Evolutionary genomics reveals variation in structure and genetic content implicated in virulence and lifestyle in the genus *Gaeumannomyces*

Rowena Hill[1]*, Michelle Grey[1], Mariano Olivera Fedi[1], Daniel Smith[2], Sabrina J. Ward[1], Gail Canning[2], Naomi Irish[1], Jade Smith[2], Vanessa E. McMillan[3], Jess Hammond[2], Sarah-Jane Osborne[2,4], Tania Chancellor[2,5], David Swarbreck[1], Neil Hall[1,6], Javier Palma-Guerrero[2,7], Kim E. Hammond-Kosack[2*], Mark McMullan[1]*

[1] Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK

[2] Rothamsted Research, Harpenden, AL5 2JQ, UK

[3] NIAB, 93 Lawrence Weaver Road, Cambridge, CB3 0LE, UK

[4] AHDB, Middlemarch Business Park, Siskin Parkway East, Coventry, CV3 4PE, UK

[5] Crop Science Centre, Department of Plant Sciences, University of Cambridge, 93 Lawrence Weaver Road, Cambridge, CB3 0LE, UK

[6] School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

[7] Research Institute of Organic Agriculture FiBL, Frick, 5070, Switzerland.

*Authors for correspondence:

Rowena.Hill@earlham.ac.uk; Kim.Hammond-Kosack@rothamsted.ac.uk;

Mark.McMullan@earlham.ac.uk

ABSTRACT

Gaeumannomyces tritici is responsible for take-all disease, one of the most important wheat root threats worldwide. High-quality annotated genome resources are sorely lacking for this pathogen, as well as for the closely related antagonist and potential wheat take-all biocontrol agent, *G. hyphopodioides*. As such, we know very little about the genetic basis of the interactions in this host-pathogen-antagonist system. Using PacBio HiFi sequencing technology we have generated nine near-complete assemblies, including two different virulence lineages for *G. tritici* and the first assemblies for *G. hyphopodioides* and *G. avenae* (oat take-all). Genomic signatures support the presence of two distinct virulence lineages in *G. tritici* (types A and B), with A strains potentially employing a mechanism to prevent gene copy-number expansions. The CAZyme repertoire was highly conserved across *Gaeumannomyces*, while candidate secreted effector proteins and biosynthetic gene clusters showed more variability and may distinguish pathogenic and non-pathogenic lineages. A transition from self-sterility (heterothallism) to self-fertility (homothallism) may also be a key innovation implicated in lifestyle. We did not find evidence for transposable element and effector gene compartmentalisation in the genus, however the presence of *Starship* giant transposable elements likely contributes to genomic plasticity in the genus. Our results depict *Gaeumannomyces* as an ideal system to explore interactions within the rhizosphere, the nuances of intraspecific virulence, interspecific antagonism, and fungal lifestyle evolution. The foundational genomic resources provided here will enable the development of diagnostics and surveillance of understudied but agriculturally important fungal pathogens.

INTRODUCTION

Gaeumannomyces is a broadly distributed genus of *Poaceae* grass-associated root-fungi (Hernández-Restrepo et al. 2016), best known for the species *Gaeumannomyces tritici* (*Gt*) which causes take-all disease, the most serious root disease of wheat (Palma-Guerrero et al. 2021). *Gaeumannomyces* is a comparatively understudied genus despite belonging to the *Magnaporthales*, an economically important order of pathogens including the rice and wheat

blast fungus *Pyricularia oryzae* (syn. *Magnaporthe oryzae* (Zhang et al. 2016)). This is perhaps due to a historical research bias towards above-ground pathogens, in part simply due to the fact that characteristic symptoms of root pathogen diseases are hidden from view (Raaijmakers et al. 2009; Balmer and Mauch-Mani 2013). Recently the rhizosphere has received more research attention as its key role in plant health and productivity has become apparent (van der Heijden et al. 2008). There have also been considerable difficulties in producing a reliable transformation system for *Gt*, preventing gene disruption experiments to elucidate function (Freeman and Ward 2004).

Although genetic studies of *Gt* have been limited, single-locus phylogenetic analyses of *Gt* have consistently recovered two distinct lineages within the species (Daval et al. 2010), which we will refer to using the 'A/B' characterisation established by Freeman et al. (2005) based on *ITS2* polymorphism. Although very little is known about the dynamics of these two lineages, each is found across the world and both lineages persistently co-occur in the same field, prompting the suggestion that the two lineages may actually be cryptic species (Daval et al. 2010; Palma-Guerrero et al. 2021). Although variation within lineages is high, there is also some evidence that type A strains are more virulent (Bateman et al. 1997; Lebreton et al. 2004, 2007), which is a major impetus for improving our understanding of these two lineages. The sister species to *Gt*, *G. avenae* (*Ga*), can also infect wheat, but is not the predominant agent of wheat take-all, and is distinguished by the fact that production of avenacinase enables *Ga* to additionally infect oat roots (Osbourn et al. 1991; Bowyer et al. 1995).

Magnaporthales are also home to several commensal and/or mutualistic fungi (Xu et al. 2014), including those with the potential to inhibit take-all (Chancellor 2022). For instance, *G. hyphopodioides* (*Gh*) — a species closely related to *Gt* that also grows on wheat roots— is not only non-pathogenic, but actually capable of suppressing take-all to varying degrees (Osborne et al. 2018). It is now apparent that prior *Gh* colonisation primes the host plant's immune response (Chancellor et al. 2023), a mechanism that has been reported in various other plant–microbe interactions associated with disease prevention (Van Wees et al. 2008;

Zamioudis and Pieterse 2012). This has prompted interest in *Gh* as a potential biocontrol agent, for instance by adding *Gh* inoculant to wheat seedstock via seed coating (Accinelli et al. 2016) and/or selecting for wheat cultivars that support enhanced levels of *Gh* root system colonisation (Osborne et al. 2018). Novel disease prevention approaches for take-all are especially desirable as up to 30% of *Gt* strains are found to be naturally resistant to the seed-dressing fungicide routinely used to treat take-all, silthiofam (Freeman et al. 2005).

Understanding the genetic machinery underpinning virulence and lifestyle in *Gaeumannomyces* has previously been hampered by a lack of genomic data. Prior to the present study, a single annotated *Gt* assembly (strain R3-111a-1), sequenced using the 454 platform, was available on NCBI (accession GCF_000145635.1) (Okagaki et al. 2015) – one other more recent PacBio assembly has been released for the same strain, but remains unannotated (GCA_016080095.1). This scarcity of genomic resources has not only limited our understanding of the genetics of the system, but also accounts for a lack of molecular diagnostics for take-all. Given the increase in research activities since 2005 following the production of genomic resources for *P. oryzae* (Sperr 2023; Dean et al. 2005), we are optimistic that providing similar high-quality assemblies for *Gaeumannomyces* species will bolster research efforts in the global take-all community.

Here, we have addressed the gap in genomic resources for *Gaeumannomyces* by generating near-complete assemblies for nine strains, including both type A and B *Gt* lineages and the first assemblies for *Gh* and *Ga*. Using an evolutionary genomics approach, we identified variation in structure as well as gene features known to be involved in plant-fungal interactions — candidate secreted effector proteins (CSEPs), carbohydrate-active enzymes (CAZymes) and biosynthetic gene clusters (BGCs) — to address the questions: (1) Are there genomic signatures distinguishing *Gt* A/B virulence lineages? (2) How do gene repertoires differ between pathogenic *Gt* and non-pathogenic *Gh*? and (3) Is there evidence of genome compartmentalisation in *Gaeumannomyces*? In the process of doing so, we also identified

giant cargo-carrying transposable elements belonging to the recently established *Starship* superfamily (Gluck-Thaler et al. 2022).

RESULTS

Evidence of greater take-all severity caused by *G. tritici* type A strains

As the five *Gt* strains sequenced in this study included representatives of both the type A and B lineages, we performed a season long inoculation experiment to determine the relative capacity for each strain to cause take-all disease symptoms. From general visual inspection, inoculation of *Gt*A strains into the highly susceptible winter wheat cultivar Hereward resulted in notably depleted roots compared to a control and, to a lesser extent, *Gt*B strains (Fig. 1a). Inoculation with *Gt*A strains also resulted in a visible reduction of overall plant size compared to the control, while *Gt*B-inoculated plants were less easily distinguished from the control (Fig. 1b). Although above- and below-ground characteristics of wheat varied depending on *Gt* strain, our statistical analysis showed that the *Gt*A strains had a greater capacity to reduce plant height and reduce root length, and both *Gt*A strains consistently produced the greatest root disease symptoms, i.e. highest Take-all Index (TAI) scores (Bateman et al. 2004) (Fig. 1c). Furthermore, five out of six wheat plants that died during the experiment were inoculated with *Gt*A strains. Several characteristics were inconsistently affected by *Gt* inoculation, including mean floral spike (ear) length; dried root biomass; number of roots; and number of roots per tiller.

Nine near-complete *Gaeumannomyces* assemblies, including first genome assemblies for *G. avenae* and *G. hyphopodioides*

We used PacBio HiFi sequencing technology to produce highly contiguous genome assemblies for five *Gt*, two *Gh* and two *Ga* strains. All nine assembled genomes had N50 values of more than 4 Mb (Supplemental Table S1), a 100-fold increase on the N50 of the existing annotated *Gt* RefSeq assembly (NCBI accession GCF_000145635.1). In addition, transcriptomes were sequenced for all nine strains to inform gene prediction, and between

22–29% of annotated gene models had two or more isoforms across all strains (Supplemental Fig. S1). Contigs corresponding to mitochondrial genomes were identified from all assemblies (Supplemental Table S1), however circularisation was only successfully detected for two strains (Gt-23d and Ga-CB1). For most strains the overall mitogenome size, GC content and number of genes fell within the expected range for ascomycetes (Fonseca et al. 2021), however the mitogenome assembly for Gt-LH10 is likely incomplete, as it was a third of the size of the other *Gt*B strains, and only had 23 genes annotated compared to the 38–40 genes found for all other strains (Table S1).

Combined GENESPACE (Lovell et al. 2022) and telomere prediction results suggested six chromosomes for *Gaeumannomyces* (Fig. 2), one less than *P. oryzae* (Dean et al. 2005). Telomere-to-telomere sequences were assembled for at least five out of six pseudochromosomes for most strains. By plotting GC content alongside transposable element (TE) and gene density, we also identified AT- and TE-rich but gene-poor regions, which are putative candidates for centromeres (Supplemental Fig. S2). Some of these regions additionally correspond well with points of fragmentation in other strains, presumably due to the difficulties associated with assembly of such highly repetitive regions. Other than these occasional splits into two fragments, in most cases pseudochromosomes were entire, the exception being Gh-1B17 pseudochromosome 2 which was fragmented across five contigs.

Both *Gt*A and, to a slightly lesser degree, *Gt*B were broadly syntenic across whole pseudochromosomes, with the exception of a major chromosomal translocation between pseudochromosomes 2 and 3 in Gt-LH10 (Fig. 2). Visualisation of the spanning reads and coverage across the regions of the apparent translocation suggests the depicted arrangement is correct and not an artefact due to misassembly (Supplemental Fig. S3a), moreover there was no evidence of a block of repeats consistent with a telomere anywhere but at the ends of the pseudochromosomes (Supplemental Fig. S3c). *Ga* was also largely syntenic with *Gt*, although there were a number of inversions in Ga-CB1 pseudochromosome 3 (Fig. 2). The more distantly related *Gh* showed chromosomal translocations involving

pseudochromosomes 1, 2 and 5, which were again supported by spanning reads and the absence of intrachromosomal telomeric repeats (Supplemental Fig. S3b, c).

No evidence for significant colocalisation of transposable elements and effectors

Compartmentalisation of effectors within genomic regions enriched in transposable elements (TEs) has previously been reported for various fungal phytopathogens (Dong et al. 2015). In all the *Gaeumannomyces* strains sequenced here, however, we did not observe that predicted CSEPs were more likely to occur in regions of high TE density (Fig. 3a). We found a weak significant positive correlation between CSEP density and TE density for a minority of strains, however the scatterplot and local polynomial regression lines were unconvincing (Fig. 3b). CSEP density was more frequently found to significantly correlate with gene density, although this was still only a weak association (Fig. 3b). For all but one strain, there was no significant difference in mean distance to closest TE for CSEPs versus other genes (Fig. 3c). For strain Gt-19d1, the mean distance from a CSEP to the closest TE was marginally lower (10,036 bp) than for other genes (12,565 bp), which permutation analysis confirmed was closer than expected based on the overall gene universe ($p=0.03$), although this only remained significant for pseudochromosomes 2 and 6 when testing pseudochromosomes separately (Supplemental Fig. S4a). Individual pseudochromosomes for other strains also had lower than expected CSEP–TE distances, but with low z-scores (a proxy for strength) across the board. Comparing across strains, mean gene–TE distance was significantly different both within and between lineages, and lowest in *Gt*B (Fig. 3c). Within *Gt*B, Gt-LH10 had significantly lower mean gene–TE distance, and the same strain has also undergone an apparent expansion in total number of TEs compared to all other strains (Supplemental Fig. S5).

Although CSEPs were not broadly colocalised with TEs, we did observe that they appeared to be non-randomly distributed in some pseudochromosomes (Fig. 3a). Permutation analyses confirmed that overall CSEPs were significantly closer to telomeric regions in all strains ($p=<0.008$), although by testing pseudochromosomes separately we found that this pattern varied across the genome (Supplemental Fig. S4b). CSEPs on pseudochromosomes 1, 2 and

5 were consistently closer to telomeric regions, whereas for pseudochromosomes 3 and 4 CSEPs were no closer than expected based on the gene universe. CSEPs were also closer to telomeres in pseudochromosome 6, but only in *Gt* strains.

Core gene content in *Gaeumannomyces*

The total number of genes was relatively similar for all strains, although, as indicated in Fig. 2, *Gt*B and *Gh* strains had 3–6% more genes than *Gt*A or *Ga* (Fig. 4a). *Gt*A and *Gt*B had a very similar number of CSEPs, CAZymes and BGCs, however, and more CSEPs and BGCs than either *Ga* or *Gh*. Almost all total genes, CSEPs and CAZymes were core or soft-core (i.e. present in all but one strain) in *Gt*, while there was a greater proportion of BGCs that were accessory or strain-specific. From a pangenome perspective, the core gene content for *Gt* from sampling these five strains amounted to ~10,000 genes (Fig. 4b), which equates to ~88% of genes per strain being core, consistent with reports in other fungi (McCarthy and Fitzpatrick 2019). The majority of BUSCO genes found to be missing in the assemblies were missing from all strains (Supplemental Fig. S6), suggesting that they are not present in the genus, rather than being missed as a result of sequencing or assembly errors. Three of these 18 missing core genes belonged to the *Snf7* family, which is involved in unconventional secretion of virulence factors in fungi (da C. Godinho et al. 2014), and is essential for pathogenicity in *P. oryzae* (Cheng et al. 2018). The next greatest set of missing BUSCOs (8) also seemed to be lineage specific – i.e. missing in *Gh* but present in *Gt*/*Ga* (Supplemental Fig. S6).

The avenacinase gene required for virulence on oat roots (Osbourn et al. 1991; Bowyer et al. 1995) was identified in all strains in a conserved position on pseudochromosome 4 (Supplemental Fig. S7a). Two mating-type (MAT) loci were identified in *Gt* and *Ga*, with homologues of *Pyricularia grisea MAT1-1* and *MAT1-2* idiomorphs located in conserved but unlinked positions on pseudochromosomes 2 and 3, while only one MAT locus and idiomorph, *MAT1-1*, was identified in *Gh* on pseudochromosome 3 (Supplemental Fig. S8).

Differences in effectors and secondary metabolite production potential between pathogenic and non-pathogenic *Gaeumannomyces* species

The predicted BGC content ranged from 9 to 17 per strain, which is low compared to many other ascomycete fungi (Gluck-Thaler et al. 2020; Franco et al. 2021; Llewellyn et al. 2023). Using a phylogenetically-informed permutational multivariate analysis of variance (PERMANOVA) method (Mesny and Vannier 2020) to identify associations between gene variance and lifestyle, we also found BGCs to have the lowest level of variance described purely by ancestry, 61% compared to 75%–85% for all genes, CSEPs and CAZymes (Fig. 4a). This was coupled with a relatively high proportion of BGC variance described by lifestyle (17%), which was also the case for all genes (17%) and CSEPs (14%), while lifestyle had much less descriptive power for CAZymes (7%). CAZymes that are known to act on plant cell wall substrates were highly conserved across the genus, and there were highly similar numbers of each CAZyme family across all strains (Supplemental Fig. S9a). The only discernible pattern was marginally more copies of GH55 and GH2 (hemicellulose and pectin) in *Gh* versus the other lineages.

In total, 9% of CSEP genes could be attributed to a known gene in the Pathogen-Host Interactions database (PHI-base) (Urban et al. 2020), most of which only had one copy in all strains (Supplemental Fig. S9b). Sixteen of the 19 'named' CSEPs have been associated with virulence via reverse genetics experiments, including five from *P. oryzae* infecting *Oryza sativa* (rice) — *MHP1* (ID PHI:458); *MoAAT* (PHI:2144); *MoCDIP4* (PHI:3216); *MoHPX1* (PHI:5188); and *MoMAS3* (PHI:123065). The latter two were assigned to genes that were only present in *Gh*, although a separate gene present in *Gt*B was also characterised as *MoHPX1*. Six CSEPs in total were present in all lineages except *Gh* or vice versa. *PBC1*, also a CAZyme, the disruption of which causes complete loss of pathogenicity of *Pyrenopeziza brassicae* in *Brassica napus*, was present in *Gt* and *Ga* but not *Gh*. While *PBC1* was absent in *Gh*, all *Gaeumannomyces* strains did have some genes belonging to the same CAZyme family (CE5; Supplemental Fig. S9a).

The BGC families were predominantly classified as terpenes, type 1 polyketide synthases (PKSI) and hybrid polyketide synthase-nonribosomal peptides (PKS-NRP) (Supplemental Fig. S9c). Presence-absence of each BGC was highly variable across strains, most notably amongst PKSI families which also had high copy-number for certain families. There were five BGCs that were present or absent in *Gh* versus other lineages, four of which belonged to the PKS-NRP hybrids (Supplemental Fig. S9c).

