

Article

Repeat-Induced Point mutations (RIP) drives the formation of distinct sub-genomic compartments in *Fusarium circinatum*

Stephanie van Wyk¹, Brenda D. Wingfield², Nicolaas A. van der Merwe³, Lieschen De Vos⁴, Mkhululi Maphosa⁵, Lizel Potgieter⁶, Emma T. Steenkamp^{7*}

¹Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, South Africa, 0028; stephanie.vanwyk@fabi.up.ac.za

²Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, South Africa, 0028; brenda.wingfield@fabi.up.ac.za

³Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, South Africa, 0028; albe.vdmerwe@up.ac.za

⁴Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, South Africa, 0028; lbahlman@fabi.up.ac.za

⁵Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, South Africa, 0028; mkhululi.maphosa@fabi.up.ac.za

⁶Department of Evolutionary Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany; Environmental Genomics, Christian-Albrechts University of Kiel, Kiel Germany; potgieter@evolbio.mpg.de

⁷Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, South Africa, 0028; emma.steenkamp@fabi.up.ac.za

*Correspondence: emma.steenkamp@fabi.up.ac.za; +27 (12) 420-3937 (ETS)

Abstract: Repeat-Induced Point mutations (RIP) serves as a genome defence mechanism that impedes the deleterious consequences of repeated motifs such as transposable elements in fungi. Genomic regions with RIP are biased for adenosine and thymine transitions and the cumulative influence of RIP is thought to have a considerable impact on genome composition. We investigated the impact of RIP on localized genomic regions and whole-genome sequences for representatives of the pine pathogen, *Fusarium circinatum*. We set out to determine the intraspecific variation in acquired RIP and the role of RIP in the development of diverse *F. circinatum* sub-genomic compartments. The results of the study show that the AT-enriched sub-genomic compartment accounts for *ca.* 97% of the calculated RIP and was further prominent in both core and accessory genomic regions. However, more extensive RIP was observed in the accessory sub-compartment and more variable regions of the genome. Regions with RIP indicated increased intrinsic curvature of the DNA which may influence DNA-protein interactions and may promote constitutive heterochromatin formation. The results show that RIP is an important source of functional novelty and genome variation. RIP contributes to the evolution of the genetic landscape and differentiation of diverse sub-genomic compartments of this important fungal pathogen.

Keywords: Repeat-Induced Point Mutations; RIP; Accessory Chromosome; Genome Compartmentalization; GC Content; Genetic Variation; *Fusarium circinatum*; Transposable Elements; Interstitial Telomeric Repeats; Effector Genes.

1. Introduction

The genome architecture of many fungi in the subphylum Pezizomycotina is characterized by local GC content biases that have important implications for genome evolution [1, 2]. In these fungi, telomeric regions, dispensable chromosomes, and accessory sub-genomic compartments often also display greater variability in genetic composition compared to that of the remainder of the genome [3, 4]. These more variable regions usually represent fast-evolving reservoirs of functional novelty, the

distribution of which often coincides with unusual GC content. However, the exact mechanisms underlying such biased base compositions and the development of unique genomic regions remain ill defined.

A prominent mechanism in fungi that has been linked to variable genomic regions and biased GC content is the Repeat-Induced Point mutation (RIP) pathway [2]. RIP is a genome defence mechanism that causes cytosine to thymine transition mutations in repeated sequences [5]. It functions via a homology dependent mechanism that detects repeated sequences in transposable elements (TEs) and duplicated genomic regions [6, 7]. As a consequence, genomic regions in which RIP is common are usually repeat-rich, GC-poor and these regions can be maintained over long genomic stretches with higher nucleotide sequence variability [3, 8-10]. RIP thus contributes to the formation of distinct GC-rich and AT-rich regions of the genome [2, 8, 11], which correspond broadly to the so-called core and accessory sub-genomic compartments respectively [4]. By contrast, the core sub-genomic compartment is more conserved, repeat-poor and evolves more slowly [11].

RIP has further been implicated in the gene content biases observed in the core and accessory sub-genomic compartments [8, 10]. Although the latter is usually gene-poor, it is thought to contain genes that promote adaptation [3, 12], while the core has an overall higher gene coding capacity, and contains genes that are essential and widely conserved for maintaining basic biological and housekeeping functions [4]. The introduction of RIP may thus lead to the formation of unique genes in repeat-rich regions of the accessory sub-genomic compartment [3, 4, 10, 13, 14]. The RIP process is also “leaky” and may extend into single copy regions that neighbour multicopy genomic regions [8, 10]. Notably, such processes likely have facilitated the diversification of effector genes, avirulence genes and genes coding for small secreted proteins [11, 15, 16]. RIP may accordingly have a significant impact on many aspects of the fungal lifestyle, including reproductive biology, host range and pathogenicity [16-18].

By changing the base composition of DNA sequences, the products of the RIP pathway can also affect the physical properties of DNA polymers [19], as well as the three-dimensional organization of chromatin [20]. Specifically, AT-rich biases caused by the RIP pathway affect the formation of heterochromatin [20]. Similar to the centromeres and telomeres of chromosomes, regions rich in RIP-mutations are maintained in a constitutive heterochromatic state [20]. Therefore, regions with RIP are thought to remain tightly coiled and condensed around nucleoprotein complexes, thereby preventing their participation in recombination [21] and transcription (*i.e.*, heterochromatic silencing) [20]. In terms of the physical properties of DNA polymers, the position of particular nucleotides relative to one another determines intrinsic curvature and flexibility of the DNA [22]. These properties may thus be sequence dependent, where they have important implications for gene regulation [19], nucleosome binding [22, 23], recombination [24], genome stability [25, 26] and DNA replication [25].

