chromosome contigs with telomere sequences at both ends, 9 partial chromosomal contigs with one telomere end, and 13 smaller fragments with no telomere. The total assembled genome length was 40.9 Mb and genome annotation predicted 11,257 gene models.

## MATERIALS AND METHODS

### Fungal culture and DNA extraction

*A. rabiei* isolate ArME14 was collected from chickpea at the Department of Agriculture and Food Western Australia (DAFWA) field station, Medina, Western Australia in 2004. Fungal cultures were grown for three days in potato dextrose liquid media, at approx. 22° with shaking (150 rpm). For Illumina sequencing, DNA was prepared using a standard CTAB extraction method. RNA was removed by incubation with DNase-free RNase A and DNA was resuspended in TE buffer (10 mM Tris-HCl 1 mM EDTA, pH 8). DNA concentration was determined by NanoDrop Spectrophotometer (Thermo-Fisher Scientific, Waltham, MA, USA) and quality and purity were assessed by agarose gel electrophoresis. For PacBio sequencing, maxi-prep DNA extractions were produced using a modified method from Xin

and Chen (2012). Fungal material was grown in Yeast Extract Glucose liquid media at approx. 22° with shaking (180 rpm), for 72 hr. DNA was resuspended in 2 mL Tris-HCl, pH 8.0 and treated with 20 µg.mL<sup>-1</sup> DNase-free RNase A. DNA was purified using Ampure XP beads (Agencourt, Beckman-Coulter, USA) in a 96-well microtitre plate. The final DNA solution was quantified using Qubit and Nanodrop (Thermo Fisher Scientific) assays. Gel electrophoresis on a 1% agarose gel was used to assess DNA quality.

### Genome sequencing

Short-read DNA sequencing was performed at the Allan Wilson Genome Centre (Massey University, Palmerston North, New Zealand) using an Illumina Genome Analyzer (Illumina Inc., San Diego, CA, USA). Illumina TruSeq paired-end libraries were prepared for *A. rabiei* isolate ArME14 DNA, size-selected for 200 bp fragments, from which 75 bp reads were sequenced. Single-Molecule, Real-Time (SMRT) PacBio sequencing was performed by Genome Quebec (McGill University, Montreal, Canada). Libraries were prepared with size-selected 17 Kb fragments from sheared genomic DNA using P6-C4 chemistry, and sequenced on six SMRT cells using a PacBio RSII instrument (Pacific Biosciences, Menlo Park, CA, USA).

### Reference genome assembly

PacBio SMRT reads were assembled using the CANU v 1.2 assembler (Berlin *et al.* 2015) and the resulting intermediate assembly sequences were corrected using Illumina reads via PILON v 1.2.1 (Walker *et al.* 2014). A single mitochondrial genome contig from the assembly was identified by homology with other published *Dothideomycete* genome sequences and was designated 'Mitochondrion MT' in the new assembly. Sequencing statistics for the final corrected *A. rabiei* ArME14 reference assembly including overall percent GC content, were calculated using QUAST v 4.6.2 (Gurevich *et al.* 2013). Telomere sequences were manually recorded based on the presence of TTAGGG tandem repeat sequences at contig ends (Schechtman 1990). Computational steps were coordinated using Nextflow (Di Tommaso *et al.* 2017).

### Genome annotation and analysis

Gene prediction for the reference *A. rabiei* ArME14 assembly was performed using the annotation program, AUGUSTUS v 3.3 (Stanke *et al.* 2004, 2006; König *et al.* 2016), based on sequence homology of *in vitro* and *in planta* RNASeq and Massive Analysis of cDNA Ends (MACE) libraries from the *A. rabiei* BioProject, PRJNA288273 (Fondevilla *et al.* 2015). We used BUSCO version 3.0 (Simão *et al.* 2015) to assess assembly and annotation completeness by running protein fasta files with benchmarking against the Ascomycota\_odb9 single-copy orthologs file downloaded from the BUSCO website September 2019 (<https://busco.ezlab.org/>). We used the program OcculterCut v 1.1 (Testa *et al.* 2016) to scan the genome assembly to determine its percent GC content distribution.

For detection and assessment of transposable element and repeat sequences, the suite of detection and classification programs in the PiRATE-Galaxy pipeline virtual machine as described by Berthelie *et al.* (Berthelie *et al.* 2018) was used. PiRATE-Galaxy uses similarity-based detection programs RepeatMasker (Smit *et al.* 2013) and TE-HMMER (Berthelie *et al.* 2018), a custom program based on HMMER (Eddy 1995) and tBLASTn (Altschul *et al.* 1990); structure-based programs MITE-Hunter (Han and Wessler 2010), SINE-Finder (Wenke *et al.* 2011), Helsearch (Yang and Bennetzen 2009) and LTRharvest (Ellinghaus *et al.* 2008); and repetitiveness-based programs,

TEdenovo (Flutre *et al.* 2011) and RepeatScout (Price *et al.* 2005). CD-HIT-EST (Li and Godzik 2006) in PiRATE-Galaxy was used to reduce redundancy in the combined TE and repeat sequence dataset by removing duplicated sequences with 100% identity to other longer sequences in the data set. Short sequences of less than 500 nucleotides were removed. Classification of repeat and TE sequences was implemented using the program PASTEC (Hoede *et al.* 2014), using nucleotide, protein and profile HMM databanks from the PiRATE-Galaxy server (Bertheliet *et al.* 2018).

Calculating coverage of ArME14 by the previous ArD2 assembly was performed with NUCMER (Kurtz *et al.* 2004) using the max-match argument and plotted using ggplot (Wickham 2016). Prediction of secreted proteins from the *A. rabiei* ArME14 set of annotated protein sequences was accomplished using SignalP version 5.0 (Almagro Armenteros *et al.* 2019), and DeepSig (Savojardo *et al.* 2018). From the set of annotated proteins, we applied effector selection criteria, including mature polypeptide molecular weight less than 25 kDa, number of cysteines, presence of a secretion signal and EffectorP 2.0 score greater than 0.8 (Sperschneider *et al.* 2016, 2018), to predict putative effector proteins using a custom pipeline written in Python [Johannes Debler. (2019, November 4). JWDebler/effector\_selection: First working release (Version v1.0). Zenodo. <http://doi.org/10.5281/zenodo.3526820>]. Where there was disagreement between SignalP and DeepSig on the signal peptide processing site, the custom pipeline chose the site determined by SignalP.

CAZymes were identified from the ArME14 annotated set of proteins using the dbCAN2 web-based meta server (Zhang *et al.* 2018), implementing HMMER v 3.2.1 (Eddy 1996) with the HMMdb release 8.0, DIAMOND (Buchfink *et al.* 2015) and Hotpep (Busk *et al.* 2017). Secondary metabolite clusters were identified using the anti-SMASH fungal version v 5.0, web-based prediction server (Medema *et al.* 2011; Blin *et al.* 2019). A CIRCOS plot that illustrates all annotated genome features was produced using the CIRCOS software v 0.69-9 (Krzywinski *et al.* 2009).

