Haplotype-phased and chromosome-level genome assembly of *Puccinia polysora*, a giga-scale fungal pathogen causing southern corn rust

Junmin Liang[1], Yuanjie Li [1,2], Peter N. Dodds[3], Melania Figueroa[3], Jana Sperschneider[3], Shiling Han[1,2], Clement K.M. Tsui[4,5,6], Keyu Zhang[7], Leifu Li[7], Zhanhong Ma[7], Lei Cai[1,2*]

[1] State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

[2] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

[3] Commonwealth Scientific and Industrial Research Organization, Agriculture and Food, Canberra, 10 ACT, Australia

[4] Division of Infectious Diseases, Faculty of Medicine, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

[5] National Centre for Infectious Diseases, Tan Tock Seng Hospital, 308433, Singapore

[6] LKC School of Medicine, Nanyang Technological University, 308232, Singapore

[7] Department of Plant Pathology, China Agricultural University, Beijing, 100193, China

Correspondence: Lei Cai, State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China.

Email: cail@im.ac.cn

Abstract

Rust fungi are characterized by large genomes with high repeat content and have two haploid nuclei in most life stages, which makes achieving high-quality genome assemblies challenging. Here, we described a pipeline using HiFi reads and Hi-C data to assemble a gigabase-sized fungal pathogen, *Puccinia polysora* f.sp. *zeae*, to haplotype-phased and chromosome-scale. The final assembled genome is 1.71 Gbp, with ~850 Mbp and 18 chromosomes in each haplotype, being currently one of the two giga-scale fungi assembled to chromosome level. Transcript-based annotation identified 47,512 genes for dikaryotic genome with a similar number for each haplotype. A high level of interhaplotype variation was found with 10% haplotype-specific BUSCO genes, 5.8 SNPs/kbp and structural variation accounting for 3% of the genome size. The *P. polysora* genome displayed over 85% repeat contents, with genome-size expansion and copy number increasing of species-specific orthogroups. Interestingly, these features did not affect overall synteny with other *Puccinia* species having smaller genomes. Fine-time-point transcriptomics revealed seven clusters of co-expressed secreted proteins that are conserved between two haplotypes. The fact that candidate effectors interspersed with all genes indicated

the absence of a "two-speed genome" evolution in *P. polysora*. Genome resequencing of 79 additional isolates revealed a clonal population structure of *P. polysora* in China with low geographic differentiation. Nevertheless, a minor population differentiated from the major population by having mutations on secreted proteins including *AvrRppC*, indicating the ongoing virulence to evade recognition by *RppC*, a major resistance gene in Chinese corn cultivars. The high-quality assembly provides valuable genomic resources for future studies on disease management and the evolution of *P. polysora*.

Keywords: rust fungi, *de novo* genome assembly, phasing, population genomics, *AvrRppC*

1 | INTRODUCTION

Rust fungi (Pucciniales) constitute one of the largest orders in the kingdom of fungi, with more than 8,000 species grouped into 18 families and approximately 170 accepted genera (Zhao et al. 2020, 2021; Aime and McTaggart 2021). Rust fungi are obligate biotrophs, which are usually recalcitrant to in vitro culturing and are host-specific to particular species, genera, or families of vascular plants. Their life cycles are diverse and complex, including up to five types of spores (spermatia, aeciospores, urediniospores, teliospores, and basidiospores) (Zhao et al. 2021). Spores of the most abundant life stage are dikaryon, which contains two physically separated and genetically different haploid nuclei. The Pucciniales contains species with the largest known genomes, with an average haplotype genome size of 380 Mbp estimated by flow cytometry, far larger than that of all fungi (37.7 Mbp) (Tavares et al. 2014). Based on flow cytometry, the largest known rust genome (*Uromyces bidentis*) was estimated at up to 2.4 Gbp (Ramos et al. 2015). The repeat contents of rust genomes range from 30% to 91% (Amie et al. 2017; Tobias et al. 2021). Their dikaryotic nature and highly repetitive sequences pose substantial challenges for high-quality genome assembly as well as an in-depth understanding of evolution within the rust group.

In addition to the above intriguing biological features, rust fungi have also received wide attention for causing major diseases on agricultural and forest crops worldwide. The most devastating rust pathogen include *Puccinia striiformis* f.sp. *tritici* (*Pst*), *P. graminis* f.sp. *tritici* (*Pgt*) and *P. triticina* f.sp. *tritici* (*Ptt*), causing three of the world's most serious wheat rust diseases (Kolmer 2005); *Puccinia sorghi* f.sp. *zeae* (*Psz*) and *P. polysora* f.sp. *zeae* (*Ppz*), also

63 seriously threaten global crop production by causing common and southern corn rust of maize
64 (Crouch and Szabo 2011; Ramirez-Cabral et al. 2017); and *Melampsora larici-populina* and
65 *Austropuccinia psidii*, two notorious forest pathogens leading to poplar rust and myrtle rust,
66 respectively (Pinon and Frey 1997; Winzer et al. 2019). Effective strategies for management of
67 rust disease rely heavily on host resistance breeding programs. In host-rust interactions, rust
68 fungi deliver specific effectors (avirulence genes, *Avr*) into host cells, which can be recognized
69 by resistance proteins from resistant plants and trigger defense responses (Dodds and Rathjen
70 2010). However, rust pathogens can acquire mutations in the *Avr* gene to avoid this recognition
71 and facilitate the pathogen infection (Cui et al. 2015). For instance, the Ug99 race (TTKSK) of *P.*
72 *graminis* evolved virulence to the widely deployed *Sr31* resistance gene in wheat has become a
73 big threat to global wheat production (Singh et al. 2011). Therefore, understanding the molecular
74 mechanism of *Avr-R*, or gene-for-gene relationship, is critical for durable disease control. Given
75 their biotrophic lifestyle, the genetic transformation of rust fungi is difficult and challenging.
76 Nevertheless, genome sequencing data of rust species has enabled the prediction of candidate
77 effectors which facilitate the identification of *Avr* genes (Figueroa et al. 2016; Anderson et al.
78 2016; Maia et al. 2017; Salcedo et al. 2017; Chen et al. 2017; Miller et al. 2018; 2020;
79 Upadhyaya et al 2021).

80 Early rust genome assemblies were haploid representations and different homologous
81 haplotypes were collapsed into a consensus assembly, which did not fully capture sequences
82 from both nuclei. Using single-molecule real-time (SMRT) sequencing technology and
83 haplotype resolution software, FALCON-Unzip (Chin et al. 2016), allowed higher contiguity and
84 partially haplotype-phased genomes of a few *Puccinia* species (Miller et al. 2018; Schwessinger
85 et al. 2018, Wu et al. 2021). Other softwares have been developed to obtain sub-assemblies from
86 diploid assemblies, such as HaploMerger2 (Huang et al. 2017) and Purge_Haplotigs (Roach et al.
87 2018), but many duplicate contigs cannot be correctly switched leading to mis-joins in the final
88 sub-assemblies. Recently, several softwares have been developed to phase two haplotypes from
89 the diploid genome using Hi-C data, such as TrioCanu (Koren et al. 2018), NuclearPhaser (Duan
90 et al. 2022) and FALCON-Phase (Kronenberg et al. 2022). Three *Puccinia* species, causing
91 wheat or oat rust have been assembled to a fully-phased and chromosome level using
92 NuclearPhaser (Li et al. 2019; Duan et al. 2022; Henningsen et al. 2022). Unlike above rusts,
93 which have been studied extensively, the genome information of corn rust pathogens is largely

unknown. Since 2012, global maize production has exceeded that of wheat and rice, becoming the food crop with the highest world production yield (FAO, https://www.fao.org/faostat/en/#data/QCL), meanwhile, the distribution of the two major rust diseases (common corn rust and southern corn rust) in corn is projected to expand to temperate regions with increasing global temperatures (Ramirez-Cabral et al. 2017), which poses a greater threat to global food safety. A draft genome sequence of common corn rust (*P. sorghi*) was released in 2016, although it was highly fragmented (Rochi et al. 2016). However, genomic information for southern corn rust pathogens has not been reported, which seriously hampers avirulence gene identification and breeding of disease resistance. The genome sizes of recent phased *Puccinia* species range from ~170 Mbp to 250 Mbp (Li et al. 2019; Miller et al. 2018; Wu et al. 2021), whereas the genome of *P. polysora* was estimated over gigabase in our preliminary test (Figure S1), which poses additional challenges for genome assembly.

In this study, we assembled the genome of *P. polysora* to haplotype-resolved and chromosome-scale levels using a modified haplotype-phasing assembly pipeline. To our knowledge, the current study and the report on the myrtle rust pathogen, *Austropuccinia psidii* (Edwards et al. 2022) provide the first giga-scale fungal genome ever assembled to complete-chromosome level. The *P. polysora* genome was annotated by fine-time-point gene expression data from germinated spores and infected issues, which supplied robust data for coexpression analyses of secreted proteins and prediction of candidate secreted effectors. Also, using the haplotype-phasing genome as a reference, we investigated the genetic and population structure of *P. polysora* based on genome sequencing of 79 additional samples from north, south and central parts of China. The high-quality genome information represents a valuable resource for better understanding genome evolution, host adaption, and genes involved in host-pathogen interactions.

2 | MATERIALS AND METHODS

2.1 | Fungal isolates and plant inoculation

The isolate, GD1913, collected from Guangdong province in the 2019 annual rust survey, was chosen for reference genome sequencing. Based on a preliminary inoculation experiment, GD1913 represented the most frequent virulence type. Additional 79 isolates, collected from the south, central, and north China were selected and used for genome resequencing (Table S1). All isolates were purified by selecting a single pustule from infected leaves and amplified by 2–3

125 rounds of infection on highly susceptible variety Zhengdan 958. Corn seeds immersed in
126 Chlormequat chloride (1.5 g/L) were sown in 10-cm wide square pots 10 days before inoculation.
127 When 10-cm tall, seedling growth was reduced by adding 15 mL maleic hydrazide acid (1.5 g/L)
128 per pot. Urediniospores were amplified by spraying on 10-day-old corn seedlings. The inoculated
129 seedlings were incubated with dew at 27°C in the dark for 24h before transferring to a climate-
130 controlled chamber (12h, 25°C dark period and a 12h, 27°C light period). To prevent airborne
131 contamination, 75% ethyl alcohol was sprayed on the area after each inoculation and each pot
132 was protected by a cellophane bag. Once the pots were heavily infected (ca. 15–20 days), spores
133 were harvested to clean cellophane by scraping them with sterile needles. For long-term storage,
134 fresh spores were dried (10% relative humidity) in a desiccator for 1 day at 4°C and then
135 maintained at -80°C or in liquid nitrogen.

136 **2.2 | DNA extraction, library preparation, and sequencing**

137 About 500 mg of fresh urediniospores were ground for 4-5 batches in liquid nitrogen and
138 genomic DNA was extracted using the lysis buffer and cetyltrimethylammonium bromide
139 (CTAB) method (Justesen et al. 2002). The integrity of genomic DNA was assessed by Agilent
140 4200 Bioanalyzer (Agilent Technologies, Palo Alto, California, USA). The average insert size of
141 15 kb PacBio library was concentrated with AMPure PB magnetic beads (Pacific Biosciences,
142 California, USA) following the manufacturer's instruction. Sequencing of high fidelity (HiFi)
143 long reads was carried out by the Pacific Bioscience Sequel II platform. For genome
144 resequencing, ~10 mg fresh urediniospores were placed in 2 mL screw-gap tubes filled with
145 Lysing Matrix C and ground twice in MP FastPrep-24 TM 5G (Mp Biomedicals, USA) with a
146 speed setting of 4 for 20 seconds. The paired-end library with 150 bp was prepared and
147 sequenced using Illumina NovaSeq 6000. All sequencing was performed at Annoroad Gene
148 Technology Co., Ltd, (Beijing, China).

149 To generate a chromosome-level and haplotype phased assembly of *P. polysora* genome, a
150 Hi-C library was generated following in situ ligation protocols. In brief, fresh urediniospores (~
151 100 mg) of GD1913 were used for crosslinking reaction by 2% formaldehyde at room
152 temperature for 15 min. After Glycine quenching, the supernatant was removed and spores were
153 then ground with liquid nitrogen for DNA extraction. The purified DNA was digested with *Mbo*
154 I restriction enzyme (New England Biolabs Inc. Beijing, China) and was labeled by incubating
155 with Biotin-14-dATP (Thermo Fisher Scientific, Massachusetts, USA) and then ligated by T4

156 DNA Ligase (Thermo Fisher Scientific, Massachusetts, USA). After incubating overnight to
157 reverse crosslinks, the ligated DNA was sheared into ~350 bp fragments. Finally, the Hi-C
158 library was quantified by Bioanalyzer and sequenced on the Illumina NovaSeq 6000 platform
159 using paired-end 150 cycles.

160 **2.3 | *De novo* assembly, haplotype phasing and Hi-C scaffolding**

161 HiFi reads were assembled by Canu 2.1.1 with -pacbio-hifi (Nurk et al. 2020). The coverage
162 depth was calculated by genomeCoverageBed in BEDtools (v2.29.2) (Quinlan and Hall 2010)
163 and small contigs (< 20 kbp) with low coverage (< 2×) were excluded from the further assembly.
164 The remaining contigs were examined by BLASTN search (v2.7.2) against the NCBI nt/nr
165 database (downloaded on May 20, 2021) with E-value set as 1e-10. To move the contaminants
166 e.g. plant rDNA and chloroplast sequences, we utilized a ContaminantScreening pipeline
167 (https://github.com/JanaSperschneider/GenomeAssemblyTools/tree/master/ContaminantScreenin
168 g). Contigs of high coverage and low GC content were defined as potential mitochondrial contigs,
169 and they were extracted from the assembled output from Canu and searched against the
170 mitochondrial database from NCBI (https://ftp.ncbi.nlm.nih.gov/refseq/release/mitochondrion/).
171 The contigs (63 contigs in Table S2) with perfect matches to mitochondria were also excluded
172 and they were assembled separately by Canu v2.1.1 to obtain mtDNA. The assembled sequences
173 were searched against itself database to test the ring formation.

174 To obtain a haplotype-phased assembly, HaploMerger2 (Huang et al. 2017), a tool to
175 rebuild both haploid sub-assemblies from the high-heterozygosity diploid genome, was used.
176 The heterozygosity evaluated by Jellyfish v2.1.3 (Marçais and Kingsford 2011) and
177 Genomescope.R (Vurture et al. 2017) was 1.08% leading to the identity setting as 95% (Huang et
178 al. 2017). To better distinguish allelic and non-allelic combinations, the scoring scheme for
179 alignments was recalculated by a Perl script (lastz_D_Wrapper.pl) that came with HaploMerger2.
180 The top 22 (in length) contigs accounting for ~10% length of all contigs were assigned to
181 part1.fasta and others were in part2.fasta. Mis-joins were detected by three rounds and all breaks
182 were manually checked combined with Hi-C validation. Based on all-vs-all alignment
183 (hm.new_scaffolds), contigs were assigned separately to haplotype A and haplotype B.
184 For scaffolding, the raw Hi-C data were trimmed by removing adapters and low-quality bases,
185 and then two-paired reads were separately mapped to two haplotypes independently using BWA-
186 MEM v0.7.8 (Li and Durbin 2010). To move experimental artifacts, the alignments went through

187 the mapping workflow from the Arima Genomics pipeline
188 (https://github.com/ArimaGenomics/mapping_pipeline/blob/master/01_mapping_arima.sh).
189 Then SALSA v2.2 (Ghurye et al. 2017) was run to cluster initial contigs into groups. SALSA
190 scaffolding was performed independently on haplotype A and B contigs. Hi-C contact matrices
191 for each scaffold were calculated by HiC-Pro (Servant et al. 2015) and contact maps are
192 visualized by HiCPlotter (Akdemir and Chin 2015). Contig reversal and breaking of mis-joins
193 were corrected manually according to Hi-C contact map and global alignment resulted from
194 HaploMerger2 (hm.new_scaffolds). Hicexplorer v3.4.1 (Wolff et al. 2018) was used to calculate
195 Hi-C links of read pairs to each haplotype. To guide scaffolds to chromosomes, telomeres were
196 identified by telomere_identification.pl (https://github.com/jimie0311/Puccinia-polysora-
197 genome/blob/main/Figure-1/telomere_identification.pl). The possible tandem repeats,
198 (CCTAAA/TTAGGG)n in most filamentous ascomycete fungi (Lue 2021), and other irregular
199 type reported in closely related rust fungi, (CCCTAA/TTTAGG)n, (CCCCTAA/TTAGGGG)n,
200 (CCCTAAA/TTTAGGG)n, (CCCTAA/TTAGGG)n, were all tested (Li et al. 2019; Tobias et al.
201 2021). Finally (CCCTAAA/TTTAGGG)n was identified as the telomere repeat units for *P.*
202 *polysora*. The corrected scaffolds were rebuilt to chromosome level according to telomeres. The
203 centromeres were identified by the cross-link signal based on Hi-C contact map (arrows in
204 Figure 2B). The completeness of each haplotype was evaluated using BUSCOs of
205 basidiomycota_odb9 and *Ustilago maydis* as the selected species for AUGUSTUS gene
206 prediction (Stanke and Morgenstern 2005) in BUSCO v3.0.2 (Simão et al. 2015). The workflow
207 of chromosome assembly has been illustrated in Figure 1A and the codes and pipelines used for
208 the assembly pipeline were available in github (https://github.com/jimie0311/Puccinia-polysora-
209 genome/blob/main/Figure-1/Figure-1A.md).