Previous work on the economically important pine pathogen [27], *Fusarium circinatum* (*Fusarium fujikuroi* species complex, FFSC [28]), has shown that RIP mutations constitute a substantial proportion of its genome [10]. However, relatively few other studies have set out to determine whether RIP is also associated with the intraspecific variation of this or other fungi, particularly within a whole-genome context. Genome-wide data can provide answers on numerous questions related to intraspecific synteny in regions with biased base composition, the role of RIP in the evolution of distinct genomic compartments, changes to the physical properties of DNA brought about by RIP, and the potential role of RIP in the development and evolution of pathogens such as *F. circinatum*. We therefore investigated the cumulative influence of RIP on the overall base composition in the different genomic regions and sub-genomic compartments of *F. circinatum*. The five specific questions addressed were as follows: i) What was the genomic distribution of RIP in the *F. circinatum* genome? ii) Are the various sub-genomic compartments and chromosomes of *F. circinatum* differentially affected by the RIP pathway? iii) Are there differences in the intrinsic DNA curvature between regions rich in RIP mutations and the remainder of the genome? iv) Is the RIP pathway associated with loss of synteny among the genomes of different *F. circinatum* isolates? v) Does RIP contribute to the diversification of *F. circinatum* effector genes? In order to answer these questions, the genome information for five isolates of *F. circinatum* were utilized, of which four were sequenced specifically for this study. The findings of this research will

allow important conclusions to be drawn regarding the way in which RIP drives genomic variation and contributes to the evolution of the genetic landscape and differentiation of sub-genomic compartments of this and other fungal pathogens.

2. Materials and Methods

2.1. Genome sequences and gene annotations

The genome sequences for five isolates of *F. circinatum* were used in this study (Table 1). Of these, the sequence of FSP34 were available from a previous study [29] and is available from the National Centre of Biotechnological Information (NCBI). The genome sequences for isolates CMWF567, UG27, FSOR, and FFRA were sequenced in the current study, after their species identity was confirmed using phylogenetic analyses (see Figure S1 and associated references). High quality genomic DNA was extracted from seven-day old cultures grown on half-strength PDA as described previously [33]. Genome sequencing for isolates FSOR and FFRA was performed using paired-end (350 bp median insert size) and mate-pair (5 kilobase pairs [Kb] median insert size) libraries on the Illumina HiSeq XTen and Hiseq2000 platforms, respectively, at Macrogen (Seoul, Korea). Those of CMWF567 and UG27 were performed using PacBio long read sequencing (10 Kb and 20 Kb libraries), and Illumina HiSeq2000 250 bp paired-end sequencing.

Following quality filtering and trimming of raw read sequences using CLC Genomics Workbench v. 8.0.1 (CLCBio, Aarhus, Denmark), *de novo* genome assembly was performed using ABySS v. 1.3.7 [34]. Contigs were scaffolded using SSPACE v.2.0 [35] and gaps filled using GapFiller v. 1.11 [36] to yield chromosome-sized scaffolds. These were then assembled into pseudochromosomes (Table S1). This involved ordering and orienting the assembled scaffolds of the newly sequenced genomes by making use of information generated from LASTZ 1.02.00 [37] alignments implemented in Geneious R8 [38] using the FSP34 genome as reference. See Table 1 for DDBJ/EMBL/GenBank accession numbers for the genome sequence data for the five isolates.

Table 1. Summary of the genome information and statistics for the five *F. circinatum* isolates investigated in this study.

Genome information	Isolates of <i>Fusarium circinatum</i>				
	FFRA	FSOR	FSP34	UG27	CMWF567
Geographic origin	Colombia	Colombia	California (USA)	South Africa	Mexico
Reference	This study, [39]	This study, [39]	[29]	This study, [40]	This study, [41]
DDBJ/EMBL/GenBank accession number	JAEOB000000000	JAELVJ000000000	GCA_000497325.3	JAELVK000000000	JADZLS000000000
Size (Mb)	46.83	45.58	45.10	45.69	46.01
GC content (%)	47.1	47.1	47	46.9	46.8
Predicted open reading frames	15 803	16 509	15 517	15 340	14 868
Number of contigs	57	57	24	68	28
BUSCO ¹	99.0	97.8	98.8	96.1	96.5

¹ Benchmarking Universal Single-Copy Orthologs with the Sordariomyceta data set.

The level of completeness was evaluated for the five *F. circinatum* assemblies using the software BUSCO (Benchmarking Universal Single-Copy Orthologs; v2) with the Sordariomyceta data set [42, 43]. All genomes were annotated using CLC software. The annotations were predicted by WebAUGUSTUS v.2.5.5 [44] using gene models for *F. graminearum* (<http://bioinf.uni-greifswald.de/augustus>) together with mRNA sequence data generated for *F. circinatum* FSP34 [29].