### Data availability

Illumina and PacBio genome sequencing data for *A. rabiei* ArME14 described herein, and the reference genome assembly have been deposited in the Sequence Read Archive and NCBI database, under the BioProject accession number PRJNA510692. *A. rabiei* ArME14 BioSample number is SAMN10613128. Illumina SRA entries are deposited under SRX5179494 and PacBio SRA data under SRX5172972. The GenBank assembly accession number is GCA\_004011695.1. Supplemental material available at figshare: <https://doi.org/10.25387/g3.11589420>

## RESULTS AND DISCUSSION

PacBio SMRT sequencing of ArME14 produced 34 contigs, including one mitochondrial genomic contig, at 166x sequencing depth (Table 1). The ArME14 assembly was 18% larger at 40,927,385 bp compared with 34,658,250 bp for the ArD2 genome assembly (Verma *et al.* 2016). Telomeres were identified manually by sequence observation and all were TTAGGG repeats of approximately 100 bp in length. Repeated TTAGGG sequences are reported to characterize the telomere regions in filamentous fungi such as *Neurospora crassa* (Schechtman 1990), *Cladosporium fulvum* (Coleman *et al.* 1993) and *Magnaporthe oryzae* (Rehmeyer *et al.* 2006; Farman 2007), and this sequence motif is a conserved feature in ArME14. Of the 33 nuclear contigs, 12 had TTAGGG telomere sequences at both ends (contig sizes 3,373,759 – 1,223,093 bp). Nine others had one telomere (contig sizes 2,532,578 – 1,278,587 bp). Values for L50 and N50 (Table 1)

were 9 and 1,812,190 bp, compared with 64 and 154,808 bp for ArD2, respectively (Verma *et al.* 2016) (Figure 1). Akamatsu *et al.* (2012) used pulsed-field gel electrophoresis to determine chromosome number and size for multiple *A. rabiei* isolates from 21 countries. The number of chromosomes ranged from 12 to 16, and total genome size estimates ranged from 23 Mb to 34 Mb (Akamatsu *et al.* 2012). Our whole genome sequencing suggests that *A. rabiei* ArME14 possesses at least 17 chromosomes and has significantly higher genome size than previously estimated (Akamatsu *et al.* 2012). The mitochondrial genomic sequence was assembled as a single contig of 74,173 bp length (Figure 1) and was identified by homology with other fungal and *Dothideomycete* mitochondrial genome sequences. PacBio genome sequencing for fungi facilitates the assembly of long contigs by resolving the repetitive and AT-rich regions that characterize these species. For ArME14, we have been able to assemble 12 end-to-end chromosomal contigs as evidenced by the telomere sequences that terminate the contigs of the assembly. Recently produced, highly resolved genome assemblies for phytopathogenic fungi include: *Verticillium dahliae* (Faino *et al.* 2016), *Botrytis cinerea* (Van Kan *et al.* 2017), *Sclerotinia sclerotiorum* (Derbyshire *et al.* 2017) *Pyrenophora teres* (Syme *et al.* 2018) and *Pyrenophora tritici-repentis* (Moolhuijzen *et al.* 2018). For each of these PacBio sequencing was implemented but in addition to this, optical mapping, and in some cases genetic mapping were used to confirm the assembly, particularly across repetitive DNA within their genomes. For *V. dahliae*, optical mapping combined with PacBio sequencing improved the assembly from 119 contigs before optical mapping to 8 contigs after optical mapping (Faino *et al.* 2015). Even without optical and genetic mapping, but having deep sequencing coverage at 166x and Illumina correction, we are confident in the ArME14 genome sequence and the organization of GC-equilibrated and AT-rich sections in the assembly.

AUGUSTUS (v 3.3) (Stanke *et al.* 2004, 2006; König *et al.* 2016) and expressed gene sequence data from a published *A. rabiei* (isolate P4) transcriptome project (Fondevilla *et al.* 2015) were used to annotate transcribed gene features of the ArME14 genome assembly. Figure 1 shows GC-balanced, gene-rich regions as thick black bars interspersed between AT-rich, gene-sparse regions. Homology-based alignment of the ArD2 and ArME14 genome assemblies (Figure 2) shows that unique and near-exact matches between the two genomes cover the majority of the ArME14 genome. However, a substantial proportion of the homologous regions are between non-uniquely matching sequences, which are likely to be repetitive regions in both assemblies. Prediction of transposable elements and other repetitive DNA sequences for ArME14 identified that these regions comprise approximately 6.14 Mb or 15% of the genome (Table 2) and this value is roughly equal to the amount of non-unique matches indicated by genome alignment (Figure 2). The most abundant of the transposable and repetitive element types present in the *A. rabiei* genome were the Class I, long terminal repeats (LTR) with 3.3 Mb (54%), Class II terminal inverted repeats (TIR), with 1.8 Mb (30%) and long interspersed nuclear element (LINE) with 0.45 Mb (7.3%). The ArME14 genome assembly had a lower overall GC content compared with the ArD2 assembly (Table 1). Using OcculterCut (Testa *et al.* 2016) we found that the content of AT-rich DNA sequence was higher for the ArME14 assembly than for the Illumina ArD2 assembly. Around 20% of the ArME14 genome has a low GC content (between 29% and 37% GC) compared to 9.4% for ArD2. The content of GC-equilibrated regions was 32 Mb for ArME14 and 30 Mb for ArD2. Overestimation of the amount of the repetitive DNA in the ArME14 genome assembly due to mis-assembly is possible, but the difference in

**Table 1 Summary assembly and annotation statistics for Illumina sequencing of *A. rabiei* isolate, ArD2 (Verma et al. 2016) and PacBio SMRT sequencing for ArME14**

| Assembly statistics                | ArD2 Illumina (13)                     | ArME14 PacBio SMRT     |
|------------------------------------|--|------------------------|
| Genome size (bp)                   | 34,658,250                             | 40,927,385             |
| Total sequenced bases              | 100 Gb                                 | ~6.8 Gb                |
| Coverage                           | 178x                                   | 166x (928,353 reads)   |
| Number of scaffolds/contigs        | 338 <sup>a</sup>                       | 33 <sup>b</sup>        |
| Largest scaffold/contig size (bp)  | 1,160,210 <sup>a</sup>                 | 3,373,759 <sup>b</sup> |
| L50                                | 64 <sup>a</sup>                        | 9 <sup>b</sup>         |
| N50 (bp)                           | 154,808 <sup>a</sup>                   | 1,812,190 <sup>b</sup> |
| GC (%)                             | 51.6                                   | 49.2                   |
| % Repetitive sequence              | 9.9                                    | 12.6                   |
| Complete chromosomes               | —                                      | 12                     |
| <b>Annotation statistics</b>       |  |                        |
| Number of protein coding genes     | 10,596                                 | 11,257                 |
| Predicted secreted proteins        | 758 <sup>c</sup> (1,111 <sup>d</sup> ) | 1,145 <sup>c</sup>     |
| Predicted effectors                | 328 <sup>c</sup> (36 <sup>d</sup> )    | 39 <sup>c</sup>        |
| Predicted sec. metabolite clusters | 26 <sup>e</sup>                        | 26                     |
| Predicted no. of CAZymes           | 1,727 <sup>c</sup> (441 <sup>f</sup> ) | 451 <sup>c f</sup>     |

<sup>a</sup>Scaffolds for ArD2 Illumina assembly (GCA\_001630375.1).

<sup>b</sup>Contigs for ArME14 PacBio SMRT assembly.

<sup>c</sup>Differences in numbers likely due largely to different selection criteria.

<sup>d</sup>Secretome and effector predictions for ArD2 assembly using the same methods applied to ArME14.

