

Article

Characterization of Host-Specific Genes from Pine- and Grass-Associated Species of the *Fusarium Fujikuroi* Species Complex

Claudette Dewing ^{1,*}, Magrieta A. van der Nest ^{1,2}, Quentin C. Santana ¹, Robert H. Proctor ³, Brenda D. Wingfield ¹, Emma T. Steenkamp ¹, and Lieschen De Vos ¹

- ¹ Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private bag X20, Pretoria 0028, South Africa; lbahlman@up.ac.za (L.D.V.); emma.steenkamp@up.ac.za (E.T.S.); brenda.wingfield@up.ac.za (B.D.W.); quentin.santana@up.ac.za (Q.C.S.)
- ² Biotechnology Platform, Agricultural Research Council, Onderstepoort, Pretoria 0110, South Africa; VanDerNestM@arc.agric.za (M.v.d.N.)
- ³ United States Department of Agriculture (USDA), Agriculture Research Service, National Center for Agricultural Utilization Research, Mycotoxin Prevention and Applied Microbiology Unit, 1815 N. University St. Peoria, Illinois 61604, USA; robert.proctor@usda.gov (R.H.P.)
- * Correspondence: lbahlman@fabi.up.ac.za

Abstract: The *Fusarium fujikuroi* species complex (FFSC) includes socioeconomically important pathogens that cause disease and/or mycotoxin contamination on numerous crops. Here, we used comparative genomics to elucidate processes underlying the ability of pine-associated and grass-associated FFSC species to colonize tissues of their respective plant hosts. We characterized the identity, possible functions, evolutionary origins, and chromosomal positions of the host-range-associated genes encoded by the two groups of fungi. The 72 and 47 genes identified as unique to the respective genome groups were potentially involved in diverse processes, ranging from transcription, regulation, and substrate transport, through to virulence/pathogenicity. Most emerged early during the evolution of *Fusarium*/FFSC and were subsequently retained only in some lineages, while some had origins outside *Fusarium*. Although differences in the densities of these genes were especially noticeable on the conditionally dispensable chromosome of *F. temperatum* (representing the grass-associates) and *F. circinatum* (representing the pine-associates), the host-range-associated genes tended to be located towards the subtelomeric regions of chromosomes. Taken together, these results demonstrate that multiple mechanisms drive the emergence of genes in grass- and pine-associated FFSC taxa examined and highlighted the diversity of molecular processes potentially underlying niche-specificity in these and other *Fusarium* species.

Keywords: *Fusarium fujikuroi* species complex; *Fusarium circinatum*; *Fusarium temperatum*; pitch canker; comparative genomics; host-specificity; horizontal gene transfer; subtelomeres

1. Introduction

The genus *Fusarium* represents a diverse group of Ascomycetes that are important in industry, agriculture, and medicine [1, 2]. While known species in the genus include pathogens of humans [3-5], animals [6] and insects [7, 8], the vast majority are destructive plant pathogens [9-14]. These pathogenic species affect the quality and quantity of most agriculturally important crops, resulting in significant risks to food security [15].

Fusarium species associate with diverse plant species, ranging from gymnosperms to angiosperms, including many monocots in the grass family, Poaceae [16-18]. Examples of these species are included in the phylogenetically defined lineage of *Fusarium* known as the *Fusarium fujikuroi* species complex (FFSC) [1, 2]. FFSC taxa that associate with gymnosperms (e.g., those in the pine family Pinaceae) include *Fusarium circinatum*, an economically important pathogen of *Pinus* species, as well as *Fusarium fracticaudum* and *Fusarium pininemorale* that colonize the tissue of these plants without any apparent disease symp-

toms [11, 13]. Examples of grass-host-associated species include the prairie grass endophyte *Fusarium konzum* [19] and the maize pathogens *Fusarium subglutinans* and *Fusarium temperatum* [10, 20]. However, these associations are not always strict, as in the case of *F. circinatum*, which can also colonize maize and grasses (from Poaceae) [21–24], various herbaceous dicots (besides Poaceae) [25], as well as certain other non-pine conifers [26, 27]. Despite their diverse host associations, these six species are closely related [13], their genomes are characterized by high levels of synteny [28, 29], and some species can hybridize [29, 30]. The ability of *F. circinatum* and *F. temperatum* to hybridize suggests that since they diverged from a recent common ancestor, they have co-evolved with their respective plant hosts in overlapping geographic ranges [10, 17].

Advances in genomics have enhanced the discovery and analysis of the mechanisms of host-specificity and processes underlying plant-fungus interactions [31]. These data showed that fungal genomes, including those of *Fusarium*, are divided into core and accessory compartments [32, 33]. Core genes typically play an integral part in primary metabolism, reproductive strategies, and development. Accessory genes, on the other hand, are typically non-essential for growth and development but encode for adaptive and niche-associated traits, such as pathogenicity and virulence factors [33]. Among closely related *Fusarium* species, the core compartment is usually conserved, syntenous and gene rich [14], whereas the accessory compartment is gene-poor and often also includes subtelomeric regions and supernumerary/accessory chromosomes [34]. Both of the latter are typically associated with loss of synteny and have been linked to genes promoting the adaptation and survival of pathogens [35].

Previous comparative genomics studies have enhanced our knowledge of how genome evolution affects host-specificity and pathogenicity [36–38]. Also, various evolutionary processes are now known to shape the outcome of the plant-fungus interaction [39–42]. Additionally, genes or genomic regions that are not under selection can be lost or accumulate mutations, which can influence adaptation of fungi to different host environments [43]. However, little information regarding such processes is available for the FFSC, especially those determining their ability to colonize certain plant tissues.

In the current study, therefore, we exploited the publicly available whole genome sequences for FFSC species (NCBI; <https://www.ncbi.nlm.nih.gov/>) to study molecular processes that may underlie the association of these fungi with their plant hosts. For this purpose, we employed a comparative genomics approach using the data available for six closely related FFSC species associated with pine (i.e., *F. circinatum*, *F. fracticaudum* and *F. pininemorale*) and grass (i.e., *F. konzum*, *F. subglutinans* and *F. temperatum*). Our specific aims were to identify genes unique to each group of species, and then characterize the genes in terms of ontology, predicted pathways and mechanisms, chromosomal location, and potential ancestral origin. Our findings contribute to the current understanding of the diversity and evolution of these species.

2. Results

2.1. Genome sequences, assembly completeness and gene prediction

The three genomes for pine-host-associated species were from *F. circinatum* FSP34, *F. fracticaudum* CMW 25245 and *F. pininemorale* CMW 25243. Genomes for grass-host-associated species were from *F. konzum* NRRL 11616, *F. subglutinans* NRRL 66333 and *F. temperatum* CMW 40964.

Analysis with the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool v. 3.0.2 [44] using the “Sordariomyceta” database, showed that all six genomes were 97.3% to 99.1% complete (Supplementary Table S1). These genomes were also similar in size, gene density, and G+C content (Table 1). Comparison of the four genomes assembled into pseudochromosomes (*F. circinatum*, *F. fracticaudum*, *F. pininemorale* and *F. temperatum*) further showed that chromosome size seemed to be conserved throughout Chromosomes 1–11, as they did not differ by more than a factor of 1.00–1.10 (Supplementary Table 2). How-

ever, Chromosome 12 showed extreme chromosome length polymorphism (CLP) and differed by as much as a factor of 2.08, ranging from 0.5 Mb in *F. circinatum* to 1.1 Mb in *F. fructicaudum* (Supplementary Table 2).

Table 1. Genome statistics for the *Fusarium* genomes examined in this study.

<i>Fusarium</i> species	Genome size (bp)	Number of scaffolds	Number of chromosomes	Gene density (ORFs/Mb ¹)	Genes with coding sequences ²	GC content (%)
<i>F. circinatum</i>	45 018 643	49	12	344.50	15 509	47.0
<i>F. fructicaudum</i>	46 252 763	50	12	352.28	16 294	47.6
<i>F. pininemorale</i>	47 778 485	153	12	332.35	15 879	46.0
<i>F. konzum</i>	43 487 959	2 262	N/A	374.49	16 286	48.8
<i>F. subglutinans</i>	44 190 517	733	N/A	356.23	15 742	48.1
<i>F. temperatum</i>	45 458 781	43	12	341.67	15 532	47.0

¹ORFs/Mb = number of open reading frames per million base pairs.

²As determined using MAKER2 [45].