210 **2.4 | RNA extraction and sequencing**

211 For gene annotation, transcriptome sequencing of *P. polysora* from various infection time points
212 was performed. Corn seedlings of 18 pots (10 seedlings/pot) were inoculated with 10 mg
213 spores/mL mixed of Tween 20 (v/v: 0.05%). About five infected leaves were cut off per
214 biological replicate at 1 day post inoculation (dpi), 2 dpi, 4 dpi, 7dpi, 10 dpi, and 14 dpi. In
215 addition, germinated spore samples were prepared by placing 30 mg of fresh urediniospores on
216 the surface of sterile water at 25–28°C in dark for12 hours. Three biological replicates were
217 performed for each condition. All samples were collected in 4 mL Eppendorf tubes, frozen in

218 liquid nitrogen before storing at -80°C. Before RNA extraction, samples were ground in liquid
219 nitrogen and RNA was extracted using the RNeasy Plant Minikit (Qiagen) according to the
220 manufacturer's protocols (https://www.qiagen.com/us/resources/download.aspx?id=246847e7-
221 0095-43e4-8d1d-41df3f9153dd&lang=en). After checking the RNA quality by Bioanalyzer,
222 ~350 bp library was constructed and sequenced by Illumina NovaSeq 6000 platform (Annoroad
223 Gene Technology Co., Ltd, Beijing, China). Before transcriptome assembly, raw reads were
224 trimmed using Trimmomatic v0.36 (Bolger et al. 2014) with the settings: ILLUMINICLIP
225 2:30:10 LEADING 3, TRAILING 3 SLIDINGWINDOW 4:10 and MINLEN 50. A reference-
226 guided transcriptome assembly was performed in Trinity v2.8.5 (Grabherr et al. 2011) using
227 combined reads from germinated spores and infected leaves.

228 **2.5 | Repeats annotation and LTR aging**

229 To compare the repeat contents between *P. polysora* (*Ppz*_GD1913) and other rust species, the
230 genome data of *Melampsora larici_populina* (*Mlp*_98AG31), *P. sorghi* (*Psz*_RO10H11247), *P.*
231 *coronata* (Pca203), *P. graminis* (*Pgt*21-0), *P. triticina* (*Pt*76) and *P. striiformis* (*Pst*104E) were
232 downloaded from NCBI or MycoCosm (DOE Joint Genome Institute), and *Austropuccina psidii*
233 (*Ap*_Au_3) was downloaded from European Nucleotide Archive (ENA) at EMBL-EBI. The
234 repeat contents were identified by a merged library including a self library in RepeatMasker
235 v4.0.8 (Smit et al. 2015) and a genome-trained library from RepeatModeler v1.0.11 (Smit and
236 Hubley 2008). The retrotransposons with long terminal repeats (LTR-RTs) of GD1913 were
237 identified with an in-house pipeline that uses *LTRharvest* (Ellinghaus et al. 2008) packaged in
238 LTR_retriever v2.9.0 (Ou and Jiang 2018). The classification of LTR-RTs was annotated based
239 on REXdb (http://repeatexplorer.org) (Neumann et al. 2019) performed by TEsorter (Zhang et al.
240 2022).The genetic distances (GD) of each pair of LTR-RTs were calculated by *dist.dna* in R
241 package of *ape*. The insert times (T) of LTR-RTs were estimated following the formula, T =
242 GD/(2*mutation_rate). We used a mutation rate of 2.0×10^{-8} from a closely related fungus,
243 *Schizophyllum commune* (Baranova et al. 2015).

244 **2.6 | Gene prediction and genome annotation**

245 The two haplotypes were annotated independently using Funannotate v1.8.7
246 (https://github.com/nextgenusfs/funannotate/releases/tag/v1.8.7). A pipeline of core modules was
247 applied as mask-> train->predict->update-> fix->annotate. The repeats of the assembled genome
248 were soft-masked according to the results of 2.5. The prediction step (funannotate predict) was

run with transcript evidence from Minimap 2 (Li 2018) RNA-seq alignments and genome-guided Trinity assemblies. Transcript evidence was aligned to two haplotypes separately and the protein evidence was aligned to the genome via Diamond (Buchfink et al. 2015)/Exonerate (Slater and Birney 2005) with the default UniProtKb/SwissProt protein database (http://legacy.uniprot.org/uniprot/?query=reviewed%3Ayes) from funannotate. The PASA gene models were parsed to train AUGUSTUS v3.2.3 (Stanke and Morgenstern 2005), snap and GlimmerHMM. In addition, GeneMark_ES (Lomsadze et al. 2005) was self-trained using two haplotypes' sequences. All above evidence was combined with default weight settings using Evidence Modeler (Haas et al. 2008) and filtered by removing genes with short length (< 50 aa), spanning gaps and transposable elements. The tRNA genes were predicted using tRNAscan-SE v1.3.1 (Lowe and Chan 2016). Funannotate update command to add UTR data to the predictions and fix gene models that are in disagreement with the RNA-seq data. Functional annotation was performed based on available databases including Pfam v34.0) (Finn et al. 2014), InterPro (v86.0) (Jones et al. 2014), eggNOG (v5.0) (Huerta-Cepas et al. 2019), UniProtKB (v.2021_03), MEROPS (v12.3) (Rawlings et al. 2016), carbohydrate hydrolyzing enzymatic domains (CAZymes) (Terrapon et al. 2017) and a set of transcription factors based on InterProScan domains to assign functional annotations.

2.7 | Interhaplotype variation analysis

Small variants including SNPs and indels were identified by mapping trimmed short reads from DNA against haplotype A with BWA-MEM v0.7.8 (Li and Durbin 2010). After removing PCR duplicates using Picard v2.18.27 (https://github.com/broadinstitute/picard), the bam file was input to GATK 4.1.9 (https://github.com/broadinstitute/gatk) to call SNPs. SNPs and Indels were filtered using gatk VariantFiltration with parameters QD > 2.0 & MQ > 40.0 & FS < 50.0 & SQR < 3.0 for SNPs and QD > 2.0 & QUAL > 30.0 & FS < 200.0 & ReadPosRankSum>-20.0. Variants were annotated from genome location and functional impact using SnpEff v4.3 (Cingolani et al. 2012). To estimate structure variations (SVs) between two haplotigs, we aligned 18 chromosomes of haplotype A to their corresponding chromosomes in haplotype B by using the nucmer program in Mummer 4 (--maxmatch -t 100 -l 100 -c 500) (Marçais et al. 2018). The delta file was filtered by parameters, delta-filter -i 95 -l 1000. To make sure the filter parameters acceptable, *show-coords* and *show-aligns* were used to check alignment details. Finally, the chromosome-pair synteny was visualized by *mummerplot*. In addition, the chromosome-pair

280 alignments were also analyzed using Assemblytics (http://assemblytics.com/) to estimate SVs of
281 three major categories, including insertion/deletion, tandem repeats contraction/expansion and
282 repeats contraction/expansion. The running parameters of Assemblytics were set with the unique
283 sequence length=10,000, maximum variant size=10,000 and minimum variant size=1. To
284 identify the orthologous genes between two haplotypes, we used OrthoFinder v2.5.4 with the
285 default parameters (Emms and Kelly 2015).

286 **2.8 | Prediction of candidate effectors**

287 Secreted proteins were predicted based on two rules 1) the presence of a predicted signal peptide
288 using SignalP 4.0 (Petersen et al. 2011); and 2) the absence of predicted transmembrane domains
289 outside the first 60 amino acids (with TMHMM 2.0). To predict candidate effectors, we use
290 EffectorP 3.0 (Sperschneider and Dodds 2021), which applied two machine learning models
291 trained on apoplastic and cytoplasmic effectors. The density plots for genes, repeats, and secreted
292 proteins from chromosomes in two haplotypes were generated using KaryoploteR (Gel and Serra
293 2017) with the scripts available at https://github.com/jimie0311/Puccinia-polysora-
294 genome/blob/main/Figure-4/Figure-4B.md. A recent study identified the first effector of *P.*
295 *polysora*, *AvrRppC* (Deng et al. 2022). To locate its genome position, we blasted the CDS
296 sequence of $AvrRppC^{ref}$ against genome sequences of two haplotypes in this study.

297 To understand the expression pattern of secreted proteins, we used all expression data from
298 germinated spores and six time-points to perform a differential expression analysis. The allele
299 specific mapping of Kallisto v0.48 (Bray et al. 2016) was applied to obtain the expression levels
300 in transcripts per million (TPM) of genes on both haplotypes. FeatureCounts v1.5.3 (Liao et al.
301 2014) was used to generate read counts for each gene model of secreted protein. Differentially
302 expressed genes were identified by expression in plants relative to germinated spores (|log fold
303 change| > 1.5; adjusted $P < 0.1$) using the DESeq2 R package (Love et al. 2014). The average
304 rlog-transformed values for each gene were used for clustering using the *k*-means method. The
305 optimal number of clusters was defined using the elbow bend method
306 (https://www.delftstack.com/howto/r/k-means-clustering-r) and circular heatmaps were plotted
307 using the Circlize R package (Gu et al. 2014).

308 **2.9 | Comparative genome analysis**

309 The high quality of *P. polysora* genome provides the opportunity for comparison with other
310 chromosome-level references, e.g., *Pgt*21-0 (Li et al. 2019), *Pt*76 (Duan et al. 2022) and *Pca*203

311 (Henningsen et al. 2022). Syntenic gene pairs among *Puccinia* species were identified using the
312 MCSCAN toolkit (Wang et al. 2012). Figures were plotted using MCscan (Python version).
313 (https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)). Orthofinder v2.5.4 (Emms
314 and Kelly 2019) was applied to identify the orthogroups of six rust species with available
315 haplotype genomes. Besides the above three *Puccinia* species and *P. polysora* in this study, the
316 primary haplotype of *P. striiformis* (*Pst*104E) and *Austropuccinia psidii* (Au_3) were also
317 included. The phylogenetic relationship of these species was visualized by FigTree v1.4.3
318 (http://tree.bio.ed.ac.uk/software/figtree/). The gained (+)/lost (–) gene numbers were calculated
319 by comparing the two most close species according to Orthogroups.GeneCount.tsv in
320 Orthofinder output.

321 **2.10 | Population genetic analyses**

322 To understand the population differentiation of *P. polysora* in China, 79 isolates were
323 resequenced. Reads were qualified by Trimmomatic v0.36 (Bolger et al. 2014) using the
324 parameters described in 2.4 RNA extraction and sequencing. To assess whether each isolate
325 comprises a single genotype free of contamination, the distribution of read counts for bi-allelic
326 SNPs was calculated by R package vcfR v1.8.0 (Knaus and Grunwald 2017) and plotted using
327 ggplot2 (v2.2.1) (Wickham 2016). The normal distribution of read allele frequencies at
328 heterozygous positions is expected to rule out contamination from other *Ppz* genotypes and only
329 the pure isolates were applied in the following population genomic analyses. All clean reads of
330 pure isolates were mapped to haplotype A and SNPs calling and filtration were performed by
331 GATK v4.1.9 as described in 2.7 Interhaplotype variation analysis section. Population structure
332 was detected and quantified using principal component analysis (PCA), which was performed
333 using Plink v1.9 (Purcell et al. 2007). Pairwise FST values were estimated between geographical
334 regions using the PopGenome R package (v2.6.1) (Pfeifer et al. 2014). To assess the effects of
335 variants, the vcf file including selected SNPs was run in SnpEff v4.3 (Cingolani et al. 2012).
336 Because the sexual stage of *P. polysora* has not been reported, we used the standardized index of
337 association (*rd*) to test linkage disequilibrium in the Chinese *Ppz* population. We constructed 100
338 sets of 10,000 random SNPs by samp.ia function from R package *poppr* (Kamvar et al. 2014) to
339 generate a distribution of *rd* values. The observed *rd* distribution of the Chinese population was
340 compared to the distribution of 10,000 *rd* values constructed using fully randomly simulated
341 datasets with 0%, 50%, 75% and 100% linkage.

342 **3 | RESULTS**

343 **3.1 | Chromosome-level genome assembly and haplotype-phasing of *P. polysora***

344 We generated a total of 53 Gbp of circular consensus reads (32× coverage) from single-molecule
345 real-time sequences on the PacBio Sequel II. A total of 4056 contigs were assembled with a
346 genome size of 1.76 Gbp and a contig N50 of 1.9 Mbp (Table 1). A total of 1627 contigs,
347 accounting for ~2% of the *de novo* assembly size were excluded due to small size, low coverage
348 or high mitochondrial similarity (Table S2). By using Haplomerger2, the remaining contigs were
349 assigned to two haplotypes, A and B, with 613 contigs representing ~862 Mbp and 1321 contigs
350 representing ~852 Mbp, respectively. These contigs were further connected to 173 and 344
351 scaffolds in haplotypes A and B by Hi-C scaffolding (Table 1), after manual checking of 33
352 duplicated scaffolds between (17) and within (16) haplotypes (Table S3). By considering both
353 Hi-C contact maps and the all-vs-all alignment from Haplomerger2, 18 super scaffolds were
354 binned for each haplotype manually. The Hi-C contact map showed evidence of a single
355 centromere on each scaffold and all scaffolds contained telomere sequences at either end (Table
356 1, Figures 1B, 1C), indicating that each haplotype contains 18 chromosomes. This is consistent
357 with the observations of 18 chromosomes in the genome of other *Puccinia* species (Boehm et al.
358 1992; Li et al. 2019; Wu et al. 2021; Duan et al. 2022). In addition, 15 (6.5 Mbp) and 20 (4.5
359 Mbp) scaffolds in haplotype A and B respectively could not be correctly connected to any
360 chromosomes due to short length or insufficient Hi-C signal. The 18 chromosomes of haplotype
361 A showed an average of over 80% Hi-C links to haplotype A, suggesting high levels of nuclear
362 phasing (Figure 1D). The chromosomes were numbered according to the synteny with the other
363 *Puccinia* species (see below). In addition, a cyclized mitochondrial sequence (67 Kb) was
364 assembled.