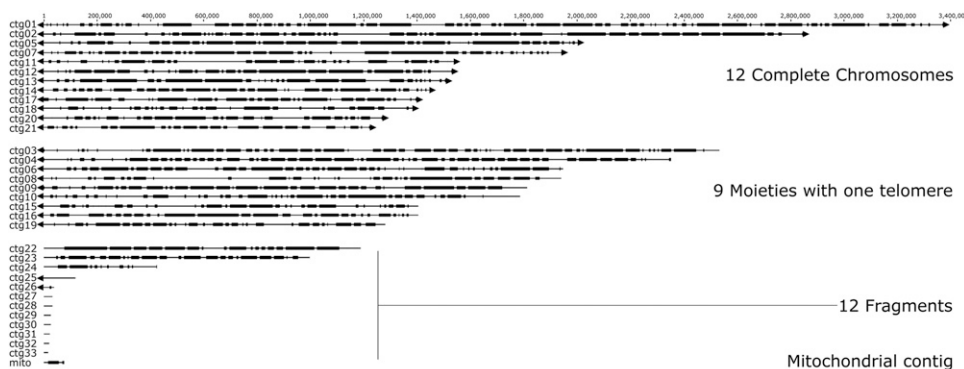
<sup>e</sup>Unknown prediction method for secondary metabolite clusters.

<sup>f</sup>CAZyme prediction using dbCAN2 meta server in this study.

AT-rich, repetitive DNA, between ArD2 and ArME14 can be explained by the more complete sequencing and assembly using PacBio sequencing. Distribution of GC content varies widely among the *Pleosporales* plant pathogenic fungi (Figure 3), with *Parastagonospora nodorum*, *Pyrenophora tritici-repentis* and *Zymoseptoria tritici* having mostly 50–55% GC content and the canola blackleg disease pathogen, *Leptosphaeria maculans*, having approximately one third of its genome as AT-rich DNA (Rouxel et al. 2011; Testa et al. 2016). *A. rabiei* has a similar GC content distribution to the barley pathogen, *Pyrenophora teres* f.sp *teres* (Figure 3 A). Size distributions for both the AT-rich and GC-balanced regions for ArME14 were highly variable, with average sizes of approximately 6,200 and 25,000 bp, respectively (Figure 3 B). Filamentous plant pathogen genomes tend to be characterized as having substantial proportions of repetitive and AT-rich sequence and their complement of genes includes a large number that encode secreted proteins. The genome architecture of ArME14 revealed by PacBio sequencing fits the “two-speed genome” model as proposed by Dong et al. (2015). The striking feature of this model, is that

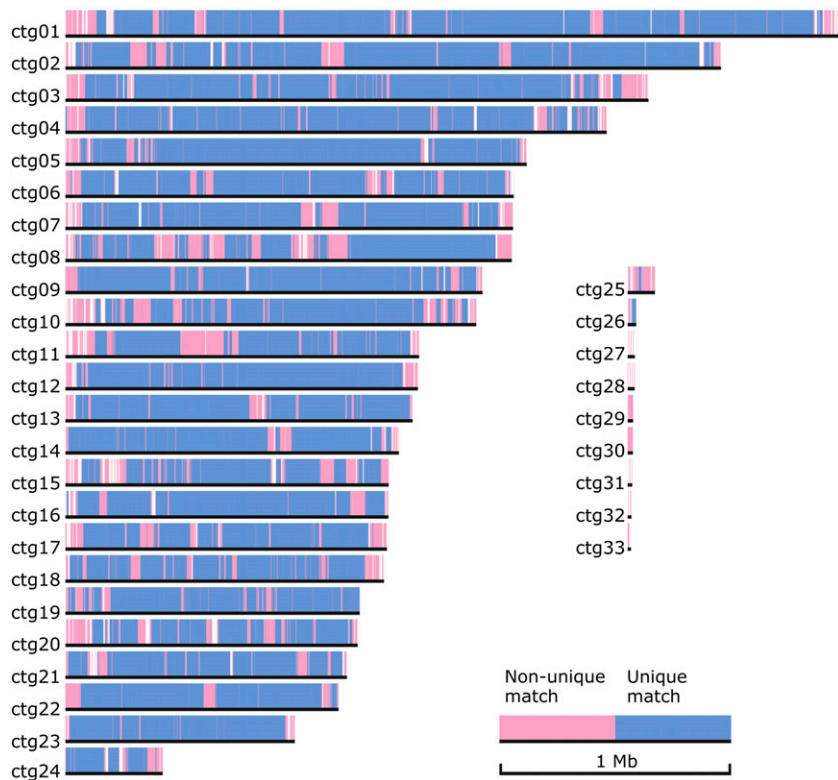
positive selection in genes located near repetitive DNA regions leads to higher rates of evolution in species for which genome architecture fits this model (Dong et al. 2015).

We predicted 11,257 protein coding genes in ArME14, 661 more than for ArD2, and again this discrepancy is likely a result of the different sequencing methods used and differences in the gene model prediction and annotation results. Using tBLASTn (Altschul et al. 1990), we identified 405 annotated ArME14 protein coding genes that were not found in the ArD2 genome assembly and almost all of these were located near contig ends or near annotated transposable element sequences. It is unclear whether the observed differences in the complement of genes between the two isolates is due to lack of sequence data for these regions in ArD2, difficulty in assembling such regions, or due to real deletions or insertions of gene-encoding sequence at these locations. Each of these possibilities can be explained by the AT-rich and repetitive nature of DNA sequence where these missing genes are located. BUSCO analysis indicated a substantial improvement for the sequencing and annotation for *A. rabiei* with only three missing and 12 fragmented



**Figure 1** Genome contigs for the reference assembly of *A. rabiei* ArME14, produced from PacBio SMRT sequencing with polishing using Illumina sequencing. Nuclear contigs are labeled ctg01 to ctg33 as archived in NCBI BioProject PRJNA510692 and the mitochondrial contig is labeled mito. Gene-dense regions of the genome are shown as dark-shaded blocks, joined by gene-sparse and interspersed repeat-rich regions indicated by thin lines. Telomeres are indicated in the figure by triangles at the ends of respective contigs.





**Figure 2** Alignment of *A. rabiei* ArD2 scaffolds to the 33 nuclear ArME14 contigs using NUCMER. Unique matches representing homologous nucleotide sequences between the two assemblies are indicated in blue, and repeat-rich nucleotide sequence that characterizes repetitive and AT-rich genomic regions are indicated by non-unique matches shown in red. Presumed non-assembled or absent sections from the ArD2 genome are represented by white space along each of the ArME14 reference contigs.

genes for ArME14, compared with 37 missing and 17 fragmented for ArD2 (Supplementary Figure: Figure\_S1).