For further analysis, we used the *F. circinatum* genome as a reference for pine-host-associated species and the *F. temperatum* genome as a reference for grass-host-associated species. Analysis of the telomere-associated repeat sequence (TTAGGG/CCCTAA) [46, 47] in the reference genomes indicated that most of the chromosomes had a telomeric cap (Supplementary Table S3). In the *F. circinatum* assembly, seven chromosomes had two telomeric caps and five had only one telomeric cap. In the *F. temperatum* assembly, nine chromosomes had two telomeric caps, two chromosomes had one telomeric cap and one chromosome had no telomeric cap.

2.2. Genes unique to pine- and grass-host-associated species

Analysis with OrthoFinder v. 2.3.1 [48] facilitated identification of 72 and 47 genes that were unique to pine- or grass-host-associated *Fusarium* species, respectively (Supplementary Tables S4 and S5). Multiple paralogs of a few of these genes occurred in some genomes: one paralog of one gene in *F. circinatum* and *F. pininemorale*, two paralogs of two genes in *F. fructicaudum* and *F. subglutinans* and three paralogs of one gene in *F. konzum*. Interestingly, the paralogs of a particular gene were present together on the same chromosome, except in *F. fructicaudum*, where two of these genes were located on different chromosomes. No inferences could be made regarding the chromosomal location of the multiple-copy genes of *F. konzum* and *F. subglutinans* due to the fragmented nature of these assemblies. The phylogenetic relationship between the paralogous genes and the unique gene could be determined for both *F. fructicaudum* paralogs, the one *F. pininemorale* paralog and only one of the two *F. subglutinans* paralogs. The phylogenetic relationship for the remaining *F. subglutinans* paralog, along with the three *F. konzum* paralogs could not be determined as the ancestral origins of these genes were unknown and/or had no significant BLAST hit in the NCBI database (Supplementary Figure S1). All paralogs shared the same ancestral origins with the unique gene, except for *F. fructicaudum*. The paralog of the unique gene was located on a different chromosome and grouped within the ancestral origin group FFSC and FO SC, whereas the unique gene shared an ancestral origin outside *Fusarium* but in the Ascomycetes.

Most of the host-range-associated genes had a gene ontology (GO) description available from Blast2GO [49]. Of the 72 unique *F. circinatum* genes, no GO terms were available for 11 of the genes. Apart from these 11 genes, 26 genes were either hypothetical or predicted to encode uncharacterized proteins. The remaining 35 could be organized into four broad groups based on the proteins they encode, i.e., those involved in virulence, transcriptional regulators, substrate transporters and permeases, proteins potentially involved the metabolism of carbohydrates, fatty acids, and steroids (Supplementary Table

S4). Of the 47 unique *F. temperatum* genes, five did not have a hit in BLAST analysis. Apart from these five genes, 17 were either hypothetical or predicted to encode uncharacterized proteins. The remaining 25 genes could be organized into five broad groups based on the likely functions of their protein products, i.e., those involved in virulence, substrate transporters and permeases, proteins potentially involved the metabolism of carbohydrates, fatty acids, amino acids, and steroids, as well as a HET-domain protein (Supplementary Table S5).

To determine whether certain GO terms in the respective unique sets were significantly enriched relative to the rest of each respective genome, we employed Fisher's Exact test implemented in Blast2GO [49]. In the pine-host-associated set, five GO terms were significantly ($P < 0.05$) enriched in comparison to the whole genome (Supplementary Table S6A), whereas 21 GO terms were significantly ($P < 0.05$) enriched in the grass-host-associated set (Supplementary Table S6B). Of the genes associated with these enriched GO terms, one gene was classified as having a "biologically relevant" role (i.e., glutamate metabolism), and four as having "essential molecular functions", such as coding for a gene product involved in glutamate decarboxylase activity, steroid dehydrogenase activity and RNA-DNA hybrid ribonuclease activity. Furthermore, one gene coded for a gene product involved in cellular components, such as the microtubule and kinesin complex.

We also used previous expression data for *F. circinatum* [50, 51] to search for evidence of expression of the pine-host-associated genes (Supplementary Table 7). The combined results indicated that 62 of the 72 pine-host-associated genes were transcribed under the conditions used to generate the expression data [50, 51]. No expression data are currently available for any of the grass-host-associated *Fusarium* species.

2.3. Phylogenetic origins of the host-range-associated genes

Ancestry of genes in the pine- and grass-host-associated sets were inferred using their protein sequences in BLAST searches against NCBI's non-redundant protein database analyses (performed on 04 October 2021), which were supplemented in some cases by several rounds of alignment of the retrieved sequences followed by phylogenetic analyses. Based on these data, we assigned the 119 genes from the two sets into eight groups (Table 2, Supplementary Figures S2-S8, Supplementary Table S8). Genes included in the first seven groups had evolutionary origins within the FFSC, the broader *Fusarium* genus, or lineages outside *Fusarium* but within the Ascomycetes. The last group contained those genes lacking significant BLAST hits in the NCBI database, as well as those that did not comply with the cut-off values stated in section 4.3.

Table 2. Numbers of genes, and their inferred evolutionary origins, included in the pine- and grass-host-associated sets of host-range-associated genes.

Ancestral Origin Group ¹	Number of genes ²	
	<i>F. circinatum</i>	<i>F. temperatum</i>
1. FFSC	21 (10; 9)	13 (9)
2. FFSC and FOSSC	6 (4; 4)	3 (1)
3. <i>Fusarium</i> but outside FFSC and FOSSC	2 (0; 0)	2 (2)
4. Less than 10 NCBI BLAST hits and mostly <i>Fusarium</i>	10 (7; 5)	7 (5)
5. Less than 10 NCBI BLAST hits and mostly not <i>Fusarium</i>	2 (2; 2)	1 (1)
6. Outside <i>Fusarium</i> , but in the Ascomycetes	9 (7; 7)	5 (2)
7. Outside Fungi	1 (1; 1)	1 (1)
8. Unknown; no significant BLAST hit in the NCBI database	21 (7; 6)	15 (9)

¹ Ancestral origin information from the phylogenetic trees in Supplementary Figures S2-S8. FFSC = *Fusarium fujikuroi* species complex and FOSSC = *Fusarium oxysporum* species complex.

² Number of host-range-associated genes with each type of phylogenetic origin. Pine-host-associated genes are from *F. circinatum*, and grass-host-associated genes are from *F. temperatum*. Values

within parentheses indicate numbers of genes located in subtelomeric regions. For *F. circinatum* the value after the semicolons are numbers of genes for which RNAseq data are available from *F. circinatum* grown on media [50, 51].

The results of these analyses indicated that the ancestral origin for most of the genes were in members of the FFSC or other lineages of *Fusarium*. Indeed, based on their phylogenies, the genes included in groups 1–4 (i.e., 39 genes in the pine-associated set and 25 in the grass-associated set) likely emerged early during the evolution of *Fusarium* and/or the FFSC, after which they were retained in only some lineages. By contrast, 12 pine-host-associated and seven grass-host-associated genes lacked homologs in *Fusarium* and were more closely related to homologs in other Ascomycetes (groups 5 and 6) or outside of the kingdom Fungi (group 7). A considerable proportion of the genes had no detectable homologs in GenBank.

2.4. Genomic distribution of the host-range-associated genes

All *F. circinatum* chromosomes included at least one pine-host-associated gene, while all but two *F. temperatum* chromosomes included at least one grass-host-associated genes (Figures 1 and 2; Supplementary Figures S9 and S10). Chi-squared tests indicated that significantly ($P < 0.05$) more of these genes were located in subtelomeric regions than outside of them (Table 2). That is, about half of the host-associated genes in each species were restricted to subtelomeric regions, which constitute approximately one quarter of each genome (Supplementary Tables S9 and S10). Of the 38 host-range-associated genes in subtelomeric regions of *F. circinatum*, 25 yielded significant hits in BLAST analysis (and evidence of expression; see Supplementary Table 7). By contrast, 16 of the 30 host-range-associated genes in subtelomeric regions in *F. temperatum* had BLAST hits. Phylogenetic analyses of these genes indicated that most had evolutionary origins within Fungi (see Table 2). However, the origins for a substantial number of subtelomeric genes remained unknown as there were no homologs for them in GenBank, while one gene in both sets apparently have origins outside of Fungi (see groups 7 and 8, respectively in Table 2).

