

UC Berkeley

UC Berkeley Previously Published Works

Title

Analysis of the genetic diversity of the soybean rust pathogen *Phakopsora pachyrhizi* reveals two major evolutionary lineages

Permalink

<https://escholarship.org/uc/item/05b288k8>

Journal

Fungal Genetics and Biology, 179

ISSN

1087-1845

Authors

Rocha, Vinicius Delgado da
Ferreira, Everton Geraldo Capote
Castanho, Fernanda Machado
et al.

Publication Date

2025-06-01

DOI

10.1016/j.fgb.2025.103990

Peer reviewed

The genetic diversity of the soybean rust pathogen *Phakopsora pachyrhizi* has been driven by two major evolutionary lineages

Vinicius Delgado da Rocha¹, Everton Geraldo Capote Ferreira^{1,2,3}, Fernanda Machado Castanho¹, Marcia Kamogae Kuwahara¹, Cláudia Vieira Godoy¹, Maurício Conrado Meyer¹, Kerry F. Pedley⁴, Ralf T. Voegelé⁵, Anna Lipzen⁶, Kerrie Barry⁷, Igor V. Grigoriev^{6,7}, Marco Loehrer⁸, Ulrich Schaffrath⁸, Catherine Sirven⁹, Sebastien Duplessis¹⁰, Francismar Corrêa Marcelino-Guimarães¹

¹Brazilian Agricultural Research Corporation - National Soybean Research Center (Embrapa Soja), Paraná, Brazil.

²Blades, Evanston, Illinois, USA.

³The Sainsbury Laboratory, University of East Anglia, Norwich, UK.

⁴United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Foreign Disease-Weed Science Research Unit, Ft. Detrick, Maryland, U.S.A.

⁵Department of Phytopathology, Institute of Phytomedicine, Faculty of Agricultural Sciences, University of Hohenheim, 70599 Stuttgart, Germany

⁶United States Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A.

1 26 ⁷Department of Plant and Microbial Biology, University of California Berkeley,
2
3
4 27 Berkeley, California, U.S.A.

5
6 28

7
8 29 ⁸Department of Plant Physiology, RWTH Aachen University, Aachen, Germany
9

10
11 30

12
13 31 ⁹Bayer SAS, Crop Science Division, Lyon, France
14

15
16 32

17
18 33 ¹⁰Université de Lorraine, INRAE, IAM, Nancy, F-54000, France
19

20 34 ORCID <https://orcid.org/0000-0002-2072-2989>
21
22

23 35

24
25 36 **Disclaimer** – Mention of trade names or commercial products in this publication is
26
27 37 solely for the purpose of providing specific information and does not imply
28
29 38 recommendation or endorsement by the U.S. Department of Agriculture. The USDA
30
31 39 is an equal opportunity provider and employer.
32
33
34

35 40

36
37 41 Corresponding author: Francismar Corrêa Marcelino-Guimarães
38
39

40 42 E-mail: francismar.marcelino@embrapa.br
41
42

43 43

44 44

45
46
47 45

48
49
50 46

51
52 47

53
54 48

55
56
57 49

58
59 50
60
61
62
63
64
65

51 **ABSTRACT**

52 *Phakopsora pachyrhizi*, an obligate biotrophic rust fungus, is the causal agent of
53 Asian Soybean Rust (ASR) disease. Here, we utilized whole-genome data to explore
54 the evolutionary patterns and population structure across 45 *P. pachyrhizi* isolates
55 collected from 1972 to 2017 from diverse geographic regions worldwide. We also
56 characterized *in-silico* mating-type (*MAT*) genes of *P. pachyrhizi*, in the predicted
57 proteome of three isolates, to investigate the sexual compatibility system. Our
58 molecular phylogenetic analysis in *P. pachyrhizi* inferred two distinct evolutionary
59 lineages structured on a temporal scale, with lineage Pp1 grouping isolates obtained
60 from 1972 to 1994, while more recently collected isolates formed a second lineage,
61 Pp2. We found high levels of genetic diversity in lineage Pp1, whereas lineage Pp2
62 exhibited a strong clonal genetic structure, with a significant lower diversity. The
63 widespread propagation of *P. pachyrhizi* clonal spores across soybean-growing
64 regions likely explains the absence of a large-scale spatial genetic structure within
65 each lineage. Two independent isolates (TW72-1 and AU79-1) showed moderate
66 levels of genetic admixture, suggesting potential somatic hybridization between the
67 two *P. pachyrhizi* lineages. We observed no clear congruence between virulence
68 levels of *P. pachyrhizi* isolates and their phylogenetic patterns. Our findings support a
69 probable tetrapolar mating system in *P. pachyrhizi*. Taken together, our study offers
70 new insights into the evolutionary history of *P. pachyrhizi* and demonstrates that
71 multiple *MAT* genes are highly expressed during the later stages of soybean
72 infection, suggesting their potential role in the formation of urediniospores within the
73 life cycle of *P. pachyrhizi*.

74 **Keywords:** genetic admixture, mating-type genes, plant-pathogens, population
75 genomics, virulence

76 INTRODUCTION

77 The rust fungus *Phakopsora pachyrhizi* (Phakopsoraceae) is an obligate
78 biotrophic plant-pathogen known for its ability to infect around 150 plant species
79 within the family Fabaceae, including soybean (*Glycine max*) (Hartman et al., 2011).
80 The broad host range of *P. pachyrhizi* may play a crucial role in its successful
81 adaptability, serving as an inoculum reservoir for survival and dissemination of the
82 pathogen on other crops and between planting seasons when the primary host (i.e.
83 soybean) is not available (Jarvie, 2009). Notably, *P. pachyrhizi* is the causal agent of
84 Asian Soybean Rust (ASR), one of the most devastating foliar diseases to soybean
85 production, especially in South America, where favorable environmental conditions
86 contribute to the development of the disease. Without adequate management, ASR
87 can lead to severe yield losses, reaching up to 80% (Hartman et al., 2016; Beruski et
88 al., 2020). ASR symptoms consist of grayish-brown lesions that form predominantly
89 on the abaxial surface of leaves, prompting intense premature leaf senescence
90 (Hossain et al., 2024). The control and management of ASR is particularly
91 challenging due to the high variability in virulence levels of *P. pachyrhizi*, with the
92 potential occurrence of multiple races (Yamaoka et al., 2002; Castanho, 2022).
93 Although seven loci (*Rpp1-Rpp7*) conferring resistance to ASR have been described
94 in soybeans, *P. pachyrhizi* has rapidly evolved to overcome individual resistance
95 genes (Chicowski et al., 2024).

96 *Phakopsora pachyrhizi* likely originated on the Asian continent and
97 subsequently spread to other worldwide soybean-growing regions, including Africa
98 and the Americas (Hossain et al., 2024). The first documented occurrence of the
99 pathogen dates back to 1902 in Japan, where it was initially identified as *Uredo sojae*
100 (Kitani and Inoue, 1960). Since then, the presence of *P. pachyrhizi* has been

1 101 observed in several geographically distinct regions. The pathogen was reported in
2
3 102 India in 1970 (Sarbhoy et al., 1972), in Africa during the 1990s (Akinsanmi et al.,
4
5 103 2001; Levy et al., 2005), South America in 2001 (Yorinori et al., 2005), and North
6
7 104 America in 2004 (Schneider et al., 2005).
8
9

10
11 105 The global dispersion of *P. pachyrhizi* has been associated with
12
13 106 intercontinental dispersals of airborne spores (Schneider et al., 2005; Jorge et al.,
14
15 107 2015). Indeed, evidence-based studies suggest that transatlantic air currents may
16
17 108 have promoted the widespread dissemination of *P. pachyrhizi* spores from Africa to
18
19 109 Brazil (Freire et al., 2008). The occurrence of long-distance dispersal events for *P.*
20
21 110 *pachyrhizi* has shaped its patterns of genetic diversity and population structure,
22
23 111 leading to intense gene flow among spatially separated populations (Freire et al.,
24
25 112 2008; Twizeyimana et al., 2011; Jorge et al., 2015). Previous investigations based on
26
27 113 molecular markers, such as microsatellites (Twizeyimana et al., 2011), AFLPs
28
29 114 (Amplified Fragment Length Polymorphism) (Darben et al., 2020), ITS (Internal
30
31 115 transcribed spacer), and ARF (ADP-Ribosylation Factor) gene sequences (Zhang et
32
33 116 al., 2012), have demonstrated that *P. pachyrhizi* isolates from various localities
34
35 117 exhibited low levels of genetic differentiation across extensive geographic areas in
36
37 118 Africa, North America, and South America, implying a predominantly clonal
38
39 119 population structure. Similarly, comparisons among three *P. pachyrhizi* isolates from
40
41 120 South America (K8108, MT2006, and UFV02), based on single nucleotide variations
42
43 121 distributed across their genomes, demonstrated that they potentially represent a
44
45 122 single clonal lineage with high levels of heterozygosity (Gupta et al, 2023).
46
47
48
49
50
51
52
53

54 123 Whether sexual recombination contributes to the diversity of *P. pachyrhizi*
55
56 124 remains an open question, as its sexual cycle is completely unknown. *P. pachyrhizi*
57
58
59 125 reproduces primarily, if not exclusively, via asexual urediniospores that harbor two
60
61
62
63
64
65

1 126 separate haploid nuclei (N + N) (Bromfield, 1984). Teliospores are the only other
2
3
4 127 spore type that has been identified in *P. pachyrhizi*; however, their germination has
5
6 128 never been observed in nature (Bromfield, 1984; Goellner et al., 2010).
7

8 129 The sexual compatibility patterns of *P. pachyrhizi*, particularly involving
9
10 130 mating-type (MAT) genes, remain to be elucidated. Conversely, comparative
11
12 131 genomic analyses have identified MAT genes in other rust fungi, including
13
14 132 *Melampsora larici-populina* (Duplessis et al., 2011a), various *Puccinia* species
15
16 133 (Duplessis et al., 2011; Cuomo et al., 2017; Luo et al., 2024), and *Austropuccinia*
17
18 134 *psidii* (Ferrarezi et al., 2022). In general, available genomes of plant-pathogenic rust
19
20 135 fungi, (e.g. *Puccinia* species), show a tetrapolar compatibility system, with MAT
21
22 136 genes organized into two loci — P/R and HD locus — located on distinct
23
24 137 chromosomes (Luo et al., 2024). Specifically, the P/R locus contains genes-encoding
25
26 138 pheromone peptide precursors (MFA), as well as multiple genes that encode
27
28 139 pheromone receptor proteins known as STE3. The HD locus carries alternative
29
30 140 alleles that encode two divergent homeodomain transcription factors (HD1 and HD2)
31
32 141 (Coelho et al., 2017). Within the Basidiomycota, the phylum in which *P. pachyrhizi* is
33
34 142 included, sexual mating compatibility involves molecular recognition between mature
35
36 143 pheromone peptides and their corresponding transmembrane G-protein coupled
37
38 144 receptors (Coelho et al., 2017). This recognition triggers the fusion of compatible
39
40 145 gametes and subsequently leads to the formation of a heterodimeric complex
41
42 146 composed of HD1 and HD2 homeodomain transcription factors (Coelho et al., 2017;
43
44 147 Cuomo et al., 2017). The HD1-HD2 heterodimer acts as a transcriptional regulator. It
45
46 148 binds to promoter sequences and enhances the expression of specific genes, thus
47
48 149 promoting sexual development (Vonk and Ohm et al., 2018).
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 150 The first genome assemblies of *P. pachyrhizi* have been recently published
2
3
4 151 (Gupta et al., 2023). By combining two different sequencing platforms — short-read
5
6 152 data and long-read data — robust bioinformatics tools constructed *de novo* genome
7
8 153 assemblies of three distinct isolates of *P. pachyrhizi* (K8108, MT2006, and UFV02)
9
10 154 (Gupta et al., 2023). These genomic resources can be used to gain insight into the
11
12 155 evolutionary landscape of the pathogen over spatial and temporal dimensions.
13
14

15 156 In this study, we elucidated the evolutionary patterns and population structure
16
17 157 of *P. pachyrhizi*. To achieve this objective, we utilized the Illumina sequencing
18
19 158 platform to re-sequence whole-genomes of 42 diverse isolates of *P. pachyrhizi*. In
20
21 159 addition, we compiled genomic data for three previously investigated isolates: K8108,
22
23 160 MT2006, and UFV02 (Gupta et al., 2023). Our collection of isolates was acquired
24
25 161 within a wide timeframe (1972-2017) and covers a global geographic range across
26
27 162 various soybean-growing areas, comprising Africa, Asia, Australia, South America,
28
29 163 and the United States of America (USA). We identified highly-curated SNPs at the
30
31 164 whole-genome level by mapping the *P. pachyrhizi* isolates to the UFV02 reference
32
33 165 genome. Next, we implemented phylogenetic reconstructions and population
34
35 166 genomic approaches to determine whether multiple evolutionary lineages have
36
37 167 influenced the distribution patterns of genetic diversity among *P. pachyrhizi* isolates
38
39 168 over the course of time. To explore whether distinct evolutionary lineages of *P.*
40
41 169 *pachyrhizi* may potentially outcross through sexual compatibility, we performed *in-*
42
43 170 *silico* functional annotation of MAT genes of *P. pachyrhizi*, and their expression
44
45 171 profiles were investigated during the interaction between the pathogen and soybean
46
47 172 host. We also carried out phenotyping experiments to assess the virulence levels of
48
49 173 *P. pachyrhizi* isolates on differential soybean genotypes containing seven loci that
50
51 174 confer resistance to *P. pachyrhizi* (*Rpp1* to *Rpp7*). Our study enables us to expand
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 175 our current understanding of how evolutionary processes have shaped the genomic
2
3
4 176 diversity of *P. pachyrhizi*.

