Research

Phenotypic and Genotypic Diversity of *Puccinia sorghi* in Eastern Australia: Implications for Maize Breeding Programs

Aurelie Quade[1,†] | Robert F. Park[2] | Benjamin J. Stodart[3] | Yi Ding[2] | Peng Zhang[2] | Mike Thang[4] | Barsha Poudel[1] | Niloofar Vaghefi[1] | Gavin J. Ash[1,3] |

[1] Centre for Crop Health, Institute for Life Sciences and the Environment, University of Southern Queensland, Toowoomba, Queensland, 4350, Australia

[2] Plant Breeding Institute, School of Life and Environmental Sciences, University of Sydney, Cobbitty, New South Wales, 2570, Australia

[3] Graham Centre for Agricultural Innovation (Charles Sturt University and NSW Department of Primary Industries), School of Agricultural and Wine Sciences, Charles Sturt University, Wagga Wagga, New South Wales, 2678, Australia

[4] QCIF Facility for Advanced Bioinformatics, Institute for Molecular Bioscience, The University of Queensland, Brisbane, 4072, Australia

† Corresponding author: A. Quade; Aurelie.Quade@usq.edu.au

Accepted for publication 7 March 2023.

Abstract

Common rust of maize (CR), caused by *Puccinia sorghi* Schwein., is found in all maize growing regions across the globe. In recent years, an increase in the incidence and severity of CR has been observed in Australia. The challenges posed by a diverse population of *P. sorghi* to maize breeding programs and the strategies required to deploy durable resistance were explored. Preliminary studies of diversity were conducted by phenotyping the five most diverse isolates against maize lines carrying no resistance genes and one to four resistance genes. The results showed that further genotyping studies were required. Short reads of 17 CR isolates were collected from an array of agroecological regions in Eastern Australia, and these were mapped back against the hybrid assembly of Isolate 1. A hybrid assembly between short (Illumina paired-end) and long (PacBio) reads was conducted in MaSuRCA of the first isolate collected. A draft genome of 145,963,792 bp was obtained after QUAST analysis, with a total length of 145,815,044 bp (when removing base pairs below ≥1,000). The prediction from the BRAKER pipeline identified 20,438 gene models. All 17 isolates sequenced represented unique genotypes, and no clones were detected. The 17 isolates grouped into three clusters. The diversity observed within the population of *P. sorghi* in Australia suggests that monogenic resistance might not be durable.

Keywords: fungal pathogen diversity, genome assembly, hybrid assembly, long reads, Pucciniales, rust, short reads, *Zea mays*

Funding

Funding was provided by the University of Southern Queensland (453040) and Grains Research and Development Corporation (9175942).

e-Xtra: Supplementary material is available online.

The author(s) declare no conflict of interest.

Puccinia sorghi Schwein., the causal agent of maize common rust (CR), belongs to the Pucciniales. Most species within this large group of biotrophic fungi have evolved to have narrow host ranges. To survive and reproduce, rust fungi must infect living plant tissues and colonize plant cells via the production of haustoria. To adapt to the constraints imposed by a biotrophic lifestyle, rusts possess a genome with great plasticity (Aime et al. 2017). A key contributor to genetic diversity of rust fungi is their complex life cycle, which may comprise up to five different spore stages: urediniospores, teliospores,

Copyright © 2023 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

basidiospores, spermacia, and aeciospores (Bettgenhaeuser et al. 2014). Basidiospores and aeciospores are produced through meiosis; more specifically, though, aeciospores are dikaryotic and the product of plasmogamy. The recombination of alleles in the pathogen population is achieved at the aecial stage on the alternate host *Oxalis* spp. for heteroecious rust species such as *P. sorghi*. Exception aside, clonally produced urediniospores, at the uredinial stage, cause economic losses of agricultural crops and are the focus of this study.

The success of rust fungi as pathogens comes from their high rate of sporulation, which maintains genome plasticity, and the rapid spread of new virulence over long distances (Roelfs and Bushnell 1985). A study of the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* estimated that at an infection rate of 50 pustules per wheat tiller in the field, approximately 10^{12} spores per hectare per day were produced (Rowell and Roelfs 1971). Potentially, one in every 100,000 to 10 million cells/spores will carry mutations beneficial to the pathogen, as determined in other pathosystems involving fungi (Stam and McDonald 2018). On any given day, at similar infection levels, every infected hectare may contain about 225,000 spores carrying mutations beneficial to the pathogen. However, rust species genomic diversity is more likely to result in mutations occurring in every 100,000 spores, resulting in approximately 10^7 spores per hectare per day with beneficial mutations (Stam and McDonald 2018).

The virulence of rust pathogens has been studied by phenotyping the responses of host plants, carrying known resistance genes, when challenged with specific rust isolates. This information has guided resistance breeding programs by allowing for selection of effective resistance genes (Park 2015). In the maize-CR pathosystem, changes in virulence have led to known resistance genes being overcome in the United States (Groth et al. 1992), South America (Darino et al. 2016), South Africa (Fato 2000), and Australia (Quade et al. 2021). Preemptive breeding studies have contributed to the identification of future virulence and/or future virulence combinations based on the data provided by existing virulence profiles. However, without such data, these studies are limited to examining virulence already present in the pathogen population. The potential for pathogens to develop new virulence is often linked to the extent of their genetic diversity. Studies of rust species have determined that intricate genetic features are involved to ensure successful infection and access to host plant resources (Barsoum et al. 2019). The genomes of rust fungi sequenced to date range from 90 Mbp to 1 Gbp in size, and over 50% of their large genome is repetitive in nature, as rust urediniospores carry two nuclei (Miller et al. 2018; Tobias et al. 2020). Rust genomes contain a large number of transposable elements (TEs); for example, 33% of the genome of *P. sorghi* consists of TEs (Rochi et al. 2018). Although the role of TEs in creating new virulence specificities in *P. sorghi* is not well understood, the ability to shift sections of DNA certainly plays a role in the emergence of new pathotypes (Rochi et al. 2018). The relationship between pathogenicity and specific TEs has been documented in other fungal plant pathogens, such as *Botrytis cinerea* (Porquier et al. 2021).

The high number of repetitive regions in rust genomes presents difficulties in de novo assemblies using short-read sequencing only, as errors such as high fragmentation, misassembly, and merging of two distinct haplotype sequences are likely to occur. Rochi et al. (2018) generated an ab initio draft genome sequence for *P. sorghi*, which was followed by a short-read comparative genome analysis to improve the predictive secretomes based on three wheat rust pathogens and the maize boil smut pathogen *Ustilago maydis.* The limitations of short-read based assemblies can be mitigated through long-read sequencing, which significantly improves the assembly of repeat-rich genomes. However, single-molecule real-time long-read sequencing possesses a higher error rate compared with short-read sequencing. Potential errors in the genome structure created with the long reads can be corrected using high-quality short reads, improving the quality of the genome assembly (De Maio et al. 2019). The cost of long-read sequencing remains high; therefore, a limited number of isolates can be processed, whereas short-read sequencing allows for the sequencing of a relatively large number of samples. A comprehensive genome assembly with long reads, with errors corrected with short reads to obtain a hybrid assembly of *P. sorghi*, has not been conducted. The genome of *P. sorghi* available on NCBI uses short reads only and is highly fragmented (Rochi et al. 2018).