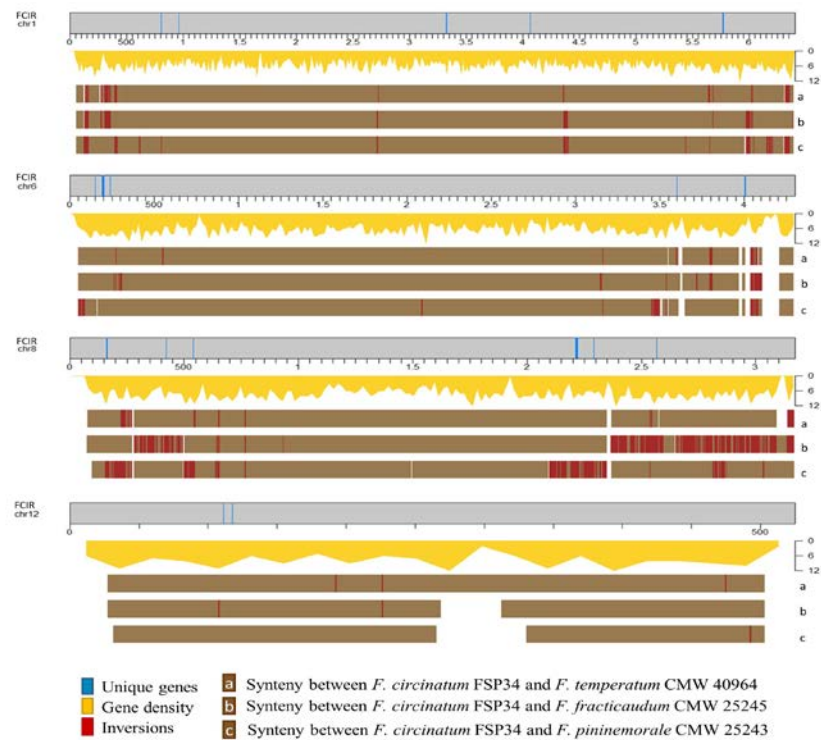


Figure 1. The distribution of host-range-associated genes from pine-host-associated *Fusarium* species and conservation of syntenicity across and among chromosomes and genomes. Distribution of pine-associated genes across each of the chromosomes are indicated by vertical blue lines. The

conservation of synteny and inversion between the relevant genomes are indicated in the brown blocks and red lines. FCIR = *F. circinatum*; chromosome size is given in Kb.

The density of host-range-associated genes differed among chromosomes in both *F. circinatum* and *F. temperatum* (Supplementary Table S11). Most chromosomes, except for Chromosome 5, 10 and 12, had higher unique gene densities for *F. temperatum* when the assemblies for *F. circinatum* and *F. temperatum* were compared. Homologous chromosomes of *F. circinatum* and *F. temperatum* had similar densities of host-range-associated genes, ranging from 1.04x (Chromosome 4) to 1.97x (Chromosome 7). Moderate differences appeared in the host-range-associated gene density (more than two-fold) of Chromosome 9 and 12, where *F. circinatum* contained more than *F. temperatum*, and vice versa, for these two chromosomes. Lastly, a major difference (more than seven-fold) occurred on Chromosome 3 and 6, where *F. circinatum* had more host-range-associated genes compared to the same chromosomes in *F. temperatum*. The sequence of *F. circinatum* Chromosome 6 did not include a telomeric cap. It is also worth noting that Chromosome 12 in *F. temperatum* contained more host-range-associated genes compared to the same chromosome in *F. circinatum*.



Figure 2. The distribution of host-range-associated genes from Poaceae-host-associated *Fusarium* species and conservation of synteny across and between chromosomes and genomes. Poaceae-associated genes distribution across each of the chromosomes as indicated by the blue lines. The conservation of synteny and inversion between the relevant genomes are indicated in the brown blocks and red lines. FTEMP = *F. temperatum*; chromosome size is given in Kb.

Our data showed that some of the host-range-associated genes occurred in clusters or were located in close proximity to one another, and that such genes could have the same or different ancestral origins (Figure 3; Supplementary Table S12). For example, *F. circinatum* genes FCIR_6_gene_2.26 and FCIR_6_gene_2.131 were adjacent to one another on Chromosome 6 and were in the same ancestral origin group (Group 5) (Figure 3A; Table 2), whereas genes FCIR_6_gene_2.2 and FCIR_6_gene_2.142 were near one another on Chromosome 6 but were in different ancestral origin groups (Groups 8 and 7, respectively) (Figure 3A and 3B; Table 2).

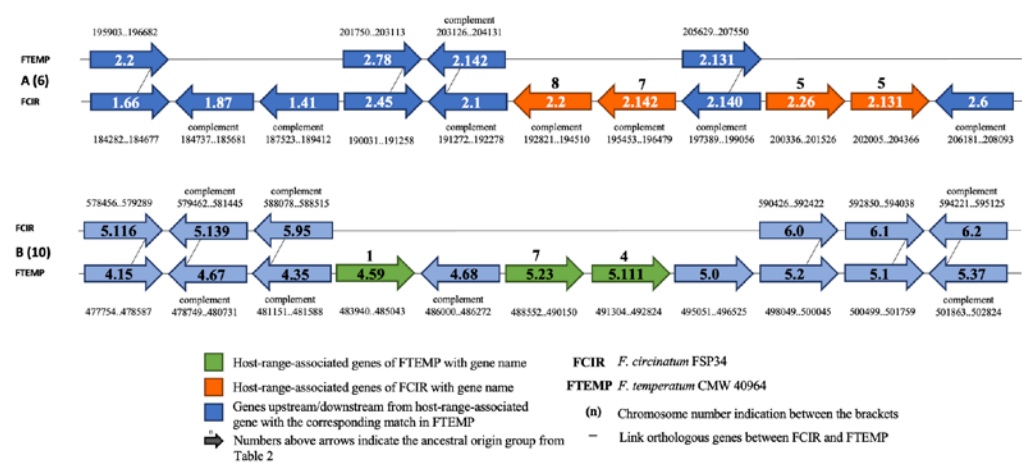


Figure 3. Examples of clusters of host-range-associated genes on Chromosome 6 of *F. circinatum* (A) and Chromosome 10 of *F. temperatum* (B). Numbers within the block arrows indicate gene identities. Bold letters above the green and orange genes indicate ancestral origin group as described in Table 2.

Using SynChro [52], we investigated whether gene areas flanking clustered host-range-associated genes exhibited synteny in *F. circinatum* and *F. temperatum* (Supplementary Figure S11 and S12). For this analysis, we included 10 genes flanking host-range-associated genes, five genes upstream and five downstream. We used the analysis to determine the frequency with which 1–10 genes flanking host-range-associated genes were syntenic in *F. circinatum* and *F. temperatum*. These analyses indicated that the genomic locations around host-range-associated genes in *F. circinatum* were normally distributed (Shapiro-Wilk’s test for departure from normality; $P > 0.05$; $W = 0.95$) [53], but not in *F. temperatum* ($P < 0.05$; $W = 0.86$) (Figure 4).

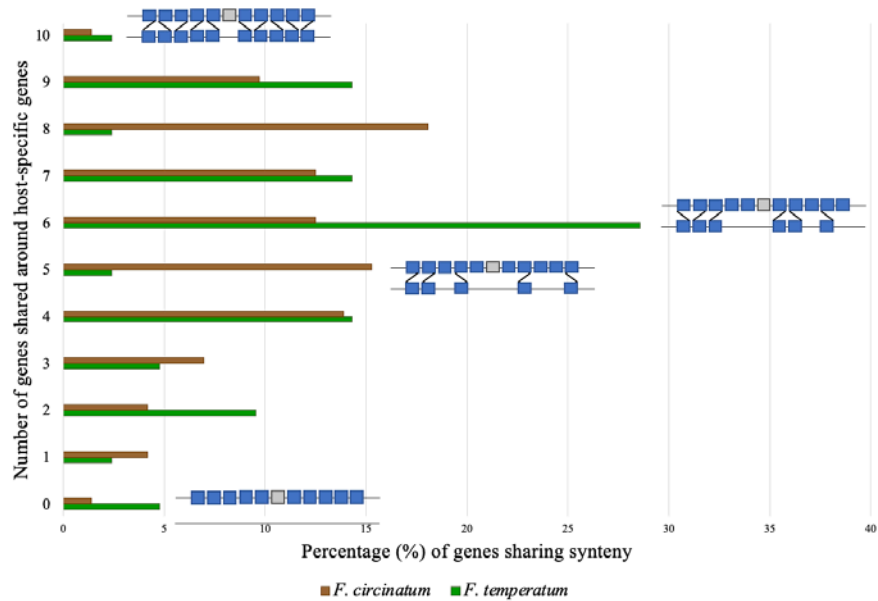


Figure 4. Data retrieved from the Shapiro-Wilk’s test for departure from normality for host-range-associated genes in *F. circinatum* and *F. temperatum*. Grey blocks represent the host-range-associated gene, whereas the blue blocks represent the genes downstream and upstream of the host-range-associated gene.