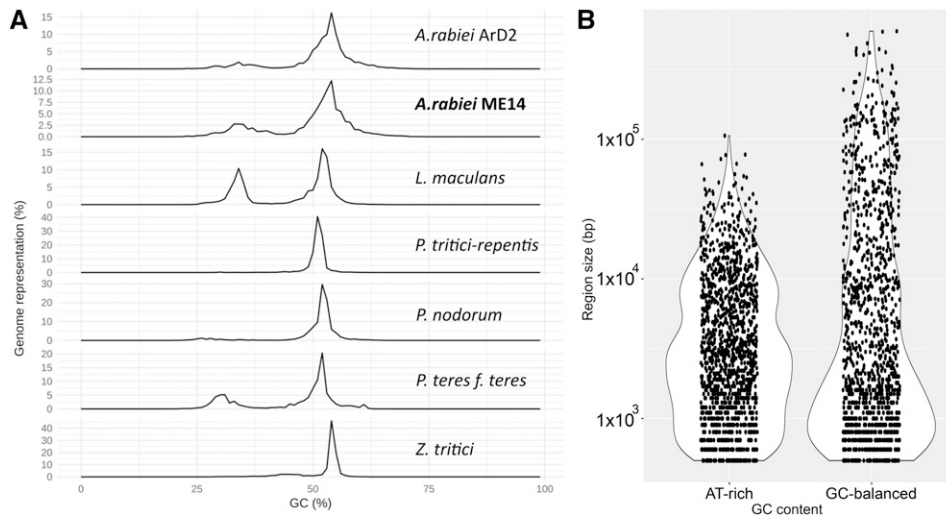
Effector proteins of plant pathogenic fungi are usually predicted based on the presence of a secretion signal, small protein size and a high proportion of cysteine residues (Jones *et al.* 2018). Therefore our first step in effector protein prediction was to determine the set of secreted proteins. SignalP v 5.0 predicted 1,145 secreted proteins for ArME14. Verma *et al.* (2016) predicted fewer secreted proteins for ArD2 (758). However, when we applied the same prediction method for secreted proteins as for ArME14, we found a similar number (1,111) of secreted proteins for ArD2, suggesting that the two genomes are highly similar with respect to their complement of secreted proteins. We compared the 1,145 ArME14 secreted proteins with the 1,111 sequences predicted for ArD2 using tBLASTn and found that 22 ArD2 proteins were not present in the ArME14 genome and 29 proteins from ArME14 that were not in ArD2 (Supplementary

File, File\_S2). Table 3 shows the number of putative effector proteins predicted from the total *A. rabiei* ArME14 proteome using different selection criteria with increasing stringency. For ArD2, 328 effectors were predicted and these were a large proportion of the secreted, non-Carbohydrate Active Enzymes (Verma *et al.* 2016). In contrast, our study used EffectorP v 2.0 (Sperschneider *et al.* 2016, 2018) as a more specific tool for fungal effector prediction. Using a mature protein size threshold of 25 KDa and EffectorP score threshold of 0.8, we nominated 39 protein sequences, designated PE01 to PE39 as putative effectors (PE). Full details of the 39 putative effector proteins are presented in the Supplementary File, File\_S4. Three of the ArME14 putative effectors were missing from the ArD2 proteome with only 36 ArD2 putative effectors being predicted using the same selection criteria as we used for ArME14. A subsequent tBLASTn search of the ArD2 assembly located one of these “missing” proteins, the ArME14 PE22 ortholog, as an un-annotated sequence in the ArD2 assembly.

**Table 2** Transposable element and repetitive DNA sequences from *A. rabiei* ArME14

| Class        | Type                   | Number of sequences | % of total sequences | Total nucleotides | % of total nucleotides | Average size |
|--------------|------------------------|---------------------|----------------------|-------------------|------------------------|--------------|
| I            | LTR                    | 780                 | 43                   | 3336780           | 54                     | 4278         |
|              | LINE                   | 177                 | 9.7                  | 450418            | 7.3                    | 2545         |
|              | LARD                   | 26                  | 1.4                  | 194939            | 3.2                    | 7498         |
|              | TRIM                   | 5                   | 0.3                  | 5668              | 0.1                    | 1134         |
|              | SINE                   | 2                   | 0.1                  | 1034              | 0.02                   | 517          |
| II           | TIR                    | 683                 | 38                   | 1820841           | 30                     | 2666         |
|              | Helitron               | 44                  | 2.4                  | 171920            | 2.8                    | 3907         |
|              | MITE                   | 7                   | 0.4                  | 4193              | 0.1                    | 599          |
| SSR          | No cat <sup>a</sup>    | 82                  | 4.5                  | 137974            | 2.2                    | 1682         |
|              | Host gene <sup>a</sup> | 4                   | 0.2                  | 8223              | 0.1                    | 2056         |
|              | SSR                    | 6                   | 0.3                  | 5102              | 0.1                    | 2056         |
| <b>Total</b> |                        | <b>1,816</b>        | <b>100</b>           | <b>6,137,092</b>  | <b>100</b>             |              |

<sup>a</sup>“No Cat” and “Host gene” are categories assigned by the PiRATE Galaxy server, and describe unclassified (no category) and potential host gene, respectively.



**Figure 3** Summary of genome compartmentalisation into GC-equilibrated and AT-rich regions of the *A. rabiei* genome assemblies. (A) GC content (%) distribution for *Dothideomycete* genome assemblies as calculated by OcculterCut (Testa *et al.* 2016). (B) Size distributions of AT-rich and GC-balanced regions for *A. rabiei* ArME14 (note logarithmic scale).

Putative effector genes PE 34 and PE36 in ArME14, were absent from the ArD2 nucleotide sequence. Both genes are located in highly repetitive regions of sub-telomeric DNA and may have been absent from ArD2 or not assembled correctly in the ArD2 Illumina genome assembly. From the set of ArD2 secreted proteins not found in ArME14, one was predicted to be an effector with mature protein molecular weight and EffectorP 2.0 score of 14.7 KDa and 0.64, respectively, although this was below our EffectorP threshold of 0.8. From the 29 ArME14 secreted proteins not in found ArD2, we predicted seven to be effectors with EffectorP score greater than 0.6, but only two having EffectorP scores above 0.8. These two proteins were PE34 and PE36 (Supplementary data) as discussed above. In the “two-speed genome” model, genes closely located to, or within highly repetitive sequence evolve at a higher rate with greater rates of positive selection (Oliver 2012; Grandaubert *et al.* 2014; Dong *et al.* 2015; Raffaele *et al.* 2015). This evolutionary process is illustrated in the case of seven small-secreted protein, avirulence effectors of *L. maculans*, where the corresponding genes are located in AT-rich regions of the *L. maculans* genome and display evidence of Repeat-Induced Point mutation (RIP) and positive selection in their sequences (Van de Wouw *et al.* 2010; Grandaubert *et al.* 2014). Similar evolutionary processes are likely to have shaped the pathogen-host relationship for *A. rabiei* and chickpea, and further insights about the molecular mechanisms of pathogenicity in this species will be uncovered through functional analysis of these predicted effector proteins.

The secondary metabolite cluster prediction tool, antiSMASH (Medema *et al.* 2011; Blin *et al.* 2019) predicted 26 clusters in both ArD2 and ArME14 (Table 3). Verma *et al.* (2016) similarly predicted 26 clusters for ArD2. The antiSMASH-predicted clusters in ArME14 matched clusters from the ArD2 genome assembly in almost all cases, with some BLAST hits spread across multiple ArD2 scaffolds. Notably, the NRPS/T1PKS cluster 10-1 on ArME14 contig 10 has a polyketide synthase gene (g5897) that was absent from the ArD2 assembly although other ortholog genes for the cluster were present. Details of the ArME14 secondary metabolite clusters are presented in the Supplementary File, File\_S4. Predicted clusters were homologous to characterized clusters designated for the biosynthesis of known fungal secondary metabolites including: cluster 16.2, melanin (Akamatsu *et al.* 2010), cluster 3.1, mellein (Chooi *et al.* 2015), and cluster 7.3, solanapyrone (Kim *et al.* 2015a,b). There were a further six clusters with characterized secondary metabolite homologs with

