A high-resolution haplotype pangenome uncovers somatic hybridization, recombination and intercontinental migration in oat crown rust

Eva C. Henningsen[1,2], David Lewis[1], Eric Nazareno[3], Yung-Fen Huang[4], Brian J. Steffenson[3], Brendan Boesen[1], Shahryar F. Kianian[3,5], Eric Stone[2], Peter N. Dodds[1], Jana Sperschneider[1], Melania Figueroa[1*]

[1]Commonwealth Scientific and Industrial Research Organisation, ACT Australia

[2]Australian National University, ACT Australia

[3]University of Minnesota, Saint Paul, MN U.S.A.

[4]National Taiwan University, Taipei, Taiwan

[5]Cereal Disease Laboratory, USDA-ARS, St. Paul, MN U.S.A.

*Corresponding author: M. Figueroa; E-mail: Melania.Figueroa@csiro.au

Keywords: haplotype, mating type, pangenome, *Puccinia coronata*, somatic hybridization, virulence

Funding

This project was supported by the Bioplatforms Australia Plant Pathogen 'Omics Initiative, GRDC project CSP2204 007RTX, USDA-NIFA BBSRC award 2022-67013-36505, the CSIRO Research Office, and grant 109-2313-B-002-028-MY3 from the National Science and Technology Council of Taiwan. ECH was supported by the ANU University Research Scholarship and ANU/CSIRO Digital Agriculture PhD Supplementary Scholarship.

Abstract

Basidiomycetes like rust fungi have complex reproductive cycles and dikaryotic life stages which influence their population structure and evolution. *Puccinia coronata* f. sp. *avenae* (*Pca*), the causal agent of oat crown rust, is a pathogen of global economic importance. To investigate the genetic diversity of the species, as well as the role of mating type system and nuclear exchange (somatic hybridization) in host adaptation of *Pca* we acquired whole genome sequencing data of Taiwanese and Australian isolates, adding to existing data for USA and South African populations. An atlas of 30 chromosome-level, fully-phased nuclear haplotypes from six USA isolates and nine Australian isolates was generated to capture the genomic composition of key oat crown rust lineages. This study provides evidence of nuclear exchange and recombination of haplotypes in both the USA and Australian *Pca* populations as mechanisms for the introduction of genetic diversity. The limitations of assuming clonal evolutionary history from virulence phenotyping is demonstrated by the detection of either sexual or cryptic genetic recombination in the *Pca* Australian population. Overall, the characterization of intercontinental migration of *Pca* at the haplotype level provides further impetus for molecular monitoring of rust pathogen populations on a global scale.

Introduction

Rust fungi are a diverse group of biotrophic plant pathogens belonging to the Basidiomycete order Pucciniales (Bauer *et al*., 2006; Aime & McTaggart, 2020). With over 7,000 species and diverse hosts, these organisms cause many significant agricultural diseases (Kirk *et al*., 2008). Notably, many rust fungal species such as *Puccinia coronata* f. sp. *avenae* (*Pca* - oat crown rust), *P. graminis* f. sp. *tritici* (*Pgt* – wheat stem rust), *P. striiformis* f. sp. *tritici* (*Pst* – wheat stripe rust), and *P. triticina* (*Pt* – wheat leaf rust) propagate as dikaryons (two haploid nuclei) while infecting their cereal hosts, which has complicated the identification of virulence factors and genomics informed surveillance (Petersen, 1974; Figueroa *et al*., 2020).

The flax – flax rust (*Melampsora lini*) pathosystem was the basis for Flor's pioneering work in plant pathology establishing the gene-for-gene model (Dodds, 2023), in which recognition of pathogen avirulence (Avr) effectors by plant resistance (*R*) proteins triggers signalling cascades and ultimately leads to cell death, an effective defence strategy against biotrophs (Jones & Dangl, 2006; Dodds & Rathjen, 2010; Chen *et al*., 2022; Dodds, 2023). However, understanding virulence evolution and host adaptation in rust populations has been hampered by their complex dikaryotic genome structures. The development of key molecular and bioinformatics technologies has resolved challenges posed by the ploidy changes in rust fungi through their complex lifecycles (Figueroa *et al*., 2022). Long-read sequencing and haplotype-aware assembly software enabled the first partial dissection of haplotypes to uncover high heterozygosity between nuclei of *Pca* and *Pst,* which highlighted the importance of nuclear phasing to create accurate rust genome references (Miller *et al*., 2018; Schwessinger *et al*., 2018). Further improvements in chromatin contact sequencing and phasing pipelines for rusts delivered the first chromosome-level and nuclear phased genomes for *Pgt*, *Pt*, and *Pca* (Li *et al*., 2019; Duan *et al*., 2022; Henningsen *et al*., 2022). In *Pgt*, this revealed that whole nuclear exchange without recombination precipitated the emergence of the devastating wheat stem rust Ug99 lineage (Li *et al*., 2019), leading to a paradigm shift that rust epidemiology must consider the movement of entire haplotypes. This has been made possible by the advent of high-fidelity long-read sequencing technology allowing the assembly of more and larger nuclear-phased

80 rust genome references (Duan *et al*., 2022). This has been explored recently in *Pt*,
81 where extensive somatic hybridization and intercontinental migration was uncovered
82 through comparative genomics (Sperschneider *et al*., 2023).
83
84 Although molecular resources for studying rust fungi continue to grow, well-
85 established areas of research for other plant pathogens remain poorly characterized
86 in these pathogens. For instance, about 90% of Basidiomycota require different mating
87 types to mate (heterothallic) (Kües *et al*., 2011), with compatibility controlled by two
88 mating type loci, *a* and *b*. These loci encode a pheromone and receptor at the *a* (*PR*)
89 locus and homeodomain transcription factor subunits at the *b* (*HD*) locus (Bölker *et*
90 *al*., 1992; Burkhard *et al*., 1990). Heterothallic fungi are further divided into two main
91 types; bipolar and tetrapolar, in which the *a* and *b* loci are linked and unlinked,
92 respectively. The presence and location of conserved mating type genes in the rust
93 species *Pgt*, *Pt*, and *Pst* suggests heterothallic and tetrapolar systems (Cuomo *et al*.,
94 2017; Holden *et al*., 2023), although the role of these presumed mating type loci has
95 not been functionally validated in rust fungi (Cuomo *et al*., 2017). It has been proposed
96 that this system also plays a role in somatic combability for nuclear exchanges to occur
97 between strains (Flor, 1964).
98
99 Oat crown rust disease caused by *Pca* results in significant yield losses worldwide and
100 is difficult to control due to rapid host adaptation by the pathogen (Nazareno *et al*.,
101 2018). Molecular evidence shows that sexuality is a clear driver of this rapid virulence
102 evolution in USA populations and influences population structure in Sweden (Berlin *et*
103 *al*., 2018; Miller *et al*., 2020; Hewitt *et al*., 2023). However, asexual *Pca* populations
104 have also been observed to gain virulence following *R* gene deployment (Nazareno *et*
105 *al*., 2018; Figueroa *et al*., 2023). Like for other rusts, key aspects of *Pca* biology and
106 epidemiology remain unclear.
107
108 The role of somatic hybridization in *Pca* virulence evolution cannot be assessed
109 without haplotype-phased genomes (Li *et al*., 2019; Sperschneider *et al*., 2023). There
110 is one fully phased, chromosome level assembly (Pca203) and two partially haplotype-
111 separated reference genomes from USA isolates (12SD80 and 12NC29) (Miller *et al*.,
112 2018; Henningsen *et al*., 2022). Given the strong genetic differentiation between USA

113 and South African isolates detected by Hewitt *et al*. (2023), the existing references are
114 unlikely to capture the genetic variation of *Pca* globally. Thus, there exists ripe
115 opportunity to address key questions for rust biology in *Pca* by further expanding
116 global resources for the pathosystem that allow genomic comparisons at the
117 haplotype-level. This study combines a haplotype-aware pangenome with the
118 extensive collection of publicly available short-read data for *Pca* to further our
119 understanding of the biology and epidemiology of *Pca* through the application of
120 contemporary genomic-based comparative approaches.
121

122 **Materials and Methods**

123 *Pca isolates, amplification, and phenotyping*

124 Australian isolates were collected from 2020 to 2023 across six states and territories
125 (ACT – Australian Capital Territory; NSW - New South Wales; QLD – Queensland; SA
126 – South Australia; VIC – Victoria; WA – Western Australia). Isolates from Taiwan (TW)
127 were collected from the College of Bioresources and Agriculture Experimental Farm
128 at the National Taiwan University in 2020 and 2021 (Taipei, Taiwan). Oat lines for
129 phenotyping were sourced from existing CSIRO seed stocks and the Australian Grains
130 Genebank (AGG) (Henningsen *et al*., 2024).
131 Differences between infection procedures for reviving field samples, single pustule
132 purification, and phenotyping are detailed in full in Methods S1. In all infections,
133 susceptible oat seedlings were grown at 23C° for 16 hours light and 18C° for 8 hours
134 dark. Plants were treated with 15 mL maleic hydrazide per pot at 9 days of growth
135 immediately prior to inoculation. Inoculations were conducted using urediniospores in
136 an oil or talc suspension and infected plants were kept in humidity chambers (90-99%
137 relative humidity) for two days before removal to growth chambers under the same
138 conditions as before. For phenotyping, infection types were recorded at 10-11 days
139 post inoculation (dpi), with the final score chosen to reflect the most prevalent infection
140 type across the biological replicates of the same differential line.
141

142 *DNA extraction and sequencing*

143 DNA extractions for Illumina sequencing were performed with the G-Biosciences
144 OmniPrep Genomic DNA isolation kit using 20-40 mg of rust spores as input. Libraries
145 were generated using either the Illumina DNA PCR-Free Prep or the IDT Prism library

146 preparation protocol depending on sample needs. Libraries were sequenced to 15-
147 30X depth, 150 bp paired end reads with Illumina Novaseq by Azenta or the Australian
148 Genome Research Facility (AGRF) (Table S1).
149 Extractions for high molecular weight (HMW) DNA from rust spores were completed
150 as described previously by Li *et al*. (2019). HMW DNA was used for PacBio HiFi library
151 preparation and sequencing to 30-50X depth at Azenta Life Sciences (formerly
152 Genewiz) through Phase Genomics (US isolates) or AGRF (Australian isolates) (Table
153 S2). Isolates 21ACT116, 20WA94, and 21WA134 were sequenced to 100-200X
154 coverage due to high yield from a single lane of PacBio Revio.
155 Spores for Hi-C were prepared as described by Sperschneider *et al*. (2023) and
156 detailed in Methods S2. Hi-C libraries were prepared at Phase Genomics, Seattle WA
157 USA and sequenced with Illumina Novaseq by Azenta Life Sciences (formerly
158 Genewiz) or were prepared and sequenced with Illumina Novaseq at the Ramaciotti
159 Centre for Genomics, NSW Australia (Methods S2).
160
161 *Genome assembly, phasing, scaffolding, and annotation*
162 Some contamination was apparent in the high-coverage HiFi read data for 20WA89,
163 21ACT116, and 21WA134 based on large and highly fragmented initial assemblies,
164 so these reads were filtered with mash v2.0 (Ondov *et al*., 2019) as described in
165 Methods S3. Raw reads for 20WA94 and 21WA139 and cleaned reads for 20WA89,
166 21ACT116, and 21WA134 were filtered on length with a 10 kb cutoff and were then
167 randomly down-sampled with seqkit (v2.7.0) sample to 30-50X coverage (Shen *et al*.,
168 2016). PacBio HiFi reads were assembled using Hifiasm with Hi-C integration and
169 resulting contigs were cleaned, phased, scaffolded, and annotated according to
170 Sperschneider *et al*. (2023) and described in detail in Methods S4. BUSCO values for
171 each haplotype were determined using compleasm v0.2.1 (Huang & Li, 2023) with the
172 odb10 Basidiomycota lineage dataset (Manni *et al*., 2021).
173
174 *Identification of putative mating type genes and population screening*
175 Protein sequences for *STE3.2.1*, *STE3.2.2*, *STE3.2.3*, *bW-HD1* and *bE-HD2* from *Pgt*
176 were BLASTed (tblastn) against the Pca203 genome reference with BLAST+ v2.13.0
177 (Camacho *et al*., 2009; Cuomo *et al*., 2017). Proteins overlapping the best hits were
178 compared to the *Pgt* alleles and examined for conserved domains known to be present

179 and intron/exon count before use in searching the other 30 haplotypes (Kües *et al*.,
180 2011; Cuomo *et al*., 2017; Methods S4). Protein sequences for *STE3.2*, *bW-HD1*, and
181 *bE-HD2* alleles were aligned with CLUSTALW, and phylogenetic trees were
182 constructed with the built-in 'raxml-bootstrap' option (Larkin *et al*., 2007). Protein trees
183 were rooted on alleles from *Pgt* CRL 75-36-700-3 (*Pgt STE3.2.1* and *Pgt bW-HD1*)
184 with R package 'ape' and visualized with iTOL (Cuomo *et al*., 2017; Paradis & Schliep,
185 2019; Letunic & Bork, 2021).
186 To screen the *Pca* population for the presence of the *STE3.2* and *HD* alleles, the 3' to
187 3' DNA sequence of the *HD* locus and 5' to 3' sequences of *STE3.2.2* and *STE3.2.3*
188 were extracted from each haplotype. These loci were sketched with mash v2.0 (-s
189 1000; Ondov *et al*., 2019) and then screened with the Illumina reads for 352 isolates
190 (Miller *et al*., 2018, 2020; Henningsen *et al*., 2022; Hewitt *et al*., 2023; Ho *et al*., 2024;
191 Table S1). Alleles were considered contained at or above 99% *k*-mer identity and 94%
192 shared *k*-mers.
193 To compare the repetitive regions on chromosome 9, haplotypes were aligned with
194 minimap2 v2.22 (-k19 -w19 -m200 -DP -r1000) and visualized with ggplot (Li, 2018;
195 Wickham, 2016).
196
197 *Population screening for haplotype containment*
198 The 352 *Pca* isolates with Illumina data (Table S1) were screened against the 32
199 available *Pca* haplotypes using mash v2.0 (Ondov *et al*., 2019). These haplotypes
200 were processed with mash sketch (-s 100000) and mash screen was run comparing
201 the Illumina data against the haplotype sketch file. Isolates with *k*-mer identity of ≥
202 99.98 % and shared *k*-mers ≥ 99.65% were considered as likely containing the
203 screened haplotype. Candidate hybrids were identified for having high identity to only
204 one haplotype, or as having high identity to two haplotypes which were not contained
205 within the same reference isolate.
206
207 *Phylogenetic analysis*

208 Illumina reads for 352 isolates (Table S1) were mapped to the diploid assemblies for
209 isolates Pca203, 90TX52, 20NSW19, 20QLD86, 20WA94, and 21ACT116 using bwa-
210 mem2 v2.2.1 (Vasimuddin *et al*., 2019). Variants were called against the diploid
211 assembly of Pca203 and individual haplotypes (3, 4, 5, 6, 13, 14, 23, 24, 25, 26), using

212 freebayes v1.3.6 (--use-best-n-alleles 6) (Garrison & Marth, 2012). Variants were
213 filtered with vcflib v1.0.1 (-f "QUAL > 20 & QUAL / AO > 10 & SAF > 0 & SAR > 0 &
214 RPR > 1 & RPL > 1 & AC > 0") and vcftools v0.1.16 (--min-alleles 2 --max-alleles 2 --
215 max-missing 0.9 --maf 0.05) (Danecek *et al*., 2011; Garrison *et al*., 2022). Variants
216 were converted to PHYLIP format using vcf2phylip (Ortiz, 2019). RAxML v8.2.12 was
217 used to construct the Maximum Likelihood (ML) tree with 500 bootstraps (-f a -m
218 GTRCAT -# 500 --no-bfgs) (Stamatakis, 2014). The resulting ML tree was visualized
219 with iTOL (Letunic & Bork, 2021).