In recent years, Australia has experienced an increase in the incidence and severity of CR. This is due to a lack of resistance in Australian commercial maize hybrids (Quade et al. 2021) and favorable environmental conditions. The aim of this research was to gain an understanding of the evolutionary potential for *P. sorghi* populations in Australia. It was hypothesized that (i) phenotyping of the most varied isolates in terms of hosts and agroecological region of collection was a suitable preliminary study to confirm or refute virulence diversity, (ii) genotyping a randomly selected set of isolates would reveal the presence of population subdivision based on agroecological region of collection, and (iii) the research would produce an improved hybrid genome assembly for *P. sorghi*. Specifically, we explored the challenges a diverse population of *P. sorghi* would pose to maize breeding programs and the strategies that would be required to deploy durable resistance.

MATERIALS AND METHODS

Phenotyping of a diverse set of isolates on a maize differential set

The focus of this experiment was to determine the overall effect of each isolate over time on a set of maize lines carrying a range of resistance genes. The maize lines in the differential set included genotypes carrying no known resistance gene and up to four different resistance genes. Isolates causing infection levels between 0 and 30% were considered to possess no to low virulence, between 30 and 60% moderate virulence, and greater than 60% high virulence.

Virulence patterns were assessed by inoculating 3-week-old seedlings carrying known resistance genes with the following five isolates. The five most diverse isolates, single pustule Isolates 1, 9, 24, 31, and 73, were arbitrarily selected from a collection of 89 isolates based on their diversity in terms of geographic origin (Victoria versus Queensland), year of collection (2018 versus 2019), and host (either maize or sweet corn) (Table 1). The single pustule increase, maize growth, and inoculation were conducted as described in Quade et al. (2021). The maize lines tested either carried a single resistance gene, *Rp1_B*, *Rp1_C*, *Rp1_E*, *Rp1_F*, *Rp1_G*, *Rp1_I*, *Rp1_J*, *Rp1_K*, *Rp1_L*, *Rp1_M*, *Rp3_D*, *Rp4_A*, *Rp5*, *Rp7*, *Rp8_A*, *Rp8_B*, or *Rpp9* (conferring resistance to *P. polysora* Underw.), or multiple *Rp* genes, *RpFJ* (*Rp1_F* and *Rp1_J*), *Rp5D* (*Rp5* and *Rp1_D*), *RpDGJ* (*Rp1_D*, *Rp1_G*, and *Rp1_J*), and *Rp5GCJ* (*Rp5*, *Rp1_G*, *Rp1_C*, and *Rp1_J*), all against unknown backgrounds. Three sweet corn lines were also used and carried no resistance genes, *Rp1_D* or *RpGJF* (*Rp1_G*, *Rp1_J*, and *Rp1_F*) (Table 2). Infection scores were recorded every 48 h from 6 days post inoculation (DPI) until 16 DPI.

The assessment of the resistance of the *Rp* lines was conducted in a split-plot design, blocked by isolates and maize lines as the plot. When experimental constraints prevented replication, the whole experiment was replicated two or three times over time.

The number of replications was determined based on the degrees of freedom (df) at the residual level, where 15 df and over were considered satisfactory.

The data were treated as continuous, and disease severity scores were recorded as the percentage of diseased area in the range of 0, 1, 5, and 10%, followed by 10% increments to 100%. NA was given to a completely dead leaf or a missing plant.

Quantitative data were analyzed using ASreml v.4 in RStudio v.1.2.5033 by fitting a linear mixed model (Butler et al. 2018; RStudio 2018). Continuous data received a logic transformation (Welham et al. 2004), and data reported in this paper were back-transformed. Differences in predicted values were reported to be significant when a P value less than 0.05 was recorded, after which post hoc analysis using Tukey's test was applied. The plot of the residuals and the test for normal distribution of the residuals were satisfactory in all the analyses of all experiments.

Fungal samples, DNA extraction, and sequencing

From a collection of 89 *P. sorghi* samples collected between 2018 and 2020 from naturally occurring populations (Quade et al. 2021), Isolate 1 was selected for PacBio sequencing, as it was the first isolate collected (in 2018), and it was hypothesized that changes within the pathogen population between 2018 and 2019 and location would be recorded. DNA was extracted from 75 mg of ungerminated urediniospores using a modified CTAB procedure (Schwessinger et al. 2018). The method described by Vaser et al. (2017) was used to remove impurities and low-molecular-weight DNA. A 20 K library was prepared and sequenced on PacBio RSII platform, using one single-molecule real-time cell at Novogen (HK).

The first isolate collected, Isolate 1, and 16 other isolates were sequenced using the Illumina Novaseq (S4) PE150 platform to represent a wide range of hosts and locations (Table 1). All isolates were established from samples collected in 2018, with the exception of Isolate 73, which originated from a sample collected in 2019. Isolates 43_1 and 43_2 were established from the same leaf sample from maize in Queensland but from different pustules (Table 1). Prior to Illumina short-read sequencing, all 17 samples were purified by increasing urediniospores from a single pustule. DNA from 0.10 mg of urediniospores was extracted following a

TABLE 2

Maize differential set entry name and corresponding resistance *Puccinia* (*Rp*) genes

Maize line entry	Resistance gene
E8[a]	NA
E9[a]	*Rp1_D*
E10[a]	*Rp1_G*, *Rp1_J*, *Rp1_F*
E11	*Rp1_B*
E12	*Rp1_C*
E13	*Rp1_E*
E14	*Rp1_F*
E15	*Rp1_G*
E16	*Rp1_I*
E17	*Rp1_J*
E18	*Rp1_K*
E19	*Rp1_L*
E20	*Rp1_M*
E21	*Rp3_D*
E22	*Rp4_A*
E23	*Rp5*
E24	*Rp7*
E25	*Rp8_A*
E26	*Rp8_B*
E27	*Rpp9*
E28	*Rp1_F*, *Rp1_J*
E29	*Rp5*, *Rp1_D*
E30	*Rp1_D*, *Rp1_G*, *Rp1_J*
E31	*Rp5*, *Rp1_C*, *Rp1_G*, *Rp1_J*

[a] Resistance genes carried in sweet corn; all other genes carried in maize.

