Parasexuality contributes to diversity and adaptive evolution of haploid (monokaryotic) fungi. However, non-sexual genetic exchange mechanisms are not defined in dikaryotic fungi (containing two distinct haploid nuclei). Newly emerged strains of the wheat stem rust pathogen, *Puccinia graminis* f. sp. *tritici* (*Pgt*), such as Ug99, are a major threat to global food security. Here, we provide genomics-based evidence supporting that Ug99 arose by somatic hybridisation and nuclear exchange between dikaryons. Fully haplotype-resolved genome assembly and DNA proximity analysis reveal that Ug99 shares one haploid nucleus genotype with a much older African lineage of *Pgt*, with no recombination or chromosome reassortment. These findings indicate that nuclear exchange between dikaryotes can generate genetic diversity and facilitate the emergence of new lineages in asexual fungal populations.
G

ereration of genetic diversity is crucial for the evolution of new traits, with mutation and sexual recombination as the main drivers of diversity in most eukaryotes. However, many species in the fungal kingdom can propagate asexually for extended periods and therefore understanding alternative mechanisms contributing to genetic diversity in asexual populations has been of great interest. Some fungi can use a parasexual mechanism to exchange genetic material independently of meiosis. This process involves anastomosis of haploid hyphae and fusion of two nuclei to generate a single diploid nucleus, which subsequently undergoes progressive chromosome loss to generate recombinant haploid offspring. Parasexuality has been described in members of the ascomycete phylum (64% of described fungal species) in which the dominant asexually propagating form is haploid. However, in basidiomycete fungi (34% of described species), the predominant life stage is generally dikaryotic, with two different haploid nuclei maintained within each individual. The role of non-sexual genetic exchange between such dikaryons in generating genetic diversity is not known.

Basidiomycetes include many fungi with critical ecosystem functions, such as wood decay and plant symbiosis, as well as agents of important human and plant diseases. Rust fungi (subphylum Pucciniomycotina) comprise over 8000 species including many pathogens of major agricultural and ecological significance. These organisms are obligate parasites with complex life cycles that can include indefinite asexual reproduction through infectious dikaryotic urediniospores. Early researchers speculated that rust fungi can exchange genetic material during the asexual phase based on the isolation of new strains, after co-infection with two potential parental isolates, with novel virulence phenotypes. However, these hypotheses could not be tested molecularly at the time. Some naturally occurring rust pathotypes have been suggested to have arisen by somatic hybridisation and genetic exchange based on limited molecular evidence of shared isozyme or random amplified polymorphic DNA (RAPD) markers. Mechanisms underlying genetic exchange are unknown, but may involve hyphal anastomosis followed by nuclear exchange and/or nuclear fusion and recombination. Recent advances in assembling complete karyon sequences in rust fungi provide the opportunity to definitively detect and discriminate between nuclear exchange and recombination.

The Ug99 strain (race TTKSK) of the wheat stem rust pathogen *Puccinia graminis f. sp. tritici* (Pgt) presents a significant threat to global wheat production. It was first detected in Uganda in 1998 and described in 1999, and has since given rise to an asexual lineage that has spread through Africa and the Middle East causing devastating epidemics. The origin of the Ug99 lineage is unknown, although it is genetically distinct from other Pgt races. This indicates that Ug99 is not likely derived by mutation of longstanding stem rust asexual lineages such as the race 21 group, which has been predominant in southern Africa at least since the 1920s and spread to Australia in the 1950s.

Here, we generate haplotype-phased genome references for the original Ug99 isolate collected in Uganda and an Australian Pgt isolate of pathotype 21-0. We show by genome comparison that Ug99 shares one haploid nucleus genotype with Pgt21-0, with no recombination or chromosome reassembly. This indicates that Ug99 arose by somatic hybridisation and nuclear exchange between an African member of the Pgt 21 lineage and an unknown isolate. Thus, nuclear exchanges between dikaryotic fungi can contribute to the emergence of new variants with significant epidemiological impacts.

**Results**

**Haplotype-phased genome assembly.** A single pustule derived from the original Ug99 isolate of Pgt was purified and its pathotype was confirmed using the standard wheat differential set. We generated polished long-read genome assemblies for both Ug99 and the Australian stem rust isolate Pgt21-0 using single-molecule real-time (SMRT) and Illumina sequence data. Genome assembly of Pgt21-0 resulted in 410 contigs with a total size of 177 Mbp and a contig \(N_50\) of 1.26 Mbp. Similarly, the size of the genome assembly of Ug99 was 176 Mbp represented in 514 contigs with contig \(N_50\) of 0.97 Mbp. Thus, both assemblies were twice the size of a collapsed haploid assembly previously generated for a North American Pgt isolate (88 Mbp), suggesting that in each assembly the sequences of the two haploid karyons were fully represented independently. Consistent with this, both genome assemblies contained over 96% of conserved fungal BUSCO genes, with the majority present in two copies (Supplementary Table 3). Furthermore, the Pgt21-0 assembly contained 69 telomeres, out of a total of 72 expected for a dikaryotic genome given the known haploid chromosome number of eighteen. To identify sequences representing alternate haplotypes within each assembly we developed a gene synteny approach to assign contigs to groups representing paired homologous haplotypes (Fig. 1). Using this approach, homologous pairs of sequences from each haplotype were assigned to 44 bins in Pgt21-0 and 62 bins in Ug99, which represented over 94% of each assembly (Supplementary Table 3 and Supplementary Data 2). Three of the 18 chromosomes in Pgt21-0, and two in Ug99 seemed to be fully assembled as these bins contained telomere sequences at each end.