Gene copy-number reduction in *G. tritici* type A

*Gt*B, *Ga* and *Gh* all had high copy-number (HCN) gene outliers (>10 copies) that were absent in *Gt*A (Fig. 4a). These 22 HCN genes were duplicated both within and across pseudochromosomes (Supplemental Fig. S10a). GO term enrichment analyses found various terms to be significantly overrepresented amongst the HCN genes, namely: vacuolar proton-transporting V-type ATPase complex assembly (Gh-1B17, Fisher's exact test, p=0.01); ubiquinone biosynthetic process (Gh-2C17, p=0.01); golgi organisation (Ga-CB1, p=0.03); mRNA cis splicing, via spliceosome (Gt-4e, p=0.03); mitochondrial respiratory chain complex I assembly (Gt-4e, p=0.05); proton-transporting ATP synthase complex assembly (Gt-LH10, p=0.03); and protein localisation to plasma membrane (Gt-LH10, p=0.03). Visualising the location of the HCN genes across the genomes (Supplemental Fig. S11) showed them to vary in terms of distribution — from relatively localised to broadly expanded — and in terms of multi-lineage versus lineage specific expansions. HCN genes were also significantly closer to TEs compared to other genes (Supplemental Fig. S10b).

Interestingly, of the 22 HCN genes, six that were shared among all species were also present in at least one *Gt*A strain but at low copy-number, while seven genes were completely absent in *Gt*A (Fig. 4c). In total, nine genes that were HCN in at least one other lineage had low-copy orthologues in *Gt*A. Moreover, these were mostly present in just one strain within the type A lineage (Gt-8d), clustered in a ~1 Mbp region on pseudochromosome 3 (Supplemental Fig. S10c). This region was flanked by repetitive regions that have been subjected to repeat induced point mutation (RIP), as measured by the composite RIP index (CRI) (Lewis et al.

2009), although the region had average CRI of -0.3 compared to an average CRI of -0.5 for the whole pseudochromosome. Average genome-wide RIP levels were highest in *Gt*A and *Gh* (13.8% and 13.6% of the genome RIP'd, respectively), compared to *Gt*B (10.8%) and *Ga* (12.4%).

***Gaeumannomyces* genomes contain *Starship* giant transposable elements**

All nine *Gaeumannomyces* strains were found to contain at least one giant TE belonging to the *Starship* superfamily of giant cargo-carrying TEs (Gluck-Thaler et al. 2022), identified using the tool starfish (Gluck-Thaler and Vogan 2023). Currently the most reliable identifying feature of *Starships* is a single 'captain' gene – a tyrosine recombinase gene containing a DUF3435 domain which is found in the first position of each *Starship* and directs the mobilisation of the element (Urquhart et al. 2023b). We found that tyrosine recombinase annotation with starfish largely overlapped with results from a separate blast search to identify DUF3435 homologues at the head of insertions. Overall, only a relatively small number of genes were in agreement as full *Starship* captains after downstream automated (starfish) or manual element inference (Fig. 5a). A gene tree of all tyrosine recombinase and putative captain genes showed the presence of two distinct lineages but no consistent clustering of either gene types or method of identifying them. Note the highly divergent nature of the genes and therefore the difficulty of alignment and subsequent poor branch support throughout the tree (Fig. 5a).

Starship size varied considerably, ranging from 34–688 kbp. *Gt*B strains harboured notably more elements, followed by *Ga* strains which included a nested element (Fig. 5b). *Gt*A and *Gh* strains each contained a single smaller (<100 kbp) element, which in both cases we predict to have been vertically transmitted based on similar gene content and conserved location within the genome (Fig. 5b, Supplemental Fig. S12). *Gt*A elements were exceptional in that each was gene-poor and positive for element-wide RIP (average CRI=0.2-0.3).

DISCUSSION

In this study we have established foundational genome resources for the genus *Gaeumannomyces*. A particular strength of the *Gt* assemblies reported here is the structural annotation methodology, which capitalised on the fact that multiple strains were sequenced, assembled and annotated in the same way, each with its own transcriptome but also employing a novel 'multiple lift-off' approach that provided additional evidence for robust gene models. Another benefit of the annotation approach is that the REAT-Mikado-minos pipeline (EI-CoreBioinformatics 2023b) provides models for gene isoforms alongside the primary transcripts. Alternative splicing has been implicated in regulation of virulence in phytopathogens (Fang et al. 2020), for instance by mediating transcriptome remodelling during pathogenesis in *P. oryzae* (Jeon et al. 2022). Alternative splicing has been reported to be more frequent in pathogens than non-pathogens (Grutzmann et al. 2014), however we found a similar overall percentage of genes with multiple isoforms in *Gh* compared to *Gt* and *Ga* (Supplemental Fig. S1). There was perhaps a skew towards a greater proportion of genes with exactly two or three isoforms in *Gt*, particularly *Gt*A, raising the question as to whether this somehow relates to their apparent higher virulence in wheat. These rich annotations resources will allow further exploration of the isoform content of *Gaeumannomyces* and its potential role in virulence.

A major finding from our synteny analyses was the presence of a large chromosomal translocation in Gt-LH10 (Fig. 2). Similar largescale translocations have been identified in *Pyricularia* (Bao et al. 2017; Gómez Luciano et al. 2019). It is entirely plausible that we have identified a genuine translocation, however confidence would be increased by obtaining Hi-C evidence and/or by corroborating with population-level data. Such rearrangements are thought to be a route to accessory chromosome formation (Croll et al. 2013; Hartmann 2022), and this has specifically been reported in *P. oryzae* (Langner et al. 2021). Although we did not find any evidence for accessory chromosomes in our *Gaeumannomyces* strains, it is interesting that the Gt-LH10 translocation resulted in one of the chromosomes being much smaller, size being a hallmark of accessory or 'mini-chromosomes'. It is also notable that this large translocation

292 occurred in the same strain we found to have an expansion of TEs (Supplemental Fig. S5), as
293 TEs have been found to mediate interchromosomal rearrangements (Bao et al. 2017; Fourie
294 et al. 2020; Langner et al. 2021). Hi-C data would also allow us to robustly locate centromeres
295 (Varoquaux et al. 2015), which are also implicated in chromosomal rearrangements (Yadav et
296 al. 2020; Guin et al. 2020). Here we used a minimal approach to estimate potential centromeric
297 regions, based simply on the fact that AT-rich regions are a common defining feature of
298 centromeres in *P. oryzae* (Yadav et al. 2019), which we also cross-checked with gene sparsity
299 (Supplemental Fig. S2) — however, we were only able to distinguish potential centromeres
300 for a subset of the pseudochromosomes.

301 In addition to the chromosomal translocation, Gt-LH10 also stood out from other strains in
302 terms of TE content, with an expansion in total number of TEs (Supplemental Fig. S5) and
303 smaller gene–TE distances (Fig. 3). Aside from the atypical features of the Gt-LH10 genome,
304 there was additional intraspecific variability within the *Gt* A/B lineages in terms of both genome
305 structure and gene content. For instance, there were strain-specific inversions (Fig. 2) and
306 many of the HCN genes were present in low copy-number in one *Gt*A strain, but completely
307 absent in the other (Fig. 4c). These findings emphasise the need for pangenome references,
308 as an individual strain alone cannot sufficiently represent the variability across the whole
309 species (Golicz et al. 2020; Badet and Croll 2020). Pangenomics is still relatively young and
310 the practicalities of how to define, store and use pangenomes and the tools necessary to do
311 so are continuously evolving (Eizenga et al. 2020). There is also the outstanding question of
312 how best to coordinate pangenome initiatives to ensure high-quality results without duplicating
313 efforts — at least three different pangenomes have been reported for the wheat pathogen
314 *Zymoseptoria tritici* to date (Plissonneau et al. 2018; Badet et al. 2020; Chen et al. 2023). The
315 five *Gt* strains reported here can act as a UK pangenome, but future research must work
316 towards building a global pangenome so that we can provide a reference for *Gt* which captures
317 a fuller representation of the species.

Another structural feature that these high-quality assemblies allowed us to explore in *Gaeumannomyces* was genome compartmentalisation. A number of fungal phytopathogens exhibit TE- and effector-rich compartments that enable rapid evolution in the plant–fungal arms race, dubbed the 'two-speed' genome model (Dong et al. 2015), which has since been extended to 'multi-speed' models (Frantzeskakis et al. 2019). Accordingly, we hypothesised that we would find CSEPs and TEs to colocalise across our assemblies, however we did not find consistent evidence for such compartments in *Gaeumannomyces* (Fig. 3). Our results are not altogether surprising as a previous study of selection signatures in *Gt* and two other *Magnaporthales* taxa also found no evidence for multi-speed genomes (Okagaki et al. 2016). We therefore consider *Gaeumannomyces* taxa to have 'one-compartment' genomes in relation to TE/effector content – a term that was introduced by Frantzeskakis et al. (2019) for genomes that do not conform to the two- or multi-speed models, and with 'compartment' suggested as an alternative to 'speed' as the defining features of these compartments does not necessarily equate to them being fast-evolving (Torres et al. 2020). With the rising number of high-quality genome resources, more examples are emerging that contradict the suggestion that phytopathogenicity is routinely accompanied by TE/effector compartmentalisation (Frantzeskakis et al. 2019). In fact, TE/effector compartmentalisation has been found in the non-pathogenic arbuscular mycorrhizal fungus *Rhizophagus irregularis* (Yildirir et al. 2022), and TE/virulence factor compartmentalisation has also been found in chytrid animal pathogens (Wacker et al. 2023), demonstrating that it is not necessarily central to phytopathogenicity, but may instead be a mechanism driving genome plasticity in fungi of various lifestyles (Torres et al. 2020). While we did not find compelling evidence for TE/effector compartmentalisation in *Gaeumannomyces*, we did observe non-random patterns in the distribution of CSEPs (Fig. 3a), which permutation analyses found to be closer to telomeric regions in a pseudochromosome-dependent manner (Supplemental Fig. S4b). This could suggest that alternative mechanisms of effector compartmentalisation may be at play.

Our results indicate conserved genetic machinery for plant cell wall deconstruction/ modification across both pathogenic and non-pathogenic *Gaeumannomyces* (Fig. 4a, S11a), suggesting that the mechanism(s) by which species first colonise roots may be similar, if not the final outcome of the plant-fungal interaction (Chancellor et al. 2023). Using spatial transcriptomics to visualise not only how *Gt* and *Gh* individually colonise wheat roots, but also how they interact with each other in the plant and the gene expression associated with this process, would undoubtedly shed light on this host-pathogen-antagonist system. Two putative orthologues of CSEP genes that have previously been implicated in pathogenicity were present in *Gt* and *Ga* pathogenic taxa but missing in non-pathogenic *Gh*, making them promising targets for future experiments to determine if either is important for *Gt* pathogenicity in wheat. *UvHrip1* (from *Ustilaginoidea virens*) is thought to be involved in suppressing host immunity and has already been reported in *Gt* (Wang et al. 2020), while *PBC1* (from *Pyrenopeziza brassicae*) is a cutinase implicated in host penetration (Li et al. 2003). It was intriguing that none of the CSEPs assigned to PHI-base genes were unique to *Gt*, perhaps suggesting that there is relatively high overlap in effector-mediated virulence mechanisms in *Gt* and *Ga*. In a similar pattern to the CSEPs, only one BGC (a PKS-NRP hybrid) was found to be specific to *Gt*, with most otherwise scattered across the genus (Supplemental Fig. S9c). There were more differences between *Gh* and the other lineages, and indeed the relative descriptive power of relatedness versus lifestyle on BGC variance (Fig. 4a) suggests that the production of secreted metabolites may be a key factor distinguishing the outcome of plant–fungal interactions in this genus. A single BGC has been shown to determine whether there is a mutualistic or pathogenic outcome of the interaction between root fungus *Colletotrichum tofieldiae* and *Arabidopsis thaliana* (Hiruma et al. 2023), demonstrating that minimal differences in metabolite repertoire can have considerable impacts on fungal lifestyle. In terms of host range, *Gt* has been shown to have low avenacinase activity relative to *Ga* (Osbourn et al. 1991), which is understood to be the reason *Gt* is incapable of also infecting oat roots (Osbourn et al. 1994). The avenacinase gene was nonetheless present in all strains across the genus; whether sequence polymorphism (Supplemental Fig. S7c) or differences in

372 regulatory machinery are responsible for the variation in avenacinase activity remains to be
373 determined. It is notable that *Gh* has also been found to be capable of colonising oat roots
374 (Osborne et al. 2018) despite greater divergence of the *Gh* avenacinase protein sequence
375 from *Ga* when compared to *Gt* (Supplemental Fig. S7b).

376 In line with the common understanding that *Gt* is self-fertile or homothallic (Palma-Guerrero et
377 al. 2021), we found both *MAT1-1* and *MAT1-2* idiomorphs to be present in the *GtA* and *GtB*
378 strains. These idiomorphs were located on two unlinked MAT loci, an atypical but occasionally
379 observed homothallic MAT locus architecture in ascomycetes (Wilken et al. 2017; Dyer et al.
380 2016; Thynne et al. 2017). Although it is homothallic, *Gt* is also capable of outcrossing
381 (Pilgeram and Henson 1992; Blanch et al. 1981), the rates of which may be underestimated
382 in many other homothallic fungi (Billiard et al. 2012; Attanayake et al. 2014). Similarly to *Gt*,
383 for *Ga* both MAT loci were identified. To our knowledge, the sex determination system of *Gh*
384 has not previously been reported, but our results indicate only one idiomorph at a single MAT
385 locus suggesting this species is self-sterile, or heterothallic. Evolutionary transitions between
386 heterothallism and homothallism are common in ascomycetes (Thynne et al. 2017; Sun et al.
387 2019; Gioti et al. 2012; Ene and Bennett 2014), but the implications on fitness are not fully
388 understood. In the scenario of a fungus infecting a crop monoculture, it may be advantageous
389 for the fungus to be homothallic when rapidly expanding across the niche, as it will not be
390 delayed by a reliance on the presence of compatible mating types. A higher rate of outcrossing
391 due to heterothallism could be unfavourable, as it could break up combinations that are
392 already well adapted to the genetically uniform host (Hill and McMullan 2023). Together, this
393 could suggest a selective pressure towards homothallism for crop fungal pathogens, and a
394 switch from heterothallism to homothallism may, therefore, have been a key innovation
395 underlying lifestyle divergence between non-pathogenic *Gh* and pathogenic *Gt* and *Ga*.

396 An unanticipated result was the absence of HCN genes in the *Gt*A lineage (Fig. 4a), despite
397 all other strains in the genus, including earlier diverging *Gh*, having genes which had
398 undergone copy-number expansions (Supplemental Fig. S11). These HCN genes were on

399 average significantly closer to TEs than other genes (Fig. 5c), which aligns with the fact that
400 TEs are known to play a role in gene duplication (Cerbin and Jiang 2018). GO enrichment
401 analysis identified a variety of fundamental biological processes to be significantly
402 overrepresented amongst HCN genes in the other lineages: regulation of cellular pH and
403 respiratory activity in non-pathogenic strains; and golgi organisation, protein localisation,
404 mRNA cis-splicing and respiratory activity in pathogenic strains. As previously mentioned,
405 alternative splicing has previously been linked to pathogenicity; respiratory activity has been
406 shown to induce a developmental switch to symbiosis in an arbuscular mycorrhizal fungus
407 (Tamasloukht et al. 2003); and mediation of cellular pH by V-ATPase has specifically been
408 linked to pathogenesis in *P. oryzae* (Chen et al. 2013), although here it was implicated in a
409 non-pathogenic *Gh* strain. Further investigation into the specific function of these genes is
410 required to determine whether any of these processes are essential to lifestyle or virulence in
411 *Gaeumannomyces*.

412 Gene duplicates are generally understood to be readily removed unless they serve to improve
413 host fitness, for instance by favourably modifying expression levels or rendering a completely
414 new function (Lynch and Conery 2000; Wapinski et al. 2007). RIP is a genome defence
415 response against unchecked proliferation of duplicated sequences (Hane et al. 2015). In
416 *Gaeumannomyces* we found 10–14% of the genome contained signatures of RIP, which is a
417 moderate level relative to other ascomycetes, e.g. *Pyronema confluens* (0.5%) (Traeger et al.
418 2013), *Fusarium* spp. (<1–6%) (Van Wyk et al. 2019), *Neurospora* spp. (8–23%) (Gioti et al.
419 2013), *Zymoseptoria tritici* (14–35%) (Lorrain et al. 2021) and *Hymenoscyphus* spp. (24–41%)
420 (Elfstrand et al. 2021). Genome-wide RIP was highest in *Gt*A, which was consistent with its
421 low level of gene duplication, but not fully explanatory as *Gh* had only marginally lower levels
422 of RIP while still maintaining HCN outliers. We can only presume that *Gt*A strains have been
423 under stronger selective pressures to remove duplicates, although the evolutionary
424 mechanisms driving this requires further investigation.

There was a similar pattern when exploring the RIP patterns across giant transposable *Starship* elements. We found only a single *Starship* in *Gt*A strains, which was gene-poor and had undergone extensive RIP (Fig. 5b), supporting the idea that this lineage employs stringent genome defence measures. By contrast, *Gt*B strains contained a proliferation of *Starships*, including one closely approaching the largest size reported thus far (Urquhart et al. 2023a). We expect that the increased availability of highly contiguous, long-read assemblies such as we report here will make the upper size extremes of such giant TEs more feasible to detect (Arkhipova and Yushenova 2019). Giant cargo-carrying TEs that can be both vertically and horizontally transmitted were first identified in bacteria (Johnson and Grossman 2015). Recently the *Starship* superfamily was identified as specific to and widespread in ascomycetes and, aside from the characteristic 'captain' tyrosine recombinase gene, each *Starship* contains a highly variable cargo (Gluck-Thaler et al. 2022). Mobilisation of cargo genes by *Starships* has been linked to the acquisition of various adaptive traits in fungal species, such as metal resistance (Urquhart et al. 2022), formaldehyde resistance (Urquhart et al. 2023a), virulence (McDonald et al. 2019), climatic adaptation (Tralamazza et al. 2023) and lifestyle switching (Gluck-Thaler et al. 2022). However, *Starships* are not inherently beneficial to the fungal host. One of the earliest groups of genes associated with the cargo of certain *Starships* was spore-killer or Spok genes, which bias their own transmission via the process of meiotic drive (i.e. by killing spores that do not inherit them) (Vogan et al. 2019). By incorporating Spok genes, a *Starship* element also biases its transmission, leading to it being referred to as a 'genomic hyperparasite' (Vogan et al. 2021). This corresponds to the concept of TEs as selfish genetic elements, which can prevail in the genome despite being neutral or deleterious to the overall fitness of the host. Whether mobilisation of an element and associated cargo is beneficial or detrimental to the host, TEs such as *Starships* are nonetheless drivers of genome evolution. Further detailed investigation of the specific cargo in the elements we have identified in *Gaeumannomyces* is a priority to explore how these giant TEs may be contributing to lifestyle and virulence.