365 **3.2 | Repeats in the supersize genome of *P. polysora***

366 The completeness of the *Ppz* assembly was assessed based on highly conserved Basidiomycete
367 genes (Badidiomycete_odb9, 1335 BUSCOs), which suggested that the two haplotypes showed a
368 similar level of completeness, with approximately 90% complete BUSCO genes and an
369 additional 3.5% fragmented BUSCOs, respectively (Table 1). However, 126 BUSCO genes
370 were complete in haplotype A but not present in haplotype B. On the contrary, 127 complete
371 BUSCO genes are only present in haplotype B, which suggests interhaplotype variation in *P.*
372 *polysora*. Considering the two haplotypes in combination, only 11 BUSCOs were missing or

373 fragmented, resulting in 99% complete BUSCOs. That means we assembled a near-complete
374 genome of *P. polysora*.
375 Using the *de novo* library of RepeatMasker and genome-trained library of RepeatModeler,
376 RepeatMasker detected repeats accounting for 85% of the diploid assembly (both haplotypes),
377 which is significantly higher than observed in closely related species of *Puccinia* (30%–59%)
378 (Miller et al. 2018; Schwessinger et al. 2018; Wu et al. 2021), but similar to *Austropuccinia*
379 *psidii* (85% estimated in this study but 91% in its original description) (Tobias et al. 2021). The
380 repeat density and low GC content (~40%) along the two haplotypes were illustrated in Figure
381 2A. The large percentage of repeats in *P. polysora* was mainly caused by Class I
382 retrotransposons with long terminal repeats (LTR), with the LTR-Gypsy superfamily most
383 abundant (33.5%), and LTR-Copia next (17.9%). The TE family composition was similar to
384 other rust species, except for *A. psidii*, which is LTR-Gypsy dominated with LTR-Copia
385 accounting for only 1.9% (Figure 2B). In addition, about 20% and 16% of repeats of *P. polysora*
386 in two haplotypes were unclassified repeat families based on Repbase (Bao et al. 2015).
387 Since LTR-RTs occupied the majority of the *P. polysora* genome, we examined whether
388 these repetitive sequences slowly accumulated over time or alternatively were subject to sudden
389 expansion in the life history of *P. polysora*. A total of 41,634 LTR-RT pairs were extracted. An
390 LTR-RT burst was estimated at around 1.7 Mya (Figure 2C).
391 **3.3 | High levels of interhaplotype gene content and structural variation**
392 RNA-seq reads from germinated spores and infected corn leaves at 1, 2, 4, 7, 10, and 14 dpi
393 (Table S4) were pooled and used to generate genome-guided transcriptome assemblies. In total,
394 we annotated 23,270 and 24,242 gene models on haplotype A and haplotype B, respectively
395 (Table 2). The protein and CDS sequences are available from
396 https://github.com/jimie0311/Puccinia-polysora-genome. The gene space of these genes accounts
397 for only ~4% of the total genome size. The two haplotypes showed a similar level of functional
398 annotation with about 51% of proteins having at least one functional annotation. Orthofinder
399 identified a total of 13,057 common orthogroups between the two haplotypes, involving 20,802
400 genes from haplotype A and 21,519 genes from haplotype B, over 90% in each haplotype. In
401 addition, 198 and 235 orthogroups were specific to haplotype A or B respectively. About 70% of
402 haplotype-specific genes have no functional annotations, with the remainder having annotated

403 functions mainly in RNA mediated transposition, transmembrane amino acid transposition or
404 ATP binding (Table S5).
405 By mapping Illumina reads to only haplotype A, we detected a total heterozygous rate of
406 6.1 variants/kbp, in which the SNP variant rate is dominated (5.8 SNPs/kbp). About 78% of the
407 SNPs were located in intergenic regions and the rest (22%) were detected in 22,802 protein-
408 coding genes. The results of Assemblytics suggested that structural variation between haplotype
409 A and B comprised about 3.0% (26/849 Mbp) of the haploid genome size (Figure 3A). The full
410 chromosome-pair alignment between two haplotypes is illustrated in Figure 3B, with alignments
411 of chromsome09 and chromsome15 in Figure 3C and D, showing some insertions/deletions and
412 inversions, respectively. Among three types of variation, insertions/deletions and repeat
413 expansions/contractions are more prevalent than tandem expansions/contractions. Besides,
414 variation with size bins of 500 to 10,000 bp is the most prevalent, which accounts for 2.8%.
415 Because the maximal variant size was restricted to 10,000 in Assemblytics, the actual difference
416 between the two haplotypes could be even higher than estimated.

417 **3.4 | Conservation of expression patterns between secreted proteins from two**
418 **haplotypes**

419 We predicted 1183 and 1251 secreted proteins on haplotype A and haplotype B, respectively. By
420 using the machine-learning tool, EffectorP v3.0, about 14% and 27% of all secreted proteins
421 were predicted as apoplastic and cytoplasmic effectors, respectively (Table 2). The protein
422 sequences of secreted proteins are available from https://github.com/jimie0311/Puccinia-
423 polysora-genome. GO enrichment analysis and Interproscan annotation predicted some effectors
424 involved in hydrolase activity, catalytic activity, protein tyrosine phosphatase activity and metal
425 ion binding, but most effectors had no homology with known or predicted functions (Table S6,
426 Figure S2). Then we used gene expression data to predict clusters in the secretome that are
427 differentially expressed during infection. The elbow bend plot (Figure S4) suggested the total
428 within-group sum of squares tend to be level off. Therefore, the expression data of secreted
429 proteins were assigned to seven clusters. Genes in cluster 1 showed high expression in
430 germinated urediniospores and early infection (1dpi, 2dpi) but low expression at 4dpi and 7dpi
431 when haustoria form. On the contrary, genes in clusters 2, 3, 4 and 5 showed low expression in
432 germinated spores, and highest in planta expression at days 1-2 (cluster 2), days 2-7 (cluster 3)
433 and from days 4-7 (cluster 4) and days 7-14 (cluster 5) Clusters 6 and 7 were more uniform in

434 expression through these stages, although cluster 6 genes increased later in infection. About 24%
435 to 67% of the secreted proteins in these clusters were predicted as candidate effectors by
436 EffectorP (Table S7). A similar set of expression profile clusters were also detected for genes in
437 haplotype B (Figure 4A). Nevertheless, two haplotypes presented different levels of
438 carbohydrate-active enzymes (CAZymes). A total of 209 and 298 CAZymes were detected, of
439 which 9 vs 79 CAZymes were predicted to be secreted in haplotypes A and B, respectively
440 (Table S7). Among CAZymes subclasses, Glycoside hydrolase (GH) enzymes are abundant,
441 accounting for 48% (88 GH families) and 42% (126 GH families) of all CAZymes of haplotype
442 A and haplotype B, respectively. Of these, 51 GH families were predicted to be secreted. The
443 GH5 (cellulase and other diverse forms being exo-/endo-glucanases and endomannannases)
444 family (Langsto et al. 2011) was observed to be largely expanded in *P. polysora* as well as other
445 *Puccinia* species (Figure 5A). Besides, AA3 (glucose-methanol-choline oxidoreductases), CE4
446 (chitin and peptidoglycan deacetylases), GH18 (chitinases of classes III and V), CH47(α-
447 mannosidases), GT2 and GT90 are also abundant in *P. polysora*.
448 We also investigated other factors potentially associated with infection or plant immune
449 inhibition, such as transcription factors (TFs) and peptidases. We found most TF families of *P.*
450 *polysora* have low abundance except C2H2-type (IPR013087) and CCHC-type (IRP001878)
451 zinc finger class, which have 87 and 67 members in haplotypes A and B (Figure 4B). In total, we
452 annotated 286 and 304 proteases in haplotypes A and B, with A01A (aspartic proteases), C19
453 (ubiquitinyl hydrolases), C26 (gamma-glutamyl hydrolase), S08A (subtilisin-like serine
454 protease), S09X (glutamyl endopeptidase C), S10 (carboxypeptidase Y), S33 (prolyl
455 aminopeptidase) and T01A (component peptidases of the proteasome) families expanded in *P.*
456 *polysora* as well as in other three *Puccinia* species (Figure 5C). A total of 8 (2.8%) and 30
457 (10.0%) proteases were predicted to be secreted and these proteases showed a discrete
458 distribution in each cluster without an obvious clustering pattern (Table S7).

459 **3.5 | Density of candidate effectors and *AvrRppC* anchoring**

460 Analysis of the local gene density, measured as flanking distances between neighboring genes
461 showed that the flanking distances in the *P. polysora* genome are generally rather far, with an
462 average distance between genes of 100 kbp (Figure 4B). Accordingly, the surrounding genomic
463 context of most genes in *P. polysora* genome is gene-sparse and repeat-rich. Large flanking
464 distances are not specific to candidate effectors. In line with this pattern, the gene distance

465 density plots revealed very similar distributions between all genes and candidate effectors,
466 including the only known *Avr* effector gene, *AvrRppC* (Figure 4B). We further investigated
467 whether candidate effectors present a different distribution of gene distance density compared to
468 basidiomycete core ortholog genes of *P. polysora*. The results highlighted that candidate
469 effectors are not located in peculiar gene-sparse areas, and the flanking distance centers of both
470 BUSCOs and candidate effectors overlap with that of all genes (Figure 4B).

471 A previous study identified the avirulence effector, *AvrRppC*, and found six allelic variants
472 which were named as $AvrRppC^{ref}$, $AvrRppC^{A}$, $AvrRppC^{C}$, $AvrRppC^{E}$, $AvrRppC^{F}$, $AvrRppC^{J}$
473 (Deng et al. 2022). By searching the CDS sequence of $AvrRppC^{ref}$ against haplotypes A and B
474 we anchored *AvrRppC* at ~ 9 Mbp on chromosome 14. Although the gene was not present in the
475 original annotation, there was evidence of transcription from RNA-seq reads at 1 dpi and 7 dpi
476 (Figure 4C). Additionally, we found the *AvrRppC* allele on Chromosome14B is $AvrRppC^{ref}$ type
477 whereas the allele in Chromosome 14A represents a new allele type, named $AvrRppC^{1}$ which has
478 two amino acids changes compared with $AvrRppC^{ref}$ and is different from the other five
479 previously reported variants.

480 **3.6 | Comparisons of synteny and orthogroups between *P. polysora* and its close**
481 **relatives**

482 Although *P. polysora* has a 7–10 times larger genome size than other rust species, the gene
483 synteny amongst four *Puccinia* species with full chromosome assemblies is well conserved, with
484 18 chromosomes corresponding to each other clearly without any chromosome rearrangement
485 (Figure 6A). A few large inversion blocks were detected in chromosomes 1, 3, 4, 7, 8, 9 and 10
486 between *Ppz*A and *Pt*76 (Figures 6A, S3). Inversion blocks were also detected in chromosomes 5,
487 7, 10, 13 between *Pt*76 and *Pca*203, and chromosomes 2, 3, 5, 7, 8, 10 and 13 between *Pgt*21-0
488 and *Pca*203 (Figures 6A, S3).

489 The protein orthologs among four closely related *Puccinia* species were compared based on
490 haplotype A of each species. The increased and reduced genes in all orthologs between two close
491 species are labeled in the phylogenetic tree (Figure 6B). It was apparent that the reduced genes
492 were much more than gained genes when compared from *P. polysora* to other *Puccinia* species.
493 Therefore, *P. polysora* might experience gene expansion events compared with the other three
494 species (Figure 6B, Table S8). Among the four species compared, a total of 11,840 orthogroups
495 were identified. Genes of *P. polysora* were assigned to 7,148 orthogroups which was less than

496 those of over 8,000 orthogroups for the other three *Puccinia* species. Nevertheless, *P. polysora*
497 presented more species-specific orthogroups and these orthogroups are featured with a high copy
498 number of gene duplication (Figure 6C). Based on Eggnog annotation result, the orthogroups of
499 *P. polysora* with high copy numbers (>10) are mainly associated with the pathogenicity-related
500 function, energy metabolism, RNA regulation and heat shock reaction (Table S9).

501 **3.7 | Population genomics analyses suggested lineage divergence and predominant**
502 **asexual mode of reproduction**

503 To understand the genetic structure of *P. polysora* in China, we analyzed genome resequencing
504 data from 79 isolates of *Ppz* collected from across maize-growing regions of China. The analyses
505 of bi-allelic frequency suggested the expected normal distribution of bi-allelic frequency for
506 most isolates except GD1922-3 and GX1905-2 (Figure S5) which were excluded from the
507 analysis. After removal, a total of 7,147,489 whole-genome SNPs were obtained from the
508 remaining 77 isolates. PCA analysis indicated the presence of two genetic groups, with a major
509 group consisting of 73 isolates (blue circle in Figure 7B) and a minor group containing only four
510 isolates (red circle in Figure 7B). However, this separation was independent of the geographic
511 origin of the isolates (North, Central, and South China). Although higher genetic differentiation
512 was revealed between North China and South China populations (Fst = 8.1×10^{-4}) than in other
513 regional pairs (3.9×10^{-4} and 3.6×10^{-4}), all pair-wise *Fst* values were close to zero indicating a
514 lack of geographic differentiation. We used the standardized index of association, *rd*, to estimate
515 the linkage disequilibrium of *P. polysora* population in China. The observed *rd* distribution for
516 all isolates was between the values calculated from simulated datasets with 0% linkage, a sexual
517 population, and 50% linkage (Figure 7C). The result suggested that, although *Ppz* population in
518 China showed as a clonal population, its evolutionary history may be influenced by some extent
519 of sexual reproduction.

520 Due to limited samples in the minor group, we did not calculate the Fst between two genetic
521 groups instead of comparing the highly differentiated SNPs between two genetic clusters. A total
522 of 32,274 SNPs (Fst > 0.9) were filtered and these SNPs were evenly distributed in 18
523 chromosomes without region specificity. These variants were annotated in 6,752 protein-coding
524 genes, of which 367 genes were affected by variants. The 20 genes were highly impacted by
525 gaining or losing start/stop, and 250 genes showed missense variants affecting amino acid
526 sequences. The known functional annotations of these 367 genes were mainly related to

527 transmembrane transport, signal transduction, zinc ion/protein/nucleic acid binding, catalytic
528 activity and protein kinase activity (Table S10). In addition, 14 of these 367 genes were
529 annotated as secreted proteins including *AvrRppC.* The allele types of *AvrRppC* varied between
530 the two genetic groups. In the major group, most isolates (93%) carried allele types of $AvrRppC^{1}$
531 and $AvrRppC^{ref}$. A few isolates showed the allele combination containing new allele types,
532 $AvrRppC^{1}$ to $AvrRppC^{5}$ as well as $AvrRppC^{ref}$. Whereas in the minor group, all four isolates
533 carried alleles of $AvrRppC^{A}$ and $AvrRppC^{J}$ (Figure 7D). The CDS and amino acid sequences for
534 new allele types are listed in Table S11.

535 **4 | DISCUSSION**

536 As rust fungi are dikaryotic pathogens, obtaining their nuclear phased assembly is critical for
537 pathogenicity studies. In this study, we reported a haplotype-phased and chromosome-scale
538 genome of *P. polysora* based on HiFi reads and Hi-C data. The 18 chromosomes had telomeres
539 at both ends and showed high completeness (~90% complete BUSCOs for each nucleus and 99%
540 complete BUSCOs for two nuclei) and high continuity (N50 of scaffolds = 54 M) as well as over
541 80% nuclear phasing. These data strongly supported that we generated a high-quality reference
542 genome for *P. polysora*. In addition, our study supplied a successful case to assemble a gigabase-
543 sized fungal genome to chromosome level. The assembly pipeline we illustrated in Figure 1 will
544 inform future genome assembly for other dikaryotic or non-haploid organisms.