The difference in distribution pattern was particularly evident when comparing clusters in the two genomes that had five and six of the 10 syntenic genes. In *F. temperatum*,

the clusters with five out of 10 syntenous genes appeared to be severely depleted, indicating minimal synteny when compared to the *F. circinatum* genome. However, the clusters with six out of 10 syntenous genes in *F. temperatum* have been expanded compared to *F. circinatum*. The general trend was that genes flanking host-range-associated genes in *F. circinatum* were frequently syntenic with genes in *F. temperatum*, whereas genes flanking host-range-associated genes in *F. temperatum* were less frequently syntenic with genes in *F. circinatum*. This trend likely suggest higher conservation of the genes and/or genomic regions neighbouring or flanking the position of host-range-associated genes in *F. circinatum* compared to *F. temperatum*.

3. Discussion

In the current study, we used a comparative genomics approach to explore the molecular basis of plant-fungus interactions by making use of two groups of *Fusarium* species in the FFSC, i.e., one associated with pine and the other with grass. This allowed identification of sets of genes that are unique to each of the two groups of *Fusarium* species, and that differed markedly in terms of their identity and function of the proteins they encode. The two sets of genes also showed large differences in ancestral origins, and they tended to occur in subtelomeric regions of chromosomes. Overall, our findings highlighted the different evolutionary and molecular processes that underpin the capacity of these fungi to colonize their respective plant hosts.

To conduct this study, we used two high-quality reference genome assemblies. The one was for the pine pathogen *F. circinatum*, while the other was for the maize pathogen *F. temperatum*. Within these assemblies, most chromosomes had telomere-to-telomere coverage, which allowed us to compare placement, localization and clustering of the host-range-associated genes identified. This also allowed for comparisons of chromosomal architecture among four of the six *Fusarium* species studied here.

The host-range-associated genes identified in the genomes of the pine-host-associates (*F. circinatum*, *F. fructicaudum* and *F. pininemorale*) and in those of the grass-host-associates (*F. konzum*, *F. subglutinans* and *F. temperatum*) were localized preferentially in subtelomeric regions. The subtelomeric location of genes have been studied in several organisms which include *Homo sapiens* [54], *Drosophila melanogaster* [55], *Plasmodium falciparum* [56], *Trypanosoma brucei* [57], *Saccharomyces cerevisiae* [58], and the fungal parasite *Encephalitozoon cuniculi* [59]. Upon genome analyses of *S. cerevisiae*, Louis [58] uncovered that subtelomeric genes belong to several gene families encoding proteins involved in carbon source utilization. In humans, for example, genes encoding for olfactory receptors (i.e., proteins capable of binding odor molecules) are the most studied and largest gene family located in the subtelomeric regions of the human genome [60, 61]. The presence of subtelomeric regions have potential importance in replication and chromosome stability, but also seem to play important roles in reversible silencing of genes mediated by proteins binding to the telomere, and engagement in ectopic recombination with other subtelomeres, likely resulting in gene diversification [62-64].

In fungi, the emerging evidence suggests that genes located in subtelomeres play adaptive roles, such as contributing to niche- or host-specificity and virulence in pathogens [65-68]. For example, the fungus *Pneumocystis jiroveci* have large and highly variable gene families located in its subtelomeric regions that encode for surface proteins with tightly regulated, yet switchable, expression patterns [69]. *Fusarium* subtelomeres have been reported to contain virulence genes involved in host-specificity, e.g., in *F. graminearum* [35, 70, 71] and *F. fujikuroi* [28]. In *F. circinatum*, a locus determining growth rate region was also characterized from a subtelomeric region [67]. In the case of the *Fusarium* species examined in this study, a large proportion of the genes inferring potential host-specificity were positioned on subtelomeres, thereby pointing to their possible roles in fungal development, adaptation, and survival for these *Fusarium* species.

The presence/absence of host-range-associated genes in the two sets of genomes, as well as their placement in subtelomeric regions is consistent with what is expected for genes belonging to the accessory genomic compartment. In other words, they represent

accessory genes likely involved in niche-associated traits which are typically enriched for functions related to cell–cell interactions, secondary metabolism, and stress responses, all potential contributors to pathogenicity and virulence [33, 72]. The dynamic evolutionary processes associated with the accessory genome [33, 40, 41, 73, 74] likely allowed for preferential loss and/or gain of genes with new functions. In turn, this ultimately led to the ability of the pine- and grass-host-associates examined here to colonize and survive on the tissue of their respective hosts.

Our findings are consistent with the notion that horizontal gene transfer (HGT) plays a major role in the evolution of accessory genomes [33, 40, 41, 73–77]. As such, HGT is increasingly recognized as a significant driver of adaptive evolution in fungi and pathogenic traits of pathogens [78–80]. For example, the ability of *Valsa mali* to grow on its host and to be pathogenic to it was acquired via HGT from diverse sources [80]. HGT can even cause basic lifestyle changes; e.g., facilitating transition of the plant-associated ancestor of *Metarhizium* into the entomopathogenic taxon known today [81]. Among FFSC species, multiple independent HGT events reported caused the development of the growth-determining locus in *F. circinatum* [67]. The polyphyletic origins of many of the host-range-associated genes identified in the current study were likely also acquired from such diverse external sources (see groups 5–8 in Table 2), while most of the remainder likely originated within *Fusarium* or the FFSC and were retained only in certain lineages (groups 1–4 in Table 2). More research is needed to determine the possible role of internal genomic mutations (due to duplication, displacement, and translocation events) [29, 82, 83] in determining the host associations of the examined fungi.

Despite the existence of genes unique to the respective pine- and grass-host-associated fungi, their genomes were remarkably syntenic [29, 84]. Variability was most evident when comparisons of inversions were investigated. Inversions closer to the ends of chromosomes have been previously reported in fungi [35, 85, 86]. Inversions were most noticeable for Chromosome 8, especially when considering the host-range-associated genes on this chromosome. All host-range-associated genes from Chromosome 8 were located within regions prone to inversion. The variability of chromosome 8 is likely due to the sizable reciprocal translocation with Chromosome 11 within some FFSC taxa, including all species investigated in this study [29, 87–90]. Chromosome 11 from *F. circinatum* had only one unique gene located outside of the inversed regions, whereas the same chromosome in *F. temperatum* lacked any host-range-associated genes. This suggests that Chromosome 8 has a larger predisposition to chromosomal architectural changes. Current advances in sequencing technologies highlighted the importance of chromosome rearrangements events in cellular processes. Translocation events have been reported to drive the evolution of secondary metabolite biosynthetic gene clusters, whereas chromosome rearrangements may result in gene loss and/or gain, genetic diversity, and changes in pathogenicity [91–94]. For example, effector genes located in rearranged chromosomal regions are more prone to evolution and are more likely to contribute to adaptation and speciation [95, 96].

The importance of Chromosome 12 in possessing host-range-associated genes poses an interesting question. This chromosome is a dispensable chromosome within the FFSC [41, 97–99] and can be strain-specific [28, 98]. Chromosome 12 is not essential for pathogenicity [28, 98, 100] and its role in niche-specificity is unknown. However, the larger Chromosome 12 of *F. temperatum* contained more host-range-associated genes potentially involved in niche-specificity among *Fusarium* species that are associated with Poaceae. This occurrence may imply that some elements of the chromosome length polymorphism between *F. circinatum* and *F. temperatum* accounts for the addition of host-range-associated genes and that this chromosome is rapidly evolving. This chromosome should therefore be an important target for future studies aimed at unravelling the molecular basis of niche-adaptation in more detail.

Based on their inferred products, the host-range-associated genes could mostly be grouped being involved in virulence, substrate transport, and metabolism of carbohydrates, and in the case of *F. temperatum* also a heterokaryon incompatibility (HET)-domain

protein. It also highlights the evolutionary changes Poaceae-associated *Fusarium* species had to undergo to acquire these genes to be successful colonizers of members of Poaceae. The presence of these groups of genes is indicative of the role in the survival and adaptation of these *Fusarium* species, likely providing a competitive advantage as well as providing a protective advantage against the host defense responses.

In conclusion, this study used near-complete and high-quality genome data to identify a set of genes that potentially underpin the association of six *Fusarium* species with their pine and grass hosts. Based on their inferred products, these genes are involved in diverse processes that could potentially provide the fungi with ecological advantages in their respective niches [101-106]. Furthermore, these genes preferentially form part of the accessory genome of the fungi examined where diverse processes likely determined their evolution and development. This work thus forms a strong foundation from which future studies could explore the functional relevance of the genes identified, as well as their expression, especially given their chromosomal locations. Such studies are needed if we are to understand the ecology and biology of the fungi that threaten the health of socioeconomically important plants and crops.