While the differences in the overall appearance of the wheat plants and their root systems when infected with *Gt*A versus *Gt*B were visually compelling (Fig. 1A), our sample size was extremely limited and the quantitative data did not show such a strong distinction (Fig. 1C). A study by Lebreton et al. (2004) with a much larger sample size found *Gt* type A strains to be significantly more aggressive *in vitro* despite high intraspecific variability in take-all severity (type A=G2 in their study (Daval et al. 2010)). The dominance of type A strains in a site has also been reported to positively correlate with disease severity (Lebreton et al. 2007). It is also notable that five out of six wheat plants which died were inoculated with *Gt*A strains. Our phylogenomic analysis confirmed with significant branch support that the two lineages are indeed monophyletic (Supplemental Fig. S13b) and, together with our comparative genomics results, the question naturally arises as to whether *Gt*A and *Gt*B are in fact distinct species. We did not find evidence that genetic divergence between *Ga* and *Gt* species was more pronounced than between the *Gt*A and *Gt*B lineages, and host alone is not a sufficient distinction since, despite being a separate species, *Ga* is also able to infect wheat (Freeman and Ward 2004). Lebreton et al. (2004) suggested that 'genetic exchanges between [A and B] groups are rare events or even do not exist', but this was based on analysis of a limited number of genetic markers. Much broader whole-genome sequencing efforts are required to assess gene flow between lineages at the population-level, as well as the level of recombination. Understanding population dynamics could also shed light on the observed changes in ratio of *Gt*A and *Gt*B across wheat cropping years (Lebreton et al. 2004), which has implications for strategic crop protection measures.

Conclusions

We have generated near-complete assemblies with robust annotations for under-explored but agriculturally important wheat-associated *Gaeumannomyces* species. In doing so we confirmed that *Gaeumannomyces* taxa have one-compartment genomes in the context of TE/effector colocalisation, however the presence of giant cargo-carrying *Starship* TEs likely contributes to genomic plasticity. Genomic signatures support the separation of *Gt* into two

479 distinct lineages, with copy-number as a potential mechanism underlying differences in
480 virulence. We also found evidence that variation amongst the relatively low number of BGCs
481 may be a key factor contributing to lifestyle differences in *Gt* and *Gh*. In addition to providing
482 foundational data to better understand this host–pathogen–antagonist system, these new
483 resources are also an important step towards developing much-needed molecular diagnostics
484 for take-all, whether conventional amplicon sequencing, rapid *in situ* assays (Hariharan and
485 Prasannath 2021) or whole-genome/metagenomic sequencing approaches (Weisberg et al.
486 2021). Future research will require whole-genome sequencing of taxa from a broader
487 geographical range to produce a global pangenome, which will provide a comprehensive
488 reference for expression analyses to explore the role of virulence in *Gt* lineages, as well as
489 population genomics to shed light on their evolution and distribution.

490 **METHODS**

491 **Samples**

492 Nine *Gaeumannomyces* strains were selected from the Rothamsted Research culture
493 collections, including five *Gt* strains (two type A and three type B), two *Ga* strains and two *Gh*
494 strains (Supplemental Table S2). All were collected from various experimental fields at
495 Rothamsted Farm (Macdonald et al. 2018) between 2014 and 2018.

496 ***G. tritici* virulence test in adult wheat plants**

497 To test the virulence of the five *Gt* strains, we performed inoculations of each strain (six
498 replicates) into the highly susceptible winter wheat cultivar Hereward. First the roots of
499 seedling plants were inoculated with the fungus by using plastic drinking cups (7.5 cm wide x
500 11 cm tall) as pots, ensuring that all seedlings were well colonised before transferring to a
501 larger pot. Pots were drilled with four drainage holes 3 mm in diameter. A 50 cm^3 layer of damp
502 sand was added to each pot, followed by a 275 g layer of naïve soil collected from a field at
503 Rothamsted Farm after a non-legume break crop. Inoculum was prepared by taking a 9 mm
504 fungal plug with a cork borer number 6 from the outer part of a fungal colony grown on a potato

505 dextrose agar (PDA) plate and mixing with sand to make up a 25 g inoculum layer. A final 150
506 g layer of naïve soil was added on top of the inoculum layer. One wheat seed was sown on
507 the surface of the soil and covered with a 50 cm^3 layer of grit to aid germination and create a
508 humid environment for fungal colonisation. Pots were watered well and placed in a controlled
509 environment room (16 hr day, light intensity 250 µmols, 15°C day, 10°C night, watered twice
510 a week from above). A randomised block design was generated in Genstat 20th Edition to
511 take potential environmental differences across the growth room into account.

512 After two weeks of growth, each wheat seedling in a small pot was transferred by removing
513 the plastic cup and placing the entire contents undisturbed into a larger 20 cm diameter pot
514 containing a 2 cm layer of clay drainage pebbles. Three small pots were transferred to each
515 large pot and filled in with more soil, resulting in three plants per pot. There were 6 replicates
516 for each treatment, and a control pot with no fungus was also set up in the same manner, but
517 a PDA plate without fungus was used for preparing the inoculum layer. The pots were
518 transferred to a screenhouse and arranged randomly within blocks containing one pot per
519 treatment. The pots were established in September and remained outside in the screenhouse
520 to ensure exposure to winter conditions and therefore allow plant vernalisation to take place.

521 Measurements of the above-ground characteristics were first undertaken to note the severity
522 of any take-all symptoms once the floral spike (ear) was fully emerged. The height of each
523 labelled plant was measured from the stem base to the tip of the ear to the nearest 0.5 cm to
524 identify whether there was stunted growth. Additionally, the length of the ear and flag leaf were
525 recorded, again to the nearest 0.5 cm. The number of ears per plant was also recorded.

526 For below-ground measurements, the pots were washed out post full plant senescence and
527 the plants were well rinsed to remove the soil while minimising damage to the roots. Any roots
528 that broke off were collected and put into the cup with the main plants to maintain accuracy of
529 the biomass measurements. The stems were then cut about 10 cm from the base. The plants
530 were placed in a white tray filled with water to enable clear observation of the roots. The
531 number of tillers for each plant was counted. The severity of take-all infection was then

estimated by using the Take-All Index (TAI), classified through the following categories: Slight 1 (0–10% of roots infected), slight 2 (11–25%), moderate 1 (25–50%), moderate 2 (51–75%) and severe (76–100%). This was then input into the following formula: TAI = ((1 x % plants slight 1) + (2 x % plants slight 2) + (3 x % plants moderate 1) + (4 x % plants moderate 2) + (5 x % plants severe)) / 5 (Bateman et al. 2004). Following this, the length of the roots was measured to the nearest 0.5 cm. By cutting off one root at a time, the number of roots for each plant was counted and the roots transferred into cardboard trays, one per pot. These were then dried at 80°C on metal trays for 16 hours. One tray at a time was removed from the oven to reduce any moisture gain before weighing. The dried root biomass per pot was then recorded.

To statistically test for mean differences in the various characteristics between strains, we first made Q-Q plots using the ggqqplot function from ggpubr v0.6.0 (Kassambara 2020) to confirm approximate data normality. We then used the levene_test function from the package rstatix v0.7.2 (Kassambara 2021) to assess the assumption of homogeneity of variance, where a significant p value ($p < 0.05$) means that the assumption is violated. If we could ascertain homogeneity of variance, a multiple comparison test between strains was performed with the tukey_hsd rstatix function. Where homogeneity of variance was violated, the games_howell_test rstatix function was instead used for multiple comparison testing (Sauder and DeMars 2019).

Genome sequencing

For DNA and RNA extractions of all nine *Gaeumannomyces* taxa, a 4 mm plug of mycelium from axenic cultures was transferred to 500 ml of potato dextrose broth treated with penicillium/streptomycin (10,000 U/mL) using a sterile 4 mm corer. Cultures were grown at 20°C in dark conditions on an orbital shaker at 140 rpm for ~ 7–14 days. Mycelia were collected via vacuum filtration and flash frozen using liquid nitrogen and stored at -80°C, before grinding with a sterilised mortar and pestle until a fine powder was created.

DNA was extracted using one of two kits: the Phytopure Nucleon Genomic DNA kit (Cytiva, MA, USA) eluted in 50 µl low-pH TE buffer; and the NucleoBond HMW DNA kit (Macherey-Nagel, North Rhine- Westphalia, Germany) eluted in 100 µl–200 µl low-pH TE buffer. The manufacturer's protocols were modified to optimise for high molecular weight (M. Grey, personal communication). Sufficient DNA concentration (50 ng/µl DNA) was confirmed by Qubit fluorometer (Invitrogen, MA, USA) and purity (260/280 absorbance ratio of approximately 1.6–2.0 and 260/230 absorbance ratio of approximately 1.8–2.4) confirmed with a NanoDrop spectrophotometer (Thermo Fisher Scientific, MA, USA). Sufficient strand lengths (80% > 40 Kbp length) were confirmed using the Femto Pulse System (Agilent Technologies, Inc, CA, USA).

RNA from the same sample material was extracted using the Quick-RNA Fungal/Bacterial miniprep kit (Zymo Research, CA, USA) using the manufacturer's protocol and eluted in 25 µl of DNase/RNase free water. Sufficient RNA concentration (71 ng/µl RNA) was confirmed by Qubit fluorometer (Invitrogen, MA, USA) and purity (260/280 absorbance ratio of approximately 1.8–2.1 and 260/230 absorbance ratio of > 2.0) confirmed with a NanoDrop spectrophotometer (Thermo Fisher Scientific, MA, USA). An RNA integrity number > 8 was confirmed by Bioanalyzer RNA analysis (Agilent Technologies, Inc, CA, USA).

DNA and RNA extractions were sent to the Genomics Pipelines Group (Earlham Institute, Norwich, UK) for library preparation and sequencing. 2–5.5 µg of each sample was sheared using the Megaruptor 3 instrument (Diagenode, Liege, Belgium) at 18-20ng/µl and speed setting 31. Each sample underwent AMPure PB bead (PacBio, CA, USA) purification and concentration before undergoing library preparation using the SMRTbell Express Template Prep Kit 2.0 (PacBio) and barcoded using barcoded overhang adapters 8A/B (PacBio) and nuclease treated with SMRTbell enzyme cleanup kit 1.0 (PacBio). The resulting libraries were quantified by fluorescence (Invitrogen Qubit 3.0) and library size was estimated from a smear analysis performed on the Femto Pulse System (Agilent). The libraries were equimolar pooled into four multiplex pools and each pool was size fractionated using the SageELF system (Sage

Science, MA, USA), 0.75% cassette (Sage Science). The resulting fractions were quantified by fluorescence via Qubit and size estimated from a smear analysis performed on the Femto Pulse System, and 1–2 fractions per pool were selected for sequencing and pooled equimolar to have equal representation of barcodes per pool. The loading calculations for sequencing were completed using the PacBio SMRTLink Binding Calculator v10.1.0.119528 or v10.2.0.133424. Sequencing primer v2 or v5 was annealed to the adapter sequence of the library pools. Binding of the library pools to the sequencing polymerase was completed using Sequel II Binding Kit v2.0 or 2.2 (PacBio). Calculations for primer to template and polymerase to template binding ratios were kept at default values. Sequel II DNA internal control was spiked into the library pool complexes at the standard concentration prior to sequencing. The sequencing chemistry used was Sequel II Sequencing Plate 2.0 (PacBio) and the Instrument Control Software v10.1.0.119549 or 10.1.0.125432. Each pool was sequenced on 1-2 Sequel II SMRTcells 8M (PacBio) on the Sequel IIe instrument. The parameters for sequencing were as follows: CCS sequencing mode; 30-hour movie; 2-hour adaptive loading set to 0.85 or diffusion loading; 2-hour immobilisation time; 2–4-hour pre-extension time; and 70–86pM on plate loading concentration.

RNA libraries were constructed using the NEBNext Ultra II RNA Library prep for Illumina kit (New England Biolabs, MA, USA), NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Multiplex Oligos for Illumina (96 Unique Dual Index Primer Pairs) at a concentration of 10 µM. RNA libraries were equimolar pooled, q-PCR was performed, and the pool was sequenced on the Illumina NovaSeq 6000 (Illumina, CA, USA) on one lane of a NVS300S4 flowcell with v1.5 chemistry producing a total of 3,370,873,981 reads.

Genome assembly

See Supplemental Fig. S14a for a schematic summarising the bioinformatics analyses. HiFi reads were assembled using hifiasm v0.16.1-r375 (Cheng et al. 2021) with the -l 0 option to disable purging of duplicates in these haploid assemblies. The assemblies were checked for content correctness with respect to the input HiFi reads using the COMP tool from KAT v2.3.4

(Mapleson et al. 2017), and QUAST v5.0.2 (Mikheenko et al. 2018) was used to calculate contiguity statistics. BlobTools v1.0.1 (Laetsch and Blaxter 2017) was used to check for contamination (Supplemental Fig. S15) — this required a hits file, which we produced by searching contigs against the nt database (downloaded 21/05/2021) using blastn v2.10, and a BAM file of mapped HiFi reads, which we produced using minimap2 v2.21 (Li 2018) and samtools v1.13 (Li et al. 2009).

Gene set completeness was assessed using the ascomycota_odb10.2020-09-10 dataset in BUSCO v5.2.1 (Manni et al. 2021). This revealed some gene duplication due to the presence of small contigs that had exceptionally low coverage (median of 1 across each small sequence) when projecting the kmer spectra of the reads onto them using KAT's SECT tool. This was taken as evidence that the sequences did not belong in the assemblies. A custom script was written to filter out these small, low-coverage sequences, using the output of KAT SECT. KAT COMP, BUSCO and QUAST were re-run for the coverage filtered assemblies to verify that duplicated genes were removed without losing core gene content and produce final assembly contiguity statistics (Supplemental Fig. S16, Supplemental Table S1).

Genome annotation

Repeats were identified and masked using RepeatModeler v1.0.11 (Smit and Hubley 2015) and RepeatMasker v4.0.7 (Smit et al. 2015) via EIRepeat v1.1.0 (Kaithakottil and Swarbreck 2023). Gene models were annotated via the Robust and Extendable Eukaryotic Annotation Toolkit (REAT) v0.6.3 (EI-CoreBioinformatics 2023b) and MINOS v1.9 (EI-CoreBioinformatics 2023a). The REAT workflow consists of three submodules: transcriptome, homology, and prediction. The transcriptome module utilised Illumina RNA-Seq data, reads that were mapped to the genome with HISAT2 v2.1.0 (Kim et al. 2019) and high-confidence splice junctions identified by Portcullis v1.2.4 (Mapleson et al. 2018). The aligned reads were assembled for each tissue with StringTie2 v1.3.3 (Kovaka et al. 2019) and Scallop v0.10.2 (Shao and Kingsford 2017). A filtered set of non-redundant gene models were derived from the combined set of RNA-Seq assemblies using Mikado v2.3.4 (Venturini et al. 2018). The REAT homology

639 workflow was used to generate gene models based on alignment of protein sequences from
640 publicly available annotations of 27 related species (Supplemental Table S3) and a set of
641 proteins downloaded from UniProt including all the proteins from the class *Sordariomycetes*
642 (taxid:147550) and excluding all proteins from the publicly available annotation of *Gt* R3-111a-
643 1 (GCF_000145635). The prediction module generated evidence-guided models based on
644 transcriptome and proteins alignments using AUGUSTUS v3.4.0 (Stanke et al. 2006), with
645 four alternative configurations and weightings of evidence, and EVidenceModeler v1.1.1
646 (Haas et al. 2008). In addition, gene models from the *Gt* R3-111a-1 annotation were projected
647 via Liftoff v1.5.1 (Shumate and Salzberg 2021), and filtered via the multicompare script from
648 the ei-liftover pipeline (Venturini and Yanes 2020), ensuring only models with consistent gene
649 structures between the original and transferred models were retained.

650 The filtered Liftoff, REAT transcriptome, homology and prediction gene models were used in
651 MINOS to generate a consolidated gene set with models selected based on evidence support
652 and their intrinsic features. Confidence and biotype classification was determined for all gene
653 models based on available evidence, such as homology support and expression. TE gene
654 classification was based on overlap with identified repeats (> 40 bp repeat overlap).

655 To make best use of having multiple identically generated annotations for the genus, we opted
656 to additionally repeat a lift-over process projecting the gene models from each MINOS run to
657 all nine assemblies. We then removed gene models overlapping rRNA genes from the
658 multiple-lift-over annotations and the previously consolidated MINOS annotation using
659 RNAmmer v1.2 (Lagesen et al. 2007) and BEDTools v2.28 (Quinlan and Hall 2010). The
660 MINOS consolidation stage was repeated using four files as input: the high-confidence models
661 from the lift-over; the high-confidence genes of the previous MINOS run for the specific
662 assembly; the low-confidence models of the previous MINOS run for the specific assembly;
663 and the low-confidence models of the lift-over of all the closely related species. This multiple-
664 lift-over approach allowed us to cross-check gene sets across strains and determine whether
665 missing genes were truly absent from individual assemblies or had just been missed by the

annotation process. Finally, mitochondrial contigs were identified using the MitoHiFi v2.14.2 pipeline (Uliano-Silva et al. 2023), with gene annotation using MitoFinder v1.4.1 (Allio et al. 2020) and the mitochondrion sequence from *Epichloë novae-zelandiae* AL0725 as a reference (GenBank accession NC_072722.1).

Functional annotation of the gene models was performed using AHRD v3.3.3 (Hallab et al. 2023), with evidence from blastp v2.6.0 searches against the Swiss-Prot and TrEMBL databases (both downloaded on 19/10/2022), and mapping of domain names using InterProScan v5.22.61 (Jones et al. 2014). Additional annotations were produced using eggNOG-mapper v2.1.9 (Cantalapiedra et al. 2021) with sequence searches against the eggNOG orthology database (Huerta-Cepas et al. 2019) using DIAMOND v2.0.9 (Buchfink et al. 2021). CAZymes were predicted using run_dbcan v3.0.1 (Le and Yohe 2021) from the dbCAN2 CAZyme annotation server (Zhang et al. 2018) this process involved (i) HMMER v3.3.2 (Mistry et al. 2013) search against the dbCAN HMM (hidden Markov model) database; (ii) DIAMOND v2.0.14 search against the CAZy pre-annotated CAZyme sequence database (Drula et al. 2022) and (iii) eCAMI (Xu et al. 2020) search against a CAZyme short peptide library for classification and motif identification. A gene was classified as a CAZyme if all three methods were in agreement.