TABLE 1

Details of *Puccinia sorghi* isolates used in the experiment

Isolate[a]	Collection date	Region	State[b]	Agroecological region climate	Latitude	Longitude	BBCH[c]	Host[d]	Rust score (infection type)[e]
1[f]*	**15/02/2018**	**Darling Downs**	**QLD**	**Subhumid, subtropical slopes and plains**	**−27.509568**	**151.389859**	**18**	**Maize**	**NA**
9*	**1/03/2018**	**East Gippsland**	**VIC**	**Wet temperate coast**	**−37.722071**	**148.414755**	**83**	**Sweet corn**	**9**
12	1/03/2018	East Gippsland	VIC	Wet temperate coast	−37.722071	148.414755	79	Sweet corn	5
13	1/03/2018	East Gippsland	VIC	Wet temperate coast	−37.722071	148.414755	80	Sweet corn	4
15	1/03/2018	East Gippsland	VIC	Wet temperate coast	−37.722071	148.414755	79	Sweet corn	9
24*	**27/03/2018**	**Southern Downs**	**QLD**	**Wet subtropical coast**	**−28.046435**	**151.962135**	**73**	**Maize**	**7**
27	27/03/2018	Southern Downs	QLD	Wet subtropical coast	−28.046435	151.962135	73	Maize	7
31*	**28/03/2018**	**Lockyer Valley**	**QLD**	**Wet tropical coast and tableland**	**−27.431075**	**152.584398**	**79**	**Sweet corn**	**2**
39	27/03/2018	Darling Downs	QLD	Subhumid, subtropical slopes and plains	−27.641913	151.869093	83	Maize	2
43_1	29/03/2018	Darling Downs	QLD	Subhumid, subtropical slopes and plains	−27.751234	151.447163	79	Maize	3
43_2	29/03/2018	Darling Downs	QLD	Subhumid, subtropical slopes and plains	−27.751234	151.447163	79	Maize	3
51	30/04/2018	East Gippsland	VIC	Wet temperate coast	−37.724023	148.433945	83	Unknown	7
55	6/05/2018	North Burnett	QLD	Wet subtropical coast	−24.792465	151.123412	87	Maize	4
56	25/05/2018	South East QLD	QLD	Wet subtropical coast	−27.375303	152.443875	87	Maize	7
57	22/05/2018	South East QLD	QLD	Wet subtropical coast	−27.419467	152.470468	87	Maize	7
58	22/05/2018	South East	QLD	Wet subtropical coast	−27.419467	152.470468	87	Maize	5
73*	**9/05/2019**	**East Gippsland**	**VIC**	**Wet temperate coast**	**−37.722071**	**148.414755**	**89**	**Maize**	**3**

[a] 43_1 and 43_2 were retrieved from the same leaf sample, but from different pustules.
[b] QLD: Queensland, VIC: Victoria.
[c] BBCH: Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie (codifies key phenological developmental stages of agricultural plants).
[d] Maize is classified as an agricultural crop, whereas sweet corn is classified as a horticultural crop.
[e] Disease scores taken in the field ranged from 1 to 9, where 1 was no disease and 9 was dead.
[f] * and bold indicate phenotyped isolates.

CTAB extraction method as described by Bailey et al. (2013). All DNA samples were concentrated and purified. DNA concentration was quantified using Qubit 3 Fluorometer (Thermo Fisher Scientific) and a NanoDrop ND-1000 (Thermo Fisher Scientific). Genomic DNA was sent to Novogen for library preparation and Illumina sequencing at 8 Gb of raw data per sample.

Short read preparation and assembly

The raw Illumina paired-end reads were uploaded to the publicly available Galaxy Platform (Afgan et al. 2018). The quality of the raw reads was checked using FastQC v.0.72 (Andrews 2010) and trimmed in Trimmomatic v.0.36.6 (Bolger et al. 2014) with parameters for paired-end reads, using a sliding window of 4 bp and average quality of 30.

Long-read DNA sequencing and de novo assembly

A hybrid genome of Isolate 1 was assembled de novo using two pipelines, MaSuRCA v.4.0.3 (Zimin et al. 2013) and Flye v.2.9 (Kolmogorov et al. 2019). The raw reads of PacBio and Illumina were assembled in MaSuRCA v. 4.0.3 with two rounds of polishing with POLCA (POLishing by Calling Alternatives). The preassembly polishing for Flye was conducted using FMLRC v.2 (Wang et al. 2018), followed by Flye assembly with default parameters in Galaxy v.2.8.3 (Lin et al. 2016). SAM files were created with BWA-mem v.0.7.17.2 and post-assembly polishing using RACON Galaxy v.1.4.20 (Vaser et al. 2017), with default parameters. Genome completeness was assessed by identifying Benchmarking Universal Single-Copy Orthologs (BUSCOs) based on the Basidiomycota_odb9-29222 database using BUSCO v.3 (Seppey et al. 2019).

Annotation of hybrid assembly

For the hybrid assembly of Isolate 1, all steps leading to annotation were performed de novo in GenSAS (Humann et al. 2019). Families of TEs were identified using RepeatModeler v.1.0.11 (Smit and Hubley 2008). Regions of all assembled contigs possessing similarities between sequences were identified in BLAST (NCBI Resource Coordinators 2016). Sequences from rRNA and tRNA were identified using RNAmmer v.1.2 (Lagesen et al. 2007) and tRNAScan v.2.0 (Chan and Lowe 2019). Eukaryotic genes were predicted in GeneMarkES v.4.33 (max contig = 5,000,000 bp, min contig = 5,000 bp, and max gap = 5,000 bp) (Borodovsky and Lomsadze 2011), and a consensus of gene structures was obtained based on ab initio gene prediction and protein from GeneMarkES and BLAST.

Orthology to the Basidiomycota_odb9-29222 database was measured with BUSCO v.3 for completeness and assessment of the gene set. Protein-coding genes found with GeneMarkES were aligned with DIAMOND v.0.9.22 (Buchfink et al. 2015) against the NCBI refseq fungi (Protein) (O'Leary et al. 2016). Another pipeline was followed for gene prediction: BRAKER v2.1.6 (Hoff et al. 2019) using ProtHint and GeneMark-EP (Brůna et al. 2020) with an in-house modified OrthoDB fungal protein database (Kriventseva et al. 2018), followed by fungal training with AUGUSTUS (Stanke and Morgenstern 2005). Lastly, a gene prediction with contig evidence files from *P. striiformis*, BioSample SAMN00025370 (Cantu et al. 2011) was performed in GenSAS.

Variant calling and analysis of genetic distance

Adapter- and quality-filtered Illumina paired-end reads of the 17 sequenced isolates were independently mapped to the hybrid assembly of Isolate 1 using BWA-mem v.0.7.17.2. After coordinate-based sorting, duplicates were removed in sortSAMtools view v.2.0.3 galaxy (Li and Durbin 2009). Alignments from all isolates were merged, and variant calling was conducted in FreeBayes parallel v.1.3.5 (Garrison and Marth 2012; Tange 2011) using default settings (*–min-mapping-quality* 30 *–min-base-quality* 20 *–pvar* 0.0001). The VCF file was filtered using VCFfilter v.1.0.0 (Garrison 2015) using parameters DP > 10 and QUAL > 20. The identified single-nucleotide polymorphisms were further filtered in TASSEL v.5.2.75 (Bradbury et al. 2007) for a minimum allele frequency of 5% and no missing data in order to maintain informative loci. This resulted in the identification of 928,211 single-nucleotide polymorphisms for the 17 isolates. Due to the small number of isolates ($n = 17$), many routine population analyses of genetic diversity and population differentiation could not be carried out; however, the genetic distance among isolates and population structuring were investigated as follows. Nei's genetic distance was used to calculate pairwise comparisons of the genetic distance between isolates to obtain a distance matrix. The matrix was then used via the UPGMA method with 1,000 bootstrap replicates using the *aboot* function of the *poppr* v.2.9.1 package in R to generate a dendrogram (Kamvar et al. 2014). SplitsTree v4.18.3 (Moulton 2003), which is more suitable to depict genetic relationships in the presence of recombination, was used to construct a neighbor-net network to visualize the genetic relationships among isolates with 1,000 bootstraps. The existence of an underlying population structure without a priori population assignment was investigated using discriminant analysis of principal components (DAPC) in R package *adegenet* v.2.1.5 (Jombart 2008). The optimal number of clusters was determined using the function *find.clusters*, and DAPC was used to assign individuals to clusters, retaining the number of principal components encompassing 86% of the cumulative variance.