The *AvrSr50* and *AvrSr35* genes encode dominant avirulence factors recognised by wheat resistance genes. These two genes are located in close proximity (<15 kbp) to each other in the genome assemblies and both haplotypes of this locus were assembled as alternate contigs in Pgt21-0 and Ug99 (Fig. 2a). Both isolates were heterozygous for *AvrSr50* with one allele containing a ~26-kbp-insertion. Pgt21-0 was also heterozygous for *AvrSr35*, with one allele containing a 400-bp-MITE insertion previously described. PCR amplification from the Ug99 strain had previously identified only a single *AvrSr35* allele, suggesting that it was homozygous in Ug99. However, the Ug99 genome assembly contained a second allele of this gene, which contained a 57-kbp-insertion that would have prevented its PCR amplification in the Salcedo et al. study. The presence of this insertion in Ug99 was supported by DNA read (PacBio and Illumina) alignments across this genomic region and confirmed by DNA amplification and amplicon sequencing of flanking border regions (Supplementary Fig. 1). Thus, Ug99 is also heterozygous for avirulence on *Sr35*, and may therefore mutate to virulence on this wheat resistance gene more readily than if it were homozygous. This is an important finding as it will inform *Sr35* deployment strategies against Ug99. Strikingly, the *AvrSr35-virSr50* haplotype of this locus is very similar in structure in Ug99 and Pgt21-0 (Fig. 2a) and shares >99% sequence identity, while the two alternate haplotypes are quite different. We therefore compared the larger genomic regions containing these loci in each isolate: namely bin 06 in Pgt21-0, which is ~3.5 Mbp and includes telomeres at both ends, and bins 15 (1.8 Mbp) and 23 (1.2 Mbp) in Ug99 (Supplementary Fig. 2a). One haplotype (designated A) was >99.7% identical in Ug99 and Pgt21-0, while the other two haplotypes (B and C) were highly divergent from each other and from haplotype A (Fig. 2b, Supplementary Fig. 2b and Supplementary Table 4). Only 71–78% of the sequences from each haplotype aligned, with an average identity of ~95% across the aligned regions, yielding total identities of only 68–76%
between these three distinct haplotypes. The high similarity between the A haplotypes of this chromosome suggested that Ug99 and Pgt21–0 may share large portions of their genomes, potentially up to an entire haploid genome copy.

Whole-genome haplotype assignment and comparison. We used a read subtraction and mapping approach (Fig. 3a, b) to identify genome regions in the Ug99 and Pgt21–0 assemblies that showed high similarity and may be derived from a shared haplotype. Illumina reads from each isolate were mapped to the genome reference of the other isolate. Reads that failed to map were retained, and this subtracted read set, in which sequences common to both isolates had been removed, was re-mapped to their original genome reference. The subtracted read coverage depth was compared to the coverage depth obtained when using all reads (Supplementary Fig. 3a, b). Contigs representing sequences shared by both isolates (designated as haplotype A) displayed a very low subtracted read coverage depth (Supplementary Data 3). In contrast, sequences unique to Pgt21–0 or Ug99 (designated as haplotypes B or C, respectively) retained a relatively high subtracted read coverage depth. Some contigs in each assembly appeared to be chimeric with distinct regions assigned to opposite haplotypes, and these contigs were divided into separate fragments (Supplementary Data 4) for subsequent haplotype comparisons. Approximately half of each genome assembly was assigned to either the A, B or C haplotypes (Fig.3c) and importantly one set of homologous sequences from each bin was assigned to each haplotype (Supplementary Data 3). The A, B and C haplotype sets contained 95–96% of conserved fungal BUSCO genes (Fig. 3d), indicating that each represents a full haploid genome equivalent. Consistent with this, the haplotypes were highly contiguous (Fig. 4). Overall sequence identity between the A haplotypes of Pgt21–0 and Ug99 was over 99.5%, with structural variation (large insertions/deletions) representing only 0.46% of the haplotype sizes, and with just 0.08% sequence divergence in aligned regions (Fig.4a, Table 1, and Supplementary Table 5). In contrast, only 91–93% of the A, B or C haplotypes could be aligned to each other, with an average sequence identity of ~96% across the aligned regions, giving total identities of 87–91% (Fig. 4b–d, Table 1 and Supplementary Table 5).
Structural variation between these haplotypes accounted for 6.7–8.6% of the haploid genome sizes. There were only ~9000 SNPs (0.1/kbp) between the two A haplotypes, versus 876,000 to 1.4 million SNPs (11–18/kbp) between the A, B and C pairs, which is consistent with estimates of heterozygosity levels in *Pgt* 21-0 (haplotypes A and B) based on variant detection from Illumina read mapping.[21] The high similarity between the A haplotypes, and divergence between A, B and C haplotypes was also supported by Illumina read coverage and SNP calling analysis (Supplementary Fig. 3c–e) showing that Ug99 and *Pgt* 21–0 share one nearly identical haplotype genome copy.

**Assessment of inter-nuclear recombination.** We tested two hypotheses that could explain the shared haplotype between Ug99 and *Pgt* 21-0: (1) Ug99 arose by a somatic hybridisation event in which an isolate of the race 21 lineage donated an intact nucleus of the A haplotype (Fig. 5a); and (2) Ug99 arose by a sexual cross in which one haploid pycnial parent was derived from a race 21 lineage isolate after meiosis (Fig. 5b). Under both scenarios, the A haplotype of Ug99 represents one entire haploid nuclear exchange scenario, the nucleus that was derived from the race 21 lineage isolate. In the sexual cross hypothesis. Furthermore, only four structural variants larger than 10 kbp were detected between the A haplotype.

