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], Keyu Zhang[6], Leifu Li[6], Zhanhong Ma[6], 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] 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 describe 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 the largest fungal genome assembled to chromosome scale. Transcript-based annotation identified 47,512 genes 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 content, with genome-size expansion, gene losses and gene family expansions suggested by multiple copies of species-specific orthogroups. Interestingly, these features did not affect overall synteny with other *Puccinia* species with 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 drifted from the major population by having mutations on secreted proteins including *AvrRppC*, indicating the ongoing evolution and population differentiation. The high-quality assembly provides valuable genomic resources for future studies on 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; Aime and McTaggart 2021; Zhao et al. 2021). Rust fungi are obligate biotrophs, which are usually recalcitrant to in vitro culturing and show host-specificity 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 dikaryons, which contain two physically separated and genetically different haploid nuclei. The Pucciniales contains fungal species with the largest known genomes, with an average of haplotype genome size 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 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 in-depth understanding of evolution within the rust group.

In addition to 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*, *P. graminis* and *P. triticina*, causing three of the world's most serious wheat rust diseases (Kolmer et al. 2005); *Puccinia sorghi* and *P. polysora,* also seriously threaten global crop production by causing common and southern corn rust of maize (Crouch and Szabo 2011; Ramirez-Cabral et al. 2017); and *Melampsora larici-populina* and *Austropuccinia psidii*, two notorious forest pathogens leading to myrtle rust and

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

77 Early rust genome assemblies were haploid representations and different homologous
78 haplotypes were collapsed into a consensus assembly, which did not fully capture sequences
79 from both nuclei. Using single-molecule real-time (SMRT) sequencing technology and
80 haplotype resolution software, FALCON-Unzip (Chin et al. 2016), allowed higher contiguity and
81 partially haplotype-phased genomes of a few *Puccinia* species (Miller et al. 2018; Schwessinger
82 et al. 2018, Wu et al. 2021). Other softwares have been developed to obtain sub-assemblies from
83 diploid assemblies, such as HaploMerger2 (Huang et al. 2017) and Purge_Haplotigs (Roach et al.
84 2018), but many duplicate contigs cannot be correctly switched leading to mis-joins in the final
85 sub-assemblies. A recent pipeline, NuclearPhaser, was designed to phase two haplotypes from
86 the diploid genome using Hi-C data (Duan et al. 2022). Three *Puccinia* species, causing wheat or
87 oat rust have been assembled to a fully-phased and chromosome level using NuclearPhaser (Li et
88 al. 2019; Duan et al. 2022; Henningsen et al. 2022). Unlike wheat-like rusts, which have been
89 studied extensively, the genome information of corn rust pathogens is largely unknown. In
90 changed global cereal settings, maize has become the world's most productive food crop (FAO,
91 https://www.fao.org/faostat/en/#data/QCL), meanwhile, the distribution of the two major rust
92 diseases (common corn rust and southern corn rust) on corn is projected to expand to temperate
93 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 of 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 up to 1.7 G in our preliminary test, which poses new 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, this is the largest fungal genome as well as the first giga-scale fungal genome ever assembled at the chromosomal 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, we investigated the genetic divergence and population structure of *P. polysora* using the genome resequencing of 79 samples from different 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 selected for reference genome sequencing. A total of 79 isolates, representing three possible populations of *P. polysora* 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 rounds of infection on highly susceptible variety Zhengdan 958. Corn seeds immersed in Chlormequat chloride (1.5 g/L) were sown in 10-cm wide square pots 10 days before inoculation. When 10-cm tall, seedling growth was reduced by adding 15 mL maleic hydrazide acid (1.5 g/L) per pot. Urediniospores were amplified by spraying on 10-day-old corn seedlings. The inoculated seedlings were incubated with dew at 27°C in the dark for 24h before transferring to a climate-controlled chamber (12h, 25°C dark period and a 12h, 27°C light period). To prevent airborne contamination, 75% ethyl alcohol was sprayed on the area after each inoculation and each pot was protected by a cellophane bag. Once the pots were

125 heavily infected (ca. 15–20 days), spores were harvested to clean cellophane by scraping them
126 with sterile needles. For long-term storage, fresh spores were dried (10% relative humidity) in a
127 desiccator for 1 day at 4°C and then maintained at -80°C or in liquid nitrogen.

128 **2.2 | DNA extraction, library preparation, and sequencing**

129 About 500 mg of fresh urediniospores were ground for 4-5 batches in liquid nitrogen and
130 genomic DNA was extracted using the lysis buffer and cetyltrimethylammonium bromide
131 (CTAB) method (Justesen et al. 2002). The integrity of genomic DNA was assessed by Agilent
132 4200 Bioanalyzer (Agilent Technologies, Palo Alto, California, USA). The average insert size of
133 15 kb PacBio library was concentrated with AMPure PB magnetic beads (Pacific Biosciences,
134 California, USA) following the manufacture's instruction. Sequencing of high fidelity (HiFi)
135 long reads was carried out by the Pacific Bioscience Sequel II platform. For genome
136 resequencing, ~10 mg fresh urediniospores were placed in 2 mL screw-gap tubes filled with
137 Lysing Matrix C and ground twice in MP FastPrep-24 TM 5G (Mp Biomedicals, USA) with a
138 speed setting of 4 for 20 seconds. The paired-end library with 150 bp was prepared and
139 sequenced using Illumina Novaseq 6000. All sequencing was performed at Annoroad Gene
140 Technology Co., Ltd, (Beijing, China).

141 To generate a chromosome-level and haplotype phased assembly of *P. polysora* genome, a
142 Hi-C library was generated following in situ ligation protocols. In brief, fresh urediniospores (~
143 100 mg) of GD1913 were used for crosslinking reaction by 2% formaldehyde at room
144 temperature for 15 min. After Glycine quenching, the supernatant was removed and spores were
145 then ground with liquid nitrogen for DNA extraction. The purified DNA was digested with *Mbo*
146 I restriction enzyme (New England Biolabs Inc. Beijing, China) and was labeled by incubating
147 with Biotin-14-dATP (Thermo Fisher Scientific, Massachusetts, USA) and then ligated by T4
148 DNA Ligase (Thermo Fisher Scientific, Massachusetts, USA). After incubating overnight to
149 reverse crosslinks, the ligated DNA was sheared into ~350 bp fragments. Finally, the Hi-C
150 library was quantified by Bioanalyzer and sequenced on the Illumina NovaSeq 6000 platform
151 using paired-end 150 cycles.

152 **2.3 | *De novo* assembly, haplotype phasing and Hi-C scaffolding**

153 HiFi reads were assembled by Canu 2.1.1 with -pacbio-hifi (Nurk et al. 2020). The coverage
154 depth was calculated by genomeCoverageBed in BEDtools (v.2.29.2) (Quinlan and Hall 2010)
155 and t oronmll contigs (< 20 kbp) with low coverage (< 2×) were excluded from the further

156 assembly. The remaining contigs were examined by BLASTN search (v.2.7.2) against the NCBI
157 nt/nr database (downloaded on May 20, 2021) with E-value set as 1e-10. Contigs with significant
158 matches to mitochondrial, plant rDNA, or chloroplast sequences were discarded in the final
159 assembly.
160 To obtain a haplotype-phased assembly, HaploMerger2 (Huang et al. 2017), a tool to
161 rebuild both haploid sub-assemblies from the high-heterozygosity diploid genome, was used.
162 The heterozygosity evaluated by Jellyfish v. 2.1.3 (Marcai and Kingsford 2011) and
163 Genomescope.R (Vurture et al. 2017) was 1.08% leading to the identity setting as 95% (Huang et
164 al. 2017). To better distinguish allelic and non-allelic combinations, the scoring scheme for
165 alignments was recalculated by a Perl script (lastz_D_Wrapper.pl) that came with HaploMerger2.
166 The top 22 (in length) contigs accounting for ~10% length of all contigs were assigned to
167 part1.fasta and others were in part2.fasta. Mis-joins were detected by three rounds and all breaks
168 were manually checked combined with Hi-C validation. Based on all-vs-all alignment
169 (hm.new_scaffolds), contigs were assigned separately to haplotype A and haplotype B.
170 For scaffolding, the raw Hi-C data were trimmed by removing adapters and low-quality bases,
171 and then two-paired reads were separately mapped to two haplotypes independently using BWA-
172 MEM v.0.7.8 (Li and Durbin 2010). To move experimental artifacts, the alignments went
173 through the mapping workflow from the Arima Genomics pipeline
174 (https://github.com/ArimaGenomics/mapping_pipeline/blob/master/01_mapping_arima.sh).
175 Then SALSA v.2.2 (Ghurye et al. 2017) was run to cluster initial contigs into groups. SALSA
176 scaffolding was performed independently on haplotype A and B contigs. Hi-C contact matrices
177 for each scaffold were calculated by HiC-Pro (Servant et al. 2015) and contact maps are
178 visualized by HiCPlotter (Akdemir and Chin 2015). Contig reversal, breaking of mis-joins were
179 corrected manually according to Hi-C contact map and global alignment resulted from
180 HaploMerger2 (hm.new_scaffold). Hicexplorer v.3.4.1 (Wolff et al. 2018)) was used to calculate
181 Hi-C links of read pairs to each haplotype. To guide scaffolds to chromosomes, telomeres were
182 identified by a custom Perl script. The possible tandem repeats, (CCTAAA/TTAGGG)n in most
183 filamentous ascomycete fungi (Lue 2021), and other irregular type reported in closely related
184 rust fungi, (CCCTAA/TTTAGG)n, (CCCCTAA/TTAGGGG)n, (CCCTAAA/TTTAGGG)n,
185 (TTAGGG/CCCTAA)n, were all tested (Li et al. 2019; Tobias et al. 2021). Finally
186 (CCCTAAA/TTTAGGG)n was identified as the telomere repeat units for *P. polysora*. Then

187 corrected scaffolds were rebuilt to chromosome-level by a Perl script. The workflow of
188 chromosome assembly has been illustrated in Figure 1A. The completeness of each haplotype
189 was evaluated using BUSCOs of basidiomycota_odb9 and *Ustilago maydis* as the selected
190 species for AUGUSTUS gene prediction (Stanke and Morgenstern 2005) in BUSCO v. 3.0.2
191 (Simão et al. 2015).

192 **2.4 | RNA extraction and sequencing**

193 For gene annotation, transcriptome sequencing of *P. polysora* from various infection time points
194 was performed. Corn seedlings of 18 pots (10 seedlings/pot) were inoculated with 10 mg
195 spores/mL mixed of Tween 20 (v/v: 0.05%). About five infected leaves were cut off per
196 biological replicate at 1 day post inoculation (dpi), 2 dpi, 4 dpi, 7dpi, 10 dpi, and 14 dpi. In
197 addition, germinated spore samples were prepared by placing 30 mg of fresh urediniospores on
198 the surface of sterile water at 25–28°C in dark for12 hours. Three biological replicates were
199 performed for each condition. All samples were collected in 4 mL Eppendorf tubes, frozen in
200 liquid nitrogen before storing at -80°C. Before RNA extraction, samples were ground in liquid
201 nitrogen and RNA was extracted using the RNeasy Plant Minikit (Qiagen) according to the
202 manufactuer's protocols (https://www.qiagen.com/us/resources/download.aspx?id=246847e7-
203 0095-43e4-8d1d-41df3f9153dd&lang=en). After checking the RNA quality by Bioanalyzer,
204 ~350 bp library was constructed and sequenced by Illumina Novaseq 6000 platform (Annoroad
205 Gene Technology Co., Ltd, Beijing, China). Before transcriptome assembly, raw reads were
206 trimmed using Trimmomatic v.0.36 (Bolger et al. 2014) with the settings: ILLUMINICLIP
207 2:30:10 LEADING 3, TRAILING 3 SLIDINGWINDOW 4:10 and MINLEN 50. A reference-
208 guided transcriptome assembly was performed in Trinity v.2.8.5 (Grabherr et al. 2011) using
209 combined reads from germinated spores and infected leaves.

210 **2.5 | Gene prediction and genome annotation**

211 The two haplotypes were annotated independently using Funannotate v1.8.7
212 (https://github.com/nextgenusfs/funannotate/releases/tag/v1.8.7). A pipeline of core modules was
213 applied as mask-> train->predict->update-> fix->annotate. The repeats of the assembled genome
214 were soft-masked according to a merged library including a self library in RepeatMasker v.4.0.8
215 (Smit et al. 2015) and a genome-trained library from RepeatModeler v.1.0.11 (Smit and Hubley
216 2008). The retrotransposons with long terminal repeats (LTR-RTs) were annotated with an in-
217 house pipeline that uses LTRharvest (Ellinghaus et al. 2008) packaged in LTR_retriever v2.9.0

218 (Ou and Jiang 2018). LTRharvest was also used to estimate the insert time of LTR-RTs
219 according to the mutation rate of 2.0×10^{-8} of a closely related fungus, *Schizophyllum commune*
220 (Baranova et al. 2015). The prediction step (funannotate predict) was run with transcript
221 evidence from HISAT2 (Kim et al. 2015) RNA-seq alignments and genome-guided Trinity
222 assemblies. Transcript evidence was aligned to two haplotypes separately using Minimap2 (Li
223 2018) and the protein evidence was aligned to the genome via Diamond (Buchfink et al.
224 2015)/Exonerate (Slater and Birney 2005) with the default UniProtKb/SwissProt protein
225 database (http://legacy.uniprot.org/uniprot/?query=reviewed%3Ayes) from funannotate. The
226 PASA gene models were parsed to train AUGUSTUS v3.2.3 (Stanke and Morgenstern 2005),
227 snap and GlimmerHMM. In addition, GeneMark_ES (Lomsadze et al. 2005) was self-trained
228 using two haplotypes' sequences. All above evidence was combined with default weight settings
229 using Evidence Modeler (Haas et al. 2008) and filtered by removing genes with short length (<
230 50 aa), spanning gaps and transposable elements. The tRNA genes were predicted using
231 tRNAscan-SE v1.3.1 (Lowe and Chan 2016). Funannotate update command to add UTR data to
232 the predictions and fix gene models that are in disagreement with the RNA-seq data. Functional
233 annotation was performed based on available databases including Pfam v. 34.0) (Finn et al.
234 2014), InterPro (v. 86.0) (Jones et al. 2014), eggNOG (v5.0) (Huerta-Cepas et al. 2019),
235 UniProtKB (v. 2021_03), MEROPS (v. 12.3) (Rawlings et al. 2016), carbohydrate hydrolyzing
236 enzymatic domains (CAZymes) (Terrapon et al. 2017) and a set of transcription factors based on
237 InterProScan domains to assign functional annotations.