CSEPs were predicted using a similar approach to Hill et al. (2022), with some additions/substitutions of tools informed by Jones et al. (2021); see Supplemental Fig. S14b for a schematic overview. Briefly, we integrated evidence from SignalP v3.0 (Dyrløv Bendtsen et al. 2004), v4.1g (Petersen et al. 2011), v6.0g (Teufel et al. 2022); TargetP v2.0 (Almagro Armenteros et al. 2019); DeepSig v1.2.5 (Savojardo et al. 2018); Phobius v1.01 (Käll et al. 2004); TMHMM v2.0c (Krogh et al. 2001); Deeploc v1.0 (Almagro Armenteros et al. 2017); ps_scan v1.86 (Gattiker et al. 2002); and EffectorP v1.0 (Sperschneider et al. 2016), v2.0 (Sperschneider et al. 2018) and v3.0 (Sperschneider and Dodds 2021). CSEPs were then matched to experimentally verified genes in the PHI-base database (Urban et al. 2020) (downloaded 21/07/2023) using a BLAST v2.10 blastp search with an e-value cutoff of 1e-25.

In the event of multiple successful hits, the hit with the top bitscore was used. Secondary metabolites were predicted using antiSMASH v6.1.1 (Blin et al. 2021). Reference protein sequences for avenacinase from *Ga* (GenBank accession AAB09777.1) and mating-type locus idiomorphs *MAT1-1* and *MAT1-2* from *Pyricularia grisea* (Latorre et al. 2022) were used to identify their respective genes in each of the nine assemblies using a blastp search (e-value cutoff 1e-25).

Phylogenetic classification of *G. tritici* types

To confirm the classification of *Gt* strains within established genetic groups — *sensu* Daval et al. (2010) and Freeman et al. (Freeman et al. 2005) — gene trees were produced for gentisate 1,2-dioxygenase (*gdo*; GenBank accessions FJ717712–FJ717728) and *ITS2*. GenePull (Hill 2021) was used to extract the two marker sequences from the new assemblies reported here. *ITS2* could not be found in the existing *Gt* R3-111a-1 assembly (RefSeq accession GCF_000145635.1), so that strain was only included in the *gdo* gene tree. We aligned each marker gene separately using MAFFT v7.271 (Katoh and Standley 2013) and manually checked the gene alignments. The gene trees were estimated using RAxML-NG v1.1.0 (Kozlov et al. 2019) and the GTR+G nucleotide substitution model (Supplemental Fig. S13a). Branch support was computed using 1,000 Felsenstein's bootstrap replicates, or until convergence according to the default 3% cutoff for weighted Robinson-Foulds distances (Pattengale et al. 2009), whichever occurred first. An avenacinase gene tree was produced in the same way but using the JTT+G4 amino acid substitution model.

Phylogenomics of *Gaeumannomyces*

A genome-scale species tree was produced to provide evolutionary context for comparative analyses. We used OrthoFinder v2.5.4 (Emms and Kelly 2019) to cluster predicted gene models for primary transcripts into orthogroups — in addition to the newly sequenced *Gaeumannomyces* taxa, this also included *Gt* R3-111a-1 and the outgroup *Magnaporthiopsis poae* ATCC 64411 (GenBank accession GCA_000193285.1). Alongside the coalescent

719 species tree produced within OrthoFinder by STAG (Emms and Kelly 2018), we also used a
720 concatenation-based approach. We used MAFFT to produce gene alignments for 7,029
721 single-copy phylogenetic hierarchical orthogroups or HOGs (hereafter, genes) that were
722 present in all taxa. These were trimmed using v1.4.rev15 (Capella-Gutiérrez et al. 2009),
723 concatenated using AMAS and run in RAxML-NG with genes partitioned and the JTT+G4
724 amino acid substitution model. Branch support was calculated as above.

725 Alongside the species tree we visualised assembly N50; the number of gene models; the
726 proportion of these that were functionally annotated by AHRD; and the number of unassigned
727 gene models from OrthoFinder (Supplemental Fig. S17). Due to concerns regarding the
728 comparability of the existing *Gt* R3-111a-1 annotation to the strains reported in this study, and
729 to avoid introducing computational bias, the existing *Gt* R3-111a-1 annotation was excluded
730 from downstream comparative analyses for the sake of consistency.

731 **Genome structure and synteny**

732 To identify both potential misassemblies and real structural novelty in our strains, we used
733 GENESPACE v1.1.8 (Lovell et al. 2022) to visualise syntenic blocks across the genomes.
734 Fragments were considered to have telomeres at the ends if Tapestry v1.0.0 (Davey et al.
735 2020) identified at least five telomeric repeats (TTAGGG), and this was used together with the
736 GENESPACE results to inform pseudochromosome designation. Telomeric repeats were also
737 cross-checked with results from tidk v0.2.31 (Brown 2023). We calculated GC content across
738 pseudochromosomes in 100,000 bp windows using BEDTools v2.29.2 (Quinlan and Hall
739 2010), and TE, gene and CSEP density were calculated in 100,000 bp windows with a custom
740 script, plot_ideograms.R. The composite RIP index (CRI) (Lewis et al. 2009) was calculated
741 in 500 bp windows using RIP_index_calculation.pl (Stajich 2023).

742 To statistically test for correlations between CSEP density and TE and /or gene density, we
743 again made Q-Q plots using the ggqqplot function to assess approximate data normality. This
744 being violated, we calculated Kendall's tau for each strain (rstatix cor_test function,

method="kendall"). The assumption of normality being similarly violated for distances from CSEPs/other genes to the closest TE, we performed a Wilcoxon rank sum test (wilcox_test function) to compare mean distances for CSEPs versus other genes for each strain. To compare the mean gene–TE distance across strains, we used a Games-Howell test (games_howell_test function) for multiple comparison testing. Comparison of distances between HCN genes and TEs versus other genes and TEs was tested in the same way.

We also performed permutation tests of CSEP–TE distances using the meanDistance evaluation function from the R package regioneR v1.32.0 (Gel et al. 2016), with the resampleRegions function used for randomisation of the gene universe over 1,000 permutations. Permutation tests of CSEP–telomere distances were performed in the same way, having assigned the first and last 10,000 bp of each pseudochromosome as telomeric regions.

Comparative genomics

Functional annotations were mapped to orthogroups using a custom script, orthogroup_assigner.R, adapted from Hill et al. (2022), which also involved retrieval of CAZyme names from the ExplorEnz website (McDonald et al. 2009) using the package rvest v1.0.3 (Wickham 2020). CAZyme families known to act on the major plant cell wall substrates were classified as by Hill et al. (2022) based on the literature (Glass et al. 2013; Levasseur et al. 2013; Zhao et al. 2013; Miyauchi et al. 2020; Hage and Rosso 2021; Mesny et al. 2021). For *Gt*, gene content was categorised as core (present in all strains), soft-core (present in all but one strain), accessory (present in at least two strains) and specific (present in one strain).

Broadscale differences in gene repertoires due to lifestyle (pathogenic *Gt* and *Ga* and non-pathogenic *Gh*) were statistically tested using a permutational analysis of variance (PERMANOVA) approach to estimate residual variance of gene content after accounting for variance explained by phylogenetic distance (Mesny and Vannier 2020). To analyse the potential for secondary metabolite production with this PERMANOVA approach, a presence-

771 absence matrix for biosynthetic gene cluster families was produced from the antiSMASH
772 results using BiG-SCAPE v1.1.5 (Navarro-Muñoz et al. 2020).

773 Gene duplicates were categorised as intrachromosomal (on the same pseudochromosome)
774 or interchromosomal (on a different pseudochromosome) using the pangenes output files from
775 GENESPACE. We conducted gene ontology (GO) enrichment analysis for high copy-number
776 (HCN) genes using the R package topGO v2.50.0 (Alexa and Rahnenfuhrer 2022) with
777 Fisher's exact test and the weight01 algorithm.

778 **_Starship_ element identification**

779 Giant transposable *Starship* elements were identified in our assemblies after noting dense
780 blocks of transposons forming gaps between annotated genes. Manual inspection of these
781 regions via synteny plots built with OMA v2.5.0 (Altenhoff et al. 2019) and Circos v0.69
782 (Krzywinski et al. 2009) revealed *Starship*-sized insertions (Gluck-Thaler et al. 2022), and an
783 NCBI blastp search of the first gene in one such insertion in strain Gt-8d (Gt-
784 8d_EIv1_0041140) returned 85% identity with an established *Gt* R3-111a-1 DUF3435 gene
785 (GenBank accession EJT80010.1). These two genes were then used for a local blastp v2.13.0
786 search against all nine *Gaeumannomyces* assemblies reported here, which identified 33 full
787 length hits (>95% identity) that were associated with insertions when visualised in Circos plots.
788 This manual approach was then compared to *Starship* element identification using starfish
789 v1.0 (Gluck-Thaler and Vogan 2023). One element identified by starfish was discounted as it
790 consisted solely of a single predicted captain gene with no cargo or flanking repeats. A gene
791 tree of all tyrosine recombinases predicted by starfish (including *Starship* captains), blastp-
792 identified DUF3435 homologues, and previously reported *Starship* captain genes (Gluck-
793 Thaler et al. 2022) was built using the same methods described above for phylogenetic
794 classification and the JTT+G4 amino acid substitution model, with the addition of alignment
795 trimming using trimAl v1.4.rev15 (Capella-Gutiérrez et al. 2009) with the -gappyout parameter.

Data visualisation was completed in R v4.3.1 (R Core Team 2022) using the packages ape v5.7-1 (Paradis and Schliep 2019), aplot v0.2.2 (Yu et al. 2023), ComplexUpset v1.3.3 (Krassowski 2022), cowplot v1.1.1 (Wilke 2020), data.table v1.14.8 (Dowle and Srinivasan 2023), eulerr v7.0.0 (Larsson 2020), ggforce v0.4.1 (Pedersen 2021), ggh4x v0.2.6 (van den Brand 2023), gggenomes v0.9.12.9000 (Hackl et al. 2023), ggmsa v1.6.0 (Zhou et al. 2022), ggnewscale v0.4.9 (Campitelli 2020), ggplot2 v3.4.4 (Wickham 2016), ggplotify v0.1.2 (Yu 2021), ggpubr v0.6.0 (Kassambara 2020), ggrepel v0.9.3 (Slowikowski 2020), ggtree v3.9.1 (Yu et al. 2017), Gviz v1.44.2 (Hahne and Ivannek 2016), matrixStats v1.0.0 (Bengtsson 2021), multcompView v0.1-9 (Graves et al. 2019), patchwork v1.1.3 (Pederson 2022), rtracklayer v1.60.1 (Lawrence et al. 2009), scales v1.2.1 (Wickham and Seidel 2020), seqmagick v0.1.6 (Yu 2023), tidyverse v2.0.0 (Wickham et al. 2019). All analysis scripts are available at https://github.com/Rowena-h/GaeumannomycesGenomics.

DATA ACCESS

WGS data and annotated genome assemblies are available on GenBank under the BioProject accession PRJNA935249 (assemblies pending release). Additional data files are deposited in Zenodo doi:10.5281/zenodo.10277622 (pending release). All bioinformatics scripts are available at https://github.com/Rowena-h/GaeumannomycesGenomics.

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

ACKNOWLEDGEMENTS

We thank Jonathan Storkey at Rothamsted Research for grass species identification following the sampling of the Park Grass long term pasture experiment at Rothamsted. Many thanks to Suzanne Clark in the Computational and Analytical Sciences group at Rothamsted Research for the randomised block design in the inoculation experiments and for the subsequent data analyses. Thanks to Rachel Rusholme Pilcher for help and advice with GENESPACE, and other members of the Anthony Hall lab group for general discussion. The authors acknowledge

funding from the Biotechnology and Biological Sciences Research Council (BBSRC), part of UK Research and Innovation, Core Capability Grants BB/CCG1720/1 and the work delivered via the Scientific Computing group, as well as support for the physical HPC infrastructure and data centre delivered via the NBI Computing infrastructure for Science (CiS) group. This research was funded by the BBSRC grants Designing Future Wheat (BB/P016855/1) and Delivering Sustainable Wheat (BB/X011003/1) and the constituent work packages (BBS/E/ER/230003B Earlham Institute and BBS/E/RH/230001B Rothamsted Research). Part of this work was delivered via the BBSRC National Capability in Genomics and Single Cell Analysis (BBS/E/T/000PR9816) at the Earlham Institute by Suzanne Henderson of the Genomics Pipelines Group. The long-term Park Grass field experiment at Rothamsted Research and the entire Rothamsted Experimental Farm is supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the Lawes Agricultural Trust. D Smith is supported by the BBSRC Core Capability Grant (BB/CCG2280/1). GC was supported by the DEFRA funded Wheat Genetic Improvement Network (WGIN) phase 3 (CH0106) and GC and JS were supported by WGIN phase 4 (CH0109). S-JO and TC were supported by the BBSRC funded University of Nottingham Doctoral Training Programme (BB/M008770/1). JH was supported by a UK government–Rothamsted Research level 3 Laboratory Technician Apprenticeship scheme.

Author contributions

RH, MM, NH, KH-K and JP-G conceived, managed and/or coordinated the work. NH, MM and KH-K were involved in funding acquisition. MM, JPG and KH-K supervised different aspects of the research. GC, VEM, S-JO, JH, MG and MM collected the samples and/or isolated strains. JP-G, JS and JH performed *Gt* inoculation experiments. MG and NI performed molecular lab work. SJW performed genome assembly analyses. MOF performed genome annotation analyses, supervised by D Swarbreck. RH performed functional annotation analyses; designed and performed phylogenetic, comparative and statistical analyses; and performed data visualisations. D Smith performed the exploratory *Starship* analyses. RH and

MM wrote the manuscript with contributions from MG, MOF, TC, D Smith, NI, NH, JP-G, and KH-K. All authors read and approved the manuscript.

REFERENCES

Accinelli C, Abbas HK, Little NS, Kotowicz JK, Mencarelli M, Shier WT. 2016. A liquid bioplastic formulation for film coating of agronomic seeds. *Crop Prot* **89**: 123–128. doi:10.1016/j.cropro.2016.07.010.

Alexa A, Rahnenfuhrer J. 2022. topGO: Enrichment Analysis for Gene Ontology. https://bioconductor.org/packages/release/bioc/html/topGO.html.

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour* **20**: 892–905. doi:10.1111/1755-0998.13160.

Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, Von Heijne G, Elofsson A, Nielsen H. 2019. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance* **2**: e201900429. doi:10.26508/lsa.201900429.

Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. 2017. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**: 3387–3395. doi:10.1093/bioinformatics/btx431.

Altenhoff AM, Levy J, Zarowiecki M, Tomiczek B, Warwick Vesztrocy A, Dalquen DA, Müller S, Telford MJ, Glover NM, Dylus D, et al. 2019. OMA standalone: orthology inference among public and custom genomes and transcriptomes. *Genome Res* **29**: 1152–1163. doi:10.1101/gr.243212.118.

Arkhipova IR, Yushenova IA. 2019. Giant Transposons in Eukaryotes: Is Bigger Better? *Genome Biol Evol* **11**: 906–918. doi:10.1093/gbe/evz041.

Attanayake RN, Tennekoon V, Johnson DA, Porter LD, Del Río-Mendoza L, Jiang D, Chen W. 2014. Inferring outcrossing in the homothallic fungus *Sclerotinia sclerotiorum* using linkage disequilibrium decay. *Heredity* **113**: 353–363. doi:10.1038/hdy.2014.37.

Badet T, Croll D. 2020. The rise and fall of genes: origins and functions of plant pathogen pangenomes. *Curr Opin Plant Biol* **56**: 65–73. doi:10.1016/j.pbi.2020.04.009.

Badet T, Oggenfuss U, Abraham L, McDonald BA, Croll D. 2020. A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biol* **18**: 12. doi:10.1186/s12915-020-0744-3.

Balmer D, Mauch-Mani B. 2013. More beneath the surface? Root *versus* shoot antifungal plant defenses. *Front Plant Sci* **4**: 256. doi:10.3389/fpls.2013.00256.

Bao J, Chen M, Zhong Z, Tang W, Lin L, Zhang X, Jiang H, Zhang D, Miao C, Tang H, et al. 2017. PacBio Sequencing Reveals Transposable Elements as a Key Contributor to Genomic Plasticity and Virulence Variation in *Magnaporthe oryzae*. *Mol Plant* **10**: 1465–1468. doi:10.1016/j.molp.2017.08.008.

887 Bateman GL, Gutteridge RJ, Jenkyn JF. 2004. Take-all and grain yields in sequences of
888 winter wheat crops testing fluquinconazole seed treatment applied in different
889 combinations of years. *Ann Appl Biol* **145**: 317–330. doi:10.1111/j.1744-
890 7348.2004.tb00389.x.

891 Bateman L, Ward E, Hornby D, Gutteridge RJ. 1997. Comparisons of isolates of the take-all
892 fungus, *Gaeumannomyces graminis* var. *tritici*, from different cereal sequences using
893 DNA probes and non-molecular methods. *Soil Biol Biochem* **29**: 1225–1232.
894 doi:10.1016/S0038-0717(97)00025-4.

895 Bengtsson H. 2021. matrixStats: Functions that Apply to Rows and Columns of Matrices
896 (and to Vectors). https://cran.r-project.org/package=matrixStats.

897 Billiard S, López-Villavicencio M, Hood ME, Giraud T. 2012. Sex, outcrossing and mating
898 types: unsolved questions in fungi and beyond. *J Evolution Biol* **25**: 1020–1038.
899 doi:10.1111/j.1420-9101.2012.02495.x.

900 Blanch PA, Asher MJC, Burnett JH. 1981. Inheritance of pathogenicity and cultural
901 characters in *Gaeumannomyces graminis* var. *tritici*. *T Brit Mycol Soc* **77**: 391–399.
902 doi:10.1016/S0007-1536(81)80042-3.

903 Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, Van Wezel GP, Medema MH, Weber
904 T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities.
905 *Nucleic Acids Res* **49**: W29–W35. doi:10.1093/nar/gkab335.