![Diagram of genomic regions containing *AvrSr35* and *AvrSr50*](image)

**Fig. 2** A common haplotype containing *AvrSr50* and *AvrSr35* is shared between *Pgt* 21-0 and Ug99. **a** Diagram of genomic regions containing *AvrSr50* and *AvrSr35* alleles in *Pgt* 21-0 and Ug99. Numbers above tracks correspond to contig coordinates and the sense of the DNA strand is indicated as + or −. Predicted gene models (including introns) are depicted as dark grey boxes and intergenic spaces are shown in light grey. Coloured arrows indicate location and direction of *AvrSr50* and *AvrSr35* genes, with size and position of insertions shown in yellow. Intergenic distances between *AvrSr50* and *AvrSr35* are indicated by brackets. **b** Total sequence identity between contigs representing homologous chromosomes of different haplotypes (coloured bars) containing the *AvrSr50*/*AvrSr35* locus (dotted white boxes). Telomere sequences are represented in grey. Chromosome size = ~3.5 Mbp.

![Diagram of genomic regions containing *AvrSr35* and *AvrSr50*](image)
Chromosome assembly and assessment of reassortment. Combining Hi-C scaffolding data with the bin and haplotype assignment information for the *Pgt*21-0 assembly allowed us to construct 18 chromosome pseudomolecules for each of the A and B haplotypes (Fig. 6a, Supplementary Table 6 and Supplementary Data 6). These covered a total of 170 Mbp and ranged from 2.8 to 7.3 Mbp in size, consistent with relative chromosome sizes from karyotype analysis. The A and B chromosomes were collinear.
except for two translocation events (Fig. 6b). In each case these were supported by contigs that spanned the translocation breakpoints. Rescaffolding the separated fragments of these contigs using Hi-C data supported the original contig assembly, indicating that these are true translocation events within the A or B genomes. The haplotype A chromosomes showed high collinearity with the Ug99 A haplotype contigs (Fig. 6c).

Approximately 65% of the total Hi-C read pairs represented links between physically contiguous sequences on the same chromosome, while the remaining pairs connected sites distributed across the genome. Because Hi-C DNA crosslinking is performed in intact cells, these non-scaffolding linkages should preferentially form between chromosomes that are located in the same nucleus. Indeed, all chromosomes of the A haplotype showed a much higher proportion of Hi-C read pair links to other chromosomes of the A haplotype (~85%) than to chromosomes of the B haplotype (~15%) (Fig. 6d), indicating that they are all located in the same nucleus. Thus, the A haplotype of Ug99 derives from a single nucleus of *Pgt*21-0. Similarly, 17 of the B haplotype chromosomes in *Pgt*21-0 showed stronger linkage to other B chromosomes (~90%) than to A chromosomes (~10%) (Fig. 6e). However, chromosome 11B showed the inverse, suggesting that both homologues of this chromosome are in the same nucleus. This implies that a single chromosome exchange event occurred during asexual propagation of the *Pgt*21-0 isolate, after its divergence from the race 21 lineage branch giving rise to Ug99.

Overall the whole-genome comparison data demonstrate that Ug99 shares one full haploid nuclear genome with the *Pgt*21-0 isolate with no recombination events within chromosomes and no reassortment of chromosomes from different nuclei. These facts are inconsistent with a sexual origin, and strongly support that the Ug99 lineage arose by a somatic hybridisation event involving one parent derived from the African race 21 lineage and another parent of unknown origin exchanging whole nuclei (Fig. 7).

### Comparison of gene content between haplotypes

Annotation of the *Pgt*21-0 and Ug99 genome assemblies predicted similar gene numbers (~18,000) in each haplotype (Supplementary Table 7). Gene orthology analysis indicated that 65–70% of genes in each of

---

**Table 1 Intra- and inter-isolate sequence comparison of entire haplotypes in Ug99 and *Pgt*21-0**

| Isolate comparison   | Bases aligned (%) | Sequence divergence (%) | Number of variants | Total variant size |
|----------------------|-------------------|-------------------------|--------------------|-------------------|
|                      |                   |                         |                    | Mbp               |
| **Isolate comparison** | **Bases aligned (%)** | **Sequence divergence (%)** | **Number of variants** | **% of genome** |
| 21-0A vs Ug99 A      | 99.64             | 0.08                    | 491                | 0.82             | 0.46          |
| Ug99 A vs Ug99 C     | 91.52             | 4.08                    | 2571               | 13.69            | 7.88          |
| 21-0A vs *Pgt*21-0 B | 91.38             | 4.19                    | 2696               | 15.01            | 8.56          |
| 21-0 B vs Ug99 C     | 93.44             | 2.4                     | 1910               | 11.50            | 6.69          |

---

**Fig. 4** *Pgt*21-0 and Ug99 share one nearly identical haploid genome. **a-d** Dot plots illustrating sequence alignment of complete haplotypes. X- and y-axes show cumulative size of the haplotype assemblies depicted by coloured bars to the right and top of the graphs. Colour key indicates sequence identity ratios for all dot plots.
the A, B and C haplotypes were shared and represent a core Pgt gene set, while the remainder were present in only one or two haplotypes (Supplementary Table 8). Mapping of orthologous gene pairs supported the overall synteny of the Pgt21-0 A and B chromosome assemblies and confirmed the translocations observed between chromosomes 3 and 5, and between 8 and 16 (Fig. 8a). Genes encoding secreted and non-secreted proteins showed a similar distribution across the chromosomes, while repeat sequences displayed an inverse distribution to genes (Fig. 8b, Supplementary Fig. 5). The location of secreted protein genes in gene-rich rather than repeat-rich regions is consistent with the absence of two-speed genome architecture in the related rust fungal species P. coronata and P. striiformis12,13.