238 **2.6 | Interhaplotype variation analysis**

239 Small variants including SNPs and indels were identified by mapping trimmed short reads from
240 DNA against haplotype A with BWA-MEM v0.7.8 (Li and Durbin 2010). After removing PCR
241 duplicates using Picard v.2.18.27 (https://github.com/broadinstitute/picard), the bam file was
242 input to GATK 4.1.9 (https://github.com/broadinstitute/gatk) to call SNPs. SNPs and Indels were
243 filtered using gatk VariantFiltration with parameters QD > 2.0 & MQ > 40.0 & FS < 50.0 &
244 SQR < 3.0 for SNPs and QD > 2.0 & QUAL > 30.0 & FS < 200.0 & ReadPosRankSum>-20.0.
245 Variants were annotated from genome location and functional impact using SnpEff v4.3
246 (Cingolani et al. 2012). To estimate structure variations (SVs) between two haplotigs, we aligned
247 18 chromosomes of haplotype A to their corresponding chromosomes in haplotype B by using
248 the nucmer program in Mummer 4 (--maxmatch -t 100 -l 100 -c 500) (Marçais et al. 2018). The

249 chromosome-pair alignments were then analysed using Assemblytics (http://assemblytics.com/)
250 with the unique sequence length=10,000, maximum variant size=10,000 and minimum variant
251 size=1. To identify the orthologous genes between two haplotypes, we used OrthoFinder v2.5.4
252 with the default parameters (Emms and Kelly 2015).

253 **2.7 | Prediction of candidate effectors**

254 Secreted proteins were predicted based on two rules 1) the presence of a predicted signal peptide
255 using SignalP 4.0 (Petersen et al. 2011); and 2) the absence of predicted transmembrane domains
256 outside the first 60 amino acids (with TMHMM 2.0). To predict candidate effectors, we use
257 EffectorP 3.0 (Sperschneider and Dodds 2021), which applied two machine learning models
258 trained on apoplastic and cytoplasmic effectors. The density plots for genes, repeats, and secreted
259 proteins from chromosomes in two haplotypes were generated using KaryoploteR (Gel and Serra
260 2017). A recent study identified the first effector of *P. polysora*, *AvrRppC* (Deng et al. 2022). To
261 locate its genome position, we blasted the CDS sequence of $AvrRppC^{ref}$ against genome
262 sequences of two haplotypes in this study.

263 To understand the expression pattern of secreted proteins, we used all expression data from
264 germinated spores and six timepoints to perform a differential expression analysis. RNA-seq
265 reads were mapped to haplotype A by HISAT2 v2.2.1 (Kim et al. 2015) and FeatureCounts
266 v1.5.3 (Liao et al. 2014) was used to generate read counts for each gene model of secreted
267 protein. Differentially expressed genes were identified by expression in plants relative to
268 germinated spores (|log fold change| > 1.5; adjusted $P < 0.1$) using the DESeq2 R package (Love
269 et al. 2014). The average rlog-transformed values for each gene were used for clustering using
270 the *k*-means method. The optimal number of clusters was defined using the elbow plot method
271 and circular heatmaps were plotted using the Circlize R package (Gu et al. 2014).

272 **2.8 | Comparative genome analysis**

273 The high quality of *P. polysora* genome provides the opportunity for comparison with other
274 chromosome-level references, such as wheat stem rust (*Pgt*21-0), wheat left rust (*Pt*76) and oat
275 crown rust (*Pca*203) (Li et al. 2019; Duan et al. 2022; Henningsen et al. 2022). Syntenic gene
276 pairs among four *Puccinia* species were identified using the MCSCAN toolkit (Wang et al.
277 2012). Figures were plotted using MCscan (Python version).
278 (https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)). Orthofinder v2.5.4 (Emms

279 and Kelly 2015) was applied to compare the orthogroups and the proteome between *P. polysora*
280 and its close rust species with available high-quality genome.

281 **2.9 | Population genetic analyses**

282 To understand the population differentiation of *P. polysora* in China, 79 isolates were
283 resequenced. Reads were qualified by Trimmomatic v.0.36 (Bolger et al. 2014) using the
284 parameters described in the RNA extraction and sequencing paragraph. To assess whether each
285 isolate comprises a single genotype free of contamination, the distribution of read counts for bi-
286 allelic SNPs was calculated by R package vcfR v.1.8.0 (Knaus and Grunwald 2017) and plotted
287 using ggplot2 (v2.2.1) (Wickham 2016). The normal distribution of read allele frequencies at
288 heterozygous positions is expected to rule out contamination from other *Ppz* genotypes and only
289 the pure isolates were applied in the following population genomic analyses. SNPs calling and
290 filtration were performed by GATK v4.1.9 as described in the interhaplotype variation analysis
291 section. Population structure was detected and quantified using principal component analysis
292 (PCA), which was performed using Plink v.1.9 (Purcell et al. 2007). Pairwise FST values were
293 estimated between geographical regions using the PopGenome R package (v2.6.1) (Pfeifer et al.
294 2014). To assess the effects of variants, the vcf file including selected SNPs was run in SnpEff
295 v4.3 (Cingolani et al. 2012). Because the sexual stage of *P. polysora* has not been reported, we
296 used the standardized index of association (*rd*) to test linkage disequilibrium in the Chinese *Ppz*
297 population. We constructed 100 sets of 10,000 random SNPs by samp.ia function from R
298 package *poppr* (Kamvar et al. 2014) to generate a distribution of *rd* values. The observed *rd*
299 distribution of the Chinese population was compared to the distribution of 10,000 *rd* values
300 constructed using fully randomly simulated datasets with 0%, 50%, 75% and 100% linkage.

301 **3 | RESULTS**

302 **3.1 | Chromosome-level genome assembly and haplotype-phasing of *P. polysora***

303 We generated a total of 53 Gbp of circular consensus reads (32× coverage) from single-molecule
304 real-time sequences on the PacBio Sequel II. A total of 4056 contigs were assembled with a
305 genome size of 1.76 Gbp and a contig N50 of 1.9 Mbp (Table 1). A total of 1627 contigs,
306 accounting for ~2% of the *de novo* assembly size were excluded due to small size, low coverage
307 or high mitochondrial similarity (Table S2). By using Haplomerger2, the remaining contigs were
308 assigned to two haplotypes, A and B, with 613 contigs representing ~862 Mbp and 1321 contigs
309 representing ~852 Mbp, respectively. These contigs were further connected to 173 and 344

310 scaffolds in haplotypes A and B by Hi-C scaffolding (Table 1), after manual checking of 33
311 duplicated scaffolds between (17) and within (16) haplotypes (Table S3). By considering both
312 the Hi-C contact maps and the all-vs-all alignment from Haplomerger2, 18 super scaffolds were
313 binned for each haplotype manually. The Hi-C contact map showed evidence of a single
314 centromere on each scaffold and all scaffolds contained telomere sequences at either end (Table
315 1, Figures 1B, 1C), indicating that each haplotype contains 18 chromosomes. This is consistent
316 with the observations of 18 chromosomes in the genome of other *Puccinia* species (Boehm et al.
317 1992; Li et al. 2019; Wu et al. 2021; Duan et al. 2022). In addition, 15 (6.5 Mbp)) and 20 (4.5
318 Mbp) small scaffolds in haplotype A and B respectively could not be assigned to any
319 chromosome. The 18 chromosomes of haplotype A showed an average of over 80% Hi-C links
320 to haplotype A, suggesting high levels of nuclear phasing (Figure 1D). The chromosomes were
321 numbered according to the synteny with the other *Puccinia* species (see below).

322 **3.2 | Assessment of repetitive DNA content**

323 The completeness of the *Ppz* assembly was assessed based on highly conserved Basidiomycete
324 genes (1335 BUSCOs), which suggested that the two haplotypes showed a similar level of
325 completeness, with approximately 90% complete BUSCO genes and an additional 3.5%
326 fragmented BUSCOs (Table 1). However, 126 BUSCO genes were complete in haplotype A but
327 not present in haplotype B. On the contrary, 127 complete BUSCO genes are only present in
328 haplotype B, which suggests interhaplotype variation in *P. polysora*. Considering the two
329 haplotypes in combination, only 11 BUSCOs were missing or fragmented, resulting in 99%
330 complete BUSCOs. That means we assembled a near-complete genome of *P. poplysora*.

331 Using the *de novo* library of RepeatMasker and genome-trained library of RepeatModeler,
332 RepeatMasker detected repeats accounting for 85% of the diploid assembly (both haplotypes),
333 which is significantly higher than observed in closely related species of *Puccinia* (30%–59%)
334 (Miller et al. 2018; Schwessinger et al. 2018; Wu et al. 2021), but similar to *Austropuccinia*
335 *psidii* (85% estimated in this study but 91% in its original description) (Tobias et al. 2021). The
336 repeat density and GC content along the two haplotypes were illustrated in Figure 2A. The large
337 percentage of repeats in *P. polysora* was mainly caused by Class I retrotransposons with long
338 terminal repeats (LTR), with the LTR-Gypsy superfamily most abundant (33.5%), and LTR-
339 Copia next (17.9%). The TE family composition was similar to other rust species, with the
340 exception of *A. psidii*, which is LTR-Gypsy dominated with LTR-Copia accounting for only

341 1.9% (Figure 2B). In addition, about 20% and 16% of repeats of *P. polysora* in two haplotypes
342 were unclassified repeat families based on Repbase (Bao et al. 2015).
343 Since LTR-RTs occupy the majority of the *P. polysora* genome, we examined whether
344 these repetitive sequences slowly accumulated over time or alternatively were subject to sudden
345 expansion in the life history of *P. polysora*. A total of 41,634 LTR-RT pairs were extracted. An
346 LTR-RT burst was estimated at around 1.7 Mya (Figure 2C).

347 **3.3 | High levels of interhaplotype gene content and structural variation**

348 RNA-seq reads from germinated spores and infected corn leaves at 1, 2, 4, 7, 10, and 14 dpi
349 (Table S4) were pooled and used to generate genome-guided transcriptome assemblies. In total,
350 we annotated 23,270 and 24,242 gene models on haplotype A and haplotype B, respectively
351 (Table 2). However, the gene space accounts for only ~4% of the total genome size. The two
352 haplotypes showed a similar level of functional annotation with about 51% of proteins having at
353 least one functional annotation. Orthofinder identified a total of 13,057 common orthogroups
354 between the two haplotypes, involving 20,802 genes from haplotype A and 21,519 genes from
355 haplotype B, over 90% in each haplotype. In addition, 198 and 235 orthogroups were specific to
356 haplotype A or B respectively. About 70% of haplotype-specific genes have no functional
357 annotations, with the remainder having annotated functions mainly in RNA mediated
358 transposition, transmembrane amino acid transposition or ATP binding (Table S5).
359 By mapping Illumina reads to only haplotype A, we detected a total heterozygous rate of
360 6.1 variants/kbp, in which SNP variant rate is dominated (5.8 SNPs/kbp). About 78% of the
361 SNPs were located in intergenic regions and the rest (22%) were detected in 22,802 protein
362 coding genes.The results of Assemblytics suggested that structural variation between haplotype
363 A and B comprised about 3.0% (26/849 Mbp) of the haploid genome size (Figure 3A). The full
364 chromosome-pair alignment between two haplotypes is illustrated in Figure 3B, with alignments
365 of chromsome09 and chromsome15 in Figure 3C and D), showing some insertions/deletions and
366 inversions. Among three types of variation, insertions/deletions and repeat
367 expansions/contractions are more prevalent than tandem expansions/contractions. Besides,
368 variation with size bins of 500 to 10,000 bp is the most prevalent, which accounts for 2.8%.
369 Because the maximal variant size was restricted to 10,000 in Assemblytic, the actual difference
370 between the two haplotypes could be even higher than estimated.

371 **3.4 | Prediction of secretome and candidate effectors.**

372 We predicted 1183 and 1251 secreted proteins on haplotype A and haplotype B, respectively. By
373 using the machine-learning tool, EffectorP3.0, about 14% and 27% of all secreted proteins were
374 predicted as apoplastic and cytoplasmic effectors, respectively (Table 2). GO enrichment
375 analysis and Interproscan annotation predicted some effectors involved in hydrolase activity,
376 catalytic activity, protein tyrosine phosphatase activity, metal ion binding, but most effectors had
377 no homology with known or predicted functions (Table S6, Figure S1). Then we used gene
378 expression data to predict clusters in the secretome that are differentially expressed during
379 infection. By using *k*-means clustering, seven clusters of genes with different expression profiles
380 were detected in haplotype A (Figure 4A). Genes in cluster 1 showed high expression in
381 germinated urediniospores and early infection (1dpi, 2dpi) but low expression at 4dpi and 7dpi
382 when haustoria form. On the contrary, genes in clusters 2, 3, 4 and 5 showed low expression in
383 germinated spores, and highest in planta expression at days 2-4 (cluster 2), days 2-7 (cluster 3)
384 and from day 7 (cluster 4 and 5). Clusters 6 and 7 were more uniform in expression through
385 these stages, although cluster 6 genes increased later in infection. About 33% to 80% of the
386 secreted proteins in these clusters were predicted as candidate effectors by EffectorP (Table S7).
387 A similar set of expression profile clusters were also detected for genes in haplotype B (Figure
388 4A). Nevertheless, two haplotypes presented different levels of carbohydrate-active enzymes
389 (CAZymes). A total of 179 and 289 CAZymes were detected, of which 8 vs 77 CAZymes were
390 predicted to be secreted in haplotypes A and B, respectively (Table S7). Among CAZymes
391 subclasses, Glycoside hydrolase (GH) enzymes are abundant, accounting for 48% (88 GH
392 families) and 42% (126 GH families) of all CAZymes of haplotype A and haplotype B,
393 respectively. Of these, 51 GH families were predicted to be secreted. The GH5 (cellulase and
394 other diverse forms being exo-/endo-glucanases and endomannannases) family (Langsto et al.
395 2011) was observed to be largely expanded in *P. polysora* as well as other *Puccinia* species
396 (Figure 5A). Besides, AA3 (glucose-methanol-choline oxidoreductases), CE4 (chitin and
397 peptidoglycan deacetylases), GH18 (chitinases of classes III and V), CH47(α-mannosidases),
398 GT2 and GT90 are also abundant in *P. polysora*.
399 We also investigated other factors potentially associated with infection or plant immune
400 inhibition, such as transcription factors (TFs) and peptidases. We found most TF families of *P.*
401 *polysora* have low abundance except C2H2-type (IPR013087) and CCHC-type (IRP001878)
402 zinc finger class, which have 87 and 67 members in haplotypes A and B (Figure 4B). In total, we

403 annotated 286 and 304 proteases in haplotypes A and B, with A01A (aspartic proteases), C19
404 (ubiquitinyl hydrolases), C26 (gamma-glutamyl hydrolase), S08A (subtilisin-like serine
405 protease), S09X (glutamyl endopeptidase C), S10 (carboxypeptidase Y), S33 (prolyl
406 aminopeptidase) and T01A (component peptidases of the proteasome) families expanded in *P.*
407 *polysora* as well as in other three *Puccinia* species (Figure 4C). A total of 8 (2.8%) and 30
408 (10.0%) proteases were predicted to be secreted and these proteases showed a discrete
409 distribution in each cluster without an obvious clustering pattern (Table S7).