RESULTS

Phenotypic diversity

Pustule initiations were visible with all five isolates at 6 DPI (Fig. 1). By 8 DPI, it was possible to classify the isolates as having low or high virulence on the differential maize set for each resistance gene (Fig. 1). From an agronomy perspective, the overall infection level/pattern for Isolates 73 and 1 was over 60%; by comparison, infection levels remained below 40% from 10 DPI for most maize lines. On the contrary, Isolates 9, 24, and 31 displayed diverse infection levels on the differential set.

At 10 DPI, Isolate 31 infection levels were above 25% for *Rp1_K*, whereas infection levels remained below 8% for Isolates 73, 24, 9, and 1. Isolate 31 had infection levels of 10% on sweet corn line E9 carrying *Rp1_D*, whereas infection levels were over 20% for Isolates 24, 73, and 1. Isolate 24 was able to cause infection on sweet corn lines E10 that carry resistance genes *Rp1_G* and *Rp1_J*, *Rp1_F*, whereas Isolates 73, 31, 9, and 1 were not. Isolate 1 and 73 infection levels for the maize line E29 carrying *Rp5* and *Rp1_D* were above 38%; however, they remained below 25% for Isolates 24 and 9.

There was no pattern of association of virulence with location of collection, suggesting that the variation in infection level on the differential set and infection levels over time were not related to the agroecological region of origin. Similarly, there was no association between patterns of virulence and year of collection. Rust samples collected in 2018 had contrasting virulence patterns on the *Rp* genes investigated. Interestingly, Isolates 1 and 73 displayed the most similar patterns of infection, despite being collected from locations separated by over 1,300 km and during

different years, and belonged to the same clade. The only differences were reported on Isolate 24 causing infection on *Rp1_E* and Isolate 73 on *Rp1_I*. Although early phenotyping data determined that some diversity was present between the isolates, they did not indicate the extent of the genomic divergence between the isolates.

Population subdivision

All 17 isolates sequenced represented unique genotypes, and no clones were detected (genotypic diversity = 1). Following clustering of genotypes based on genetic distance, the 17 isolates grouped into three different clusters (Figs. 2 and 3). Isolates 9, 12, 13, and 58 belonged to clade one; Isolates 9, 12, and 13 all collected on the same date and in the same field in Victoria (VIC), were most closely related. Isolate 58 was collected in 2018 from Queensland (QLD). Interestingly, Isolate 15 was also collected at the same location as Isolates 9, 12, and 13 but belonged to the third clade. Isolate 15 was most closely related to Isolate 73, despite the two originating from different hosts, sweet corn and maize, respectively. These two isolates originated from the same field, although 14 months apart. The second clade comprised the most diverse group of isolates. Most originated from diverse agroecological regions in QLD, except for Isolate 51 (VIC): subhumid, subtropical slopes and wet subtropical coast. The third clade comprised two isolates from VIC (Isolates 15 and 73) and two from QLD (Isolates 1 and 56). Results from the DAPC analysis were in line with the genetic distance tree, identifying three main clusters (Supplementary Fig. S1). Cluster 1, the main cluster, consisted of Isolates 24, 27, 31, 39, 43-1, 43-2, 51, 55, and 57; Cluster 2 included Isolates 1, 15, 56, and 73; and Cluster 3 consisted of Isolates 9, 12, 13, and 58. The same clusters were identified using the neighbor-net algorithm implemented in SplitTree (Fig. 3). Reticulation of the network in the main cluster (Cluster 1) indicated the presence of recombination among isolates, whereas no reticulation was detected in Clusters 2 and 3.

De novo hybrid assembly of Isolate 1

In MaSuRCA, the total number of contigs was 2,476, compared with 2,627 for Flye (Table 3). In BUSCO, the completeness of the *P. sorghi* draft genome constructed using MaSuRCA was estimated to be 90.2%, compared with 89% using Flye, and was therefore retained. Following the 20× de novo hybrid assembly of Isolate 1 in MaSuRCA, a draft genome of 145,963,792 bp was obtained after QUAST analysis (Table 3), with a total length of 145,815,044 bp (≥1,000 bp), compared with 99,534,058 bp in Rochi et al. (2018). The draft genome comprised 2,702 contigs for the long-read assembly, compared with 28,117 contigs in the short-read assembly generated by Rochi et al. (2018). Of the total number of contigs, 2,476 were over 1,000 bp. The N50 was 123,460 and 4,987 for Rochi et al. (2018), and the L50 was 354, compared with 3,829 for Rochi et al. (2018).

Following Repeat Masker, a majority of the TEs were classified as LTR (Long Terminal Repeat) 28.32% (Table 4). When LTR repeats (40,828,212 bp) were removed, the resulting total genome length was 104,986,831 bp. The collapse of abundant repeats of the rust genome in short-read assemblies often obscures their "true" genome size. The de novo assembly resulted in a total interspersed repeats content of 72.77%. RNAmmer (Lagesen