5
6 177
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

MATERIAL AND METHODS

179

Fungal isolate sampling

Our study encompassed 37 *P. pachyrhizi* isolates, which were deposited in the Collection of Microorganisms of the Embrapa Soybean (CMES), Londrina, Paraná, Brazil. The isolates were sampled over 15 years (2003-2017) and obtained from soybean leaves collected in four geographically distant regions, including Africa, Asia, South America, and the USA (Supplementary Table S1). Each isolate was subjected to single-spore purification on detached soybean leaflets. In summary, a drop of water was carefully placed over a single uredinium. Single spores were isolated via a Zeiss Axio Scope A1 microscope. A single spore was transferred to healthy leaflets of the susceptible genotype Embrapa 48, kept in a Petri dish containing 1% agar-water medium, and then maintained in a growth chamber, at a temperature of approximately 21°C and a humidity level of 60% for approximately 15 days for the production of new lesions. To guarantee the purity of the isolates, three cycles of single spore isolation were carried out. Detached leaflets of the same susceptible genotypes were used for spore multiplication, and the spores were transferred to a 2 mL microtube kept open in a desiccator for 24 hours for dehydration and were then stored in a freezer at -80°C for preservation.

Additionally, we also included five *P. pachyrhizi* isolates representing early strains that were originally collected between the years 1972 and 2004 (Supplementary Table S1). These five isolates [Taiwan-1972 (TW72-1), India-1973 (IN73-1), Australia-1979 (AU79-1), Hawaii-1994 (HW94-1), and Colombia-2004 (CO04-2)] were maintained at the United States Department of Agriculture (USDA), Foreign Disease-Weed Science Research Unit (FDWSRU), Frederick, MD, USA.

203

204 **DNA extraction and whole-genome sequencing**

205 Total genomic DNA was extracted from 2-3 mg of spore per isolate, following
206 an existing protocol with adaptations (Dellaporta et al., 1983). DNA concentration
207 was quantified by a Nanodrop 8000 spectrophotometer (Thermo Scientific,
208 Wilmington, DE). Whole-genome sequencing of each isolate was performed at DOE
209 Joint Genome Institute (JGI), using paired-end sequencing (2 x 151 bp) on an
210 Illumina NovaSeq S4 platform. To evaluate the quality of raw sequence data, we
211 employed a JGI QC pipeline to remove potential contaminants, adapters, and low-
212 quality sequences.

213

214 **Mapping and variant calling**

215 Both read-mapping and variant detection were conducted according to the
216 procedures developed by Gupta et al., (2023). Illumina sequence data from each
217 isolate were mapped to the *P. pachyrhizi* UFV02 v2.1 reference genome (Gupta et
218 al., 2023) using the BWA-MEM algorithm with options -M-R in BWA v. 0.7.17
219 software (Li and Durbin, 2009). The SAMtools v1.9 software (Li et al., 2009) was
220 employed to convert alignment files into compressed BAM format. Subsequently,
221 bam files were input into the Picard package (<https://broadinstitute.github.io/picard/>)
222 to remove duplicate reads. Local realignments of sequencing-reads around indel
223 regions were performed using two algorithms (RealignerTargetCreator and
224 IndelRealigner) implemented in the GATK v3.8.1 software (McKenna et al., 2010).

225 SNPs and indels were identified using the HaplotypeCaller algorithm in GATK
226 software, producing a single VCF file for 42 *P. pachyrhizi* isolates. To ensure high-
227 quality variants, the resulting VCF file was filtered in GATK v.4.2 software (McKenna

et al., 2010), with the following thresholds: QUAL < 30.00, MQ < 40.00, SOR > 3.00, QD < 2.00, FS > 60.00, MQRankSum < -12.500, ReadPosRankSum < -8.00. Utilizing VCFtools v0.1.16 software (Danecek et al., 2011), the VCF file underwent additional filtering parameters to remove indels and retain biallelic SNPs with a minimum depth of read coverage (minDP = 5) and minimum allele frequency (MAF ≥ 5%).

To incorporate additional genomic data into our study, we generated a SNP set based on publicly available genomic assemblies of three *P. pachyrhizi* isolates, K8108, MT2006, and UFV02 (Gupta et al., 2023). This analysis utilized the NUCmer algorithm implemented in MUMmer v3.23 (Kurtz et al., 2004), with options `--show-snps -Clr`. Using the “merge” function in the BCFtools software (Li, 2011), the predicted SNPs in the three isolates were concatenated with our VCF file, comprising the SNPs from 42 isolates. Finally, we obtained a large SNP panel containing 452,853 whole-genomic variants (MAF ≥ 5% and missing data ≤ 20%) across 45 *P. pachyrhizi* isolates. We referred to this SNP panel as Dataset A.

Annotation and prediction of SNP effects

We applied SnpEff v.5.1 software (Cingolani et al., 2012) to annotate and predict the putative effects of SNPs on coding-DNA regions in *P. pachyrhizi*. The UFV02 v2.1 reference genome and its gene annotation GFF file (Gupta et al., 2023) were utilized to build a SnpEff database through the “built” function and default parameters. Dataset A served as an input file for SnpEff.

Phylogenetic analyses and population structure

Maximum likelihood (ML) approaches reconstructed the phylogenetic patterns among 45 *P. pachyrhizi* isolates by analyzing Dataset A. Before achieving

1 253 phylogenetic analyses, the best-fit model of nucleotide evolution was inferred through
2
3 254 the ModelFinder2 software tool (Kalyaanamoorthy et al., 2017) implemented in IQ-
4
5 255 TREE v1.6.12 (Nguyen et al., 2015). The Bayesian Information Criterion (BIC) chose
6
7 256 TVMe as the best-fit model. We conducted an ML phylogeny in IQ-TREE software,
8
9
10 257 with ten independent runs and 1,000 ultrafast bootstrap replicates. The bootstrap
11
12 258 convergence criterion was 0.99 and followed the previous standards described by
13
14 259 Hoang et al., (2018). The resulting phylogenetic tree was edited using FigTree v1.2.4
15
16 260 (<http://tree.bio.ed.ac.uk/software/figtree/>).
17
18
19

20 261 To reduce linkage disequilibrium (LD) bias on the population structure
21
22 262 analyses, we selected randomly SNPs separated within a distance of 5 kbp utilizing
23
24 263 parameter --thin 5000 in VCFtools v0.1.16 (Danecek et al., 2011). The SNPs were
25
26 264 filtered with $MAF \geq 5\%$ and missing data $\leq 20\%$, generating a final VCF file
27
28 265 containing 115,065 whole-genomic SNPs. From this data, we assembled two distinct
29
30 266 datasets — Dataset B and Dataset C — to explore population structure in *P.*
31
32 267 *pachyrhizi*. Dataset B came from a full set of 45 isolates, while Dataset C included
33
34 268 exclusively 41 isolates belonging to lineage Pp2 uncovered during the preceding
35
36 269 phylogenetic analysis (Fig. 1A).
37
38
39
40
41

42 270 Population structure analyses were conducted with two independent methods:
43
44 271 principal component analysis (PCA) and Bayesian clustering. PCA was performed
45
46 272 using R v4.4 (<https://www.r-project.org/>), with the “glPca” function implemented in the
47
48 273 adegenet package v2.1.10 (Jombart, 2008). The 2D PCA plot was generated in the R
49
50 274 ggplot package v3.5.1(Wickham, 2016). The fastStructure v1.0 (Raj et al., 2014) was
51
52 275 utilized to determine Bayesian model-based clustering, ranging K from 1 to 10, with
53
54 276 10 independent replications for each K. The optimal number of K was chosen based
55
56 277 on a FastStructure chooseK.py script. The Q-matrix files from 10 replications,
57
58
59
60
61
62
63
64
65

representing membership probabilities for the optimal K, were merged and results were visualized in the R pophelper package v2.3.1 (Francis, 2017).

280

281 Genetic diversity analyses

282 To calculate the pairwise genetic differentiation index (F_{ST}) between two *P.*
283 *pachyrhizi* lineages identified in our initial phylogenetic inference (Fig. 1A), we
284 carefully divided Dataset A into two subsets, Dataset D and Dataset E, each
285 harboring 307,752 whole-genomic SNPs from distinct lineage P1 and P2,
286 respectively. For these datasets, we applied a new filter with $MAF \geq 5\%$ and sites
287 with missing data were discarded. The pairwise F_{ST} values (Weir and Cockerham,
288 1984) were estimated in VCFtools v0.1.16 (Danecek et al., 2011), with non-
289 overlapping sliding windows of 10 kbp.

290 To assess genetic diversity levels among isolates of each *P. pachyrhizi*
291 lineage, we calculated two independent measures: nucleotide diversity per site (π)
292 and Watterson's theta per site (θ_w). We also computed the Tajima's D test neutrality
293 to detect potential signals of selective pressure acting on *P. pachyrhizi* lineages.
294 These analyses were conducted in the R PopGenome package v2.7.5 (Pfeifer et al.,
295 2014) and were structured into two stages. First stage, we included Dataset D and
296 Dataset E containing SNPs distributed across the whole-genome. Second stage, we
297 included exclusively SNPs located in genes (both predicted effector and non-
298 effector), utilizing the annotated GFF3 file of the UFV02 isolate as a reference gene
299 annotation (Gupta et al., 2023). The classification of candidate effector genes in the
300 UFV02 isolate was based on its predicted secretome, with well-established criteria
301 (see Gupta et al., 2023). Specifically, we considered a total of 1,168 candidate

effector genes predicted in the UFV02 isolate, while 21, 299 remaining genes were categorized as non-effectors (Gupta et al., 2023).

The non-parametric Mann-Whitney-Wilcoxon test inferred statistical differences between the two *P. pachyrhizi* lineages for genetic diversity measures and Tajima's D values, utilizing R v4.4 (<https://www.r-project.org/>).

Identification of mating-type (*MAT*) genes

For *in-silico* characterization of mating-type (*MAT*) genes in *P. pachyrhizi*, we used the predicted proteomes from the three isolates K8108, MT2006, and UFV02 (Gupta et al., 2023). We ran local BLASTp searches (cutoff e-value < 1e-5) to predict putative *MAT* genes, encompassing pheromone receptors (*STE3.2*), pheromone precursors (*MFA*), and homeodomains (*HD*). As queries, BLASTp searches utilized *MAT* protein sequences previously identified in two closely related rust fungal species: *Puccinia triticina* and *Puccinia graminis* f. sp. *tritici* (Cuomo et al., 2017) (Supplementary Table S2). Furthermore, we also searched for *MAT* homologs through HMMER v3.3.2 software (<http://hmmer.org/>), with an e-value < 0.001, employing independent profile Hidden Markov Models for Pheromone A receptor (PF02076) and Homeobox KN domain (PF05920), downloaded from the Pfam database v33.0 (Mistry et al., 2021).

After identifying putative *MAT* genes in *P. pachyrhizi*, we examined their phylogenetic relationships among the isolates K8108, MT2006, and UFV02, and two *Puccinia* species were used as outgroups. For this purpose, we created three datasets, each consisting of protein sequences from a different *MAT* gene in *P. pachyrhizi*. The *MAT* protein sequences were aligned utilizing MAFFT v7.453 (Katoh and Standley, 2013) under the L-INS-i algorithm, with default parameters. For each

set of *MAT* genes in *P. pachyrhizi*, independent ML phylogenies were performed in IQ-TREE v1.6.12 (Nguyen et al., 2015), with ten separate runs and 1,000 ultrafast bootstrap replicates.