410 **3.5 | No evidence to support the "two-speed genome" in *P. polysora***

411 Analysis of the local gene density, measured as flanking distances between neighboring genes
412 showed that the flanking distances in the *P. polysora* genome are generally rather high, with an
413 average distance between genes of 100 kbp (Figure 4B). Accordingly, the surrounding genomic
414 context of most genes in *P. polysora* genome is gene-sparse and repeat-rich. Large flanking
415 distances are not specific to candidate effectors. In line with this pattern, the gene distance
416 density plots revealed very similar distributions between all genes and candidate effectors,
417 including the only known *Avr* effector gene, *AvrRppC* (Figure 4B). We further investigated
418 whether candidate effectors present a different distribution of gene distance density compared to
419 basidiomycete core ortholog genes of *P. polysora*. The results highlighted that candidate
420 effectors are not located in peculiar gene-sparse areas, and the flanking distance centers of both
421 BUSCOs and candidate effectors overlap with that of all genes (Figure 4B).

422 **3.6 | Anchoring the *AvrRppC* in *Ppz*-GD1913**

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

432 **3.7 | Comparative genomic analysis**

433 Although *P. polysora* has a 7–10 times larger genome size than other rust species, the gene
434 synteny amongst four *Puccinia* species with full chromosome assemblies is well conserved, with
435 18 chromosomes corresponding to each other clearly without any chromosome rearrangement
436 (Figure 6A). A few large inversion blocks were detected in chromosomes 1, 3, 4, 7, 8, 9 and 10
437 between *Ppz*A and *Pt*76 (Figures 6A, S2). Inversion blocks were also detected in chromosomes 5,
438 7, 10, 13 between *Pt*76 and *Pca*203, and in chromosomes 2, 3, 5, 7, 8, 10 and 13 between *Pgt*21-
439 0 and *Pca*203 (Figures 6A, S2).
440 The protein orthologs among four closely related *Puccinia* species were compared based on
441 haplotype A of each species. It seems that *Puccinia polysora* experienced gene expansion events
442 compared with the other three species (Figure 6B, Table S8). Among 11,840 orthogroups
443 identified from the four species, genes of *P. polysora* were assigned to 7,148 orthogroups which
444 was less than those of over 8,000 orthogroups for the other three species. Nevertheless, *P.*
445 *polysora* presented more species-specific orthogroups and these orthogroups are featured with a
446 high copy number of gene duplication (Figure 6C). Based on Eggnog annotation result, the
447 orthogroups of *P. polysora* with high copy numbers (>10) are mainly associated with the
448 pathogenicity-related function, energy metabolism, RNA regulation and heat shock reaction
449 (Table S9).
450 **3.8 | Population genetics**
451 To understand the population differentiation of *P. polysora*, we analyzed genome resequencing
452 data from 79 isolates of *Ppz* collected from across maize-growing regions of China and
453 performed population genetic analysis. The analyses of bi-allelic frequency suggested the
454 expected normal distribution of bi-allelic frequency for most isolates except GD1922-3 and
455 GX1905-2 (Figure S3) which were excluded from the analysis. After removal, a total of
456 7,147,489 whole-genome SNPs were obtained from the remaining 77 isolates. PCA analysis
457 separated the isolates into 2 groups, with a major group consisting of 73 isolates (blue circle in
458 Figure 7B) and a minor group containing only four isolates (red circle in Figure 7B). However,
459 this separation was independent of the geographic origin of isolates (North, Central and South
460 China). Correspondingly, although higher genetic differentiation was revealed between North
461 China and South China populations (Fst $=8.1\times10^{-4}$) than in other regional pairs (3.9×10^{-4} and
462 3.6×10^{-4}), all pair-wise *Fst* values were close to zero indicating a lack of geographic
463 differentiation. We used the standardized index of association, *rd*, to estimate the linkage

464 disequilibrium of *P. polysora* population in China. The observed *rd* distribution for all isolates
465 was between the values calculated from simulated datasets with 0% linkage, a sexual population,
466 and 50% linkage (Figure 7C). The result suggested that, although *Ppz* population in China
467 showed as a clonal population, its evolutionary history may be influenced by some level of
468 sexual reproduction.

469 Due to limited samples in the minor group, we did not calculate the *F*st between two genetic
470 groups instead of comparing the highly differentiated SNPs between two genetic clusters. A total
471 of 32,281 SNPs (*F*st > 0.9) were filtered and these SNPs were evenly distributed in 18
472 chromosomes without region specificity. These variants were annotated in 5,975 protein-coding
473 genes, of which 305 genes were affected by the moderate or high impact on amino acid
474 sequences. The known functional annotations of these 305 genes were mainly related to
475 transmembrane transport, signal transduction, zinc ion/protein/nucleic acid binding, catalytic
476 activity and protein kinase activity (Table S10). Interestingly, 16 secreted protein including
477 *AvrRppC* also showed mutations. The allele types of *AvrRppC* varied between the two genetic
478 groups. In the major group, most isolates (93%) carried allele types of $AvrRppC^{1}$ and $AvrRppC^{ref}$.
479 A few isolates showed the allele combination containing new allele types, $AvrRppC^{1}$ to
480 $AvrRppC^{5}$ as well as $AvrRppC^{ref}$. Whereas in the minor group, all four isolates carried the allele
481 of $AvrRppC^{A}$ and $AvrRppC^{J}$ (Figure 7D). The CDS and amino acid sequences for new allele
482 types are listed in Table S11.

483 **4 | DISCUSSION**

484 As rust fungi are dikaryotic pathogens, obtaining their nuclear phased assembly is critical for
485 pathogenicity studies. In this study, we reported a haplotype-phased and chromosome-scale
486 genome of *P. polysora* based on HiFi reads and Hi-C data. The 18 chromosomes had telomeres
487 at both ends and showed high completeness (~90% complete BUSCOs for each nucleus and 99%
488 complete BUSCOs for two nuclei) and high continuity (N50 of scaffolds = 54 M) as well as over
489 80% nuclear phasing. These data strongly supported that we generated a high-quality reference
490 genome for *P. polysora*. In addition, this is the first case to assemble a gigabase-sized fungal
491 genome to chromosome level. The assembly pipeline we illustrated in Figure 1 will inform
492 future genome assembly for other dikaryotic or non-haploid organisms. Although NuclearPhaser
493 has been developed and successfully applied in three rust other species (*Pgt*21-0, *Pt*76 and

494 *Pca*203) (Li et al. 2019; Duan et al. 2022; Henningsen et al. 2022), whether it can handle large
495 genome size (> 1 G) needs to be further confirmed.
496 Compared to its close relatives, *P. polysora* has experienced a large genome expansion,
497 with its genome size of 1.71 Gbp equivalent to 7–10 times of four other *Puccinia* species
498 assemblies (*Pst*-104E, *Pgt*21-0, *Pt*76, and *Pca*203) (Schwessinger et al. 2018; Li et al. 2019; Wu
499 et al. 2021; Henningsen et al. 2022). Based on a limited set of rust species, Tavares et al (2014)
500 inferred that rusts infecting monocot Poaceae hosts have considerably smaller genomes than
501 rusts infecting dicot Fabaceae hosts (556.6 Mbp on average). Thus, our study supplied an
502 exception to this observation, indicating it may not be a general model. Gene duplication, repeat
503 content variation and ploidy variation are common mechanisms to account for genome expansion
504 of fungi (Castanera et al. 2016; Sipos et al. 2017; Todd et al. 2017). In this case, although gene
505 expansion or loss has happened in the evolutionary process from *P. polysora* to other wheat-like
506 *Puccinia* species (Figure 6B) and correspondingly the gene number predicted in *Ppz*-GD1913
507 (~24000 per haplotype) is much more than those of its two close relatives, *P. coronata* (~18000)
508 and *P. striiformis* (~14000) (Miller et al. 2018; Schwessinger et al. 2018), the expanded genes
509 only contribute to a size difference of 7 Mbp and 17 Mbp (Table S12). On the contrary, repeat
510 analyses suggested that genome expansion of *P. polysora* is due to its large repeat contents (85%
511 of the total genome), notably a proliferation of LTR-RTs, which is consistent with the common
512 evidence for rust genome expansion (Tobias et al. 2021). The evidence of genome expansion has
513 also been detected in other obligate fungi, such as powdery mildew (PM) fungi (Spanu et al.
514 2010; Frantzeskakis et al. 2020) and symbiotic fungi (Miyauchi et al. 2020). PM fungi lose
515 carbohydrate-active enzymes, transporters, etc. which are probably redundant genes in strict
516 parasitism as tradeoffs (Spanu et al. 2010). The low abundance of PCWDEs and transcription
517 factors in *P. polysora* as well as other *Puccinia* species was in line with the above pattern (Figure
518 5) (Duplessis et al. 2011, Miller et al. 2018). Despite the massive TE proliferation, *P. polysora*
519 shows high gene synteny with other *Puccinia* species (Figure 5), as suggested by comparing
520 BUSCO genes in another study (Henningsen et al. 2022).
521 In fungi, TEs have been implicated in coevolution with the host, genome architecture,
522 plasticity, and adaptation (Lorrain et al. 2021). The "TE-thrust hypothesis" proposed that TEs
523 can act to generate genetic novelties for organisms leading to speciation or may be related to
524 adaptation to new challenges, e.g. new host or rapidly changing climate (Oliver and Greene 2012;

525 Zeh et al. 2009). We detected a TE burst of *P. polysora* at 1.7 Mya (Figure 2) and the
526 phylogenetic tree (Figure 6) suggested that *P. polysora* speciated earlier than the divergence time
527 between *P. coronata* and *P. striiformis* (40 Mya) (Aime et al. 2018). So, TE burst of *P. polysora*
528 may be unlikely for speciation. Other explanation could be "nearly neutral theory", in which
529 varied genomic features (e.g. large genome size, TE content, introns) are not initially harmful but
530 are passively fixed by mutation or random genetic drift (Lynch and Conery 2003). With more
531 available genome resources in the future, the relationship between effective genome size and
532 adaptative evolution of *P. polysora* needs to be answered.

533 The "two-speed genome" concept has been put forward to highlight the over-representation
534 of effector-like genes in the repeat-rich and gene-sparse genome in many filamentous pathogens
535 (Dong et al. 2015, Frantzeskakis et al. 2019). These TE-rich invasion/blocks may contribute to
536 extensive chromosomal reshuffling and the rise of accessory chromosomes (de Jonge et al. 2013).
537 Pathogens can resolve the evolutionary conflict through rapid evolution of effector genes to
538 adapt to changing environment but maintain the housekeeping genes in the core genome with a
539 moderate evolution rate (Presti et al. 2015). However, the situation is clearly different in rust
540 fungi. There was no signal of effector compartments detected in Oat crown rust, wheat stripe rust
541 and myrtle rust pathogens when comparing the intergenic distance between all genes and
542 effectors (Schwessinger et al. 2018; Miller et al. 2018; Tobias et al. 2021). In *P. polysora*, the
543 numerous TEs are evenly dispersed throughout the genome (Figure 2). The intergenic distance
544 distribution of candidate effectors overlapped with that of all genes and BUSCOs (Figure 7).
545 This genomic architecture with TEs and genes interspersed was also reported in another
546 biotrophic fungal group, powdery mildew (PM) fungi (Frantzeskakis et al. 2020). Similarly, AT-
547 rich isochores or large-scale compartmentalization are also missing in PM fungi (Frantzeskakis
548 et al. 2020). When compared with other *Puccinia* species, the high repeat content of *P. polysora*
549 leads to much lower relative gene space (4% in *Ppz* vs 20%–35% for currently reported *Puccinia*
550 species) and large intergenic distance expansion (~100 kbp of *Ppz* vs 1kbp for *Pst* and *Pca*)
551 (Figure 6B) (Miller et al. 2018; Schwessinger et al. 2018).

552 In the "arms race" between rust species and plants, the *Avr* genes in rust pathogens have
553 been the subject of mutations to avoid the recognition by host resistance genes (Cui et al. 2015).
554 Therefore, monitoring the variation of *Avr* genes in rust populations is a priority for disease
555 management. Based on whole genome SNPs, we detected two genetic groups of *P. polysora* in

556 China. The highly differentiated SNPs between these two groups suggested mutations in 16
557 secreted proteins including the only verified avirulence gene, *AvrRppC*. A recent study
558 investigated the allele types of *AvrRppC* in a Chinese *P. polysora* population (Deng et al. 2022).
559 Isolates with allele types of *AvrRppC*^{A}, *AvrRppC*^{F} and *AvrRppC*^{J} could escape the recognition of
560 *RppC* causing southern corn rust but *AvrRppC*^{ref}, *AvrRppC*^{E}, and *AvrRppC*^{C} are avirulence alleles
561 that trigger *RppC*-mediated resistance. In our study, isolates in the major group almost all carried
562 the avirulence allele, *AvrRppC*^{ref} and four isolates in the minor genetic group carried the
563 virulence allele types, *AvrRppC*^{A} and *AvrRppC*^{J} (Figure 7D). These results suggested that
564 differentiation between two genetic groups could be related to virulence evolution and such
565 relation was more prominent in the wheat stripe rust pathogen (Hubbard et al. 2015). Upadhyaya
566 et al (2021) suggested that avirulence genes of *P. graminis* showed a similar expression pattern.
567 In this study, *AvrRppC* was classified in cluster 3 (Figure 4A) with low expression in germinated
568 spores and early infection but the high expression in intermediate infection (4 dpi and 7 dpi).
569 Among 16 secreted proteins with *F*st > 0.9 between two genetic groups, FUNA_005021 (cluster
570 2) and FUNA_020823 (cluster 3) showed similar expression patterns to *AvrRppC* and may be
571 candidate avirulence genes. In other rust species, the rust population could drastically shift
572 towards a wider or new spectrum of virulence over time (Miller et al. 2020; Bai et al. 2021),
573 therefore, performing long-term surveillance as well as developing effective virulence markers
574 are particularly critical for *P. polysora*.
575 Rust fungi have complex life cycles involving asexual as well as sexual reproduction
576 (Figueroa et al. 2020; Zhao et al. 2021). In some species, sexual reproduction is rare or cryptic
577 that it can remain undiscovered. For example, it took a century to discover the alternate host of *P.*
578 *striiformis* (Jin et al. 2010). In our case, the sexual stage of *P. polysora* is still a mystery;
579 teliospores are rarely found in natural fields and germination tests of teliospores were not
580 successful in past attempts (Cammack 1959). Also, studies on population genetics of *P. polysora*
581 are very limited, possibly due to the lack of effective molecular markers and the reference
582 genome. Our data using whole-genome SNPs supported that the clonal Chinese population of *P.*
583 *polysora*, however, the low *rd* between those from simulated with 0% and 50% linkage
584 suggested potential influence of sexual reproduction. The absence of a sexual stage or alternate
585 host may be reflected by unequal gene numbers in the secretome. A supportive example is two
586 species in *Melampsora*. *Melampsora lini* (autoecious, no alternate host, ~800 secreted proteins)