220 Variants called against the Pca203 complete reference for the 137 Australian isolates
221 were filtered to include only biallelic single-nucleotide polymorphisms (SNPs). Variants
222 were converted to PHYLIP format with vcf2phylip (Ortiz, 2019), and heterozygous calls
223 were converted to missing (N) resulting in 391,118 SNPs. Splitstree CE v.6.2.1-beta
224 was used to calculate Hamming distances and generate a neighbor net with 301 splits
225 (Bryant & Huson, 2023; https://github.com/husonlab/splitstree6). The network was
226 evaluated with the phi test for recombination in Splitstree CE v6.2.1-beta (Bruen *et al*.,
227 2005).
228
229 *Comparison of pathotype clustering and phylogenetic tree structures*
230 Infection types were converted to a binary system of 0 (avirulent, infection types 0 to
231 2) and 1 (virulent, infection types 3 to 4) and clustered with R v4.3.2 hclust() before
232 conversion to a dendrogram (R Core Team, 2023). The Pca203 ML tree was pruned
233 to contain only phenotyped Australian isolates with iTOL (Letunic & Bork, 2021) and
234 was modified to be binary and ultrametric, with R package 'ape' before conversion to
235 a dendrogram (Paradis & Schliep, 2019). The two dendrograms were compared with
236 untangle_step_rotate_1side() from 'dendextend', with the phenotype clustering
237 dendrogram being rotated to find the best structural match to the RAxML-derived
238 dendrogram (Galili, 2015). The comparison was visualized with
239 tanglegram(rank_branches = TRUE) from 'dendextend' to visualize topology without
240 considering branch lengths (Galili, 2015). Finally, the two dendrograms were
241 compared with cor_bakers_gamma() from 'dendextend' to calculate the correlation
242 between the tree structures (Galili, 2015).
243
244 *Haplotype comparisons*

245 Nucmer and DNAdiff from mummer v4.0.0 were run to determine the identity between
246 all nuclear haplotype pairs (Kurtz *et al*., 2004; Marçais *et al*., 2018). Haplotypes were
247 aligned with D-Genies v1.5.0 and visualized with ggplot2 (Wickham, 2016; Cabanettes
248 & Klopp, 2018). The cactus-pangenome pipeline from cactus v2.6.6 was run on all
249 haplotypes and variants were called for each haplotype as the reference (Hickey *et*
250 *al*., 2023). The full procedure for determining shared haplotype blocks is detailed in
251 Methods S5. Briefly, variant counts were binned and a cutoff of <= 50 non-ref SNPs
252 per 100kb was applied. Adjacent shared regions for each sample-reference pair were
253 merged, subjected to an ordered bedtools (v2.31.1; Quinlan & Hall, 2010) subtraction,
254 and regions <= 50 Kb were removed before visualization.
255
256 **Results**
257 *Contextualization of Australian and Taiwanese Pca isolates within the global*
258 *population*
259
260 From 2020 to 2023, 137 Australian *Pca* isolates were collected through community
261 submissions across six Australian states and territories (ACT = 9, NSW = 36, QLD =
262 9, WA = 55, SA = 10, VIC = 18), and isolates were recovered from both wild (*n*=111)
263 and cultivated oats (*n*=26) (Table S1). Furthermore, four Taiwanese isolates were
264 collected from cultivated oat in 2020 and 2021. Whole genome sequence (WGS) data
265 (Illumina) was generated from genomic DNA from all isolates (42X average genome
266 coverage; Table S1). We generated a ML phylogenetic tree using the WGS data of all
267 isolates and including published data of 211 isolates from the USA, South Africa, and
268 Taiwan (total *n*=352) (Miller *et al*., 2018, 2020; Henningsen *et al*., 2022; Hewitt *et al*.,
269 2023; Ho *et al*., 2024). Reads were mapped to both haplotypes of the Pca203 genome
270 reference (Henningsen *et al*., 2022) and filtered variants (*n*=376,646) were used to
271 construct the ML phylogenetic tree.
272 A total of 18 lineages were detected in the Australian *Pca* collection, which is
273 unexpected as the population is proposed to consist of only four asexually-reproducing
274 (clonal) genetic groups (Fig. **2a**; Park *et al*., 2022). Although clonality is influential in
275 Australia as shown by the size and temporal persistence of the largest lineages (L18
276 – 48 individuals, L1 – 35 individuals), nine lineages consist of only one isolate. Further,
277 only seven lineages were sampled across multiple years (4 years – L1, L9, L18; 3

278 years – L2, L11; 2 years – L5, L16), suggesting there are more lineages to be sampled
279 (Fig. **2a**). Lineage diversity was greatest in WA (*n*=15), followed by SA (*n*=5),
280 NSW/ACT (*n*=4), VIC (*n*=3), and QLD (*n*=2; Fig. **2b**). The five most abundant lineages
281 (L1, L3, L9, L11, L18) were sampled on both wild and cultivated oat, suggesting that
282 the populations on these different host types are not separate (Table S1). Only one
283 lineage was detected from the four Taiwanese isolates, which are clones of isolate
284 NTU-01.
285 To assess the robustness of virulence phenotypes as the basis for lineage assignment
286 within the Australian *Pca* population, 95 of the 137 sequenced *Pca* isolates were
287 tested on 27 genotypes from the Australian oat crown rust differential set (Henningsen
288 *et al*., 2024; Fig. S2; Table S3). Most isolates in lineage L1 (20ACT25, 22NSW08)
289 displayed broad virulence across the differential lines, while others had virulence to
290 fewer lines (22QLD110, 22NSW107) (Fig. S2). Lineage L2 had the least virulence,
291 with 21ACT116 and 23NSW13 only defeating resistance for two lines, Swan (carrying
292 *Pc1*) and Pc70 (also known as H547). The remaining lineages (L3-L18) share many
293 virulence traits. For example, isolates 22WA31 and 22WA54 from lineages L11 and
294 L5 had nearly identical reactions. Clustering isolates by these phenotypic results is
295 inconsistent with the phylogenetic relationships (Baker's Gamma = 0.64, Fig. S3).
296 Thus, virulence phenotypes (pathotypes) are poor predictors of lineage relationships
297 in the Australian population.
298
299 *Construction of a haplotype-aware pangenome of Pca*
300
301 From the phylogenetic analysis, we selected representative *Pca* isolates to generate
302 haplotype-phased chromosome-level genome references that capture genetic
303 diversity in Australia (*n*=9) and USA (*n*=6) (Fig. **S4a,b**). References were generated
304 from eight Australian clonal groups (20QLD86 – L1, 21ACT116 – L2, 20WA72 and
305 20WA95 – L9, 20WA94 – L11, 21WA139 – L13, 20WA89 – L14, 21WA134 – L16,
306 20NSW19 – L18), with lineage L9 sampled twice (20WA72, 20WA95) as a baseline
307 for comparing near-identical haplotypes (Fig. **S4a**). We also chose three contemporary
308 buckthorn-derived isolates (18MNBT34, 18MNBT36, 18MNBT50) from the USA
309 sexual population (Hewitt *et al*., 2023; Fig. **S4b**) and three isolates from USA clonal
310 lineages (90TX52, 90AR100, 90MN4B; Miller *et al*., 2018; Fig. **S4a,b**).

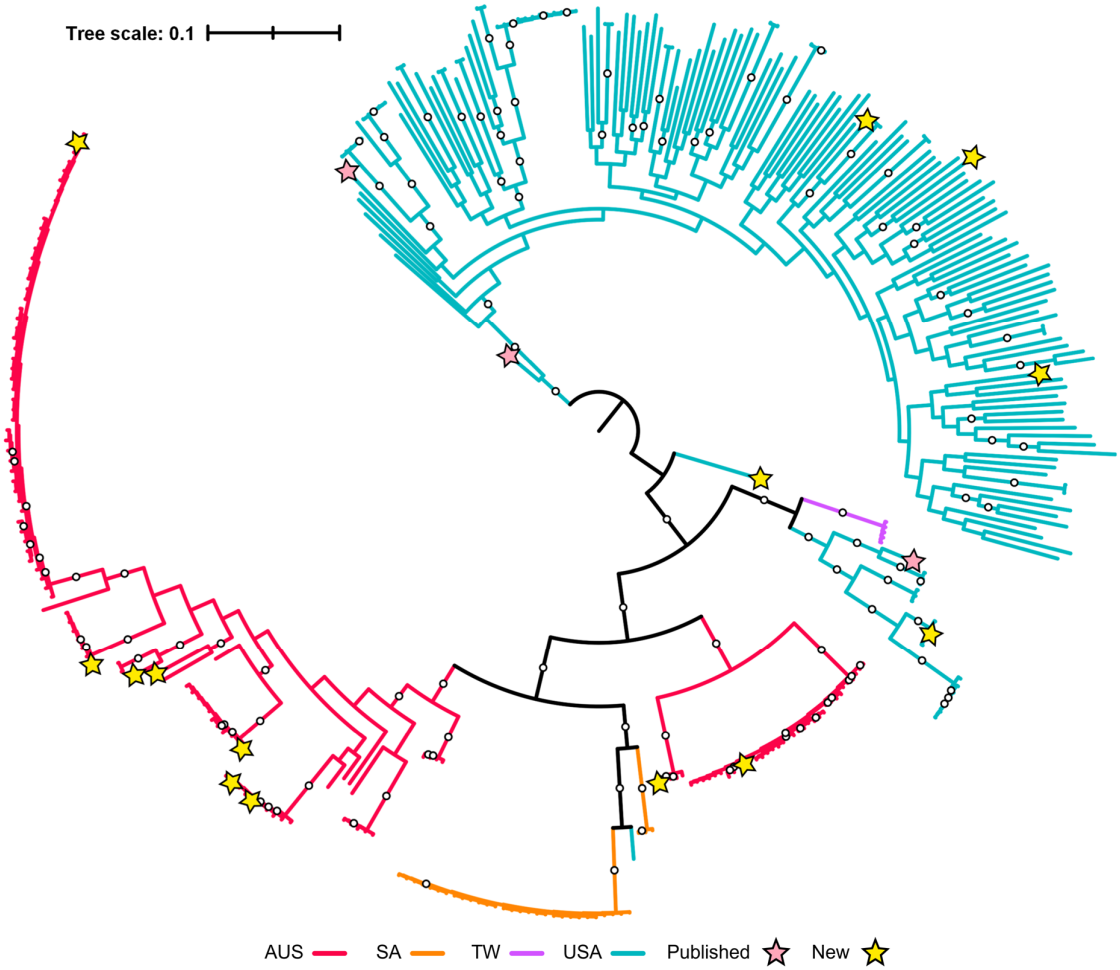


Figure 1. Midpoint rooted ML phylogenetic tree constructed by mapping reads from 352 *Puccinia coronata* f. sp. *avenae* (*Pca*) isolates and calling variants against the full Pca203 reference (hap1, hap2, and unplaced contigs). A total of 376,646 biallelic SNPs and 500 bootstraps were used for this analysis. Tree branches are colored by country of origin: AUS = Australia; SA = South Africa; TW = Taiwan; USA = United States of America. Bootstrap values above 80% shown as white circles. Yellow stars indicate isolates chosen for the haplotype atlas and red stars indicate isolates with existing references. Tree scale is mean substitutions per site.

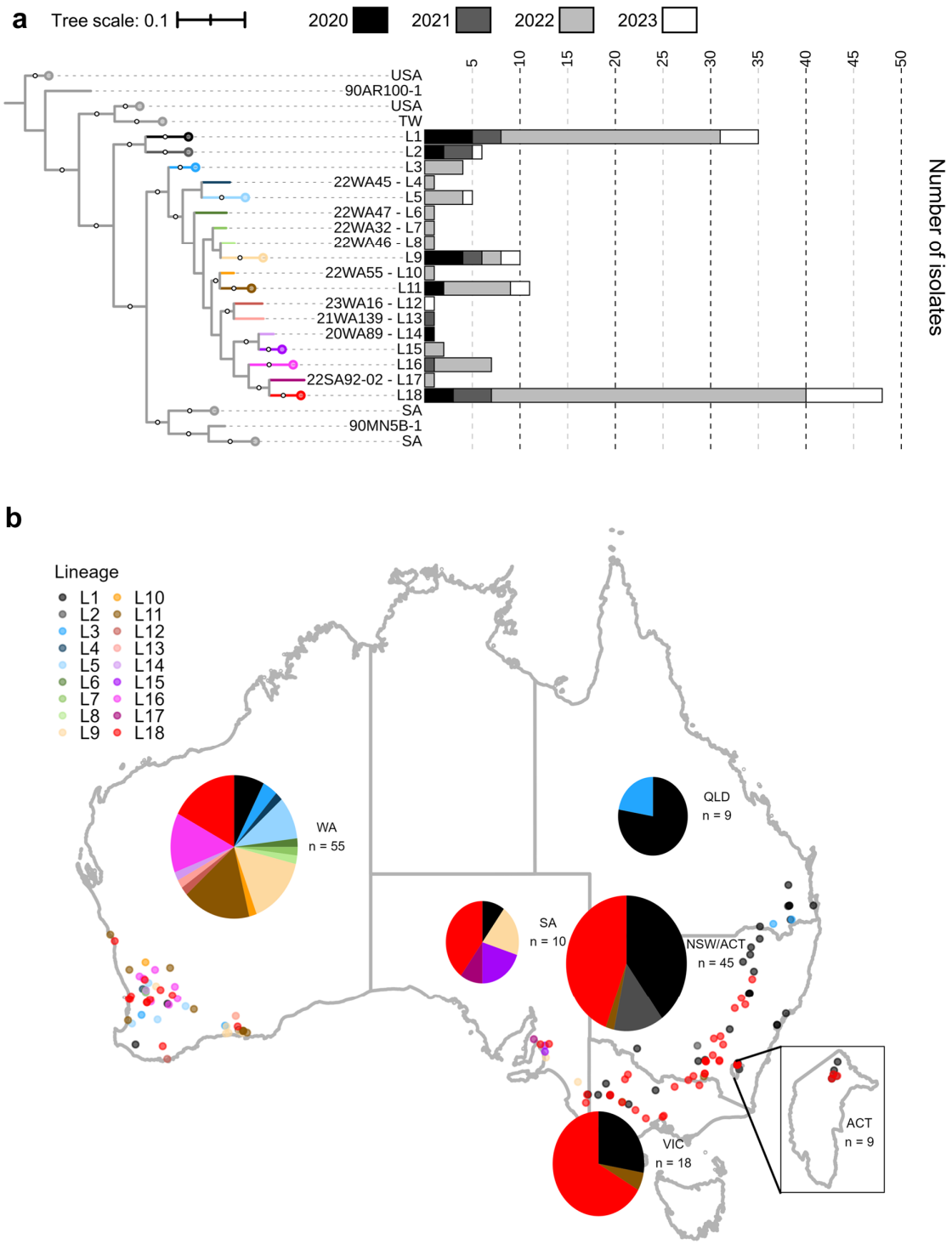


Figure 2. **a)** View of Australian lineages with clones collapsed in the midpoint rooted ML phylogenetic tree constructed for 352 *Puccinia coronata* f. sp. *avenae* (*Pca*) isolates based on 376,646 biallelic SNPs called against the full Pca203 reference. Bootstrap values (500 cycles) above 80% shown as white circles. Tree scale is mean substitutions per site. Stacked barplots show the number of isolates of each lineage

327 collected in 2020 (black), 2021 (dark grey), 2022 (light grey), and 2023 (white) across
328 Australian lineages. Tree branches for Australian lineage L1-L18 are colored by
329 lineage to match points and pie charts in panel **b)** Distribution of *Pca* isolates and
330 lineages across Australia. Pie charts are scaled on a logarithmic scale according to
331 the total number of *Pca* isolates collected in their respective states. Pie chart for
332 NSW/ACT includes isolates from both regions due to the small area of the ACT relative
333 to NSW. Colors reflect lineage assignments.
334
335 We assembled nuclear phased genome references of 15 isolates following established
336 computational workflows using hifiasm with incorporated Hi-C chromatin contact data
337 (Duan *et al*., 2022; Henningsen *et al*., 2022; Sperschneider *et al*., 2023), resulting in
338 30 complete nuclear haplotypes (hap3-hap32; Fig. **1**, see yellow stars; Fig. **3a**; Table
339 S2). Nuclearphaser detected only five phase swaps (in three isolates) that required
340 correction. Four within-haplotype mis-assemblies were also corrected in hap5, hap9,
341 hap10, and hap19. Scaffolding resulted in 18 chromosomes in each haplotype (chr1
342 to chr18; Fig. **3b**). An average of 84.44% of *trans* and 96.75% of *cis* and *trans* Hi-C
343 contacts occurred on chromosomes within the same haplotype which reflects accurate
344 haplotype separation and nuclear assignment. *Pca* haplotypes were on average 98.4
345 Mb long with 42.7% repeat coverage and 19,700 gene annotations (Table S4).
346 Haplotype BUSCO completeness averaged 94.85% with less than 2% duplication,
347 which is comparable to haplotype-phased assemblies for *Pgt* and *Pt* (Table S4; Li *et*
348 *al*., 2019; Duan *et al*., 2022; Sperschneider *et al*., 2023). Haplotypes were consistent
349 for most genome characteristics (Table S4). Chromosome sizes were likewise
350 consistent except for chr9, which harbours a large repetitive region of approximately
351 0.1 to 1.4 Mb (Fig. **3b,c**). Most assemblies break within this locus, so the exact size of
352 the region is not certain.
353 Pairwise haplotype comparisons validated the clonal relationship between 20WA72
354 and 20WA95 from L9, as their matching haplotypes (hap7 ≈ hap22; hap8 ≈ hap21)
355 had very high sequence identity (99.98%). In contrast, comparison of the two
356 haplotype sequences within any isolate showed high levels of divergence and
357 heterozygosity (98.73 - 99.45% identity), with most inter-isolate haplotype
358 comparisons similarly diverse (98.65 - 99.98% identity). The only exception to this was
359 the haplotype pair hap5 (20QLD86 from L1) and hap24 (21ACT116 from L2) which

360 showed high 99.94% sequence identity (99.78-99.86% alignment coverage, 9,228
361 SNPs) comparable to haplotypes in the clonal 20WA72 and 20WA95 isolates (99.44-
362 99.97% alignment coverage, 99.98% identity, 1,051-1,701 SNPs; Fig. **3d**). As the
363 other haplotypes from 20QLD86 (hap6) and 21ACT116 (hap23) were extremely
364 dissimilar (98.39% identity, 96.17-97.85% alignment coverage, 481,662 SNPs), it is
365 likely that lineages L1 and L2 are related through somatic hybridization and nuclear
366 exchange (Fig. **3d**).
367
368 *Characterization of mating loci in the Pca pangenome and overall population*
369
370 To better understand the role of putative *a* and *b* mating type loci in *Pca*, we
371 characterized the *bW-HD1/bE-HD2* and *STE3.2* loci using sequences identified in *Pgt*
372 (Cuomo *et al*., 2017). Pca203 has one copy of *STE3.2.3* on chr9A and one copy of
373 *STE3.2.2* on chr9B. Similarly, all isolates in the *Pca* pangenome carry one near-
374 identical copy each of *STE3.2.2* and *STE3.2.3* on opposite haplotypes of chromosome
375 9 (Fig. S**5a,b**). The *mfa* pheromone precursors were also identified close to these
376 *STE3.2* alleles (Fig. **S5a**). The related gene *STE3.2.1* was invariant and present in a
377 single copy on chr1 in all haplotypes except in hap11 (90MN4B) which had two copies,
378 supporting previous findings that it is unlikely to be involved in mating compatibility
379 (Cuomo *et al*., 2017). The *STE3.2.2* and *STE3.2.3* alleles are adjacent to the large
380 repetitive region on chr9 which is allele-specific, explaining the unusual chr9 length
381 distribution (Fig. **3b**, Fig. **S5c,d**). We then characterized the *bW-HD1 bE-HD2* (*HD*)
382 allele pairs across haplotypes in the pangenome. All isolates contained different *HD*
383 alleles in each haplotype on chr4 and a total of 13 alleles were recorded (Fig**. S5e,f**).
384 Using these characterized loci, we screened Illumina short reads from the entire *Pca*
385 collection (*n*=352) and found that all isolates contain both *STE3.2.2* and *STE3.2.3*
386 alleles and two different alleles at the *HD* locus (Fig. **S6a**, Table S5, Table S6). In *Pca*
387 isolates where one or both *HD* alleles were not characterized in the pangenome, short
388 read mapping still indicates heterozygosity at the *HD* locus (Fig. S7). The observed
389 genotype frequency for both the *PR* and *HD* loci in the USA sexual population (*n*=88
390 isolates derived as aecia from the sexual host buckthorn in the UMN nursery) are 0%
391 homozygous and 100% heterozygous (*PR* locus alleles *n*=2, *HD* locus alleles *n*=11).
392 This is a significant departure from the expected 50:50 (homozygous:heterozygous)