To study the presence and distribution of genes associated with the functioning of RIP, DNA methylation and RIP directed heterochromatic silencing (based on reports from *N. crassa*) [10, 45-49], BLASTp and tBLASTn searches were performed on the genomes investigated in this study. Fungal nucleic acid and protein sequences with similarity (Expect-value $< 1 \times 10^{-5}$) to those described in *N. crassa* were identified using CLC software. Although the exact genomic mechanism underlying RIP is still under investigation, the query sequences considered in this study were the two 5-cytosine methyltransferases RID (RIP deficient) and DIM-2 (defective in methylation). In these analyses, we also included the six additional cofactors of the DIM-2 mediated DNA methylation, HP-1 (heterochromatic protein), DIM-5, -7, DNA Damage binding protein 1 (DDB-1) and Cul4 (Cullin-4 protein) [47-49]. Further, a set of 12 effector genes, previously reported in *F. circinatum* isolate FSP34 [50], were identified using nucleotide BLAST searches as described above for the four remaining genome assemblies. These genes were expressed in *planta* and significantly upregulated in the fungus during pathogenesis on pine [50]. Functional annotations, predicted functional domains, protein family (PFAM) membership and Gene Ontology (GO) were evaluated using InterProScan v. 5 [51].

2.2. Genome-wide RIP analysis

For downstream RIP analyses, the five genomes examined here were artificially segmented to generate six types of data sets. The first of these included only the coding content for the five genomes, where the coding sequences predicted by AUGUSTUS was included in separate data sets for each isolate. Similarly, the second data set type included only the sequences for the putative 12 effector genes identified for these fungi.

The third data set type was generated to investigate the effect of RIP on AT-enriched genomic compartments and regions with enriched AT nucleotide combinations. For this purpose, data sets based on GC content (*i.e.*, 10-20%; 20-40%; 40-50%; and 50-100%) were constructed using OcculterCut v. 1.1 [2]. These data sets, comprising of different GC content classifications, were referred to as the GC-light isochore (*i.e.*, 10-20%; 20-40% GC content) and the GC-heavy isochore (*i.e.* 40-50%; and 50-100% GC content), following the notation reported previously for *Leptosphaeria maculans* [8]. OcculterCut was used to investigate genome segmentation, changes in GC content, and mono-, di-, and trinucleotide frequencies for each fungal genome, based on default parameters.

The fourth type of data set contained sequences for the accessory genomic compartment and for the core compartment. To generate these data sets, core and the accessory regions were predicted using NUCmer alignments, implemented in Spine (v0.3.1) [52]. Sequences that were shared among all five genomes with $\geq 95\%$ nucleotide identity across $\geq 10\,000$ bp were considered as core genomic regions, while regions unique to each isolate were recorded as part of the accessory sub-genomic compartment. Core and accessory data sets were accordingly generated for each of the five *F. circinatum* genome assemblies.

To generate the fifth type of data set, genome sequences were separated based on chromosome identity. For this purpose, homologous chromosomes in the various genome assemblies were identified using LASTZ alignments as described before. Accordingly, the data sets contained the respective core chromosome (chromosome 1 to 11) and the dispensable chromosome (chromosome 12) sequences. Because the RIP pathway is induced by repeated sequences, the data for the dispensable chromosomes also included information regarding the distribution of TEs. The latter was identified in the dispensable chromosome assemblies for the five isolates with REPET [53, 54]. Where relevant, REPET annotations determined previously for the FSP34 strain [10] were included in these analyses.

The sixth type of data set included the sequences for putative centromeric loci and telomeric regions. This was done as described previously [55] by using 100% similarity to the known "TTAGGG" telomeric repeat sequence described for *F. circinatum* and other fungi [55, 56]. The putative positions of centromeres were determined by identifying homologous, GC-depleted and extensively RIP affected regions that were common to the five isolates and that occurred in regions homologous to those of the centromeres of the closely related fungi, *F. fujikuroi* and *F. verticillioides* [57]. To confirm the putative locations of *F. circinatum* centromeric regions, the genomic positions of genes flanking the identified

centromeric regions of *F. fujikuroi* and *F. verticillioides* [57] were compared to those occurring adjacent to the putative centromeric regions in *F. circinatum* by using tBLASTn searches implemented using CLC software (Expect-value < 1×10^{-5}).

The whole genome assembly for each isolate, as well as the various data sets containing particular types of sequence information (described above) were subjected to analysis with The RIPper v1.0 [58]. The widely used RIP indices [20, 58-60], were calculated for all sequences using 1000 bp sliding windows with 500 bp steps, except for the data sets containing only the coding sequences and the effector gene sequences that utilized 100 bp windows and 50 bp step sizes. For the RIP analyses performed in this study, stringent parameters were applied (*i.e.*, RIP product index > 1.1, RIP substrate index ≤ 0.75 and RIP composite index > 0) to limit false positives (*i.e.*, regions erroneously suggested to contain RIP), as suggested previously for the Ascomycota [58]. Large RIP Affected Regions (LRARs; *i.e.*, genomic regions spanning at least 4 000 bp or seven consecutive RIP-affected windows) were identified using the “Calculate LRAR” tool of The RIPper by applying default parameters.