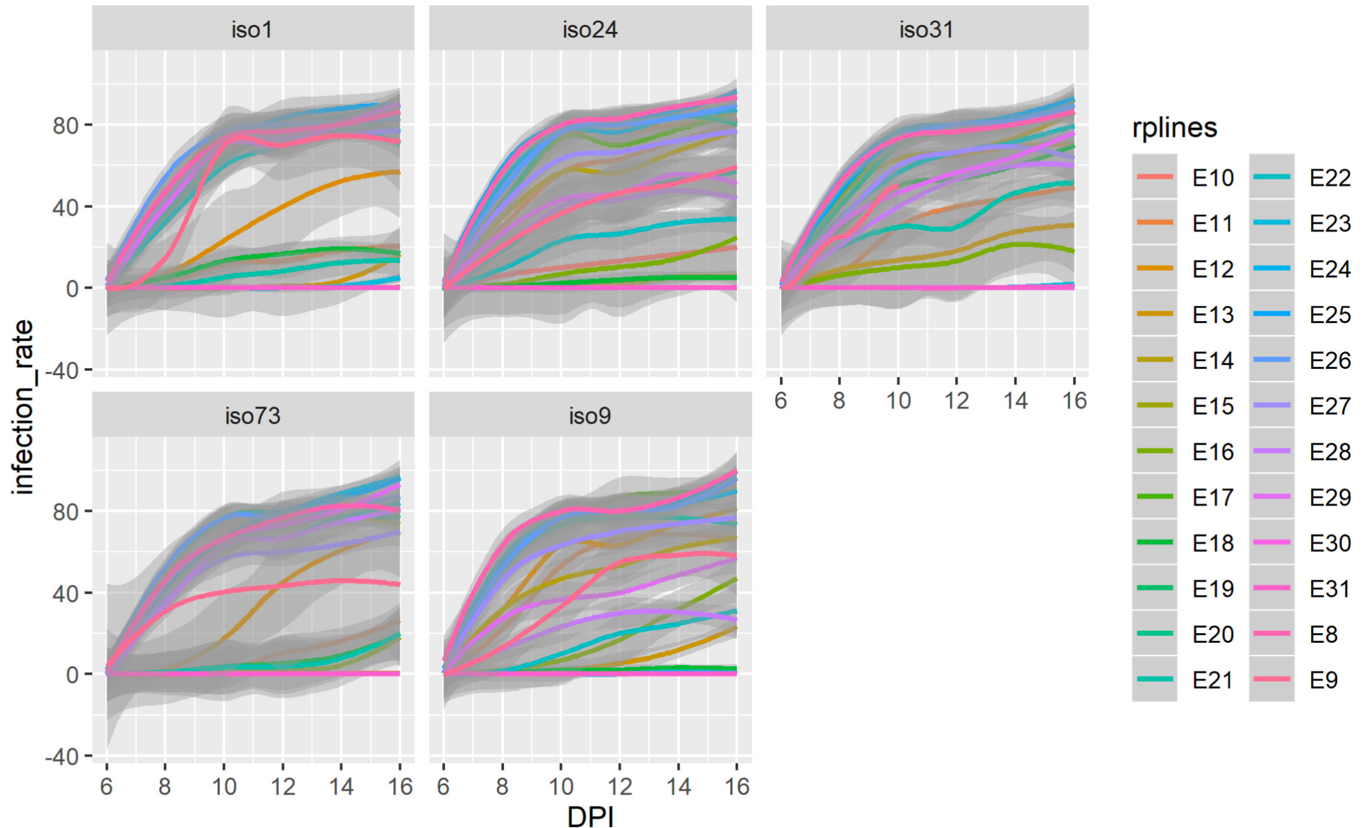

FIGURE 1

The percentage of infection from 6 to 16 days post inoculation (DPI) for the five most diverse *Puccinia sorghi* isolates in terms of collection location, year, and host (maize versus sweet corn) on a maize differential set carrying known resistance genes in an experiment conducted in Wagga Wagga, NSW, Australia.

et al. 2007) and tRNAScan (Chan and Lowe 2019) identified 13 rRNA and 1,007 tRNA, respectively. DNA elements accounted for 9.70%, and 34.20% remained unclassified. Gene prediction was achieved de novo in GenSAS with GeneMarkes and resulted in 10,357 predicted genes.

The gene prediction from the BRAKER pipeline identified 20,438 gene models and gave the most predictions. However, the gene prediction with evidence from *P. striiformis* resulted in 5,536 predicted features.

The hybrid assembly and gene predictions were conducted de novo, resulting in lower gene identification than in the assembly by Rochi et al. (2018).

DISCUSSION

Diversity in pathogen populations is the primary cause of resistance breakdown in hundreds of commercial crops (Schumann and D'Arcy 2012). The *P. sorghi* inoculum load found in Australia is likely to drive the emergence of new virulent strains. Phenotyping of the five most diverse isolates demonstrated that the rust population is diverse in Eastern Australia. The effect of the isolates not only varied on a genetic basis for their maize hosts but also between 2018 and 2019; that is, not all isolates produced urediniospores at 8 DPI on all the maize lines. Most isolates produced some pustules on all maize lines. The risk of the occurrence of a beneficial mutation in the pathogen remains low, at infection levels of 1 to 30%. The phenotyping data also suggest that virulence patterns were not correlated with a particular agroecological region. The level of diversity at the field level was not examined, as genome analyses are costly and should be the focus of further studies.

Maintaining an up-to-date and well-characterized collection of *P. sorghi* isolates with different virulence profiles is critical to assess the resistance in elite breeding lines and novel commercial lines. Furthermore, the data from the long-read sequence and mapped short reads of the 17 isolates can assist with the identification of *P. sorghi* secretomes and candidate effectors. For instance, we suspect that the gene prediction of 20,438 genes could potentially have been higher if the gene identification had not been conducted de novo and more evidence files from other fungi had been used.

The diversity observed within the population of *P. sorghi* in Australia suggests that maize breeding programs are at higher risks of experiencing the breakdown of monogenic resistance in the short to medium term and cannot rely on monogenic resistance. Although gene pyramiding offers great hope to increase the longevity of resistance in commercial crops, stacking an increasing number of resistance genes into one line comes at the