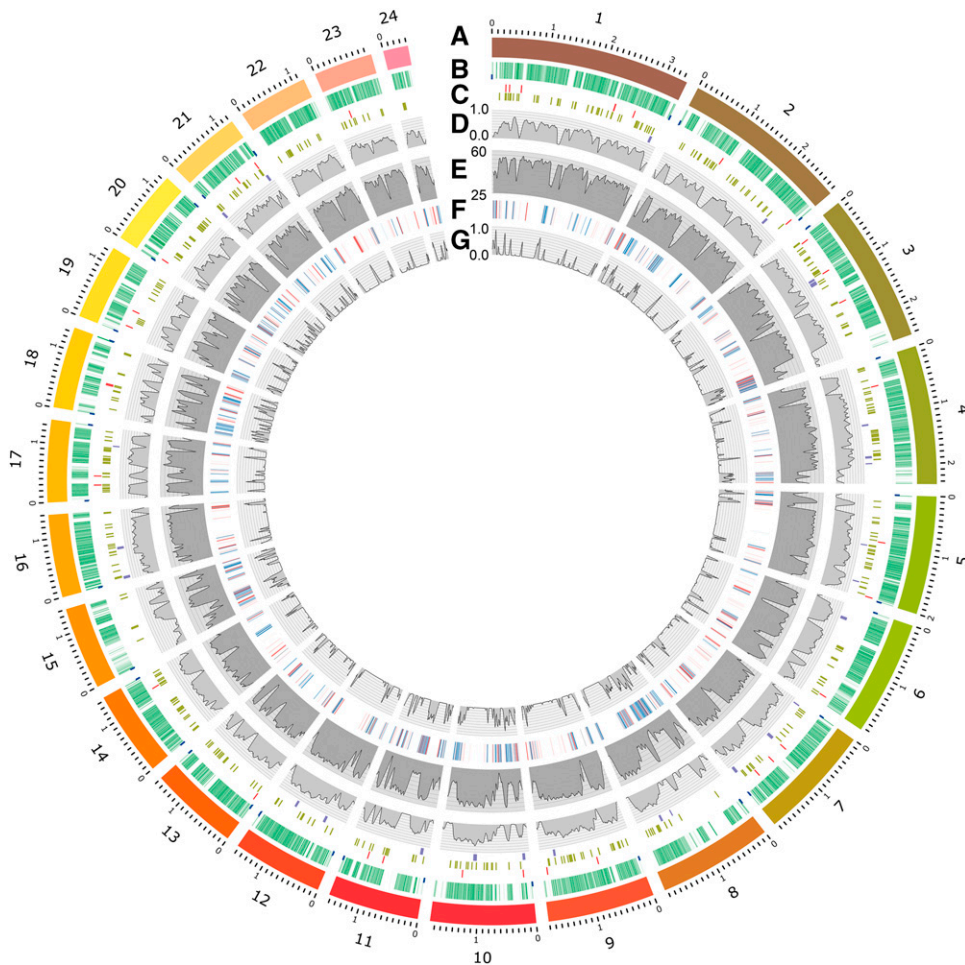
proposed roles in fungal physiology or reproduction and 17 other gene clusters putatively producing molecules with unknown structures and functions. It is likely that some of these gene clusters will have a role in producing novel molecules required for virulence and host specificity. Of the 26 secondary metabolite clusters, eight were located within sub-telomeric regions of the ArME14 assembly and two were bounded by highly repetitive regions populated by transposable elements. Similar to the predicted effectors, the presence of secondary metabolite clusters in repeat-rich regions of the genome confers mobility between species and rapid adaptation through processes such as Repeat-Induced Point mutation (RIP) (Hane and Oliver 2008; Fudal *et al.* 2009; Rouxel *et al.* 2011; Testa *et al.* 2016; Seidl and Thomma 2017). The features of the ArME14 genome are consistent with repetitive genome structure having played a role in the evolution and host adaptation of *A. rabiei*.

Carbohydrate-Active Enzymes (CAZymes) are a key feature of all fungi, and in plant pathogens these enzymes are essential for the

**Table 3 Summary of potential pathogenicity genome features, including: secondary metabolite clusters, predicted effector genes and CAZyme genes identified from the *A. rabiei* ArME14 genome assembly. Detailed tables are provided in Supplementary File, File\_S4**

| Class                                | Number |
|--------------------------------------|--------|
| <b>Putative effectors</b>            |        |
| EffP > 0.8, MW < 25KDa <sup>a</sup>  | 39     |
| EffP > 0.8, MW < 15KDa <sup>a</sup>  | 27     |
| EffP > 0.9, MW < 15KDa <sup>a</sup>  | 15     |
| <b>Secondary metabolite clusters</b> | 26     |
| T1PKS                                | 7      |
| T3PKS                                | 1      |
| NRPS                                 | 2      |
| NRPS-like                            | 7      |
| NRPS/NRPS-like - T1PKS               | 4      |
| Indole                               | 1      |
| Terpene                              | 4      |
| <b>CAZymes</b>                       | 451    |
| AA - Auxiliary activities            | 77     |
| CBM - Carbohydrate-binding module    | 3      |
| CE - Carbohydrate esterase           | 31     |
| GH - Glycoside hydrolase             | 227    |
| GT - Glycosyl transferase            | 82     |
| PL - Polysaccharide lyase            | 31     |

<sup>a</sup> mature protein MW.



**Figure 4** CIRCOS plot of key features of the *A. rabiei* ArME14 reference genome. Tracks labeled from outside: (A) The 26 largest contigs with size scale in Mb; (B) Annotated genes in scale (green) and telomeres (not to scale) (blue); (C) Locations (not to scale) of predicted effector genes (red), CAZyme genes (green), and predicted secondary metabolite clusters (purple and in scale); (D) Gene density with 20 Kb moving average window; (E) percent GC content with 50 Kb moving average window; (F) Transposable elements and repetitive DNA sequence regions. LINE, LTR, LARD, SINE and TRIM elements (blue), TIR, Helitron and MITE elements (red), SSRs (orange), No category, LTR/TIR, PLE/LARD and potential host gene (gray); (G) TE and repetitive DNA density with 20 Kb moving average window. An SVG version of this figure (Figure\_S2) is included in the Supplementary Data to enable closer inspection of genome features.

degradation of host plant polysaccharides for penetrating, colonizing and deriving nutrition from host tissues. Our CAZyme predictions from ArME14 produced 451 CAZyme sequences (Table 3), which is substantially fewer than the published number of 1,727 for ArD2 (Verma *et al.* 2016). Our search method identified only 441 CAZymes in ArD2. The dbCAN2 web server estimates of CAZyme number for *A. rabiei* are similar to those reported for other plant pathogenic fungi (Zhao *et al.* 2013). A total of 650 CAZymes were predicted by at least one of the tools and 451 to 650 is the likely range for the number of *A. rabiei* CAZymes. The main distinction of CAZyme complement among fungi is that necrotrophic fungal pathogens generally have a greater number (approx. 400-850) than biotrophs (approx. 170-320) (Zhao *et al.* 2013). The *A. rabiei* ArME14 genome has at least 450 and possibly up to 650 CAZymes, which is a similar number to those identified for other *Dothideomycete* genomes (Zhao *et al.* 2013). Fungal pathogens of dicots generally have an adapted set of CAZymes that are tailored to the types of carbohydrates found in dicot cell walls. Zhao *et al.* (Zhao *et al.* 2013) report that dicot pathogens generally have more polysaccharide lyases that degrade pectate and pectin (classes PL1 and PL3), which are more abundant in the cell walls of dicots than of monocots. In *A. rabiei* ArME14 there were nine PL1 CAZymes, which is similar to the average number reported for dicot pathogens and significantly greater than the average number of three PL1 enzymes for monocot pathogens (Zhao *et al.* 2013). In addition, *A. rabiei* ArME14 contained 12 pectin degrading polygalacturonases from GH28 where the average for dicot and monocot pathogens is

13 and 5, respectively (Zhao *et al.* 2013). Figure 4 summarizes the genome structure and locations of transposable and repetitive elements, putative effectors, CAZymes and secondary metabolite clusters in a CIRCOS plot. A plot of percent GC content in the CIRCOS format emphasizes partitioning of the genome into AT-rich gene-sparse, and GC-rich gene-dense sections. Notably 62% of the 39 predicted effector genes were located within 50 kb of repetitive regions and 23% were between 50-100 kb from the nearest repeat-rich region. Genome features are provided in a Supplementary genome feature file (Supplementary Data, File\_S3).