393 ratio for the biallelic *PR* locus in Hardy-Weinberg equilibrium (p = 6.55 x 10^{-21}).
394 Likewise, the *HD* locus genotype ratio is significantly different than expected (12%
395 homozygous, 88% heterozygous; p = 5.70 x 10^{-4}). This suggests that both loci
396 contribute to mating type compatibility resulting in exclusive heterozygosity, meaning
397 that *Pca* likely has a tetrapolar mating system given that the *PR* and *HD* loci are on
398 separate chromosomes.
399

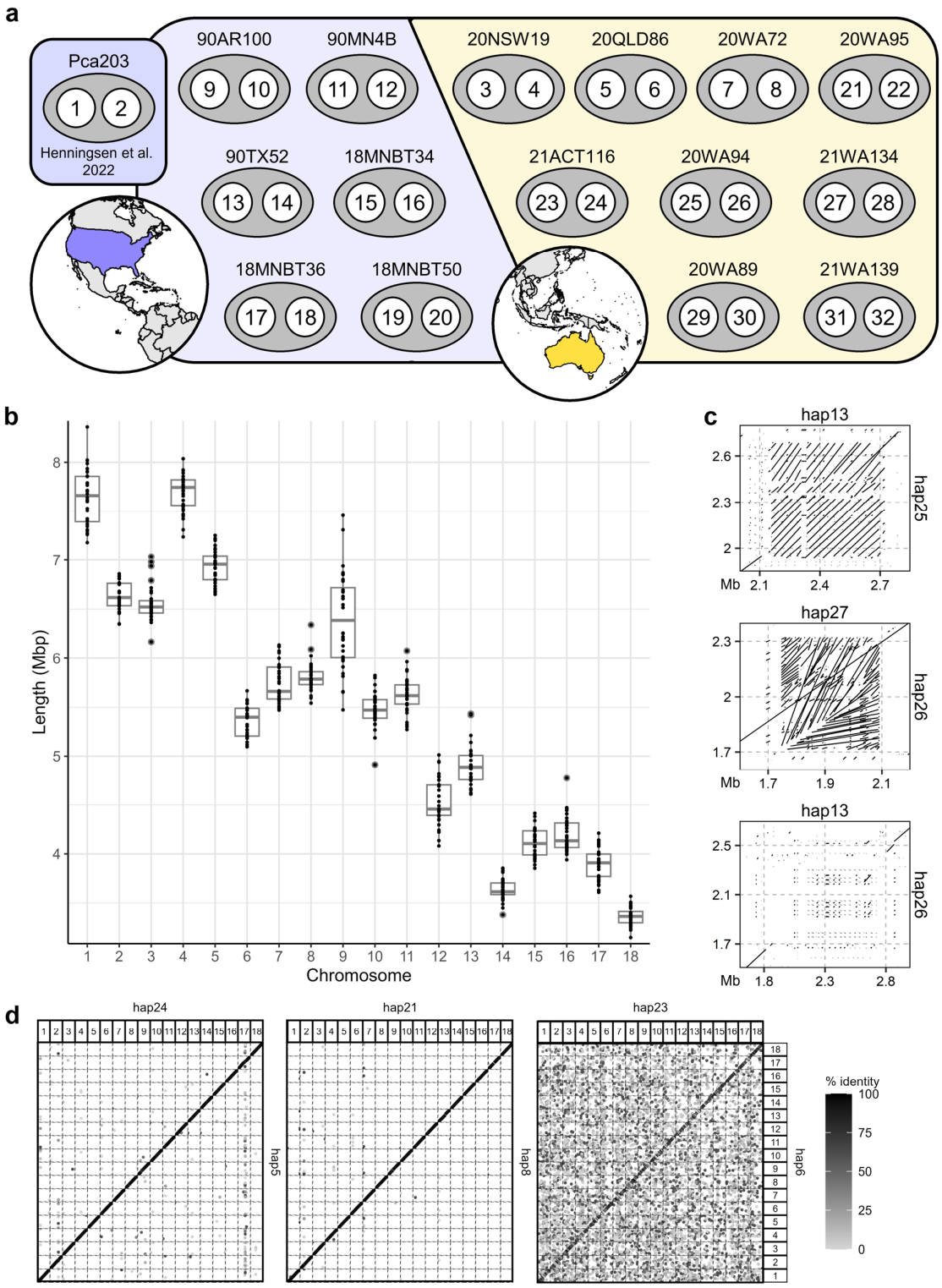


Figure 3. **a**) Haplotype numbers assigned to nuclear genotypes of 15 *Puccinia coronata* f. sp. *avenae* (*Pca*) isolates from the USA (purple) and Australia (yellow), in addition to previously published haplotypes from Pca203 (1 = “A”; 2 = “B”). **b**) Boxplots of chromosome lengths for the 32 *Pca* haplotypes with black dots indicating

405 chromosome size of each haplotype. Box top and bottom boundaries are the upper
406 and lower quartiles, respectively. Lines within boxes represent the mean. Lines
407 extending below and above boxes delimit minimum and maximum values. **c**) In
408 descending order: alignments of chromosome 9 regions containing STE3.2 genes
409 between hap13 and hap25 (*STE3.2.2*), hap26 and hap27 (*STE3.2.3*), and hap13 and
410 hap26 containing different *STE3.2* alleles. **d**) Dotplot of genome alignment between
411 *Puccinia coronata* f. sp. *avenae* (*Pca*) haplotypes hap5 and hap24 (left), hap8 and
412 hap21 (middle) and hap6 and hap23 (right). Chromosomes are numbered along the
413 x- and y-axis. Greyscale indicates percent identity along alignment.

414 *Pca isolates from the USA, Australia, and Taiwan are related through somatic*
415 *hybridization and migration*
416
417 To assess the role of somatic hybridization in the evolution of *Pca* we employed *k*-mer
418 containment, phylogenetics, and comparative genomics approaches used to identify
419 somatic hybridization events in other rust species (Li *et al*., 2019; Sperschneider *et al*.,
420 2023). In our analysis, *k*-mer identity of ≥ 99.98 % and shared *k*-mers ≥ 99.65%
421 accurately identified clones of the *Pca* reference isolates. For example, L18 isolate
422 20NSW20 has high *k*-mer containment for hap3 and hap4 from the L18 reference
423 20NSW19 (100% *k*-mer identity and ≥ 99.93% shared *k*-mers). As established in *Pgt*
424 and *Pt*, phylogenetic trees constructed from an isolate's entire genome show clonal
425 lineages as discrete clades, while single-haplotype trees merge lineages that share
426 the reference haplotype into a single clade (Li *et al*., 2019; Sperschneider *et al*., 2023).
427 Multiple putative nuclear exchange events were identified across the entire *Pca*
428 dataset using these criteria (Table S1; Table S7).
429 The full haplotype comparisons we described earlier are the strongest evidence for
430 somatic hybridization between L1 and L2 and *k*-mer containment and phylogenetic
431 tree results also support this hypothesis. L1 isolates had high containment for hap5,
432 hap6, and hap24 (> 99.98% identity, > 99.75% shared *k*-mers), but not hap23 (<
433 99.86% identity, < 96.99% shared *k*-mers) (Fig. **4a**; Table S7). Conversely, L2 isolates
434 had high containment for hap23, hap24, and hap5 (> 99.98% identity, > 99.82%
435 shared *k*-mers) and low containment for hap6 (< 99.64% identity, < 92.50% shared *k*-
436 mers) (Fig. **4a**; Table S7). Further, trees constructed from these haplotypes
437 individually support hybridization between L1 and L2. Phylogenies generated from
438 hap5 and hap24 both show L1 and L2 isolates as members of a single clade, while
439 those for hap6 and hap23 show L1 and L2 as discrete clades (Fig. **4b-e**). These
440 analyses also suggest a somatic hybridization event linking Australia and Taiwan,
441 implicating direct or indirect migration of *Pca* (Fig. **4f**). The five Taiwanese isolates (L-
442 TW) have high *k*-mer containment for hap6 (99.99% identity, > 99.84% shared *k*-mers)
443 but not hap5 (< 99.70% identity, < 93.65% shared *k*-mers) (Fig. **4a**; Table S7). This is
444 also supported by the phylogenetic trees, as L1 and L-TW isolates comprise separate
445 clades in the hap5 tree but cluster together in the hap6 tree (Fig. **4b,d**). Since we did

446 not sample the other haplotype for L-TW in the current pangenome, we cannot infer
447 the history of haplotype exchange.
448
449 The *k*-mer containment and phylogenetic analyses also support the occurrence of
450 somatic hybridization and nuclear exchange within the USA *Pca* population. Members
451 of the 90TX52 lineage (L-1990) have high containment for both 90TX52 haplotypes
452 (100% identity, > 99.92% shared *k*-mers) (Fig. **S8a**; Table S7). Seven clonal isolates
453 collected in 2017 (L-2017) from the southern USA (LA, TX, FL) contain hap14 (100%
454 identity, > 99.89% shared *k*-mers), but not hap13 (< 99.85% identity, < 96.65% shared
455 *k*-mers) (Fig. **S8a**; Table S7). Phylogenetic trees constructed separately from hap13
456 and hap14 SNPs further support somatic hybridization between L-1990 and L-2017,
457 as they appear as one clade in the hap14 phylogeny and separate clades in the hap13
458 phylogeny (Fig. **S8b,c**). As the second haplotype of the 2017 isolates was not sampled
459 in the current pangenome, we cannot establish the history of hap14 inheritance (e.g.,
460 donor in the hybridization event) (Fig. **S8d**).
461 Mating allele composition of putative hybrid isolates agrees with the postulated nuclear
462 exchange events, as somatic hybrid isolates have the same *HD* locus allele as the
463 haplotype they are proposed to contain (Fig. **4b-e**; Fig. **S8b,c**; Table S5). The
464 assemblies for hap5 and hap24 likewise have the same alleles at both mating loci
465 (*STE3.2.2* and *bW3bE3*; Fig. S5). The other pangenome haplotypes do not appear to
466 be related through somatic hybridization events to *Pca* isolates in the existing
467 collection. However, 90MN14B and 90MN7B were identified as likely clones of
468 90MN4B (100% identity, > 99.92% shared *k*-mers) for which short-read data is not
469 available (Table S7). Altogether, our results support the occurrence of somatic
470 hybridization in the broad *Pca* population.

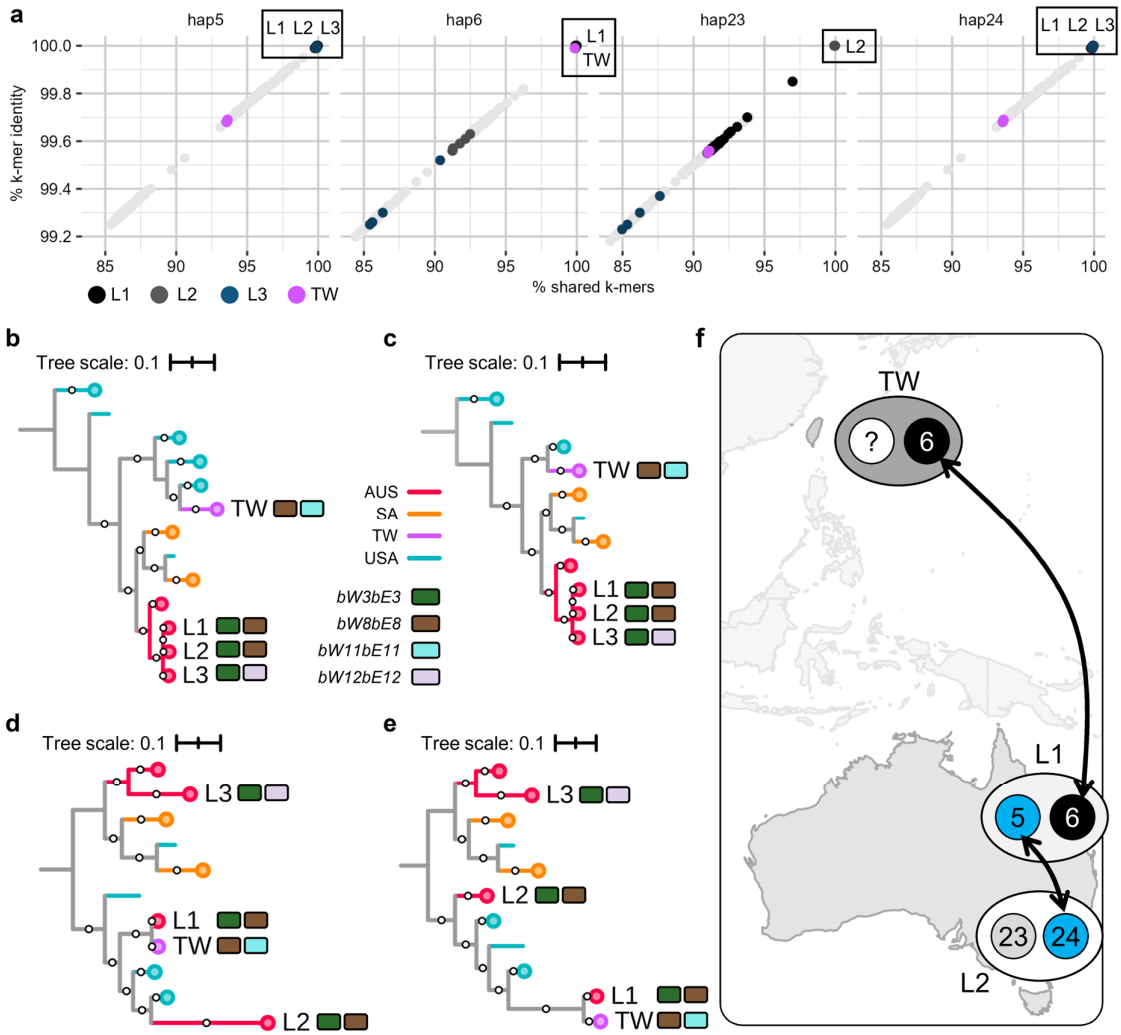


Figure 4. **a)** Plot of % shared *k*-mers (x-axis) versus % *k*-mer identity (y-axis) for 352 *Pca* isolates on hap5, hap6, hap23, and hap24. Colors indicate relevant lineages (L1-L3, TW =Taiwan); light grey points are all other isolates. **b-e)** Maximum likelihood trees for **b)** hap5 (225,727 SNPs) **c)** hap24 (239,896 SNPs) **d)** hap6 (198,467 SNPs) and **e)** hap23 (191,316 SNPs). Tree branches are colored by country of origin: AUS = Australia; SA = South Africa; TW = Taiwan; USA = United States of America. 500 bootstraps were generated for each tree and bootstraps 80% and higher are shown as white circles on branches. Collapsed branches are indicated with circles at leaf tips. Mating type alleles for hybrid lineages are shown as rectangles to the right of branches. Tree scales are mean substitutions per site. **f)** Proposed relationships between *Pca* lineages and haplotypes overlaid on a map of Oceania and Asia.

483 *The Australian and USA Pca populations are shaped by genetic recombination*

484 The *k*-mer containment analysis also yielded some results that were incompatible with
485 the hypothesis of nuclear exchanges. For instance, L18 isolates had high containment
486 for hap3 and hap4 (L18) and hap25 (L11) (100% identity, > 99.91% shared *k*-mers)
487 but not hap26 (L11) (< 99.94% identity, < 98.56% shared *k*-mers; Table S7). However,
488 L11 isolates only had high *k*-mer containment for hap25 and hap26 (100% identity, >
489 99.95% shared *k*-mers) but neither hap3 nor hap4 (< 99.95% identity, < 98.70%
490 shared *k*-mers), as would be the case if L11 and L18 were somatic hybrids (Table S7).
491 Haplotype-specific phylogenetic trees reflected the same relationships, with L18 and
492 L11 isolates forming a single clade in the hap25 phylogeny and discrete clades in the
493 hap3, hap4, and hap26 phylogenetic trees (Fig. **S9a-d**). Whole-haplotype alignments
494 between hap3, hap4, and hap25 clarified this non-reciprocal relationship. Hap25 has
495 high-identity alignments from hap3 and hap4 in large blocks covering >95% of this
496 nuclear genome, with 0 to 3 breakpoints between hap3 and hap4 sequences per
497 chromosome. This is consistent with the frequency of recombination breakpoints in a
498 single meiotic event reported in other rust fungi (Fig. **S9e**; Anderson *et al*., 2016). As
499 further confirmation, we obtained variants from a pangenome graph of all haplotypes
500 and recovered the same recombination breakpoints as before from hap3 and hap4
501 SNP-sparse regions on hap25 (Fig. **S9f**). This is consistent with hap25 being derived
502 as a haploid product of meiosis from lineage L18 as part of a sexual cross.