2.3. Loss of synteny and collinearity

In order to evaluate whether the RIP pathway can affect synteny and collinearity among homologous genomic regions, ten LRARs from FSP34 were randomly chosen using Randomizer (www.randomizer.org). The nucleic acid sequences, together with their predicted annotations of the selected LRARs, were then used as query sequences in nucleotide BLAST searches and comparisons using CLC software (Expect-value < 1×10^{-5}) to identify corresponding regions in the genomes of the other four fungi. Five Kb regions up and downstream of each of the LRARs were included in these analyses. The results were visualized with Kablam v. 1.0 [61] using the XML file output retrieved from NCBI alignment analyses.

2.4. DNA curvature

In order to explore the potential role of RIP on intrinsic DNA curvature, genomic regions with and without RIP were investigated. Using the information generated above, a set of ten 1000 bp-windows with RIP and ten such windows without RIP were randomly selected from the data generated for FSP34, by making use of Randomizer. In the same way, we also selected ten LRARs in the FSP34 genome. The three data sets were then subjected to DNA curvature analyses using Bend.it [62]. In this analysis, the default window size of 31 bp were used together with bendability parameters originally derived from nucleosome binding and DNaseI digestion data [62]. In addition, these analyses of DNA curvature were also performed on the data sets containing sequences for the GC-light and GC-heavy isochores of FSP34 using the information generated above.

3. Results

3.1. Genome sequences and annotations

The genome assemblies of the five *F. circinatum* isolates had comparable genome statistics (Table 1). The assembled draft genome for the newly sequenced isolates ranged from 45.58 million bases (Mb) (FSOR) to 46.83 Mb (FFRA), with an average GC content ranging of 46.8-47.1%, which was comparable to that of the FSP34 strain of *F. circinatum* (Table 1). The respective assemblies consisted of 24 to 57 contigs. Based on the LASTZ alignments, the four new assemblies contained 12 chromosome-sized scaffolds corresponding to the expected 11 core and 12th dispensable chromosomes (Table S1) [63].

The sets of core pseudochromosome assemblies had a high percentage of nucleotide similarity and sequence coverage and were the expected sizes compared to those known for FSP34 (Table S1) [63]. Nucleotide similarity against FSP34 ranged from 89.3% (chromosome 11, UG27) to 99.7% (chromosome 2, FFRA, for example), and query coverage ranged between 89.7% (chromosome 10, UG27) to 99.7% (chromosome 1; FSOR, for example). The assemblies for the 12th chromosomes were notably larger (*i.e.*, 978 035 bp for UG27 and CMWF567) than that of FSP34 (525 065 bp) and the Colombian strains (*i.e.*,

809 205 bp for FFRA and 805 519 bp for FSOR). The UG 27 and CMWF567 12th chromosome assemblies further displayed reduced sequence similarity (92.8% and 81.5%, respectively) and query coverage (46.5% and 81.5%, respectively) compared to that of the homologous sets of the core chromosomes of FSP34 (Table S1).

The BUSCO results showed that the assemblies were between 96.1% (UG27) to 99.0% (FFRA) complete (Table 1 and Table S2), suggesting that they included most of the expected gene content in these fungi [53, 54]. Based on the WebAUGUSTUS results, these strains encode between 14 868 (CMWF567) and 16 509 (FSOR) genes (Table 1). Of these, candidate homologous sequences for each of the query genes associated with the RIP process and DNA methylation (described for *N. crassa*; Table S3) were identified for the five *F. circinatum* assemblies and were comparable to RIP associated genes identified previously for FFSC species and that of FSP34 [10].

Additionally, putative homologs for the 12 effector genes in FSP34, which were previously shown to be upregulated during pathogenesis on pine [50], were identified in the four new genomes examined (Table S4). Gene sequences sharing high nucleotide similarity in comparable genome locations were identified for the 12 effector genes in each of the isolates (Table S4). Notably, the effector genes, FCIRG_06237 and FCIRG_08174 identified in FSP34 [50] likely originate from a tandem gene duplication event, as they shared similar protein functional domains and were located adjacent to one another (FCIRG_08174 was absent from UG27 and CMWF567). Overall, the InterProScan results showed that most of these genes had similar functional annotations and contained the requisite functional domains as those identified for the FSP34 data set (Table S5). Some differences were observed in that genes contained domains that differed compared to those described in FSP34, which included four genes in FFRA, two in FSOR, two in UG27 and two genes in CMWF567 (see Table S5 for detail).

3.2. Genome-wide RIP analysis

The overall summary of RIP statistics was evaluated for the entire set of genome assemblies for *F. circinatum* (Table 2). These analyses showed that between 5.96% (FSOR) and 7.08% (UG27) of the respective genomes were constituted by RIP mutations. Based on the criteria described previously [64], all *F. circinatum* isolates were classified as 'Moderately RIP Affected', with more than 5% of their genomes constituting RIP mutations.

For the data sets containing coding sequences, our analysis showed that RIP mutations constituted an average of 2.08% of this sub-section of the genome (Table 2, Table S6). However, more than a third (average 34.35%) of all *F. circinatum* genes contained at least one window indicating RIP. In such cases, RIP was often recorded in single windows (50 bp), with rare instances of consecutive RIP-affected windows (Table S6). Furthermore, the GC content of the data sets with coding sequences was higher (average of 51.46%) compared to that of the remainder of the genome (about 47%). Analysis of the data sets containing sequences for 12 putative effector genes of FSP34, showed that each had at least one window indicating RIP (Figure S2), although they were all located in close proximity (within 30 Kb) to regions with RIP or were in regions neighbouring LRARs (Figure S3). For the remaining effector gene data sets in the other fungi, our analyses showed that some genes did not contain evidence of RIP affected (*i.e.*, one in FFRA, and one in CMWF567) (see Figure S2 for details).