906 Bowyer P, Clarke BR, Lunness P, Daniels MJ, Osbourn AE. 1995. Host Range of a Plant
907 Pathogenic Fungus Determined by a Saponin Detoxifying Enzyme. *Science* **267**:
908 371–374. doi:10.1126/science.7824933.

909 Brown M. 2023. A Telomere Identification toolKit (tidk). https://github.com/tolkit/telomeric-
910 identifier.

911 Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using
912 DIAMOND. *Nat Methods* **18**: 366–368. doi:10.1038/s41592-021-01101-x.

913 Campitelli E. 2020. ggnewscale: Multiple Fill and Colour Scales in “ggplot2.” https://cran.r-
914 project.org/package=ggnewscale.

915 Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-
916 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at
917 the Metagenomic Scale. *Mol Biol Evol* **38**: 5825–5829. doi:10.1093/molbev/msab293.

918 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated
919 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–
920 1973. doi:10.1093/bioinformatics/btp348.

921 Cerbin S, Jiang N. 2018. Duplication of host genes by transposable elements. *Curr Opin*
922 *Genet Dev* **49**: 63–69. doi:10.1016/j.gde.2018.03.005.

923 Chancellor T. 2022. Evaluating the potential of non-pathogenic *Magnaporthaceae* species
924 for the control of take-all disease in wheat. University of Nottingham, UK.

925 Chancellor T, Smith DP, Chen W, Clark SJ, Venter E, Halsey K, McMillan V, Canning G,
926 Hammond-Kosack KE, Palma-Guerrero J. 2023.Exploring the family feud: a fungal
927 endophyte induces local cell wall-mediated resistance in wheat roots against the

928 closely related “take-all” pathogen. *bioRxiv [Preprint]*.
929 doi:10.1101/2023.11.23.568424.

930 Chen G, Liu X, Zhang L, Cao H, Lu J, Lin F. 2013. Involvement of *MoVMA11*, a Putative
931 Vacuolar ATPase c’ Subunit, in Vacuolar Acidification and Infection-Related
932 Morphogenesis of *Magnaporthe oryzae*. *PLoS ONE* **8**: e67804.
933 doi:10.1371/journal.pone.0067804.

934 Chen H, King R, Smith D, Bayon C, Ashfield T, Torriani S, Kanyuka K, Hammond-Kosack K,
935 Bieri S, Rudd J. 2023. Combined pangenomics and transcriptomics reveals core and
936 redundant virulence processes in a rapidly evolving fungal plant pathogen. *BMC Biol*
937 **21**: 24. doi:10.1186/s12915-023-01520-6.

938 Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo
939 assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**: 170–175.
940 doi:10.1038/s41592-020-01056-5.

941 Cheng J, Yin Z, Zhang Z, Liang Y. 2018. Functional analysis of MoSnf7 in *Magnaporthe*
942 *oryzae*. *Fungal Genet Biol* **121**: 29–45. doi:10.1016/j.fgb.2018.09.005.

943 Croll D, Zala M, McDonald BA. 2013. Breakage-fusion-bridge Cycles and Large Insertions
944 Contribute to the Rapid Evolution of Accessory Chromosomes in a Fungal Pathogen.
945 *PLoS Genet* **9**: e1003567. doi:10.1371/journal.pgen.1003567.

946 da C. Godinho RM, Crestani J, Kmetzsch L, De S. Araujo G, Frases S, Staats CC, Schrank
947 A, Vainstein MH, Rodrigues ML. 2014. The vacuolar-sorting protein Snf7 is required
948 for export of virulence determinants in members of the *Cryptococcus neoformans*
949 complex. *Sci Rep* **4**: 6198. doi:10.1038/srep06198.

950 Daval S, Lebreton L, Gazengel K, Guillerm-Erckelboudt A-Y, Sarniguet A. 2010. Genetic
951 evidence for differentiation of *Gaeumannomyces graminis* var. *tritici* into two major
952 groups. *Plant Pathol* **59**: 165–178. doi:10.1111/j.1365-3059.2009.02158.x.

953 Davey JW, Davis SJ, Mottram JC, Ashton PD. 2020.Tapestry: validate and edit small
954 eukaryotic genome assemblies with long reads. *bioRxiv [Preprint]*.
955 doi:10.1101/2020.04.24.059402.

956 Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R,
957 Xu J-R, Pan H, et al. 2005. The genome sequence of the rice blast fungus
958 *Magnaporthe grisea*. *Nature* **434**: 980–986. doi:10.1038/nature03449.

959 Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens:
960 waltz with plants. *Curr Opin Genet Dev* **35**: 57–65. doi:10.1016/j.gde.2015.09.001.

961 Dowle M, Srinivasan A. 2023. data.table: Extension of `data.frame`. https://CRAN.R-
962 project.org/package=data.table.

963 Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The
964 carbohydrate-active enzyme database: functions and literature. *Nucleic Acids*
965 *Research* **50**: D571–D577. doi:10.1093/nar/gkab1045.

966 Dyer PS, Inderbitzin P, Debuchy R. 2016. Mating-Type Structure, Function, Regulation and
967 Evolution in the Pezizomycotina. In *Growth, Differentiation and Sexuality* (ed. J.
968 Wendland), Vol. 1 of *The Mycota, A Comprehensive Treatise on Fungi as*

Experimental Systems for Basic and Applied Research, pp. 351–385, Springer, Cham doi:10.1007/978-3-319-25844-7.

Dyrløv Bendtsen J, Nielsen H, von Heijne G, Brunak S. 2004. Improved Prediction of Signal Peptides: SignalP 3.0. *J Mol Biol* **340**: 783–795. doi:10.1016/j.jmb.2004.05.028.

EI-CoreBioinformatics. 2023a. minos - a gene model consolidation pipeline for genome annotation projects. https://github.com/EI-CoreBioinformatics/minos.

EI-CoreBioinformatics. 2023b. REAT - Robust and Extendable eukaryotic Annotation Toolkit. https://github.com/EI-CoreBioinformatics/reat.

Eizenga JM, Novak AM, Sibbesen JA, Heumos S, Ghaffaari A, Hickey G, Chang X, Seaman JD, Rounthwaite R, Ebler J, et al. 2020. Pangenome Graphs. *Annu Rev Genom Hum Genet* **21**: 139–162. doi:10.1146/annurev-genom-120219-080406.

Elfstrand M, Chen J, Cleary M, Halecker S, Ihrmark K, Karlsson M, Davydenko K, Stenlid J, Stadler M, Durling MB. 2021. Comparative analyses of the *Hymenoscyphus fraxineus* and *Hymenoscyphus albidus* genomes reveals potentially adaptive differences in secondary metabolite and transposable element repertoires. *BMC Genomics* **22**: 503. doi:10.1186/s12864-021-07837-2.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**: 238. doi:10.1101/466201.

Emms DM, Kelly S. 2018.STAG: Species Tree Inference from All Genes. *bioRxiv [Preprint]*. doi:10.1101/267914.

Ene IV, Bennett RJ. 2014. The cryptic sexual strategies of human fungal pathogens. *Nat Rev Microbiol* **12**: 239–251. doi:10.1038/nrmicro3236.

Fang S, Hou X, Qiu K, He R, Feng X, Liang X. 2020. The occurrence and function of alternative splicing in fungi. *Fungal Biol Rev* **34**: 178–188. doi:10.1016/j.fbr.2020.10.001.

Fonseca PLC, De-Paula RB, Araújo DS, Tomé LMR, Mendes-Pereira T, Rodrigues WFC, Del-Bem L-E, Aguiar ERGR, Góes-Neto A. 2021. Global Characterization of Fungal Mitogenomes: New Insights on Genomic Diversity and Dynamism of Coding Genes and Accessory Elements. *Front Microbiol* **12**: 787283. doi:10.3389/fmicb.2021.787283.

Fourie A, De Jonge R, Van Der Nest MA, Duong TA, Wingfield MJ, Wingfield BD, Barnes I. 2020. Genome comparisons suggest an association between *Ceratocystis* host adaptations and effector clusters in unique transposable element families. *Fungal Genet Biol* **143**: 103433. doi:10.1016/j.fgb.2020.103433.

Franco MEE, Wisecaver JH, Arnold AE, Ju Y, Slot JC, Ahrendt S, Moore LP, Eastman KE, Scott K, Konkel Z, et al. 2021. Ecological generalism drives hyperdiversity of secondary metabolite gene clusters in xylarialean endophytes. *New Phytol*. doi:10.1111/nph.17873.

Frantzeskakis L, Kusch S, Panstruga R. 2019. The need for speed: compartmentalized genome evolution in filamentous phytopathogens. *Mol Plant Pathol* **20**: 3–7. doi:10.1111/mpp.12738.

1010 Freeman J, Ward E. 2004. *Gaeumannomyces graminis*, the take-all fungus and its relatives.
1011 *Mol Plant Pathol* **5**: 235–252. doi:10.1111/j.1364-3703.2004.00226.x.

1012 Freeman J, Ward E, Gutteridge RJ, Bateman GL. 2005. Methods for studying population
1013 structure, including sensitivity to the fungicide silthiofam, of the cereal take-all fungus,
1014 *Gaeumannomyces graminis* var. *tritici*. *Plant Pathol* **54**: 686–698. doi:10.1111/j.1365-
1015 3059.2005.01252.x.

1016 Gattiker A, Gasteiger E, Bairoch A. 2002. ScanProsite: a reference implementation of a
1017 PROSITE scanning tool. *Appl Bioinf* **1**: 107–108. .

1018 Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016.
1019 regioneR: an R/Bioconductor package for the association analysis of genomic
1020 regions based on permutation tests. *Bioinformatics* **32**: 289–291.
1021 doi:10.1093/bioinformatics/btv562.

1022 Gioti A, Mushegian AA, Strandberg R, Stajich JE, Johannesson H. 2012. Unidirectional
1023 Evolutionary Transitions in Fungal Mating Systems and the Role of Transposable
1024 Elements. *Mol Biol Evol* **29**: 3215–3226. doi:10.1093/molbev/mss132.

1025 Gioti A, Stajich JE, Johannesson H. 2013. *Neurospora* and the dead-end hypothesis:
1026 genomic consequences of selfing in the model genus. *Evolution* **67**: 3600–3616.
1027 doi:10.1111/evo.12206.

1028 Glass NL, Schmoll M, Cate JHD, Coradetti S. 2013. Plant cell wall deconstruction by
1029 ascomycete fungi. *Annu Rev Microbiol* **67**: 477–498. doi:10.1146/annurev-micro-
1030 092611-150044.

1031 Gluck-Thaler E, Haridas S, Binder M, Grigoriev IV, Crous PW, Spatafora JW, Bushley K,
1032 Slot JC. 2020. The Architecture of Metabolism Maximizes Biosynthetic Diversity in
1033 the Largest Class of Fungi. *Mol Biol Evol* **37**: 2838–2856.
1034 doi:10.1093/molbev/msaa122.

1035 Gluck-Thaler E, Ralston T, Konkel Z, Ocampos CG, Ganeshan VD, Dorrance AE, Niblack
1036 TL, Wood CW, Slot JC, Lopez-Nicora HD, et al. 2022. Giant *Starship* Elements
1037 Mobilize Accessory Genes in Fungal Genomes. *Mol Biol Evol* **39**: msac109.
1038 doi:10.1093/molbev/msac109.

1039 Gluck-Thaler E, Vogan AA. 2023.Systematic identification of cargo-carrying genetic
1040 elements reveals new dimensions of eukaryotic diversity. *bioRxiv [Preprint]*.
1041 doi:10.1101/2023.10.24.563810.

1042 Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D. 2020. Pangenomics Comes of Age:
1043 From Bacteria to Plant and Animal Applications. *Trends GEnet* **36**: 132–145.
1044 doi:10.1016/j.tig.2019.11.006.

1045 Gómez Luciano LB, Tsai IJ, Chuma I, Tosa Y, Chen Y-H, Li J-Y, Li M-Y, Lu M-YJ,
1046 Nakayashiki H, Li W-H. 2019. Blast Fungal Genomes Show Frequent Chromosomal
1047 Changes, Gene Gains and Losses, and Effector Gene Turnover. *Mol Biol Evol* **36**:
1048 1148–1161. doi:10.1093/molbev/msz045.

1049 Graves S, Piepho H-P, Selzer L, Dorai-Raj S. 2019. multcompView: Visualizations of Paired
1050 Comparisons. https://cran.r-project.org/package=multcompView.

1051 Grutzmann K, Szafranski K, Pohl M, Voigt K, Petzold A, Schuster S. 2014. Fungal
1052 Alternative Splicing is Associated with Multicellular Complexity and Virulence: A
1053 Genome-Wide Multi-Species Study. *DNA Res* **21**: 27–39. doi:10.1093/dnares/dst038.

1054 Guin K, Sreekumar L, Sanyal K. 2020. Implications of the Evolutionary Trajectory of
1055 Centromeres in the Fungal Kingdom. *Annu Rev Microbiol* **74**: 835–853.
1056 doi:10.1146/annurev-micro-011720-122512.

1057 Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Robin CR, Wortman
1058 JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler
1059 and the Program to Assemble Spliced Alignments. *Genome Biol* **9**: R7.
1060 doi:10.1186/gb-2008-9-1-r7.

1061 Hackl T, Ankenbrand MJ, van Adrichem B. 2023. gggenomes: A Grammar of Graphics for
1062 Comparative Genomics. https://github.com/thackl/gggenomes.

1063 Hage H, Rosso MN. 2021. Evolution of Fungal Carbohydrate-Active Enzyme Portfolios and
1064 Adaptation to Plant Cell-Wall Polymers. *J Fungi* **7**: 1. doi:10.3390/jof7030185.

1065 Hahne F, Ivannek R. 2016. Visualizing Genomic Data Using Gviz and Bioconductor. In
1066 *Statistical Genomics: Methods and Protocols* (eds. E. Mathé and S. Davis), Vol. 1418
1067 of *Methods in Molecular Biology*, Springer New York, New York, NY doi:10.1007/978-
1068 1-4939-3578-9.

1069 Hallab A, Klee K, Boecker F, Srinivas G, Schoof H. 2023. Automated Assignment of Human
1070 Readable Descriptions (AHRD). https://github.com/groupschoof/AHRD.

1071 Hane JK, Williams AH, Taranto AP, Solomon PS, Oliver RP. 2015. Repeat-Induced Point
1072 Mutation: A Fungal-Specific, Endogenous Mutagenesis Process. In *Genetic*
1073 *Transformation Systems in Fungi* (eds. M.A. Van den Berg and K. Maruthachalam),
1074 Vol. 2 of *Fungal Biology*, pp. 55–68, Springer International Publishing
1075 doi:10.1007/978-3-319-10503-1_4.

1076 Hariharan G, Prasannath K. 2021. Recent Advances in Molecular Diagnostics of Fungal
1077 Plant Pathogens: A Mini Review. *Front Cell Infect Microbiol* **10**: 600234.
1078 doi:10.3389/fcimb.2020.600234.

1079 Hartmann FE. 2022. Using structural variants to understand the ecological and evolutionary
1080 dynamics of fungal plant pathogens. *New Phytol* **234**: 43–49. doi:10.1111/nph.17907.

1081 Hernández-Restrepo M, Groenewald JZ, Elliott ML, Canning G, McMillan VE, Crous PW.
1082 2016. Take-all or nothing. *Stud Mycol* **83**: 19–48. doi:10.1016/j.simyco.2016.06.002.

1083 Hill R. 2021. GenePull. https://github.com/Rowena-
1084 h/MiscGenomicsTools/tree/main/GenePull.

1085 Hill R, Buggs RJA, Vu DT, Gaya E. 2022. Lifestyle Transitions in Fusarioid Fungi are
1086 Frequent and Lack Clear Genomic Signatures. *Mol Biol Evol* **39**: msac085.
1087 doi:10.1093/molbev/msac085.

1088 Hill R, McMullan M. 2023. Recombination triggers fungal crop disease. *Nat Ecol Evol*.
1089 doi:10.1038/s41559-023-02132-7.

1090 Hiruma K, Aoki S, Takino J, Higa T, Utami YD, Shiina A, Okamoto M, Nakamura M,
1091 Kawamura N, Ohmori Y, et al. 2023. A fungal sesquiterpene biosynthesis gene

cluster critical for mutualist-pathogen transition in *Colletotrichum tofieldiae*. *Nat Commun* **14**: 5288. doi:10.1038/s41467-023-40867-w.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**: D309–D314. doi:10.1093/nar/gky1085.

Jeon J, Kim K-T, Choi J, Cheong K, Ko J, Choi G, Lee H, Lee G-W, Park S-Y, Kim S, et al. 2022. Alternative splicing diversifies the transcriptome and proteome of the rice blast fungus during host infection. *RNA Biol* **19**: 373–386. doi:10.1080/15476286.2022.2043040.

Johnson CM, Grossman AD. 2015. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu Rev Genet* **49**: 577–601. doi:10.1146/annurev-genet-112414-055018.

Jones DAB, Rozano L, Debler JW, Mancera RL, Moolhuijzen PM, Hane JK. 2021. An automated and combinative method for the predictive ranking of candidate effector proteins of fungal plant pathogens. *Sci Rep* **11**: 19731. doi:10.1038/s41598-021-99363-0.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240. doi:10.1093/bioinformatics/btu031.

Kaithakottil GG, Swarbreck D. 2023. EIRepeat - EI Repeat Identification Pipeline. https://github.com/EI-CoreBioinformatics/eirepeat.

Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338**: 1027–1036. doi:10.1016/j.jmb.2004.03.016.

Kassambara A. 2020. ggpubr: “ggplot2” Based Publication Ready Plots. https://cran.r-project.org/package=ggpubr.

Kassambara A. 2021. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. https://cran.r-project.org/package=rstatix.

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version : Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780. doi:10.1093/molbev/mst010.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915. doi:10.1038/s41587-019-0201-4.

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**: 278. doi:10.1186/s13059-019-1910-1.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455. doi:10.1093/bioinformatics/btz305.

1134 Krassowski M. 2022. ComplexUpset. https://github.com/krassowski/complex-
1135 upset/tree/v1.3.5.

1136 Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane
1137 protein topology with a hidden Markov model: Application to complete genomes. *J*
1138 *Mol Biol* **305**: 567–580. doi:10.1006/jmbi.2000.4315.

1139 Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.
1140 2009. Circos: An information aesthetic for comparative genomics. *Genome Res* **19**:
1141 1639–1645. doi:10.1101/gr.092759.109.