545 **4.1 | TEs amplification leads to genome expansion of *P. polysora*, possibly in response**
546 **to ancestral host decay**

547 Compared to its close relatives, *P. polysora* has experienced a large genome expansion,
548 with its genome size of 1.71 Gbp equivalent to 7–10 times of four other *Puccinia* species
549 assemblies (*Pst*-104E, *Pgt*21-0, *Pt*76, and *Pca*203) (Schwessinger et al. 2018; Li et al. 2019; Wu
550 et al. 2021; Henningsen et al. 2022). Based on a limited set of rust species, Tavares et al (2014)
551 inferred that rusts infecting monocot Poaceae hosts have considerably smaller genomes than
552 rusts infecting dicot Fabaceae hosts (556.6 Mbp on average). Thus, our study supplied an
553 exception to this observation, indicating it may not be a general model. Gene duplication, repeat
554 content variation and ploidy variation are common mechanisms to account for the genome
555 expansion of fungi (Castanera et al. 2016; Sipos et al. 2017; Todd et al. 2017). Although gene
556 expansion or loss has happened in the evolutionary process from *P. polysora* to other wheat-

557 associated *Puccinia* species (Figure 6B) and correspondingly the gene number predicted in *Ppz-*
558 GD1913 (~24000 per haplotype) is much more than those of its two close relatives, *P. coronata*
559 (~18000) and *P. striiformis* (~14000) (Miller et al. 2018; Schwessinger et al. 2018), the
560 expanded genes only contribute to a size difference of 7 Mbp and 17 Mbp (Table S12). On the
561 contrary, repeat analyses suggested that genome expansion of *P. polysora* is due to its large
562 repeat contents (85% of the total genome), notably a proliferation of LTR-RTs, which is
563 consistent with the common evidence for rust genome expansion (Tobias et al. 2021). The
564 evidence of genome expansion has also been detected in other obligate fungi, such as powdery
565 mildew (PM) fungi (Spanu et al. 2010; Frantzeskakis et al. 2020) and symbiotic fungi (Miyauchi
566 et al. 2020). PM fungi lose carbohydrate-active enzymes, transporters, etc. which are probably
567 redundant genes in strict parasitism as tradeoffs (Spanu et al. 2010). The low abundance of
568 PCWDEs and transcription factors in *P. polysora* as well as other *Puccinia* species was in line
569 with the above pattern (Figure 5) (Duplessis et al. 2011, Miller et al. 2018). Despite the massive
570 TE proliferation causing a significant enlargement in genome size, *P. polysora* shows very good
571 gene synteny with other *Puccinia* species (Figure 5). This overall chromosome synteny was
572 consistent with the dominant synteny detected by comparing BUSCO genes (Henningsen et al.
573 2022). In addition, Edwards et al (2022) reported that *Austropuccinia psidii*
574 (Sphaerophragmiaceae, Pucciales), was in parallel with species in Pucciniaceae, Pucciales,
575 indicating that conservatism of chromosomal synteny can exist at a family level.
576 In fungi, TEs have been implicated in coevolution with the host, genome architecture,
577 plasticity, and adaptation (Lorrain et al. 2021). The "TE-thrust hypothesis" proposed that TEs
578 can act to generate genetic novelties for organisms leading to speciation or may be related to
579 adaptation to new challenges, e.g. new host or rapidly changing climate (Oliver and Greene 2012;
580 Zeh et al. 2009). We detected a TE burst of *P. polysora* at around 1.7 Mya (Figure 2C) and the
581 phylogenetic tree (Figure 6B) suggested that *P. polysora* speciated earlier than the divergence
582 time between *P. coronata* and *P. striiformis* (40 Mya) (Aime et al. 2018). So, TE burst of *P.*
583 *polysora* may be unlikely for speciation. Most scientists agree that maize originated in central
584 Mexico and domesticated at 1–0.5 Mya from a wild grass called teosinte (Heerwaarden et al.
585 2011). At about 1.7 Mya, teosinte but not maize could be the major host of *P. polysora*. However,
586 during the late Pleistocene, there may be a decline in the distribution of tropical plants, including
587 teosinte during the cool and drying climate in Mexico (Piperno et al. 2007), leading to a drop in

588 the effective population size of *P. polysora*. As reviewed by Lorrain et al (2021), host population
589 dynamics can impact TEs landscapes in Fungi. The de-repressed TEs will cause genome
590 expansion until they decay by being silenced or mutated by host defenses (Le Rouzic and Cayp
591 2005). Therefore, the TE burst at ~1.7 Mya could be one of the evolutionary events to support
592 and maintain the genetic variation of *P. polysora* while going through abiotic stress and
593 environmental change. Another explanation could be the "nearly neutral theory", in which varied
594 genomic features (e.g. large genome size, TE content, introns) are not initially harmful but are
595 passively fixed by mutation or random genetic drift (Lynch and Conery 2003). With more
596 available genome resources in the future, the relationship between effective genome size and the
597 adaptative evolution of *P. polysora* needs to be answered.

598 **4.2 | Obligate biotrophic fungi lost the evolutionary feature of "two-speed genome"**

599 The "two-speed genome" concept has been put forward to highlight the over-representation
600 of effector-like genes in the repeat-rich and gene-sparse genome in many filamentous pathogens
601 (Dong et al. 2015, Frantzeskakis et al. 2019). These TE-rich invasion/blocks may contribute to
602 extensive chromosomal reshuffling and the rise of accessory chromosomes (de Jonge et al. 2013).
603 Pathogens can resolve the evolutionary conflict through rapid evolution of effector genes to
604 adapt to changing environments but maintain the housekeeping genes in the core genome with a
605 moderate evolution rate (Presti et al. 2015). However, the situation is clearly different in rust
606 fungi. There was no signal of effector compartments detected in Oat crown rust, wheat stripe rust
607 and myrtle rust pathogens when comparing the intergenic distance between all genes and
608 effectors (Schwessinger et al. 2018; Miller et al. 2018; Tobias et al. 2021). In *P. polysora*, the
609 numerous TEs are widely dispersed throughout the genome (Figure 2). The intergenic distance
610 distribution of candidate effectors overlapped with that of all genes and BUSCOs (Figure 7).
611 This genomic architecture with TEs and genes interspersed was also reported in another
612 biotrophic fungal group, powdery mildew (PM) fungi (Frantzeskakis et al. 2020). Similarly, AT-
613 rich isochores or large-scale compartmentalization are also missing in PM fungi (Frantzeskakis
614 et al. 2020). When compared with other *Puccinia* species, the high repeat content of *P. polysora*
615 leads to much lower relative gene space (4% in *Ppz* vs 20%–35% for currently reported *Puccinia*
616 species) but greater intergenic distance expansion (average 100 kbp of *Ppz* vs average 1–2 kbp
617 for *Pst* and *Pca*) (Figure 4B) (Miller et al. 2018; Schwessinger et al. 2018).

618 **4.3 | Population genomics data accelerate the identification of potential avirulent genes**

619 In the "arms race" between rust species and plants, the *Avr* genes in rust pathogens have
620 been the subject of mutations to avoid the recognition by host resistance genes (Cui et al. 2015).
621 Therefore, monitoring the variation of *Avr* genes in rust populations is a priority for disease
622 management. Based on whole-genome SNPs, we detected two genetic groups of *P. polysora* in
623 China. The highly differentiated SNPs between these two groups suggested mutations in 16
624 secreted proteins including the only verified avirulence gene, *AvrRppC*. A recent study
625 investigated the allele types of *AvrRppC* in a Chinese *P. polysora* population (Deng et al. 2022).
626 Isolates with allele types of $AvrRppC^{A}$, $AvrRppC^{F}$ and $AvrRppC^{J}$ could escape the recognition of
627 *RppC* causing southern corn rust but $AvrRppC^{ref}$, $AvrRppC^{E}$, and $AvrRppC^{C}$ are avirulent alleles
628 that trigger *RppC*-mediated resistance. In our study, isolates in the major group almost all carried
629 the avirulence allele, $AvrRppC^{ref}$ and four isolates in the minor genetic group carried the virulent
630 alleles, $AvrRppC^{A}$ and $AvrRppC^{J}$ (Figure 7D). These results suggested that differentiation
631 between two genetic groups could be related to virulence evolution. Similarly, this kind of
632 relation was also detected in the wheat stripe rust pathogen (Hubbard et al. 2015). Upadhyaya et
633 al (2021) suggested that avirulent genes of *P. graminis* showed a similar expression pattern. In
634 this study, *AvrRppC* was classified in cluster 4 (Figure 4A) with low expression in germinated
635 spores and early infection but high expression in intermediate infection (7 dpi). Among 14
636 secreted proteins with Fst > 0.9 between two genetic groups, FUNA_005021 (cluster 3) and
637 FUNA_001724-T1 (cluster 5) showed similar expression patterns to *AvrRppC* and may be
638 candidate avirulent genes. In other rust species, the rust population could drastically shift
639 towards a wider or new spectrum of virulence over time (Miller et al. 2020; Bai et al. 2021),
640 therefore, performing long-term surveillance as well as developing effective virulence markers
641 are particularly critical for *P. polysora*.

642 **4.4 | Secretome could be an indicator to infer the presence of alternate hosts**

643 Rust fungi have complex life cycles involving asexual as well as sexual reproduction
644 (Figueroa et al. 2020; Zhao et al. 2021). In some species, sexual reproduction is rare or cryptic
645 and it can remain undiscovered. For example, it took a century to discover the alternate host of *P.*
646 *striiformis* (Jin et al. 2010). In our case, the sexual stage of *P. polysora* is still a mystery;
647 teliospores are rarely found in natural fields and germination tests of teliospores were not
648 successful in past attempts (Cammack 1959). Studies on population genetics of *P. polysora* are
649 very limited, possibly due to the lack of effective molecular markers and the reference genome.

650 Our data using whole-genome SNPs supported the clonal Chinese population of *P. polysora*,
651 however, the low *rd* between those from simulated with 0% and 50% linkage suggested the
652 potential influence of sexual reproduction. The absence of a sexual stage or alternate host may be
653 reflected by unequal gene numbers in the secretome. A supportive example is two species in
654 *Melampsora*. *Melampsora lini* (autoecious, no alternate host, ~800 secreted proteins) was found
655 to have much less secreted proteins and secreted plant cell wall-degrading enzymes (PCWDEs)
656 than its close relative, *M. larici-populina* (heteroecious, with both primary and alternate hosts,
657 ~1800 secreted proteins) (Presti et al. 2015). In our case, we identified 2,434 secreted proteins
658 for *P. polysora*, which was comparable to the secretomes of two heteroecious relatives, *P.*
659 *coronata* (~2500) and *P. graminis* (~2500) (Miller et al. 2018; Li et al. 2019). Of course, the
660 single *Melampsora* example did not allow for making a general conclusion. Varied quality of
661 genome assemblies and different prediction methods can lead to significant bias in secreted
662 protein predictions. However, it will be valuable to test this hypothesis with more available high-
663 quality rust genomes in future studies. If the hypothesis is true, our results may suggest the
664 existence of alternate hosts that have not been discovered yet for *P. polysora.*

665 **5 | CONCLUSION**

666 In conclusion, we successfully obtained the first nuclear phased and chromosome-scale genome
667 assembly of the serious fungal pathogen, *P. polysora*. The high-quality genome assembly and
668 fine-time-point transcriptome facilitate the comparative genomic analyses. The genome
669 expansion of *P. polysora* is driven by TEs, but this did not affect its gene synteny with close
670 relatives. TEs insertion in turn led to expanded intergenic distance. The investigation of genome
671 characteristics and functional features provided a fundamental resource for pathogenicity studies
672 as well as to understand the evolutionary mechanisms of genome expansion in rust fungi. In
673 addition, the population genomic data revealed low genetic differentiation in the Chinese *P.*
674 *polysora* population, which expanded from south to north in recent decades. However, a minor
675 genetic group with low frequency suggested the ongoing virulence evolution to evade
676 recognition by *RppC*, a major resistance gene in Chinese corn cultivars, alarming the need to
677 excavate additional resistance genes as soon as possible.
678

679 **ACKNOWLEDGEMENTS**

We would like to thank Xiufeng Liu, Peng Lu, Gongxian Liao, Shikun Pan, Xingshang Lu, Baoqin Tan, Qiusheng Luo for field diseased sample collection. We also appreciate valuable suggestions from Dr. Mao Jianfeng (Beijing Forestry University, Beijing) and Dr. Qi Wu (Institute of Microbiology, Chinese Academy of Sciences, Beijing) before submission. Clement K.M. Tsui is grateful to CAS PIFI for the award of visiting scientist fellowship.

AUTHOR CONTRIBUTIONS

The experiment was conceived and managed by Lei Cai and Junmin Liang. Field samples were collected by Junmin Liang, Yuanjie Li, Zhanhong Ma, Leifu Li and Keyu Zhang. Single-spore isolation, reproduction, DNA/RNA extraction and other biological materials used for sequencing were conducted by Junmin Liang and Yuanjie Li. Data analyses were performed by Junmin Liang. Analysis tools were assisted by Peter N. Dodds, Melania Figueroa and Jana Sperschneider. Junmin Liang wrote the manuscript and Peter N. Dodds, Jana Sperschneider, Clement K.M. Tsui, and Lei Cai reviewed the manuscript.

DATA AVAILABILITY. All raw sequence reads (PacBio, Illumina, Hi-C and RNA-seq data) generated in this study are available both in NCBI and the National Microbiology Data Center (NMDC, https://nmdc.cn/) under the BioProject of PRJNA881038 and NMDC10018113, respectively. The accession number for all raw data are listed in Table S13. The accession number for assembled genome was JAOOQA000000000 in NCBI and NMDC60042795 in NMDC. Scripts and pipelines used to construct figures and annotation files (gff3 and cds) are available at https://github.com/jimie0311/Puccinia-polysora-genome.

FUNDING. This work was financially supported by the National Sciences Foundation of China (NSFC 31972210 and NSFC 31725001) and the National Sciences and Technology Fundamental Resources Investigation Program of China (2021FY100900).

REFERENCES

1. Aime, M. C., & McTaggart, A. R. (2021). A higher-rank classification for rust fungi, with notes on genera. *Fungal Systematics and Evolution*, 7, 21–47.

711 2. Aime, M. C., Bell, C. D., Wilson, A. W. (2018). Deconstructing the evolutionary complexity
712 between rust fungi (Pucciniales) and their plant hosts. *Studies in Mycology*, 89,143–152.
713 3. Aime, M. C., McTaggart, A. R., Mondo, S. J., & Duplessis, S. (2017). Phylogenetics and
714 phylogenomics of rust fungi. *Advances in Genetics*, 100, 267–307.
715 4. Akdemir, K.C., & Chin, L. (2015). HiCPlotter integrates genomic data with interaction
716 matrices. *Genome Biololgy,* 16, 198.
717 5. Anderson, C., Khan, M. A., Catanzariti, A. M., Jack, C. A., Nemri, A., Lawrence, G. J.,
718 Upadhyaya, N. M., Hardham, A. R., Ellis, J. G., Dodds, P. N., & Jones, D. A. (2016).
719 Genome analysis and avirulence gene cloning using a high-density RADseq linkage map of
720 the flax rust fungus, *Melampsora lini*. *BMC Genomics,* 17, 667.
721 6. Bai, Q., Wan, A. M., Wang, M. N., See, D. R., & Chen, X. M. (2021). Molecular
722 characterization of wheat stripe rust pathogen (*Puccinia striiformis* f.sp. *tritici*) Collections
723 from Nine countries. *International Journal of Molecular Sciences*, 22, 9457.
724 7. Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase update, a database of repetitive
725 elements in eukaryotic genomes. *Mobile DNA*, 6, 4–9.
726 8. Baranova, M. A., Logacheva, M. D., Penin, A. A., Seplyarskiy, V. B., Safonova, Y. Y.,
727 Naumenko, S. A., Klepikova, A. V., Geraso,pv. E. D., Bazykin, G. A., James, T. Y., &
728 Kondrashov, A. S. (2015). Extraordinary genetic diversity in a wood decay
729 mushroom. *Molecular Biology and Evolution*, 32, 2775–2783.
730 9. Boehm, E. W. A., Wenstrom, J. C., McLaughlin, D. J., Szabo, L. J., Roelfs, A. P., &
731 Bushnell, W. R. (1992). An ultrastructural pachytene karyotype for *Puccinia graminis* f.sp.
732 *tritici*. *Canadian Journal of Botany*, 70, 401–413.
733 10. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for
734 Illumina sequence data. *Bioinformatics*, 30, 2114–2120.
735 11. Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near optimal probabilistic
736 RNA-seq quantification, Nature Biotechnology 34, 525–527.
737 12. Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using
738 DIAMOND. *Nature Methods*, 12, 59.
739 13. Cammack, R. H. (1959). Studies on *Puccinia polysora* Underw. III. Description and life
740 cycle of *P. polysora* in West Africa. *Transactions of the British Mycological Society*, 42,
741 55–58.