As expected for RIP-capable fungi, the frequency and distribution of GC content across the *F. circinatum* genomes was bimodal, with a major peak around 50% GC and a minor peak around 20% GC (Figure S4). In order to investigate the occurrence and extent of RIP in the AT-enriched genomic compartments, the *F. circinatum* genomes were separated into data sets representing GC-light isochores (0-20% and 20-40% GC content) and GC-heavy isochores (40-50% and 50-100%). Regions defined as "isochore" are genomic regions that are alike in their DNA base compositions bias (either GC- or AT-rich), which are maintained across the length of the isochore [8]. On average, about 3.9% and 3% of the five genomes were characterized as part of the GC-light isochore, corresponding to genomic regions with 0-20% and 20-40% GC content, respectively (Figure S5). By contrast, 83.2% and 10.2% of the respective *F. circinatum* genomes constituted the GC-heavy isochore, corresponding to genomic regions

with 40-50% and 50-100% GC content, respectively. For the GC-heavy isochore, di- and trinucleotide frequencies remained similar (Figure S5). In comparison, the di- and trinucleotide sequences in GC-light isochore were enriched for AT-rich combinations, comparable to that of other FFSC fungi [10].

Table 2. Summary of RIP statistics of *Fusarium circinatum* strains investigated in this study.

RIP statistics	Isolates of <i>Fusarium circinatum</i>				
	CMWF567	UG27	FSOR	FFRA	FSP34
Proportion (%) of genome with RIP	6.69	7.08	5.96	5.97	6.39
Number of LRARs	163	156	156	165	157
Average size (bp) of LRARs	16 056	16 831	15 177	15 300	15 537
Average % GC content of LRARs	20.23	20.24	20.17	20.02	20.8
Genomic proportion(bp) of LRARs	2 617 051	2 625 607	2 367 563	2 524 500	2 517 015
Proportion (%) of coding compartment that constitutes RIP	2.40	1.96	2.02	2.02	2.01
RIP affected proportion (%) of genomic regions with 0-20% GC	98.96	99.16	97.66	97.89	99.09
RIP affected proportion (%) of genomic regions with 20-40% GC	81.22	86.66	78.83	81.54	83.07
RIP affected proportion (%) of genomic regions with 40-50% GC	0.25	0.21	0.29	0.28	0.23
RIP affected proportion (%) of genomic region with 50-100% GC	0.05	0.04	0.06	0.03	0.09
Proportion (%) of RIP in the core genomic compartment	2.27	2.27	2.12	2.10	2.27
Proportion (%) of RIP in the accessory genomic compartment	36.27	48.48	32.87	31.23	45.24

Further analyses of the GC-light isochores showed that the majority of these sequences had RIP mutations (Table 2; Figure 1; Table S7). Between 97.66 and 99.16% of the 0-20% GC content data set were constituted by RIP mutations, as compared to 81.22% to 86.66 of the 20-40% GC content data set. Together, the total proportion of RIP in GC-light isochore represented the greater majority (*ca.* 97%) of all recorded RIP mutations. By contrast, for the data sets containing sequences representing the GC-heavy isochore, RIP mutations only constituted 0.31%, and 3.21% of the overall observed RIP recorded for *F. circinatum*. Further, the GC-light isochores contained LRARs, while LRARs were completely absent from the GC-heavy isochore sequences.

To evaluate the occurrence and extent of RIP in the core and accessory genomic compartments of *F. circinatum*, alignment-based analyses were performed to distinguish genomic regions present in all five isolates (core compartment) and those genomic regions that are unique to each isolate (*i.e.*, regions that are part of the accessory compartment) (Figure 1). The size genome unique to an isolate ranged from 3.0 Mb in FSP34 to 3.9 Mb in CMWF567, while the average size of the core genomic compartment was *ca.* 39.91 Mb (Table S8). Both the core genomic compartment and the parts of the genomes unique to an isolate contained RIP mutations (Table 2; Table S8; Figure 1). RIP mutations constituted an average of 2.17% of the overall core compartment, which ranged from 2.09% in FSP34 though to 2.27% in UG27