503 We next explored the contradictory finding that L3 isolates had high containment for
504 haplotype hap5 (= hap24) as well as hap26 from 20WA94 (> 99.98% identity, >
505 99.84% shared *k*-mers), despite these all having the *bW3bE3 HD* allele pair (Fig. **4b**;
506 Fig. S5; Table S5; Table S7). In support of the haplotype containment results, we found
507 that L3 isolates cluster with L1 and L2 isolates in the hap5 and hap24 phylogenetic
508 trees but with L11 in the hap26 tree, while these lineages all form discrete clades in
509 the hap6, hap23, and hap25 phylogenetic trees (Fig. **4c-f**; Fig. **S9a,b**). The hap5/24
510 and hap26 haplotypes all contain the *bW3bE3 HD* allele, while L3 is heterozygous for
511 b3/b12 alleles suggesting some recombination involved in the relationships between
512 these isolates. Alignment of hap5 and hap26 again revealed numerous high-identity
513 segments covering about 45% of haplotype genome, in a pattern that suggests
514 recombination (Fig. **S10a**; Table S8). A model to explain these results is that L3 shares

515 a nucleus (hap5) with lineage L1 and L2 through a nuclear exchange event, while
516 hap26 in L11 is a meiotic recombinant between the two nuclei (hap5 and an unknown
517 haplotype) in L3 (Fig. **S10b**). In this scenario, L11 results from a sexual cross between
518 L3 and L18. The unique sequence amounting to approximately half of the unknown L3
519 haplotype should also contain *bW12bE12*.

520 Given the evidence for recombination in the hap25 and hap26 genomes, we evaluated
521 the entire Australian *Pca* population for recombination by generating a splitstree
522 network which showed reticulation between most Australian lineages. The *phi* test was
523 significant ($p < 10^{-10}$), suggesting that recombination has occurred to shape the extant
524 Australian population. We next characterized haplotype recombination across the
525 entire *Pca* pangenome using the pangenome graph-derived variants approach
526 validated above in our parent-F1 comparison. A subset of six USA haplotypes were
527 assessed first, which showed a pattern of small recombination blocks that reflects
528 frequent sexual recombination in the USA facilitated by the prevalence of *R. cathartica*
529 (Fig. **5a**). Further, each haplotype added only 4 to 9% coverage, with hap18 having
530 approximately 30% coverage by the preceding five USA haplotypes. This suggests
531 the full diversity of the USA sexual population has not been sampled in the haplotype
532 pangenome.

533 Among the Australian haplotypes, hap6 and hap23 are clearly divergent from the
534 others, while all other haplotypes appear related through recombination (Fig. **5b**).
535 Hap3 and hap4 have contributed to all recombinant haplotypes, while hap5 and hap26
536 have contributed to all except hap25. The large size of recombination blocks between
537 Australian haplotypes relative to those shared between USA haplotypes suggests that
538 recombination is relatively infrequent. However, Australian haplotypes in L3-L18 are
539 approximately 75% covered by hap3, hap4, hap5, and hap26, consistent with founding
540 by four haplotypes. Considering our earlier results with hap5 and hap26, we
541 hypothesize that the remaining 25% is likely present in L3, which has not been
542 represented in the pangenome yet. Therefore, Australian lineages L4-L17 may have
543 been founded by limited recombination between the L3 and L18 lineages (Fig. **6a**).
544 The divergent hap6 and hap23 haplotypes in L1 and L2 have not contributed to the
545 L3-L18 recombining population.

546 Recombination analysis performed on all haplotypes demonstrated that the USA and
547 Australian haplotypes overall have not had recent recombination, as the total amount
548 of shared sequence between USA and Australian haplotypes was generally less than
549 5% (Table S8). The only exceptions were hap6 and hap23, which share approximately
550 50% and 15% of their sequence with the USA haplotypes, respectively (Fig. **S12b**).
551 Hap1 and hap2 from Pca203 contribute most to hap6 (31.4%), while the other USA
552 haplotypes share small regions (<5%) with the remaining hap6 sequence (Fig. **S12b**).
553 Hap9 and hap14 are divergent from other USA haplotypes, but hap14 and hap6 have
554 approximately 11% of their sequence in common (Table S8). The USA population is
555 shaped by frequent recombination between diverse haplotypes, which is illustrated by
556 hap20 having only 45% of its sequence represented by 11 other USA haplotypes (Fig.
557 **S12b**). Historic isolate Pca203 clearly had an influential founding effect on the USA
558 population, with hap1 and hap2 contributing from 21.3 to 57.6% of the sequence of
559 other USA haplotypes (Fig. **S12b**; Table S8).

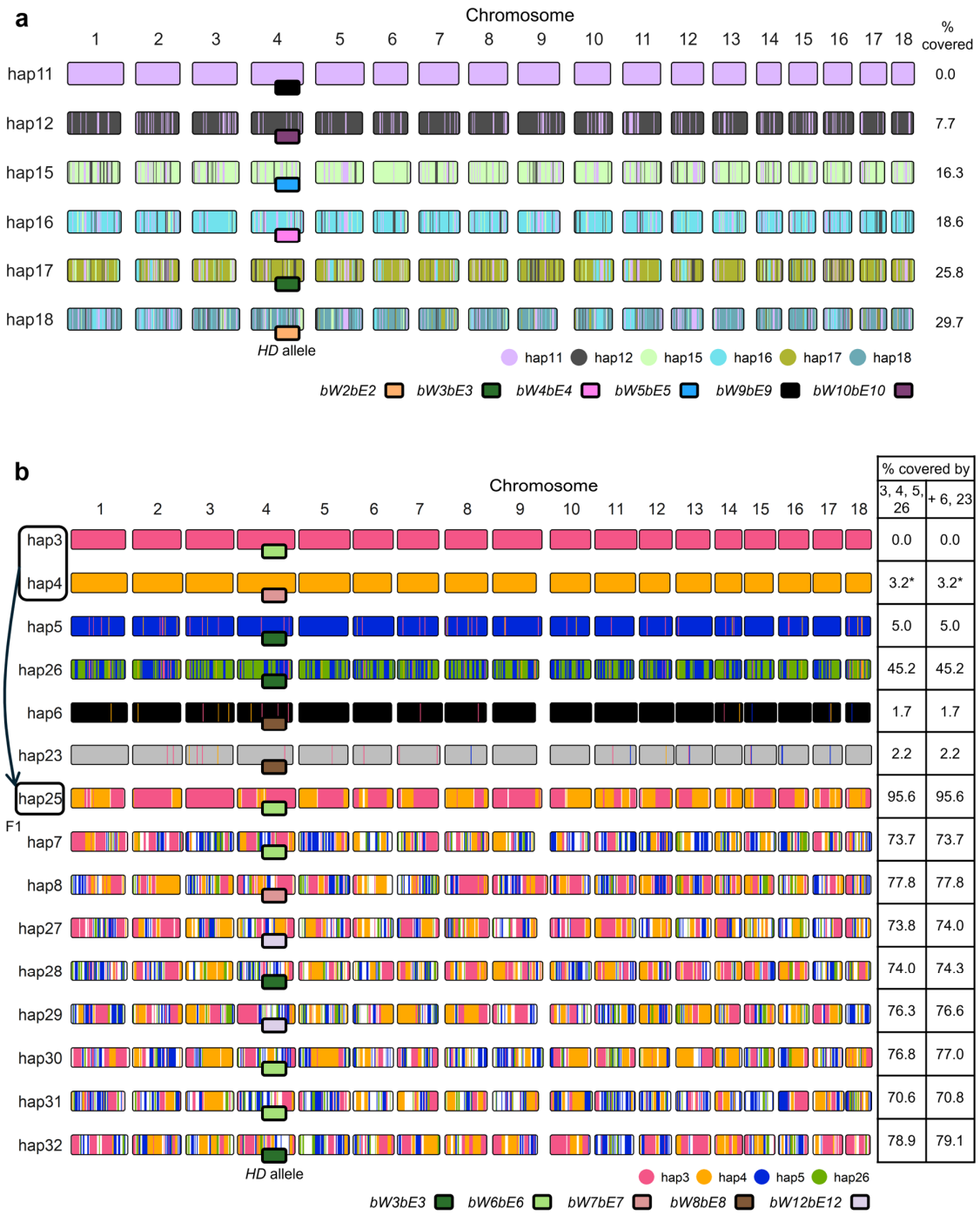


Figure 5. **a)** Shared recombination blocks across six *Puccinia coronata* f. sp. *avenae* (*Pca*) haplotypes identified by identifying 100 kb sequence bins with less than 50 non-reference variants hierarchically across hap11, hap12, hap15, hap16, hap17, and hap18. Shared sequences with hap11 were assigned across the other five haplotypes first, after which the unassigned areas of hap12 were evaluated on the four haplotypes below in the hierarchy, and so on. **b)** Shared haplotype blocks between 15 Australian

Pca haplotypes from founder representatives hap3, hap4, hap5, hap26 and divergent haplotypes hap6 and hap23 generated in a similar manner to panel **a)** with hap3 being first in the hierarchy. The haplotype with the shorter name was kept as the representative in cases where haplotypes are nearly identical, as they show redundant information in this analysis (hap8, hap7, and hap5 shown; hap21, hap22, and hap24 not shown). Chromosome fill color represents unassigned regions or regions shared with haplotypes earlier in the hierarchy. Unassigned regions from hap7, hap8, hap27, hap28, hap29, hap30, hap31, and hap32 are filled white. Table on the right side shows percent coverage of each haplotype by those listed in each column. *HD* locus alleles are indicated with fill color of rectangles at the chromosome 4 midpoint. Asterisk (*) indicates that the haplotype blocks on hap4 from hap3 are not shown for visual clarity.

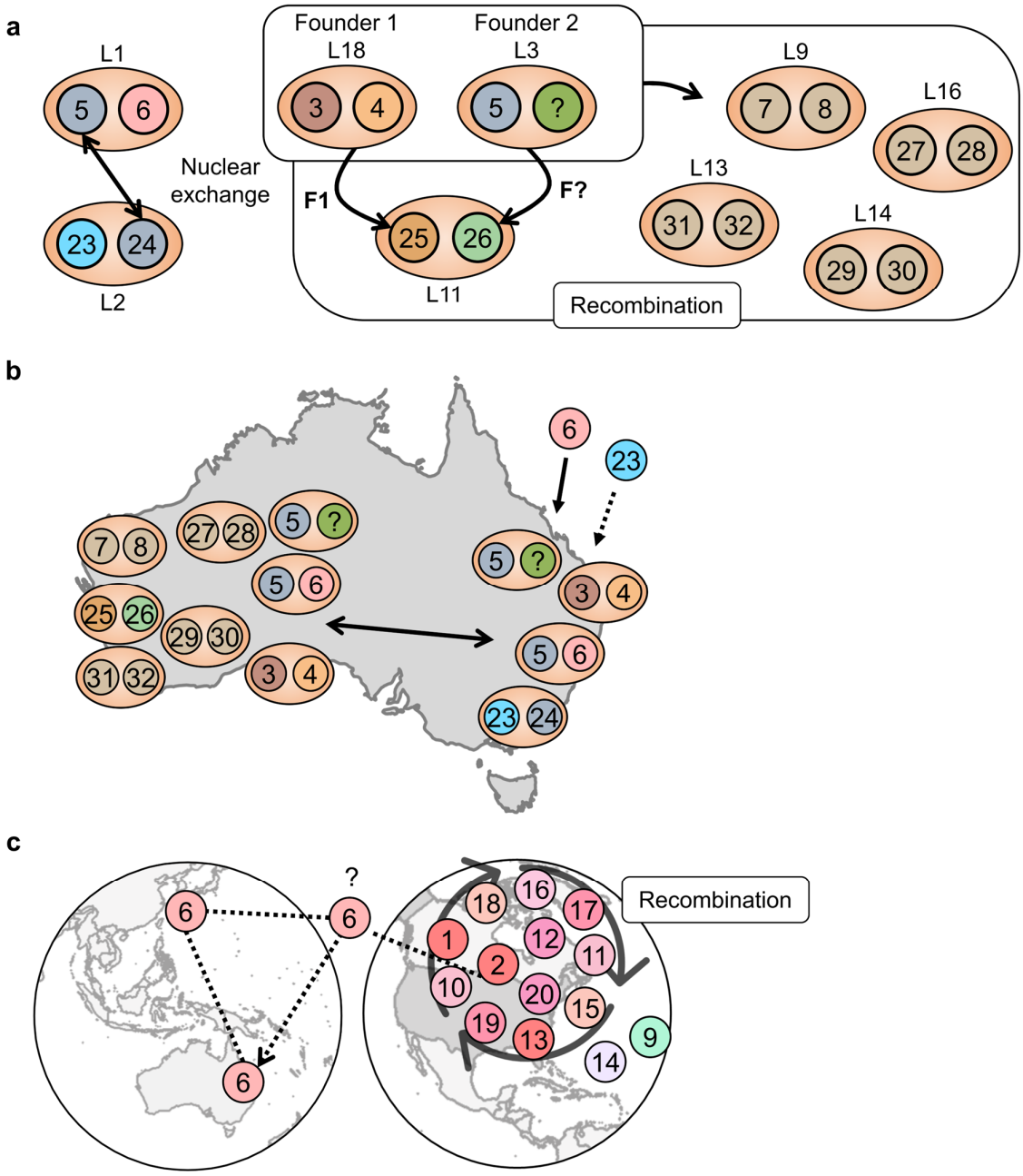


Figure 6. **a)** Diagram showing the proposed relationship of haplotypes in the Australian *Puccinia coronata* f. sp. *avenae* (*Pca*) population to the two postulated founder lineages L18 and L3. Four colors are assigned to founding haplotypes (brown – hap3, orange – hap4, blue – hap5, green – hap'unknown') and derivative haplotypes are color combinations of their ancestors. **b)** Known distribution of haplotypes in the Australian *Pca* collection in either Western Australia or eastern Australia (Australian Capital Territory, New South Wales, Queensland, South Australia, Victoria). **c)**

586 Recombination between closely related USA *Pca* haplotypes (shades of pink) and
587 hap6 with simplified possible pathways for hap6 migration.

588 **Discussion**

589

590 In this study we investigated the genetic relationships between newly sequenced
591 Australian and Taiwanese *Pca* isolates and previously evaluated populations from the
592 USA and South Africa (Miller *et al*., 2018, 2020; Hewitt *et al*., 2023). The rich genotypic
593 diversity of this pathogen necessitated the construction of a haplotype-aware
594 pangenome for the species to capture genotypic variability across different geographic
595 areas. We focused on haplotypes from fifteen isolates representing Australian and
596 USA *Pca* lineages as an initial step. This haplotype atlas allowed us to examine
597 mechanisms that contribute to virulence evolution such as migration, somatic
598 hybridization, and sexual recombination in *Pca* (Li *et al*., 2019; Figueroa *et al*., 2020;
599 Sperschneider *et al*., 2023).

600

601 The role of somatic hybridization (nuclear exchange) in the evolution of dikaryotic rust
602 fungi was proposed in the 1950-1960's (Flor, 1964; Dodds, 2023) and was
603 investigated in *Pca* under controlled inoculations and conditions (Bartos *et al*., 1969).
604 Unfortunately, most early studies assessing somatic hybridization in cereal rust fungi
605 failed to ensure that isolate cross-contamination did not explain detection of new
606 races/virulence profiles. Indisputable proof of nuclear exchange in rust fungi was found
607 by comparing the first chromosome level assembly of wheat stem rust *Pgt*21-0 and
608 sequences from epidemic isolate *Pgt* Ug99 (Li *et al*., 2019), which showed that one
609 nuclear haplotype of *Pgt*21-0 is nearly identical to one of the haplotypes of *Pgt* Ug99.
610 The power of haplotype resolution to track the migration and nuclear exchanges
611 shaping global rust populations was subsequently demonstrated for *Pt* (Sperschneider
612 *et al*., 2023). This study provides evidence that somatic hybridization is also occurring
613 in *Pca* populations in Australia, USA and possibly Taiwan, supporting that nuclear
614 exchanges are common to rust fungi.

615

616 Newly acquired and existing whole-genome sequence short-read data and haplotype
617 atlas also enabled population-level *in silico* characterization of mating type control in
618 *Pca*. In other Basidiomycota such as smuts and mushrooms, mating type has been
619 shown to maintain dikaryotic state, regulate life cycle, and enforce self/non-self
620 recognition during mating (Nieuwenhuis *et al*., 2013; Coelho *et al*., 2017). Although

621 the role of mating loci has not been functionally characterized in the rusts, our finding
622 that all 353 *Pca* individuals in this study are heterozygous at both the *PR* and *HD* loci
623 suggests that mating type controls critical biological functions in *Pca*. Consistent with
624 the findings of Luo et al. (2024), pheromone precursor *mfa2* is linked closely to
625 *STE3.2.2* in *Pca*; however, we identified a 55 amino acid *mfa3* pheromone precursor
626 linked to *STE3.2.3* which was not reported in their study. Analysis of the *HD* alleles in
627 the haplotype pangenome released by our study support most sequences imputed for
628 *Pca HD* locus alleles published by Luo et al. (2024). There is one discrepancy between
629 the *bW5* allele between the studies in which the first 12 N-terminal amino acids of *bW5*
630 defined by Luo et al. (2024) are not encoded in the haplotypes used to define *bW5* in
631 this study. In addition, the '*bW8bE8*' allele pair reported by Luo et al. (2024) was only
632 found in a single isolate (90MN5B) that was not included in our pangenome.
633
634 The *HD* locus configuration has been used in *Pt* and *Pst* to support conclusions
635 regarding migration and population structure derived from genomic and phylogenetic
636 approaches (Holden *et al*., 2023; Sperschneider *et al*., 2023). However, several
637 divergent haplotypes (~500-600K SNPs) included in the *Pca* pangenome have
638 identical alleles at both mating type loci (i.e., hap6 and hap23, hap3 and hap13, hap5
639 and hap17). Thus, *HD* locus allele containment solely should not be used to infer
640 lineage membership or haplotype composition.
641
642 We investigated the contribution of genetic recombination to the evolution of *Pca* in
643 the USA and Australia. Small and frequent recombination blocks among USA
644 haplotypes were detected and are consistent with widespread sexual reproduction
645 enable by the prevalence of *R. cathartica* in North America (Miller *et al*., 2020; Hewitt
646 *et al*., 2023). Prior to this study, molecular characterization of the Australian population
647 was limited to 12 *Pca* isolates using DNA amplification fingerprinting and virulence
648 phenotypes (Brake *et al*., 2001). Based on this, it was proposed the Australian *Pca*
649 population was evolving by mutations in clonal lineages, despite the high phenotypic
650 diversity of the pathogen.
651
652 We uncovered 18 lineages in the Australian *Pca* population, in contrast to the *Pca*
653 collections from Taiwan and South Africa, which contain only one or two clonal