742 14. Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J.,
743 Grimwood, J., Pérez, G., Pisabarro, A. G., Grigoriev, I. V., Stajich, J. E., & Ramírez, L.
744 (2016). Transposable elements versus the fungal genome: impact on whole genome
745 architecture and transcriptional profiles. *PLoS Genetics*, 12, e1006108.
746 15. Chen, J., Upadhyaya, N. M., Ortiz, D., Sperschneider, J., Li, F., Bouton, C., ... & Dodds, P.
747 N. (2017). Loss of *AvrSr50* by somatic exchange in stem rust leads to virulence for *Sr50*
748 resistance in wheat. *Science*, 358, 1607–1610.
749 16. Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion. G. T., Clum, A., Dunn,
750 C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M.,
751 Luo, C., Ecker, J. R., Cantum D., Ranmk, D, R., & Schatz, M. C. (2016). Phased diploid
752 genome assembly with single-molecule real-time sequencing. *Nature Methods*. 13, 1050–
753 1054.
754 17. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Ruden,
755 D. M., & Lu, X. Y. (2012). A program for annotating and predicting the effects of single
756 nucleotide polymorphisms, SnpEff. *Fly (Austin)*, 6, 80–92.
757 18. Crouch, J. A., & Szabo, L. J. (2011). Real-time PCR detection and discrimination of the
758 southern and common corn rust pathogens *Puccinia polysora* and *Puccinia sorghi*. *Plant*
759 *Disease*, 295, 624–632.
760 19. Cui, H., Tsuda, K., & Parker, J. E. (2015). Effector-triggered immunity: from pathogen
761 perception to robust defense. *Annual Review of Plant Biology,* 66, 487–511.
762 20. de Jonge, R., Bolton, M. D., Kombrink, A., van den Berg, G. C. M., Yadeta, K. A., Thomma,
763 B. P. H. J. (2013). Extensive chromosomal reshuffling drives evolution of virulence in an
764 asexual pathogen. *Genome Research*, 23,1271–1282.
765 21. Deng, C., Leonard, A., Cahill, J. M., Meng, L., Li, Y.R., Thatcher, S. Li, X. Y., Zhao, X. D.,
766 Du, W. J., Li, Z., Li, H. M., Llaca, V., Fengler, K., Marshall, L., Harris, C., Tabor, G., Li, Z.
767 M., Tian, Z. Q., Yang, Q. H., Chen, Y. H., Tang, J. H., Wang, X. T., Hao, J. J., Yan, J. B.,
768 Lai, Z. B., Fei, X. H., Song, W. B., Lai, J. S., Zhang, X. C., Shu, G. P., Wang, Y. B., Chang,
769 Y. X., Zhu, W. L., Xiong, W., Sun, J., Li, B. L., & Ding, J. Q. (2022).The RppC-AvrRppC
770 NLR-effector interaction mediates the resistance to southern corn rust in maize. *Molecular*
771 *Plant,* 15, 904–912.

772 22. Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-
773 pathogen interactions. *Nature Reviews Genetics*. 11, 539–548.
774 23. Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous
775 pathogens: waltz with plants. *Current Opinion in Plant Biology*, 35, 57–65.
776 24. Duan, H., Jones, A. W., Hewitt, T., Mackenzie, A., Hu, Y., Sharp, A., Lewis, D., Mago, R.,
777 Upadhyaya, N. M., Rathjen, J. P, Stone, W. A., Schwessinger, B., Figueroa, M., Dodds, P.
778 N., Periyannan, S., & Sperschneider J. (2022). Identification and correction of phase
779 switches with Hi-C data in the Nanopore and HiFi chromosome-scale assemblies of the
780 dikaryotic leaf rust fungus *Puccinia triticina*. *Current Biology*, 23, 84.
781 25. Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C., Joly, D. L.,
782 Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M.,
783 Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., Kohler, A., Kües, U.,
784 Lindquist, E. A., Lucas, S. M., Mago, R., Mauceli, E., Morin, E., Murat, C.,Pangilinan, J. L.,
785 Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A. A.,
786 Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G. A., Henrissat, B., Van de, Peer.
787 Y., Rouze, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev,
788 I. V., Szabo, L. J., & Martin, F. (2011). Obligate biotrophy features unraveled by the
789 genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*, 108,
790 9166–9171.
791 26. Edwards, R. J., Dong, C. M., Park, R. F., Tobias, P. A. (2022) A phased chromosome-level
792 genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen,
793 *Austropuccinia psidii*. https://doi.org/10.1101/2022.04.22.489119
794 27. Ellinghaus, D., Kurtz, S. & Willhoeft, U. (2008). *LTRharvest*, an efficient and flexible
795 software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics 9, 18.
796 28. Emms, D.M. & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for
797 comparative genomics. *Genome Biology,* 20, 1–14.
798 29. Figueroa, M., Dodds, P. N., & Henningsen, E. C. (2020). Evolution of virulence in rust
799 fungi–multiple solutions to one problem. *Current Opinion in Plant Biology*, 56, 20–27.
800 30. Figueroa, M., Upadhyaya, N. M, Sperschneider, J., Park, R. F., Szabo, L. J., Steffenson, B.,
801 Eills, J. G., Dodds, P. N. (2016). Changing the game: using integrative genomics to probe

802 virulence mechanisms of the stem rust pathogen *Puccinia graminis* f. sp. *tritici*. *Frontiers in*
803 *Plant Science*, 7, 205.
804 31. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A.,
805 Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., & Punta, M. (2014).
806 Pfam: the protein families database. *Nucleic Acids Research,* 42, D222–D230.
807 32. Frantzeskakis, L., Barsoum, M., Panstruga, R., Kusch, S., Kiss, L., Takamatsu, S., &
808 Panstruga, R. (2019). The *Parauncinula polyspora* draft genome provides insights into
809 patterns of gene erosion and genome expansion in powdery mildew fungi. *MBio*, 10, 1–17.
810 33. Frantzeskakis, L., Pietro, A. D., Rep, M., Schirawski, J., Wu, C. H., & Panstruga, R. (2020).
811 Rapid evolution in plant–microbe interactions—a molecular genomics perspective. *New*
812 *Phytologist*, 225, 1134–1142.
813 34. Gel, B., & Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable
814 genomes displaying arbitrary data. *Bioinformatics* 33, 3088–3090.
815 35. Ghurye, J., Pop, M., Koren, S., Bickhart, D., & Chin, C. S. (2017). Scaffolding of long read
816 assemblies using long range contact information. *BMC genomics*, 18, 527.
817 36. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis,
818 X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A.,
819 Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., &
820 Regev, A. (2011). Trinity: reconstructing a full-length transcriptome assembly from RNA-
821 Seq data. *Nature biotechnology*, 29, 644–652.
822 37. Gu, Z. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*
823 30, 2811–2812.
824 38. Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C.
825 R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using
826 EVidenceModeler and the Program to Assemble Spliced Alignments. Genome biology, 9,
827 R7.
828 39. Henningsen, E. C., Hewitt, T., Dugyala, S., Nazareno, E. S., Gilbert, E., Li, F., Kianian, S. F,
829 Steffenson, B. J., Dodds, P. N., Sperschneider, J., & Figueroa, M. (2022). A chromosome-
830 level, fully phased genome assembly of the oat crown rust fungus *Puccinia coronata* f. sp.
831 *avenae*: a resource to enable comparative genomics in the cereal rusts. *G3-Genes Genomes*
832 *Genetics*, 12(8), jkac149.

833 40. Huang, S. F., Kang, M. J., & Xu, A. L. (2017). HaploMerger2: rebuilding both haploid sub-
834 assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*, 33,2577–
835 2579.
836 41. Hubbard, A., Lewis, C. M., Yoshida, K., Ramirez-Gonzalez, R. H., de Vallavieille-Pope, C.,
837 Thomas, J., Kamoun, S., Bayles, R., Uauy, C., & Saunders, D. G. (2015). Field
838 pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome*
839 *Biology*, 16, 23. https://doi.org/10.1186/s13059-015-0590-8
840 42. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H.,
841 Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019).
842 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource
843 based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309–D314.
844 43. Jin, Y., Szabo, L. J., & Carson, M. (2010). Century-old mystery of *Puccinia striiformis* life
845 history solved with the identification of *Berberis* as an alternate host. *Phytopathology*. 100,
846 432–435.
847 44. Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H.,
848 Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A.,
849 Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-
850 scale protein function classification. Bioinformatics (Oxford, England), 30(9), 1236–1240.
851 https://doi.org/10.1093/bioinformatics/btu031
852 45. Justesen, A. F., Ridout, C. J., & Hovmøller, M. S. (2002). The recent history of *Puccinia*
853 *striiformis* f. sp. *tritici* in Denmark as revealed by disease incidence and AFLP markers.
854 *Plant pathology*, *51*, 13–23.
855 46. Kamvar, Z N., Tabima, J. F., & Grünwald, N. J. (2014) Poppr: an R package for genetic
856 analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ,* 2,
857 e281.
858 47. Knaus, B. J., & Grünwald, N, J. (2017). vcfr: a package to manipulate and visualize variant
859 call format data in R. *Molecular Ecology Resources*. 17, 44–53.
860 48. Kolmer, J. A. (2005). Tracking wheat rust on a continental scale. *Current Opinion in Plant*
861 *Biology*, 8, 441–449.
862 49. Le Rouzic, A., & Capy, P. (2005). The first steps of transposable elements invasion:
863 Parasitic strategy vs. genetic drift. *Genetics* 169 (2), 1033–1043.

50. Li, F., Upadhyaya, N. M., Sperschneider, J., Matny, O., Nguyen-Phuc, H., Mago, R., Raley, C., Miller, M. E., Silverstein, K., Henningsen, E., Hirsch, C. D., Visser, B., Pretorius, Z. A., Steffenson, B. J., Schwessinger, B., Dodds, P. N., & Figueroa, M. (2019). Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nature communications*, 10, 5068. https://doi.org/10.1038/s41467-019-12927-7
51. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics,* 34, 3094–3100.
52. Li, H., & Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26, 589–595.
53. Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923–930.
54. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm, *Nucleic Acids Research,* 33, 6494–6506.
55. Lorrain, C., Feurtey, A., Mller, M., Haueisen, J., & Stukenbrock, E. (2021). Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defences. *G3-Genes Genomes Genetics,* 11, jkab068.
56. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.
57. Lowe, T. M., & Chan, P. P. (2016). TRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, 44, W54 –W57.
58. Lue, N. F. (2021). Duplex telomere-binding proteins in fungi with canonical telomere repeats: new lessons in the rapid evolution of telomere proteins. *Frontier in Genetics*, 12,638790.
59. Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science,* 302, 1401–1404.
60. Maia, T., Badel, J. L., Marin-Ramirez, G., Rocha, C. M., Fernandes, M. B., da Silva, J. C, de Azevedo-Junior, G. M., Brommonschenkel, S. H. (2017). The *Hemileia vastatrix* effector HvEC016 suppresses bacterial blight symptoms in coffee genotypes with the SH1 rust resistance gene. *New Phytologist*, 213,1315–1329.

895 61. Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S.L., & Zimin, A.
896 (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational*
897 *Biology*, 14, e1005944.
898 62. Marçais, M., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel
899 counting of occurrences of k-mers. *Bioinformatics,* 27, 764–770.
900 63. Miller, M. E., Nazareno, E. S., Rottschaefer, S. M., Riddle, J., Dos Santos Pereira, D., Li, F.,
901 Nguyen-Phuc, H., Henningsen, E. C., Persoons, A., Saunders, D., Stukenbrock, E., Dodds, P.
902 N., Kianian, S. F., & Figueroa, M. (2020). Increased virulence of *Puccinia coronata* f. sp.
903 *avenae* populations through allele frequency changes at multiple putative *Avr* loci. *PLoS*
904 *genetics*, 16, e1009291. https://doi.org/10.1371/journal.pgen.1009291
905 64. Miller, M. E., Zhang, Y., Omidvar, V., Sperschneider, J., Schwessinger, B., Raley, C.,
906 Palmer, J. M., Garnica, D., Upadhyaya, N., Rathjen, J., Taylor, J. M., Park, R. F., Dodds, P.
907 N., Hirsch, C. D., Kianian, S. F., & Figueroa, M. (2018). *De Novo* Assembly and Phasing of
908 Dikaryotic Genomes from Two Isolates of *Puccinia coronata* f. sp. *avenae*, the Causal
909 Agent of Oat Crown Rust. *mBio*, 9, e01650-17. https://doi.org/10.1128/mBio.01650-17
910 65. Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sánchez-García, M., Morin, E.,
911 Andreopoulos, B., Barry, K. W., Bonito, G., Buée, M., Carver, A., Chen, C., Cichocki, N.,
912 Clum, A., Culley, D., Crous, P. W., Fauchery, L., Girlanda, M., Hayes, R. D., ...,Martin, F.
913 M. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the
914 early evolution of symbiotic traits. *Nature Communication,* 11, 5125.
915 https://doi.org/10.1038/s41467-020-18795-w
916 66. Neumann, P., Novák, P., Hoštáková, N., & Macas, J. (2019) Systematic survey of plant
917 LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and
918 provides a reference for element classification. *Mobile DNA* 10, 1.
919 https://doi.org/10.1186/s13100-018-0144-1
920 67. Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., ... & Koren, S.
921 (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants
922 from high-fidelity long reads. *Genome research*, *30*, 1291–1305.
923 68. Oliver, K. R.,& Greene, W. K. (2012). Transposable elements and viruses as factors in
924 adaptation and evolution: an expansion and strenthening of the TE thrust hypothesis.
925 *Ecology and Evolution*, 2, 2912–2933.

926 69. Ou, S., & Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for
927 Identification of Long Terminal Repeat Retrotransposons. *Plant Physiology*, 176, 1410–
928 1422.
929 70. Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0:
930 discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 785–786.
931 71. Pfeifer, B., Wittelsbuerger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome:
932 An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology*
933 *and Evolution*, 31, 1929–1936.
934 72. Pinon, J., & Frey, P. (1997). Structure of *Melampsora larici-populina* populations on wild
935 and cultivated poplar. *European Journal of Plant Pathology*. 103,159–173.
936 73. Piperno, D. R., Moreno, J. E., Iriarte, J., Holst, I., Lachniet, M., Jones, J. G., Tanere, A. J., &
937 Castanzo, R. (2007). Late Pleistocene and Holocene environmental history of the Iguala
938 Valley, Central Balsas Watershed of Mexico. *Proceedings of National of American of*
939 *Sciences*. 104, 11874–11881.
940 74. Presti, L. L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., & Tollot, M. (2015). Fungal
941 effectors and plant susceptibility. *Annual Review of Plant biology,* 66, 513–545.
942 75. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., maller, J.m
943 Sklar, P., Bakker, P. I. W. D., & Daly, M. J (2007). Plink: a tool set for whole-genome
944 association and population-based linkage analyses. The *American Journal of Human*
945 *Genetics*, 81, 559–575.
946 76. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing
947 genomic features. *Bioinformatics*. 26, 841–842.
948 77. Ramirez-Cabral, N. Y. Z., Kumar, L., & Shabani, F. (2017). Global risk levels for corn rusts
949 (*Puccinia sorghi* and *Puccinia polysora*) under climate change projections. Journal of
950 Phytopathology, 165, 563–574.
951 78. Ramos, A. P., Tavares, S., Tavares, D., Silva, M. D. C, Loureiro, J., & Talhinhas, P. (2015).
952 Flow cytometry reveals that the rust fungus, *Uromyces bidentis* (Pucciniales), possesses the
953 largest fungal genome reported-2489Mbp. *Molecular Plant Pathology*, 16, 1006–1010.
954 79. Rawlings, N. D., Barrett, A. J., & Finn, R. (2016). Twenty years of the Merops database of
955 proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 44, D343–
956 D350.