1142 Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies.
1143 *F1000Research* **6**: 1287. doi:10.12688/f1000research.12232.1.

1144 Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer:
1145 consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**:
1146 3100–3108. doi:10.1093/nar/gkm160.

1147 Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Malmgren A, Latorre SM, Burbano
1148 HA, Win J, Kamoun S. 2021. Genomic rearrangements generate hypervariable mini-
1149 chromosomes in host-specific isolates of the blast fungus. *PLoS Genet* **17**:
1150 e1009386. doi:10.1371/journal.pgen.1009386.

1151 Larsson J. 2020. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses.
1152 https://cran.r-project.org/package=eulerr.

1153 Latorre SM, Were VM, Langer T, Foster AJ, Win J, Kamoun S, Talbot NJ, Burbano HA.
1154 2022. A curated set of mating-type assignment for the blast fungus
1155 (Magnaporthales). doi:10.5281/zenodo.6369833.

1156 Lawrence M, Gentleman R, Carey V. 2009. rtracklayer: an R package for interfacing with
1157 genome browsers. *Bioinformatics* **25**: 1841–1842. doi:10.1093/bioinformatics/btp328.

1158 Le H, Yohe T. 2021. run_dbcan 3.0. https://github.com/linnabrown/run_dbcan.

1159 Lebreton L, Gosme M, Lucas P, Guillerm-Erckelboudt A-Y, Sarniguet A. 2007. Linear
1160 relationship between *Gaeumannomyces graminis* var. *tritici* (*Ggt*) genotypic
1161 frequencies and disease severity on wheat roots in the field. *Environ Microbiol* **9**:
1162 492–499. doi:10.1111/j.1462-2920.2006.01166.x.

1163 Lebreton L, Lucas P, Dugas F, Guillerm A-Y, Schoeny A, Sarniguet A. 2004. Changes in
1164 population structure of the soilborne fungus *Gaeumannomyces graminis* var. *tritici*
1165 during continuous wheat cropping. *Environ Microbiol* **6**: 1174–1185.
1166 doi:10.1111/j.1462-2920.2004.00637.x.

1167 Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the
1168 enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes.
1169 *Biotechnol Biofuels* **6**: 41. doi:10.1186/1754-6834-6-41.

1170 Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, Schübeler D, Selker EU.
1171 2009. Relics of repeat-induced point mutation direct heterochromatin formation in
1172 *Neurospora crassa*. *Genome Res* **19**: 427–437. doi:10.1101/gr.086231.108.

1173 Li D, Ashby AM, Johnstone K. 2003. Molecular Evidence that the Extracellular Cutinase
1174 Pbc1 Is Required for Pathogenicity of *Pyrenopeziza brassicae* on Oilseed Rape.
1175 *MPMI* **16**: 545–552. doi:10.1094/MPMI.2003.16.6.545.

1176 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:
1177 3094–3100. doi:10.1093/bioinformatics/bty191.

1178 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
1179 R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence
1180 Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
1181 doi:10.1093/bioinformatics/btp352.

1182 Llewellyn T, Nowell RW, Aptroot A, Temina M, Prescott TAK, Barraclough TG, Gaya E.
1183 2023. Metagenomics Shines Light on the Evolution of “Sunscreen” Pigment
1184 Metabolism in the *Teloschistales* (Lichen-Forming Ascomycota). *Genome Biol Evol*
1185 **15**: evad002. doi:10.1093/gbe/evad002.

1186 Lorrain C, Feurtey A, Möller M, Haueisen J, Stukenbrock E. 2021. Dynamics of transposable
1187 elements in recently diverged fungal pathogens: lineage-specific transposable
1188 element content and efficiency of genome defenses. *G3* **11**: jkab068.
1189 doi:10.1093/g3journal/jkab068.

1190 Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D,
1191 Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene
1192 copy number variation across multiple genomes. *eLife* **11**: e78526.
1193 doi:10.7554/eLife.78526.

1194 Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes.
1195 *Science* **290**: 1151–1155. doi:10.1126/science.290.5494.1151.

1196 Macdonald A, Poulton P, Clark I, Scott T, Glendining M, Perryman S, Storkey J, Bell J,
1197 Shield I, McMillan V, et al. 2018. *Guide to the classical and other long-term*
1198 *experiments, datasets and sample archive*. ed. A.J. Macdonald. Rothamsted
1199 Research, Harpenden, Hertfordshire.

1200 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel
1201 and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage
1202 for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**: 4647–
1203 4654. doi:10.1093/molbev/msab199.

1204 Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer
1205 analysis toolkit to quality control NGS datasets and genome assemblies ed. B.
1206 Berger. *Bioinformatics* **33**: 574–576. doi:10.1093/bioinformatics/btw663.

1207 Mapleson D, Venturini L, Kaithakottil G, Swarbreck D. 2018. Efficient and accurate detection
1208 of splice junctions from RNA-seq with Portcullis. *GigaScience* **7**.
1209 doi:10.1093/gigascience/giy131.

1210 McCarthy CGP, Fitzpatrick DA. 2019. Pan-genome analyses of model fungal species.
1211 *Microb Genom* **5**. doi:10.1099/mgen.0.000243.

1212 McDonald AG, Boyce S, Tipton KF. 2009. ExplorEnz: the primary source of the IUBMB
1213 enzyme list. *Nucleic Acids Res* **37**: D593–D597. doi:10.1093/nar/gkn582.

McDonald MC, Taranto AP, Hill E, Schwessinger B, Liu Z, Simpfendorfer S, Milgate A, Solomon PS. 2019. Transposon-Mediated Horizontal Transfer of the Host-Specific Virulence Protein ToxA between Three Fungal Wheat Pathogens. *mBio* **10**: e01515-19. doi:10.1128/mBio.01515-19.

Mesny F, Miyauchi S, Thiergart T, Pickel B, Atanasova L, Karlsson M, Hüttel B, Barry KW, Haridas S, Chen C, et al. 2021. Genetic determinants of endophytism in the *Arabidopsis* root mycobiome. *Nat Commun* **12**: 7227. doi:10.1038/s41467-021-27479-y.

Mesny F, Vannier N. 2020. Detecting the effect of biological categories on genome composition. https://github.com/fantin-mesny/Effect-Of-Biological-Categories-On-Genomes-Composition.

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**: i142–i150. doi:10.1093/bioinformatics/bty266.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* **41**: e121. doi:10.1093/nar/gkt263.

Miyauchi S, Kiss E, Kuo A, Drula E, Kohler A, Sánchez-García M, Morin E, Andreopoulos B, Barry KW, Bonito G, et al. 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. *Nat Commun* **11**: 5125. doi:10.1038/s41467-020-18795-w.

Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, et al. 2020. A computational framework to explore large-scale biosynthetic diversity. *Nat Chem Biol* **16**: 60–68. doi:10.1038/s41589-019-0400-9.

Okagaki LH, Nunes CC, Sailsbery J, Clay B, Brown D, John T, Oh Y, Young N, Fitzgerald M, Haas BJ, et al. 2015. Genome Sequences of Three Phytopathogenic Species of the Magnaporthaceae Family of Fungi. *G3* **5**: 2539–2545. doi:10.1534/g3.115.020057.

Okagaki LH, Sailsbery JK, Eyre AW, Dean RA. 2016. Comparative genome analysis and genome evolution of members of the magnaporthaceae family of fungi. *BMC Genomics* **17**: 135. doi:10.1186/s12864-016-2491-y.

Osborne S-J, McMillan VE, White R, Hammond-Kosack KE. 2018. Elite UK winter wheat cultivars differ in their ability to support the colonization of beneficial root-infecting fungi. *J Exp Bot* **69**: 3103–3115. doi:10.1093/jxb/ery136.

Osbourn AE, Clarke BR, Dow JM, Daniels MJ. 1991. Partial characterization of avenacinase from *Gaeumannomyces graminis* var. *avenae*. *PMPP* **38**: 301–312. doi:10.1016/S0885-5765(05)80121-3.

Osbourn AE, Clarke BR, Lunness P, Scott PR, Daniels MJ. 1994. An oat species lacking avenacin is susceptible to infection by *Gaeumannomyces graminis* var. *tritici*. *PMPP* **45**: 457–467. doi:10.1016/S0885-5765(05)80042-6.

Palma-Guerrero J, Chancellor T, Spong J, Canning G, Hammond J, McMillan VE, Hammond-Kosack KE. 2021. Take-All Disease: New Insights into an Important

Wheat Root Pathogen. *Trends Plant Sci* **26**: 836–848. doi:10.1016/j.tplants.2021.02.009.

Paradis E, Schliep K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528. doi:10.1093/bioinformatics/bty633.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2009. How Many Bootstrap Replicates Are Necessary? In *13th Annual International Conference on Research in Computational Molecular Biology* (ed. S. Batzoglou), pp. 184–200, Springer-Verlag Berlin Heidelberg, Tucson, AZ, USA doi:10.1089/cmb.2009.0179.

Pedersen TL. 2021. ggforce: Accelerating “ggplot2.” https://cran.r-project.org/web/packages/ggforce/index.html.

Pederson TL. 2022. patchwork: The Composer of Plots. https://CRAN.R-project.org/package=patchwork.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785–786. doi:10.1038/nmeth.1701.

Pilgeram AL, Henson JM. 1992. Sexual crosses of the homothallic fungus *Gaeumannomyces graminis* var. *tritici* based on use of an auxotroph obtained by transformation. *Exp Mycol* **16**: 35–43. doi:10.1016/0147-5975(92)90039-T.

Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol* **16**: 5. doi:10.1186/s12915-017-0457-4.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033.

R Core Team. 2022. R: A language and environment for statistical computing. https://www.r-project.org/.

Raaijmakers JM, Paulitz TC, Steinberg C, Alabouvette C, Moënne-Loccoz Y. 2009. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant Soil* **321**: 341–361. doi:10.1007/s11104-008-9568-6.

Sauder DC, DeMars CE. 2019. An Updated Recommendation for Multiple Comparisons. *Adv Meth Pract Psychol Sci* **2**: 26–44. doi:10.1177/2515245918808784.

Savojardo C, Martelli PL, Fariselli P, Casadio R. 2018. DeepSig: deep learning improves signal peptide detection in proteins. *Bioinformatics* **34**: 1690–1696. doi:10.1093/bioinformatics/btx818.

Shao M, Kingsford C. 2017. Accurate assembly of transcripts through phase-preserving graph decomposition. *Nat Biotechnol* **35**: 1167–1169. doi:10.1038/nbt.4020.

Shumate A, Salzberg SL. 2021. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**: 1639–1643. doi:10.1093/bioinformatics/btaa1016.

Siozios S. 2021. SioStef/panplots: A small R script for generating pangenome accumulation curves. https://github.com/SioStef/panplots.

1296 Slowikowski K. 2020. ggrepel: Automatically Position Non-Overlapping Text Labels with
1297 “ggplot2.” https://cran.r-project.org/package=ggrepel.

1298 Smit A, Hubley R. 2015. RepeatModeler Open-1.0. http://www.repeatmasker.org.

1299 Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. http://www.repeatmasker.org.

1300 Sperr E. 2023. *Magnaporthe oryzae* hits per 100,000 citations in PubMed. *PubMed by Year*.
1301 https://esperr.github.io/pubmed-by-
1302 year/?q1=magnaporthe%20oryzae&q2=pyricularia%20oryzae.

1303 Sperschneider J, Dodds PN. 2021. EffectorP 3.0: prediction of apoplastic and cytoplasmic
1304 effectors in fungi and oomycetes. *MPMI* **35**: 146–156. doi:10.1094/MPMI-08-21-
1305 0201-R.

1306 Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction
1307 of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol* **19**:
1308 2094–2110. doi:10.1111/mpp.12682.

1309 Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM,
1310 Taylor JM. 2016. EffectorP: predicting fungal effector proteins from secretomes using
1311 machine learning. *New Phytol* **210**: 743–761. doi:10.1111/nph.13794.

1312 Stajich JE. 2023. fungaltools RIP scripts.
1313 https://github.com/hyphaltip/fungaltools/tree/master/scripts.

1314 Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab*
1315 *initio* prediction of alternative transcripts. *Nucleic Acids Res* **34**: W435–W439.
1316 doi:10.1093/nar/gkl200.

1317 Sun S, Lin X, Coelho MA, Heitman J. 2019. Mating-System Evolution: All Roads Lead to
1318 Selfing. *Curr Biol* **29**: R738–R761. doi:10.1016/j.cub.2019.06.073.

1319 Tamasloukht M, Séjalon-Delmas N, Kluever A, Jauneau A, Roux C, Bécard G, Franken P.
1320 2003. Root Factors Induce Mitochondrial-Related Gene Expression and Fungal
1321 Respiration during the Developmental Switch from Asymbiosis to Presymbiosis in the
1322 Arbuscular Mycorrhizal Fungus *Gigaspora rosea*. *Plant Physiol* **131**: 1468–1478.
1323 doi:10.1104/pp.012898.

1324 Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther
1325 O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of
1326 signal peptides using protein language models. *Nat Biotechnol* **40**: 1023–1025.
1327 doi:10.1038/s41587-021-01156-3.

1328 Thynne E, McDonald MC, Solomon PS. 2017. Transition from heterothallism to
1329 homothallism is hypothesised to have facilitated speciation among emerging
1330 *Botryosphaeriaceae* wheat-pathogens. *Fungal Genet Biol* **109**: 36–45.
1331 doi:10.1016/j.fgb.2017.10.005.

1332 Torres DE, Oggenfuss U, Croll D, Seidl MF. 2020. Genome evolution in fungal plant
1333 pathogens: looking beyond the two-speed genome model. *Fungal Biol Rev* **34**: 136–
1334 143. doi:10.1016/j.fbr.2020.07.001.

1335 Traeger S, Altegoer F, Freitag M, Gabaldon T, Kempken F, Kumar A, Marcet-Houben M,
1336 Pöggeler S, Stajich JE, Nowrousian M. 2013. The Genome and Development-

1337 Dependent Transcriptomes of *Pyronema confluens*: A Window into Fungal Evolution.
1338 *PLoS Genet* **9**: e1003820. doi:10.1371/journal.pgen.1003820.

1339 Tralamazza SM, Gluck-Thaler E, Feurtey A, Croll D. 2023.Copy number variation introduced
1340 by a massive mobile element underpins global thermal adaptation in a fungal wheat
1341 pathogen. *bioRxiv [Preprint]*. doi:10.1101/2023.09.22.559077.

1342 Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Darwin Tree of Life Consortium,
1343 Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M, et al. 2023.
1344 MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high
1345 fidelity reads. *BMC Bioinformatics* **24**: 288. doi:10.1186/s12859-023-05385-y.

1346 Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, De Silva N, Martinez
1347 MC, Pedro H, Yates AD, et al. 2020. PHI-base: the pathogen-host interactions
1348 database. *Nucleic Acids Res* **48**: D613–D620. doi:10.1093/nar/gkz904.

1349 Urquhart AS, Chong NF, Yang Y, Idnurm A. 2022. A large transposable element mediates
1350 metal resistance in the fungus *Paecilomyces variotii*. *Curr Biol* **32**: 937–950.
1351 doi:10.1016/j.cub.2021.12.048.

1352 Urquhart AS, Gluck-Thaler E, Vogan AA. 2023a.Gene acquisition by giant transposons
1353 primes eukaryotes for rapid evolution via horizontal gene transfer. *bioRxiv [Preprint]*.
1354 doi:10.1101/2023.11.22.568313.

1355 Urquhart AS, Vogan AA, Gardiner DM, Idnurm A. 2023b. *Starships* are active eukaryotic
1356 transposable elements mobilized by a new family of tyrosine recombinases. *Proc*
1357 *Natl Acad Sci USA* **120**: e2214521120. doi:10.1073/pnas.2214521120.

1358 van den Brand T. 2023. ggh4x: Hacks for “ggplot2.” https://CRAN.R-
1359 project.org/package=ggh4x.

1360 van der Heijden MGA, Bardgett RD, van Straalen NM. 2008. The unseen majority: soil
1361 microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol*
1362 *Letters* **11**: 296–310. doi:10.1111/j.1461-0248.2007.01139.x.

1363 Van Wees SC, Van der Ent S, Pieterse CM. 2008. Plant immune responses triggered by
1364 beneficial microbes. *Curr Opin Plant Biol* **11**: 443–448.
1365 doi:10.1016/j.pbi.2008.05.005.

1366 Van Wyk S, Wingfield B, De Vos L, van der Merwe N, Santana Q, Steenkamp E. 2019.
1367 Repeat-Induced Point Mutations Drive Divergence between *Fusarium circinatum* and
1368 Its Close Relatives. *Pathogens* **8**: 298. doi:10.3390/pathogens8040298.

1369 Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, Vert J-P, Noble WS.
1370 2015. Accurate identification of centromere locations in yeast genomes using Hi-C.
1371 *Nucleic Acids Res* **43**: 5331–5339. doi:10.1093/nar/gkv424.

1372 Venturini L, Caim S, Kaithakottil GG, Mapleson DL, Swarbreck D. 2018. Leveraging multiple
1373 transcriptome assembly methods for improved gene structure annotation.
1374 *GigaScience* **7**. doi:10.1093/gigascience/giy093.

1375 Venturini L, Yanes L. 2020. ei-liftover. https://github.com/lucventurini/ei-liftover.

1376 Vogan AA, Ament-Velásquez SL, Bastiaans E, Wallerman O, Saupe SJ, Suh A,
1377 Johannesson H. 2021. The *Enterprise*, a massive transposon carrying *Spok* meiotic
1378 drive genes. *Genome Res* **31**: 789–798. doi:10.1101/gr.267609.120.

1379 Vogan AA, Ament-Velásquez SL, Granger-Farbos A, Svedberg J, Bastiaans E, Debets AJ,
1380 Coustou V, Yvanne H, Clavé C, Saupe SJ, et al. 2019. Combinations of *Spok* genes
1381 create multiple meiotic drivers in *Podospora*. *eLife* **8**: e46454.
1382 doi:10.7554/eLife.46454.

1383 Wacker T, Helmstetter N, Wilson D, Fisher MC, Studholme DJ, Farrer RA. 2023. Two-speed
1384 genome evolution drives pathogenicity in fungal pathogens of animals. *Proc Natl*
1385 *Acad Sci USA* **120**: e2212633120. doi:10.1073/pnas.2212633120.