587 was found to have much less secreted proteins and secreted plant cell wall-degrading enzymes
588 (PCWDEs) than its close relative, *M. larici-populina* (heteroecious, with both primary and
589 alternate hosts, ~1800 secreted proteins) (Presti et al. 2015). In our case, we identified 2434
590 secreted proteins for *P. polysora*, which was comparable to the secretomes of two heteroecious
591 relatives, *P. coronata* (~2500) and *P. graminis* (~2500) (Miller et al. 2018; Li et al. 2019). Of
592 course, this single *Melampsora* example did not allow for making a general conclusion. Varied
593 quality of genome assemblies and different prediction methods can lead to significant bias in
594 secreted protein predictions. However, it will be valuable to test this hypothesis with more
595 available high-quality rust genomes in future studies. If the hypothesis is true, our results may
596 suggest the existence of alternate hosts that have not been discovered yet for *P. polysora.*

597 **5 | CONCLUSION**

598 In conclusion, we successfully obtaining the first nuclear phased and chromosome-scale genome
599 assembly of the serious fungal pathogen, *P. polysora*. The high-quality of genome assembly and
600 fine-time-point transcriptome facilitate the comparative genomic analyses and The genome
601 expansion of *P. polysora* is driven by TEs, but this did not affect its gene synteny with close
602 relatives. The investigation of genome characteristics and functional features provided a
603 fundamental resource for pathogenicity studies as well as to understand the evolutionary
604 mechanisms of genome expansion in rust fungi. Also, the population genomic data revealed low
605 genetic differentiation in the Chinese *P. polysora* population, which expanded from south to
606 north in recent decades. However, a minor genetic group with low frequency suggested the
607 ongoing virulence evolution to evade recognition by *RppC*, a major resistance gene in Chinese
608 corn cultivars, alarming the need to excavate additional resistance genes as soon as possible.
609

610 **ACKNOWLEDG EMENTS**

611 We would like to thank Xiufeng Liu, Peng Lu, Gongxian Liao, Shikun Pan, Xingshang Lu,
612 Baoqin Tan, Qiusheng Luo for field diseased sample collection. We also appreciate valuable
613 suggestions from Dr. Mao Jianfeng and Dr. Qi Wu before submission. Clement K.M. Tsui is
614 grateful to CAS PIFI for the award of visiting scientist fellowship.
615

616 **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 in the National Microbiology Data Center (NMDC, https://nmdc.cn/) under the BioProject (NMDC10018113) and the accession number for all raw data are listed in Table S13. The assembled genome has been deposited at NMDC under the accession number NMDC60042795. Scripts used to construct figures 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 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.
2. Aime, M. C., Bell, C. D., & Wilson, A. W. (2018). Deconstructing the evolutionary complexity between rust fungi (Pucciniales) and their plant hosts. *Studies in Mycology*, 89,143–152.
3. Aime, M. C., McTaggart, A. R., Mondo, S. J., & Duplessis, S. (2017). Phylogenetics and phylogenomics of rust fungi. *Advances in Genetics*, 100, 267–307.
4. Akdemir, K.C., & Chin, L. (2015). HiCPlotter integrates genomic data with interaction matrices. *Genome Biololgy,* 16, 198.
5. Anderson, C., Khan, M. A., Catanzariti, A. M., Jack, C. A., Nemri, A., Lawrence, G. J., Upadhyaya, N. M., Hardham, A. R., Ellis, J. G., Dodds, P. N., Jones, D. A. (2016). Genome

648 analysis and avirulence gene cloning using a high-density RADseq linkage map of the flax
649 rust fungus, *Melampsora lini*. *BMC Genomics,* 17, 667.
650 6. Bai, Q., Wan, A. M., Wang, M. N., See, D. R., & Chen, X. M. (2021). Molecular
651 characterization of wheat stripe rust pathogen (*Puccinia striiformis* f.sp. *tritici*) Collections
652 from Nine countries. *International Journal of Molecular Sciences*, 22, 9457.
653 7. Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase update, a database of repetitive
654 elements in eukaryotic genomes. *Mobile DNA*, 6, 4–9.
655 8. Baranova, M. A., Logacheva, M. D., Penin, A. A., Seplyarskiy, V. B., Safonova, Y. Y.,
656 Naumenko, S. A., Klepikova, A. V., Geraso,pv. E. D., Bazykin, G. A., James, T. Y., &
657 Kondrashov, A. S. (2015). Extraordinary genetic diversity in a wood decay
658 mushroom. *Molecular Biology and Evolution*, 32, 2775–2783.
659 9. Boehm, E. W. A., Wenstrom, J. C., McLaughlin, D. J., Szabo, L. J., Roelfs, A. P., &
660 Bushnell, W. R. (1992). An ultrastructural pachytene karyotype for *Puccinia graminis* f.sp.
661 *tritici*. *Canadian Journal of Botany*, 70, 401–413.
662 10. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for
663 Illumina sequence data. *Bioinformatics*, 30, 2114–2120.
664 11. Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using
665 DIAMOND. *Nature Methods*, 12, 59.
666 12. Cammack, R. H. (1959). Studies on *Puccinia polysora* Underw. III. Description and life
667 cycle of *P. polysora* in West Africa. *Transactions of the British Mycological Society*, 42,
668 55–58.
669 13. Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J.,
670 Grimwood, J., Pérez, G., Pisabarro, A. G., Grigoriev, I. V., Stajich, J. E., & Ramírez, L.
671 (2016). Transposable elements versus the fungal genome: impact on whole genome
672 architecture and transcriptional profiles. *PLoS Genetics*, 12, e1006108.
673 14. Chen, J., Upadhyaya, N. M., Ortiz, D., Sperschneider, J., Li, F., Bouton, C., ... & Dodds, P.
674 N. (2017). Loss of *AvrSr50* by somatic exchange in stem rust leads to virulence for *Sr50*
675 resistance in wheat. *Science*, 358, 1607–1610.
676 15. Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion. G. T., Clum, A., Dunn,
677 C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M.,
678 Luo, C., Ecker, J. R., Cantum D., Ranmk, D, R., & Schatz, M. C. (2016). Phased diploid

679 genome assembly with single-molecule real-time sequencing. *Nature Methods*. 13, 1050–
680 1054.
681 16. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Ruden,
682 D. M., & Lu, X. Y. (2012). A program for annotating and predicting the effects of single
683 nucleotide polymorphisms, SnpEff. *Fly (Austin)*, 6, 80–92.
684 17. Crouch, J. A., & Szabo, L. J. (2011). Real-time PCR detection and discrimination of the
685 southern and common corn rust pathogens *Puccinia polysora* and *Puccinia sorghi*. *Plant*
686 *Disease*, 295, 624–632.
687 18. Cui, H., Tsuda, K., & Parker, J. E. (2015). Effector-triggered immunity: from pathogen
688 perception to robust defense. *Annual Review of Plant Biology,* 66, 487–511.
689 19. de Jonge, R., Bolton, M. D., Kombrink, A., van den Berg, G. C. M., Yadeta, K. A., Thomma,
690 B. P. H. J. (2013). Extensive chromosomal reshuffling drives evolution of virulence in an
691 asexual pathogen. *Genome Research*, 23,1271–1282.
692 20. Deng, C., Leonard, A., Cahill, J. M., Meng, L., Li, Y.R., Thatcher, S. Li, X. Y., Zhao, X. D.,
693 Du, W. J., Li, Z., Li, H. M., Llaca, V., Fengler, K., Marshall, L., Harris, C., Tabor, G., Li, Z.
694 M., Tian, Z. Q., Yang, Q. H., Chen, Y. H., Tang, J. H., Wang, X. T., Hao, J. J., Yan, J. B.,
695 Lai, Z. B., Fei, X. H., Song, W. B., Lai, J. S., Zhang, X. C., Shu, G. P., Wang, Y. B., Chang,
696 Y. X., Zhu, W. L., Xiong, W., Sun, J., Li, B. L., & Ding, J. Q. (2022).The RppC-AvrRppC
697 NLR-effector interaction mediates the resistance to southern corn rust in maize. *Molecular*
698 *Plant,* 15, 904–912.
699 21. Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-
700 pathogen interactions. *Nature Reviews Genetics*. 11, 539–548.
701 22. Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous
702 pathogens: waltz with plants. *Current Opinion in Plant Biology*, 35, 57–65.
703 23. Duan, H., Jones, A. W., Hewitt, T., Mackenzie, A., Hu, Y., Sharp, A., Lewis, D., Mago, R.,
704 Upadhyaya, N. M., Rathjen, J. P, Stone, W. A., Schwessinger, B., Figueroa, M., Dodds, P.
705 N., Periyannan, S., & Sperschneider J. (2022). Identification and correction of phase
706 switches with Hi-C data in the Nanopore and HiFi chromosome-scale assemblies of the
707 dikaryotic leaf rust fungus *Puccinia triticina*. *Current Biology*, 23, 84.
708 24. Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C., Joly, D. L.,
709 Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M.,

710 Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., Kohler, A., Kües, U.,
711 Lindquist, E. A., Lucas, S. M., Mago, R., Mauceli, E., Morin, E., Murat, C.,Pangilinan, J. L.,
712 Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A. A.,
713 Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G. A., Henrissat, B., Van de, Peer.
714 Y., Rouze, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev,
715 I. V., Szabo, L. J., & Martin, F. (2011). Obligate biotrophy features unraveled by the
716 genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*, 108,
717 9166–9171.
718 25. Ellinghaus, D., Kurtz, S. & Willhoeft, U. (2008). *LTRharvest*, an efficient and flexible
719 software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics 9, 18.
720 26. Emms, D.M. & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole
721 genome comparisons dramatically improves orthogroup inference accuracy. *Genome*
722 *Biology,* 16, 157.
723 27. Figueroa, M., Dodds, P. N., & Henningsen, E. C. (2020). Evolution of virulence in rust
724 fungi–multiple solutions to one problem. *Current Opinion in Plant Biology*, 56, 20–27.
725 28. Figueroa, M., Upadhyaya, N. M, Sperschneider, J., Park, R. F., Szabo, L. J., Steffenson, B.,
726 Eills, J. G., Dodds, P. N. (2016). Changing the game: using integrative genomics to probe
727 virulence mechanisms of the stem rust pathogen *Puccinia graminis* f. sp. *tritici*. *Frontiers in*
728 *Plant Science*, 7, 205.
729 29. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A.,
730 Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., & Punta, M. (2014).
731 Pfam: the protein families database. *Nucleic Acids Research,* 42, D222–D230.
732 30. Frantzeskakis, L., Barsoum, M., Panstruga, R., Kusch, S., Kiss, L., Takamatsu, S., &
733 Panstruga, R. (2019). The *Parauncinula polyspora* draft genome provides insights into
734 patterns of gene erosion and genome expansion in powdery mildew fungi. *MBio*, 10, 1–17.
735 31. Frantzeskakis, L., Pietro, A. D., Rep, M., Schirawski, J., Wu, C. H., & Panstruga, R. (2020).
736 Rapid evolution in plant–microbe interactions—a molecular genomics perspective. *New*
737 *Phytologist*, 225, 1134–1142.
738 32. Gel, B., & Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable
739 genomes displaying arbitrary data. *Bioinformatics* 33, 3088–3090.

740 33. Ghurye, J., Pop, M., Koren, S., Bickhart, D., & Chin, C. S. (2017). Scaffolding of long read
741 assemblies using long range contact information. *BMC genomics*, 18, 527.
742 34. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis,
743 X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A.,
744 Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., &
745 Regev, A. (2011). Trinity: reconstructing a full-length transcriptome assembly from RNA-
746 Seq data. *Nature biotechnology*, 29, 644–652.
747 35. Gu, Z. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*
748 30, 2811–2812.
749 36. Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C.
750 R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using
751 EVidenceModeler and the Program to Assemble Spliced Alignments. Genome biology, 9,
752 R7.
753 37. Henningsen, E. C., Hewitt, T., Dugyala, S., Nazareno, E. S., Gilbert, E., Li, F., Kianian, S. F,
754 Steffenson, B. J., Dodds, P. N., Sperschneider, J., & Figueroa, M. (2022). A chromosome-
755 level, fully phased genome assembly of the oat crown rust fungus *Puccinia coronata* f. sp.
756 *avenae*: a resource to enable comparative genomics in the cereal rusts.
757 https://doi.org/10.1101/2022.01.26.477636.
758 38. Huang, S. F., Kang, M. J., & Xu, A. L. (2017). HaploMerger2: rebuilding both haploid sub-
759 assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*, 33,2577–
760 2579.
761 39. Hubbard, A., Lewis, C. M., Yoshida, K., Ramirez-Gonzalez, R. H., de Vallavieille-Pope, C.,
762 Thomas, J., Kamoun, S., Bayles, R., Uauy, C., & Saunders, D. G. (2015). Field
763 pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome*
764 *Biology*, 16, 23. https://doi.org/10.1186/s13059-015-0590-8
765 40. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H.,
766 Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019).
767 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource
768 based on 5090 organisms and 2502 viruses. Nucleic acids research, 47(D1), D309–D314.

41. Jin, Y., Szabo, L. J., & Carson, M. (2010). Century-old mystery of *Puccinia striiformis* life history solved with the identification of *Berberis* as an alternate host. *Phytopathology*. 100, 432–435.
42. Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics (Oxford, England), 30(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031
43. Justesen, A. F., Ridout, C. J., & Hovmøller, M. S. (2002). The recent history of *Puccinia striiformis* f. sp. *tritici* in Denmark as revealed by disease incidence and AFLP markers. *Plant pathology*, *51*, 13–23.
44. Kamvar, Z N., Tabima, J. F., & Grünwald, N. J. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ,* 2, e281.
45. Kim, D., Langmead, B. & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12, 357–360.
46. Knaus, B. J., & Grünwald, N, J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*. 17, 44–53.
47. Kolmer, J. A. (2005). Tracking wheat rust on a continental scale. *Current Opinion in Plant Biology*, 8, 441–449.
48. 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
49. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics,* 34, 3094–3100.
50. Li, H., & Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26, 589–595.
51. 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.