To determine whether the *MAT* genes of *P. pachyrhizi* were differentially expressed throughout the soybean infection cycle, we retrieved transcriptome data for the three *P. pachyrhizi* isolates—K8108, MT2006, and UFV02—available as Log2 fold change values of gene expression during different timepoints of spore germination and soybean infection (Gupta et al., 2023). Heatmap plots were constructed using the R ggplot2 package to visualize the expression profiles of *MAT* genes in *P. pachyrhizi*.

337

Virulence of *P. pachyrhizi* isolates on differential soybean genotypes

We performed comprehensive phenotyping evaluations to investigate the virulence profiles of seven *P. pachyrhizi* isolates (TW72-1, IN73-1, AU79-1, HW94-1, CO04-2, MT2006, and UFV02) from the two main evolutionary lineages recognized in our study (Fig. 1A). To characterize the pathogenicity of *P. pachyrhizi* isolates, we used eight differential soybean accessions, each carrying a single *Rpp* gene. The complete list of soybean plant introduction (PI) accessions used in this study can be found in Supplementary Table S3.

Phenotyping analyses for isolates MT2006 and UFV02 were carried out at Embrapa Soybean, Londrina, Brazil. In a greenhouse, we cultivated three plants per soybean accession until they reached the V3-V4 growth stage. The third trifoliolate of each plant was collected and kept in Petri dishes with 1% agar-water medium for subsequent inoculation. For each fungal isolate, we prepared a spore suspension (4×10^4 spores/mL) consisting of sterile distilled water and 0.01% Tween® 20. The

spore suspension was spread homogeneously on the abaxial side of leaves using a glass atomizer. The inoculated leaves were kept in a grow chamber with 60% relative humidity and 24°C for 14 days. After the incubation period, leaves were photographed and evaluated based on four previous quantitative parameters proposed by Yamanaka et al., 2010. The four parameters were as follows: (1) presence or absence of lesions, (2) lesion sporulation level (SL), (3) number of uredinia per lesion (NoU), and (4) frequency of lesions showing uredinia (%LU) (Supplementary Table S4). In each soybean accession, 3 biological replicates (one leaf per plant) were analyzed, with 30 lesions per leaf, totaling 90 lesions per soybean accession. Based on the quantitative parameters, four qualitative categories were established: immune (I), highly resistant (R), intermediate resistant (IM), and susceptible (S). As a control, we used the susceptible cultivar BRS 184.

For the five remaining isolates (AU79-1, CO04-2, HW94-1, IN73-1, and TW72-1), we conducted phenotyping assays at the FDWSRU. Soybean plants were grown in a growth chamber (Environmental Growth Chambers) at a constant temperature of 20°C with a 16-h photoperiod. All plants were fertilized 3 weeks after germination with Peters Professional 20-20-20. Plants were transferred to the USDA biological safety level 3 (BSL-3) plant pathogen containment greenhouse facility at Ft. Detrick and inoculated with *P. pachyrhizi* approximately 3 weeks after germination. Respective isolates were prepared as previously described (Kendrick et al., 2011). Spore concentrations were measured with a hemacytometer and adjusted to approximately 5×10^4 spores/mL in a solution of 0.01% Tween® 20 in sterile distilled water. Inoculum was sprayed onto plants with an atomizer until leaves were saturated. Inoculated plants were incubated overnight in a 20°C dew chamber, and then transferred to a 25°C greenhouse. The susceptible line Williams 82 (Wm82) was

1 377 used as a control. Plants were photographed 14 days after inoculation. Phenotypes
2
3
4 378 were scored as immune (IM1), immune with flecking (IMF), red-brown lesions with no
5
6 379 uredinia (RB2), red-brown lesions with some lesions containing uredinia (RB3), red-
7
8 380 brown lesions with most lesions containing uredinia (RB4), or Tan lesions with
9
10 381 uredinia (TAN5). A mixed phenotype was observed for one soybean line and *P.*
11
12 382 *pachyrhizi* isolate combination.
13
14

15 383 To conduct subsequent statistical analyses, we constructed a data matrix by
16
17 384 attributing numerical scores to the phenotypic responses of soybean genotypes
18
19 385 against seven *P. pachyrhizi* isolates. This matrix classified disease severity degree
20
21 386 into six distinct classes, each of which was converted to a numerical score: 0 =
22
23 387 Immune, 1 = Immune with flecking, 2 = Red-brown lesions with no uredinia (highly
24
25 388 resistant), 3 = Red-brown lesions with some lesions containing uredinia (resistant), 4
26
27 389 = Red-brown lesions with most lesions containing uredinia (moderately resistant),
28
29 390 and 5 = Tan lesions with uredinia (susceptible). We then input the matrix data into the
30
31 391 R Pheatmap v1.0.12 package ([https://cran.r-](https://cran.r-project.org/web/packages/pheatmap/index.html)
32
33 392 [project.org/web/packages/pheatmap/index.html](https://cran.r-project.org/web/packages/pheatmap/index.html)) to generate a heatmap and perform
34
35 393 hierarchical clustering, enabling us to visualize the virulence patterns of *P. pachyrhizi*
36
37 394 isolates.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

RESULTS

396

Genome-wide SNP identification in *P. pachyrhizi* panel

398 We re-sequenced the whole-genomes of 42 diverse isolates of *P. pachyrhizi*
399 collected throughout a wide geographic range, including Africa, Asia, Australia, South
400 America, and the USA. The average genome coverage was 88.99% at 15X across
401 42 isolates, with a maximum coverage of 95.68% for isolate Pp10 (Supplementary
402 Table S1). To investigate whole-genomic diversity among these 42 isolates, we
403 identified SNPs by mapping the short-read sequences of each isolate along the
404 UFV02 v2.1 reference genome assembly. The percent of mapped reads ranged from
405 75.53% (isolate CMES1958) to 98.48% (isolate TW72-1) (Supplementary Table S1).

406 To further explore the genomic context within *P. pachyrhizi*, we also included
407 genetic variants discovered on the high-quality genome assemblies of the three
408 isolates, K8108, MT2006, and UFV02, which were sequenced in a previous
409 investigation (Gupta et al., 2023). In total, our study comprised 45 *P. pachyrhizi*
410 isolates.

411 After applying quality control filters and minimum allele frequency ($MAF \geq 5\%$)
412 (see Material and Methods section), our pipeline detected a total of 452,853 whole-
413 genomic SNPs among 45 *P. pachyrhizi* isolates. The SnpEff software revealed the
414 potential functional effects of these SNPs on the *P. pachyrhizi* genome
415 (Supplementary Table S5). The majority of identified SNPs displayed either modifier
416 or moderate impacts, with only 0.04% of variants classified as having a high
417 functional impact. The largest proportion of SNPs was found in intergenic regions,
418 while coding regions, such as exons, exhibited a relatively low number of SNPs.

419

420 **Molecular evidence reveals two major evolutionary lineages in *P. pachyrhizi***

421 An ML phylogenetic tree, using all whole-genomic SNPs, inferred phylogenetic
422 relationships among 45 *P. pachyrhizi* isolates (Fig. 1A). The topology of the
423 phylogenetic tree showed two highly-divergent clades, each with high support
424 bootstrap values (100). Here, we refer to these two clades as lineages Pp1 and Pp2,
425 respectively. Notably, the two lineages were temporally separated. While lineage Pp1
426 comprised isolates sampled between 1972 and 1994, lineage Pp2 represented a
427 large group of isolates collected more recently, from 2003 onwards.

428 Lineage Pp1 comprised four isolates: TW72-1, IN73-1, AU79-1, and HW94-1
429 (Fig. 1A). These isolates were exclusively found in specific geographic locations,
430 namely Asia (Taiwan and India), Australia, and Hawaii (Fig. 2). Within the
431 phylogenetic tree (Fig. 1A), short branch lengths characterized lineage Pp2, with a
432 total of 41 isolates. Showing a widespread distribution, members of the lineage Pp2
433 were present across Africa, Asia, South America, and the USA. (Fig. 2).

434 Consistent with the phylogenetic results, the FastStructure choose K.py
435 algorithm clustered all 45 *P. pachyrhizi* isolates into two Bayesian groups (optimal K
436 = 2; Fig. 1B). As a whole, the Bayesian clustering method highly supported the
437 presence of two divergent lineages. Importantly, the geographic origin of the isolates
438 did not influence the placement of each Bayesian group. For example, isolates from
439 different areas (Asia, Africa, South America, and the USA) shared similar ancestry
440 within a single Bayesian group. Interestingly, two isolates (TW72-1 and AU79-1)
441 contained moderate proportions of membership assigned to both Bayesian groups,
442 indicating their genomes display certain levels of genetic admixture (Fig. 1B). We
443 also produced a PCA to illustrate genomic relationships among the 45 isolates (Fig.

1 444 1C). As expected, the PCA plot revealed that the isolates form two distinct clusters,
2
3
4 445 each corresponding to a described lineage.

5
6 446 As lineage Pp2 was the most populous clade discovered in our study, we
7
8 447 decided to further investigate its population structure by conducting three additional
9
10 448 analyses focusing on 41 isolates. Collectively, the three analyses (i.e. FastStructure
11
12 449 Bayesian clustering, PCA, and Minimum spanning network) indicated a lack of
13
14
15 450 genetic structure within lineage Pp2, suggesting high levels of clonal reproduction
16
17
18 451 and long-distance dispersal of spores (Supplementary Fig. S1).

19
20
21 452

22 23 453 **Genetic diversity and evidence of selection in two *P. pachyrhizi* lineages**

24
25 454 We detected distinct patterns of molecular genetic diversity shaping the
26
27
28 455 evolution of *P. pachyrhizi* lineages (Fig. 3A-F; Supplementary Table S6A-B). Based
29
30 456 on whole-genomic SNP data, lineage Pp2 exhibited significant ($p < 2.2 \times 10^{-16}$) lower
31
32 457 levels of nucleotide diversity (mean $\pi = 0.63 \times 10^{-4}$) and Watterson's Theta (mean θ_w
33
34
35 458 $= 0.29 \times 10^{-4}$) when compared to lineage Pp1 (Fig. 3A-B). This trend was also
36
37
38 459 observed in both predicted effector and non-effector genes, in which lineage Pp2
39
40 460 showed substantially lower genetic diversity (Fig. 3C-F). We found a high value of
41
42 461 genetic differentiation between lineages Pp1 and Pp2 (Mean $F_{ST} = 0.53$ and Weight
43
44
45 462 $F_{ST} = 0.55$).

46
47 463 Notably, Tajima's D neutrality test suggested that *P. pachyrhizi* lineages may
48
49
50 464 have evolved under different effects of natural selection (Fig. 3G-H; Supplementary
51
52 465 Table S6A-B). Although both *P. pachyrhizi* lineages exhibited positive mean Tajima's
53
54 466 D values within predicted effector (Fig. 3G) and non-effector genes (Fig. 3H), lineage
55
56
57 467 Pp2 showed a significantly higher proportion of positive Tajima's D values, indicating
58
59
60 468 balancing selection.

469

470 **Genomic characterization of mating-type (MAT) loci in *P. pachyrhizi***

471 We used the assembled genomes of three *P. pachyrhizi* isolates K8108,
472 MT2006, and UFV02 (Gupta et al., 2023) to identify *MAT* genes. The three genomes
473 together contained a total of 35 putative *MAT* genes, including pheromone receptors
474 (*STE3.2*), pheromone precursors (*MFA*), and homeodomain transcription factors
475 (*HD*) (Supplementary Table S7). The *STE3.2* pheromone receptor genes from *P.*
476 *pachyrhizi* shared a canonical G protein-coupled receptor domain (PF02076). The
477 repertoire of *STE3.2* gene-encoding pheromone receptors ranged from four copies
478 (isolate K8108) to seven (isolate UFV02), while *MFA* genes were identified with only
479 two copies per isolate. On the other hand, each *P. pachyrhizi* genome harbored two
480 pairs of genes that encoded homeodomain transcription factors (namely HD1 and
481 HD2), each containing a canonical homeobox KN domain (PF05920).

482 We constructed a schematic diagram to better illustrate the predicted structure
483 of *MAT* genes in the *P. pachyrhizi* UFV02 reference genome (Fig. 4A-B). The *MAT*
484 genes of the UFV02 genome were located in physically unlinked loci designated *P/R*
485 and *HD*. These loci were present in both the primary (Fig. 4A) and secondary (Fig.
486 4B) haplotypes of the *P. pachyrhizi* UFV02 genome. The *P/R* loci carried *STE3.2*
487 genes-encoding pheromone receptors and their precursors (*MFA*). In contrast, the
488 *HD* loci contained only *HD1* and *HD2* genes-encoding homeodomain transcription
489 factors. Within the UFV02 genome assembly, HD1 and HD2 genes were connected
490 by short DNA sequences: 776 bp in the primary haplotype and 84 bp in the
491 secondary haplotype (Fig. 4A).