Both Ug99 and Pgt21-0 are heterozygous at the predicted a and b mating type loci (Supplementary Fig. 6), despite Ug99 being derived by a non-sexual mechanism. This is consistent with an expectation that formation and maintenance of a stable dikaryon requires two distinct compatible mating types41. We observed multiple alleles at the b locus on chromosome 9, which encodes the divergently transcribed transcription factors BE and BW, with variants sharing 70–80% amino acid identity. Cuomo and colleagues28 previously described two alleles, b1 and b2, for this locus in Pgt. Pgt21-0 contained the b1 allele and a novel b3 allele, while Ug99 contained b3 and another novel b4 allele. Both isolates were heterozygous for the + and – alleles of the a locus on chromosome 4, which encodes a pheromone (mfa) and pheromone receptor (STE3) pair. However, one of the receptor alleles in Ug99 contained a single-nucleotide deletion that resulted in truncation of the last 48 amino acids of the protein. Thus, the mating type system for Pgt appears to consist of two independent loci, one di-allelic and one multi-allelic.

**Phylogenetic analysis of global Pgt isolates.** We used the haplotype-phased genome references for Pgt21-0 and Ug99 to determine genetic relationships within a set of global Pgt isolates.
Maximum likelihood trees based on whole-genome SNPs (Fig. 9a and Supplementary Fig. 7a) showed a very similar overall topology to that reported previously for most of these isolates. The 5 isolates of the Ug99 lineage, and the 13 South African and Australian isolates each formed a separate tight clade, consistent with their proposed clonal nature. However, tree building using filtered SNPs from just the A haplotype resulted in the formation of a single clade containing the Ug99, South African and Australian isolates, which indicates the clonal derivation of this nucleus among these isolates (Fig. 9b, Supplementary Fig. 7b, c). The Ug99 group forms a subclade within the race 21 group consistent with a derived origin. In contrast, these groups remained in two distant clades in phylogenies inferred using filtered SNPs from the B genome. Surprisingly, in this case two isolates from the Czech Republic and three isolates from Pakistan were now...
located in a single clade with the South African and Australian isolates (Fig. 9c). This suggests that these isolates contain a haplotype closely related to the B genome of the race 21 lineage and may also have arisen by somatic hybridization and nuclear exchange. A phylogeny based on the C genome SNPs grouped isolate IR-01 from Iran with the Ug99 lineage (Fig. 9d), suggesting that these isolates share the C haplotype. IR-01 could represent a member of the parental lineage that donated the C nucleus to Ug99, or alternatively may have acquired the C nucleus from Ug99. Notably, this was the only isolate that shared the \textit{AvrSr35} 57 kbp insertion allele identified in Ug99 (Supplementary Fig. 8a). The relationships between these putative hybrid isolates were also supported by the patterns of homozygous and heterozygous SNPs detected in each haplotype (Supplementary Fig. 8b–e). The incongruities between phylogenies generated based on different haplotypes highlight the difficulty of inferring relationships between isolates based on whole-genome SNP data without haplotype resolution. Overall, these observations suggest that somatic hybridisation and nuclear exchange may be a common mechanism generating genetic diversity in global populations of \textit{Pgt}.

**Discussion**

Although sexual reproduction of \textit{Pgt} can generate individuals with novel genetic combinations, the completion of the sexual cycle requires infection of an alternate host, \textit{Berberis} spp. (barberry)\textsuperscript{30}. In parts of the world where barberry is scarce or absent, either due to eradication programmes or its natural distribution, \textit{Pgt} is restricted to asexual propagation with new diversity often arising by mutation or migration\textsuperscript{19,20}. Somatic hybridisation provides an alternative explanation for the appearance of new races not derived by stepwise mutation. Hybrids with high adaptive value in agroecosystems may establish new lineages of rust fungi. Although we did not observe any recombination between two haplotypes\textsuperscript{23}. Thus, genetic exchange between haploid nuclei may occur as rare events during asexual propagation of a single lineage in \textit{Pgt}. Whether extensive genetic exchange similar to ascomycete parasexuality\textsuperscript{2} can also occur between rust nuclei during hybridisation remains to be determined. This may require controlled infection experiments, as such recombinant hybrids would be difficult to distinguish from the products of sexual recombination in field derived-strains, especially given the potential for long-range spore dispersal.

Although there is now clear evidence of nuclear exchange between dikaryons in \textit{Pgt}, nothing is known of how this process occurs or is regulated. It differs from parasexuality in ascomycetes\textsuperscript{2}, as the dikaryotic state is maintained with no nuclear fusion or haploidisation resulting in chromosome reassortment. Wang and McCallum\textsuperscript{33} observed the formation of fusion bodies where germ tubes of different \textit{P. triticina} isolates came into contact, with the potential for nuclear exchange at these junctions. It has been proposed that mating type loci contribute to determining the compatibility between isolates for the formation of hybrids\textsuperscript{6,7}, but this also remains to be confirmed experimentally. Our findings provide a new framework to take advantage of haplotype genome resolution to understand the role of somatic exchange in population diversity of rust fungi.