800 52. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene
801 identification in novel eukaryotic genomes by self-training algorithm, *Nucleic Acids*
802 *Research,* 33, 6494–6506.
803 53. Lorrain, C., Feurtey, A., Mller, M., Haueisen, J., & Stukenbrock, E. (2021). Dynamics of
804 transposable elements in recently diverged fungal pathogens: lineage-specific transposable
805 element content and efficiency of genome defences. *G3-Genes Genomes Genetics,* 11,
806 jkab068.
807 54. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and
808 dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.
809 55. Lowe, T. M., & Chan, P. P. (2016). TRNAscan-SE On-line: integrating search and context
810 for analysis of transfer RNA genes. *Nucleic Acids Research*, 44, W54 –W57.
811 56. Lue, N. F. (2021). Duplex telomere-binding proteins in fungi with canonical telomere
812 repeats: new lessons in the rapid evolution of telomere proteins. *Frontier in Genetics*,
813 12,638790.
814 57. Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science,* 302, 1401–
815 1404.
816 58. Maia, T,, Badel, J. L., Marin-Ramirez, G., Rocha, C. M., Fernandes, M. B., da Silva, J. C, de
817 Azevedo-Junior, G. M., Brommonschenkel, S. H. (2017). The *Hemileia vastatrix* effector
818 HvEC016 suppresses bacterial blight symptoms in coffee genotypes with the SH1 rust
819 resistance gene. *New Phytologist*, 213,1315–1329.
820 59. Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S.L., & Zimin, A.
821 (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational*
822 *Biology*, 14, e1005944.
823 60. Marcais, M., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel
824 counting of occurrences of k-mers. *Bioinformatics,* 27, 764–770.
825 61. Miller, M. E., Nazareno, E. S., Rottschaefer, S. M., Riddle, J., Dos Santos Pereira, D., Li, F.,
826 Nguyen-Phuc, H., Henningsen, E. C., Persoons, A., Saunders, D., Stukenbrock, E., Dodds, P.
827 N., Kianian, S. F., & Figueroa, M. (2020). Increased virulence of *Puccinia coronata* f. sp.
828 *avenae* populations through allele frequency changes at multiple putative *Avr* loci. *PLoS*
829 *genetics*, 16, e1009291. https://doi.org/10.1371/journal.pgen.1009291

830 62. Miller, M. E., Zhang, Y., Omidvar, V., Sperschneider, J., Schwessinger, B., Raley, C.,
831 Palmer, J. M., Garnica, D., Upadhyaya, N., Rathjen, J., Taylor, J. M., Park, R. F., Dodds, P.
832 N., Hirsch, C. D., Kianian, S. F., & Figueroa, M. (2018). *De Novo* Assembly and Phasing of
833 Dikaryotic Genomes from Two Isolates of *Puccinia coronata* f. sp. *avenae*, the Causal
834 Agent of Oat Crown Rust. *mBio*, 9, e01650-17. https://doi.org/10.1128/mBio.01650-17
835 63. Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sánchez-García, M., Morin, E.,
836 Andreopoulos, B., Barry, K. W., Bonito, G., Buée, M., Carver, A., Chen, C., Cichocki, N.,
837 Clum, A., Culley, D., Crous, P. W., Fauchery, L., Girlanda, M., Hayes, R. D., ··· Martin, F.
838 M. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the
839 early evolution of symbiotic traits. *Nature Communication,* 11, 5125.
840 https://doi.org/10.1038/s41467-020-18795-w
841 64. Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., ... & Koren, S.
842 (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants
843 from high-fidelity long reads. *Genome research*, *30*, 1291–1305.
844 65. Oliver, K. R.,& Greene, W. K. (2012). Transposable elements and viruses as factors in
845 adaptation and evolution: an expansion and strenthening of the TE thrust hypothesis.
846 *Ecology and Evolution*, 2, 2912–2933.
847 66. Ou, S., & Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for
848 Identification of Long Terminal Repeat Retrotransposons. *Plant Physiology*, 176, 1410–
849 1422.
850 67. Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0:
851 discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 785–786.
852 68. Pfeifer, B., Wittelsbuerger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome:
853 An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology*
854 *and Evolution*, 31, 1929–1936.
855 69. Pinon, J., & Frey, P. (1997). Structure of *Melampsora larici-populina* populations on wild
856 and cultivated poplar. *European Journal of Plant Pathology*. 103,159–173.
857 70. Presti, L. L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., & Tollot, M. (2015). Fungal
858 effectors and plant susceptibility. *Annual Review of Plant biology,* 66, 513–545.
859 71. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., maller, J.m
860 Sklar, P., Bakker, P. I. W. D., & Daly, M. J (2007). Plink: a tool set for whole-genome

861 association and population-based linkage analyses. The *American Journal of Human*
862 *Genetics*, 81, 559–575.
863 72. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing
864 genomic features. *Bioinformatics*. 26, 841–842.
865 73. Ramirez-Cabral, N. Y. Z., Kumar, L., & Shabani, F. (2017). Global risk levels for corn rusts
866 (*Puccinia sorghi* and *Puccinia polysora*) under climate change projections. Journal of
867 Phytopathology, 165, 563–574.
868 74. Ramos, A. P., Tavares, S., Tavares, D., Silva, M. D. C, Loureiro, J., & Talhinhas, P. (2015).
869 Flow cytometry reveals that the rust fungus, *Uromyces bidentis* (Pucciniales), possesses the
870 largest fungal genome reported-2489Mbp. *Molecular Plant Pathology*, 16, 1006–1010.
871 75. Rawlings, N. D., Barrett, A. J., & Finn, R. (2016). Twenty years of the Merops database of
872 proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 44, D343–
873 D350.
874 76. Roach, M. J., Schmidt, S. A., & Borneman, A R. (2018). Purge Haplotigs: allelic contig
875 ressignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19, 460.
876 77. Rochi, L., Diéguez, M. J., Burguener, G., Darino, M. A., Pergolesi, M. F., Ingala, L. R.,
877 Cuyeu, A. R., Turjanski, A., Kreff, E. D., & Sacco, F. (2018). Characterization and
878 comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize
879 common rust. *Fungal Genetics and Biology*, 112, 31–39.
880 78. Salcedo, A., Rutter, W., Wang, S., Akhunova, A., Bolus, S., Chao, S., Anderson, N., De
881 Soto, M. F., Rouse, M.,Szabo, L., Bowden, R. L., Dubcovsky. J., & Akhunov, E. (2017).
882 Variation in the *AvrSr35* gene determines *Sr35* resistance against wheat stem rust race Ug99.
883 *Science*, 358, 1604–1606.
884 79. Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J.
885 M., Dodds, P. N., Figueroa, M., Park, R. F., & Rathjen, J. P. (2018). A near-complete
886 haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f.
887 sp. *tritici* reveals high interhaplotype diversity. *mBio*, 9, e0227517.
888 80. Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., Heard, E.,
889 Dekker, J., & Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data
890 processing. *Genome Biology*, 16, 259.

81. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics,* 31, 3210–3212.
82. Singh, R. P., Hodson, D. P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., Herrera-Foessel, S., Singh, P. K., Singh, S., & Govindan, V. (2011). The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annual Review of Phytopathology*, 49, 465–481.
83. Sipos, G., Prasanna, A. N., Walter, M. C., O'Connor, E., Bálint, B., Krizsán, K., Kiss, B., Hess, J., Varga. T., Slot, J., Riley, R., Bóka, B., Rigling, F., Barry, K., Lee, J., Mihaltcheva, S., LaButti, K., Lipzen, A., Moloney, N. M., Sperisen, C., Kredics, L., Vágvölgyi, C., Patrignani, A., Fitzpatrick, D., Nagy, I., Soyle. S., Anderson, J. B., Grigoriev, I. V., Güldener, U., Münsterkötter, M., & Nagy, L. G. (2017). Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi Armillaria. Nature Ecology & Evolution, 1, 1931–1941.
84. Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6, 1–11.
85. Smit, A., & Hubley, R. (2008). RepeatModeler open 1.0. Institute for Systems Biology, Seattle, WA. http://www.repeatmasker.org/
86. Smit, A., Hubley, R., & Green, P. (2015). RepeatMasker open 4.0. 2013–2015. Institute for Systems Biology, Seattle, WA. http://www.repeatmasker.org/
87. Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., ... & Panstruga, R. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, 330,1543–1546.
88. Sperschneider, J., & Dodds, P. N. (2022). EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Wffectors in Fungi and Oomycetes. *Molecular Plant-Microbe Interaction*, 35, 146–156.
89. Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33, W465–W467.
90. Tavares, D., Romos, A. P., Pires, A. S., Azomjeora, H. G., Caldeirinha, P., Link, T., Abranches, R., Silva, M. do. C., Voegele, R. T., Loureiro, J., & Talhinhas, P. (2014).

921 Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in Plant*
922 *Sceiences*, 5, 422.
923 91. Terrapon, N., Lombard, V., Drula, E., Coutinho, P.M., Henrissat, B. (2017). The CAZy
924 Database/the Carbohydrate-Active Enzyme (CAZy) Database: Principles and Usage
925 Guidelines. In: Aoki-Kinoshita, K. (eds) A Practical Guide to Using Glycomics Databases.
926 Springer, Tokyo.
927 92. Tobias, P. A., Schwessinger, B., Deng, C.H., Wu, C., Dong, C. M., Sperschneider, J., Jones,
928 A. Lou, Z. Y., Zhang, P., Sandhu, K., Smith, G. R., Tibbits, J., Chagne, D., & Park, R. F.
929 (2021). *Austropuccinia psidii*, causing myrtle rust, has a gigabase-sized genome shaped by
930 transposable elements. *G3*, 11, jkaa015.
931 93. Todd, R. T., Forche, A., & Selmecki, A. (2017). Ploidy Variation in Fungi: Polyploidy,
932 Aneuploidy, and Genome Evolution. *Microbiology Spectrum*, 5,10.
933 94. Upadhyaya, N. M., Mago, R., Panwar, V., Hewittm, T., Luo, M., Chen, J., Sperschneider, J.,
934 Nguyen-Phuc, H., Wang, A.H., Ortiz, D., Hac, L., Bhatt, D., Li, F., Zhang, J. P., Ayliffe, M.,
935 Figueroa, M., Kanyuka, K., Wllis, J. G., & Dodds, P. N. (2021). Genomics accelerated
936 isolation of a new stem rust avirulence gene–wheat resistance gene pair. *Nature*
937 *Plants*, 7, 1220–1228.
938 95. Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J.,
939 & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short
940 reads. *Bioinformatics,* 14,2202–2204.
941 96. Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B.,
942 Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: a toolkit for detection and
943 evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40, e49.
944 https://doi.org/10.1093/nar/gkr1293
945 97. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-
946 Verlag.
947 98. Winzer, F. L., Berthon, K. A., Carnegie, A. J., Pegg, G. S., & Leishman, M. R.
948 (2019). *Austropuccinia psidii* on the move: survey based insights to its geographical
949 distribution, host species, impacts and management in Australia. *Biological Invasions.* 21,
950 1215–1225.

99. Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramírez, F., & Grüning, B. A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Research, 46, W11–W16.

100. Wu, J. Q., Song, L., Ding, Y., Dong, C. M., Hasan, M., & Park, R. (2021). A chromosome-scale assembly of the wheat leaf rust pathogen *Puccinia triticina* provides insights into structural variations and genetic relationship with haplotype resolution. *Frontiers in Microbiology,* 12, 704253.

101. Zeh, D. W., Zeh, J. A., & Ishida, Y. (2009). Transposable elements and an epigenetic basis for punctuated equilibria. *BioEssays,* 31, 715–726.

102. 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.

103. 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.

967 **Figure Legends**

968 **FIGURE 1** Chromosome-level assembly of GD1913 isolate of *Puccinia polysora*. (A) The
969 dikaryotic phasing pipeline and switch strategies for duplicated contigs. Duplicated contigs were
970 checked based on the all-vs-all sequence alignment from HaploMerger 2 (Huang et al. 2017). (B)
971 Hi-C-based contig anchoring. The heat map showed the density of Hi-C interactions within
972 haplotype A. The 18 chromosomes are highlighted by blue squares. (C) Schematic representation
973 of assembled chromosomes for GD1913 of each haplotype. (D) Percentage of Hi-C links of each
974 chromosome of haplotype A to either haplotype A (blue) or haplotype B (red).

975 **FIGURE 2** Genomic features of *P. polysora* f.sp. *zeae* isolate, GD1913 and repeat comparison
976 analyses with its close relatives. (A) Genomic landscape of 18 chromosomes for haplotype A
977 (hapA) and haplotype B (hapB). From outer to inner circles: (i) chromosomes, (ii) repeats
978 density, (iii) gene density, (iv) candidate effector density, (v) Guanine-cytosine (GC) content. (ii-
979 iv) are drawn in non-overlapping 100 kbp sliding windows. (B) Comparison of genome size (two
980 nuclei) and repeat contents among *Puccinia polysora* f.sp. *zeae* (*Ppz*) and other rust species,
981 *Melampsora larici-populina* (*Mlp*), *Austropuccinia psidii (Ap*), *Puccinia sorghi* f.sp. *zeae* (*Psz*),
982 *Pucccnia coronata* f.sp. *avenae* (*Pca*), *Puccinia graminis* f.sp. *tritici* (*Pgt*), *Puccinia triticina* f.sp.
983 *tritici* (*Ptt*) and *Puccinia striiformis* f.sp. *tritici* (*Pst*). The isolate names are after the hyphen. (C)
984 The insertion time (Mya) distribution of intact LTRs in *P. polysora*.

985 **FIGURE 3** The interhaplotype variation of *Ppz*-GD1913 isolate. (A) Summary of interhaplotype
986 variation between haplotype A and B using Assemblytis. Six types of specific variation were
987 illustrated by schematic plot and each bar chart indicates the number of bases categorized to each
988 type. (B) Synteny plot of haplotype A and B using Mummer. (C) and (D) Two representative
989 whole-chromosome alignments.

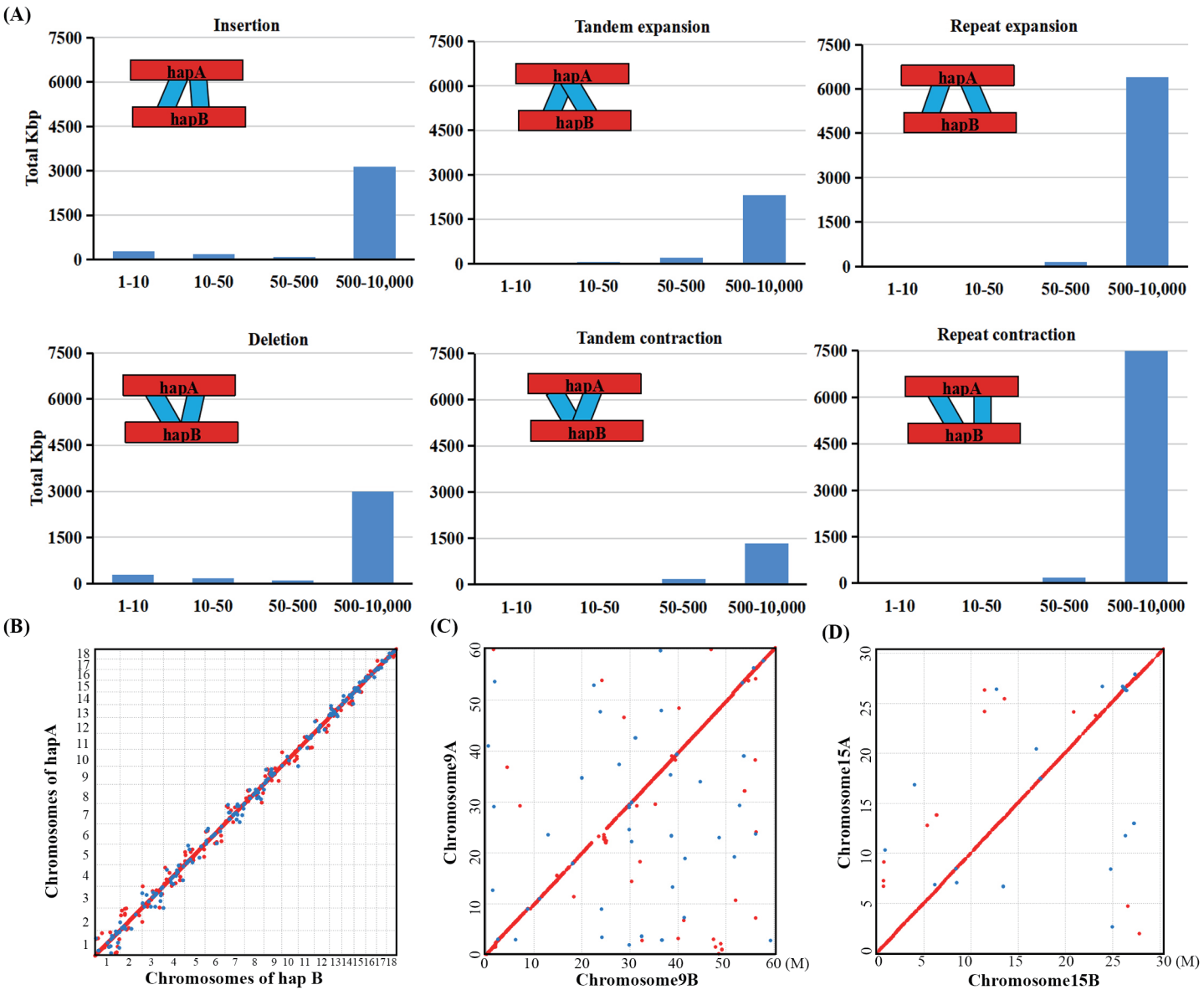
990 **FIGURE 4** Analyses of gene expression and gene distance of predicted secretome of *P. polysora*.
991 (A) Clustering analysis of gene expression of secretome on two haplotypes. Heatmaps show rlo-
992 transformed expression values. Cluster numbers are shown outside the graphs, and tracks
993 represent gene expression in germinated spores (GS), and infected tissues at 1 dpi, 2 dpi, 4 dpi, 7
994 dpi, 10 dpi and 14 dpi. (B) Hexplots for closest-neighbor gene distance density of *P. polysora* in
995 haplotype A (left) and haplotype B (right). Circle dots represent two gene categories, BUSCOs
996 (yellow) and predicted candidate effectors (red). The black dots represent the distribution of the
997 avirulent gene, *AvrRppC.* (C) Gene and repeat density as well as log2TPM (transcripts per