654 lineages. This is a contrasting result to cereal rusts in Australia, like *Pgt* and *Pt*, which
655 appear to be exclusively clonal, with the presence of only two and five lineages,
656 respectively (Li *et al*., 2019; Sperschneider *et al*., 2023). The discovery of many
657 lineages in the Australian *Pca* population was unexpected given its presumed
658 asexuality (Park *et al*., 2022) and is evidence against mutations and incursions as the
659 sole drivers of *Pca* evolution in the country. The best-described sexual host (*R.*
660 *cathartica*) is not present in Australia or Oceania; however, other *Rhamnus* species
661 (e.g., *R. alaternus, R. lycioides*, *R. palaestina*) have been reported as aecial hosts for
662 *Pca* (Dinoor, 1962; Hemmami *et al*., 2006; Nazareno *et al*., 2018). Notably, *R.*
663 *alaternus* was introduced to Australia as an ornamental and is currently managed as
664 an invasive weed (Stajsic & VicFlora, 2023).
665
666 Both network analysis and whole-genome comparisons to define recombination
667 blocks provided evidence of recombination among most Australian lineages, although
668 infrequent when compared to the USA *Pca* population. It is yet unclear where or when
669 these events occurred, and their rarity could be explained by sporadic access to a
670 sexual host or uncommon parasexual events. Parasexuality akin to somatic
671 hybridization that involves karyogamy and haploidization to generate recombinants
672 without the aecial host is also proposed to occur in rust fungi (Harder, 1984), although
673 this phenomenon has not been experimentally validated using molecular and genomic
674 tools. The use of field collections hampers validation of parasexuality as isolates
675 resulting from either somatic recombination or a sexual cross would be genotypically
676 indistinguishable. The recombination among Australian lineages likely explains the
677 high phenotypic diversity of *Pca* in Australia (Park *et al*., 2022), which was poorly
678 correlated to lineage structure as has been reported in other *Pca* populations (Hewitt
679 *et al*., 2023).
680
681 The splitstree network showed clear reticulation between lineages L3 through L18.
682 Interestingly L3 and L18, which are positioned at opposite ends of the network, are
683 also implicated in recombination events connected to L11. Hap25 is likely the meiotic
684 product of hap3 and hap4, while hap26 may be a recombinant derived from L3
685 haplotypes (hap5 and unknown). Therefore, it seems that L3 and L18 could have
686 undertaken a sexual cross to produce L11. While intercontinental migrations could

687 explain these results, the likelihood of F1 haplotypes migrating together or in separate
688 events is much lower than the hypothesis that rare sexual/parasexual cycles may
689 occur. Together with the evidence of somatic hybridization in L1 and L2, we propose
690 that the extant Australian *Pca* population was derived from three ancestral isolates.
691 Two of these have undergone sexual and/or parasexual processes and the third
692 donated nuclei via somatic nuclear exchange.
693
694 Our results suggest that it would be prudent to continue molecular monitoring of *Pca*
695 and investigate the relationship between *R. alaternus* or other *Rhamnus* species and
696 *Pca* in the Australian context, as the presence of these invasive species may present
697 significant risk for virulence evolution to the detriment of the oat industry. Ongoing
698 efforts to expand this haplotype-aware pangenome to include members from other
699 continents will further our understanding of the evolution and global movement of *Pca*.
700

Acknowledgements

We would like to thank Jakob Riddle (USDA-ARS) for his technical support. We would like to acknowledge the contribution of the Plant Pathogen 'Omics Initiative consortium in the generation of data used in this publication. The Initiative is supported by funding from Bioplatforms Australia, enabled by the Commonwealth Government National Collaborative Research Infrastructure Strategy (NCRIS).

Competing Interests

The authors declare no competing interests.

Author Contributions

MF, PD, SFK, JS, BJS, and ES planned and supported the project. DL, ECH, Y-FH, and ESN amplified and prepared pathogen materials for sequencing and phenotyping. ECH and JS analysed the data. ECH, JS, MF, and PD prepared the manuscript and figures. All authors reviewed and edited the manuscript.

Data Availability

Reference genomes are available in the CSIRO Data Access Portal (https://data.csiro.au/collection/csiro:61932). PacBio, HiC and genomic short reads are available on NCBI under BioProject PRJNA1063754. Scripts used to perform analyses and generate figures are available on GitHub (https://github.com/henni164/Pca_pangenome).

723 **References**

724 Aime MC, McTaggart AR. 2020. A higher-rank classification for rust fungi, with notes
725 on genera. *Fungal Sys. Evol.* 7: 21-47.
726 Anderson C, Khan MA, Catanzariti AM, Jack CA, Nemri A, Lawrence GJ, Upadhyaya
727 NM, Hardham AR, Ellis JG, Dodds PN, et al. 2016. Genome analysis and avirulence
728 gene cloning using a high-density RADseq linkage map of the flax rust fungus,
729 *Melampsora lini*. *BMC Genomics*. 17: 667.
730 Bartos P, Fleischmann G, Samborski DJ, Shipton WA. 1969. Studies on asexual
731 variation in the virulence of oat crown rust, *Puccinia coronata* f. sp. *avenae*, and wheat
732 leaf rust, *Puccinia recondita*. *Can. J. Botany.* 47: 1383-1387.
733 Bauer R, Begerow D, Sampaio JP, Weiß M, Oberwinkler F. 2006. The simple-septate
734 basidiomycetes: a synopsis. *Mycol. Prog.* 5: 41-66.
735 Berlin A, Wallenhammar A-C, Andersson B. 2018. Population differentiation of
736 *Puccinia coronata* between hosts – implications for the epidemiology of oat crown rust.
737 *Eur. J. Plant Pathol*. 152: 901–907.
738 Bölker M, Urban M, Kahmann R. 1992. The *a* mating type locus of *U. maydis* specifies
739 cell signaling components. *Cell*. 68: 441-450.
740 Brake VM, Irwin JAG, Park RF. 2001. Genetic variability in Australian isolates of
741 *Puccinia coronata* f. sp. *avenae* assessed with molecular and pathogenicity markers.
742 *Australas. Plant Path*. 30: 259-266.
743 Bruen TC, Philippe H, Bryant D. 2005. A simple and robust statistical test for detecting
744 the presence of recombination. *Genetics.* 17: 2665-81.
745 Bryant D, Huson DH. 2023. NeighborNet: improved algorithms and implementation.
746 *Front. Bioinform*. 3: 1178600.
747 Burkhard S, Banuet F, Dahl M, Schlesinger R, Schäfer W, Martin T, Herskowitz I,
748 Kahmann R. 1990. The *b* alleles of *U. maydis*, whose combinations program
749 pathogenic development, code for polypeptides containing a homeodomain-related
750 motif. *Cell*. 60: 295-306.
751 Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive,
752 efficient and simple way T. Tullius, ed. *PeerJ*. 6: e4958.
753 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden T.
754 L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10: 421.

755 Chen J, Zhang X, Rathjen JP, Dodds PN. 2022. Direct recognition of pathogen
756 effectors by plant NLR immune receptors and downstream signalling. *Essays*
757 *Biochem*. 66: 471–483.
758 Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal sex: the
759 Basidiomycota. *Microbiol. Spectr.* 5. 10.1128/microbiolspec.funk-0046-2016.
760 Cuomo CA, Bakkeren G, Khalil HB, Panwar V, Joly D, Linning R, Sakthikumar S,
761 Song X, Adiconis X, Fan L, et al. 2017. Comparative analysis highlights variable
762 genome content of wheat rusts and divergence of the mating loci. *G3*. 7: 361–376.
763 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,
764 Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools.
765 *Bioinformatics*. 27: 2156–2158.
766 Dinoor A. 1962. Note on germination of crown rust teliospores and infection of
767 buckthorn. *Can. J. Plant Sci*. 42: 537-538.
768 Dodds PN. 2023. From gene-for-gene to resistosomes: Flor's enduring legacy. *MPMI*.
769 36: 461-467.
770 Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant–
771 pathogen interactions. *Nat Rev Genet*. 11: 539–548.
772 Duan H, Jones AW, Hewitt T, Mackenzie A, Hu Y, Sharp A, Lewis D, Mago R,
773 Upadhyaya NM, Rathjen JP, et al. 2022. Physical separation of haplotypes in
774 dikaryons allows benchmarking of phasing accuracy in Nanopore and HiFi assemblies
775 with Hi-C data. *BMC Genome Biol*. 23: 84.
776 Figueroa M, Dodds PN, Henningsen EC. 2020. Evolution of virulence in rust fungi –
777 multiple solutions to one problem. *Curr. Opin. Plant Biol*. 56: 20-27.
778 Figueroa M, Dodds PN, Henningsen EC, Sperschneider J. 2023. Global landscape of
779 rust epidemics by *Puccinia* species: current and future perspectives. In: Scott B,
780 Mesarich C, eds. *Plant Relationships: Fungal-Plant Interactions*. Springer, 183-203.
781 Flor HH. 1964. Genetics of somatic variation for pathogenicity in *Melampsora lini*.
782 *Phytopathol*. 54: 823-826.
783 Galili T. 2015. dendextend: an R package for visualizing, adjusting, and comparing
784 trees of hierarchical clustering. *Bioinformatics*. 31: 3718-3720.
785 Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. 2022. A spectrum of
786 free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2,
787 hts-nim and slivar. *PLoS Comput. Biol*. 18: e1009123.

788 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read
789 sequencing. *arXiv*. doi: 10.48550/arXiv.1207.3907.
790 Harder DE. 1984. Developmental Ultrastructure of Hyphae and Spores. In: Bushnell
791 WR, Roelfs AP, eds. *The Cereal Rusts Volume I: Origins, specificity, structure, and*
792 *physiology*. Orlando, Florida: Academic Press.
793 Hemmami I, Allagui MB, Chakroun M, Gazzah ML. 2006. *Rhamnus lycoides* in Tunisia
794 is a new aecial host of oat crown rust. *Eur. J. Plant Pathol*. 115: 357-361.
795 Henningsen EC, Hewitt T, Dugyala S, Nazareno ES, Gilbert E, Li F, Kianian SF,
796 Steffenson BJ, Dodds PN, Sperschneider J, et al. 2022. A chromosome-level, fully
797 phased genome assembly of the oat crown rust fungus *Puccinia coronata* f. sp.
798 *avenae*: a resource to enable comparative genomics in the cereal rusts. *G3*. 12:
799 jkac149.
800 Henningsen EC, Lewis D, Nguyen DT, Sperschneider J, Kianian SF, Stone E, Dodds
801 PN, Figueroa M. 2024. Virulence patterns of oat crown rust in Australia – season 2022.
802 *bioRxiv*. doi: 10.1101/2023.09.24.558528.
803 Hewitt T, Henningsen EC, Pereira D, McElroy K, Nazareno ES, Dugyala S, Nguyen-
804 Phuc H, Li F, Miller ME, et al. 2023. Genome-enabled analysis of population
805 dynamics and virulence associated loci in the oat crown rust fungus *Puccinia*
806 *coronata* f. sp. *avenae*. *MPMI*. doi: 10.1094/MPMI-09-23-0126-FI.
807 Hickey G, Monlong J, Ebler J, Novac AM, Eizenga JM, Gao Y, Human Pangenome
808 Reference Consortium, Marschall T, Li H, Paten B. 2023. Pangenome graph
809 construction from genome alignments with Minigraph-Cactus. *Nat. Biotechnol*. doi:
810 10.1038/s41587-023-01793-w.
811 Holden S, Bakkeren G, Hubensky J, Bamrah R, Abbasi M, Qutob D, de Graaf ML,
812 Kim SH, Kutcher HR, McCallum BD, et al. 2023. Uncovering the history of
813 recombination and population structure in western Canadian stripe rust populations
814 through mating-type alleles. *BMC Biology*. 21: 233.
815 Ho C-Y, Henningsen EC, Chen S-T, Ariyawansa HA, Nazareno ES, Sperschneider
816 J, Dodds PN, Riddle J, Kianian SF, Figueroa M, et al. 2024. Detection of oat crown
817 rust disease in Taiwan (2019-2021). *bioRxiv.* doi: 10.1101/2024.03.12.584714.
818 Huang N, Li H. 2023. compleasm: a faster and more accurate reimplementation of
819 BUSCO. *Bioinformatics*. 39: btad595.
820 Jones JDG, Dangl JL. 2006. The plant immune system. *Nature*. 444: 323–329.

821 Kirk P, Cannon P, Minter D, Stalpers J. 2008. *Dictionary of the Fungi.* 10th ed.
822 Wallingford, UK: CABI, eISBN:978-1-84593-728-7.
823 Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL.
824 2004. Versatile and open software for comparing large genomes. *Genome Biol*. 5:
825 R12.
826 Kües U, James TY, Heitman J. 2011. Mating type in Basidiomycetes: Unipolar, bipolar,
827 and tetrapolar patterns of sexuality. In: Pöggeler S, Wöstemeyer J, eds. *Evolution of*
828 *Fungi and Fungal-Like Organisms. The Mycota, vol. 14*. Berlin, Heidelberg: Springer.
829 https://doi.org/10.1007/978-3-642-19974-5_6.
830 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H,
831 Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version
832 2.0. *Bioinformatics*. 23: 2947-2948.
833 Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for
834 phylogenetic tree display and annotation. *Nucleic Acids Res*. 49: W293–W296.
835 Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, Raley C,
836 Miller ME, Silverstein KAT, Henningsen EC, et al. 2019. Emergence of the Ug99
837 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nat. Commun*.
838 10: 5068.
839 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*.
840 34: 3094–3100.
841 Luo Z, McTaggart A, Schwessinger B. 2024. Genome biology and evolution of mating-
842 type loci in four cereal rust fungi. *PLoS Genet.* 20: e1011207.
843 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update:
844 novel and streamlined workflows along with broader and deeper phylogenetic
845 coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol*. 38:
846 4647-4654.
847 Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018.
848 MUMmer4: A fast and versatile genome alignment system. *PLoS Comp. Biol*. 14:
849 e1005944.
850 Miller ME, Zhang Y, Omidvar V, Sperschneider J, Schwessinger B, Raley C, Palmer
851 JM, Garnica D, Upadhyaya N, Rathjen J, et al. 2018. *De novo* assembly and phasing
852 of dikaryotic genomes from two isolates of *Puccinia coronata* f. sp. *avenae*, the causal
853 agent of oat crown rust. *mBio*. 9: 10.1128/mbio.01650-17.

854 Miller ME, Nazareno ES, Rottschaefer SM, Riddle J, Dos Santos Pereira D, Li F,
855 Nguyen-Phuc H, Henningsen EC, Persoons A, Saunders DGO, et al. 2020. Increased
856 virulence of *Puccinia coronata* f. sp. *avenae* populations through allele frequency
857 changes at multiple putative *Avr* loci. *PLoS Genet*. 16: e1009291.
858 Nazareno ES, Li F, Smith M, Park RF, Kianian SF, Figueroa M. 2018. *Puccinia*
859 *coronata* f. sp. *avenae*: a threat to global oat production. *Mol. Plant Pathol*. 19: 1047–
860 1060.
861 Nieuwenhuis BPS, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. 2013.
862 Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity.* 111:
863 445-455.
864 Ondov BD, Starrett GJ, Sappington A, Kostic A, Koren S, Buck CB, Phillippy AM. 2019.
865 Mash Screen: high-throughput sequence containment estimation for genome
866 discovery. *Genome Biol*. 20: 232.
867 Ortiz EM. 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for
868 phylogenetic analysis. doi: 10.5281/zenodo.2540861.
869 Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and
870 evolutionary analyses in R. *Bioinformatics*. 35: 526-528.
871 Park RF, Chhetri M, Singh D, Ding Y. 2022. *Cereal rust situation, August 2022*.
872 Cereal Rust Report 19. Plant Breeding Institute, University of Sydney. URL
873 https://www.sydney.edu.au/content/dam/corporate/documents/faculty-of-
874 science/research/life-and-environmental-sciences/rust-reports/cereal-rust-report-
875 2022-vol-19-2-revised-updated.pdf.
876 Petersen RH. 1974. The rust fungus life cycle. *Bot. Rev*. 40: 453-513.
877 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
878 features. *Bioinformatics*. 26: 841–842.
879 R Core Team. 2023. *R: A language and environment for statistical computing v4.3.2*.
880 R Foundation for Statistical Computing, Vienna, Austria.
881 Schwessinger B, Sperschneider J, Cuddy WS, Garnica DP, Miller ME, Taylor JM,
882 Dodds PN, Figueroa M, Park RF, Rathjen JP. 2018. A near-complete haplotype-
883 phased genome of the dikaryotic wheat stripe rust fungus *Puccnia striiformis* f. sp.
884 *tritici* reveals high interhaplotype diversity. *mBio*. 9: 10.1128/mbio.02275-17.
885 Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A cross-platform and ultrafast toolkit for
886 FASTA/Q file manipulation. *PLoS One*. 11: e0163962.