1386 Wang Y, Li J, Xiang S, Zhou J, Peng X, Hai Y, Wang Y, Li S, Wei S. 2020. A putative
1387 effector UvHrip1 inhibits BAX-triggered cell death in *Nicotiana benthamiana* , and
1388 infection of *Ustilaginoidea virens* suppresses defense-related genes expression.
1389 *PeerJ* **8**: e9354. doi:10.7717/peerj.9354.

1390 Wapinski I, Pfeffer A, Friedman N, Regev A. 2007. Natural history and evolutionary
1391 principles of gene duplication in fungi. *Nature* **449**: 54–61. doi:10.1038/nature06107.

1392 Weisberg AJ, Grünwald NJ, Savory EA, Putnam ML, Chang JH. 2021. Genomic Approaches
1393 to Plant-Pathogen Epidemiology and Diagnostics. *Annu Rev Phytopathol* **59**: 311–
1394 332. doi:10.1146/annurev-phyto-020620-121736.

1395 Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org.

1396 Wickham H. 2020. rvest: Easily Harvest (Scrape) Web Pages. https://cran.r-
1397 project.org/package=rvest.

1398 Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes
1399 A, Henry L, Hester J, et al. 2019. Welcome to the Tidyverse. *Journal of Open Source*
1400 *Software* **4**: 1686. doi:10.21105/joss.01686.

1401 Wickham H, Seidel D. 2020. scales: Scale Functions for Visualization. https://cran.r-
1402 project.org/package=scales.

1403 Wilke CO. 2020. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.”
1404 https://cran.r-project.org/package=cowplot.

1405 Wilken PM, Steenkamp ET, Wingfield MJ, De Beer ZW, Wingfield BD. 2017. Which *MAT*
1406 gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature reconsidered.
1407 *Fungal Biol Rev* **31**: 199–211. doi:10.1016/j.fbr.2017.05.003.

1408 Xu J, Zhang H, Zheng J, Dovoedo P, Yin Y. 2020. eCAMI: simultaneous classification and
1409 motif identification for enzyme annotation. *Bioinformatics* **36**: 2068–2075.
1410 doi:10.1093/bioinformatics/btz908.

1411 Xu X-H, Su Z-Z, Wang C, Kubicek CP, Feng X-X, Mao L-J, Wang J-Y, Chen C, Lin F-C,
1412 Zhang C-L. 2014. The rice endophyte *Harpophora oryzae* genome reveals evolution
1413 from a pathogen to a mutualistic endophyte. *Sci Rep* **4**: 5783.
1414 doi:10.1038/srep05783.

1415 Yadav V, Sun S, Coelho MA, Heitman J. 2020. Centromere scission drives chromosome
1416 shuffling and reproductive isolation. *Proc Natl Acad Sci USA* **117**: 7917–7928.
1417 doi:10.1073/pnas.1918659117.

1418 Yadav V, Yang F, Reza MdH, Liu S, Valent B, Sanyal K, Naqvi NI. 2019. Cellular Dynamics
1419 and Genomic Identity of Centromeres in Cereal Blast Fungus. *mBio* **10**: e01581-19.
1420 doi:10.1128/mBio.01581-19.

1421 Yildirir G, Sperschneider J, Malar C M, Chen ECH, Iwasaki W, Cornell C, Corradi N. 2022.
1422 Long reads and Hi-C sequencing illuminate the two-compartment genome of the
1423 model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytol* **233**:
1424 1097–1107. doi:10.1111/nph.17842.

1425 Yu G. 2021. ggplotify: Convert Plot to “grob” or “ggplot” Object. https://cran.r-
1426 project.org/package=ggplotify.

1427 Yu G. 2023. seqmagick: Sequence Manipulation Utilities. https://CRAN.R-
1428 project.org/package=seqmagick.

1429 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. GGTREE: an R package for visualization
1430 and annotation of phylogenetic trees with their covariates and other associated data.
1431 *Methods Ecol Evol* **8**: 28–36. doi:10.1111/2041-210X.12628.

1432 Yu G, Xu S, Hackl T. 2023. aplot: Decorate a “ggplot” with Associated Information.
1433 https://github.com/YuLab-SMU/aplot.

1434 Zamioudis C, Pieterse CMJ. 2012. Modulation of Host Immunity by Beneficial Microbes.
1435 *MPMI* **25**: 139–150. doi:10.1094/MPMI-06-11-0179.

1436 Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018.
1437 dbCAN2: a meta server for automated carbohydrate-active enzyme annotation.
1438 *Nucleic Acids Res* **46**: W95–W101. doi:10.1093/nar/gky418.

1439 Zhang N, Luo J, Rossman AY, Aoki T, Chuma I, Crous PW, Dean R, de Vries RP, Donofrio
1440 N, Hyde KD, et al. 2016. Generic names in *Magnaporthales*. *IMA Fungus* **7**: 155–
1441 159. doi:10.5598/imafungus.2016.07.01.09.

1442 Zhao Z, Liu H, Wang C, Xu JR. 2013. Comparative analysis of fungal genomes reveals
1443 different plant cell wall degrading capacity in fungi. *BMC Genomics* **14**.
1444 doi:10.1186/1471-2164-15-6.

1445 Zhou L, Feng T, Xu S, Gao F, Lam TT, Wang Q, Wu T, Huang H, Zhan L, Li L, et al. 2022.
1446 ggmsa: a visual exploration tool for multiple sequence alignment and associated
1447 data. *Brief Bioinform* **23**: bbac222. doi:10.1093/bib/bbac222.

1448

1449 **Figure 1** Intraspecific variation in *Gaeumannomyces tritici* (*Gt)* virulence assessed from

1450 inoculation of wheat plants. Representative photos of wheat roots (**a**) and above-ground

1451 features (**b**) following inoculation treatment. Inoculated strains from top left to bottom right:

1452 no *Gt* (control), Gt-8d, Gt-19d1, Gt-23d, Gt-4e and Gt-LH10. **c** Box and violin plots showing

1453 the impact of the five *Gt* strains sequenced in this study on above- and below-ground
1454 characteristics in winter wheat. Control, *Gt* type A and type B groups are indicated by
1455 different colours. Strains with a significant mean difference for the characteristic as
1456 calculated by either the Tukey HSD or Games-Howell test are shown by letters above the
1457 plots.

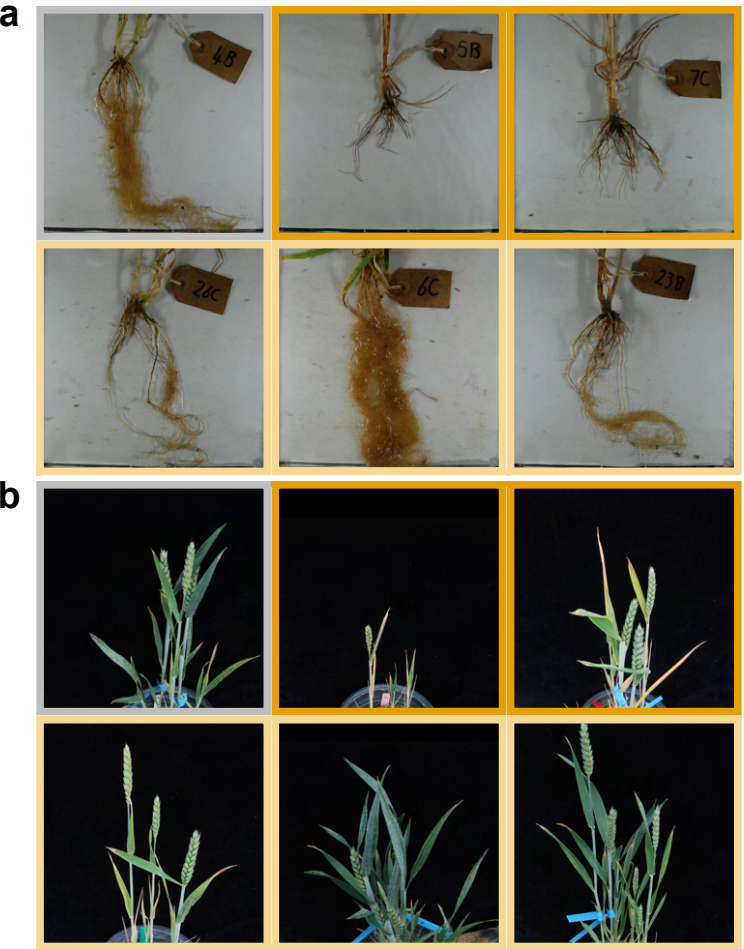
1458 **Figure 2** GENESPACE plot showing synteny across the nine *Gaeumannomyces* strains.
1459 A/B lineages are indicated for *G. tritici* strains. Fragments are labelled with numbers
1460 corresponding to pseudochromosomes, and an asterisk indicates that a fragment was
1461 inverted in the visualisation. Black bars on the ends of fragments indicate telomeres
1462 predicted using Tapestry.

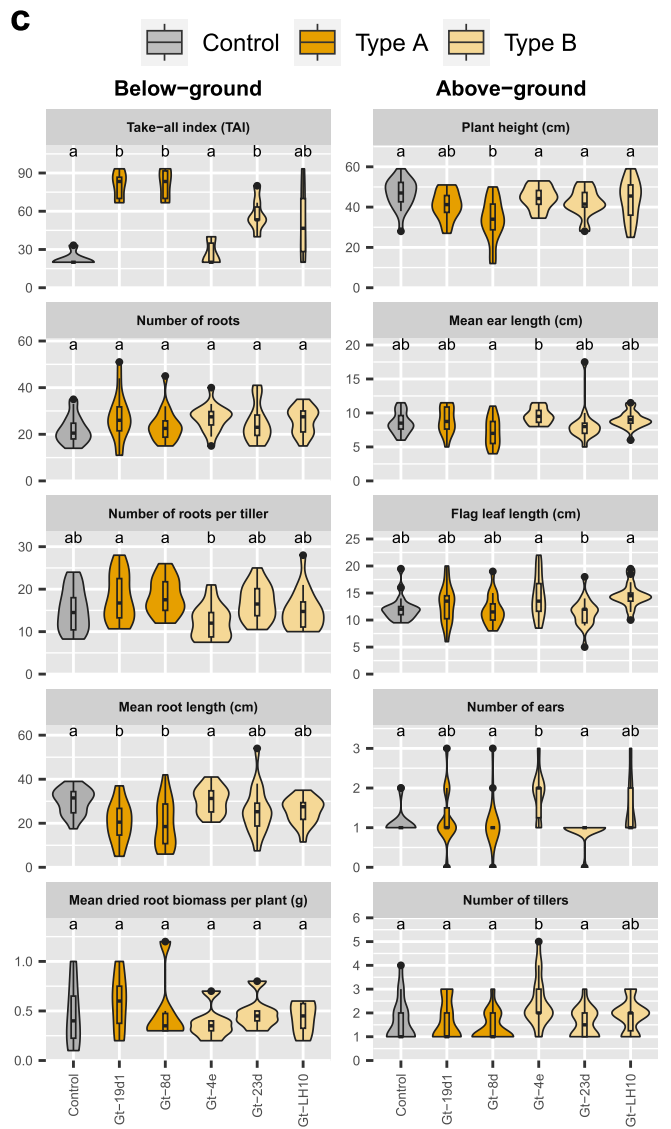
1463 **Figure 3** The relationship between candidate secreted effector proteins (CSEPs) and
1464 transposable elements (TEs) in *Gaeumannomyces*. **a** TE density (per 100,000 bp) and the
1465 location of CSEPs (black ticks) across fragments. Fragments are ordered syntenically
1466 according to GENESPACE (Fig. 2). **b** Scatterplot showing the relationship between CSEP
1467 density versus TE and gene density (per 100,000 bp) with local polynomial regression lines
1468 (ggplot2 function geom_smooth, method = “loess”). Significant correlation is indicated with
1469 Kendall’s tau (τ). **c** Box and violin plots showing the distance of genes to the closest TE, with
1470 CSEPs and other genes distinguished by colour. An asterisk indicates where a Wilcoxon
1471 rank sum test found the mean TE distance to be significantly different for CSEPs versus
1472 other genes. Strains with a significant mean difference in overall gene-TE distance as
1473 calculated by the Games-Howell test are shown by letters above the plots.

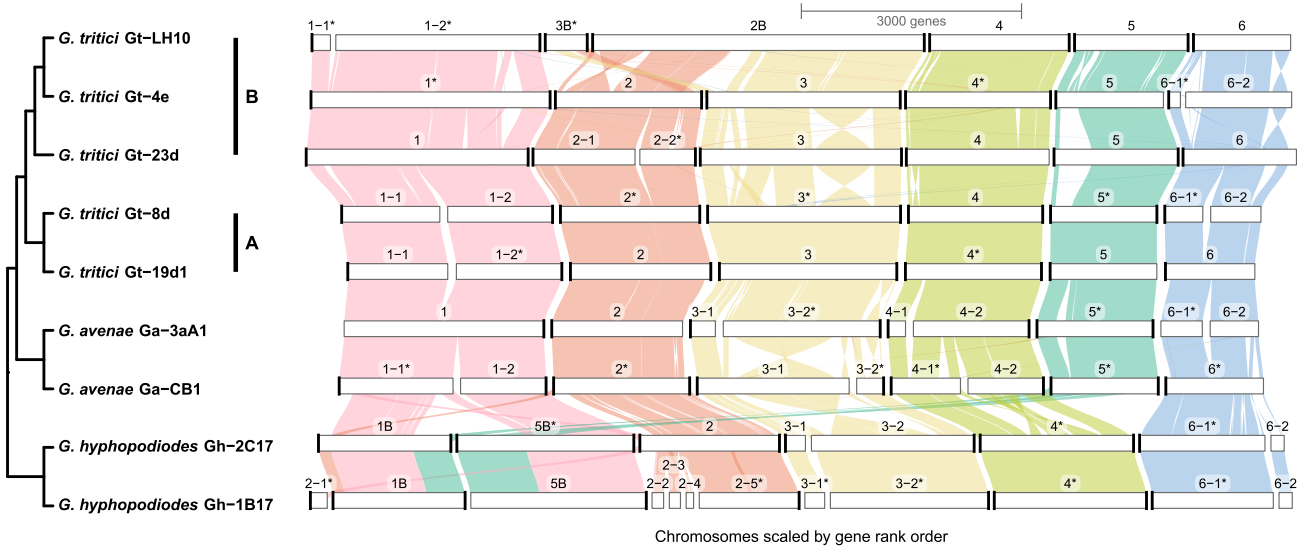
1474 **Figure 4** Summary of predicted gene content for the *Gaeumannomyces* strains reported in
1475 this study. **a** Number of total genes, candidate secreted effector proteins (CSEPs),
1476 carbohydrate-active enzymes (CAZymes) and biosynthetic gene clusters (BGCs) for each
1477 *Gaeumannomyces* strain. The A/B lineages are indicated for *Gaeumannomyces tritici* (*Gt*)
1478 strains. The dashed line in the phylogeny indicates bootstrap support <70 found within the *Gt*B
1479 lineage (see Supplemental Fig. S13b for the full genome-scale *Gaeumannomyces* species

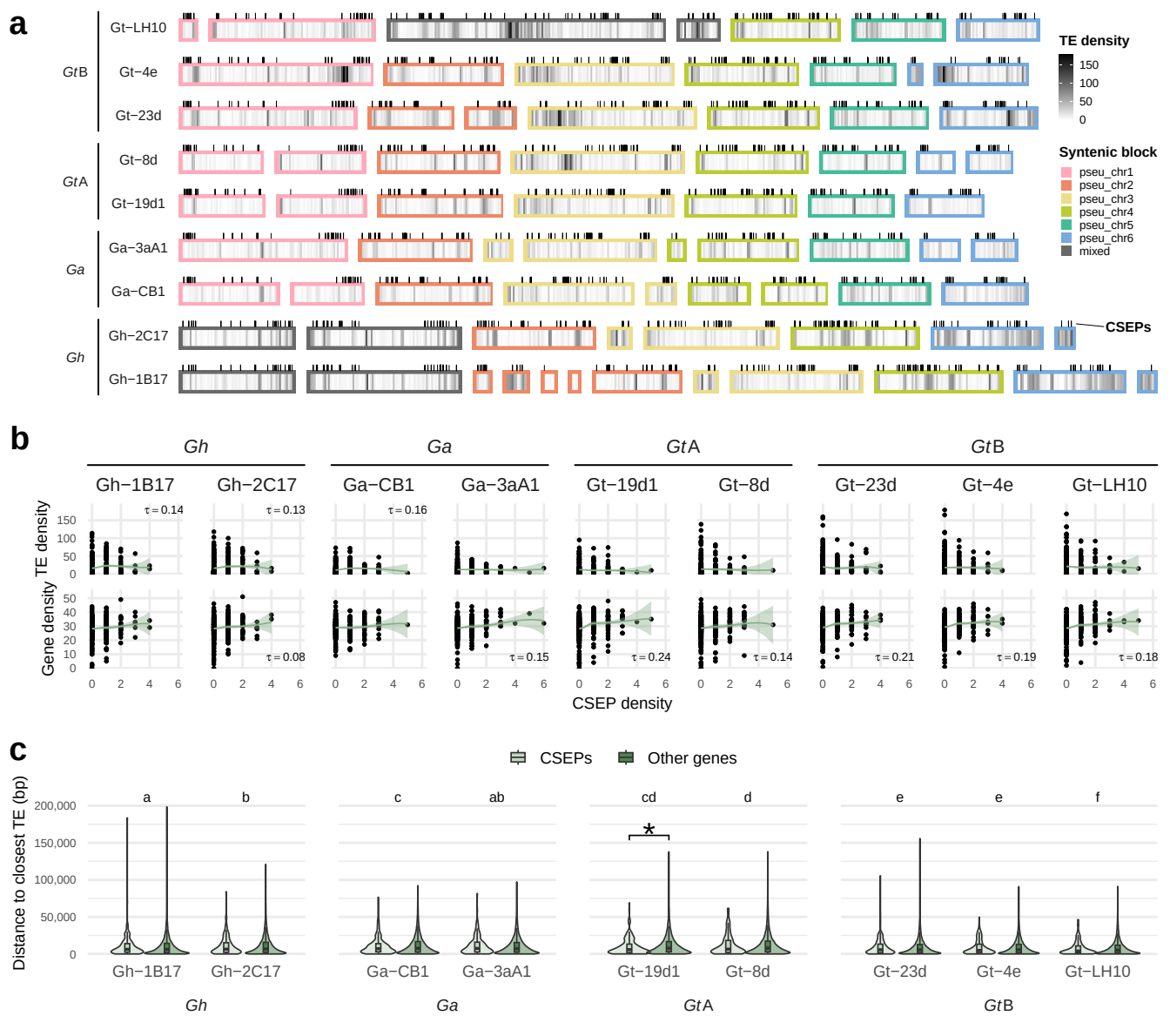
tree). The *Gt* pangenome (within dashed box) is categorised as core (present in all strains), soft-core (present in all but one strain), accessory (present in at least two strains) and specific (present in one strain). The lefthand inset box shows the results of PERMANOVA statistical tests which calculate the descriptive power of relatedness (phylogeny) versus lifestyle categorisation (*Gt* and *G. avenae* as pathogenic in wheat, *G. hyphopodioides* as non-pathogenic) on gene variance. Gene copy-number is shown on a scatterplot to the right, with points jittered vertically to improve visualisation. **b** Accumulation curves of pan and core genes for the *Gt* pangenome (Siozios 2021). **c** Euler diagram summarising whether high copy-number genes in each lineage are present but in low copy-number in *Gt*A, or completely absent.