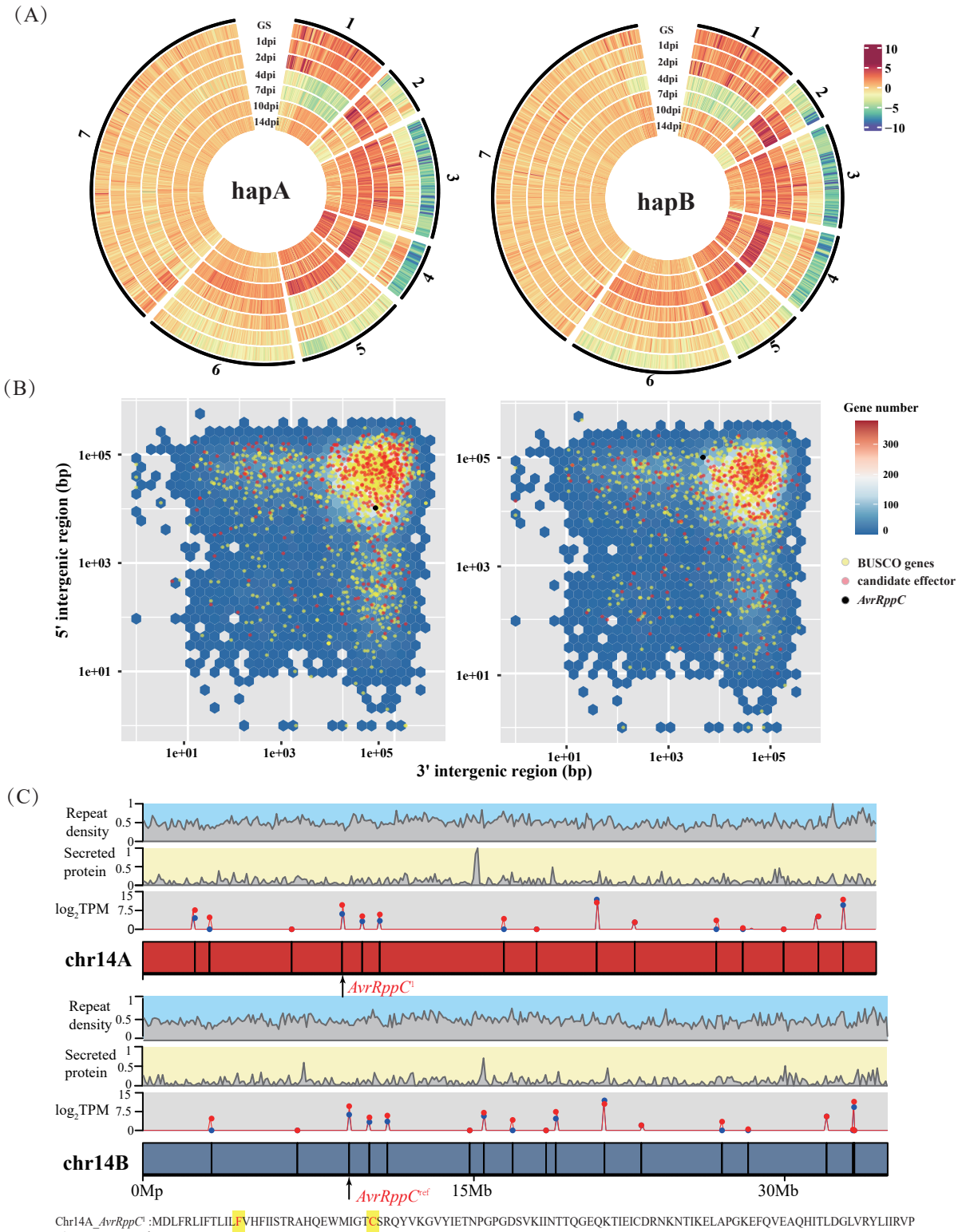
998 million) of candidate effectors from chromosomes 14A and 14B. The red and blue points
999 represent expression values of candidate effectors at 1 dpi and 7dpi. Positions and protein
1000 sequences of *AvrRppC* genes are presented.
1001 **FIGURE 5** Functional annotation of CAZymes, transcription factors and Merops proteases in
1002 *Ppz*-GD1913 isolate. (A) CAZyme families comparison of *P. polysora* (*Ppz,* GD1913), *P.*
1003 *triticina* (*Ptt*, *Pt76* isolate), *P. graminis* (*Pgt*, *Pgt*21-0 isolate) and *P. coronata* (*Pca*, *Pca*203).
1004 Heat maps showing gene numbers annotated in the following classes: auxiliary activities (AA),
1005 carbohydrate-binding modules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH),
1006 glycosyltransferases (GTs), and polysaccharide lyases (PL). (B) Percentages of genes from *P.*
1007 *polysora* are predicted to encode members of various fungal transcription factor classes based on
1008 InterProScan and Eggnog annotation. (C) Merops families comparison of *Ppz*, *Ptt*, *Pgt* and *Pca*.
1009 Heat map showing gene numbers annotated in the classes of aspartic acid (A), cysteine(C),
1010 metalloprotease (M), serine protease (S), and threonine protease (T) or peptidase inhibitors (I).
1011 **FIGURE 6** Comparative analyses of macro-synteny and orthogroups among *Puccinia polysora* f.
1012 sp. *zeae* (GD1913) and its relatives, *Austropuccinia psidii* (Au_3), *P. triticina* f. sp. *tritic*i (*Pt*76),
1013 *P. graminis* f.sp. *tritici* (*Pgt*21-0) and *P. coronata* f.sp *avenae* (*Pca*203). (A) Macro-synteny of
1014 four species in *Puccinia* with available haplotype-phased and chromosome-scale genome. Only
1015 the chromosomes from haplotype A/primary haplotype were compared. Grey lines are
1016 homologous proteins identified by MCscan with default parameters and black lines present the
1017 genome inversions between species. (B) Numbers of duplicated (+) and lost (-) genes inferred
1018 from OrthoFinder. The typical symptoms caused by each species are listed accordingly. (C)
1019 Distribution of species-specific orthogroups with different gene copy numbers among four
1020 *Puccinia* species.
1021 **FIGURE 7** Population genomics analyses of *P. polysora* from China. (A) Geographic
1022 distribution of 79 *P. polysora* isolates collected from north, central and south of China. (B)
1023 Principal component analysis of Chinese *P. polysora* population showing two genetic clusters.
1024 (C) Linkage disequilibrium test of Chinese *P. polysora* population. (D) *AvrRppC* genotypes of *P.*
1025 *polysora* isolates in minor (red) and major (blue) genetic groups.

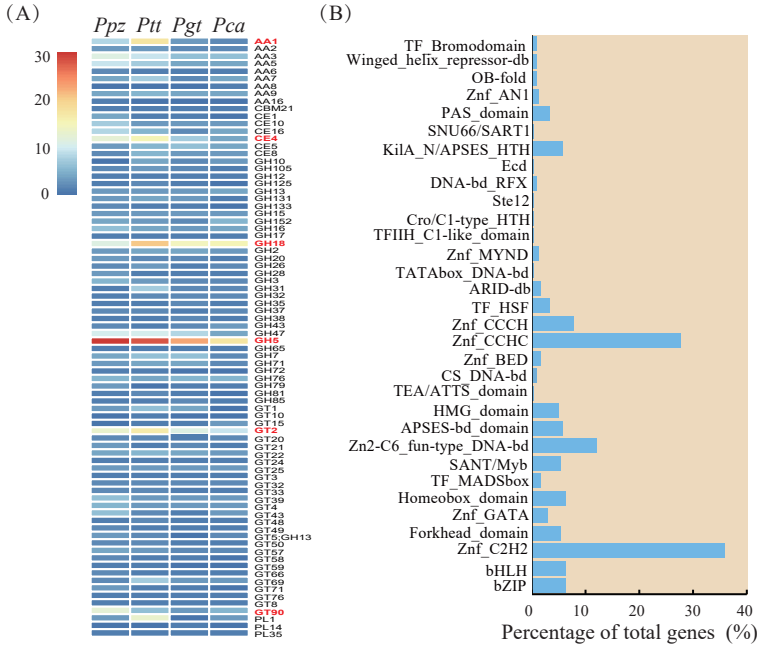
1026 **Supplementary files**

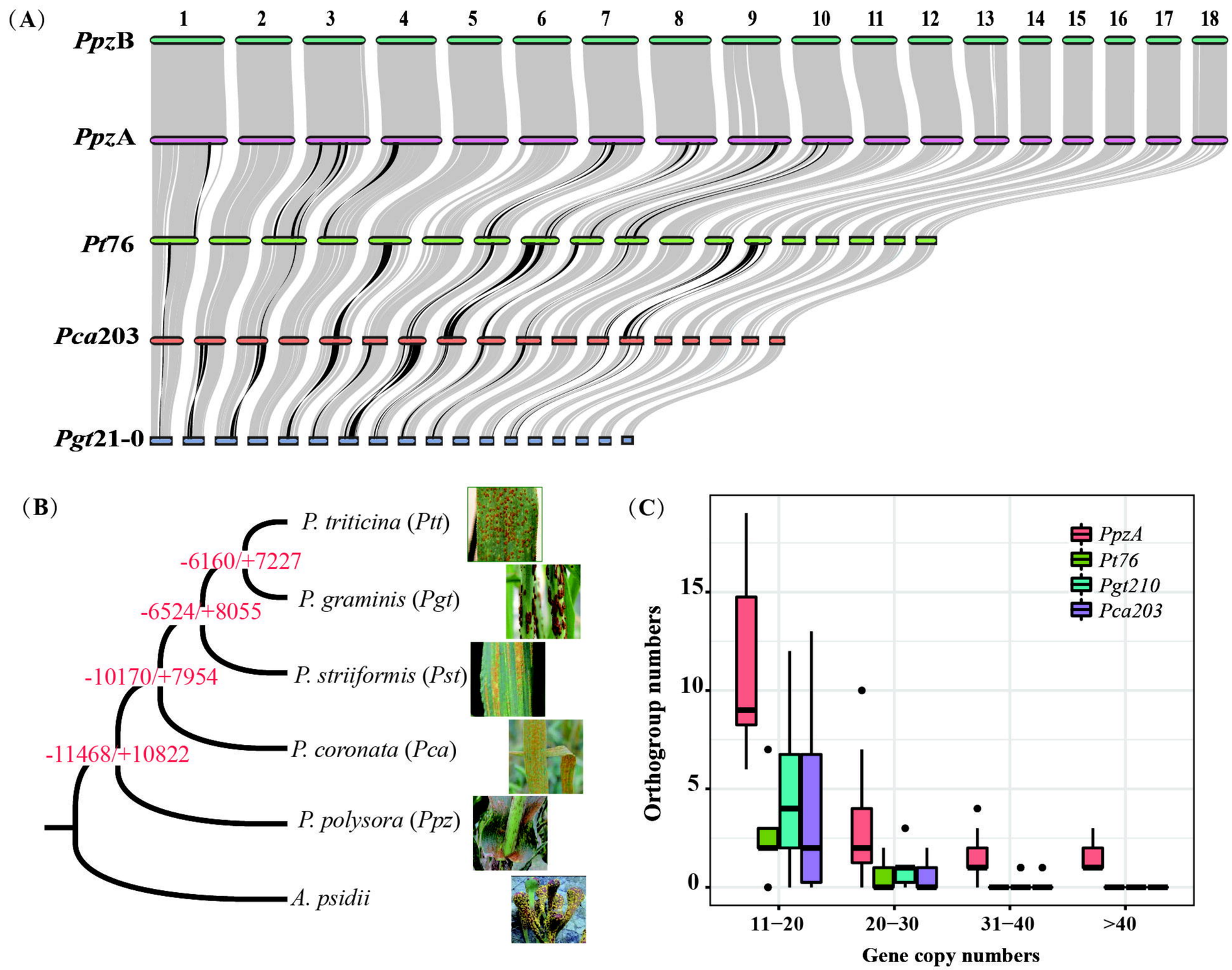
1027 Table S1 Information of 79 isolates from China used for genome resequencing.