4. Materials and Methods

4.1. Genome sequences, assembly completeness and gene prediction

The completeness of each genome assembly was evaluated through the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool v. 3.0.2 [44] using the “Sordariomyceta” database. Using the significant synteny found within the FFSC [28, 29], scaffolds from the genomes of *F. pininemorale* (isolated from diseased *P. tecunumanii* tissue [13, 88]) and *F. fracticaudum* (isolated from diseased *Pinus maximinoi* tissue [13, 89]) were compared against the genomes of *F. circinatum* [90] (isolated from diseased *Pinus radiata* tissue [11, 107]), *F. fujikuroi* [28] and *F. temperatum* [87] (isolated from diseased teosinte tissue [10, 87]) as these have been assembled into chromosomes. The scaffolds were ordered and orientated into 12 contiguous pseudochromosomes [representing the core chromosomes (1-11) and the accessory chromosome (12)] [29]. The genomes of *F. subglutinans* (733 contigs; isolated from diseased maize tissue [20, 108]) and *F. konzum* (2 262 contigs; isolated from prairie grass [19, 108]) were considered too fragmented to be assembled into pseudochromosomes.

A reference genome was selected to represent either the pine- or grass-host-associates. The purpose of this was to achieve a detailed comparison with two well-assembled genomes between *Fusarium* species colonizing different plant hosts. *Fusarium circinatum* was chosen to represent the pine-host-associates and *F. temperatum* represented grass-host-associates. The abundance of the telomere-associated repeat sequence (TTAGGG/CCCTAA) [46, 47] was evaluated using a motif search performed in CLC Genomics Workbench v. 11 (CLC bio, Aarhus, Denmark), to further investigate the completeness of the two reference genomes (*F. circinatum* and *F. temperatum*). For the motif search, a window size of 10 000 bp with 5 000 bp increments was used. Only repeats with $\geq 80\%$ similarity to the telomere repeat were considered in this analysis.

The MAKER2 pipeline [45] was utilized for structural annotations of all six genomes in order to identify protein-coding genes. Gene prediction was performed in MAKER using SNAP [109], GeneMark ES [110] and AUGUSTUS [111]. As additional evidence, gene model data from *F. circinatum* [90], *F. fujikuroi* [28], *F. verticillioideus* and *F. graminearum* [41], as well as *F. mangiferae* and *F. proliferatum* [18], were included. These isolates were selected due to the availability of their genomic information on the NCBI public database (<https://www.ncbi.nlm.nih.gov/>).

4.2. Genes unique to pine- and grass-host-associated species

The gene content of all six *Fusarium* genomes was evaluated to determine which genes are unique to the species associated with the two groups of plant hosts. For this

purpose, OrthoFinder v. 2.3.1 was implemented [48]. The genes occurring only in the genomes of the pine- and grass-host-associates, respectively, were labeled as “unique” genes.

Functional annotation was performed using the Blast2GO [49] plugin for CLC Genomics Workbench. A two-tailed Fisher exact test was implemented ($P < 0.05$) in Blast2GO [49] to detect the GO terms that were overrepresented in the host-range-associated genes set of both *Fusarium* representatives, using the whole genome of each as reference.

4.3. Phylogenetic origins of the host-range-associated genes

The two unique gene sets were uploaded onto NCBI to perform BLASTp searches against the non-redundant database using the online position-specific iterative (psi) BLAST tool [112]. All highly divergent protein sequences were excluded by considering only those sequences with at least 40% amino acid identity over 70% of the query sequence length and that had E-values $\leq 1 \times 10^{-5}$ [67]. All the sequences were aligned using the constraint-based alignment tool (COBALT) [113] and phylogenetic trees were viewed to determine the ancestral origins of the host-range-associated genes. The host-range-associated genes with unclear ancestral origin were selected to construct phylogenetic trees. To infer phylogenies, all sequences were aligned with MAFFT (Multiple sequence Alignment based on Fast Fourier Transform) v. 7.0 with default settings [114]. These alignments included the relevant *Fusarium* sequences, together with those from other Ascomycota. MEGA v. 7.0.26 [115] was used to draw initial tree(s) for the heuristic search by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances to estimate the best substitution model to use. The Maximum Likelihood branch support was estimated using bootstrap analyses based on 100 pseudoreplicates.

4.4. Genomic distribution of the host-range-associated genes

The location and distribution of the host-range-associated genes were plotted across the 12 chromosomes in each of the two reference genomes using KaryoploteR v. 3.9 [116]. The difference in the distribution of genes in the different genomic regions was evaluated using Chi-squared tests ($P < 0.05$), by comparing those within and outside subtelomeric regions, with the null hypothesis that they are characterized by similar frequencies. Here we regarded the first and last 500 kb from chromosome ends as subtelomeres [67, 117–119]. Furthermore, the synteny and conservation of the location of the different host-range-associated genes were studied with SynChro which revealed the synteny break-points between the reference and query genomes [52]. Lastly, the Shapiro-Wilks test was performed to test for normality by detecting all departures from normality for genes belonging to *F. circinatum* and *F. temperatum*.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1. Phylogenetic relationship of paralogs with regards to the respective host-range-associated gene. FCIR = *Fusarium circinatum*; FFRAC = *Fusarium fructicaudum*; FPIN = *Fusarium pininemorale*; FSUB = *Fusarium subglutinans*, Figure S2. Host-range-associated genes with ancestral origins that emerged within the FFSC. Host-range-associated genes investigated are highlighted in yellow; FCIR = *Fusarium circinatum* and FTEMP = *Fusarium temperatum*, Figure S3. Host-range-associated genes with ancestral origins that emerged within the FFSC and FOSC. Host-range-associated genes investigated are highlighted in yellow; FCIR = *Fusarium circinatum* and FTEMP = *Fusarium temperatum*, Figure S4. Host-range-associated genes with ancestral origins that emerged within the broader *Fusarium* outside the FFSC and FOSC. Host-range-associated genes investigated are highlighted in yellow; FCIR = *Fusarium circinatum* and FTEMP = *Fusarium temperatum*, Figure S5. Host-range-associated genes with less than 10 ancestral origin hits and mostly *Fusarium*. Host-range-associated genes investigated are highlighted in yellow; FCIR = *Fusarium circinatum* and FTEMP = *Fusarium temperatum*, Figure S6. Host-range-associated genes with ancestral origins hits and mostly not *Fusarium*. Host-range-associated genes investigated are highlighted in yellow; FCIR = *Fusarium circinatum* and FTEMP = *Fusarium temperatum*, Figure S7. Host-range-associated genes with ancestral origins outside *Fusarium* but in the Ascomycetes. Host-range-associated genes investigated

are highlighted in yellow; FCIR = *Fusarium circinatum* and FTEMP = *Fusarium temperatum*, Figure S8. Host-range-associated genes with ancestral origins outside Fungi. Host-range-associated genes investigated are highlighted in yellow; FCIR = *Fusarium circinatum* and FTEMP = *Fusarium temperatum*, Figure S9. The distribution of host-range-associated genes from pine-host-associated *Fusarium* species and conservation of synteny across and between chromosomes and genomes. Pine-host-associated genes distribution across each of the chromosomes as indicated by the blue lines. The conservation of synteny and inversion between the relevant genomes are indicated in the brown blocks and red lines. FCIR = *F. circinatum*; chromosome size is given in kbp., Figure S10. The distribution of host-range-associated genes from Poaceae-host-associated *Fusarium* species and conservation of synteny across and between chromosomes and genomes. Poaceae-host-associated genes distribution across each of the chromosomes as indicated by the blue lines. The conservation of synteny and inversion between the relevant genomes are indicated in the brown blocks and red lines. FTEMP = *F. temperatum*; chromosome size is given in kbp., Figure S11. The syntenous relationship between genes from *F. circinatum* versus *F. temperatum*., Figure S12. The syntenous relationship between genes from *F. temperatum* and *F. circinatum*., Table S1. BUSCO results for the relevant *Fusarium* genomes. Table S2. The size difference between the chromosomes of four of the six *Fusarium* species., Table S3. Presence of telomeres at chromosomal ends for the two representative *Fusarium* species examined., S4 Table. The Blast2GO data for the 72 unique pine-host-associated genes specifically for (A) *Fusarium circinatum*, (B) *F. fructicaudum* and (C) *F. pininemorale*., Table S5. The Blast2GO data for the 47 unique Poaceae-host-associated genes specifically for (A) *Fusarium konzum*, (B) *F. subglutinans* and (C) *F. temperatum*., Table S6. The Fischer exact test data for (A) the 72 unique pine-host-associated genes and (B) the 47 unique Poaceae-host-associated genes., Table S7. The EST and RNA-seq data for *F. circinatum*, obtained from Wingfield et al. [51] and van Wyk et al. [50], respectively., Table S8. The placement of host-range-associated genes in groups that infer their evolutionary origins., Table S9. The gene information for the unique *F. circinatum* genes, in terms of chromosome location, subtelomeric placement, ancestral origin and BLAST description., Table S10. The gene information for the unique *F. temperatum* genes, in terms of chromosome location, subtelomeric placement, ancestral origin and BLAST description., Table S11. The host-range-associated gene density for both *F. circinatum* and *F. temperatum*., Table S12. The SynChro data for genes downstream and upstream of the host-range-associated genes of both the pine- and Poaceae-host-associated *Fusarium* species; FCIR = *Fusarium circinatum* and FTEMP = *Fusarium temperatum*.