492 Our ML phylogenetic reconstructions elucidated the evolutionary relationships
493 for each set of *MAT* protein sequences from *P. pachyrhizi* isolates (K8108, MT2006,

1 494 and UFV02), with two *Puccinia* species serving as outgroups (Fig. 4C-E). The
2
3 495 STE3.2 pheromone receptors from both *P. pachyrhizi* and *Puccinia* species were
4
5 496 clustered into two major highly-supported clades, with bootstrap values of 100. (Fig.
6
7 497 4C). In agreement with previous findings (Cuomo et al., 2017), one such clade was
8
9 498 named STE3.2.1, while the other was further divided into two sub-clades, STE3.2.2
10
11 499 and STE3.2.3. With a low bootstrap value (>80), the MAF proteins from *P. pachyrhizi*
12
13 500 constituted a single clade, (Fig. 4D). We found that HD proteins from *P. pachyrhizi*
14
15 501 were closely related to those from *Puccinia* species, forming two main clades, each
16
17 502 representing a homologous gene (*HD1* and *HD2*) (Fig. 4E).

22
23 503 To gain new insights into the expression patterns of *MAT* genes during the
24
25 504 interaction between *P. pachyrhizi* and soybean plants, we analyzed publicly
26
27 505 accessible transcriptome data for the three isolates K8108, MT2006, and UFV02
28
29 506 (Gupta et al., 2023). The expression of *MAT* genes was evaluated throughout various
30
31 507 time points, including *in-vitro* stages (spore germination) and *in-plant* stages
32
33 508 (soybean infection process). Overall, *MAT* genes in those three isolates exhibited
34
35 509 varying levels of expression over the examined timeframe (Supplementary Table S8,
36
37 510 Fig. S2). We detected a notable trend for *MAT* genes showing increasing expression
38
39 511 levels from 72 hours post-spore inoculation. This finding was most obvious in isolate
40
41 512 UFV02 (Supplementary Fig. S2 C). The expression of several *STE3.2* pheromone
42
43 513 receptors and *HD* genes were significantly up-regulated in the late stages of soybean
44
45 514 infection, namely during sporulation.

51
52 515 Finally, we examined polymorphisms among the predicted *MAT* genes across
53
54 516 whole-genome data from all 45 *P. pachyrhizi* isolates (Dataset A). Our analysis
55
56 517 identified six SNPs within *STE3.2.1* pheromone receptor genes, while no SNPs were
57
58 518 observed in the remaining *MAT* genes (Supplementary Table S9). By applying

1 519 SnpEff software, we found that four out of six SNPs are located within intron regions
2
3
4 520 of *STE3.2.1* genes, while the other two SNPs were predicted to be synonymous or
5
6 521 missense variants.
7

8 522
9

10 523 **Virulence profiles of *P. pachyrhizi***

11
12
13 524 Our phenotyping assays identified distinct virulence patterns among four *P.*
14
15 525 *pachyrhizi* isolates (AU79-1, HW94-1, IN73-1, and TW72-1) from lineage Pp1 and
16
17 526 three isolates (CO04-2, MT2006, and UFV02) from lineage Pp2 (Fig. 5;
18
19 527 Supplementary Fig. S3; Supplementary Table S10A-B). The virulence traits of *P.*
20
21 528 *pachyrhizi* isolates were analyzed on eight differential soybean PI lines, plus a
22
23 529 susceptible soybean accession (BRS 184 or Wm82) utilized as a positive control.
24
25 530 Both isolates, HW94-1 and IN73-1, from lineage Pp1 exhibited low profiles of
26
27 531 virulence, with all eight soybean PI lines demonstrating immune reactions or some
28
29 532 resistance response to these isolates (Fig. 5). On the contrary, the highest levels of
30
31 533 virulence were present in the isolate UFV02, a member of the lineage Pp2. The
32
33 534 isolate UFV02 caused susceptible reactions in eight soybean lines, although PI
34
35 535 200487 (*Rpp5*) was moderately resistant to it. In terms of virulence levels, the
36
37 536 hierarchical clustering dendrogram (Fig. 5) did not reflect the phylogenetic
38
39 537 placements of the isolates as shown in Fig. 1A
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

DISCUSSION

540

Two evolutionary lineages drove the diversification of *P. pachyrhizi*

Previous studies of genetic diversity in *P. pachyrhizi* have relied on microsatellite markers (Twizeyimana et al., 2011), AFLPs (Amplified Fragment Length Polymorphism) (Darben et al., 2020), as well as specific genomic regions such as ITS (Internal transcribed spacer) (Freire et al., 2008; Jorge et al., 2015). This study represents a pioneering effort to acquire whole-genome re-sequencing data from multiple *P. pachyrhizi* isolates. Our results provide a comprehensive view of the genetic diversity in *P. pachyrhizi* at the genome-wide level.

In our study, phylogenetic reconstructions, based on whole-genomic SNP data, uncovered two divergent evolutionary lineages of *P. pachyrhizi* distributed across diverse global regions. It is well-established that populations of plant-pathogenic fungi undergo rapid evolutionary changes, often leading to the emergence of locally adapted lineages in response to selection pressure from host species (Gilbert and Parker, 2010). Indeed, the evolution of virulent clonal lineages has been documented in invasive rust fungal populations, including *Puccinia graminis* f. sp. *tritici* (Brar et al., 2018), *Uromyces transversalis* (DeLong et al., 2019), and *Puccinia triticina* (Fellers et al., 2021). Our findings align with previous studies that identified genetically distinct clusters in *P. pachyrhizi* (Twizeyimana et al., 2011; Zhang et al., 2012). Utilizing both ITS and ARF (ADP-Ribosylation Factor) sequences, molecular analyses showed that *P. pachyrhizi* genotypes from four geographic regions — Africa, America, Asia, and Australia — formed distinct clades that were not influenced by geographic locations of the *P. pachyrhizi* isolates (Zhang et al., 2012).

1 564 The long-distance transport of airborne spores via air currents has been a key
2
3
4 565 dispersal mechanism for *P. pachyrhizi* (Schneider et al., 2005; Hossain et al., 2024).
5
6 566 This phenomenon has potentially contributed to the widespread distribution of
7
8 567 members of the two evolutionary lineages (Pp1 and Pp2) in *P. pachyrhizi* across
9
10 568 large geographic distances. Notably, lineage Pp1 was predominantly dispersed in the
11
12 569 Eastern Hemisphere, with its members established in countries such as Australia,
13
14 570 India, and Taiwan. Long-distance dispersal events likely explain the presence of one
15
16 571 member of lineage Pp1, observed in the Hawaii archipelago. With a larger
17
18 572 geographic distribution, lineage Pp2 exhibits a remarkable ability to disseminate
19
20 573 fungal spores across vast distances. Through airborne inoculum dispersal, members
21
22 574 of the lineage Pp2 may have reached major soybean-producing centers in Africa,
23
24 575 South America, and the continental USA. Consistent with a general trend observed in
25
26 576 prior works (Jorge et al., 2015; Darben et al., 2020), we found a lack of genetic
27
28 577 structure across broad geographic ranges within each of the evolutionary lineages,
29
30 578 Pp1 and Pp2, in *P. pachyrhizi*. This absence of spatial genetic structure in *P.*
31
32 579 *pachyrhizi* lineages is likely due to the high propagation rates of clonal spores among
33
34 580 soybean-growing areas located on large geographic scales. The maintenance of *P.*
35
36 581 *pachyrhizi* clonal lineages in fields may be facilitated by alternative host species,
37
38 582 which serve as inoculum reservoirs during planting seasons (Jarvie, 2009)
39

40 583 Despite their close phylogenetic relationships, limited gene flow was observed
41
42 584 between the two *P. pachyrhizi* lineages, Pp1 and Pp2, as indicated by high F_{ST}
43
44 585 values. Within lineage Pp2, low levels of polymorphism were detected, with only a
45
46 586 few SNPs distinguishing different isolates. The reduced genetic diversity suggests
47
48 587 that lineage Pp2 displays highly clonal reproduction. In contrast, a significant
49
50 588 accumulation of genomic polymorphism was observed among isolates in lineage
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 589 Pp1. The high levels of genomic variability within lineage Pp1 may be attributed to its
2
3
4 590 presence in regions with low-intensity soybean production, where relaxed selective
5
6 591 constraints could have promoted increased genetic variation. The genetic diversity of
7
8 592 *P. pachyrhizi* may also reflect mechanisms such as somatic recombination via hyphal
9
10
11 593 anastomosis (Vittal et al., 2012) and intense activity of transposable elements in the
12
13 594 dikaryotic urediniospores (Gupta et al., 2023).

15 595 Temporal shifts are thought to drive the emergence of new genotypes and the
16
17
18 596 selection of specific gene loci to elevate virulence in plant-pathogen populations
19
20 597 (Miller et al., 2020). The temporal divergence between the two lineages Pp1 and Pp2
21
22
23 598 in *P. pachyrhizi* is an interesting discovery. Our findings indicate that lineage Pp1
24
25 599 cluster *P. pachyrhizi* isolates obtained from several decades ago (1972 -1994), while
26
27
28 600 lineage Pp2 includes more recently collected isolates (2003-2017). We propose two
29
30 601 possible scenarios to explain the temporal separation observed in the two *P.*
31
32 602 *pachyrhizi* lineages. The first scenario suggests that our sampling efforts may have
33
34
35 603 been insufficient to catch recent isolates of *P. pachyrhizi* belonging to lineage Pp1.
36
37 604 This scenario implies that lineage Pp1 continues to circulate in the current soybean-
38
39
40 605 producing zones. On the other hand, the second scenario postulates that lineage
41
42 606 Pp1 may represent an ancestral group, and its parental members could have evolved
43
44
45 607 in response to host interactions over time, subsequently giving rise to lineage Pp2,
46
47 608 which is now prevalent across the world. The impact of temporal differentiation on the
48
49
50 609 population structure of plant-pathogens has been reported. For example, in the
51
52 610 crown rust fungus (*Puccinia coronata* f. sp. *avenae*), populations sampled in the USA
53
54
55 611 during 1990 and 2015 formed two distinct genetic clusters, each comprising isolates
56
57 612 from a single collection year (Miller et al., 2020). In this case, the latter *P. coronata*
58
59
60
61
62
63
64
65

population (2015) likely derived from the 1990 population and acquired a substantial increase in virulence levels (Miller et al., 2020).

Previously, evidence indicated that *P. pachyrhizi* appears to be originated in Asia (Hossain et al., 2024), with its earliest records from 1902 in Japan (Kitani and Inoue, 1960). We speculate that lineages Pp1 and Pp2 may have initially emerged in Asia and later spread to geographic regions worldwide. The co-existence of both evolutionary lineages Pp1 and Pp2 in Asia implies that their evolutionary histories may have been interconnected over ancient periods in that region. However, extensive sampling of additional *P. pachyrhizi* isolates, especially in the Asian continent, is required to determine the precise geographic origin for lineages Pp1 and Pp2 of *P. pachyrhizi*.

Genetic admixture among *P. pachyrhizi* lineages

The origin of two *P. pachyrhizi* isolates, TW72-1 (from Taiwan) and AU79-1 (from Australia), is likely related to genomic admixture involving the lineages Pp1 and Pp2. fastStructure analysis revealed that TW72-1 and AU79-1 share ancestral origins to both lineages Pp1 and Pp2 (Fig. 1B). Genetic admixture in rust fungi may be a consequence of multiple mechanisms, including sexual recombination and somatic hybridization (Guo et al., 2022). A recent study utilizing fully nuclear-phased genome assemblies demonstrated that somatic hybridization through nuclear exchange led to the emergence of wheat leaf rust (*Puccinia triticina*) lineages (Sperschneider et al., 2023). The mechanisms involving the genetic admixture of TW72-1 and AU79-1 are not fully explained in our study. However, one possible explanation is that members of lineages Pp1 and Pp2 may be mating-compatible partners and can outcross, producing new recombinant genotypes (e.g. TW72-1 and AU79-1). This hypothesis

would be supported by the presence of different *MAT* alleles among isolates of lineages Pp1 and Pp2, although our analysis revealed reduced allelic variation within *MAT* genes of *P. pachyrhizi* lineages (Supplementary Table S9). Specifically, only the *STE3.2.1* pheromone receptor genes showed some polymorphisms, while other *MAT* genes exhibited no allelic variations among isolates in lineages Pp1 and Pp2. Therefore, sexually compatible outcrossing between lineages Pp1 and Pp2 appears a less likely hypothesis to elucidate the origin of TW72-1 and AU79-1. Alternatively, somatic hybridization through nuclear exchanges may be a more plausible scenario to explain the genetic admixture observed in TW72-1 and AU79-1. Considering this possible scenario, isolates TW72-1 and AU79-1 may harbor dikaryotic spores containing two different nuclear haplotypes that were derived from lineages Pp1 and Pp2, independently. The recognized occurrence of hyphal anastomosis followed by nuclear migration in germ tubes of *P. pachyrhizi* spores (Vittal et al., 2012) further reinforces the possibility of a somatic hybridization event among *P. pachyrhizi* lineages. Similar incidents associated with somatic hybridization via nuclear exchange have been documented in other fungi rust, for instance, *Puccinia graminis* f. sp. *tritici* (Li et al., 2019), *Puccinia triticina* (Sperschneider et al., 2023) and *Puccinia coronata* f. sp. *avenae* (Henningesen et al., 2024). Although the sexual cycle of *P. pachyrhizi* has not been characterized, its role should not be ruled out as a factor to explain the genomic admixture observed in TW72-1 and AU79-1.