887 Sperschneider J, Hewitt T, Lewis DC, Periyannan S, Milgate AW, Hickey LT, Mago R,
888 Dodds PN, Figueroa M. 2023. Nuclear exchange generates population diversity in the
889 wheat leaf rust pathogen *Puccinia triticina*. *Nat. Microbiol*. 8: 2130-2141.
890 Stajsic T, VicFlora. 2023. *VicFlora: Rhamnus alaternus*. Royal Botanic Gardens
891 Victoria. [WWW document] URL: https://vicflora.rbg.vic.gov.au/flora/taxon/1e3789d3-
892 489c-4071-ae6e-05f722c10a18. [accessed 6 March 2024].
893 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
894 analysis of large phylogenies. *Bioinformatics*. 30: 1312–1313.
895 Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration
896 of BWA-MEM for multicore systems. In: *2019 IEEE International Parallel and*
897 *Distributed Processing Symposium (IPDPS)*. Rio de Janeiro, Brazil: IEEE, 314-324.
898 Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New
899 York. https://ggplot2-book.org/.
900

Supplementary Figures

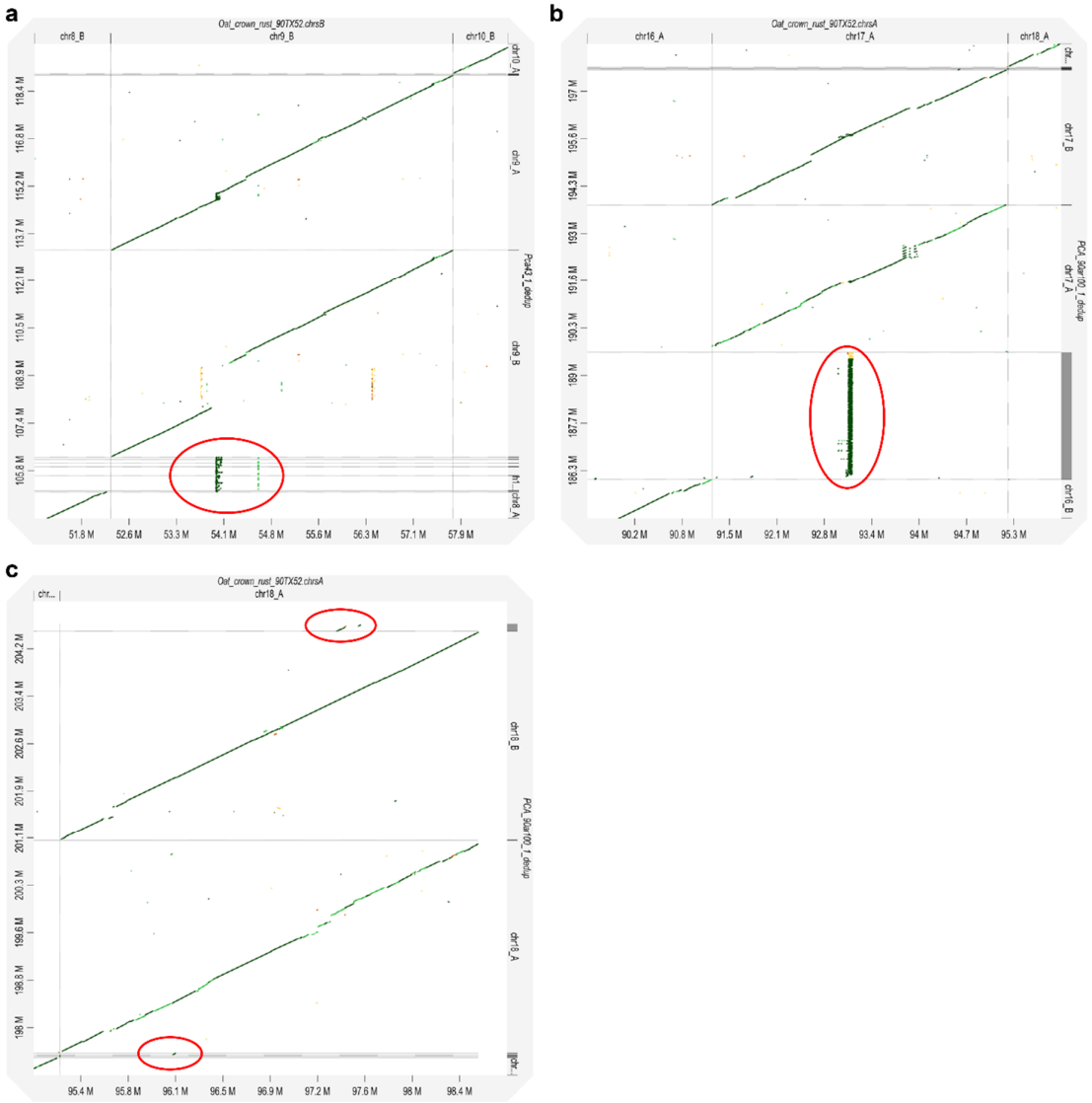
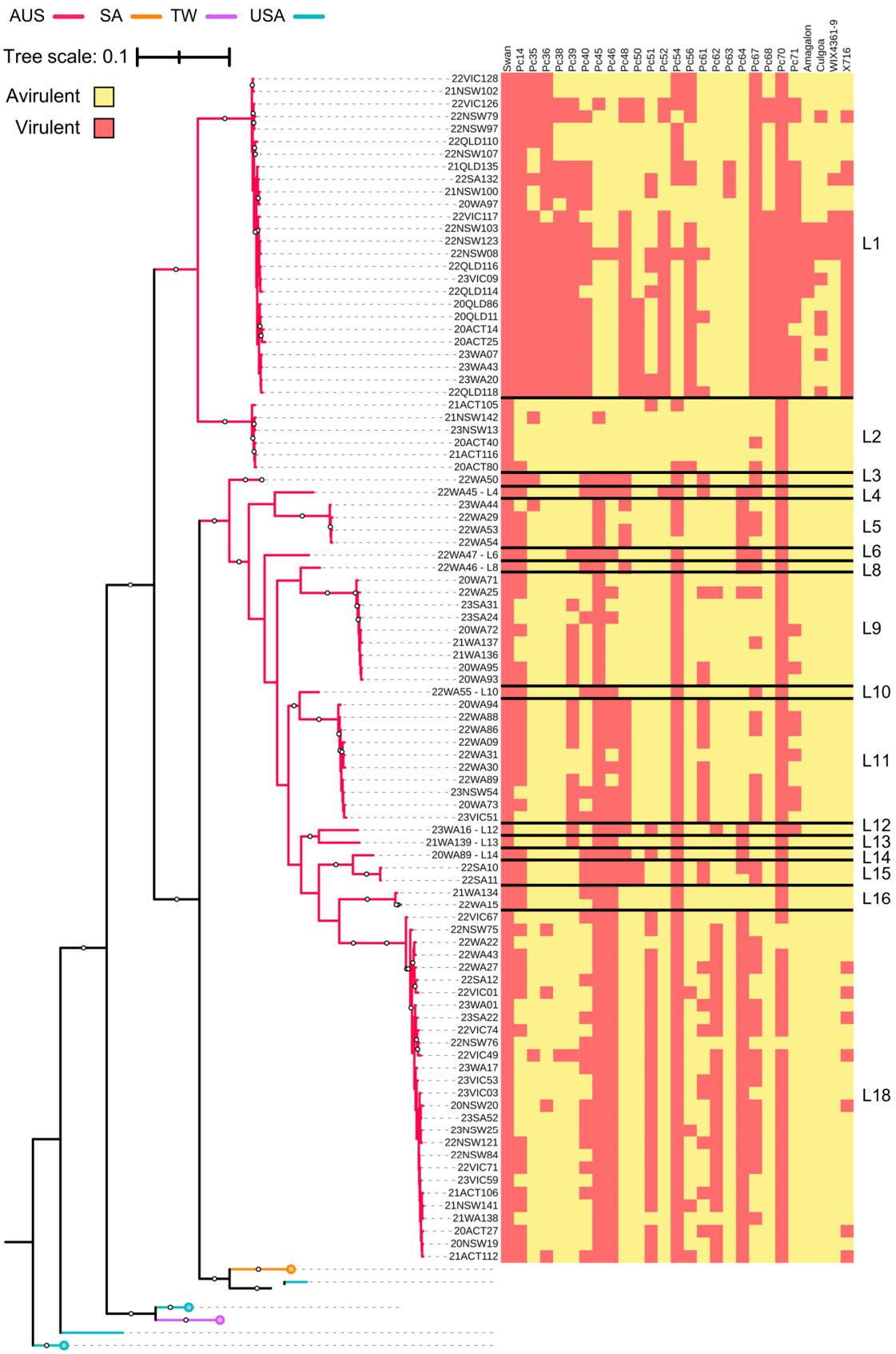


Figure S1. Examples of three main unplaced contig types: **a**) extra copies of chromosome 9 repeat-rich region **b**) extra copies of chromosome 17 ribosomal repeats **c**) extra copies of other genome sequences which are already represented in chromosomes.


AUS
SA
TW
USA
Tree scale: 0.1
Avirulent
Virulent
Swan
Pc14
Pc35
Pc36
Pc38
Pc39
Pc40
Pc45
Pc46
Pc48
Pc50
Pc51
Pc52
Pc54
Pc56
Pc61
Pc62
Pc63
Pc64
Pc67
Pc68
Pc70
Pc71
Amagalon
Culgoa
WIX4361-9
X716
22VIC128
21NSW102
22VIC126
22NSW79
22NSW97
22QLD110
22NSW107
21QLD135
22SA132
21NSW100
20WA97
22VIC117
22NSW103
22NSW123
22NSW08
22QLD116
23VIC09
22QLD114
20QLD86
20QLD11
20ACT14
20ACT25
23WA07
23WA43
23WA20
22QLD118
21ACT105
21NSW142
23NSW13
20ACT40
21ACT116
20ACT80
22WA50
22WA45 - L4
23WA44
22WA29
22WA53
22WA54
22WA47 - L6
22WA46 - L8
20WA71
22WA25
23SA31
23SA24
20WA72
21WA137
21WA136
20WA95
20WA93
22WA55 - L10
20WA94
22WA88
22WA86
22WA09
22WA31
22WA30
22WA89
23NSW54
20WA73
23VIC51
23WA16 - L12
21WA139 - L13
20WA89 - L14
22SA10
22SA11
21WA134
22WA15
22VIC67
22NSW75
22WA22
22WA43
22WA27
22SA12
22VIC01
23WA01
23SA22
22VIC74
22NSW76
22VIC49
23WA17
23VIC53
23VIC03
20NSW20
23SA52
23NSW25
22NSW121
22NSW84
22VIC71
23VIC59
21ACT106
21NSW141
21WA138
20ACT27
20NSW19
21ACT112
L1
L2
L3
L4
L5
L6
L8
L9
L10
L11
L12
L13
L14
L15
L16
L18

907

908 Figure S2. Phylogenetic tree constructed by mapping reads from 352 *Puccinia*
909 *coronata* f. sp. *avenae* isolates to Pca203 hap1, and hap2, and unplaced contigs.
910 376,646 biallelic SNPs and 500 bootstraps were used to produce the maximum
911 likelihood phylogenetic tree. Tree branches are colored by country of origin: AUS =
912 Australia; SA = South Africa; TW = Taiwan; USA = United States of America. Bootstrap
913 values are percentages (100 = 100%). Tree scale is mean substitutions per site.
914 Heatmap represents isolate virulence on the differential lines (virulent = red, avirulent
915 = yellow). Isolates that were not phenotyped were pruned from the tree, resulting in
916 the removal of L7 and L17.

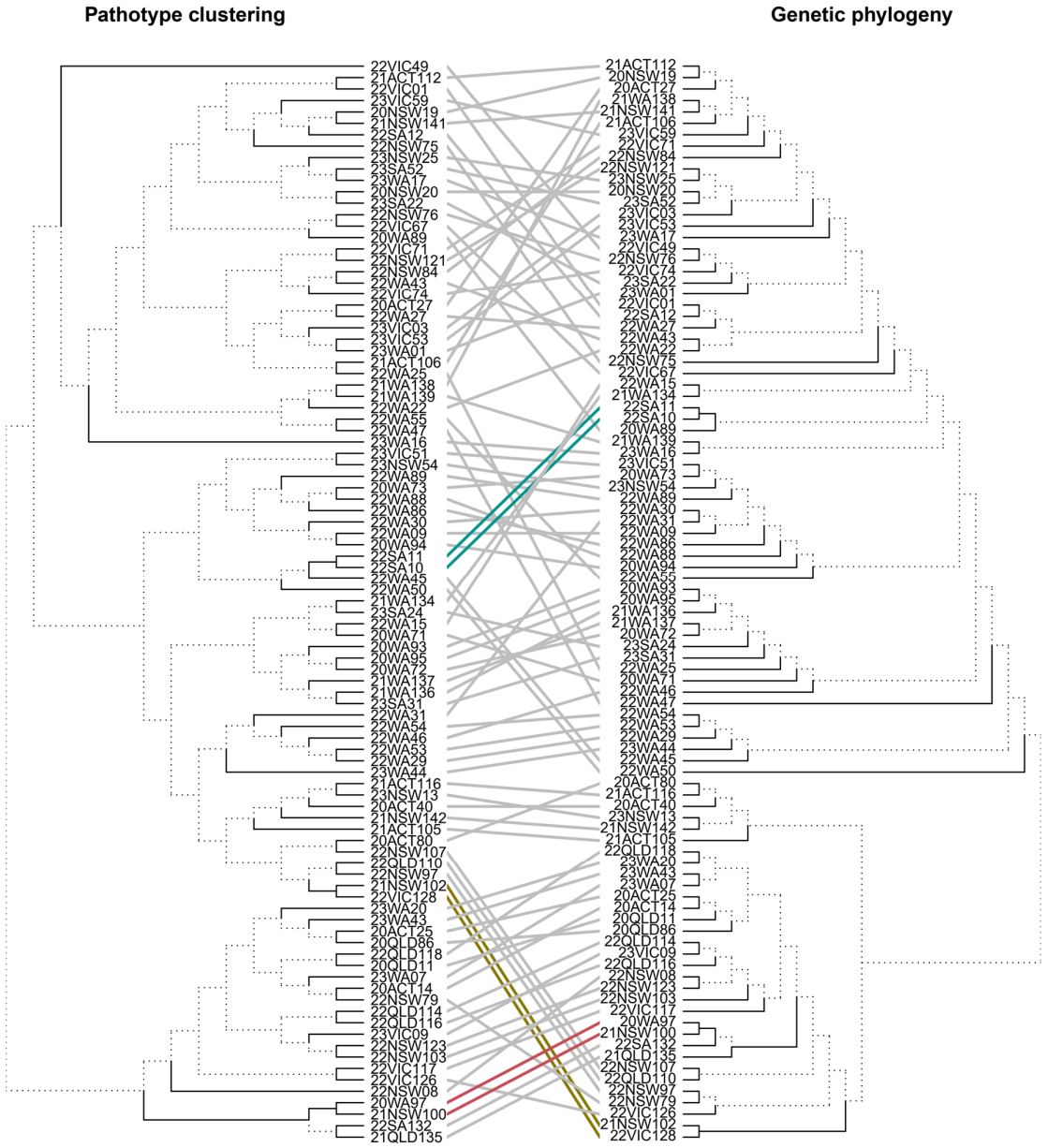
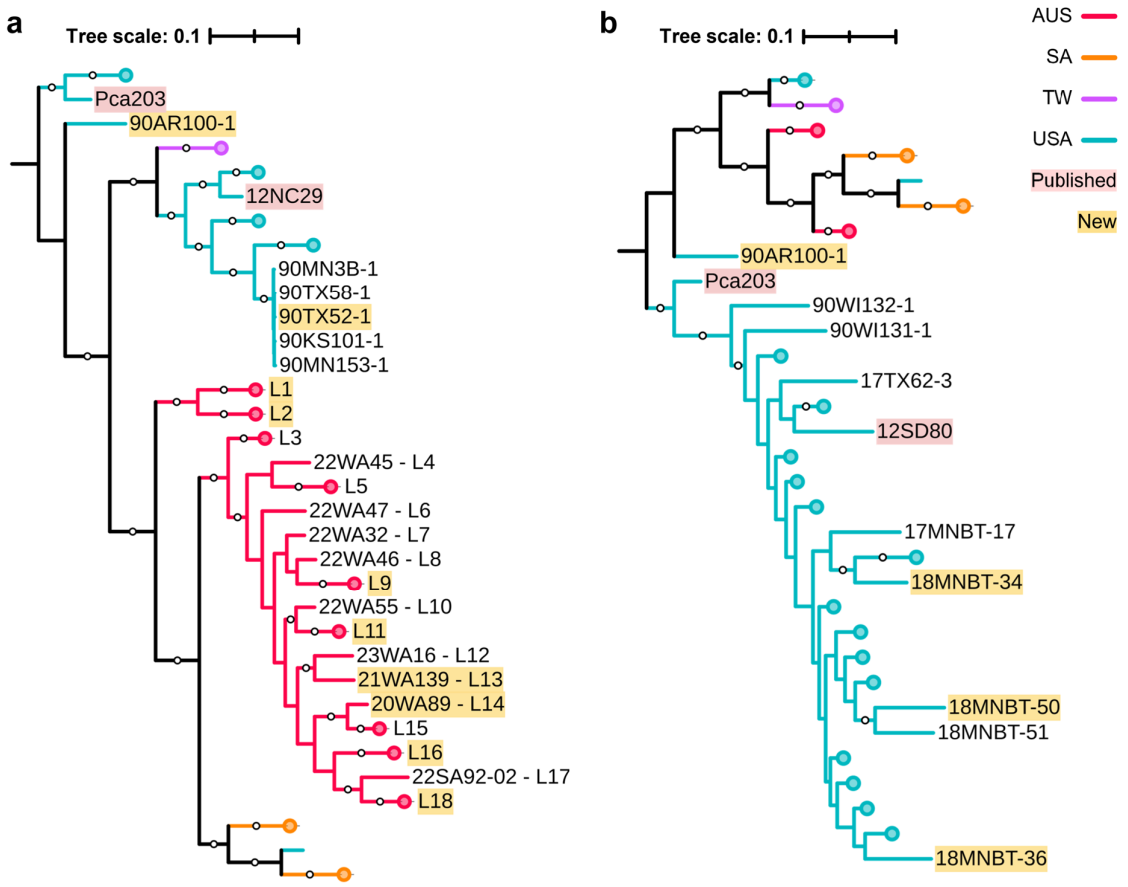


Figure S3. Comparison between tree topologies of 16 Australian *Puccinia coronata* f. sp. *avenae* lineages from clustering by pathotype (left) versus phylogenetic relationships (right). The ML tree (right) was generated with 376,646 SNPs from 352 *Pca* isolates against the complete Pca203 genome (hap1, and hap2, and unplaced contigs), which was pruned to contain only phenotyped Australian isolates and midpoint rooted. The R package ‘tanglegram’ was used to rotate the pathotype clustering tree branches until the best match to the phylogenetic tree was found. Branch lengths are arbitrary in the visualization and assessment of similarity. Solid black lines in the tree structure indicate edges found in both trees. Lines in the center

connect the same isolate across trees, with colored lines showing clusters containing more than one isolate which are structurally identical.