Author Contributions: Conceptualization, C.D., L.D.V., E.T.S., M.A.v.d.N. and B.D.W.; methodology, C.D., L.D.V., E.T.S., M.A.v.d.N. and B.D.W.; software, C.D., L.D.V., E.T.S., M.A.v.d.N., B.D.W. and Q.C.S.; validation, C.D., L.D.V., and E.T.S.; formal analysis, C.D., L.D.V. and M.A.v.d.N.; investigation, C.D., L.D.V., M.A.v.d.N.; resources, L.D.V., E.T.S. and B.D.W.; data curation, C.D., L.D.V., E.T.S., M.A.v.d.M., and B.D.W.; writing—original draft preparation, C.D. and L.D.V.; writing—review and editing, C.D., L.D.V., E.T.S., M.A.v.d.M., Q.C.S. and B.D.W.; visualization, C.D. L.D.V. and M.A.v.d.N.; supervision, L.D.V., E.T.S., M.A.v.d.M. and B.D.W.; project administration, L.D.V.; funding acquisition, B.D.W. and E.T.S.; critical revisions and writing, C.D., L.D.V. and E.T.S.

Funding: This research was funded by the South African Department of Science and Innovation's South African Research Chair Initiative (SARChI grant number 98353), through the research chair of Prof. BD Wingfield. We would like to acknowledge the DSI-NRF Centre of Excellence in Plant Health Biotechnology (CPHB grant number 40945), at the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria. RHP was supported by the Food Safety National Program (Project Number 5010-42000-053-00D) at the United States Department of Agriculture, Agricultural Research Service. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgements: The authors are grateful for the excellent technical assistance of Amy E. McGovern and Crystal E. Probyn.

References

- Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharyya MK, Brandt ME, et al. One fungus, one name: Defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology*. 2013;103(5):400-408.
- Geiser DM, Al-Hatmi AM, Aoki T, Arie T, Balmas V, Barnes I, et al. Phylogenomic analysis of a 55.1-kb 19-gene dataset resolves a monophyletic *Fusarium* that includes the *Fusarium solani* species complex. *Phytopathology*. 2021;111(7):1064-1079.
- Chang DC, Grant GB, O'Donnell K, Wannemuehler KA, Noble-Wang J, Rao CY, et al. Multistate outbreak of *Fusarium* keratitis associated with use of a contact lens solution. *American Journal of Ophthalmology*. 2006;142(5):896-897.
- Sutton DA, Brandt ME. *Fusarium* and other opportunistic hyaline fungi. 10th ed. ed. Washington, DC: ASM Press; 2011.
- O'Donnell K, Sutton DA, Rinaldi MG, Magnon KC, Cox PA, Revankar SG, et al. Genetic diversity of human pathogenic members of the *Fusarium oxysporum* complex inferred from multilocus DNA sequence data and amplified fragment length polymorphism analyses: Evidence for the recent dispersion of a geographically widespread clonal lineage and nosocomial origin. *Journal of Clinical Microbiology*. 2004;42(11):5109-5120.
- Ortoneda M, Guarro J, Madrid MP, Caracul Z, Roncero MIG, Mayayo E, et al. *Fusarium oxysporum* as a multihost model for the genetic dissection of fungal virulence in plants and mammals. *Infection and Immunity*. 2004;72(3):1760-1766.
- Navarro-Velasco GY, Prados-Rosales RC, Ortiz-Urquiza A, Quesada-Moraga E, Di Pietro A. *Galleria mellonella* as model host for the trans-kingdom pathogen *Fusarium oxysporum*. *Fungal Genetics and Biology*. 2011;48(12):1124-1129.
- Pelizza S, Stenglein S, Cabello M, Dinolfo M, Lange C. First record of *Fusarium verticillioides* as an entomopathogenic fungus of grasshoppers. *Journal of Insect Science*. 2011;11(1):70.
- Parry D, Jenkinson P, McLeod L. *Fusarium* ear blight (scab) in small grain cereals—A review. *Plant Pathology*. 1995;44(2):207-238.
- Desjardins AE, Plattner RD, Gordon TR. *Gibberella fujikuroi* mating population A and *Fusarium subglutinans* from teosinte species and maize from Mexico and Central America. *Mycological Research*. 2000;104(7):865-872.
- Gordon TR, Storer AJ, Okamoto D. Population structure of the pitch canker pathogen, *Fusarium subglutinans* f. sp. *pini*, in California. *Mycological Research*. 1996;100(7):850-854.
- Bottalico A, Perrone G. Toxigenic *Fusarium* species and mycotoxins associated with head blight in small-grain cereals in Europe. *European Journal of Plant Pathology*. 2002;611-624.
- Herron DA, Wingfield MJ, Wingfield BD, Rodas CA, Marincowitz S, Steenkamp ET. Novel taxa in the *Fusarium fujikuroi* species complex from *Pinus* spp. *Studies in Mycology*. 2015;80:131-150.
- Armitage AD, Taylor A, Sobczyk MK, Baxter L, Greenfield BPJ, Bates HJ, et al. Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *Scientific Reports*. 2018;8(1):13530.
- Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, et al. Emerging fungal threats to animal, plant and ecosystem health. *Nature*. 2012;484(7393):186-194.
- O'Donnell K, Cigelnik E, Nirenberg HI. Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia*. 1998;90(3):465-493.
- Steenkamp ET, Wingfield BD, Desjardins AE, Marasas WF, Wingfield MJ. Cryptic speciation in *Fusarium subglutinans*. *Mycologia*. 2002;94(6):1032-1043.
- Niehaus E-M, Kim H-K, Münsterkötter M, Janevska S, Arndt B, Kalinina SA, et al. Comparative genomics of geographically distant *Fusarium fujikuroi* isolates revealed two distinct pathotypes correlating with secondary metabolite profiles. *PLoS Pathogens*. 2017;13(10):e1006670.
- Zeller KA, Summerell BA, Bullock S, Leslie JF. *Gibberella konza* (*Fusarium konzum*) sp. nov. from prairie grasses, a new species in the *Gibberella fujikuroi* species complex. *Mycologia*. 2003;95(5):943-954.
- Desjardins AE, Maragos CM, Proctor RH. Maize ear rot and moniliformin contamination by cryptic species of *Fusarium subglutinans*. *Journal of Agricultural and Food Chemistry*. 2006;54(19):7383.
- Swett CL, Gordon TR. Endophytic association of the pine pathogen *Fusarium circinatum* with corn (*Zea mays*). *Fungal Ecology*. 2015;13:120-129.
- Swett CL, Gordon TR. First report of grass species (Poaceae) as naturally occurring hosts of the pine pathogen *Gibberella circinata*. *Plant Disease*. 2012;96(6):908-908.
- Carter JW, Gordon TR. Infection of the native California grass, *Bromus carinatus*, by *Fusarium circinatum*, the cause of pitch canker in pines. *Plant Disease*. 2020;104(1):194-197.
- Herron DA, Wingfield MJ, Fru F, Wingfield BD, Steenkamp ET. Grasses as a refuge for *Fusarium circinatum* L. – evidence from South Africa. *Southern Forests*. 2020;82(3):253-262.
- Hernandez-Escribano L, Iturrutxa E, Elvira-Recuenco M, Berbegal M, Campos JA, Renobales G, et al. Herbaceous plants in the understory of a pitch canker-affected *Pinus radiata* plantation are endophytically infected with *Fusarium circinatum*. *Fungal Ecology*. 2018;32:65-71.
- Martín-García J, Lukačevićová A, Flores-Pacheco JA, Diez JJ, Dvořák M. Evaluation of the susceptibility of several Czech conifer provenances to *Fusarium circinatum*. *Forests*. 2018;9(2):72.
- Drenkhan R, Ganley B, Martín-García J, Vahalík P, Adamson K, Adamčíková K, et al. Global geographic distribution and host range of *Fusarium circinatum*, the causal agent of pine pitch canker. *Forests*. 2020;11(7):724.