Tetrapolar mating types in *P. pachyrhizi*

In basidiomycete fungi, the role of *MAT* genes in controlling sexual compatibility is well-established (Coelho et al., 2017). In line with previous findings in other basidiomycete fungi, such as myrtle rust, *Austropuccinia psidii* (Ferrarezi et al.,

2022), and *Puccinia* species (Luo et al., 2024), our analyses confirmed that *P. pachyrhizi* possesses a tetrapolar mating system that likely regulates sexual compatibility. Indeed, within the phylum Basidiomycota including rust fungi, a tetrapolar mating system appears to be an ancestral trait (Luo et al., 2024). The tetrapolar mating system of *P. pachyrhizi*, composed of physically unlinked HD and P/R loci, is present in both primary and secondary haplotypes of the genome assemblies. This finding supports a common pattern observed in rust fungi (Ferrarezi et al., 2022; Luo et al., 2024). Typically, in the cereal rust fungi genomes, *MFA* pheromone genes are found close to their respective *STE3.2* gene-encoding receptors on the chromosomes (Luo et al., 2024). On the contrary, the *MFA* genes of *P. pachyrhizi* were located on distinct contigs separated from the *STE3.2* receptor genes. This unexpected finding likely reflects fragmented genome assemblies of *P. pachyrhizi*, which may have prevented us from achieving the full understanding of the genomic architecture of *MAT* genes in this species. While two *MFA* pheromone genes were identified in each examined *P. pachyrhizi* isolate, it is possible that additional *MFA* copies are present within their genomes. The short length of *MFA* proteins (80 amino acids) in *P. pachyrhizi* further complicates their identification through bioinformatic tools like BlastP.

The predicted *MAT* genes in *P. pachyrhizi* are potentially functional, as evidenced by their significant expression levels observed in the transcriptome data. The high expression of *MAT* genes during the later stages of soybean-*P. pachyrhizi* infection is consistent with previous observations in the poplar leaf rust fungus *Melampsora larici-populina* (Duplessis et al., 2011b) and wheat rust fungi, e.g. *Puccinia* species (Luo et al., 2024). The notable expression patterns of *P. pachyrhizi* *MAT* genes raise interesting questions about how *MAT* genes might contribute to

1 688 urediniospores formation over the late stages of the *P. pachyrhizi* infection cycle on
2
3 689 soybean. A recent study, focusing on rust fungi within the genus *Puccinia*, suggested
4
5 690 that *MAT* genes may play a crucial role in regulating the pairing of two haploid nuclei
6
7 691 during the development of new dikaryotic urediniospores in the late infection phases
8
9 692 (Luo et al., 2024). If *MAT* genes are indeed involved in the formation of *P. pachyrhizi*
10
11 693 asexual urediniospores, experimental analyses are needed to confirm their functional
12
13 694 roles.
14
15
16
17
18
19

20 695

21 696 **Virulence profiles of *P. pachyrhizi*: implications for disease management**

22
23 697 Experiment-based evidence has revealed the existence of several potential
24
25 698 pathogenic races of *P. pachyrhizi*, with varying levels of virulence (Yamaoka et al.,
26
27 699 2002; Castanho, 2022). Our findings demonstrated that different isolates from
28
29 700 lineages Pp1 and Pp2 are virulent in most differential soybean lines, indicating their
30
31 701 notable adaptive ability to overcome resistance conferred by *Rpp* genes. The large
32
33 702 genome size of *P. pachyrhizi*, coupled with a high abundance of transposable
34
35 703 elements, is thought to play a key role in the adaptability and virulence of the
36
37 704 pathogen (Gupta et al., 2023).
38
39
40
41

42 705 The emergence of new *P. pachyrhizi* lineages demand the attention and
43
44 706 surveillance of phytopathologists to update ASR-resistance breeding or crop
45
46 707 protection approaches. To control ASR, the primary strategies have been fungicide
47
48 708 applications and the use of resistant soybean varieties (Chicowski et al., 2024). The
49
50 709 effective monitoring and managing of ASR is indispensable since *P. pachyrhizi*
51
52 710 populations may rapidly adapt to local conditions, manipulating host defense
53
54 711 mechanisms and giving rise to virulent strains that can quickly spread to neighboring
55
56 712 regions via wind-dispersed spores. When screening soybean germplasm for
57
58
59
60
61
62
63
64
65

1 713 resistance to ASR, breeding programs should consider the genomic variation among
2
3
4 714 *P. pachyrhizi* isolates within the lineages Pp1 and Pp2 as described in our study.
5

6 715
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

ACKNOWLEDGEMENTS

The work (proposal: 10.46936/10.25585/60000959) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. The study was supported by the Brazilian Agricultural Research Corporation - EMBRAPA (proposal 03.17.03.004.00.00) and Bayer SAS/EMBRAPA (proposal 33.17.00.113.00.00). The National Council of Scientific and Technological Development - CNPq and the Brazilian Ministry of Science, Technology and Innovation - MCTI funded project 440782/2022-8 and supported FCMG (PQ 799 312214/2022-7) and VDR (DTI-A 383533/2023-6). This work was also partially funded by the USDA-ARS project 8044-22000-051-000D. SD was supported by a grant of the French Plan Investissement d'Avenir (PIA) Lab of Excellence ARBRE [ANR-11-LABX-0002- 01]. We acknowledge Dr. Glen Hartman and Dr. H. M. Murithi, who kindly provided the isolates from Africa, Argentina and the US; Dr. Y. Yamaoka for the Japanese isolates, and Dr. Sergio Herminio Brommonschenkel for the UFV02 isolate.

Credit authorship contribution statement

F.C.M.G, M.L, R.T.V, I.V.G, U.S, C.S, and S.B: Conceptualization, Funding acquisition, Manuscript revision. **F.M.C and M.K.K:** Methodology, Isolation and cultivation of fungal isolates. **M.K.K:** Methodology, DNA extraction and quality analysis, Phenotypic characterization **K.F.P.:** Cultivation of fungal isolates, Phenotypic characterization, Manuscript revision. **I.V.G, A.L and K.B:** Methodology, Re-sequencing analysis and In-silico analysis in the JGI platform. **C.V.G and M.C.M:**

1 741 Fungal spore sampling. **E.G.C F:** Methodology, Identification of genetic variants.
2
3 742 **V.D.R:** Data curation, Formal analysis, Visualization, Writing – original draft. **F.C.M.G:**
4
5 743 Writing – review & editing and Supervision. All authors read and approved the
6
7 744 submitted version.
8
9

10 745

11 746

12 747 **CONFLICT OF INTEREST**

13 748 All authors declare no conflicting interest

14 749

15 750 **DATA AVAILABILITY**

16 751 Whole-genome re-sequencing data have been deposited at the JGI portal
17
18 752 (<https://genome.jgi.doe.gov/portal/>) under project IDs available in Supplementary
19
20 753 Table S1.
21
22

23 754

24 755 **REFERENCES**

25 756 Akinsanmi, O.A., Ladipo, J.L., Oyekan, P.O., 2001. First report of soybean rust
26
27 757 *Phakopsora pachyrhizi* in Nigeria. Plant Dis 85, 97–97.
28
29 758 <https://doi.org/10.1094/PDIS.2001.85.1.97B>
30
31
32 759 Beruski, G.C., Sentelhas, P.C., Pereira, A.B., Câmara, G.M.S., Junior, I.P.A.,
33
34 760 Schiebelbein, L.M., 2020. Soybean rust epidemics as affected by weather
35
36 761 conditions in Brazil. Journal of Agricultural Science 12, 213–225.
37
38 762 <https://doi.org/10.5539/jas.v12n1p213>
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

763 Brar, G.S., Ali, S., Qutob, D., Ambrose, S., Lou, K., Maclachlan, R., Pozniak, C.J.,
 764 Fu, Y., Sharpe, A.G., Kutcher, H.R., 2018. Genome re-sequencing and simple
 765 sequence repeat markers reveal the existence of divergent lineages in the
 766 Canadian *Puccinia striiformis* f. sp. *tritici* population with extensive DNA
 767 methylation. *Environ Microbiol* 20, 1498–1515. [https://doi.org/10.1111/1462-](https://doi.org/10.1111/1462-2920.14067)
 768 2920.14067
 769 Bromfield, K.R., 1984. Soybean Rust. Monograph n. 11. St Paul, Minnesota. The
 770 American Phytopathological Society.
 771 Castanho, F.M., 2022. Population genomics and race characterization reveal genes
 772 associated with (a)virulence in *Phakopsora pachyrhizi*. Universidade Estadual de
 773 Londrina. Ph.D. Thesis.
 774 Chicowski, A.S., Bredow, M., Utiyama, A.S., Marcelino-Guimarães, F.C., Whitham,
 775 S.A., 2024. Soybean- *Phakopsora pachyrhizi* interactions: towards the
 776 development of next-generation disease-resistant plants. *Plant Biotechnol J* 22,
 777 296–315. <https://doi.org/10.1111/pbi.14206>
 778 Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu,
 779 X., Ruden, D.M., 2012. A program for annotating and predicting the effects of
 780 single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila*
 781 *melanogaster* strain *w¹¹¹⁸*; *iso-2*; *iso-3*. *Fly* (Austin) 6, 80–92.
 782 <https://doi.org/10.4161/fly.19695>

- 783 Coelho, M.A., Bakkeren, G., Sun, S., Hood, M.E., Giraud, T., 2017. Fungal Sex: The
784 Basidiomycota. *Microbiol Spectr* 5. [https://doi.org/10.1128/microbiolspec.FUNK-](https://doi.org/10.1128/microbiolspec.FUNK-0046-2016)
785 0046-2016
- 786 Cuomo, C.A., Bakkeren, G., Khalil, H.B., Panwar, V., Joly, D., Linning, R.,
787 Sakthikumar, S., Song, X., Adiconis, X., Fan, L., Goldberg, J.M., Levin, J.Z.,
788 Young, S., Zeng, Q., Anikster, Y., Bruce, M., Wang, M., Yin, C., McCallum, B.,
789 Szabo, L.J., Hulbert, S., Chen, X., Fellers, J.P., 2017. Comparative analysis
790 highlights variable genome content of wheat rusts and divergence of the mating
791 loci. *G3 Genes|Genomes|Genetics* 7, 361–376.
792 <https://doi.org/10.1534/g3.116.032797>
- 793 Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A.,
794 Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R.,
795 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.
796 <https://doi.org/10.1093/bioinformatics/btr330>
- 797 Darben, L.M., Yokoyama, A., Castanho, F.M., Lopes-Caitar, V.S., Cruz Gallo de
798 Carvalho, M.C., Godoy, C.V., Carvalho, S., Gonela, A., Marcelino-Guimarães,
799 F.C., 2020. Characterization of genetic diversity and pathogenicity of *Phakopsora*
800 *pachyrhizi* mono-uredinial isolates collected in Brazil. *Eur J Plant Pathol* 156, 355–
801 372. <https://doi.org/10.1007/s10658-019-01872-2>
- 802 Dellaporta, S.L., Wood, J., Hicks, J.B., 1983. A plant DNA miniprep: Version II.
803 *Plant Mol Biol Report* 1, 19–21. <https://doi.org/10.1007/BF02712670>