1

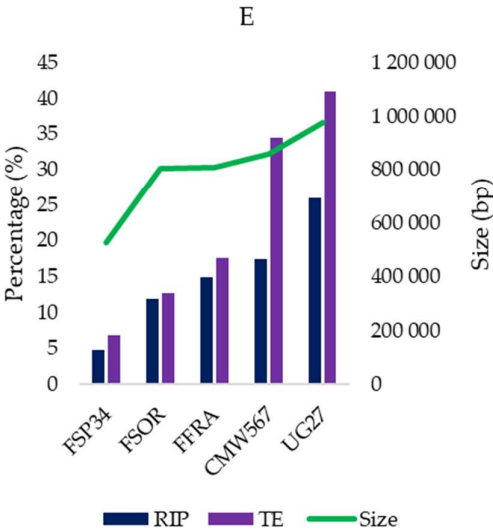
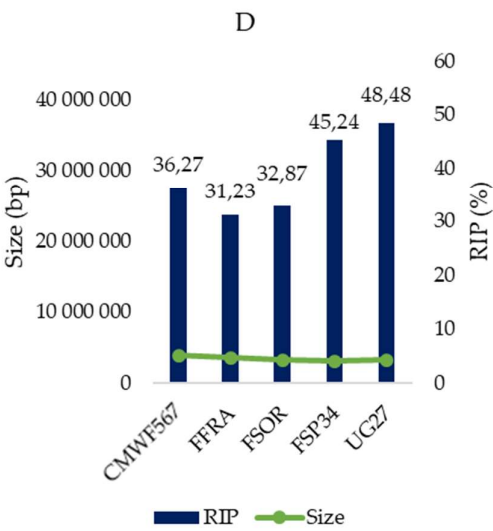
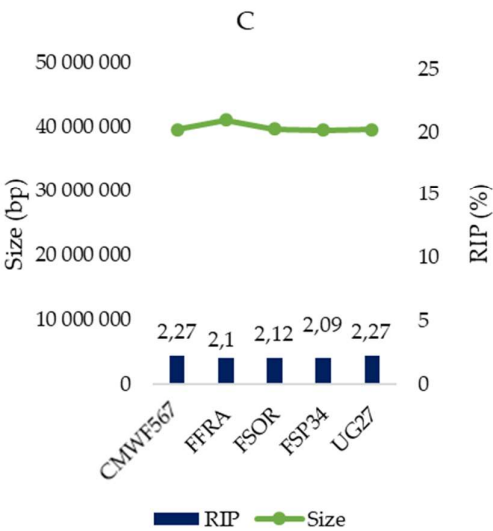
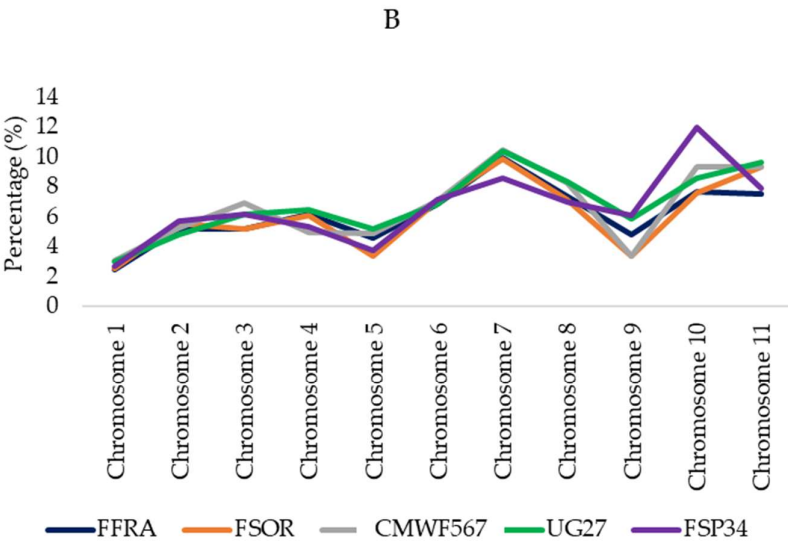
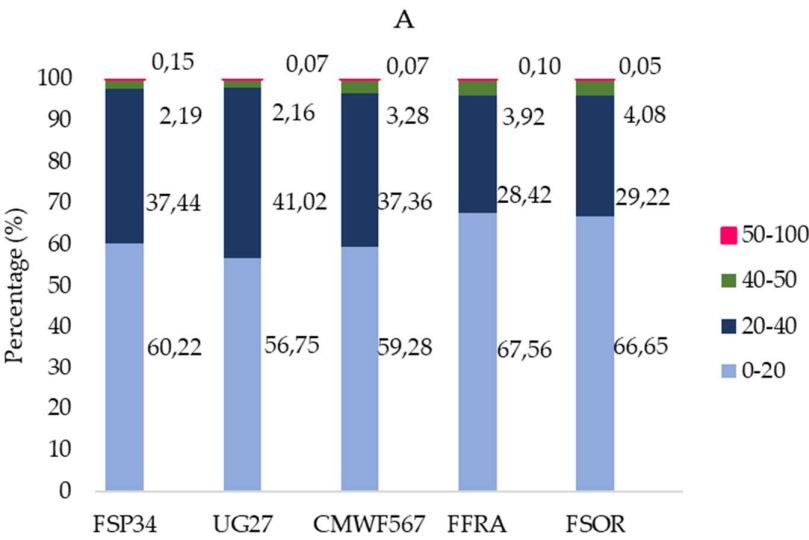


Figure 1: Summary of RIP statistics in the five *F. circinatum* data sets investigated. (A) Sum of the total RIP observed for the five *F. circinatum* strains (expressed as percentage) depicted for each GC sub-genomic components. (B) Proportion (%) of the core chromosomes affected by RIP, for the five *F. circinatum* isolates investigated in this study. (C-D) Summary of the total genomic proportions (blue; percentage [%]) constituting RIP and total size (bp) of the core (C) and accessory (D) sub-genomic compartments of *F. circinatum* isolates investigated in this study. (E) Percentage of the dispensable chromosome sequences affected by RIP (y-axis; blue) and the total number of Transposable Elements (TEs) for each isolate investigated (right x-axis; purple). The size of each dispensable chromosome sequence is indicated in base pairs (left x-axis; green).

and CMWF567. However, in the regions unique to a specific isolate, the proportion constituting RIP mutations ranged from 31.23% in FFRA to 48.48% in UG27.