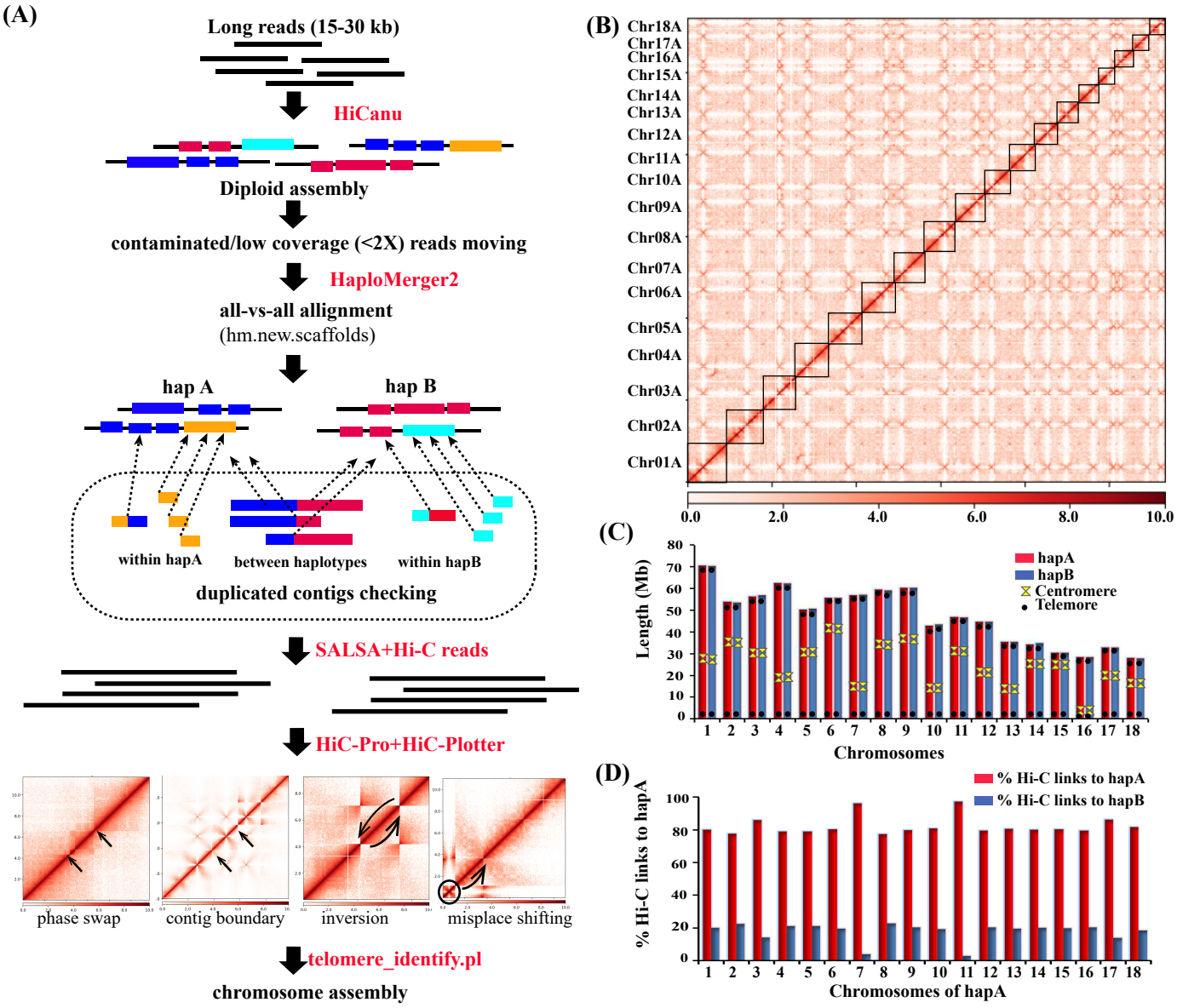
1028 Table S2 Contig information removed from the final assembly due to low-quality assessment.
1029 Table S3 Treatments for duplicate contigs between/within haplotype A and haplotype B
1030 Table S4 Sample information and data used for transcript analyses.
1031 Table S5 Functional annotation of specific orthologs between two haplotypes.
1032 Table S6 GO and IPR annotation for predicted candidate effectors.
1033 Table S7 Features of secreted proteins in different expression clusters of *P. polysora* f.sp. *zeae.*
1034 Table S8 Proteome comparison of four *Puccinia* species.
1035 Table S9 GO annotation of top 42 *P. polysora* specific orthogroups with high gene copy number.
1036 Table S10 Functional annotation of SNPs with *F*st > 0.9 between two genetic groups.
1037 Table S11 The CDS and amino acid sequences of *AvrRppC* gene.
1038 Table S12 Comparison of Gene number and gene length among three *Puccinia* species.
1039 Table S13 NMDC numbers for sequencing raw data used in this study.
1040 Figure S1. GO enrichment analysis of secreted proteins on two haplotypes.
1041 Figure S2 Pair-wise synteny of *P. polysora* and three close relatives.
1042 Figure S3 The bi-allele distribution of 79 isolates used for population genomics. Two isolates
1043 (GD1922-3 and GX1905-2) in red squares are removed from the final analysis due to abnormal
1044 distribution.

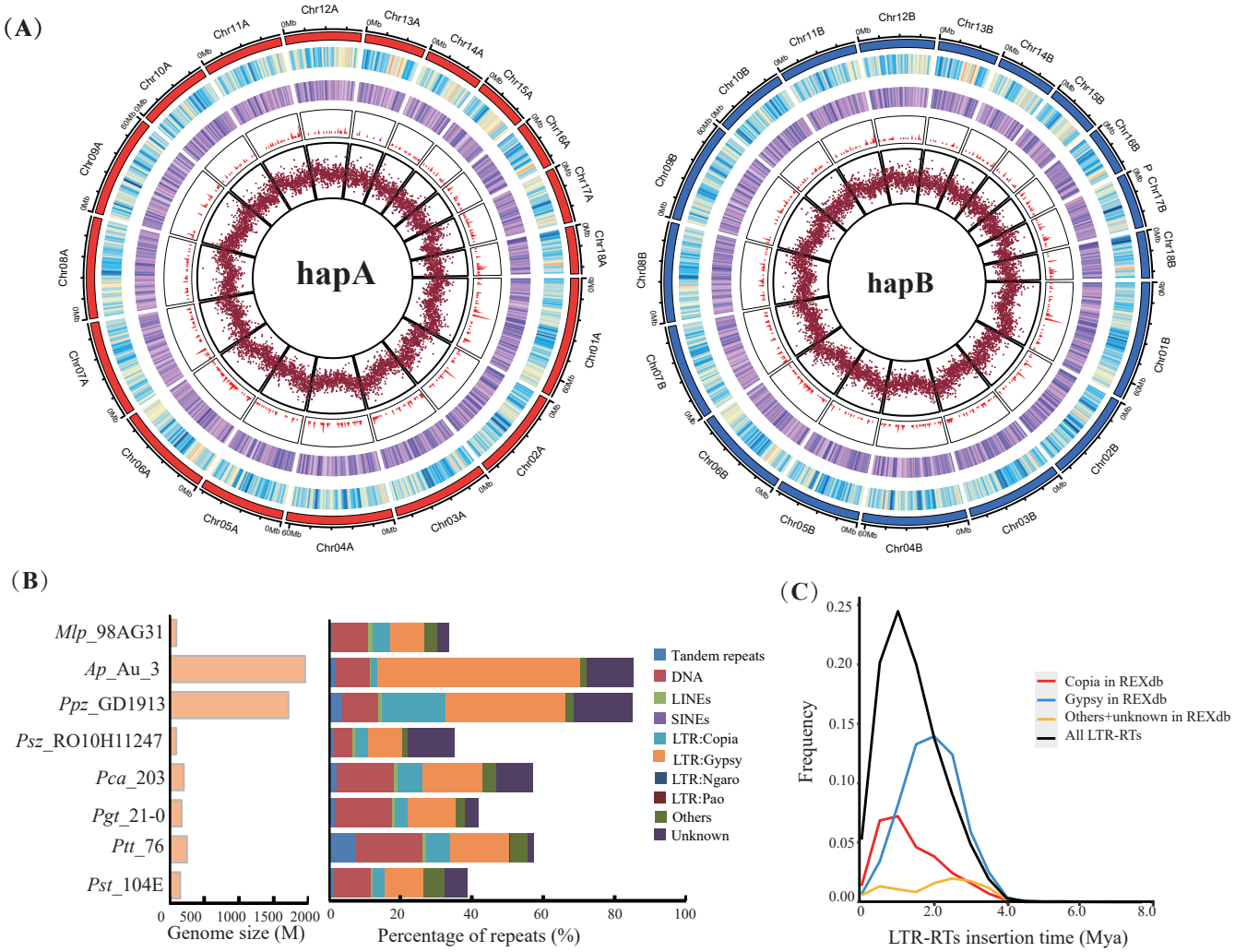

(A)
Insertion
Tandem expansion
Repeat expansion
Deletion
Tandem contraction
Repeat contraction
hapA
hapB
Total Kbp
7500
6000
4500
3000
1500
0
1-10
10-50
50-500
500-10,000
(B)
Chromosomes of hapA
Chromosomes of hap B
(C)
Chromosome9A
Chromosome9B
60 (M)
(D)
Chromosome15A
Chromosome15B
30 (M)

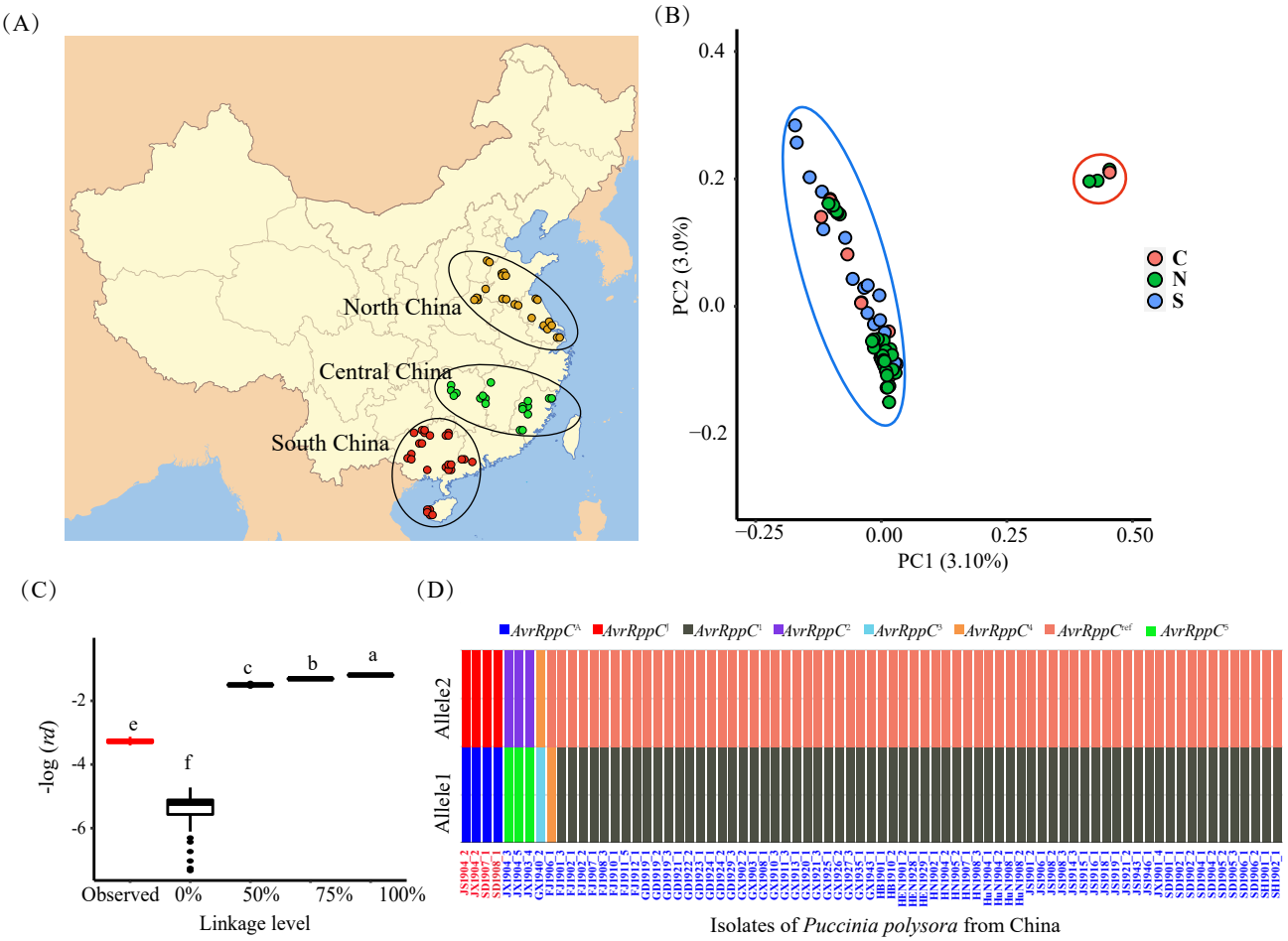

(A)
GS
1dpi
2dpi
4dpi
7dpi
10dpi
14dpi
hapA
hapB
1
2
3
4
5
6
7
10
5
0
−5
−10
(B)
5' intergenic region (bp)
3' intergenic region (bp)
1e+01
1e+03
1e+05
Gene number
300
200
100
0
BUSCO genes
candidate effector
AvrRppC
(C)
Repeat density
Secreted protein
log2TPM
chr14A
AvrRppC1
chr14B
AvrRppCref
0Mp
15Mb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

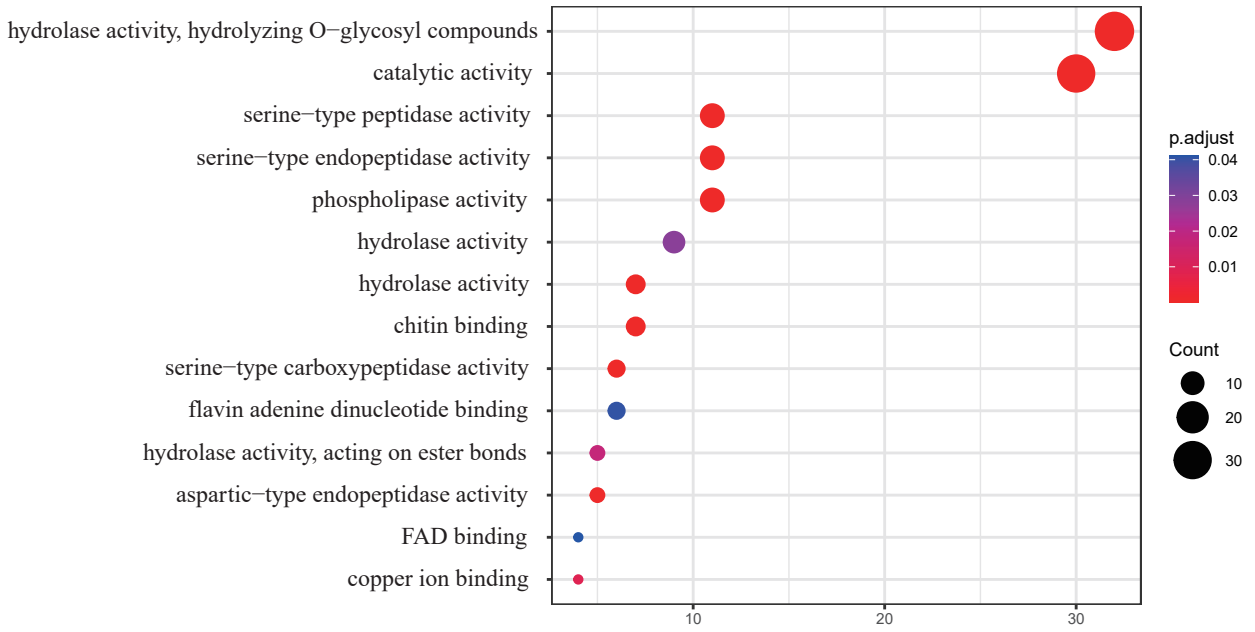

(A)
Ppz Ptt Pgt Pca
30
20
10
0
AA1
AA2
AA3
AA5
AA6
AA7
AA8
AA9
AA16
CBM21
CE1
CE10
CE16
CE4
CE5
CE8
GH10
GH105
GH12
GH125
GH13
GH131
GH133
GH15
GH152
GH16
GH17
GH18
GH2
GH20
GH26
GH28
GH3
GH31
GH32
GH35
GH37
GH38
GH43
GH47
GH5
GH65
GH7
GH71
GH72
GH76
GH79
GH81
GH85
GT1
GT10
GT15
GT2
GT20
GT21
GT22
GT24
GT25
GT3
GT32
GT33
GT39
GT4
GT43
GT48
GT49
GT5;GH13
GT60
GT57
GT58
GT59
GT66
GT69
GT71
GT76
GT8
GT90
PL1
PL14
PL35
(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
A02A
A11A
A11B
A22B
A28A
C01B
C02A
C115
C12
C13
C14B
C15
C19
C26
C44
C48
C50
C54
C56
C76B
C83
C85
C85A
C85B
C86
C97
I09
I32
I51
M01
M03A
M12B
M16A
M16B
M16C
M17
M18
M20D
M20F
M24A
M24B
M24X
N28A
M28E
M28X
M36
M38
M41
M48A
M48C
M49
M50B
M07A
M67C
M67X
M76
M79
M80
S08A
S08B
S09A
S09B
S09C
S09X
S10
S14
S16
S26A
S26B
S33
S49A
S53
S54
S59
T01A
T01X
T03
T05