Transcription factors are an important feature of the *A. rabiei* genome and of the 381 identified in the ArD2 genome assembly, three were found using tBLASTn searches to be absent from ArME14 (KZM27601.1, KZM27726.1 and KZM27745.1). Functional annotation of predicted ArME14 proteins using InterProScan showed 126 proteins described as being transcription factors in addition to the 378 ArD2 transcription factor orthologs. Most of these were present in the ArD2 assembly but either they were not annotated as transcription factors or they were not annotated as protein-encoding genes for ArD2. We found three putative transcription factor genes in ArME14 for which there was no homologous DNA sequence in the ArD2 assembly. These were ArME14 g29, g427 and g4943, each described as containing fungal transcription factor domains. Putative transcription factor sequences from the comparisons between *A. rabiei* ArD2 and ArME14 are provided in the Supplementary Material, File\_S2.



Developing an understanding of the mechanisms of virulence of plant pathogens is critical to the effective control of plant disease in crop production. The *Pleosporales* order of filamentous fungi including *P. nodorum*, *P. tritici-repentis*, *Cochliobolus heterostrophus* and *P. teres* f. *teres* among others, have many common overarching features that govern their primary functions as plant pathogens. Notwithstanding the similarities in genome structure and function among these species, there are also differences in virulence genes and effectors that determine the very important phenomenon of host specialization in plant pathogens. Furthermore, regulation of gene expression is critical to the production of virulence factors and the interaction of pathogen and plant host (Verma *et al.* 2017). The publication of a near-complete, high-fidelity genome assembly for *A. rabiei* complements the previously published genome assembly (Verma *et al.* 2016) and provides the basis for further work in the field of chickpea ascochyta blight research.

## ACKNOWLEDGMENTS

This work was funded by the Australian Grains Research and Development Corporation (GRDC) research grants UMU00021 and UMU00022 at Murdoch University, and CUR00014 and CUR00023 at Curtin University. RMS acknowledges the Malaysian Ministry of Higher Education and Universiti Malaysia Terengganu for providing a scholarship. This work was supported by resources from the Pawsey Supercomputing Centre, Kensington, Western Australia and the National Computational Infrastructure (NCI) funded by the Australian Government. The authors gratefully acknowledge the contributions of Judith Lichtenzweig to the early conception and establishment of research projects CUR00014 and the pulse pathogen program of CUR00023, isolate selection and initial genome sequencing. Robert Syme is acknowledged for producing the PacBio genome assembly, genome annotation and NUCMER comparison.

## LITERATURE CITED

Akamatsu, H. O., M. I. Chilvers, W. J. Kaiser, and T. L. Peever, 2012 Karyotype polymorphism and chromosomal rearrangement in populations of the phytopathogenic fungus, *Ascochyta rabiei*. *Fungal Biol.* 116: 1119–1133. <https://doi.org/10.1016/j.funbio.2012.07.001>

Akamatsu, H. O., M. I. Chilvers, J. E. Stewart, and T. L. Peever, 2010 Identification and function of a polyketide synthase gene responsible for 1,8-dihydroxynaphthalene-melanin pigment biosynthesis in *Ascochyta rabiei*. *Curr. Genet.* 56: 349–360. <https://doi.org/10.1007/s00294-010-0306-2>

Alam, S. S., J. N. Bilton, A. M. Z. Slawin, D. J. Williams, R. N. Sheppard *et al.*, 1989 Chickpea blight: Production of the phytotoxins solanapyrones A and C by *Ascochyta rabiei*. *Phytochemistry* 28: 2627–2630. [https://doi.org/10.1016/S0031-9422\(00\)98054-3](https://doi.org/10.1016/S0031-9422(00)98054-3)

Almagro Armenteros, J. J., K. D. Tsigos, C. K. Sønderby, T. N. Petersen, O. Winther *et al.*, 2019 SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37: 420–423. <https://doi.org/10.1038/s41587-019-0036-z>

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic Local Alignment Search Tool. *J. Mol. Biol.* 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)

Armstrong, C. L., G. Chongo, B. D. Gossen, and L. J. Duczek, 2001 Mating type distribution and incidence of the teleomorph of *Ascochyta rabiei* (*Didymella rabiei*) in Canada. *Can. J. Plant Pathol.* 23: 110–113. <https://doi.org/10.1080/07060660109506917>

Aveskamp, M. M., J. de Gruyter, J. H. C. Woudenberg, G. J. M. Verkley, and P. W. Crous, 2010 Highlights of the *Didymellaceae*: A polyphasic approach to characterise *Phoma* and related pleosporalean genera. *Stud. Mycol.* 65: 1–60. <https://doi.org/10.3114/sim.2010.65.01>

Berlin, K., S. Koren, C. S. Chin, J. P. Drake, J. M. Landolin *et al.*, 2015 Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat. Biotechnol.* 33: 623–630. <https://doi.org/10.1038/nbt.3238>

Berthelier, J., N. Casse, N. Daccord, V. Jamilloux, B. Saint-Jean *et al.*, 2018 A transposable element annotation pipeline and expression analysis reveal potentially active elements in the microalga *Tisochrysis lutea*. *BMC Genomics* 19: 378. <https://doi.org/10.1186/s12864-018-4763-1>

Blin, K., S. Shaw, K. Steinke, R. Villebro, N. Ziemert *et al.*, 2019 antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47: W81–W87. <https://doi.org/10.1093/nar/gkz310>

Buchfink, B., C. Xie, and D. H. Huson, 2015 Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12: 59–60. <https://doi.org/10.1038/nmeth.3176>

Busk, P. K., B. Pilgaard, M. J. Lezyk, A. S. Meyer, and L. Lange, 2017 Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. *BMC Bioinformatics* 18: 214. <https://doi.org/10.1186/s12859-017-1625-9>

Chooi, Y., C. Krill, R. A. Barrow, S. Chen, R. Trengove *et al.*, 2015 An *In Planta*-Expressed Polyketide Synthase Produces (R)-Mellein in the Wheat Pathogen *Parastagonospora nodorum*. *Applied and Environmental Microbiology* 81: 177–186.

Coleman, M. J., M. T. McHale, J. Arnau, A. Watson, and R. P. Oliver, 1993 Cloning and characterisation of telomeric DNA from *Cladosporium fulvum*. *Gene* 132: 67–73. [https://doi.org/10.1016/0378-1119\(93\)90515-5](https://doi.org/10.1016/0378-1119(93)90515-5)

Derbyshire, M., M. Denton-Giles, D. Hegedus, S. Seifbarghi, J. Rollins *et al.*, 2017 The complete genome sequence of the phytopathogenic fungus *Sclerotinia sclerotiorum* reveals insights into the genome architecture of broad host range pathogens. *Genome Biol. Evol.* 9: 593–618. <https://doi.org/10.1093/gbe/evx030>

Dong, S., S. Raffaele, and S. Kamoun, 2015 The two-speed genomes of filamentous pathogens: Waltz with plants. *Curr. Opin. Genet. Dev.* 35: 57–65. <https://doi.org/10.1016/j.gde.2015.09.001>

Eddy, S. R., 1996 Hidden Markov models. *Curr. Opin. Struct. Biol.* 6: 361–365. [https://doi.org/10.1016/S0959-440X\(96\)80056-X](https://doi.org/10.1016/S0959-440X(96)80056-X)

Eddy, S. R., 1995 Multiple alignment using hidden Markov models. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 3: 114–120.