Extended dikaryotic developmental stages are common in many other fungi, especially basidiomycetes. Indeed, separation of karyogamy (fission of haploid nuclei to form a diploid nucleus) from gamete fusion is a feature unique to the fungal kingdom\textsuperscript{1}. However, it is unclear why fungi maintain an extended dikaryotic stage prior to formation of a diploid nucleus as a precursor to sexual reproduction\textsuperscript{34}. One possibility is that the ability to exchange haploid nuclei offers an advantage over the diploid state due to the enhanced genetic variation in long-lived asexual...
There is also evidence for somatic exchange of genetic markers in dikaryotes of the mushroom *Schizophyllum commune*, which belongs to another Basidiomycete subphylum, Agaricomycotina. Arbuscular mycorrhizae (AM) fungi are another ancient fungal lineage whose spores contain hundreds of nuclei and for which no sexual stages have been described, raising questions of how these lineages have survived. Recently some dikaryotic-like AM isolates possessing two divergent classes of nuclei have been observed. Nuclear exchange between dikaryotes could be another driver of genetic variation in these fungi. Evidently, the members of the fungal kingdom display remarkable genetic plasticity and further investigation is required to reveal the mechanism, prevalence and evolutionary importance of nuclear exchange in dikaryotic and multinucleate fungi.

**Methods**

**Fungal stocks and plant inoculation procedures.** Pgt isolates Ug99, UVPgt55, UVPgt59, UVPgt60 and UVPgt61 collected in South Africa were transferred to the Biosafety Level 3 (BSL-3) containment facility at the University of Minnesota for growth and manipulation. Samples were purified by single pustule isolation and then amplified by 2–3 rounds of inoculation on the susceptible wheat cultivar McNair. Virulence pathotypes and purity of each isolate were confirmed by inoculation onto the standard wheat differential set (Supplementary Data 1). Other isolates used in this study were Pgt21-0, which was first isolated in Australia in 1954, the North American isolate CRL 75-36-700-3 (pathotype SCCL) and Kenyan isolate 04KEN156/04 (pathotype TTKSK). For rust inoculations, urediniospores retrieved from −80 °C were heat treated (45 °C for 15 min) and suspended in mineral oil (Soltrol 170, Phillips Petroleum, Borger, TX, USA) at 14 mg/ml. Seven-day-old seedlings were spray-inoculated at 50 μl/plant and kept in a dark mist chamber at 22–25 °C, 100% humidity for 16 h. Subsequently, plants were exposed to light (150–250 μmol photons s^-1 m^-2) for 5 h in the mist chamber and then transferred to a controlled growth chamber (18 h/6 h of light/dark, 24 °C/
DNA extraction and sequencing of rust isolates. High molecular weight DNA of Ug99 and Pgt21-0 was extracted from 300 to 350 mg urediniospores as described\(^\text{38}\), with the following modifications: (1) for Phenol:Chloroform:isoamyl alcohol extractions, samples were centrifuged at 4 °C and 5000 × g for 20 mins; (2) a wide-bore 1-ml pipette tip was used to transfer the DNA pellet; (3) samples were incubated for 1 h at 28 °C with 200 μl of 10 × loading buffer before re-centrifugation. DNA concentration was quantified using a NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA). Approximately 10 μg DNA from Ug99 and Pgt21-0 was sequenced using PacBio SMRT sequencing (Pacific Bioscience, Menlo Park, CA, USA) at either the Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc. (Frederick, MD, USA) or the Ramaciotti Centre (Sydney, Australia), respectively; DNA was concentrated and cleaned using AMPure PB beads for Ug99 or AMPure XP beads for Pgt21-0 (Pacific Biosciences). DNA quantification and size assessment were conducted using a NanoDrop (Thermo Fisher Scientific) and 2200 TapeStation instruments (Agilent Technologies, Santa Clara, CA, USA). DNA was sheared to a targeted average size of 20 kb using Q5 (New England Biolabs, Beverly, MA, USA). Libraries were constructed following the 20 kb Template Preparation BluePippin Size-Selection System protocol (Pacific Biosciences) using a BluePippin instrument (Sage Science, Beverly, MA, USA) with a 0.75% agarose gel. In addition, 4 SMRT cells were run on PacBio RSII platform using P6-C4 chemistry, the Sequel Binding Kit 2.0 (Pacific Biosciences), loading 0.15 nM MagBead loading and 240-min movie lengths. For Pgt21-0, 17 SMRT cells were sequenced on a PacBio Sequel platform using P6-C4 chemistry, the Sequel Binding Kit 2.0 (Pacific Biosciences), diffusion loading, 10-h movie lengths and Magbead loading at 2 pM (3 cells) or 4 pM (2 cells). In addition, 4 SMRT cells were run on PacBio RSII platform using P6-C4 chemistry, 0.15 nM MagBead loading and 360-min movie lengths. For Pgt21-0, 17 SMRT cells were sequenced on the RSII platform using P6-C4 chemistry, Magbead loading (0.12–0.18 nM) and ≥240-min movie lengths.

Genomic DNA for Illumina sequencing was extracted from 10 to 20 mg urediniospores of Ug99, UVPt55, 59, 60 and 61 using the OmniPrep™ kit (G-Biosciences, St. Louis, MO, USA) following the manufacturer’s instructions. TruSeq Nano DNA libraries were prepared from 300 ng of DNA and 150 bp paired-end sequence reads were generated at the University of Minnesota Genomics Center on the Illumina NextSeq 550 platform using Illumina Real-Time Analysis software version 1.18.64 for quality-scored base calling.

18 °C for day/night, 50% relative humidity). Spores were collected 9 and 14 days post inoculation (dpi) and maintained at −80 °C.