Homologous sets of *F. circinatum* chromosomes shared high degrees of sequence similarity (Table S1) and were also very similar in their RIP mutation content (Figure 1). RIP constituted an average of 6.45% of the core chromosomes and varied from 2.48% for chromosome 1 of FFRA to 11.99% for chromosome 10 of the CMWF567 strain. The dispensable chromosome of the isolates examined (*i.e.*, chromosome 12) were comparatively much richer in RIP mutations (Figure 1), where these mutations constituted between 11.85% (FSOR) and 25.96% (UG27) of the respective assemblies (Table 3). However, as reported previously [10], the RIP affected proportion of the 12th chromosome of FSP34 was much lower (4.85%). The total proportion of the dispensable chromosomes that constituted RIP also showed a strong positive correlation to both chromosome size and TE density (correlation coefficient [r]=0.9 and 0.9 respectively; Figure 1; Table S9).

Table 3. RIP statistics for the 12th chromosome assemblies of the *F. circinatum* strains analysed.

RIP statistics	Isolates of <i>Fusarium circinatum</i>				
	CMWF567	UG27	FSP34	FSOR	FFRA
Size (bp)	978 035	978 035	525 065	805 519	809 205
GC content (%)	42.75	41.06	46.36	44.48	43.61
RIP affected proportion (%)	17.43	25.96	4.85	11.85	14.95
Number of windows affected by RIP	299	506	50	188	242
Number of LRARs	6	11	1	9	7
Total size of LRARs (bp)	139 500	248 535	10 000	88 500	106 500

The ends of almost all chromosome scaffolds of the five *F. circinatum* genome assemblies investigated were characterized by reduced GC content (*ca.* 28.21%) and high frequencies of the telomeric repeat sequence (Table 4 and Table S10, Figure 2 and Figure S6). The only exceptions were the left end of Chromosome 2 of FSOR and FFRA (Figure S6). The average number of telomeric repeats (per isolate) at the chromosome ends varied from 62 (FFRA) to 104 (UG27). These putative telomeric regions were also rich in RIP mutations (Table 4; Table S10; Figure S6), where the average size of such RIP-affected regions was *ca.* 33.98 Kb. Surprisingly, however, telomeric repeats were also enriched for in many of the LRARs in the *F. circinatum* chromosomes investigated (Figure 2; Figure S6).

The putative locations of all 11 centromeric regions of the *F. circinatum* core chromosomes sequences were identified (Table S11). These genomic positions corresponded to those previously characterized in the closely related species, *F. fujikuroi* and *F. verticillioides* [57]. The data showed a high level of conservation in the sequences surrounding the centromeric regions in *F. fujikuroi*, *F. verticillioides* and *F. circinatum* (Table S12). However, as expected [65], the reciprocal chromosomal translocation between the distal ends of chromosome 8 and 11, reported previously [65], were the only exception recorded that deviated from the observed synteny and collinearity between these species. In *F. circinatum*, the putative centromeric regions were on average *ca.* 36 Kb in size, ranging

Table 4: Summary of RIP statistics and genetic features of telomeric regions identified in the five isolates of *Fusarium circinatum*.

RIP statistics and features	Isolates of <i>Fusarium circinatum</i>				
	FSP34	FSOR	FFRA	UG27	CMWF567
Number of telomeres identified	22	23	23	23	23
Average number of telomeric repeats ¹	83	66	62	104	77
Average size (bp)	42 477	32 283	30 065	41 152	24 386
RIP product index value ²	1.50	1.61	1.53	1.62	1.68
RIP substrate index value ²	0.50	0.45	0.45	0.41	0.39
RIP Composite index value ²	1.00	1.16	1.04	1.21	1.29
GC content (%)	29.18	28.73	29.1	27.26	26.82

¹Telomeric repeat sequence= 5'-TTAGGG-3' and 3'-CCCTAA-5' [56]. ²RIP product index values $[(TpA/ ApT) > 1.1]$, substrate index values $[(CpA + TpG/ ApC + GpT) 0.75 \geq x]$, and composite index values $[(TpA/ ApT) - (CpA + TpG/ ApC + GpT)] x > 0$ are indicative of RIP [58].