929

930 Figure S4. Midpoint rooted maximum likelihood phylogenetic tree of 352 *Puccinia*
931 *coronata* f. sp. *avenae* (*Pca*) isolates constructed by mapping reads and calling
932 variants against the full Pca203 reference (hap1, hap2, and unplaced contigs).
933 376,646 biallelic SNPs and 500 bootstraps were used. **a)** Collapsed tree view of
934 primarily Australian lineages (L1 to L18). Except for Pca203, the first two digits of the
935 name of the *Pca* isolate reflect year of collection, followed by state and sample
936 identifier. **b)** Collapsed tree view showing USA lineages. USA *Pca* lineages were not
937 numbered as the population is highly diverse. Tree branches are colored by country
938 of origin: AUS = Australia; SA = South Africa; TW = Taiwan; USA = United States of
939 America. Bootstrap values above 80% shown as white circles. Yellow labels indicate
940 isolates chosen for the haplotype atlas and red labels indicate isolates with published
941 references. Tree scales are mean substitutions per site.

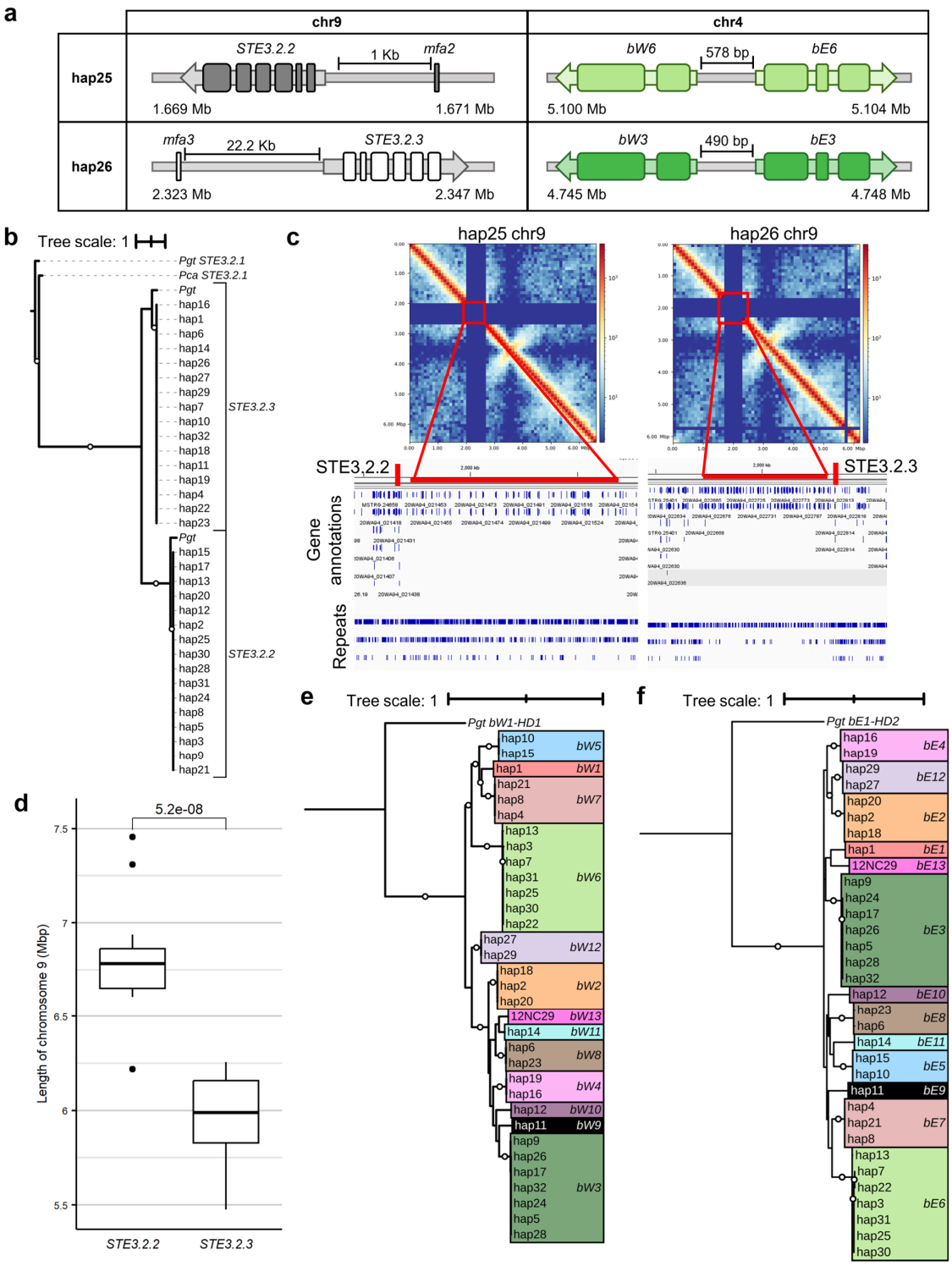


Figure S5. **a**) Orientation and arrangement of *STE3.2* and *mfa* (*PR*) alleles on chromosome 9 (chr9) and *HD* alleles on chromosome 4 (chr4) of hap25 and hap26 from *Puccinia coronata* f. sp. *avenae* (*Pca*) isolate 20WA94. **b)** Phylogenetic tree of *STE3.2* alleles in 32 *Pca* haplotypes, rooted at *Puccinia graminis* f. sp. *tritici* (*Pgt*)

947 *STE3.2.1.* **c**) Diagram showing chromatin contact maps, and gene and annotations
948 around the *STE3.2* locus (approximate position shown by labelled red vertical lines)
949 on chromosome 9 in hap25 and hap26. Repetitive sequences adjacent to the *STE3.2*
950 alleles correspond to multimapping reads resulting in loss of signal in the contact
951 maps. **d**) Boxplots showing chromosome 9 length distribution separated by which
952 *STE3.2* allele is present. **e**) Phylogenetic tree of *bW-HD1* alleles rooted at *Pgt bW1-*
953 *HD1* and **f)** *bE-HD2* alleles rooted at *Pgt bE1-HD2* in 32 *Pca* haplotypes. Colors
954 indicate identical *HD* alleles within each tree and allele pairs across trees (i.e. *bW3*
955 and *bE3* are always found in the same haplotype), as recombination between *bW* and
956 *bE* was not detected in our haplotype atlas**.** Branches with bootstraps over 80% are
957 shown with white circles at the midpoint. Tree scales are mean substitutions per site.

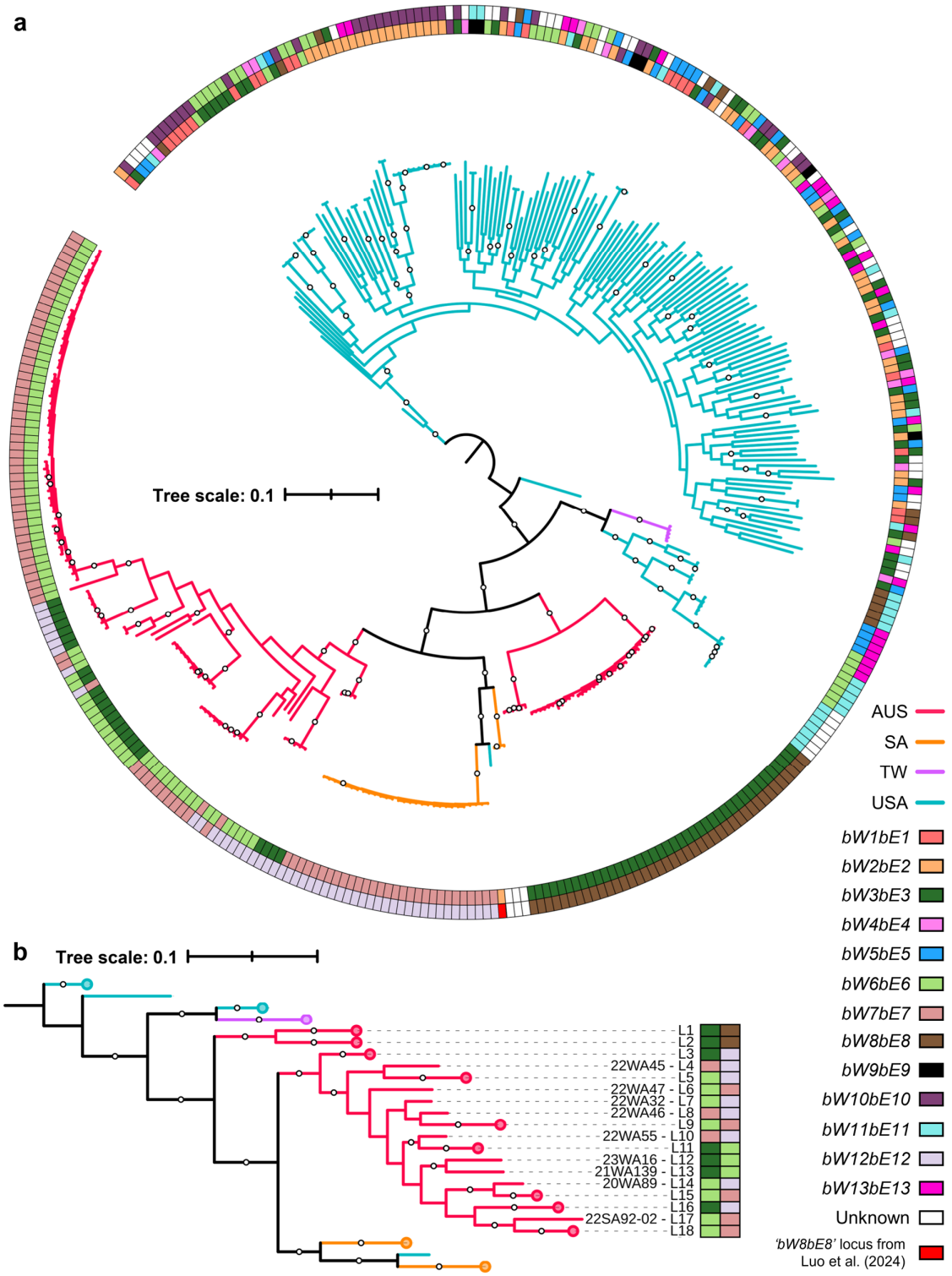


Figure S6. **A)** Midpoint rooted phylogenetic tree constructed by mapping short reads from 352 *Puccinia coronata* f. sp. *avenae* isolates and calling variants against the complete Pca203 genome (hap1, hap2, unplaced contigs). 376,646 biallelic SNPs were used over 500 bootstraps to produce the maximum likelihood tree. Bootstrap

963 values are percentages (100 = 100%). **b)** Collapsed view showing Australian lineages
964 L1-L18. *HD* alleles are shown as rectangles next to branches. Tree branches are
965 colored by country of origin: AUS = Australia; SA = South Africa; TW = Taiwan; USA
966 = United States of America. Tree scales are mean substitutions per site.

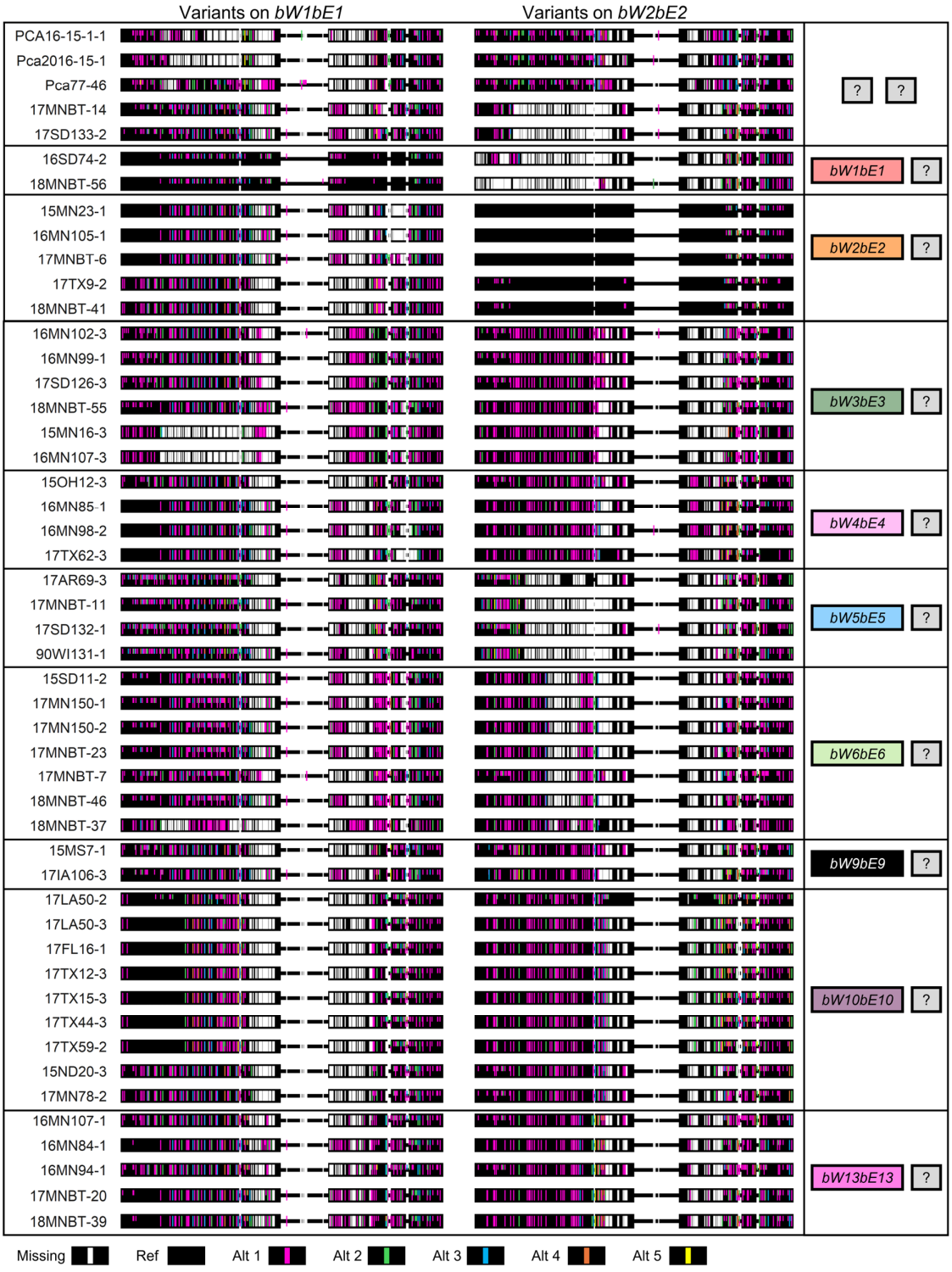


Figure S7. Variants called for 49 *Puccinia coronata* f. sp. *avenae* isolates with one or two unknown *HD* locus alleles in the *HD* locus regions of Pca203 hap1 (*bW1bE1*) and hap2 (*bW2bE2*). Line colors indicate missing (white), reference (black) and alternative

(pink, green, blue, orange, yellow) genotypes. Half-length lines indicate heterozygous sites and full-length lines indicate homozygous sites.

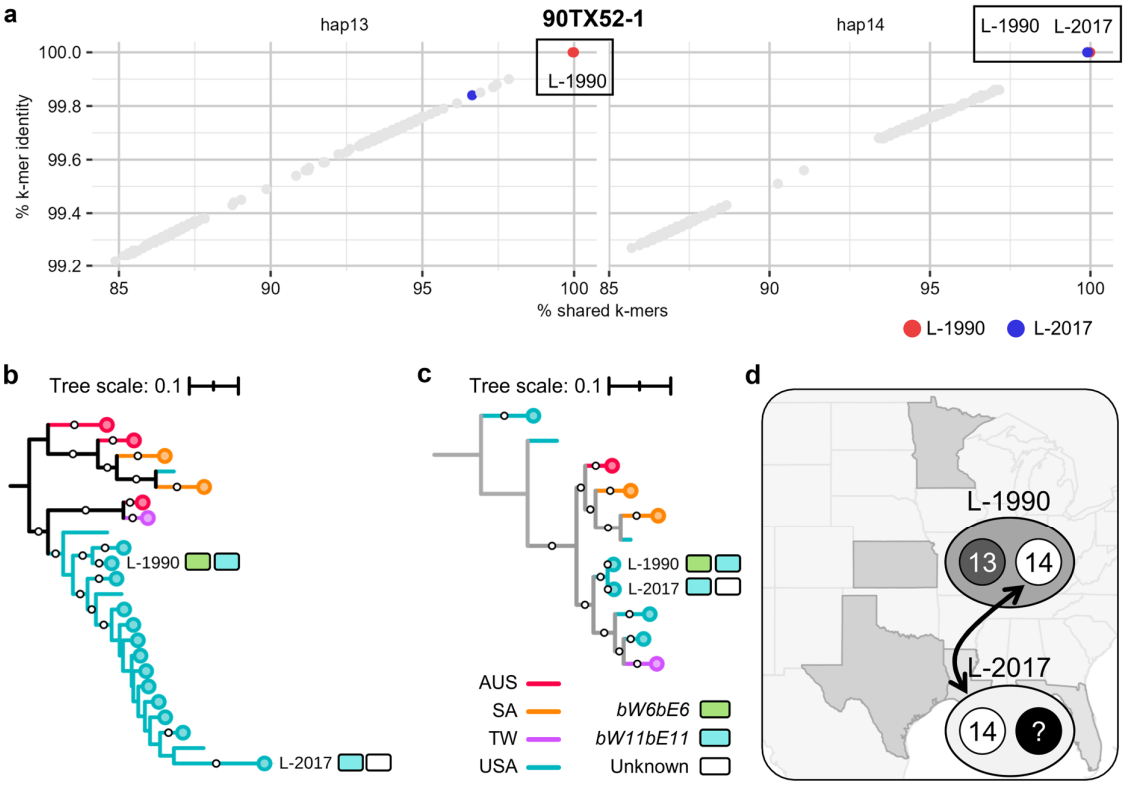


Figure S8. **a)** Plot of % *k*-mer identity (y-axis) versus % shared *k*-mers (x-axis) for 352 *Puccinia coronata* f. sp. *avenae* isolates on hap13 and hap14. Colors indicate relevant lineages; light grey points are all other isolates. **b-c)** Maximum likelihood phylogenetic trees constructed from variants from individual haplotypes **b**) hap13 (194,540 SNPs), **c)** hap14 (209,542 SNPs) with percentages for 500 bootstraps shown on branches. Tree branches are colored by country of origin: AUS = Australia; SA = South Africa; TW = Taiwan; USA = United States of America. Collapsed clades are indicated by circles at tips. *HD* alleles are shown as colored rectangles next to relevant clonal lineages. Tree scales are mean substitutions per site. **d)** Diagram of the proposed relationship between USA lineages and haplotypes involved in somatic hybridization.