Ellinghaus, D., S. Kurtz, and U. Willhoft, 2008 LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* 9: 18. <https://doi.org/10.1186/1471-2105-9-18>

Faino, L., M. F. Seidl, E. Datema, G. C. M. van den Berg, A. Janssen *et al.*, 2015 Single-Molecule Real-Time Sequencing Combined with Optical Mapping Yields Completely Finished Fungal Genome. *MBio* 6. <https://doi.org/10.1128/mBio.00936-15>

Faino, L., M. F. Seidl, X. Shi-kunne, M. Pauper, G. C. M. van den Berg *et al.*, 2016 Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* 26: 1091–1100. <https://doi.org/10.1101/gr.204974.116>

Farman, M. L., 2007 Telomeres in the rice blast fungus *Magnaporthe oryzae*: The world of the end as we know it. *FEMS Microbiol. Lett.* 273: 125–132. <https://doi.org/10.1111/j.1574-6968.2007.00812.x>

Flutre, T., E. Duprat, C. Feuillet, and H. Quesneville, 2011 Considering transposable element diversification in *de novo* annotation approaches. *PLoS One* 6: e16526. <https://doi.org/10.1371/journal.pone.0016526>

Fondevilla, S., N. Krezdorn, B. Rotter, G. Kahl, and P. Winter, 2015 *In planta* Identification of Putative Pathogenicity Factors from the Chickpea Pathogen *Ascochyta rabiei* by *De novo* Transcriptome Sequencing Using RNA-Seq and Massive Analysis of cDNA Ends. *Front. Microbiol.* 6: 1–15. <https://doi.org/10.3389/fmicb.2015.01329>

Fudal, I., S. Ross, H. Brun, A.-L. Besnard, M. Ermel *et al.*, 2009 Repeat-Induced Point mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. *Mol. Plant Microbe Interact.* 22: 932–941. <https://doi.org/10.1094/MPMI-22-8-0932>

Grandaubert, J., R. G. T. Lowe, J. L. Soyer, C. L. Schoch, I. Fudal *et al.*, 2014 Transposable Element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species



- complex of fungal pathogens. *Biomed Cent. Genomics* 15: 891. <https://doi.org/10.1186/1471-2164-15-891>
- de Gruyter, J., M. M. Aveskamp, J. H. C. Woudenberg, G. J. M. Verkley, J. Z. Groenewald *et al.*, 2009 Molecular phylogeny of *Phoma* and allied anamorph genera: Towards a reclassification of the *Phoma* complex. *Mycol. Res.* 113: 508–519. <https://doi.org/10.1016/j.mycres.2009.01.002>
- Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler, 2013 QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29: 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Hamid, K., and R. N. Strange, 2000 Phytotoxicity of solanapyrones A and B produced by the chickpea pathogen *Ascochyta rabiei* (Pass.) Labr. and the apparent metabolism of solanapyrone A by chickpea tissues. *Physiol. Mol. Plant Pathol.* 56: 235–244. <https://doi.org/10.1006/pmpp.2000.0272>
- Han, Y., and S. R. Wessler, 2010 MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* 38: e199. <https://doi.org/10.1093/nar/gkq862>
- Hane, J. K., and R. P. Oliver, 2008 RIPCAL: A tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics* 9: 478. <https://doi.org/10.1186/1471-2105-9-478>
- Hoede, C., S. Arnoux, M. Moisset, T. Chaumier, O. Inizan *et al.*, 2014 PASTEC: An automatic transposable element classification tool. *PLoS One* 9: e91929. <https://doi.org/10.1371/journal.pone.0091929>
- Jones, D. A. B., S. Bertazzoni, C. J. Turo, R. A. Syme, and J. K. Hane, 2018 Bioinformatic prediction of plant – pathogenicity effector proteins of fungi. *Curr. Opin. Microbiol.* 46: 43–49. <https://doi.org/10.1016/j.mib.2018.01.017>
- de Jonge, R., H. P. van Esse, A. Kombrink, T. Shinya, Y. Desaki *et al.*, 2010 Conserved fungal LysM effector ECP6 prevents chitin-triggered immunity in plants. *Science* 329: 953–955. <https://doi.org/10.1126/science.1190859>
- Van Kan, J. A. L., J. H. M. Stassen, A. Mosbach, T. A. J. van der Lee, L. Faino *et al.*, 2017 A gapless genome sequence of the fungus *Botrytis cinerea*. *Mol. Plant Pathol.* 18: 75–89. <https://doi.org/10.1111/mpp.12384>
- Kim, W., J. J. Park, D. R. Gang, T. L. Peever, and W. Chen, 2015a A novel type pathway-specific regulator and dynamic genome environments of a solanapyrone biosynthesis gene cluster in the fungus *Ascochyta rabiei*. *Eukaryot. Cell* 14: 1102–1113. <https://doi.org/10.1128/EC.00084-15>
- Kim, W., C. Park, J. Park, H. O. Akamatsu, T. L. Peever *et al.*, 2015b Functional Analyses of the Diels-Alderase Gene sol5 of *Ascochyta rabiei* and *Alternaria solani* Indicate that the Solanapyrone Phytotoxins Are Not Required for Pathogenicity. *Mol. Plant Microbe Interact.* 28: 482–496. <https://doi.org/10.1094/MPMI-08-14-0234-R>
- König, S., L. W. Romoth, L. Gerischer, and M. Stanke, 2016 Simultaneous gene finding in multiple genomes. *Bioinformatics* 32: 3388–3395.
- Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne *et al.*, 2009 Circos: An information aesthetic for comparative genomics. *Genome Res.* 19: 1639–1645. <https://doi.org/10.1101/gr.092759.109>
- Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway *et al.*, 2004 Versatile and open software for comparing large genomes. *Genome Biol.* 5: R12.
- Leo, A. E., R. Ford, and C. C. Linde, 2015 Genetic homogeneity of a recently introduced pathogen of chickpea, *Ascochyta rabiei*, to Australia. *Biol. Invasions* 17: 609–623. <https://doi.org/10.1007/s10530-014-0752-8>
- Li, W., and A. Godzik, 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Lichtenzweig, J., E. Gamliel, O. Frenkel, S. Michaelido, S. Abbo *et al.*, 2005 Distribution of mating types and diversity in virulence of *Didymella rabiei* in Israel. *Eur. J. Plant Pathol.* 113: 15–24. <https://doi.org/10.1007/s10658-005-8914-2>
- Lombard, V., H. G. Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat, 2014 The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42: D490–D495. <https://doi.org/10.1093/nar/gkt1178>
- Medema, M. H., K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski *et al.*, 2011 AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39: W339–W346. <https://doi.org/10.1093/nar/gkr466>
- Mehmood, Y., P. Sambasivam, S. Kaur, J. Davidson, A. E. Leo *et al.*, 2017 Evidence and Consequence of a Highly Adapted Clonal Haplotype within the Australian *Ascochyta rabiei* Population. *Front. Plant Sci.* 8: 1029. <https://doi.org/10.3389/fpls.2017.01029>
- Moolhuijzen, P., P. T. See, J. K. Hane, G. Shi, Z. Liu *et al.*, 2018 Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC Genomics* 19: 1–23.
- Murray, G. M., and J. P. Brennan, 2012 *The current and potential costs from diseases of pulse crops in Australia*, Grains Research and Development Corporation, Canberra, Australia.
- Oliver, R., 2012 Genomic tillage and the harvest of fungal phytopathogens. *New Phytol.* 196: 1015–1023. <https://doi.org/10.1111/j.1469-8137.2012.04330.x>
- Peever, T. L., S. S. Salimath, G. Su, W. J. Kaiser, and F. J. Muehlbauer, 2004 Historical and contemporary multilocus population structure of *Ascochyta rabiei* (teleomorph: *Didymella rabiei*) in the Pacific Northwest of the United States. *Mol. Ecol.* 13: 291–309. <https://doi.org/10.1046/j.1365-294X.2003.02059.x>
- Lo Presti, L., D. Lanver, G. Schweizer, S. Tanaka, L. Liang *et al.*, 2015 Fungal Effectors and Plant Susceptibility. *Annu. Rev. Plant Biol.* 66: 513–545. <https://doi.org/10.1146/annurev-arplant-043014-114623>
- Price, A. L., N. C. Jones, and P. A. Pevzner, 2005 *De novo* identification of repeat families in large genomes. *Bioinformatics* 21: i351–i358. <https://doi.org/10.1093/bioinformatics/bti1018>
- Raffaele, S., R. A. Fairer, L. M. Cano, D. J. Studholme, M. Thines *et al.*, 2015 Genome Evolution Following Host Jumps in the Irish Potato Famine. *Genome Biol. Evol.* 330: 1540–1543.
- Rehmeyer, C., W. Li, M. Kusaba, Y.-S. Kim, D. Brown *et al.*, 2006 Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. *Nucleic Acids Res.* 34: 4685–4701. <https://doi.org/10.1093/nar/gkl588>
- Rouxel, T., J. Grandaubert, J. K. Hane, C. Hoede, A. P. van de Wouw *et al.*, 2011 Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat. Commun.* 2: 202. <https://doi.org/10.1038/ncomms1189>
- Savojarjo, C., P. L. Martelli, P. Fariselli, and R. Casadio, 2018 DeepSig: Deep learning improves signal peptide detection in proteins. *Bioinformatics* 34: 1690–1696. <https://doi.org/10.1093/bioinformatics/btx818>
- Schechtman, M. G., 1990 Characterization of telomere DNA from *Neurospora crassa*. *Gene* 88: 159–165. [https://doi.org/10.1016/0378-1119\(90\)90027-O](https://doi.org/10.1016/0378-1119(90)90027-O)
- Seidl, M. F., and B. P. H. J. Thomma, 2017 Transposable Elements Direct The Coevolution between Plants and Microbes. *Trends Genet.* 33: 842–851. <https://doi.org/10.1016/j.tig.2017.07.003>
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Smit, A. F., R. Hubley, and P. Green, 2013 RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
- Sperschneider, J., P. N. Dodds, D. M. Gardiner, K. B. Singh, and J. M. Taylor, 2018 Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol. Plant Pathol.* 19: 2094–2110. <https://doi.org/10.1111/mpp.12682>
- Sperschneider, J., D. M. Gardiner, P. N. Dodds, F. Tini, L. Covarelli *et al.*, 2016 EffectorP: Predicting fungal effector proteins from secretomes using machine learning. *New Phytol.* 210: 743–761. <https://doi.org/10.1111/nph.13794>
- Stanke, M., O. Schöffmann, B. Morgenstern, and S. Waack, 2006 Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7: 62. <https://doi.org/10.1186/1471-2105-7-62>
- Stanke, M., R. Steinkamp, S. Waack, and B. Morgenstern, 2004 AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32: W309–W312. <https://doi.org/10.1093/nar/gkh379>