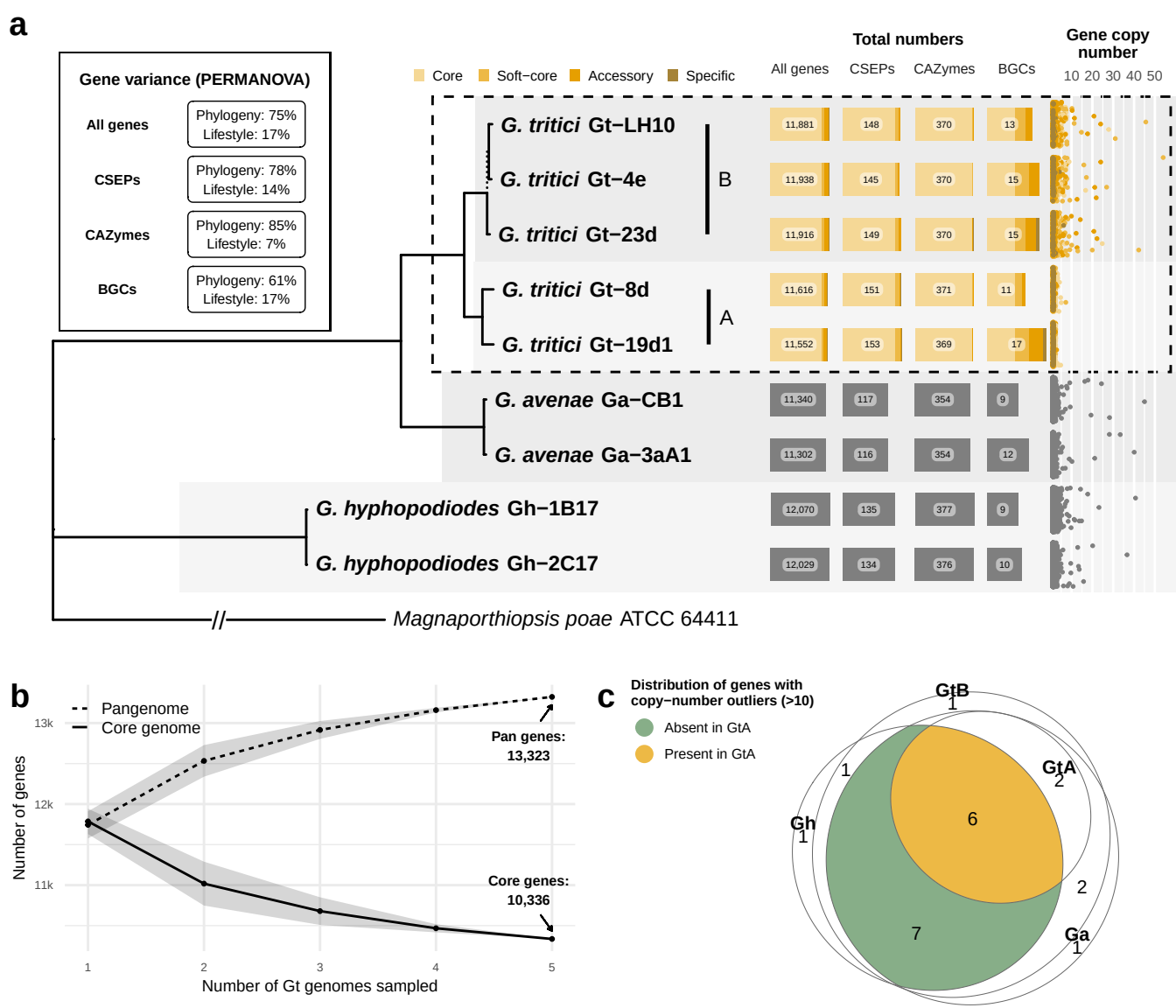
Figure 5 *Gaeumannomyces* genomes contain *Starship* giant transposable elements. **a** Gene tree of *Starship* 'captain' genes, including captains and other tyrosine recombinases identified from our assemblies via starfish, captain homologues identified via blastp, and previously published captain genes. **b** A summary of the *Starship* elements identified by starfish with the composite RIP index (CRI) shown above each element. The yellow highlight distinguishes a nested element. cap=captain gene, DR=direct repeat, RIP=repeat-induced point mutation, TE=transposable element gene, TIR=terminal inverted repeat, tyr=tyrosine recombinase gene.

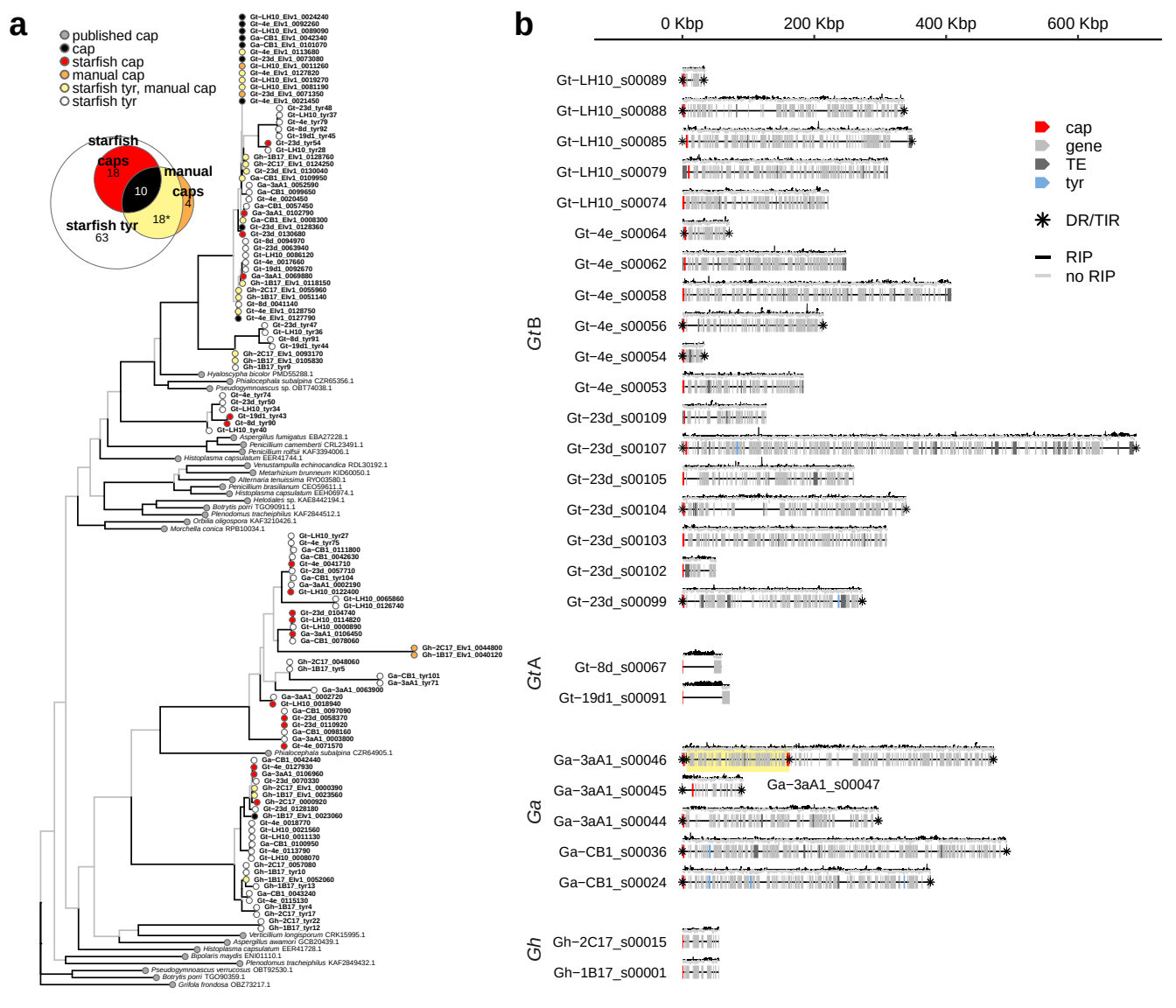

a
b
4B
5B
7C
26C
6C
23B


c
Control
Type A
Type B
Below−ground
Above−ground
Take−all index (TAI)
Plant height (cm)
Number of roots
Mean ear length (cm)
Number of roots per tiller
Flag leaf length (cm)
Mean root length (cm)
Number of ears
Mean dried root biomass per plant (g)
Number of tillers
Control
Gt−19d1
Gt−8d
Gt−4e
Gt−23d
Gt−LH10


G. tritici Gt−LH10
G. tritici Gt−4e
G. tritici Gt−23d
G. tritici Gt−8d
G. tritici Gt−19d1
G. avenae Ga−3aA1
G. avenae Ga−CB1
G. hyphopodiodes Gh−2C17
G. hyphopodiodes Gh−1B17
B
A
3000 genes
Chromosomes scaled by gene rank order


a
GtB
Gt−LH10
Gt−4e
Gt−23d
GtA
Gt−8d
Gt−19d1
Ga
Ga−3aA1
Ga−CB1
Gh
Gh−2C17
Gh−1B17
CSEPs
TE density
150
100
50
0
Syntenic block
pseu_chr1
pseu_chr2
pseu_chr3
pseu_chr4
pseu_chr5
pseu_chr6
mixed
b
Gh
Ga
GtA
GtB
Gh−1B17
Gh−2C17
Ga−CB1
Ga−3aA1
Gt−19d1
Gt−8d
Gt−23d
Gt−4e
Gt−LH10
TE density
Gene density
τ = 0.14
τ = 0.13
τ = 0.16
τ = 0.08
τ = 0.15
τ = 0.24
τ = 0.14
τ = 0.21
τ = 0.19
τ = 0.18
CSEP density
c
CSEPs
Other genes
Distance to closest TE (bp)
200,000
150,000
100,000
50,000
0
a
b
c
ab
cd
d
e
e
f
*
Gh−1B17
Gh−2C17
Ga−CB1
Ga−3aA1
Gt−19d1
Gt−8d
Gt−23d
Gt−4e
Gt−LH10
Gh
Ga
GtA
GtB


a
Gene variance (PERMANOVA)
All genes
Phylogeny: 75%
Lifestyle: 17%
CSEPs
Phylogeny: 78%
Lifestyle: 14%
CAZymes
Phylogeny: 85%
Lifestyle: 7%
BGCs
Phylogeny: 61%
Lifestyle: 17%
Core
Soft−core
Accessory
Specific
Total numbers
All genes
CSEPs
CAZymes
BGCs
Gene copy number
10 20 30 40 50
G. tritici Gt−LH10
G. tritici Gt−4e
G. tritici Gt−23d
B
11,881
148
370
13
11,938
145
370
15
11,916
149
370
15
G. tritici Gt−8d
G. tritici Gt−19d1
A
11,616
151
371
11
11,552
153
369
17
G. avenae Ga−CB1
G. avenae Ga−3aA1
11,340
117
354
9
11,302
116
354
12
G. hyphopodiodes Gh−1B17
G. hyphopodiodes Gh−2C17
12,070
135
377
9
12,029
134
376
10
Magnaporthiopsis poae ATCC 64411
b
Pangenome
Core genome
13k
12k
11k
Number of genes
Pan genes:
13,323
Core genes:
10,336
1
2
3
4
5
Number of Gt genomes sampled
c
Distribution of genes with copy−number outliers (>10)
Absent in GtA
Present in GtA
GtB
1
GtA
2
Gh
1
1
6
2
7
Ga
1


a
published cap
cap
starfish cap
manual cap
starfish tyr, manual cap
starfish tyr
starfish caps 18
10
manual caps 4
18*
starfish tyr 63
Gt-LH10_Elv1_0024240
Gt-4e_Elv1_0092260
Gt-LH10_Elv1_0089090
Ga-CB1_Elv1_0042340
Ga-CB1_Elv1_0101070
Gt-4e_Elv1_0113680
Gt-23d_Elv1_0073080
Gt-LH10_Elv1_0011260
Gt-4e_Elv1_0127820
Gt-LH10_Elv1_0019270
Gt-LH10_Elv1_0081190
Gt-23d_Elv1_0071350
Gt-4e_Elv1_0021450
Gt-23d_tyr48
Gt-LH10_tyr37
Gt-4e_tyr79
Gt-8d_tyr92
Gt-19d1_tyr45
Gt-23d_tyr54
Gt-LH10_tyr28
Gh-1B17_Elv1_0128760
Gh-2C17_Elv1_0124250
Gt-23d_Elv1_0130040
Ga-CB1_Elv1_0109950
Ga-3aA1_0052590
Ga-CB1_0099650
Gt-4e_0020450
Ga-CB1_0067450
Ga-3aA1_0102790
Ga-CB1_Elv1_0008300
Gt-23d_Elv1_0128360
Gt-23d_0130680
Gt-8d_0094970
Gt-23d_0063940
Gt-LH10_0086120
Gt-4e_0017660
Gt-19d1_0092670
Ga-3aA1_0069880
Gh-1B17_Elv1_0118150
Gh-2C17_Elv1_0055960
Gh-1B17_Elv1_0051140
Gt-8d_0041140
Gt-4e_Elv1_0128750
Gt-4e_Elv1_0127790
Gt-23d_tyr47
Gt-LH10_tyr36
Gt-8d_tyr91
Gt-19d1_tyr44
Gh-2C17_Elv1_0093170
Gh-1B17_Elv1_0105830
Gh-1B17_tyr9
Hyaloscypha bicolor PMD55288.1
Phialocephala subalpina CZR65356.1
Pseudogymnoascus sp. OBT74038.1
Gt-4e_tyr74
Gt-23d_tyr50
Gt-LH10_tyr34
Gt-19d1_tyr43
Gt-8d_tyr90
Gt-LH10_tyr40
Aspergillus fumigatus EBA27228.1
Penicillium camemberti CRL23491.1
Penicillium roftsii KAF3940061.1
Histoplasma capsulatum EER41744.1
Verustampulla echinocandica RDL30192.1
Metarhizium brunneum KID60050.1
Alternaria tenuissima RYO03580.1
Penicillium brasilianum CEO59611.1
Histoplasma capsulatum EEH06974.1
Helotiales sp. KAE8442194.1
Botrytis porri TGO90911.1
Plenodomus tracheiphilus KAF2844512.1
Orbilia oligospora KAF3210426.1
Morchella conica RPB10034.1
Gt-LH10_tyr27
Gt-4e_tyr75
Ga-CB1_0111800
Ga-CB1_0042630
Gt-4e_0041710
Gt-23d_0057710
Ga-CB1_tyr104
Ga-3aA1_0002190
Gt-LH10_0122400
Gt-LH10_0065860
Gt-LH10_0126740
Gt-23d_0104740
Gt-LH10_0114820
Gt-LH10_0000890
Ga-3aA1_0106450
Ga-CB1_0078060
Gh-2C17_Elv1_0044800
Gh-1B17_Elv1_0040120
Gh-2C17_0048060
Gh-1B17_tyr5
Ga-CB1_tyr101
Ga-3aA1_tyr71
Ga-3aA1_0063900
Ga-3aA1_0002720
Gt-LH10_0018940
Ga-CB1_0097090
Gt-23d_0058370
Gt-23d_0110920
Ga-CB1_0098160
Ga-3aA1_0003800
Gt-4e_0071570
Phialocephala subalpina CZR64905.1
Ga-CB1_0042440
Gt-4e_0127930
Ga-3aA1_0106960
Gt-23d_0070330
Gh-2C17_Elv1_0000390
Gh-1B17_Elv1_0023560
Gh-2C17_0000920
Gt-23d_0128180
Gh-1B17_Elv1_0023060
Gt-4e_0018770
Gt-LH10_0021560
Gt-LH10_0011130
Ga-CB1_0100950
Gt-4e_0133790
Gt-LH10_0008070
Gh-2C17_0057080
Gh-1B17_tyr10
Gh-1B17_Elv1_0052060
Gh-1B17_tyr13
Ga-CB1_0043240
Gt-4e_0115130
Gh-1B17_tyr4
Gh-2C17_tyr17
Gh-2C17_tyr22
Gh-1B17_tyr12
Verticillium longisporum CRK15995.1
Aspergillus awamori GCB20429.1
Histoplasma capsulatum EER41728.1
Bipolaris maydis ENI01110.1
Plenodomus tracheiphilus KAF2849432.1
Pseudogymnoascus verrucosus OBT92530.1
Botrytis porri TGO90359.1
Grifola frondosa OBZ73217.1
b
0 Kbp
200 Kbp
400 Kbp
600 Kbp
cap
gene
TE
tyr
DR/TIR
RIP
no RIP
GtB
Gt-LH10_s00089
Gt-LH10_s00088
Gt-LH10_s00085
Gt-LH10_s00079
Gt-LH10_s00074
Gt-4e_s00064
Gt-4e_s00062
Gt-4e_s00058
Gt-4e_s00056
Gt-4e_s00054
Gt-4e_s00053
Gt-23d_s00109
Gt-23d_s00107
Gt-23d_s00105
Gt-23d_s00104
Gt-23d_s00103
Gt-23d_s00102
Gt-23d_s00099
GtA
Gt-8d_s00067
Gt-19d1_s00091
Ga
Ga-3aA1_s00046
Ga-3aA1_s00045
Ga-3aA1_s00047
Ga-3aA1_s00044
Ga-CB1_s00036
Ga-CB1_s00024
Gh
Gh-2C17_s00015
Gh-1B17_s00001