- 804 DeLong, J.A., Stewart, J.E., Valencia-Botín, A., Pedley, K.F., Buck, J.W., Brewer,
805 M.T., 2019. Invasions of gladiolus rust in North America are caused by a widely-
806 distributed clone of *Uromyces transversalis*. PeerJ 7, e7986.
807 <https://doi.org/10.7717/peerj.7986>
- 808 Duplessis, S., Cuomo, C.A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C.,
809 Joly, D.L., Hacquard, S., Amselem, J., Cantarel, B.L., Chiu, R., Coutinho, P.M.,
810 Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M.G., Kodira,
811 C.D., Kohler, A., Kues, U., Lindquist, E.A., Lucas, S.M., Mago, R., Mauceli, E.,
812 Morin, E., Murat, C., Pangilinan, J.L., Park, R., Pearson, M., Quesneville, H.,
813 Rouhier, N., Sakthikumar, S., Salamov, A.A., Schmutz, J., Selles, B., Shapiro, H.,
814 Tanguay, P., Tuskan, G.A., Henrissat, B., Van de Peer, Y., Rouzé, P., Ellis, J.G.,
815 Dodds, P.N., Schein, J.E., Zhong, S., Hamelin, R.C., Grigoriev, I. V., Szabo, L.J.,
816 Martin, F., 2011a. Obligate biotrophy features unraveled by the genomic analysis
817 of rust fungi. Proceedings of the National Academy of Sciences 108, 9166–9171.
818 <https://doi.org/10.1073/pnas.1019315108>
- 819 Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., Kohler,
820 A., 2011b. *Melampsora larici-populina* transcript profiling during germination and
821 timecourse infection of poplar leaves reveals dynamic expression patterns
822 associated with virulence and biotrophy. Molecular Plant-Microbe Interactions 24,
823 808–818. <https://doi.org/10.1094/MPMI-01-11-0006>
- 824 Fellers, J.P., Sakthikumar, S., He, F., McRell, K., Bakkeren, G., Cuomo, C.A.,
825 Kolmer, J.A., 2021. Whole-genome sequencing of multiple isolates of *Puccinia*

1 826 *triticina* reveals asexual lineages evolving by recurrent mutations. G3 11.
2
3 827 <https://doi.org/10.1093/g3journal/jkab219>
4
5
6
7 828 Ferrarezi, J.A., McTaggart, A.R., Tobias, P.A., Hayashibara, C.A.A., Degnan, R.M.,
8
9 829 Shuey, L.S., Franceschini, L.M., Lopes, M.S., Quecine, M.C., 2022.
10
11 830 *Austropuccinia psidii* uses tetrapolar mating and produces meiotic spores in older
12
13 831 infections on *Eucalyptus grandis*. Fungal Genetics and Biology 160, 103692.
14
15 832 <https://doi.org/10.1016/j.fgb.2022.103692>
16
17
18
19
20 833 Francis, R.M., 2017. pophelper: an R package and web app to analyse and visualize
21
22 834 population structure. Mol Ecol Resour 17, 27–32. <https://doi.org/10.1111/1755->
23
24 835 0998.12509
25
26
27
28 836 Freire, M.C.M., Oliveira, L.O. de, Almeida, Á.M.R. de, Schuster, I., Moreira, M.A.,
29
30 837 Liebenberg, M.M., Mienie, C.M.S., 2008. Evolutionary history of *Phakopsora*
31
32 838 *pachyrhizi* (the Asian Soybean Rust) in Brazil based on nucleotide sequences of
33
34 839 the internal transcribed spacer region of the nuclear ribosomal DNA. Genet Mol
35
36 840 Biol 31, 920–931. <https://doi.org/10.1590/S1415-47572008005000026>
37
38
39
40
41
42 841 Gilbert, G.S., Parker, I.M., 2010. Rapid evolution in a plant-pathogen interaction and
43
44 842 the consequences for introduced host species. Evol Appl 3, 144–156.
45
46 843 <https://doi.org/10.1111/j.1752-4571.2009.00107.x>
47
48
49
50 844 Goellner, K., Loehrer, M., Langenbach, C., Conrath, U., Koch, E., Schaffrath, U.,
51
52 845 2010. *Phakopsora pachyrhizi*, the causal agent of Asian Soybean Rust. Mol Plant
53
54 846 Pathol 11, 169–177. <https://doi.org/10.1111/j.1364-3703.2009.00589.x>
55
56
57
58
59
60
61
62
63
64
65

847 Guo, Y., Betzen, B., Salcedo, A., He, F., Bowden, R.L., Fellers, J.P., Jordan, K.W.,
 848 Akhunova, A., Rouse, M.N., Szabo, L.J., Akhunov, E., 2022. Population genomics
 849 of *Puccinia graminis* f.sp. *tritici* highlights the role of admixture in the origin of
 850 virulent wheat rust races. Nat Commun 13, 6287. [https://doi.org/10.1038/s41467-](https://doi.org/10.1038/s41467-022-34050-w)
 851 022-34050-w
 852 Gupta, Y.K., Marcelino-Guimarães, F.C., Lorrain, C., Farmer, A., Haridas, S.,
 853 Ferreira, E.G.C., Lopes-Caitar, V.S., Oliveira, L.S., Morin, E., Widdison, S.,
 854 Cameron, C., Inoue, Y., Thor, K., Robinson, K., Drula, E., Henrissat, B., LaButti,
 855 K., Bini, A.M.R., Paget, E., Singan, V., Daum, C., Dorme, C., van Hoek, M.,
 856 Janssen, A., Chandat, L., Tarriotte, Y., Richardson, J., Melo, B. do V.A.,
 857 Wittenberg, A.H.J., Schneiders, H., Peyrard, S., Zanardo, L.G., Holtman, V.C.,
 858 Coulombier-Chauvel, F., Link, T.I., Balmer, D., Müller, A.N., Kind, S., Bohnert, S.,
 859 Wirtz, L., Chen, C., Yan, M., Ng, V., Gautier, P., Meyer, M.C., Voegelé, R.T., Liu,
 860 Q., Grigoriev, I. V., Conrath, U., Brommonschenkel, S.H., Loehrer, M., Schaffrath,
 861 U., Sirven, C., Scalliet, G., Duplessis, S., van Esse, H.P., 2023. Major proliferation
 862 of transposable elements shaped the genome of the soybean rust pathogen
 863 *Phakopsora pachyrhizi*. Nat Commun 14, 1835. [https://doi.org/10.1038/s41467-](https://doi.org/10.1038/s41467-023-37551-4)
 864 023-37551-4
 865 Hartman, G.L., Hill, C.B., Twizeyimana, M., Miles, M.R., Bandyopadhyay, R., 2011.
 866 Interaction of soybean and *Phakopsora pachyrhizi*, the cause of soybean rust.
 867 CABI Reviews 1–13. <https://doi.org/10.1079/PAVSNNR20116025>

- 868 Hartman, G.L., Rupe, J.C., Sikora, E.J., Domier, L.L., Davis, J.A., Steffey, K.L., 2016.
869 Compendium of Soybean Diseases and Pests, 5th ed. The American
870 Phytopathological Society, St Paul, Minnesota.
- 871 Henningsen, E.C., Lewis, D., Nazareno, E., Huang, Y.F., Steffenson, B.J., Boesen,
872 B., Kianian, S.F., Stone, E., Dodds, P.N., Sperschneider, J., Figueroa, M., 2024. A
873 high-resolution haplotype pangenome uncovers somatic hybridization,
874 recombination and intercontinental migration in oat crown rust. bioRxiv.
875 <https://doi.org/10.1101/2024.03.27.583983>.
- 876 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018.
877 UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35, 518–
878 522. <https://doi.org/10.1093/molbev/msx281>
- 879 Hossain, Md.M., Sultana, F., Yesmin, L., Rubayet, Md.T., Abdullah, H.M., Siddique,
880 S.S., Bhuiyan, Md.A.B., Yamanaka, N., 2024. Understanding *Phakopsora*
881 *pachyrhizi* in soybean: comprehensive insights, threats, and interventions from the
882 Asian perspective. Front Microbiol 14. <https://doi.org/10.3389/fmicb.2023.1304205>
- 883 Jarvie, J.A., 2009. A review of soybean rust from a South African perspective. S Afr J
884 Sci 105. <https://doi.org/10.4102/sajs.v105i3/4.55>
- 885 Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic
886 markers. Bioinformatics 24, 1403–1405.
887 <https://doi.org/10.1093/bioinformatics/btn129>
- 888 Jorge, V.R., Silva, M.R., Guillin, E.A., Freire, M.C.M., Schuster, I., Almeida, A.M.R.,
889 Oliveira, L.O., 2015. The origin and genetic diversity of the causal agent of Asian

- 1 890 Soybean Rust, *Phakopsora pachyrhizi*, in South America. Plant Pathol 64, 729–
2
3
4 891 737. <https://doi.org/10.1111/ppa.12300>
5
6
7 892 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermin, L.S.,
8
9 893 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat
10
11 894 Methods 14, 587–589. <https://doi.org/10.1038/nmeth.4285>
12
13
14
15 895 Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software
16
17 896 Version 7: Improvements in Performance and Usability. Mol Biol Evol 30, 772–
18
19 897 780. <https://doi.org/10.1093/molbev/mst010>
20
21
22
23 898 Kendrick, M.D., Harris, D.K., Ha, B.-K., Hyten, D.L., Cregan, P.B., Frederick, R.D.,
24
25 899 Boerma, H.R., Pedley, K.F., 2011. Identification of a second Asian Soybean Rust
26
27 900 Resistance Gene in Hyuuga Soybean. Phytopathology 101, 535–543.
28
29 901 <https://doi.org/10.1094/PHYTO-09-10-0257>
30
31
32
33
34 902 Kitani, K., Inoue, Y., 1960. Studies on the soybean rust and its control measure, part.
35
36 903 1. Studies on the soybean rust. Bull. Shikoku Agric. Exp. Station, 5, 319–342.
37
38
39
40 904 Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C.,
41
42 905 Salzberg, S.L., 2004. Versatile and open software for comparing large genomes.
43
44 906 Genome Biol 5, R12. <https://doi.org/10.1186/gb-2004-5-2-r12>
45
46
47
48 907 Levy, C., 2005. Epidemiology and Chemical Control of Soybean Rust in Southern
49
50 908 Africa. Plant Dis 89, 669–674. <https://doi.org/10.1094/PD-89-0669>
51
52
53
54 909 Li, F., Upadhyaya, N.M., Sperschneider, J., Matny, O., Nguyen-Phuc, H., Mago, R.,
55
56 910 Raley, C., Miller, M.E., Silverstein, K.A.T., Henningsen, E., Hirsch, C.D., Visser,
57
58
59
60
61
62
63
64
65

911 B., Pretorius, Z.A., Steffenson, B.J., Schwessinger, B., Dodds, P.N., Figueroa, M.,
 912 2019. Emergence of the Ug99 lineage of the wheat stem rust pathogen through
 913 somatic hybridization. *Nat Commun* 10, 5068. [https://doi.org/10.1038/s41467-019-](https://doi.org/10.1038/s41467-019-12927-7)
 914 12927-7
 915 Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association
 916 mapping and population genetical parameter estimation from sequencing data.
 917 *Bioinformatics* 27, 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
 918 Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–
 919 Wheeler transform. *Bioinformatics* 25, 1754–1760.
 920 <https://doi.org/10.1093/bioinformatics/btp324>
 921 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
 922 Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and
 923 SAMtools. *Bioinformatics* 25, 2078–2079.
 924 <https://doi.org/10.1093/bioinformatics/btp352>
 925 Luo, Z., McTaggart, A., Schwessinger, B., 2024. Genome biology and evolution of
 926 mating-type loci in four cereal rust fungi. *PLoS Genet* 20, e1011207.
 927 <https://doi.org/10.1371/journal.pgen.1011207>
 928 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A.,
 929 Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The
 930 Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation
 931 DNA sequencing data. *Genome Res* 20, 1297–1303.
 932 <https://doi.org/10.1101/gr.107524.110>

- 933 Miller, M.E., Nazareno, E.S., Rottschaefer, S.M., Riddle, J., Dos Santos Pereira, D.,
934 Li, F., Nguyen-Phuc, H., Henningsen, E.C., Persoons, A., Saunders, D.G.O.,
935 Stukenbrock, E., Dodds, P.N., Kianian, S.F., Figueroa, M., 2020. Increased
936 virulence of *Puccinia coronata* f. sp. *avenae* populations through allele frequency
937 changes at multiple putative Avr loci. PLoS Genet 16, e1009291.
938 <https://doi.org/10.1371/journal.pgen.1009291>
- 939 Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer,
940 E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman,
941 A., 2021. Pfam: The protein families database in 2021. Nucleic Acids Res 49,
942 D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- 943 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast
944 and Effective Stochastic Algorithm for Estimating Maximum-Likelihood
945 Phylogenies. Mol Biol Evol 32, 268–274. <https://doi.org/10.1093/molbev/msu300>
- 946 Pfeifer, B., Wittelsburger, U., Ramos-Onsins, S.E., Lercher, M.J., 2014. PopGenome:
947 an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol
948 31, 1929–1936. <https://doi.org/10.1093/molbev/msu136>
- 949 Raj, A., Stephens, M., Pritchard, J.K., 2014. fastSTRUCTURE: variational inference
950 of population structure in large SNP data Sets. Genetics 197, 573–589.
951 <https://doi.org/10.1534/genetics.114.164350>
- 952 Sarbhoy, A., Thapliyal, P.N., Payak, M.M., 1972. *Phakopsora pachyrhizi* on soybean
953 in India. Science and Culture, 38, 198.