957 80. Roach, M. J., Schmidt, S. A., & Borneman, A R. (2018). Purge Haplotigs: allelic contig
958 ressignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19, 460.
959 81. Rochi, L., Diéguez, M. J., Burguener, G., Darino, M. A., Pergolesi, M. F., Ingala, L. R.,
960 Cuyeu, A. R., Turjanski, A., Kreff, E. D., & Sacco, F. (2018). Characterization and
961 comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize
962 common rust. *Fungal Genetics and Biology*, 112, 31–39.
963 82. Salcedo, A., Rutter, W., Wang, S., Akhunova, A., Bolus, S., Chao, S., Anderson, N., De
964 Soto, M. F., Rouse, M.,Szabo, L., Bowden, R. L., Dubcovsky. J., & Akhunov, E. (2017).
965 Variation in the *AvrSr35* gene determines *Sr35* resistance against wheat stem rust race Ug99.
966 *Science*, 358, 1604–1606.
967 83. Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J.
968 M., Dodds, P. N., Figueroa, M., Park, R. F., & Rathjen, J. P. (2018). A near-complete
969 haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f.
970 sp. *tritici* reveals high interhaplotype diversity. *mBio*, 9, e0227517.
971 84. Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., Heard, E.,
972 Dekker, J., & Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data
973 processing. *Genome Biology*, 16, 259.
974 85. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. (2015).
975 BUSCO: assessing genome assembly and annotation completeness with single-copy
976 orthologs. *Bioinformatics,* 31, 3210–3212.
977 86. Singh, R. P., Hodson, D. P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., Herrera-
978 Foessel, S., Singh, P. K., Singh, S., & Govindan, V. (2011). The emergence of Ug99 races
979 of the stem rust fungus is a threat to world wheat production. *Annual Review of*
980 *Phytopathology*, 49, 465–481.
981 87. Sipos, G., Prasanna, A. N., Walter, M. C., O'Connor, E., Bálint, B., Krizsán, K., Kiss, B.,
982 Hess, J., Varga. T., Slot, J., Riley, R., Bóka, B., Rigling, F., Barry, K., Lee, J., Mihaltcheva,
983 S., LaButti, K., Lipzen, A., Moloney, N. M., Sperisen, C., Kredics, L., Vágvölgyi, C.,
984 Patrignani, A., Fitzpatrick, D., Nagy, I., Soyle. S., Anderson, J. B., Grigoriev, I. V.,
985 Güldener, U., Münsterkötter, M., & Nagy, L. G. (2017). Genome expansion and lineage-
986 specific genetic innovations in the forest pathogenic fungi *Armillaria*. Nature Ecology &
987 Evolution, 1, 1931–1941.

988 88. Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological
989 sequence comparison. *BMC Bioinformatics*, 6, 1–11.
990 89. Smit, A., & Hubley, R. (2008). RepeatModeler open 1.0. Institute for Systems Biology,
991 Seattle, WA. http://www.repeatmasker.org/
992 90. Smit, A., Hubley, R., & Green, P. (2015). RepeatMasker open 4.0. 2013–2015. Institute for
993 Systems Biology, Seattle, WA. http://www.repeatmasker.org/
994 91. Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., ... &
995 Panstruga, R. (2010). Genome expansion and gene loss in powdery mildew fungi reveal
996 tradeoffs in extreme parasitism. *Science*, 330,1543–1546.
997 92. Sperschneider, J., & Dodds, P. N. (2022). EffectorP 3.0: Prediction of Apoplastic and
998 Cytoplasmic Wffectors in Fungi and Oomycetes. *Molecular Plant-Microbe Interaction*, 35,
999 146–156.
1000 93. Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in
1001 eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33, W465–W467.
1002 94. Tavares, D., Romos, A. P., Pires, A. S., Azomjeora, H. G., Caldeirinha, P., Link, T.,
1003 Abranches, R., Silva, M. do. C., Voegele, R. T., Loureiro, J., & Talhinhas, P. (2014).
1004 Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in Plant*
1005 *Sceiences*, 5, 422.
1006 95. Terrapon, N., Lombard, V., Drula, E., Coutinho, P.M., Henrissat, B. (2017). The CAZy
1007 Database/the Carbohydrate-Active Enzyme (CAZy) Database: Principles and Usage
1008 Guidelines. In: Aoki-Kinoshita, K. (eds) A Practical Guide to Using Glycomics Databases.
1009 Springer, Tokyo.
1010 96. Tobias, P. A., Schwessinger, B., Deng, C.H., Wu, C., Dong, C. M., Sperschneider, J., Jones,
1011 A. Lou, Z. Y., Zhang, P., Sandhu, K., Smith, G. R., Tibbits, J., Chagne, D., & Park, R. F.
1012 (2021). *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by
1013 transposable elements. *G3*, 11, jkaa015.
1014 97. Todd, R. T., Forche, A., & Selmecki, A. (2017). Ploidy Variation in Fungi: Polyploidy,
1015 Aneuploidy, and Genome Evolution. *Microbiology Spectrum*, 5,10.
1016 98. Upadhyaya, N. M., Mago, R., Panwar, V., Hewittm, T., Luo, M., Chen, J., Sperschneider, J.,
1017 Nguyen-Phuc, H., Wang, A.H., Ortiz, D., Hac, L., Bhatt, D., Li, F., Zhang, J. P., Ayliffe, M.,
1018 Figueroa, M., Kanyuka, K., Wllis, J. G., & Dodds, P. N. (2021). Genomics accelerated

1019 isolation of a new stem rust avirulence gene–wheat resistance gene pair. *Nature*
1020 *Plants*, 7, 1220–1228.
1021 99. van Heerwaarden, J., Doebley, J., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., de Jesus
1022 Sanchez Gonzalez, J., & Ross-Ibarra, J. (2011). *Proceedings of National of American of*
1023 *Sciences* 108, 1088–1092.
1024 100. Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J.,
1025 & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short
1026 reads. *Bioinformatics,* 14,2202–2204.
1027 101. Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B.,
1028 Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: a toolkit for detection and
1029 evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40, e49.
1030 https://doi.org/10.1093/nar/gkr1293
1031 102. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-
1032 Verlag.
1033 103. Winzer, F. L., Berthon, K. A., Carnegie, A. J., Pegg, G. S., & Leishman, M. R.
1034 (2019). *Austropuccinia psidii* on the move: survey based insights to its geographical
1035 distribution, host species, impacts and management in Australia. *Biological Invasions.* 21,
1036 1215–1225.
1037 104. Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T.,
1038 Backofen, R., Ramírez, F., & Grüning, B. A. (2018). Galaxy HiCExplorer: a web server for
1039 reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids
1040 Research, 46, W11–W16.
1041 105. Wu, J. Q., Song, L., Ding, Y., Dong, C. M., Hasan, M., & Park, R. (2021). A chromosome-
1042 scale assembly of the wheat leaf rust pathogen *Puccinia triticina* provides insights into
1043 structural variations and genetic relationship with haplotype resolution. *Frontiers in*
1044 *Microbiology,* 12, 704253.
1045 106. Zeh, D. W., Zeh, J. A., & Ishida, Y. (2009). Transposable elements and an epigenetic basis
1046 for punctuated equilibria. *BioEssays,* 31, 715–726.
1047 107. Zhang, R. G., Li, G. Y., Wang, X. L., Dainat J., Wang, Z. X., Ou, S. J., & Ma, Y. P. (2022).
1048 TEsorter: An accurate and fast method to classify LTR-retrotransposons in plant
1049 genomes, *Horticulture Research*, 9, uhac017, https://doi.org/10.1093/hr/uhac017

108. Zhao, P., Qi, X. H., Crous, P. W., Duan, W. J., & Cai, L. (2020). Gymnosporangium species on Malus: species delineation, diversity and host alternation. *Persoonia*, 45, 68–100.
109. Zhao, P., Zhang, Z. F., Hu, D. M., Tsui, K. M., Qi, X. H., Phurbu, D., Gafforov, Y., & Cai, L. (2021). Contribution to rust flora in China I, tremendous diversity from natural reserves and parks. *Fungal Diversity*, 110, 1–58.

1055 **Figure Legends**

1056 **FIGURE 1** Chromosome-level assembly of GD1913 isolate of *Puccinia polysora*. (A) The
1057 dikaryotic phasing pipeline and switch strategies for duplicated contigs. Duplicated contigs were
1058 checked based on the all-vs-all sequence alignment from HaploMerger 2 (Huang et al. 2017). (B)
1059 Hi-C-based contig anchoring. The heat map showed the density of Hi-C interactions within
1060 haplotype A. The 18 chromosomes are highlighted by blue squares with arrows pointed at
1061 centromeres. (C) Schematic representation of assembled chromosomes for GD1913 of each
1062 haplotype. (D) Percentage of Hi-C links of each chromosome of haplotype A to either haplotype
1063 A (blue) or haplotype B (red).

1064 **FIGURE 2** Genomic features of *P. polysora* f.sp. *zeae* isolate, GD1913 and repeat comparison
1065 analyses with its close relatives. (A) Genomic landscape of 18 chromosomes for haplotype A
1066 (hapA) and haplotype B (hapB). From outer to inner circles: (i) Chromosomes, (ii) Repeat
1067 density, (iii) Gene density, (iv) Candidate effector density, (v) Guanine-cytosine (GC) content.
1068 The tracks from ii to iv are estimated by non-overlapping 100 kbp sliding windows. (B)
1069 Comparison of genome size (two nuclei) and repeat contents among *Puccinia polysora* f.sp. *zeae*
1070 (*Ppz*_GD1913) and other rust species, *Melampsora larici-populina* (*Mlp*_98AG31),
1071 *Austropuccinia psidii (Ap*_Au_3*), Puccinia sorghi* f.sp. *zeae (Psz*_RO10H11247*), Pucccnia*
1072 *coronata* f.sp. *avenae (Pca*_203*)*, *Puccinia graminis* f.sp. *tritici (Pgt*_21-0*)*, *Puccinia triticina*
1073 f.sp. *tritici (Ptt*_76*)* and *Puccinia striiformis* f.sp. *tritici* (*Pst*_104E*)*. (C) The insertion time (Mya)
1074 distribution of intact LTR-RTs in *P. polysora*.

1075 **FIGURE 3** The interhaplotype variation of *Ppz*-GD1913 isolate. (A) Summary of interhaplotype
1076 variation between haplotype A and B using Assemblytics. Six types of specific variation were
1077 illustrated by schematic plot and each bar chart indicates the number of bases categorized to each
1078 type. (B) Synteny plot of haplotype A and B using Mummer. (C) and (D) Full-chromosome
1079 alignments of chromosomes 9 and 15 showing insertion/deletion and inversion (arrows) between
1080 haplotypes.

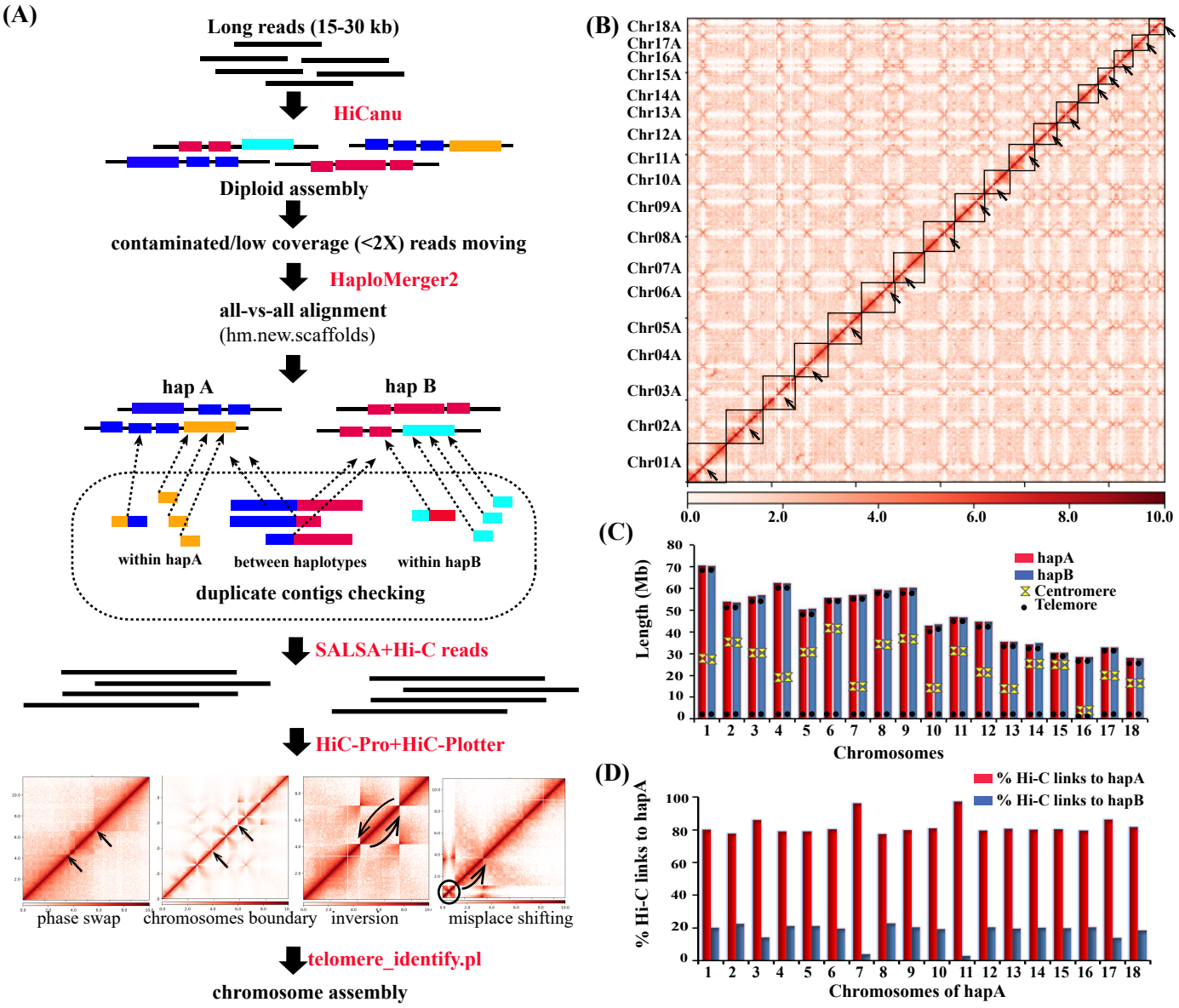
1081 **FIGURE 4** Analyses of gene expression and gene distance of predicted secretome of *P. polysora*.
1082 (A) Clustering analysis of gene expression of secretome on two haplotypes. Heatmaps show rlo-
1083 transformed expression values. Cluster numbers are shown outside the graphs, and tracks
1084 represent gene expression in germinated spores (GS), and infected tissues at 1 dpi, 2 dpi, 4 dpi, 7
1085 dpi, 10 dpi and 14 dpi. (B) Hexplots for closest-neighbor gene distance density of *P. polysora* in