28. Wiemann P, Sieber CMK, von Bargen KW, Studt L, Niehaus E-M, Espino JJ, et al. Deciphering the cryptic genome: Genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLoS Pathogens*. 2013;9(6):e1003475.
29. De Vos L, Steenkamp ET, Martin SH, Santana QC, Fourie G, van der Merwe NA, et al. Genome-wide macrosynteny among *Fusarium* species in the *Gibberella fujikuroi* complex revealed by amplified fragment length polymorphisms. *PloS One*. 2014;9(12):e114682.
30. De Vos L, Myburg AA, Wingfield MJ, Desjardins AE, Gordon TR, Wingfield BD. Complete genetic linkage maps from an inter-specific cross between *Fusarium circinatum* and *Fusarium subglutinans*. *Fungal Genetics and Biology*. 2007;44(8):701-714.
31. Nirmala J, Dahl S, Steffenson BJ, Kannangara CG, Von Wettstein D, Chen X, et al. Proteolysis of the barley receptor-like protein kinase RPG1 by a proteasome pathway is correlated with Rpg1-mediated stem rust resistance. *Proceedings of the National Academy of Sciences of the United States of America*. 2007;104(24):10276-10281.
32. Covert SF. Supernumerary chromosomes in filamentous fungi. *Current Genetics*. 1998;33(5):311-319.
33. Croll D, McDonald BA. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens*. 2012;8(4):e1002608.
34. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nature Reviews Microbiology*. 2012;10(6):417-430.
35. Zhao C, Waalwijk C, de Wit PJGM, Tang D, van der Lee T. Relocation of genes generates non-conserved chromosomal segments in *Fusarium graminearum* that show distinct and co-regulated gene expression patterns. *BMC Genomics*. 2014;15(1):191.
36. Hardison RC. Comparative genomics. *PLoS Biology*. 2003;1(2):e58.
37. Wittenberg AHJ, Van der Lee TAJ, M'Barek SB, Ware SB, Goodwin SB, Kilian A, et al. Meiosis drives extraordinary genome plasticity in the haploid fungal plant pathogen *Mycosphaerella graminicola*. *PLoS One*. 2009;4(6):e5863.
38. Goodwin SB, Ben M'Barek S, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, et al. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics*. 2011;7(6):e1002070.
39. Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, et al. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(38):13826-13831.
40. Akagi Y, Akamatsu H, Otani H, Kodama M. Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryotic Cell*. 2009;8(11):1732-1738.
41. Ma L, Van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. 2010;464(7287):367-373.
42. Bäuml A, Fang FC. Host specificity of bacterial pathogens. *Cold Spring Harbor Perspectives in Medicine*. 2013;3(12):a010041.
43. Kisiela DI, Chattopadhyay S, Libby SJ, Karlinsey JE, Fang FC, Tchesnokova V, et al. Evolution of *Salmonella enterica* virulence via point mutations in the fimbrial adhesin. *PLoS Pathogens*. 2012;8(6):e1002733.
44. Waterhouse RM, Seppely M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*. 2018;35(3):543-548.
45. Capek P, Alfoldi J, Liskova D. An acetylated galactoglucomanan from *Picea abies* L. Karst. *Carbohydrate Research*. 2002;337(11):1033-1037.
46. Garcia-Pedrajas MD, Roncero MI. A homologous and self-replicating system for efficient transformation of *Fusarium oxysporum*. *Current Genetics*. 1996;29(2):191-198.
47. Fulnecková J, Sevcíková T, Fajkus J, Lukesová A, Lukes M, Vlcek C, et al. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biology and Evolution*. 2013;5(3):468-483.
48. Emms DM, Kelly S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*. 2015;16(1):157.
49. Conesa A, Gotz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*. 2008;2008:619832.
50. Van Wyk S, Wingfield BD, De Vos L, Van der Merwe NA, Santana QC, Steenkamp ET. Repeat-induced point mutations drive divergence between *Fusarium circinatum* and its close relatives. *Pathogens*. 2019;8(4):E298.
51. Wingfield BD, Steenkamp ET, Santana Q, Coetzee M, Bam S. First fungal genome sequence from Africa: A preliminary analysis. *South Africa Journal of Science*. 2012;108:104-112.
52. Drillon G, Carbone A, Fischer G. SynChro: A fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. *PLoS One*. 2014;9(3):e92621.
53. Shapiro SS, Wilk MB. Analysis of variance test for normality (complete samples). *Biometrika*. 1965;52:591-611.
54. Mefford HC, Trask BJ. The complex structure and dynamic evolution of human subtelomeres. *Nature Reviews Genetics*. 2002;3(2):91-102.
55. Mason JM, Biessmann H. The unusual telomeres of *Drosophila*. *Trends in Genetics*. 1995;11(2):58-62.
56. Scherf A, Figueiredo LM, Freitas-Junior LH. *Plasmodium* telomeres: a pathogen's perspective. *Current Opinion in Microbiology*. 2001;4(4):409-414.

57. Bringaud F, Biteau N, Melville SE, Hez S, El-Sayed NM, Leech V, et al. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryotic Cell*. 2002;1(1):137-151.
58. Louis EJ. The chromosome ends of *Saccharomyces cerevisiae*. *Yeast*. 1995;11(16):1553-1573.
59. Peyret P, Katinka MD, Duprat S, Duffieux F, Barbe V, Barbazanges M, et al. Sequence and analysis of chromosome I of the amitochondriate intracellular parasite *Encephalitozoon cuniculi* (Microspora). *Genome Research*. 2001;11(2):198-207.
60. Trask BJ, Friedman C, Martin-Gallardo A, Rowen L, Akinbami C, Blankenship J, et al. Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Human Molecular Genetics*. 1998;7(1):13-26.
61. Glusman G, Yanai I, Rubin I, Lancet D. The complete human olfactory subgenome. *Genome Research*. 2001;11(5):685-702.
62. Barry J, Ginger ML, Burton P, McCulloch R. Why are parasite contingency genes often associated with telomeres? *International Journal for Parasitology*. 2003;33(1):29-45.
63. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP, et al. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*. 2005;121(1):25-36.
64. Figueiredo LM, Janzen CJ, Cross GA. A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS Biology*. 2008;6(7):e161.
65. Rahnama M, Wang B, Dostart J, Novikova O, Yackzan D, Yackzan A, et al. Telomere roles in fungal genome evolution and adaptation. *Frontiers in Genetics*. 2021;12(1303).
66. Huang X. Host-specific subtelomere: Genomic architecture of pathogen emergence in asexual filamentous fungi. *bioRxiv*. 2019:721753.
67. Van Wyk S, Wingfield BD, De Vos L, Santana QC, Van der Merwe NA, Steenkamp ET. Multiple independent origins for a subtelomeric locus associated with growth rate in *Fusarium circinatum*. *IMA Fungus*. 2018;9(1):27-36.
68. Juárez-Reyes A, Castaño I. Chromatin architecture and virulence-related gene expression in eukaryotic microbial pathogens. *Current Genetics*. 2019;65(2):435-443.
69. Keely SP, Renauld H, Wakefield AE, Cushion MT, Smulian AG, Fosker N, et al. Gene arrays at *Pneumocystis carinii* telomeres. *Genetics*. 2005;170(4):1589-1600.
70. Galagan JE, Calvo SE, Cuomo C, Ma L, Wortman JR, Batzoglou S, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*. 2005;438(7071):1105.
71. Perrin RM, Fedorova ND, Bok JW, Cramer Jr RA, Wortman JR, Kim HS, et al. Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by LaeA. *PLoS Pathogens*. 2007;3(4):e50.
72. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*. 2018;556(7701):339-344.
73. Ma LJ, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, Trail F, et al. *Fusarium* pathogenomics. *Annual Review of Microbiology*. 2013;67:399-416.
74. Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, et al. The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics*. 2009;5(8):e1000618.
75. Donald MG, Megan CM, Lorenzo C, Peter SS, Anca GR, Mhairi M, et al. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLoS Pathogens*. 2012; 8(9):1002952.
76. Schönknecht G, Weber AP, Lercher MJ. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *Bioessays*. 2014;36(1):9-20.
77. Sieber CM, Lee W, Wong P, Münsterkötter M, Mewes HW, Schmeitzl C, et al. The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS One*. 2014;9(10):e110311.
78. Richards TA, Leonard G, Soanes DM, Talbot NJ. Gene transfer into fungi. *Fungal Biology Reviews*. 2011;25(2):98-110.
79. Soanes D, Richards TA. Horizontal gene transfer in eukaryotic plant pathogens. *Annual Review of Phytopathology*. 2014;52:583-614.
80. Yin Z, Zhu B, Feng H, Huang L. Horizontal gene transfer drives adaptive colonization of apple trees by the fungal pathogen *Valsa mali*. *Scientific Reports*. 2016;6(1):33129.
81. Zhang Q, Chen X, Xu C, Zhao H, Zhang X, Zeng G, et al. Horizontal gene transfer allowed the emergence of broad host range entomopathogens. *Proceedings of the National Academy of Sciences of the United States of America*. 2019;116(16):7982-7989.
82. Le Gac M, Giraud T. Existence of a pattern of reproductive character displacement in Homobasidiomycota but not in Ascomycota. *Journal of Evolution and Biology*. 2008;21(3):761-772.
83. Proctor RH, McCormick SP, Alexander NJ, Desjardins AE. Evidence that a secondary metabolic biosynthetic gene cluster has grown by gene relocation during evolution of the filamentous fungus *Fusarium*. *Molecular Microbiology*. 2009;74(5):1128-1142.
84. Wiemann P, Sieber CM, von Bargaen KW, Studt L, Niehaus EM, Espino JJ, et al. Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLoS Pathogens*. 2013;9(6):e1003475.