- Syme, R. A., A. Martin, N. A. Wyatt, J. A. Lawrence, M. J. Muria-Gonzalez *et al.*, 2018 Transposable element genomic fissuring in *Pyrenophora teres* is associated with genome expansion and dynamics of host-pathogen genetic interactions. *Front. Genet.* 9: 130. <https://doi.org/10.3389/fgene.2018.00130>
- Tan, K., and R. P. Oliver, 2017 Regulation of proteinaceous effector expression in phytopathogenic fungi. *PLoS Pathog.* 13: e1006241. <https://doi.org/10.1371/journal.ppat.1006241>
- Testa, A. C., R. P. Oliver, and J. K. Hane, 2016 OcculterCut: A comprehensive survey of AT-rich regions in fungal genomes. *Genome Biol. Evol.* 8: 2044–2064. <https://doi.org/10.1093/gbe/evw121>
- Di Tommaso, P., M. Chatzou, E. W. Floden, P. P. Barja, E. Palumbo *et al.*, 2017 Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* 35: 316–319. <https://doi.org/10.1038/nbt.3820>
- Trapero-Casas, A., and W. J. Kaiser, 1992 Development of *Didyella rabiei*, the teleomorph of *Ascochyta rabiei*, on chickpea straw. *Phytopathology* 82: 1261–1266. <https://doi.org/10.1094/Phyto-82-1261>
- Verma, S., R. K. Gazara, S. Nizam, S. Parween, D. Chattopadhyay *et al.*, 2016 Draft genome sequencing and secretome analysis of fungal phytopathogen *Ascochyta rabiei* provides insight into the necrotrophic effector repertoire. *Sci. Rep.* 6: 24638. <https://doi.org/10.1038/srep24638>
- Verma, S., R. K. Gazara, and P. K. Verma, 2017 Transcription Factor Repertoire of Necrotrophic Fungal Phytopathogen *Ascochyta rabiei*: Predominance of MYB Transcription Factors As Potential Regulators of Secretome. *Front. Plant Sci.* 8: 1037. <https://doi.org/10.3389/fpls.2017.01037>
- Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.*, 2014 Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9: e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Wenke, T., T. Dobel, T. R. Sorensen, H. Junghans, B. Weisshaar *et al.*, 2011 Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. *Plant Cell* 23: 3117–3128. <https://doi.org/10.1105/tpc.111.088682>
- Wickham, H., 2016 *ggplot2 Elegant Graphics for Data Analysis*. Springer, New York.
- Winnenburg, R., T. Baldwin, M. Urban, C. Rawlings, J. Kohler *et al.*, 2006 PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.* 34: D459–D464. <https://doi.org/10.1093/nar/gkj047>
- Van de Wouw, A. P., A. J. Cozijnsen, J. K. Hane, P. C. Brunner, B. A. McDonald *et al.*, 2010 Evolution of linked avirulence effectors in *Lep-tosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. *PLoS Pathog.* 6: e1001180. <https://doi.org/10.1371/journal.ppat.1001180>
- Xin, Z., and J. Chen, 2012 A high throughput DNA extraction method with high yield and quality. *Plant Methods* 8: 26. <https://doi.org/10.1186/1746-4811-8-26>
- Yang, L., and J. L. Bennetzen, 2009 Structure-based discovery and description of plant and animal Helitrons. *Proc. Natl. Acad. Sci. USA* 106: 12832–12837. <https://doi.org/10.1073/pnas.0905563106>
- Zhang, H., T. Yohe, L. Huang, S. Entwistle, P. Wu *et al.*, 2018 DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46: W95–W101. <https://doi.org/10.1093/nar/gky418>
- Zhao, Z., H. Liu, C. Wang, and J. R. Xu, 2013 Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* 14: 274. <https://doi.org/10.1186/1471-2164-14-274>

Communicating editor: S. Smith