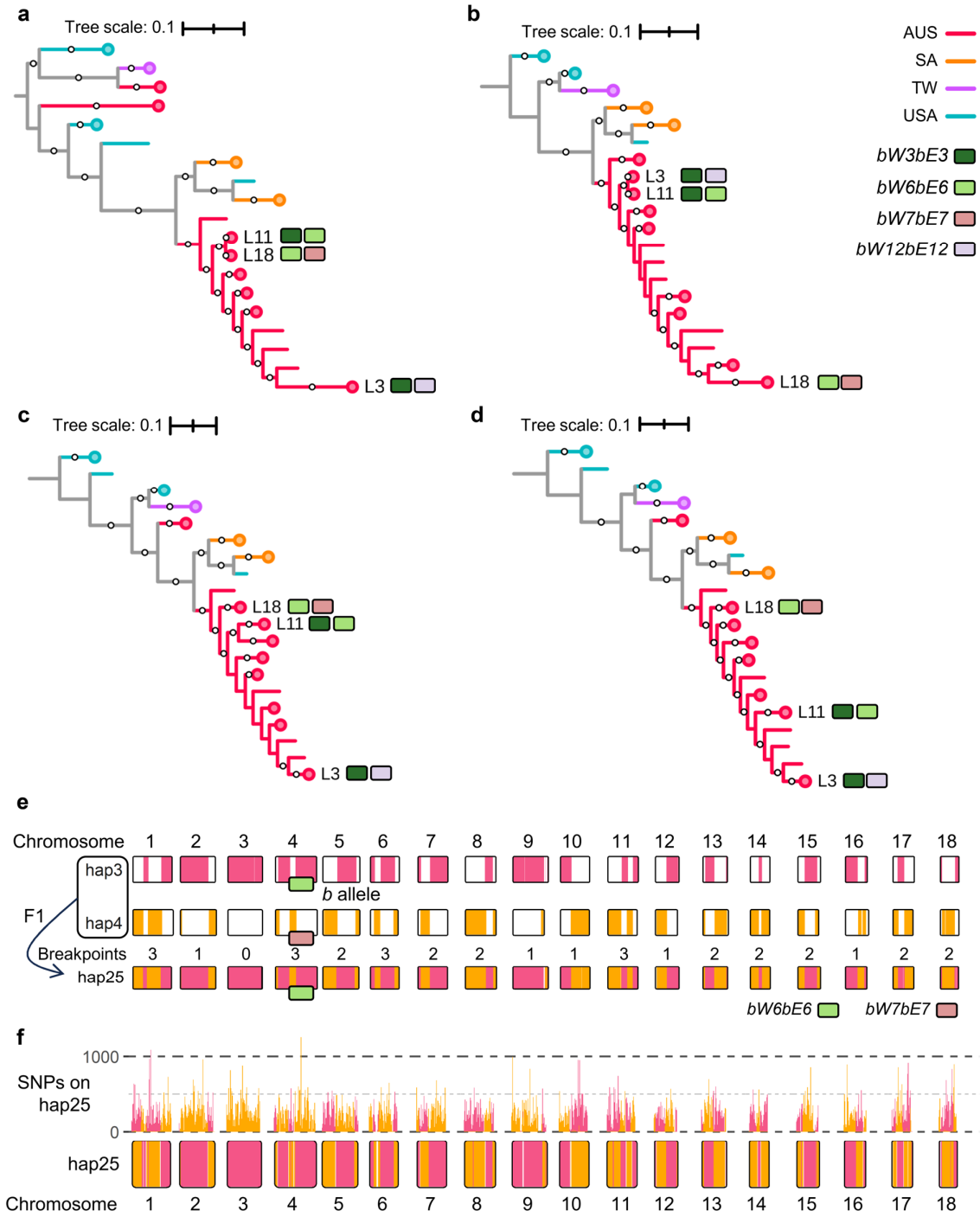


Figure S9. Maximum likelihood phylogenetic trees for 352 *Puccinia coronata* f. sp. *avenae* isolates constructed from variants from individual haplotypes **a**) hap25 (205,993 SNPs) **b**) hap26 (188,803 SNPs) **c**) hap3 (228,300 SNPs) **d**) hap4 (226,617 SNPs). Bootstrap supports 80% or higher are shown as circles at branch midpoints. Tree branches are colored by country of origin: AUS = Australia; SA = South Africa; TW = Taiwan; USA = United States of America. Collapsed branches are indicated by

991 circles at tips. *HD* alleles are shown as colored rectangles next to relevant clonal
992 lineages. Tree scales are mean substitutions per site. **e)** Genome alignment between
993 20NSW19 haplotypes (hap3, hap4) and hap25 from 20WA94. Chromosomes for the
994 three haplotypes are shown with high-identity alignments shown with colored fill (hap3
995 = pink, hap4 = orange). **f)** Non-reference variant counts within 100 kb bins from hap3
996 (pink) and hap4 (orange) on hap25 are shown in the histogram above hap25
997 chromosomes. Hap25 chromosome fill color is determined by identifying bins with a
998 low density (<50 SNPs/100 kb) of non-reference variants from hap3 (pink) and hap4
999 (orange).

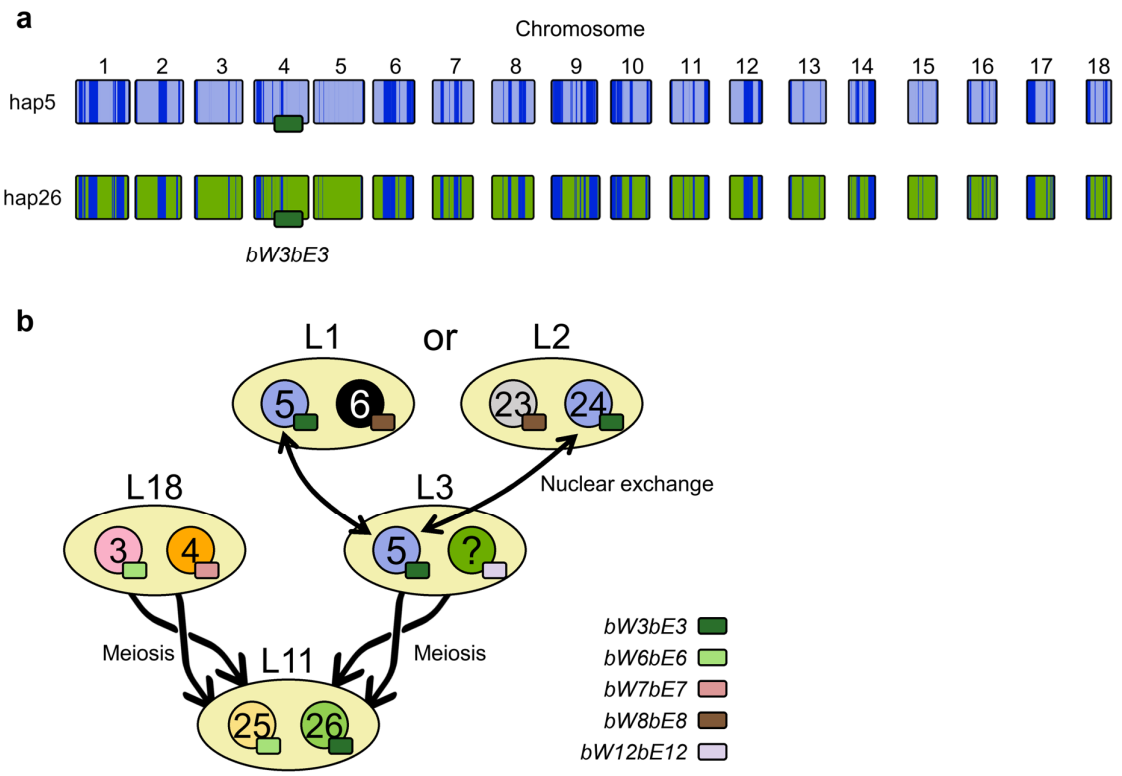


Figure S10. **a)** Regions of 20QLD86 hap5 against 20WA94 hap26 with over 95% identity when aligned are shown in dark blue. Unique sequences are shown in light blue (hap5) and green (hap26). **b)** Considering the mating type composition and shared sequences between hap5 and hap26, it is possible that hap26 from L11 is the progeny of hap5 and an unknown haplotype from L3. L3 may have been founded by somatic hybridization with L1 or L2.

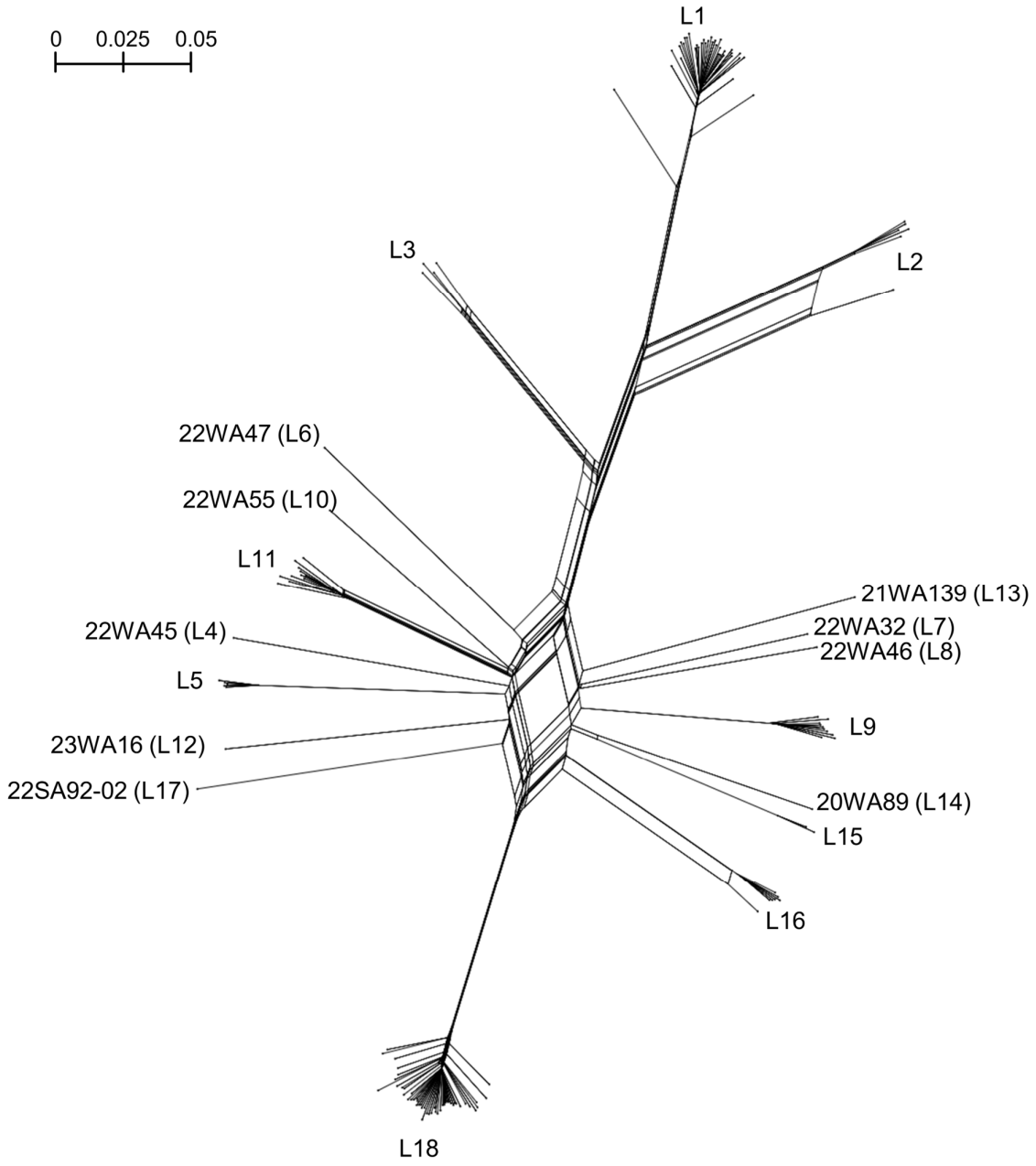


Figure S11. Splitstree network of 137 Australian *Puccinia coronata* f. sp. *avenae* isolates from 18 lineages generated from 391,118 SNPs from the entire Pca203 genome. Heterozygous sites were converted to missing. Scale indicates nucleotide substitutions per site.

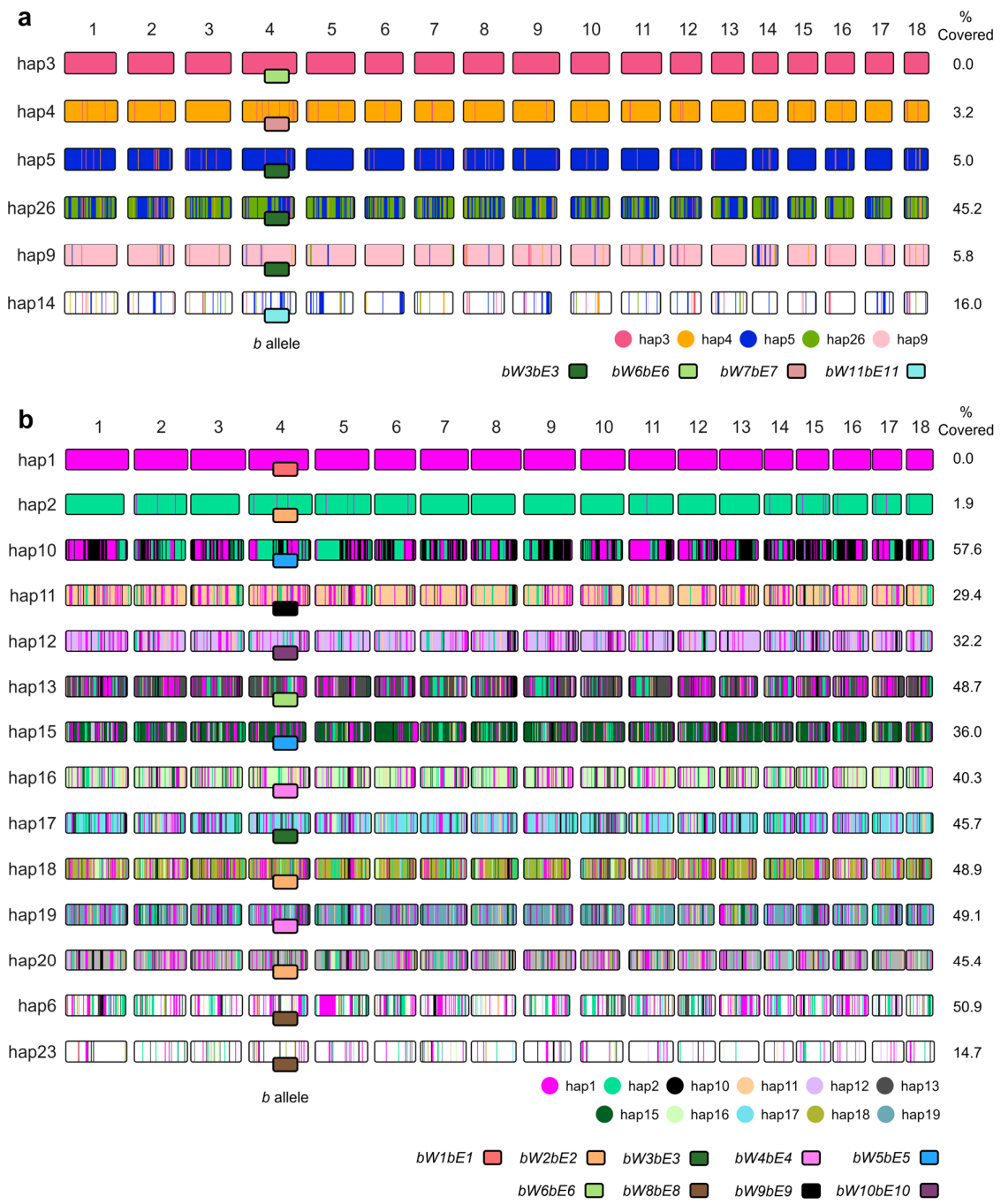


Figure S12. **a)** Shared haplotype blocks across six *Puccinia coronata* f. sp. *avenae* (*Pca*) haplotypes from 100 kb bins with less than 50 non-reference variants which were assigned hierarchically across hap3, hap4, hap5, hap26, hap9, and hap14. Shared sequences with hap3 were assigned across the other five haplotypes first, after which the unassigned areas of hap4 were evaluated on the four haplotypes below in the hierarchy, and so on. Regions in hap14 were not assigned to other haplotypes, so the unassigned hap14 sequence has a white fill. **b)** Generated the same way as panel **a)**,

1020 except shared haplotype blocks are assigned across 14 *Pca* haplotypes (hap1, hap2,
1021 hap10, hap11, hap12, hap13, hap15, hap16, hap17, hap18, hap19, hap20, hap6,
1022 hap23) with hap1 being the first in the hierarchy. Regions from hap6 were not assigned
1023 across hap23, so both haplotypes have unassigned sequences filled in white.
1024 Chromosome fill color represents unassigned regions or regions shared with
1025 haplotypes earlier in the hierarchy. Percent coverage by preceding haplotypes is
1026 shown on the right side. *HD* locus alleles indicated by fill color of rectangles at the
1027 chromosome 4 midpoint.