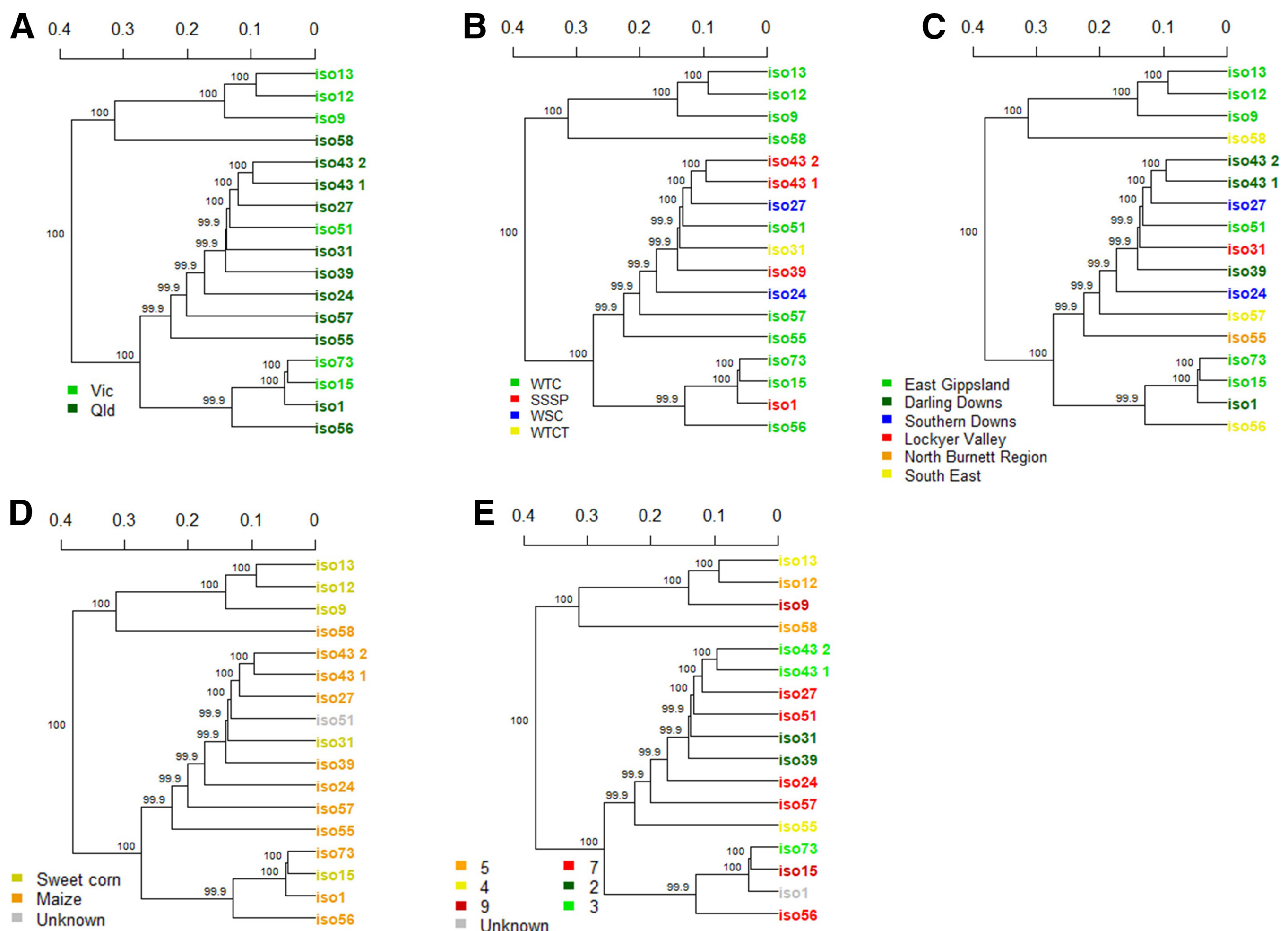

FIGURE 2

UPGMA dendrogram constructed based on Nei's distance among *Puccinia sorghi* isolates collected from *Zea maydis* in Australia. Bootstrap support values greater than 50 are shown above the branches. The colors of the isolates correspond to the **A,** state of collection (Qld: Queensland, Vic: Victoria); **B,** climate (WTC: wet temperate coast; SSSP: subhumid, subtropical slopes and plains; WSC: wet subtropical coast; WTCT: wet tropical coast and tableland); **C,** region; **D,** host; and **E,** virulence (scores taken in the field).

expense of other desirable traits and is time consuming and costly. In maize, resistance genes located in the *Rp1* complex may have the ability to create new resistance specificities as rapidly as *P. sorghi* engenders new virulence. The level of diversity in the *P. sorghi* population in Australia has the potential to overcome any newly deployed resistance genes as soon as they are deployed in the field. The likelihood that one spore may carry a beneficial mutation for the pathogen is 10^7 per day, per hectare, and under optimal conditions for disease development (Stam and McDonald 2018). According to Stam and McDonald (2018), the likelihood that a spore carries mutations that would overcome two *R* genes is $10^{-6} \times 10^{-6} = 10^{-12}$, representing 10 double mutants per day, per hectare. Additionally, if three genes are to be overcome, one mutant would be found in every 10,000 hectares of infected maize fields. However, for the above to be true, the pathogen must be avirulent to all three resistance genes. In the lines carrying three resistance genes *RpGDJ* (*Rp1_D*, *Rp1_G*, and *Rp1_J*), *RpGJF* (*Rp1_G*, *Rp1_J*, and *Rp1_F*) and four resistance genes *Rp5GCJ* (*Rp5*, *Rp1_C*, *Rp1_G*, and *Rp1_J*), only two resistance genes were effective: *Rp1_G* and *Rp5* (Quade et al. 2021). All other genes were ineffective. It is unknown whether the combination of those genes into one plant would have a synergistic effect and provide durable resistance. This leaves the maize industry in Australia with few candidates for long-term protection.

The high level of genetic diversity among genotypes in 2018 suggests that the population of *P. sorghi* in Eastern Australia is genetically diverse and may also support the theory of long-distance dispersal of genotypes between states. In 2018, weather conditions were not conducive to disease development, and yet sporulation occurred, and three distinct clades were identified following mapping (Figs. 2 and 3). We previously thought that the three clades would group rust isolates collected from the same field and/or in the same year. However, this was not the case. For instance, Isolates 9, 12, and 13 from VIC belonged to Clade 1, together with Isolate 58, which was collected from QLD 2 months later. Clade 2 comprised the most diverse group of isolates. Although most isolates in Clade 2 originated from QLD, except for Isolate 51, they all came from different locations of diverse agroecology. Clade 3 comprised two isolates from VIC (Isolates 15 and 73) and isolates from QLD (Isolates 1 and 56). Isolate 15 was expected to be genetically more closely related to Isolates 9, 12, and 13 but belonged to Clade 3, together with Isolates 1, 56, and 73. Isolates 15 and 73 originated from different hosts. As demonstrated by other studies (Shah and Dillard 2004), *P. sorghi* infects maize and sweet corn indiscriminately. Although collected 14 months apart, Isolates 15 and 73 also originated from the same field. It is suspected that individual strains may persist from one year to the next, but this remains to be confirmed.

Such high levels of genotypic diversity and lack of clones in the isolates suggest that the *P. sorghi* population in Australia may undergo sexual reproduction and/or experience incursions from overseas. The alternative host, *Oxalis* spp., is present in Australia. If the sexual cycle is not active in Australia, it would be expected that clonally produced isolates would group within a clade based

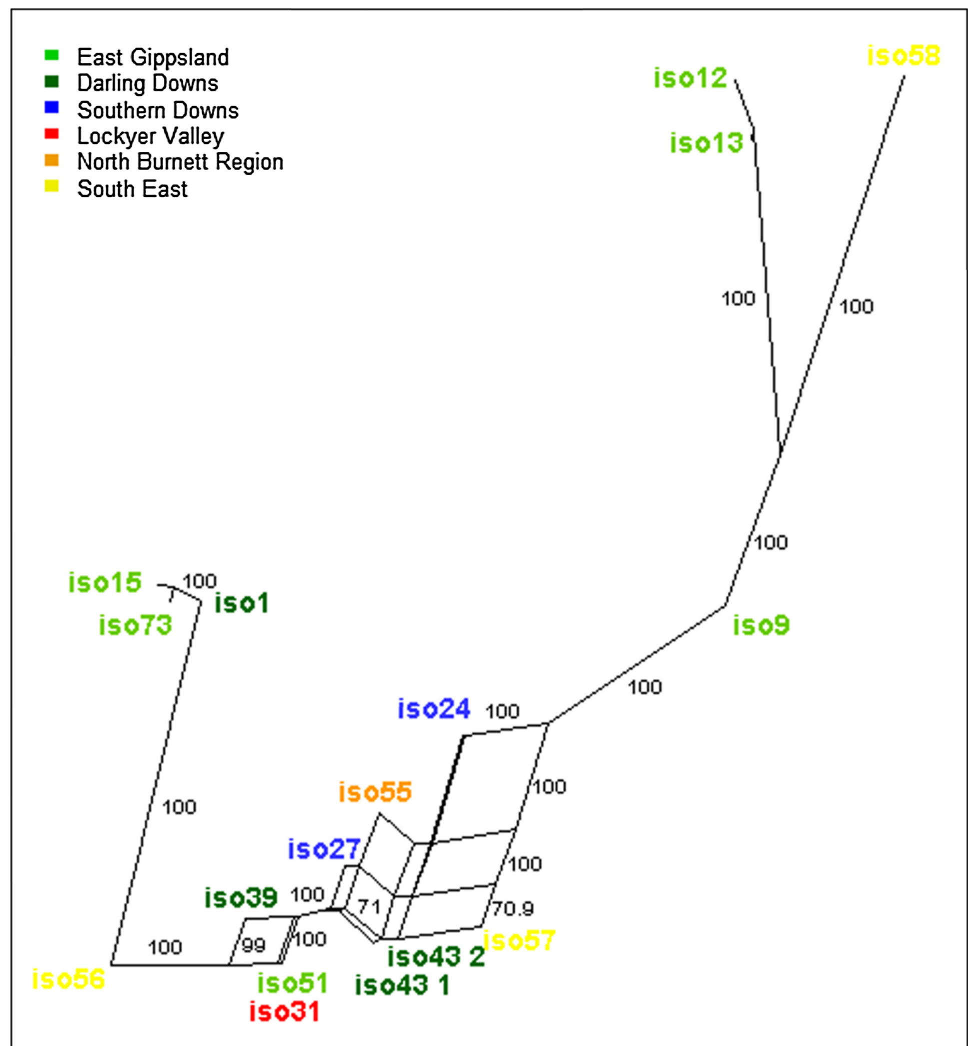

FIGURE 3 Neighbor-net network constructed for *Puccinia sorghi* isolates collected from *Zea maydis* in Australia, with 1,000 bootstrap replicates, constructed using genome-wide single-nucleotide polymorphism data. The color of isolates corresponds to the region from which the isolates were collected.