- 954 Schneider, R.W., Hollier, C.A., Whitam, H.K., Palm, M.E., McKemy, J.M., Hernández,
955 J.R., Levy, L., DeVries-Paterson, R., 2005. First report of soybean rust caused by
956 *Phakopsora pachyrhizi* in the Continental United States. Plant Dis 89, 774–774.
957 <https://doi.org/10.1094/PD-89-0774A>
- 958 Sperschneider, J., Hewitt, T., Lewis, D.C., Periyannan, S., Milgate, A.W., Hickey,
959 L.T., Mago, R., Dodds, P.N., Figueroa, M., 2023. Nuclear exchange generates
960 population diversity in the wheat leaf rust pathogen *Puccinia triticina*. Nat Microbiol
961 8, 2130–2141. <https://doi.org/10.1038/s41564-023-01494-9>
- 962 Twizeyimana, M., Ojiambo, P.S., Haudenshield, J.S., Caetano-Anollés, G., Pedley,
963 K.F., Bandyopadhyay, R., Hartman, G.L., 2011. Genetic structure and diversity of
964 *Phakopsora pachyrhizi* isolates from soybean. Plant Pathol 60, 719–729.
965 <https://doi.org/10.1111/j.1365-3059.2011.02428.x>
- 966 Vittal, R., Yang, H.C., Hartman, G.L., 2012. Anastomosis of germ tubes and
967 migration of nuclei in germ tube networks of the soybean rust pathogen,
968 *Phakopsora pachyrhizi*. Eur J Plant Pathol 132, 163–167.
969 <https://doi.org/10.1007/s10658-011-9872-5>
- 970 Vonk, P.J., Ohm, R.A., 2018. The role of homeodomain transcription factors in fungal
971 development. Fungal Biology Reviews 32, 219–230.
972 <https://doi.org/10.1016/j.fbr.2018.04.002>.
- 973 Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of
974 Population Structure. Evolution 38, 1358. <https://doi.org/10.2307/2408641>

- Wickham, H., 2016. ggplot2. Springer International Publishing, Cham.
<https://doi.org/10.1007/978-3-319-24277-4>
- Yamanaka, N., Yamaoka, Y., Kato, M., Lemos, N.G., Passianotto, A.L. de L., Santos, J.V.M. dos, Benitez, E.R., Abdelnoor, R. V., Soares, R.M., Suenaga, K., 2010. Development of classification criteria for resistance to soybean rust and differences in virulence among Japanese and Brazilian rust populations. Trop Plant Pathol 35, 153–162. <https://doi.org/10.1590/S1982-56762010000300003>
- Yamaoka, Y., Fujiwara, Y., Kakishima, M., Katsuya, K., Yamada, K., Hagiwara, H., 2002. Pathogenic races of *Phakopsora pachyrhizi* on soybean and wild host plants collected in Japan. Journal of General Plant Pathology 68, 52–56. <https://doi.org/10.1007/PL00013053>
- Yorinori, J.T., Paiva, W.M., Frederick, R.D., Costamilan, L.M., Bertagnolli, P.F., Hartman, G.E., Godoy, C. V., Nunes, J., 2005. Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. Plant Dis 89, 675–677. <https://doi.org/10.1094/PD-89-0675>
- Zhang, X.C., Freire, M.C.M., Le, M.H., Oliveira, L.O. De, Pitkin, J.W., Segers, G., Concibido, V.C., Baley, G.J., Hartman, G.L., Upchurch, G., Pedley, K.F., Stacey, G., 2012. Genetic diversity and origins of *Phakopsora pachyrhizi* isolates in the United States. Asian J Plant Pathol 6, 52–65. <https://doi.org/10.3923/ajppaj.2012.52.65>

LEGEND

Fig. 1. Phylogenetic reconstruction and population structure in *Phakopsora pachyrhizi*. The analyses utilized whole-genome SNP data from 45 isolates collected from six geographic regions (Asia, Africa, Australia, Hawaii, South America, and the USA). **(A)** The maximum likelihood tree (unrooted) reveals two major clades, designated as lineage Pp1 and lineage Pp2. Branch lengths are proportional to the scale bar that corresponds to the expected number of substitutions per site. Bootstrap support values are displayed above the branches (when > 80). The arrow indicates the phylogenetic placement of the UFV02 reference genome. Isolate names are color-coded according to geographic regions, as indicated. **(B)** fastSTRUCTURE analysis points out the presence of two Bayesian clusters, as supported by optimal K number (K=2). Each colored horizontal bar represents the ancestry proportion of an isolate to its respective Bayesian cluster. **(C)** Plot of the first two principal components, with eigenvalues for each component shown in parentheses.

Fig. 2. Geographic distribution of two evolutionary lineages uncovered in *Phakopsora pachyrhizi*. Color-codes: red (lineage Pp1) and blue (lineage Pp2). Each pie diagram represents a single isolate. Pie diagrams containing both colors (red and blue) represent isolates that exhibited genetic admixture levels according to fastSTRUCTURE analysis shown in Fig. 1B.

Fig. 3. Molecular genetic diversity and Tajima's D values for two evolutionary lineages of *Phakopsora pachyrhizi*. The analyses involved SNPs localized on three distinct regions — whole-genome, effector gene, and non-effector gene — using the UFV02 isolate as a reference genome. **(A-B)** Violin plots showing the distribution of nucleotide diversity (π) and Watterson's theta (θ_w) across the whole-genome. **(C-H)** Violin plots showing the distribution of π , θ_w , and Tajima's D values for predicted effector and non-effector genes. Box plots were drawn inside the violin plots, with back circles representing the median. Statistical differences between the two lineages were based on the Mann-Whitney-Wilcoxon test (p-values are indicated).

Fig. 4. Mating-type (*MAT*) loci in *Phakopsora pachyrhizi*. **(A-B)** Schematic diagram showing the structure of *MAT* loci on the primary and secondary haplotypes of the UFV02 reference genome. The gene lengths were drawn proportional to the scale bar indicating the number of nucleotides in base pairs (pb). **(C-E)** Maximum likelihood phylogenies (unrooted tree) for three sets of *MAT* proteins in *P. pachyrhizi* isolates (UFV02, K8108, and MT2006), with *Puccinia* species as outgroups. **C.** *STE3* pheromone receptor. **(D)** Pheromone precursors (*MFA*). **(E)** Homeodomain (*HD*). Branch lengths are drawn to the scale bar that corresponds to the expected number of substitutions per site. Bootstrap support values are displayed above the branches (when > 80). Protein IDs are indicated in parentheses. Arrows represent the placements of *MAT* genes for the UFV02 isolate.

Fig. 5. Heatmap illustrating the virulence levels of *Phakopsora pachyrhizi* isolates from two major lineages (Pp1 and Pp2) on differential soybean genotypes. The x-axis shows eight soybean PI accessions with distinct *Rpp* genes: PI 200492 (*Rpp1*), PI 594538A (*Rpp1b*), PI 230970 (*Rpp2*), PI 462312 (*Rpp3*), PI 459025B (*Rpp4*), PI 200487 (*Rpp5*), PI 567129 (*Rpp6*), and PI 605823 (*Rpp7*). A susceptible genotype (BRS 184 or Wm82) was included as a positive control. The y-axis represents six *P. pachyrhizi* isolates: AU79-1, CO04-2, HW94-1, IN73-1, MT2006, TW72-1, and UFV02. Color intensity reflects the virulence level of each *P. pachyrhizi* isolate on the respective soybean genotype. Dendrograms are included to illustrate the hierarchical clustering of soybean genotypes (columns) and *P. pachyrhizi* isolates (rows).