85. Van Dam P, Fokkens L, Ayukawa Y, Van der Gragt M, Horst A, Brankovics B, et al. A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Scientific Reports*. 2017;7:9042.
86. King R, Brown NA, Urban M, Hammond-Kosack KE. Inter-genome comparison of the Quorn fungus *Fusarium venenatum* and the closely related plant infecting pathogen *Fusarium graminearum*. *BMC Genomics*. 2018;19(1):269.
87. Wingfield BD, Barnes I, De Beer ZW, De Vos L, Duong TA, Kanzi AM, et al. Draft genome sequences of *Ceratocystis eucalypticola*, *Chrysosporthe cubensis*, *C. deuterocubensis*, *Davidsoniella virescens*, *Fusarium temperatum*, *Graphilbum fragrans*, *Penicillium nordicum*, and *Thielaviopsis musarum*. *IMA Fungus*. 2015b;6(2):493-506.
88. Wingfield BD, Berger DK, Steenkamp ET, Lim HJ, Duong TA, Bluhm BH, et al. Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntiaella decipiens* and *Ophiostoma ips*. *IMA Fungus*. 2017;8(2):385-396.
89. Wingfield BD, Bills GF, Dong Y, Huang W, Nel WJ, Swalaski-Parry BS, et al. Draft genome sequence of *Annulohypoxyton stygium*, *Aspergillus mulundensis*, *Berkeleyomyces basicola* (syn. *Thielaviopsis basicola*), *Ceratocystis smalleyi*, two *Cercospora beticola* strains, *Coleophoma cylindrospora*, *Fusarium fructicaudum*, *Phialophora* cf. *hyalina*, and *Morchella septimelata*. *IMA Fungus*. 2018a;9(1):199-223.
90. Wingfield BD, Liu M, Nguyen HDT, Lane FA, Morgan SW, De Vos L, et al. Nine draft genome sequences of *Claviceps purpurea* s.lat., including *C. arundinis*, *C. humidiphila*, and *C. cf. spartinae*, pseudomolecules for the pitch canker pathogen *Fusarium circinatum*, draft genome of *Davidsoniella eucalypti*, *Grossmannia galeiformis*, *Quambalaria eucalypti*, and *Teratosphaeria destructans*. *IMA Fungus*. 2018b;9:401-418.
91. Olarte RA, Menke J, Zhang Y, Sullivan S, Slot JC, Huang Y, et al. Chromosome rearrangements shape the diversification of secondary metabolism in the cyclosporin producing fungus *Tolypocladium inflatum*. *BMC Genomics*. 2019;20(1):1-23.
92. Tralamazza SM, Rocha LO, Oggenfuss U, Corrêa B, Croll D. Complex evolutionary origins of specialized metabolite gene cluster diversity among the plant pathogenic fungi of the *Fusarium graminearum* species complex. *Genome Biology and Evolution*. 2019;11(11):3106-3122.
93. Shi-Kunne X, Faino L, van den Berg GC, Thomma BP, Seidl MF. Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangement and gene loss. *Environmental Microbiology*. 2018;20(4):1362-1373.
94. Moolhuijzen P, See PT, Hane JK, Shi G, Liu Z, Oliver RP, et al. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC Genomics*. 2018;19(1):1-23.
95. Wang Q, Sun M, Zhang Y, Song Z, Zhang S, Zhang Q, et al. Extensive chromosomal rearrangements and rapid evolution of novel effector superfamilies contribute to host adaptation and speciation in the basal ascomycetous fungi. *Molecular Plant Pathology*. 2020;21(3):330-348.
96. Rajeh A, Lv J, Lin Z. Heterogeneous rates of genome rearrangement contributed to the disparity of species richness in Ascomycota. *BMC Genomics*. 2018;19(1):1-13.
97. Xu J, Yan K, Dickman MB, Leslie F. Electrophoretic karyotypes distinguish the biological species of *Gibberella fujikuroi* (*Fusarium* section *Liseola*). *Molecular Plant-Microbe Interactions*. 1995;8:7.
98. Van der Nest MA, Beirn LA, Crouch JA, Demers JE, De Beer ZW, De Vos L, et al. Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntiaella omanensis*, *Leptographium procerum*, *Rutstroemia sydowiana*, and *Sclerotinia echinophila*. *IMA Fungus*. 2014;5(2):473-486.
99. Jurgenson JE, Zeller KA, Leslie JF. Expanded genetic map of *Gibberella moniliformis* (*Fusarium verticillioides*). *Applied and Environmental Microbiology*. 2002;68(4):1972-1979.
100. Slinski SL, Kirkpatrick SC, Gordon TR. Inheritance of virulence in *Fusarium circinatum*, the cause of pitch canker in pines. *Plant Pathology*. 2016;65(8):1292-1296.
101. Shin JY, Bui DC, Lee Y, Nam H, Jung S, Fang M, et al. Functional characterization of cytochrome P450 monooxygenases in the cereal head blight fungus *Fusarium graminearum*. *Environmental Microbiology*. 2017;19(5):2053-2067.
102. Brunke S, Mogavero S, Kasper L, Hube B. Virulence factors in fungal pathogens of man. *Current Opinion in Microbiology*. 2016;32:89-95.
103. Soyer JL, Hamiot A, Ollivier B, Balesdent MH, Rouxel T, Fudal I. The APSES transcription factor LmStuA is required for sporulation, pathogenic development and effector gene expression in *Leptosphaeria maculans*. *Molecular Plant Pathology*. 2015;16(9):1000-1005.
104. Cain CW, Lohse MB, Homann OR, Sil A, Johnson AD. A conserved transcriptional regulator governs fungal morphology in widely diverged species. *Genetics*. 2012;190(2):511-521.
105. Coleman JJ, Mylonakis E. Efflux in fungi: La piece de resistance. *PLoS Pathogens*. 2009;5(6):e1000486.
106. Hückelhoven R. Cell wall-associated mechanisms of disease resistance and susceptibility. *Annual Review of Phytopathology*. 2007;45(1):101-127.
107. De Vos L, Van der Nest MA, Santana QC, Van Wyk S, Leeuwendaal KS, Wingfield BD, et al. Chromosome-level assemblies for the pitch canker fungus, *Fusarium circinatum*. 2021, unpublished.
108. Kim H, Lohmar JM, Busman M, Brown DW, Naumann TA, Divon HH, et al. Identification and distribution of gene clusters required for synthesis of sphingolipid metabolism inhibitors in diverse species of the filamentous fungus *Fusarium*. *BMC Genomics*. 2020;21(1):510.
109. Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004;5:59.

110. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Research*. 2008;18:1979-1990.
111. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. *Methods in Molecular Biology*. 2019;1962:65-95.
112. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*. 1997;25(17):3389-3402.
113. Papadopoulos JS, Agarwala R. COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics*. 2007;23(9):1073-1079.
114. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. 2017;20(4):1160-1166.
115. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*. 2016;33(7):1870-1874.
116. Gel B, Serra E. karyoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*. 2017;33(19):3088-3090.
117. Ridenour JB, Möller M, Freitag M. Polycomb repression without bristles: Facultative heterochromatin and genome stability in fungi. *Genes*. 2020;11(6):638.
118. Gacek A, Strauss J. The chromatin code of fungal secondary metabolite gene clusters. *Applied Microbiology and Biotechnology*. 2012;95(6):1389-1404.
119. Louis EJ. In: Louis EJ, Becker MM, editors. *Subtelomeres*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014.