Fig. 9 Somatic hybridisation in Pgt evolution. a Phylogenetic analysis of Pgt isolates from diverse countries of origin (colour key) using a RAxML model and biallelic SNPs called against the full dikaryotic genome of Pgt21-0. Scale bar indicates number of nucleotide substitutions per site. Red asterisks indicate P. graminis f. sp. avenae isolates used as an outgroup. b Dendrogram inferred using biallelic SNPs detected against haplotype A of Pgt21-0. c Dendrogram inferred using SNPs detected against haplotype B of Pgt21-0. d Dendrogram inferred from SNPs detected in haplotype C of Ug99
De novo long read assembly. Genome assemblies of Ug99 and Pgt21-0 were built from PacBio reads using Canu version 1.6\textsuperscript{21} with default parameters and an estimated genome size of 170 Mbp. Assemblies were polished with the Arrow algorithm using the raw PacBio reads in the sa3\_ds sequencing pipeline in pbsmrtpipe workflow within SMRTLINK/5.10 (Pacific Biosciences). Assemblies were further polished by two rounds of Pilon\textsuperscript{40} with the option --all using Illumina reads from Ugh99 (this work) or Pgt21-0 (NCBI SRA run Accession SR26420301) and BLASTN search (version 2.2.17) against the NCBI nr database (downloaded on 4/11/2018) with E-value set as 1\textsuperscript{–}10 identified two contigs in the Ug99 assembly with significant hits to plant RNA and chloroplast sequences and these were removed.

Covered and Illumina reads were mapped to the assembly using BWA-MEM (version 0.7.17)\textsuperscript{41} and BAM files were indexed and sorted using SAMtools (version 1.9)\textsuperscript{42}. Read coverage analysis using genomeCoverageBed in BEDTools (version 2.27.1)\textsuperscript{43} identified 144 small contigs (<50 kb) in the Ug99 assembly with low coverage (<2x) for both short and long reads and these contigs were also excluded from the final assembly. Genome assembly metrics were assayed using QUAST (version 4.3)\textsuperscript{44} and completeness was assessed via benchmarking universal single-copy orthologs (BUSCOs) of the basidiomycota as fungal lineage and \textit{Ustilago maydis} as the reference species for AUGUSTUS gene prediction\textsuperscript{45} in the software BUSCO v2.0 (gene mode)\textsuperscript{46}. Telomeric sequences were identified using either a high stringency BLAST with 32 repeats of TTAGGG as query or a custom python script to detect at least five CCCTAA or TTAGGG repeats in the assemblies. Repeats of at least 60 bp length and occurring within 100 bp of the contig end were defined as telomeric sequences.

Detection of alternate contigs and bin assignment. To identify contigs representing corresponding haplotypes (Fig. 1) 22,484 predicted Pgt gene coding sequences\textsuperscript{21} were screened against the genome assemblies using BLITZ (Blat-like examined by alignment and similarity plotting using D-genies\textsuperscript{47}. Contigs repre-sented Sanger sequencing and alignment to \textit{AvrSr50} and \textit{AvrSr35} (version 0.7.17)\textsuperscript{41} and \textit{AvrSr35} gene sequences were identified by BLASTN search against customised databases for the Ug99 and Pgt21-0 genome assemblies. Illuminia and PacBio reads of Ug99 mapped to the genome assembly were visualised in the Integrative Genomics Viewer (IGV). To validate the presence of the 57-kbp insert in \textit{AvrSr35}, flanking and internal sequences were amplified from genomic DNA extracted using the \textit{Omniprep} kit (G-Biosciences) from \textit{Ug99}, 04KEN156/04, and CRL 75-36-700-3. PCR was performed using Phusion high-fidelity DNA polymerase according to the manufacturer’s recommendations (New England BioLabs Inc., Ipswich, MA, USA) and primers listed in Supplementary Table 9. The amplification products were separated by agarose gel and stained using SYBR Safe DNA gel stain (Invitrogen). Specific bands were cleaned using NucleoSpin gel clean-up kit (Takara Bio, Mountain View, CA, USA) for subsequent Sanger sequencing and alignment to \textit{AvrSr35} alleles. Gene models in the \textit{AvrSr35} and \textit{AvrSr50} locus were depicted using GenomicFeatures\textsuperscript{48} and ggbioc\textsuperscript{49} in a custom R script.

Haplotype assignment by read cross-mapping and subtraction. Illumina reads from Pgt21-0 (NCBI SRA SR26420301) were trimmed (“Trim sequences” quality limit = 0.01) and mapped to the Ug99 reference assembly using the “map reads to reference” tool in CLC Genomics Workbench version 10.0.1 or later with high stringency parameters (similarity fraction 0.99, length fraction 0.98, global align-ment). Unmapped reads (Ug99-subtracted reads) were retained and mapped back to the Pgt21-0 assembly contigs using the same parameters. The original Pgt21-0 reads were also mapped to the Pgt21-0 assembly and the read coverage for each contig compared to the Ug99-subtracted reads. Contigs with very low coverage (<2X total and <10% of the original read coverage) with the Ug99-subtracted reads were designated as karyon A (Fig. 5, Supplementary Data 3). Contigs with substantial overlap with the subtracted reads (2–3X of the original coverage) were designated as karyon B. Contigs with ambiguous read mapping data, including those with low coverage in the original unsubtracted reads or covered by largely non-uniquely mapping reads were left as unassigned. Read mapping to all contigs was confirmed via visual inspection of coverage graphs and read alignments in the CLC Genomics Workbench browser. Potential chimeric contigs were identified as containing distinct regions with either high or no coverage with the Ug99-subtracted reads (Supplementary Fig. 4). For subsequent comparison and analyses, these contigs were manually split into their component fragments which were designated as haplotype A or B accordingly (Supplementary Data 4). The same process was followed in reverse for the assignment of the A and C haplotype contigs in Ug99.