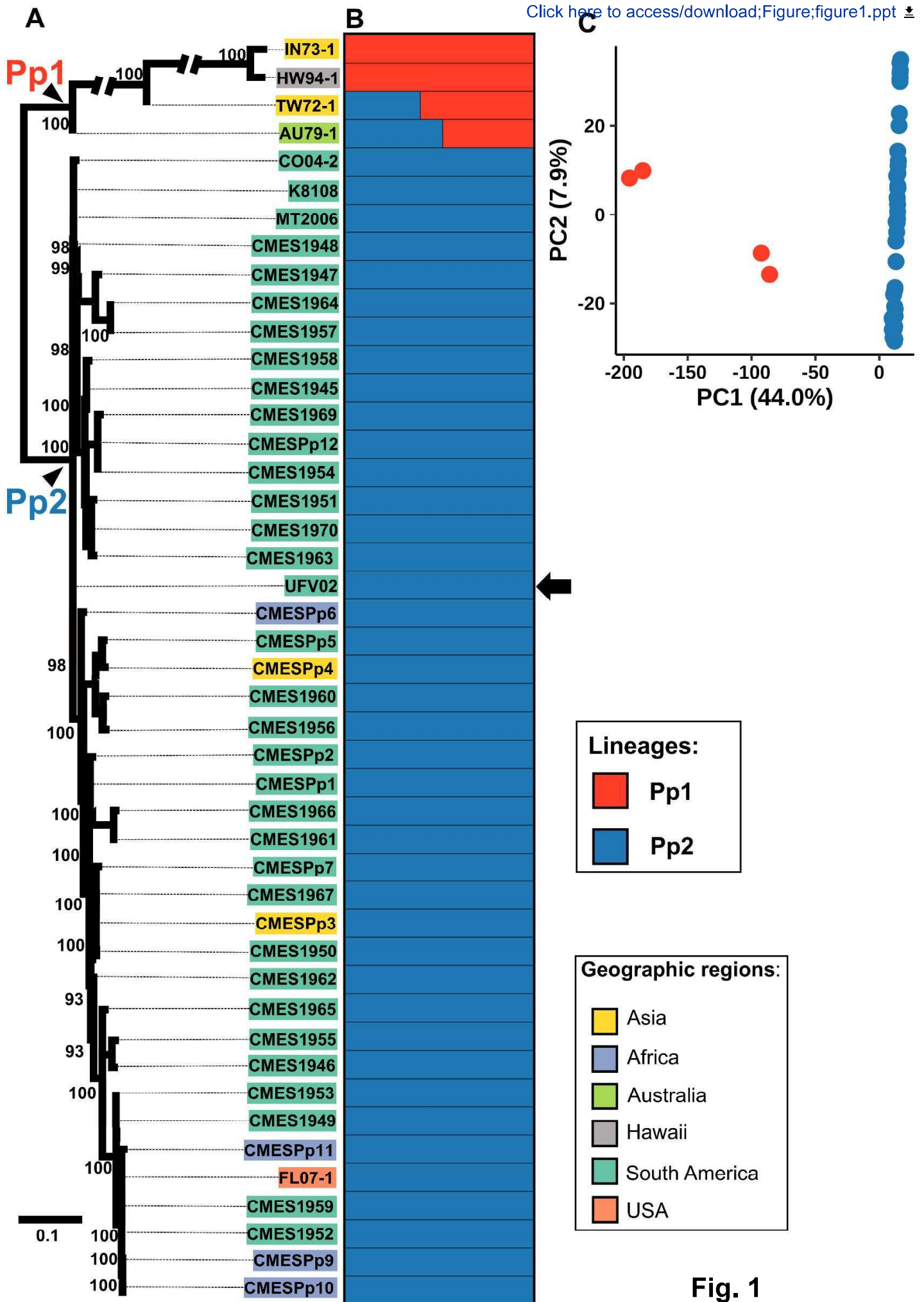


Fig. 1

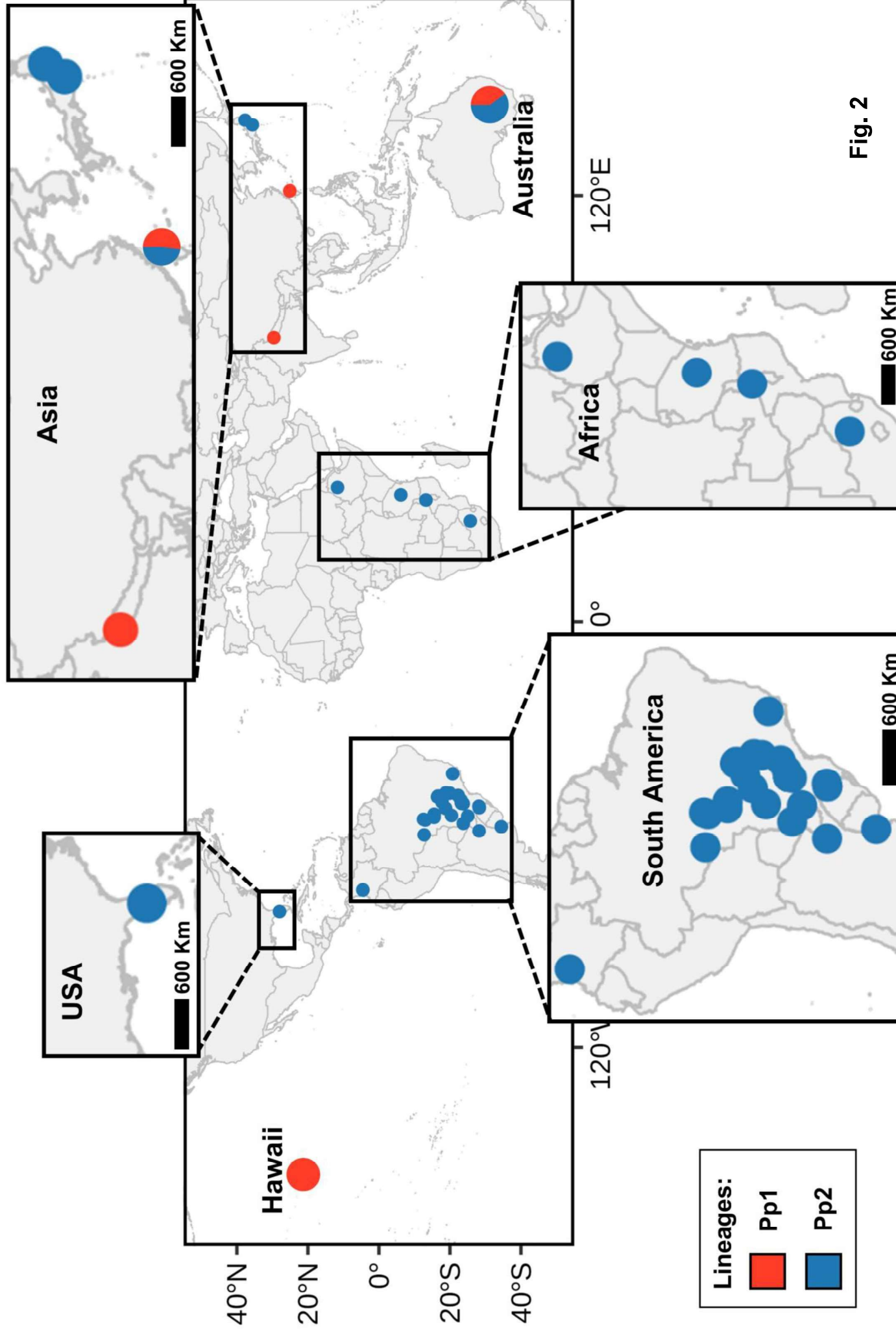


Fig. 2

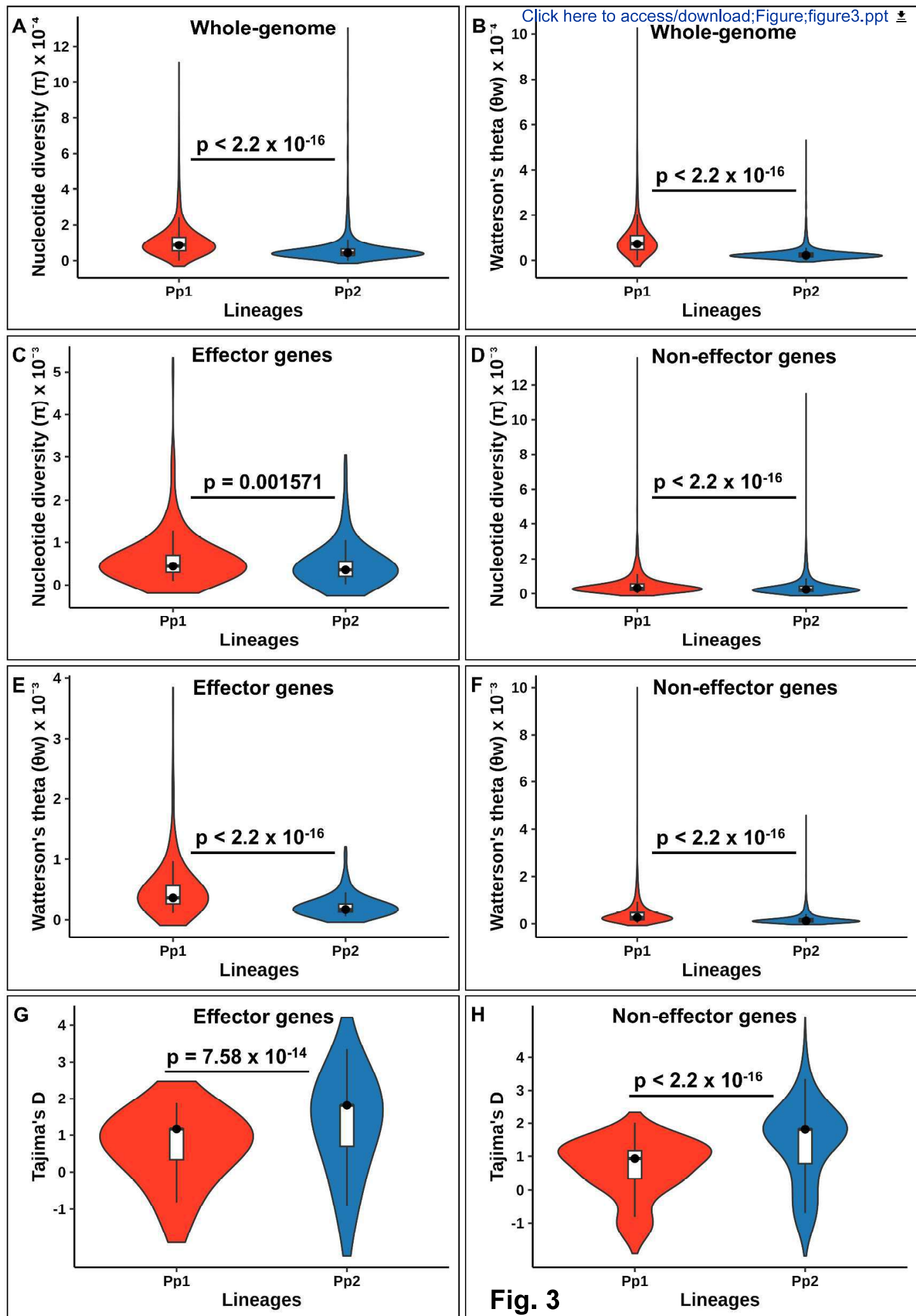


Fig. 3

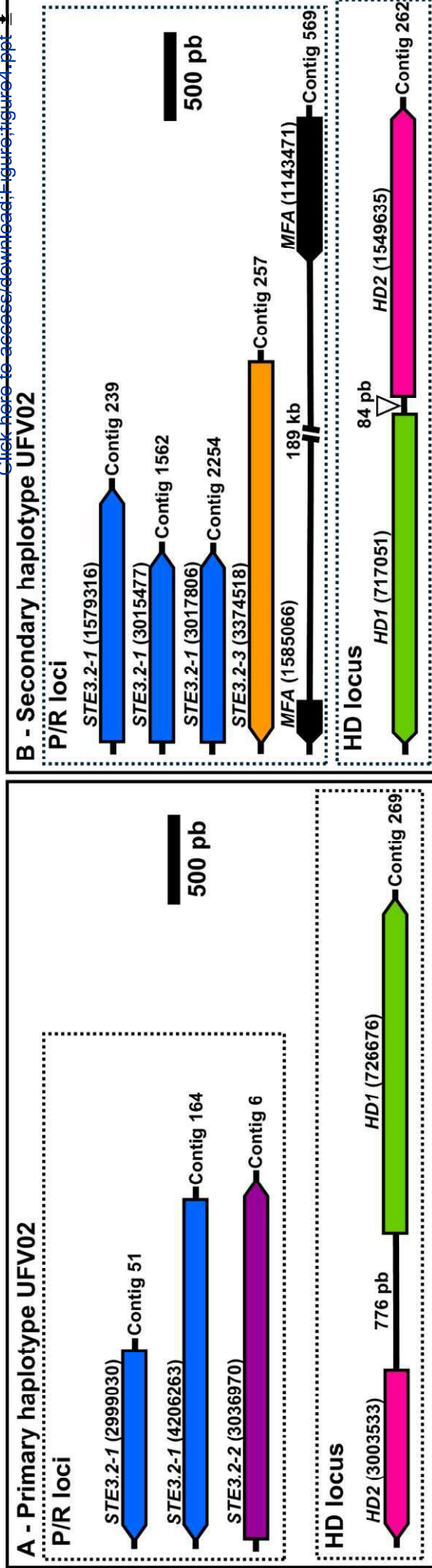
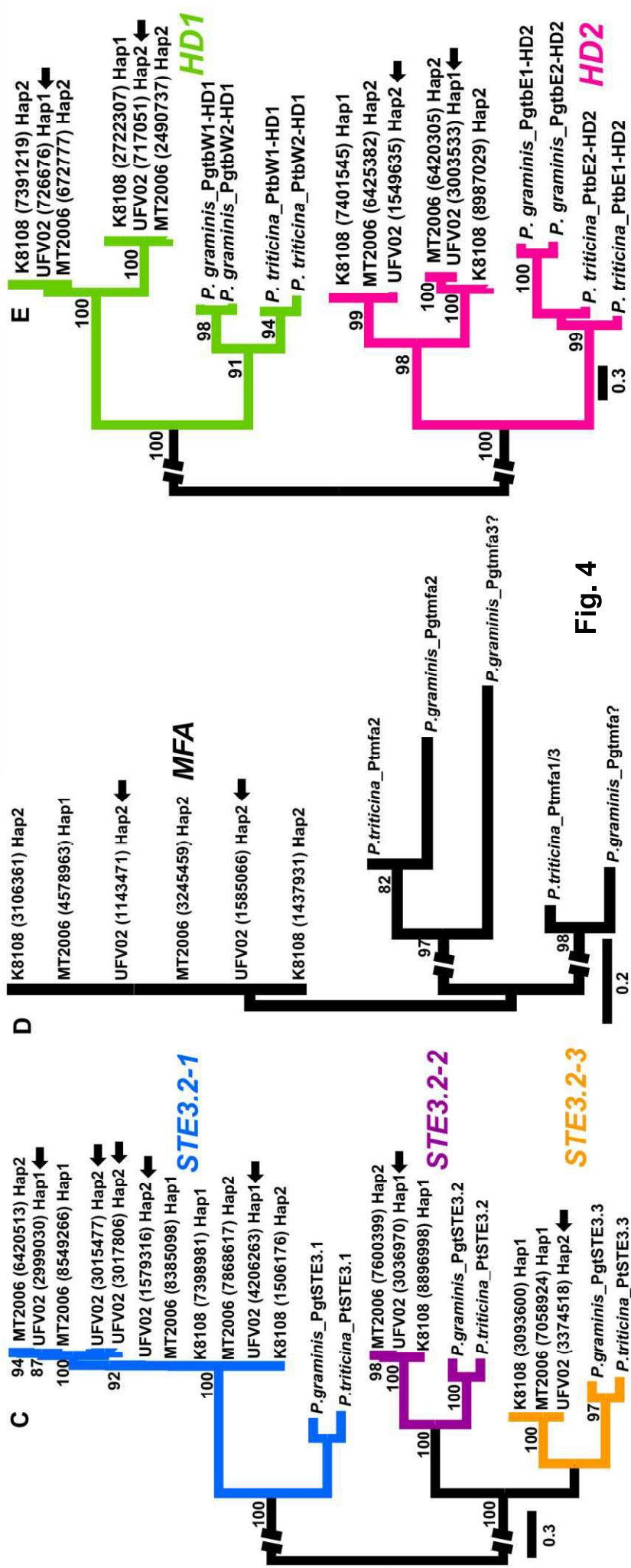


Fig. 4