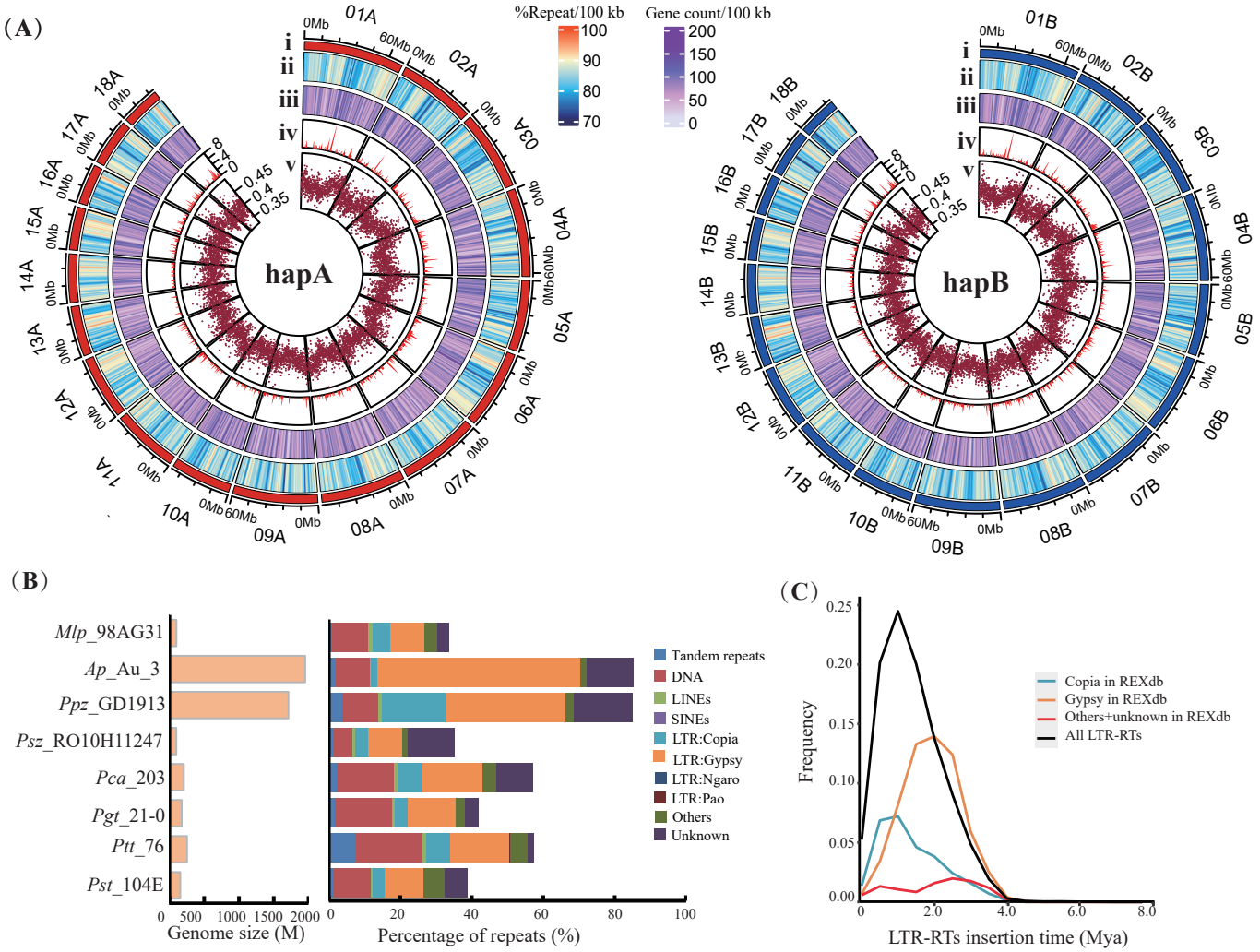
1086 haplotype A (left) and haplotype B (right). Circle dots represent two gene categories, BUSCOs
1087 (yellow) and predicted candidate effectors (red). The black dots represent the distribution of the
1088 avirulent gene, *AvrRppC.* (C) Gene and repeat density as well as log2TPM (transcripts per
1089 million) of candidate effectors from chromosomes 14A and 14B. The red and blue points
1090 represent expression values of candidate effectors at 1 dpi and 7dpi. Positions and protein
1091 sequences of *AvrRppC* genes are presented.
1092 **FIGURE 5** Functional annotation of CAZymes, transcription factors and Merops proteases of
1093 *Puccinia polysora* GD1913 (*Ppz*-GD1913) and its comparison with other *Puccinia* species. (A)
1094 CAZyme families comparison of *P. polysora* (*Ppz,* GD1913), *P. triticina* (*Ptt*, *Pt76*), *P. graminis*
1095 (*Pgt*, *Pgt*21-0) and *P. coronata* (*Pca*, *Pca*203). Heat maps showing gene numbers annotated in
1096 the following classes: auxiliary activities (AA), carbohydrate-binding modules (CBM),
1097 carbohydrate esterases (CE), glycoside hydrolases (GH), glycosyltransferases (GTs), and
1098 polysaccharide lyases (PL). Expanded families of AA1, CE4, GH18, GH5, GT2 and GT90 were
1099 highlighted in red. (B) Percentages of genes from *P. polysora* are predicted to encode members
1100 of various fungal transcription factor classes based on InterProScan and Eggnog annotation. (C)
1101 Merops families comparison of *Ppz*, *Ptt*, *Pgt* and *Pca*. Heat map showing gene numbers
1102 annotated in the classes of aspartic acid (A), cysteine(C), peptidase inhibitors (I), metalloprotease
1103 (M), serine protease (S), and threonine protease (T) which were labeled in the right of the heat
1104 map and expanded families were highlighted in red.
1105 **FIGURE 6** Comparative analyses of macro-synteny and orthogroups among *Puccinia. polysora*
1106 f. sp. *zeae* (*Ppz*GD1913A and *Ppz*GD1913B) and its close relatives, *Austropuccinia psidii*
1107 (Au_3), *P. triticina* f. sp. *tritic*i (*Pt*76), *P. graminis* f.sp. *tritici* (*Pgt*21-0) and *P. coronata* f.sp
1108 *avenae* (*Pca*203). (A) Macro-synteny of four species in *Puccinia* with available haplotype-
1109 phased and chromosome-scale genome. Only the chromosomes from haplotype A/primary
1110 haplotype were compared. Grey lines are homologous proteins identified by MCscan Python
1111 version with default parameters and black lines present the genome inversions between species.
1112 For better visualization, the genome size ratio of *Ppz*B:*Ppz*A:*Pt*76:*Pca*203:*Pgt*21-0 =
1113 0.2:0.2:1:1:1. (B) Phylogenetic relationship of *P. polysora* and other rust species was inferred by
1114 1454 single-copy orthologue genes. The numbers indicated the sum of gained (+) and lost (–)
1115 genes from all orthogroups by comparing the two most close species. The typical symptoms

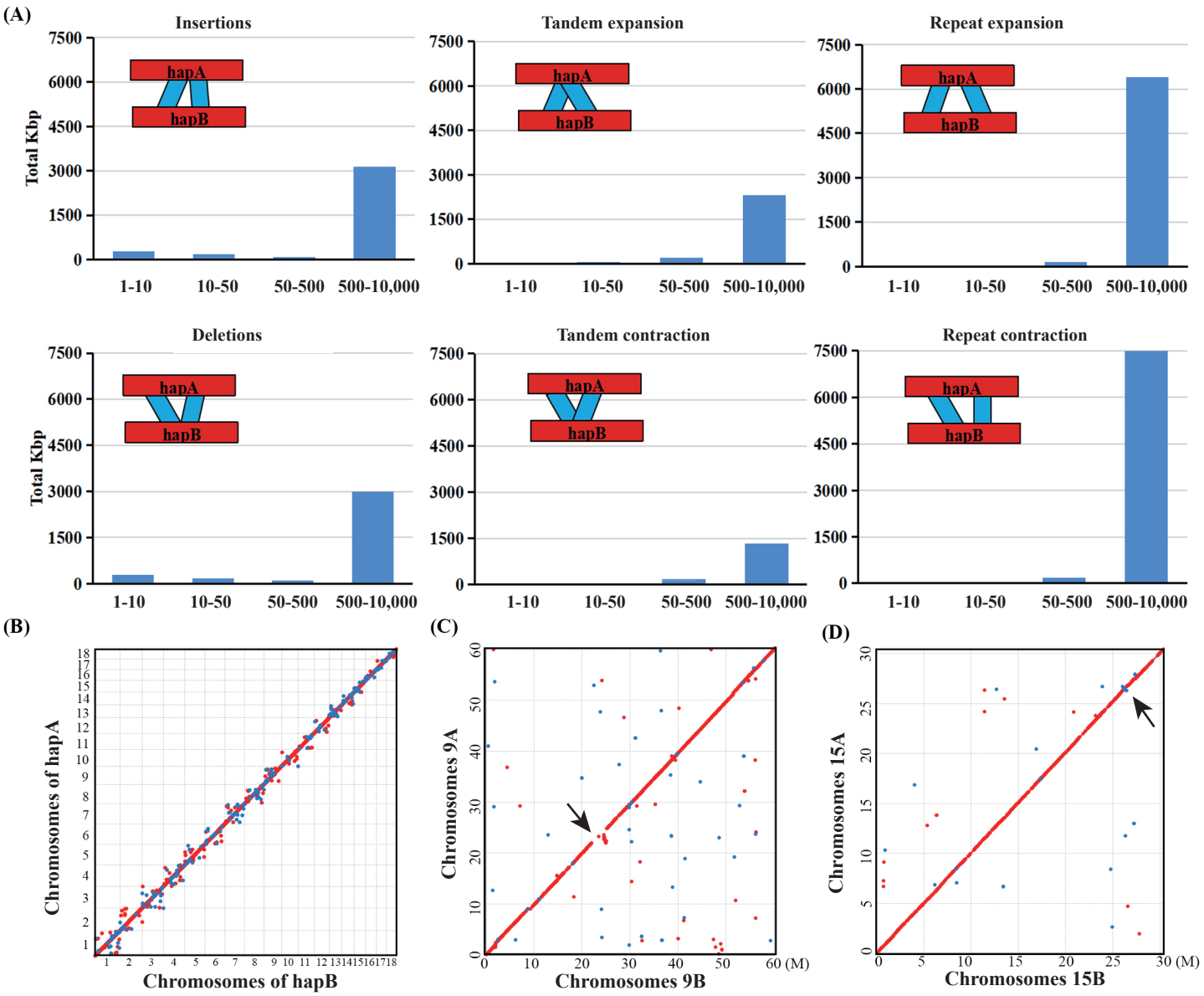
1116 caused by each species are listed accordingly. (C) Distribution of species-specific orthogroups
1117 with varied ranges of gene copy number among four *Puccinia* species.
1118 **FIGURE 7** Population genomics analyses of *P. polysora* from China. (A) Geographic
1119 distribution of 79 *P. polysora* isolates collected from north, central and south of China. (B)
1120 Principal component analysis of Chinese *P. polysora* population showing two genetic groups. (C)
1121 Linkage disequilibrium test of Chinese *P. polysora* population. (D) *AvrRppC* genotypes of *P.*
1122 *polysora* isolates in minor (red) and major (blue) genetic groups.

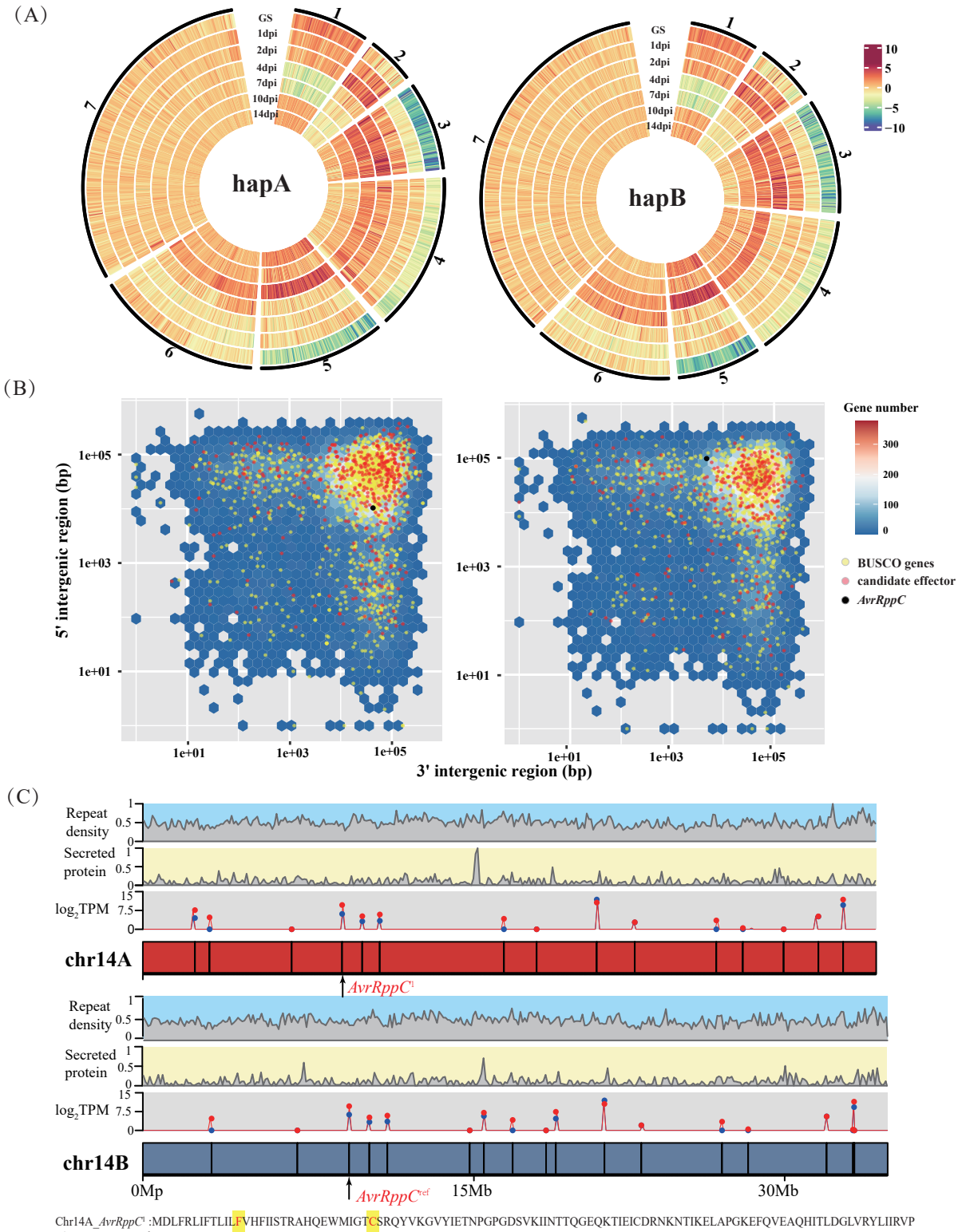
1123 **Supplementary files**

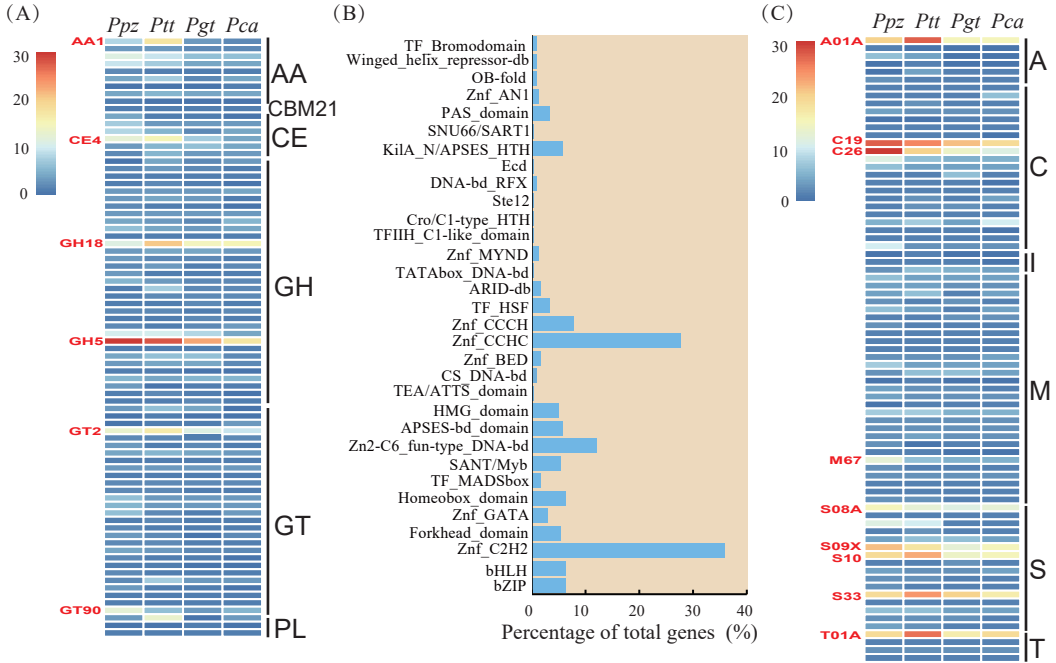
1124 Table S1 Information of 79 isolates from China used for genome resequencing.
1125 Table S2 Contig information removed from the final assembly due to low-quality assessment.
1126 Table S3 Treatments for duplicate contigs between/within haplotype A and haplotype B
1127 Table S4 Sample information and data used for transcript analyses.
1128 Table S5 Functional annotation of specific orthologs between two haplotypes.
1129 Table S6 GO and IPR annotation for predicted candidate effectors.
1130 Table S7 Features of secreted proteins in different expression clusters of *P. polysora* GD1913.
1131 Table S8 Proteome comparison of four *Puccinia* species.
1132 Table S9 GO annotation of top 42 *P. polysora* specific orthogroups with high gene copy number.
1133 Table S10 Table S10 Variants impact of SNPs located in 367 genes with *F*st > 0.9 between two
1134 groups.
1135 Table S11 The CDS and amino acid sequences of *AvrRppC* gene.
1136 Table S12 Comparison of Gene number and gene length among three *Puccinia* species.
1137 Table S13 NMDC numbers for sequencing raw data used in this study.
1138 Figure S1 Genome size estimation of *Ppz*-GD1913 using 90 Gb illumina short reads.
1139 Figure S2 GO enrichment analysis of secreted proteins on two haplotypes.
1140 Figure S3 Pair-wise synteny of *P. polysora* and three close relatives.
1141 Figure S4 The elbow bend plot.
1142 Figure S5 The bi-allele distribution of 79 isolates used for population genomics. Two isolates
1143 (GD1922-3 and GX1905-2) in red squares are removed from the final analysis due to abnormal
1144 distribution.