on year and/or location. It was originally hypothesized that rust isolates collected from the same field or the same region would belong to the same cluster, but this was obviously not the case. Surprisingly, these genetic differences were observed even in a year of reduced disease pressure and therefore low sporulation. In years of environmental conditions being conducive to urediniospore production, it is expected that the likelihood of virulence favorable to the pathogen will increase, as the number of spores produced increases. The results also warrant a closer look at the occurrence of the alternate host, *Oxalis* spp., growing in maize producing areas in Australia.

Isolates with emerging virulence will have the potential to produce more spores and spread from new infection sites. The question remains as to why Isolate 58 from QLD belonged to the same clade as three isolates from VIC, but Isolate 51 from VIC was in the Clade 2 with isolates from QLD. It is unclear whether these isolates migrated in wind currents or human intervention transported them from one state to the next.

The hybrid assembly resulted in an increase in read length. Rochi et al. (2018) obtained a genome size of 78 Mb for *P. sorghi* after assembly with SPAdes, once most of the reads had been assembled. A total length of 145,815,044 bp (≥1,000 bp) was obtained. This is similar to long reads based on de novo assemblies of several other rust fungi, for example, *P. triticina* (140 Mb) (Wu et al. 2020) and *P. coronata* f. sp. *avenae* (99.2 Mb) (Miller et al. 2018), and Rochi et al. (2018) estimated that the genome size of *P. sorghi* is 102 Mb. The higher percentage of LTR found by Rochi et al. (2018) (15.76% compared with 28.32% in the findings of this research) suggests that the short-read assembly is less effective in handling repeat regions. The repetitive nature of the genomes of rust species limits the ability to estimate their true genome size based on assembled sequence data. LTR are responsible, among other mechanisms, for the integration of the newly created cDNA into another location and contribute to genome plasticity. DNA elements accounted for 9.70% of the repeats, compared with 7.41% for Rochi et al. (2018). However, it was unclear why 34.20% remained unclassified, compared with 22.13% in the study by Rochi et al. (2018) and 16.02% in the study by Miller et al. (2018). It is suspected that the higher-quality genome assembly resulted in the identification of a greater number of repeat regions that are yet to be categorized. The total interspersed repeat content of 72.77% found in the de novo assembly is close to that found by Tobias et al. (2020) in their study of *Austropuccinia psidii*, with over 80% interspersed repeats.

Maize breeding programs are urged to orientate their research toward multigenic resistance and/or quantitative resistance. The likelihood of a single gene continuing to provide effective resistance is extremely low, considering the diversity present in the Australian population of *P. sorghi*. Ongoing monitoring of the rust population across maize growing regions is necessary to make informed decisions on which maize lines should be removed from the market as soon as they become susceptible to prevent exponential spore production. A targeted effort to understand the role of the alternate host, *Oxalis* spp., in the emergence of new pathotypes is required. Virulence studies will also help us understand the origin of new virulence and alert us to potential overseas incursions (e.g., transported by human and/or wind currents).

TABLE 3

Genome statistics for the draft hybrid assembly of *Puccinia sorghi* Isolate 1

Isolate 1 statistics	MaSuRCA assembly	Flye assembly
# contigs (≥0 bp)	2,702	2,728
# contigs (≥1,000 bp)	2,476	2,627
Total length (≥0 bp)	145,963,792	155,488,878
Total length (≥1,000 bp)	145,815,044	155,417,575
# contigs	2,702	2,723
Largest contig	685,541	981,868
Total length	145,963,792	155,486,859
Estimated reference length	77,000,000	155,486,859
GC (%)	44.38	44.58
N50[a]	123,460	165,816
NG50[b]	203,904	na[c]
N75[d]	67,607	75,180
NG75[e]	152,782	na
L50[f]	354	269
LG50[g]	134	na
L75[h]	750	612
LG75[i]	243	na
# N's per 100 kbp	0.2	0.23

[a] N50 = 50% of the total assembled sequences is on contigs larger than that size.
[b] NG50 = The size of the smallest contig in the set of largest contigs that make up 50% of the estimated genome size.
[c] na = not available.
[d] N75 = 75% of the total assembled sequences is on contigs larger than that size.
[e] NG75 = The size of the smallest contig in the set of largest contigs that make up 75% of the estimated genome size.
[f] L50 = Count of the smallest number of contigs whose length makes up 50% of the genome assembly size.
[g] LG50 = Count of the smallest number of contigs whose length makes up 50% of the estimated genome size.
[h] L75 = Count of the smallest number of contigs whose length makes up 75% of the genome assembly size.
[i] LG75 = Count of the smallest number of contigs whose length makes up 75% of the estimated genome size.

TABLE 4

Proportions of repeated sequence content in *Puccinia sorghi* hybrid assembly

Repeat class	Percentage (%)
SINEs	0
LINES	0.55
LTR elements	28.32
DNA elements	9.70
Unclassified	34.20
Total interspersed repeats	72.77
Small RNA	0
Satellites	0
Simple repeats	0.42
Low complexity	0

ACKNOWLEDGMENTS

We thank the University of Southern Queensland and the Grains Research and Development Corporation for the Ph.D. scholarship that enabled this study, Charles Sturt University for the use of their facilities, and John Xie for his help with statistical analysis.

LITERATURE CITED

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., and Blankenberg, D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46:W537-W544.

Aime, M. C., McTaggart, A. R., Mondo, S. J., and Duplessis, S. 2017. Phylogenetics and phylogenomics of rust fungi. Adv. Genet. 100:267-307.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. https://qubeshub.org/resources/fastqc

Bailey, J., Karaoglu, H., Wellings, C., and Park, R. 2013. Isolation and characterisation of 25 genome-derived simple sequence repeat markers for *Puccinia striiformis* f. sp. *tritici*. Mol. Ecol. Resour. 13:760-762.