Sequence comparisons of genome assemblies. Haploype sequences of the \textit{AvrSr50} or \textit{AvrSr35} chromosome as well as the full haplode genomes were aligned using MUMmer4.x\textsuperscript{50} with nucler -maxmatch and other parameters set as default and the alignment metrics summarised in MUMmer dmdiff. Structural variation between the two assemblies was determined using Assemblytics\textsuperscript{50} from the MUMmer delta file with a minimum variant size of 50 bp, a maximum variant size of 100 kbp, and a unique sequence length for anchor filtering of 10 kbp. Haploype dot plot alignments were generated using D-genies\textsuperscript{46}.

Read coverage analysis and SNP calling on haplotypes. Illumina reads from Ug99 and Pgt21-0 were each mapped against the Ug99 and Pgt21-0 assemblies in CLC Genomics Workbench (similarity fraction 0.98, length fraction 0.95). For each assembly, SNP coverage was calculated (proportion of reads calculated per base in “window”) using samtools bedcov. Read coverage normalised to the mean coverage of each haplotype was graphed as a violin plot using seaborn 0.9.0 package (https://seaborn.pydata.org/) using a custom python script. To detect SNPs between two haplotypes, Illumina read pairs of Pgt21-0 that mapped uniquely to either the Pgt21-0 A or B haplotype were identified. The filtered read pairs that uniquely mapped to either the A or C haplotype contigs of Ug99 were extracted. These read sets were then separately mapped to the two assemblies in CLC Genomics Workbench (similarity fraction 0.99, length fraction 0.98). Variant calling was performed using FreeBayes v1.1.0\textsuperscript{51} with default parameters in the A or B haplotype set. Variant calling was performed using FreeBayes v.1.1.0 with default parameters in the A or B haplotype set.
and carbohydrate hydrolysing enzymatic domains (CAZYmes) were annotated using InterProScan5 (v5.23-62.0) to identify InterPro terms, GO ontology and single genotype free of contamination, read allele frequencies at heterozygous coverage were verified.

Received: 27 June 2019; Accepted: 3 October 2019; Published online: 07 November 2019

Data availability
All sequence data, assemblies and gene annotation files generated in this study are available in NCBI under BioProject PRJNA516922. Assemblies and annotations are also available at the DOE-JGI MycosCos Portal (https://mycoscos.jgi.doe.gov/mycoscos/home). Metadata for RNAseq libraries of Pg21-0 and Illumina DNAseq libraries from all isolates are available in Supplementary Data 7 and 8, respectively. Data underlying Figs 3c and Supplementary Fig. 3a is provided in Supplementary Data 3. Data underlying Fig. 6 is available in Supplementary Data 6. All other relevant data is available upon request from the corresponding authors.

Code availability
Unless specified otherwise, all scripts and files are available at https://github.com/figueruelalab/Pgt_genomes.