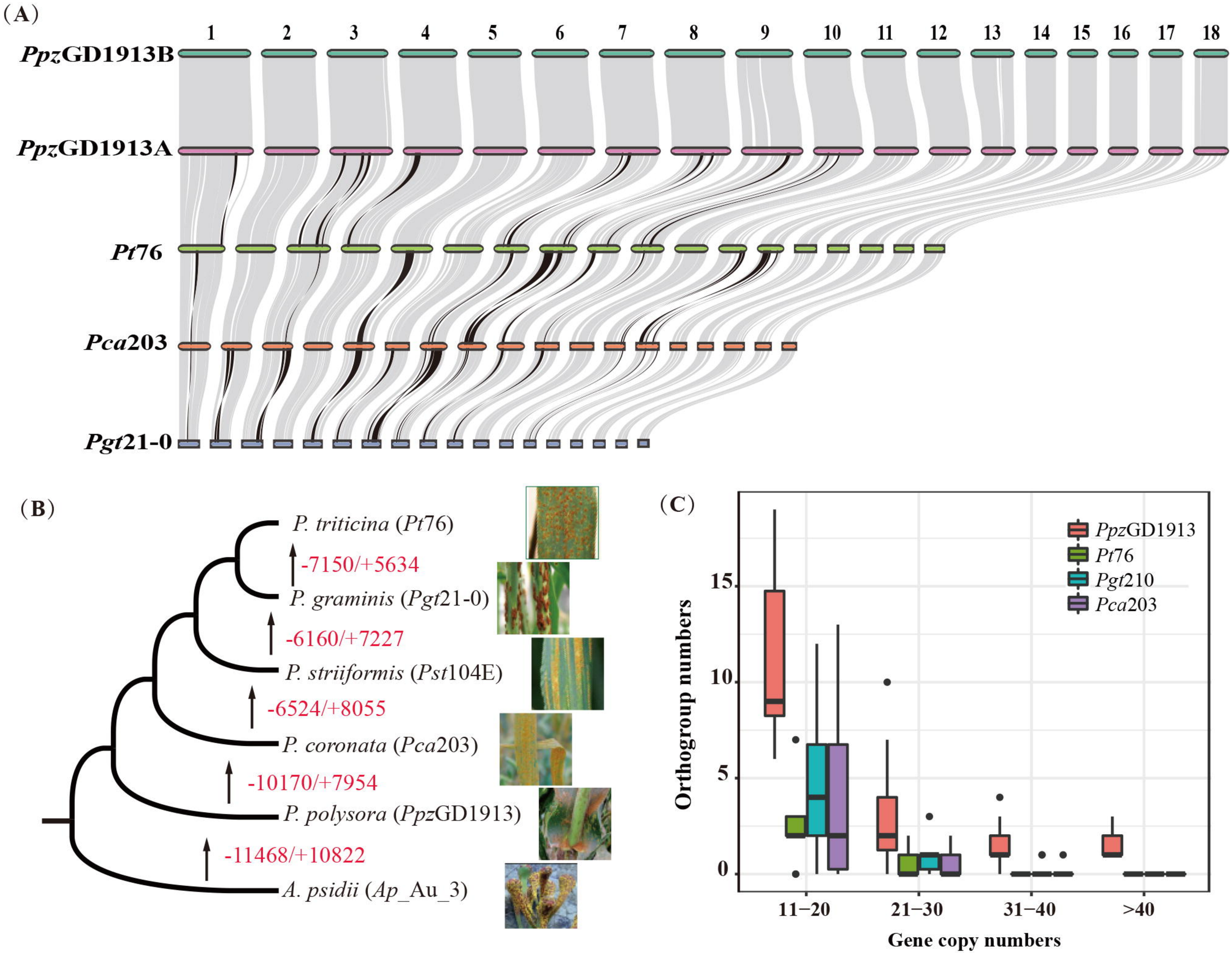

(A)
Long reads (15-30 kb)
HiCanu
Diploid assembly
contaminated/low coverage (<2X) reads moving
HaploMerger2
all-vs-all alignment
(hm.new.scaffolds)
hap A
hap B
within hapA
between haplotypes
within hapB
duplicate contigs checking
SALSA+Hi-C reads
HiC-Pro+HiC-Plotter
phase swap
chromosomes boundary
inversion
misplace shifting
telomere_identify.pl
chromosome assembly
(B)
Chr18A
Chr17A
Chr16A
Chr15A
Chr14A
Chr13A
Chr12A
Chr11A
Chr10A
Chr09A
Chr08A
Chr07A
Chr06A
Chr05A
Chr04A
Chr03A
Chr02A
Chr01A
0.0
2.0
4.0
6.0
8.0
10.0
(C)
Length (Mb)
hapA
hapB
Centromere
Telemore
Chromosomes
(D)
% Hi-C links to hapA
% Hi-C links to hapB
Chromosomes of hapA


(A)
%Repeat/100 kb
Gene count/100 kb
hapA
hapB
(B)
Mlp_98AG31
Ap_Au_3
Ppz_GD1913
Psz_RO10H11247
Pca_203
Pgt_21-0
Ptt_76
Pst_104E
Genome size (M)
Percentage of repeats (%)
Tandem repeats
DNA
LINEs
SINEs
LTR:Copia
LTR:Gypsy
LTR:Ngaro
LTR:Pao
Others
Unknown
(C)
Frequency
LTR-RTs insertion time (Mya)
Copia in REXdb
Gypsy in REXdb
Others+unknown in REXdb
All LTR-RTs


(A)
Insertions
Tandem expansion
Repeat expansion
Deletions
Tandem contraction
Repeat contraction
Total Kbp
hapA
hapB
1-10
10-50
50-500
500-10,000
(B)
Chromosomes of hapA
Chromosomes of hapB
(C)
Chromosomes 9A
Chromosomes 9B
(M)
(D)
Chromosomes 15A
Chromosomes 15B
(M)


(A)
GS
1dpi
2dpi
4dpi
7dpi
10dpi
14dpi
hapA
hapB
1
2
3
4
5
6
7
10
5
0
−5
−10
(B)
5' intergenic region (bp)
3' intergenic region (bp)
1e+01
1e+03
1e+05
Gene number
300
200
100
0
BUSCO genes
candidate effector
AvrRppC
(C)
Repeat density
Secreted protein
log2TPM
chr14A
AvrRppC1
chr14B
AvrRppCref
0Mp
15Mb
30Mb
Chr14A_AvrRppC1 :MDLFRLIFTLILFVHFIISTRAHQEWMIGTCSRQYVKGVYIETNPGPGDSVKIINTTQGEQKTIEICDRNKNTIKELAPGKEFQVEAQHITLDGLVRYLIIRVP
Chr14B_AvrRppCref:MDLFRLIFTLILCVHFIISTRAHQEWMIGTRSRQYVKGVYIETNPGPGDSVKIINTTQGEQKTIEICDRNKNTIKELAPGKEFQVEAQHITLDGLVRYLIIRVP


(A)
Ppz
Ptt
Pgt
Pca
30
20
10
0
AA1
CE4
GH18
GH5
GT2
GT90
AA
CBM21
CE
GH
GT
PL
(B)
TF_Bromodomain
Winged_helix_repressor-db
OB-fold
Znf_AN1
PAS_domain
SNU66/SART1
KilA_N/APSES_HTH
Ecd
DNA-bd_RFX
Ste12
Cro/C1-type_HTH
TFIIH_C1-like_domain
Znf_MYND
TATAbox_DNA-bd
ARID-db
TF_HSF
Znf_CCCH
Znf_CCHC
Znf_BED
CS_DNA-bd
TEA/ATTS_domain
HMG_domain
APSES-bd_domain
Zn2-C6_fun-type_DNA-bd
SANT/Myb
TF_MADSbox
Homeobox_domain
Znf_GATA
Forkhead_domain
Znf_C2H2
bHLH
bZIP
0
10
20
30
40
Percentage of total genes (%)
(C)
Ppz
Ptt
Pgt
Pca
30
20
10
0
A01A
C19
C26
M67
S08A
S09X
S10
S33
T01A
A
C
I
M
S
T


(A)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
PpzGD1913B
PpzGD1913A
Pt76
Pca203
Pgt21-0
(B)
P. triticina (Pt76)
-7150/+5634
P. graminis (Pgt21-0)
-6160/+7227
P. striiformis (Pst104E)
-6524/+8055
P. coronata (Pca203)
-10170/+7954
P. polysora (PpzGD1913)
-11468/+10822
A. psidii (Ap_Au_3)
(C)
PpzGD1913
Pt76
Pgt210
Pca203
Orthogroup numbers
0
5
10
15
11−20
21−30
31−40
>40
Gene copy numbers

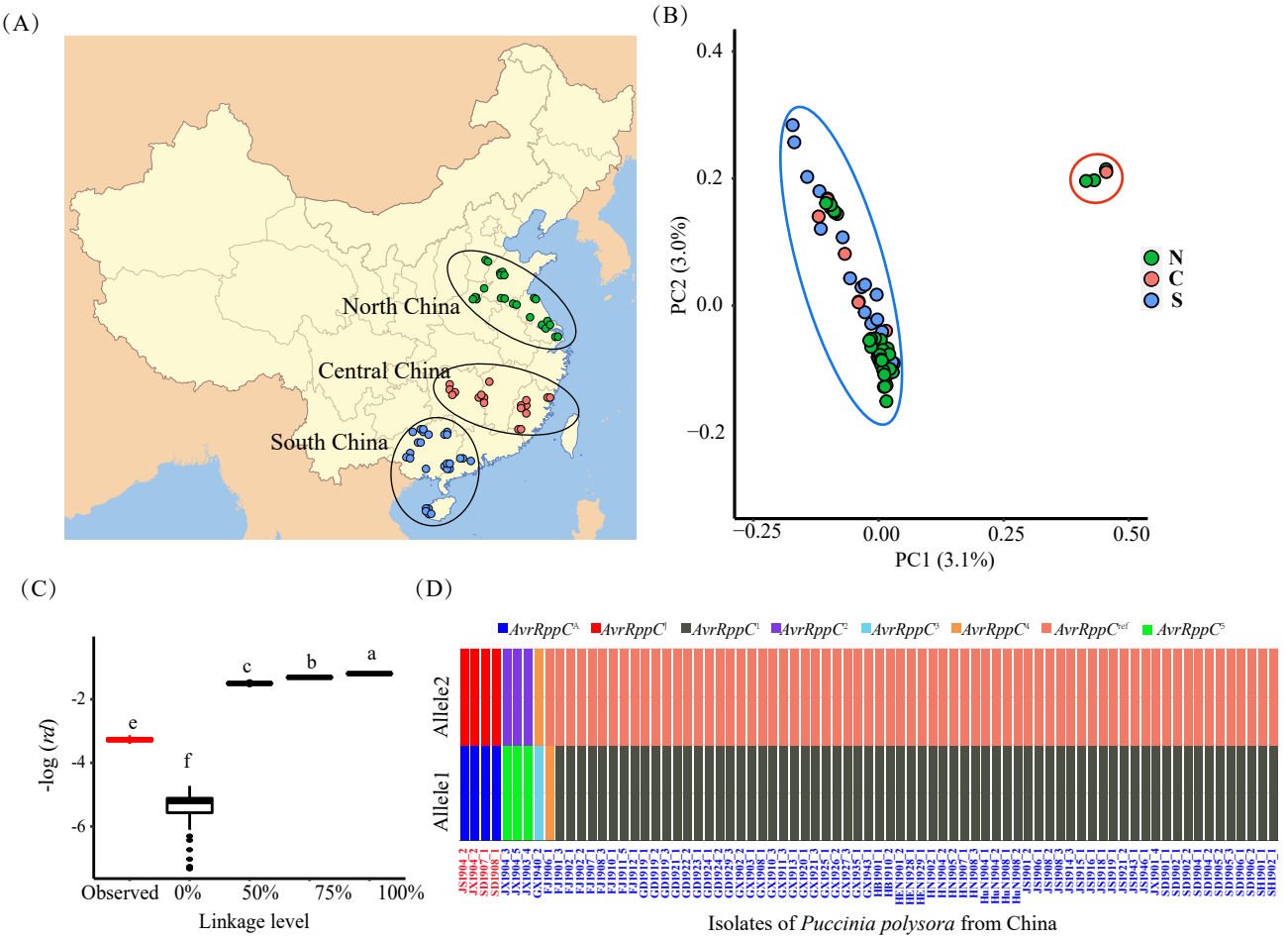

(A)
North China
Central China
South China
(B)
PC2 (3.0%)
PC1 (3.1%)
N
C
S
(C)
-log (rd)
Observed
0%
50%
75%
100%
Linkage level
(D)
Allele2
Allele1
Isolates of Puccinia polysora from China

Table 1 Assembly information and completeness evaluation of *Puccinia polysora* f.sp. *tritici,* GD1913.

Parameters	before haplotype-phased	
No. of contigs	4056	
No. of bases (Gbp)	1.76	
Size of longest contig (Mbp)	14.7	
N50 (Mbp) of contigs	1.9	
	after Hi-C scaffolding	
	Haplotype A	Haplotype B
No. of scaffolds	173	344
Size of scaffolds (Mbp)	862	855
Size of longest scaffold (Mbp)	31	39
N50 (Mbp) of scaffolds	13.2	9.5
	after Hi-C contact map binning	
Total size (Mbp)	856	855
No. of chromosomes	18	18
No. of unplaced scaffolds	15	20
Size of chromosomes (Mbp)	849	850
Size of longest chromosome (Mbp)	70.5	70.3
GC content	40.1%	40.0%
Complete BUSCOs	89.7%	89.7%
Complete and Single-copy BUSCOs	88.6%	88.6%
Complete and Duplicated BUSCOs	1.1%	1.1%
Fragmented BUSCOs	3.5%	3.5%
Missing BUSCOs	6.8%	6.8%

Table 2 Genes, orthogroups, and effectors for two haplotypes of *Puccinia polysora* f.sp. *tritici,* GD1913.

	haplotype A	haplotype B
No. of genes	23270	24242
No.(%) of genes located in 18 chromosomes	23038 (99.0%)	24080 (99.3%)
No.(%) of genes located in unplaced scaffolds	232 (1.0%)	162 (0.7%)
Mean gene length (bp)	1509	1480
% of genome covered by genes	4.1	4.2
No. of orthogroups on A and B	13057	
No. of genes in orthogroups	19815	20428
No. of secreted proteins	1183	1251
No. of apoplastic effectors	169	190
No. of cytoplasmic effectors	314	341
No. of effectors in orthogroups	463	501
No. of haplotype-specific effectors	20	30