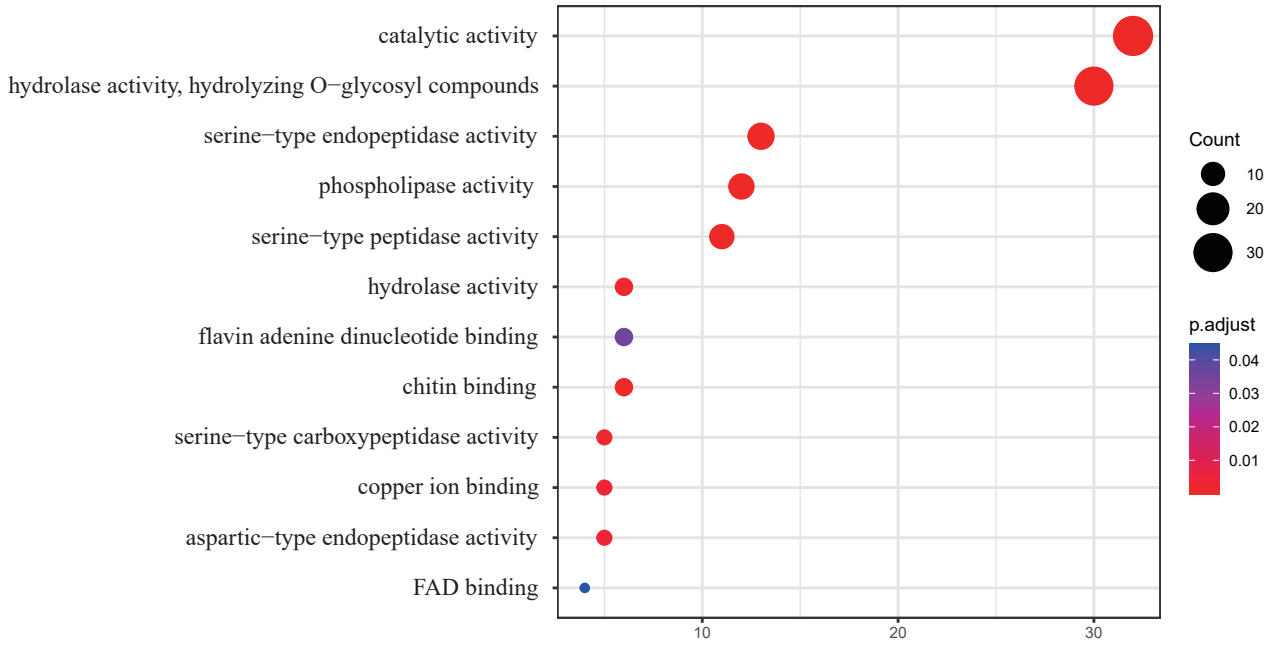

(A)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
PpzB
PpzA
Pt76
Pca203
Pgt21-0
(B)
P. triticina (Ptt)
P. graminis (Pgt)
P. striiformis (Pst)
P. coronata (Pca)
P. polysora (Ppz)
A. psidii
-6160/+7227
-6524/+8055
-10170/+7954
-11468/+10822
(C)
PpzA
Pt76
Pgt210
Pca203
Orthogroup numbers
0
5
10
15
11−20
20−30
31−40
>40
Gene copy numbers

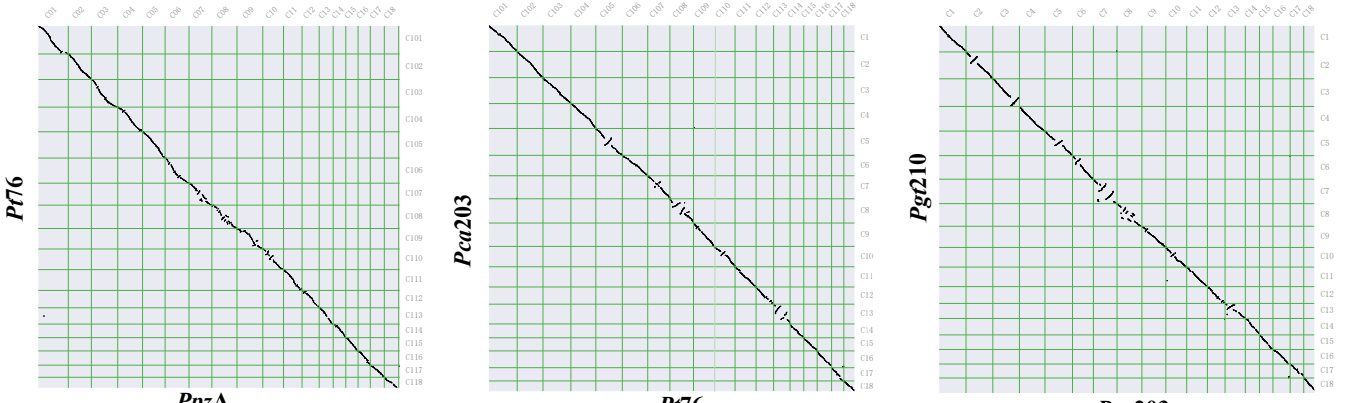

(A)
Long reads (15-30 kb)
HiCanu
Diploid assembly
contaminated/low coverage (<2X) reads moving
HaploMerger2
all-vs-all allignment
(hm.new.scaffolds)
hap A
hap B
within hapA
between haplotypes
within hapB
duplicated contigs checking
SALSA+Hi-C reads
HiC-Pro+HiC-Plotter
phase swap
contig boundary
inversion
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)
hapA
hapB
Chr01A
Chr02A
Chr03A
Chr04A
Chr05A
Chr06A
Chr07A
Chr08A
Chr09A
Chr10A
Chr11A
Chr12A
Chr13A
Chr14A
Chr15A
Chr16A
Chr17A
Chr18A
Chr01B
Chr02B
Chr03B
Chr04B
Chr05B
Chr06B
Chr07B
Chr08B
Chr09B
Chr10B
Chr11B
Chr12B
Chr13B
Chr14B
Chr15B
Chr16B
Chr17B
Chr18B
(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)
North China
Central China
South China
(B)
0.4
0.2
0.0
−0.2
PC2 (3.0%)
−0.25
0.00
0.25
0.50
PC1 (3.10%)
C
N
S
(C)
-log (rd)
-2
-4
-6
e
f
c
b
a
Observed
0%
50%
75%
100%
Linkage level
(D)
AvrRppC^A
AvrRppC^J
AvrRppC^1
AvrRppC^2
AvrRppC^3
AvrRppC^4
AvrRppC^ref
AvrRppC^5
Allele2
Allele1
Isolates of Puccinia polysora from China


hydrolase activity, hydrolyzing O−glycosyl compounds
catalytic activity
serine−type peptidase activity
serine−type endopeptidase activity
phospholipase activity
hydrolase activity
hydrolase activity
chitin binding
serine−type carboxypeptidase activity
flavin adenine dinucleotide binding
hydrolase activity, acting on ester bonds
aspartic−type endopeptidase activity
FAD binding
copper ion binding
10
20
30
p.adjust
0.04
0.03
0.02
0.01
Count
10
20
30


catalytic activity
hydrolase activity, hydrolyzing O−glycosyl compounds
serine−type endopeptidase activity
phospholipase activity
serine−type peptidase activity
hydrolase activity
flavin adenine dinucleotide binding
chitin binding
serine−type carboxypeptidase activity
copper ion binding
aspartic−type endopeptidase activity
FAD binding
10
20
30
Count
10
20
30
p.adjust
0.04
0.03
0.02
0.01


Pt76
PpzA
4347 gene pairs
Pca203
Pt76
4971 gene pairs
Pgt210
Pca203
5560 gene pairs

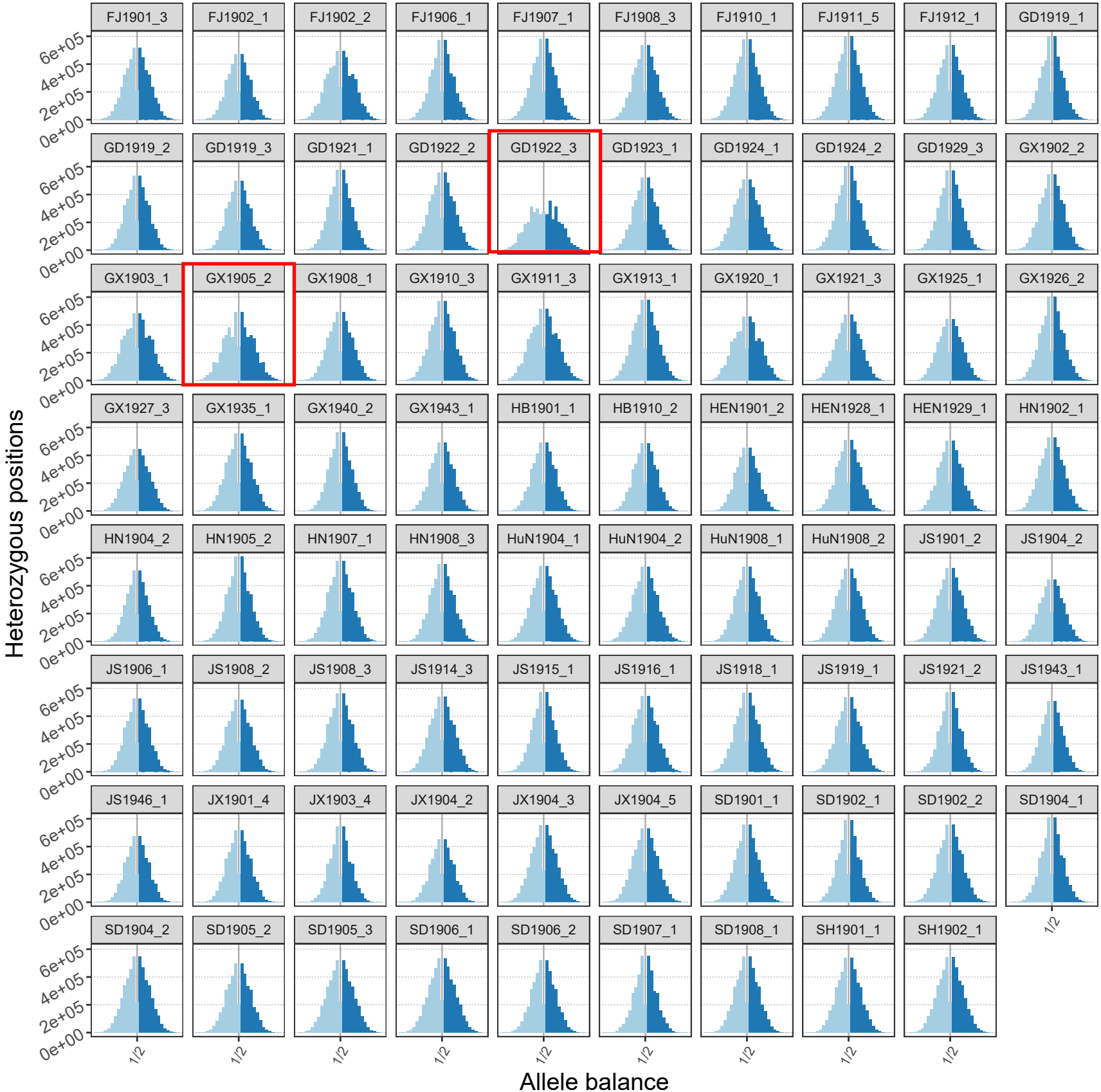

FJ1901_3
FJ1902_1
FJ1902_2
FJ1906_1
FJ1907_1
FJ1908_3
FJ1910_1
FJ1911_5
FJ1912_1
GD1919_1
GD1919_2
GD1919_3
GD1921_1
GD1922_2
GD1922_3
GD1923_1
GD1924_1
GD1924_2
GD1929_3
GX1902_2
GX1903_1
GX1905_2
GX1908_1
GX1910_3
GX1911_3
GX1913_1
GX1920_1
GX1921_3
GX1925_1
GX1926_2
GX1927_3
GX1935_1
GX1940_2
GX1943_1
HB1901_1
HB1910_2
HEN1901_2
HEN1928_1
HEN1929_1
HN1902_1
HN1904_2
HN1905_2
HN1907_1
HN1908_3
HuN1904_1
HuN1904_2
HuN1908_1
HuN1908_2
JS1901_2
JS1904_2
JS1906_1
JS1908_2
JS1908_3
JS1914_3
JS1915_1
JS1916_1
JS1918_1
JS1919_1
JS1921_2
JS1943_1
JS1946_1
JX1901_4
JX1903_4
JX1904_2
JX1904_3
JX1904_5
SD1901_1
SD1902_1
SD1902_2
SD1904_1
SD1904_2
SD1905_2
SD1905_3
SD1906_1
SD1906_2
SD1907_1
SD1908_1
SH1901_1
SH1902_1
0e+00
2e+05
4e+05
6e+05
1/2
Heterozygous positions
Allele balance

Table 1 Assembly information and completeness evaluation of *Puccinia polysora*

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	40.10%	40.00%
Complete BUSCOs	89.70%	89.70%
Complete and Single-copy BUSCOs	88.60%	88.60%
Complete and Duplicated BUSCOs	1.10%	1.10%
Fragmented BUSCOs	3.50%	3.50%
Missing BUSCOs	6.80%	6.80%

Table 2 Gene, ortholog content, and effectors in *Puccinia polyosra* genome assemblies

	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