References
1. Stajich, J. E. et al. Primer-the fungi. Curr. Biol. 19, R840 (2009).
2. Fleißner, A., Serrano, A. Growth, Differentiation and Sexuality (eds Wendland, J.) (Springer, 2016).
3. Spatafora, J. W. et al. The fungal tree of life: from molecular systematics to genome-scale phylogenies. Microbiol. Spectr. 5, 1–32 (2017).
4. Aime, M. C., McTaggart, A. R., Mundo, S. J. & Dusselis, S. Phylogenetics and phylogenomics of rust fungi. Adv. Genet. 100, 267–307 (2017).
5. Watson, I. Further studies on the production of new races from mixtures of races of Puccinia graminis var. tritici on wheat seedlings. Phytopathology 47, 510–512 (1957).
6. Ellingboe, A. Somatic recombination in Puccinia graminis var. tritici. Phytopathology 51, 13–15 (1961).
7. Flor, H. Genetics of somatic variation for pathogenicity in Melampsora lini. Phytopathology 54, 823–826 (1964).
8. Bartos, P., Fleischmann, G., Samborski, D. & Shipton, W. Studies on asexual variation in the virulence of oat crown rust, Puccinia coronata f. sp.avenae, and wheat leaf rust, Puccinia recondita. Can. J. Bot. 47, 1383–1387 (1969).
9. Burdon, J. J., Marshall, D. R. & Luig, N. H. Isozyme analysis indicates that a virulent rust pathogen is a somatic hybrid. Nature 293, 565 (1981).
10. Park, R., Burdon, J. & Jahoorn, A. Evidence for somatic hybridisation in the leaf rust pathogen of wheat (Puccinia recondita f. sp. tritici). Mycological Res. 103, 715–723 (1999).
11. Park, R. F. & Wellings, C. R. Somatic hybridization in the Uredinales. Annu. Rev. Phytopathol. 50, 219–239 (2012).
12. Miller, M. E. et al. De novo assembly and phasing of dikaryotic genomes from two isolates of Puccinia coronata f. sp. avennae, the causal agent of oat crown rust. BioRxiv 9, e01650–e01617 (2018).
13. Schuessinger, B. et al. A near-complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus Puccinia striiformis f. sp. tritici reveals high interhaplotype diversity. mBio 9, e02275–e02217 (2018).
14. Singh, R. P. et al. Emergence and spread of new races of wheat stem rust fungus revealed threat to food security and prospects of genetic control. Phytopathology 105, 872–884 (2015).
15. Pretorius, Z., Singh, R., Wagoire, W. & Payne, T. Detection of virulence to wheat stem rust resistance gene Sr33 in Puccinia graminis f. sp. tritici in Uganda. Plant Dis. 84, 203–208 (2000).
16. Visser, B. et al. Characterization of two new Puccinia graminis f. sp. tritici races within the Ug99 lineage in South Africa. Euphytica 179, E119–E127 (2012).
17. Olivera, P. et al. Phenotypic and genotypic characterization of race TKTTF of Puccinia graminis f. sp. tritici that caused a wheat stem rust epidemic in southern Ethiopia in 2013–14. Phytopathology 105, 917–928 (2015).
18. Visser, B. et al. Microsatellite analysis and u-radiosporic dispersal simulations support the movement of Puccinia graminis f. sp. tritici from Southern Africa to Australia. Phytopathology 109, 133–144 (2019).
19. Pretorius, Z. A., Pakendorf, K. W., Maraiss, G. F., Prins, R. & Komen, J. S. Challenges for sustainable cereal rust control in South Africa. Aust. J. Agric. Res. 58, 593–601 (2007).
20. Park, R. F. Stem rust of wheat in Australia. Aust. J. Agric. Res. 58, 558–566 (2007).
21. Upadhyaya, N. M. et al. Comparative genomics of Australian isolates of the wheat stem rust pathogen Puccinia graminis f. sp. tritici reveals extensive polymorphism in candidate effector genes. Front. Plant Sci. 5, 759 (2014).
22. Duplessis, S. et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc. Natl Acad. Sci. USA 108, 9166–9171 (2011).
23. Chen, J. et al. Loss of AvrSr50 by somatic exchange in stem rust leads to virulence for Sr50 resistance in wheat. Science 358, 1607–1610 (2017).
24. Salcedo, A. et al. Variation in the AvrSr35 gene determines Sr35 resistance against wheat stem rust race Ug99. Science 358, 1604–1606 (2017).
25. Anderson, C. et al. Genome analysis and avirulence gene cloning using a high-density RADseq linkage map of the flax rust fungus, Melampsora lini. BMC Genomics 17, 667 (2016).
26. Belton, J. M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58, 268–276 (2012).
27. Boehm, E. W. A. et al. An ultrastructural pachytene karyotype for Puccinia graminis f. sp. tritici. Can. J. Bot. 70, 401–413 (1992).
28. Cembrowski, C. A. et al. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. G3: Genes, Genomes. Genetics 3, 367–371 (2016).
29. Lewis, C. M. et al. Potential for re-emergence of wheat stem rust in the United Kingdom. Commun. Biol. 1, 13 (2018).
30. Figueroa, M. et al. Changing the game: using integrative genomics to probe virulence mechanisms of the stem rust pathogen Puccinia graminis f. sp. tritici. Front. Plant Sci. 7, 205 (2016).
31. Lei, Y. et al. Virulence and molecular characterization of experimental isolates of the stripe rust pathogen (Puccinia striiformis) indicate somatic recombination. Phytopathology 107, 329–344 (2017).
32. Barr, R., Caldwell, R. & Amacher, R. Examination of vegetative recombination of aedispore color and virulence in mixtures of certain races of Puccinia recondita. Phytopathology 54, 104–109 (1964).
33. Wang, X. & McCallum, B. Fusion body formation, germ tube anastomosis, and nuclear migration during the germination of urediospores of the wheat leaf rust fungus, Puccinia triticina. Phytopathology 99, 1355–1364 (2009).
34. Krukal, E. K. & Anderson, J. B. Dikaryons of the basidiomycete fungus Schizophyllum commune: evolution in long-term culture. Genetics 167, 1663–1675 (2004).
35. Corradi, N. & Brachmann, A. Fungal mating in the most widespread plant pathogenic fungus Puccinia graminis f. sp. tritici. Proc. Natl. Acad. Sci. 22, 175–183 (2017).
36. Pretorius, Z., Szabo, L., Boshoff, W., Herselman, L. & Visser, B. First report of a new TTKSF race of wheat stem rust (Puccinia graminis f. sp. tritici) in South Africa and Zimbabwe. Plant Dis. 96, 590–590 (2012).
37. Schwingsberger, B., Rathjen, J. P. Wheat Rust Diseases (eds Perrenynan, S.) (Springer, 2017).
38. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736 (2017).
39. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9, e112963 (2014).
40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
41. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
42. Quinlan, A. R. & Hall, I. M. BEDTools: a fast flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
43. Gerstewich, A., Saveliev, V., Vyáhi, N. & Telider, G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075 (2013).
44. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467 (2005).
45. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212 (2015).
46. Cabanneet, F. & Klop, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6, e6958 (2018).
47. Lawrence, M. et al. Software for calculating and annotating genomic ranges. PLoS Comput. Biol. 9, e1003118 (2013).
48. Yin, T., Cook, D. & Lawrence, M. ggbiom: an R package for extending the grammar of graphics for genomic data. Genome Biol. 13, R77 (2012).
49. Kurtz, S. et al. Versatile and open software for comparing large genomes. Genome Biol. 5, R12 (2004).
50. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. arXiv 1207, 3907 (2012).
51. Wingett, S., et al. HICUP: pipeline for mapping and processing Hi-C data. F1000Research 4, 1310 (2015).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357 (2012).
53. Ghurye, J., Pop, M., Koren, S., Bickhart, D. & Chin, C.-S. Scaffolding of long reads assemblies using long range contact information. BMC Genomics 18, 527 (2017).
54. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6, 11 (2015).
55. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357 (2015).
56. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290 (2015).
Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-12927-7.

Correspondence and requests for materials should be addressed to P.N.D. or M.F.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.