Barsoum, M., Sabelleck, B., Spanu, P. D., and Panstruga, R. 2019. Rumble in the effector jungle: Candidate effector proteins in interactions of

plants with powdery mildew and rust fungi. Crit. Rev. Plant Sci. 38: 255-279.
Bettgenhaeuser, J., Gilbert, B., Ayliffe, M., and Moscou, M. J. 2014. Nonhost resistance to rust pathogens - a continuation of continua. Front. Plant Sci. 5:664.
Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.
Borodovsky, M., and Lomsadze, A. 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Curr. Protoc. Bioinform. 4:4.6.10.
Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.
Brůna, T., Lomsadze, A., and Borodovsky, M. 2020. GeneMark-EP+: Eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom. Bioinform. 2:lqaa026.
Buchfink, B., Xie, C., and Huson, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59-60.
Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., and Thompson, R. 2018. ASReml-R Reference Manual Version 4. VSN International, Hemel Hempstead, U.K.
Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K. K., Jurka, J., Michelmore, R. W., and Dubcovsky, J. 2011. 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.
Chan, P. P., and Lowe, T. M. 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol. Biol. 1962:1-14.
Darino, M. A., Rochi, L., Lia, V. V., Kreff, E. D., Pergolesi, M. F., Ingala, L. R., Dieguez, M. J., and Sacco, F. 2016. Virulence characterization and identification of maize lines resistant to *Puccinia sorghi* Schwein. present in the Argentine corn belt region. Plant Dis. 100:770-776.
De Maio, N., Shaw, L. P., Hubbard, A., George, S., Sanderson, N. D., Swann, J., Wick, R., AbuOun, M., Stubberfield, E., Hoosdally, S. J., Crook, D. W., Peto, T. E. A., Sheppard, A. E., Bailey, M. J., Read, D. S., Anjum, M. F., Walker, A. S., and Stoesser, N. 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. Microb. Genom. 5:e000294.
Fato, P. 2000. Pathogenic variability in *Puccinia sorghi* on maize in South Africa. Ph.D. Thesis, Orange Free State, Faculty of Natural and Agricultural Sciences.
Garrison, E. 2015. *vcflib*, GitHub. https://github.com/ekg/vcflib
Garrison, E. P., and Marth, G. T. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv:* Genomics. https://doi.org/10.48550/arXiv.1207.3907
Groth, J. V., Pataky, J. K., and Gingera, G. R. 1992. Virulence in eastern North-American populations of *Puccinia sorghi* to *Rp* resistance genes in corn. Plant Dis. 76:1140-1144.
Hoff, K. J., Lomsadze, A., Borodovsky, M., and Stanke, M. 2019. Whole-genome annotation with BRAKER. Methods Mol. Biol. 1962:65-95.
Humann, J. L., Lee, T., Ficklin, S., and Main, S. D. 2019. Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. Humana Press, New York, NY.
Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.
Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.
Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37:540-546.
Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., and Zdobnov, E. M. 2018. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47:D807-D811.
Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. 2007. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35:3100-3108.
Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.
Lin, Y., Yuan, J., Kolmogorov, M., Shen, M. W., Chaisson, M., and Pevzner, P. A. 2016. Assembly of long error-prone reads using de Bruijn graphs. Proc. Natl. Acad. Sci. U.S.A. 113:E8396-E405.
Miller, M. E., Zhang, Y., Omidvar, V., Sperschneider, J., Schwessinger, B., Raley, C., Palmer, J. M., Garnica, D., Upadhyaya, N., Rathjen, J., Taylor, J. M., Park, R. F., Dodds, P. N., Hirsch, C. D., Kianian, S. F., and Figueroa, M. 2018. Assembly and phasing of dikaryotic genomes from two isolates of *Puccinia coronata* f. sp. *avenae* the causal agent of oat crown rust. mBio 9:e01650-17.
Moulton, V. 2003. SplitsTree: A network-based tool for exploring evolutionary relationships. Pages 312-328 in: The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. M. Salemi and A.-M. Vandamme, eds. Cambridge University Press, Cambridge, U.K.
NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44:D7-D19.
O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., and Pruitt, K. D. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44:D733-D745.
Park, R. F. 2015. Long term surveys of pathogen populations underpin sustained control of the rust diseases of wheat in Australia. J. Proc. R. Soc. N. S. W. 148:455-456.
Porquier, A., Tisserant, C., Salinas, F., Glassl, C., Wange, L., Enard, W., Hauser, A., Hahn, M., and Weiberg, A. 2021. Retrotransposons as pathogenicity factors of the plant pathogenic fungus *Botrytis cinerea*. Genome Biol. 22:225.
Quade, A., Ash, G. J., Park, R. F., and Stodart, B. J. 2021. Resistance in maize (*Zea mays*) to isolates of *Puccinia sorghi* from Eastern Australia. Phytopathology 111:1751-1757.
Rochi, L., Diéguez, M. J., Burguener, G., Darino, M. A., Pergolesi, M. F., Ingala, L. R., Cuyeu, A. R., Turjanski, A., Kreff, E. D., and Sacco, F. 2018. Characterization and comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize common rust. Fungal Genet. Biol. 112:31-39.
Roelfs, A. P., and Bushnell, W. R. 1985. The Cereal Rusts: Diseases, Distribution, Epidemiology and Control, Volume 2. Academic Press, Cambridge, MA.
Rowell, J. B., and Roelfs, A. P. 1971. Evidence for an unrecognized source of overwintering wheat stem rust in the United States. Plant Dis. Rep. 55:990-992.
RStudio. 2018. Integrated Development for R. Boston, MA. https://www.rstudio.com/
Schumann, G. L., and D'Arcy, C. J. 2012. Hungry planet: Stories of plant diseases. American Phytopathological Society, St. Paul, MN.
Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., Dodds, P. N., Figueroa, M., Park, R. F., and Rathjen, J. P. 2018. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high interhaplotype diversity. mBio 9:e02275-17.
Seppey, M., Manni, M., and Zdobnov, E. M. 2019. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962:227-245.
Shah, D., and Dillard, H. 2004. Possible sources of *Puccinia sorghi* inoculum initiating rust epidemics on sweet corn in western New York, 2002. 94:S95.
Smit, A. F. A., and Hubley, R., 2008. RepeatModeler Open-1.0. http://www.repeatmasker.org
Stam, R., and McDonald, B. A. 2018. When resistance gene pyramids are not durable—The role of pathogen diversity. Mol. Plant Pathol. 19:521-524.
Stanke, M., and Morgenstern, B. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33:W465-W467.
Tange, O. 2011. GNU Parallel - The command-line power tool. USENIX Magazine 2011:42-47.
Tobias, P. A., Schwessinger, B., Deng, C. H., Wu, C., Dong, C., Sperschneider, J., Jones, A., Lou, Z., Zhang, P., Sandhu, K., Smith, G. R., Tibbits, J., Chagné, D., and Park, R. F. 2020. *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. G3 Genes Genom. Genet. 11:jkaa015.
Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27:737-746.
Wang, J. R., Holt, J., McMillan, L., and Jones, C. D. 2018. FMLRC: Hybrid long read error correction using an FM-index. BMC Bioinform. 19:50.
Welham, S., Cullis, B., Gogel, B., Gilmour, A., and Thompson, R. 2004. Prediction in linear mixed models. Aust. N. Z. J. Stat. 46:325-47.
Wu, J. Q., Dong, C., Song, L., and Park, R. F. 2020. Long-read–based de novo genome assembly and comparative genomics of the wheat leaf rust pathogen *Puccinia triticina* identifies candidates for three avirulence genes. Front. Genet. 11:521.
Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., and Yorke, J. A. 2013. The MaSuRCA genome assembler. Bioinformatics 29: 2669-2677.