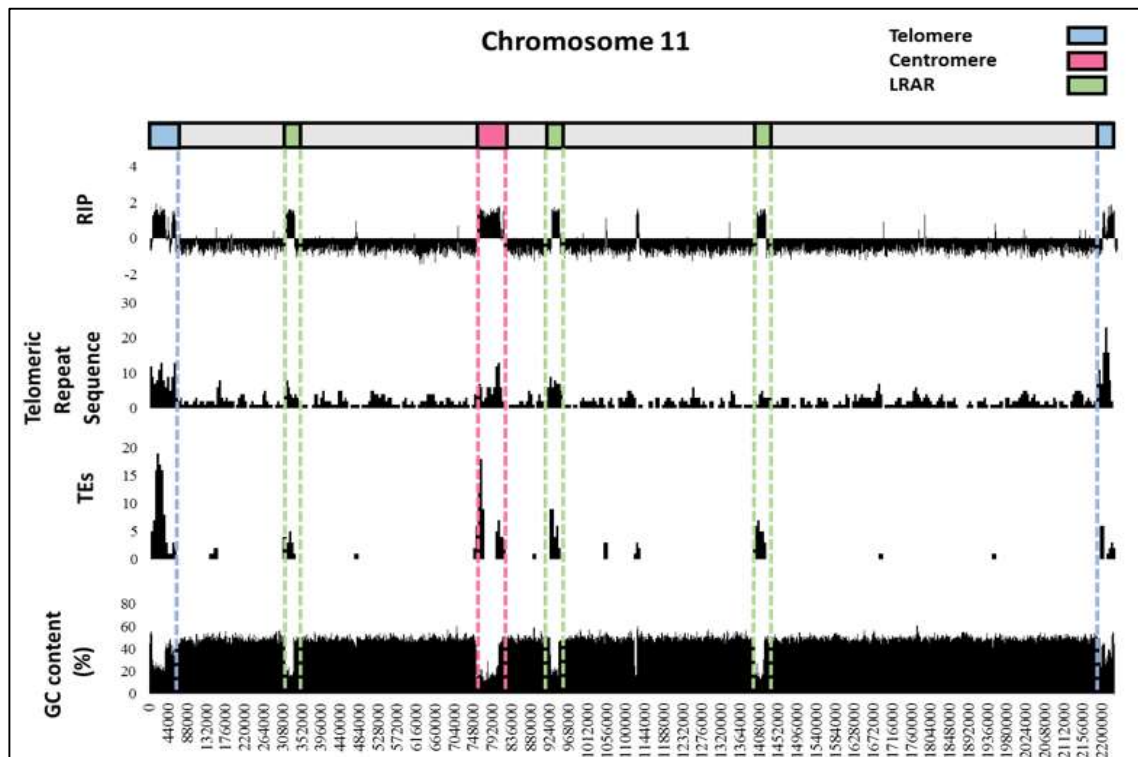


Figure 2: Summary of the genetic features of chromosome 11 of *F. circinatum* (FSP34). The first panel represents the chromosome showing putative positions of the telomeric and centromeric regions, as well as Large RIP affected Regions (LRAR; > 10 000 bp). The second panel depicts the changes in RIP composite index values across the length of chromosome using a 1000 base pairs (bp) window and increments of 500 bp. Values above 0 indicate RIP. The third panel depicts a summary of changes in the number of telomeric repeat sequences (5'-TTAGGG-3') calculated for every 10 Kb and 5 Kb increments. (D) The fourth panel depicts the distribution of TEs and repeated sequences (> 100 bp) across the length of the chromosome, (10Kb and 5Kbp increments) window. The bottom panel depicts the changes in % GC content calculated with a 1000 bp step-size and 500 bp increments.

from 16.4 Kb in chromosome five of UG27 to 117 Kb in chromosome eight of UG27 (Table S11, Figure S6). In all cases, these regions were extensively affected by RIP and had reduced GC content at an average of 17.1% (Table S11).

The results of this study showed that the *F. circinatum* isolates had many instances of consecutive RIP-positive windows, and that these extensively RIP affected regions were similar in size and mostly had comparable genomic locations for each homologous set of chromosomes (Table S13). For example, the LRARs of homologous sets of chromosomes 5 had similar distribution patterns, occurred in comparable genomic locations and generally had similar sizes. Regional biases amongst the position of LRARs in specific chromosomes were also observed (Table S13). For example, in chromosome 5 of FSP34, seven predicted LRARs were recorded and the same chromosome in UG27 had 14 LRARs. Overall, between 156 (FSOR and UG27) to 165 (FFRA) LRARs were identified among the five genomes examined (Table 2), where the average size varied between *ca.* 15 177 bp (FSOR) and 16 056 bp (CMWF567). These LRARs further constituted large proportions of the individual genome assemblies (Table 2). For all isolates, the LRARs displayed reduced GC content compared to that of the remainder of the genome sequences (Table 2). Numerous (6-11) LRARs were also recorded in each of the assemblies for the 12th chromosome (Table 3), although that of FSP34 contained only one, as reported previously [10].

3.3. Loss of synteny and collinearity

Homologous regions associated with ten randomly chosen LRARs were investigated in the five genomes to evaluate whether RIP affected synteny and collinearity in these regions. Except for one of the LRARs that were absent from the CMWF567 genome, homologous locations and regions for all 10 LRARs were identified in the five genomes (Table S14). Overall, the homologous loci were characterized by high levels of nucleotide similarity (91%-100%), but such similarity could stretch across as little as 10% of the compared regions or could span the entire sequence. In other words, the overall query coverage ranged from 10% (*e.g.*, the LRAR on chromosome 6 of UG27) to 100% (*e.g.*, the LRAR located on chromosome 11 of UG27). Two of the ten LRARs (*i.e.*, “LRAR 2” on chromosome 2 and LRAR 3 on chromosome 4; Figure S7) had overall high levels of nucleotide similarity and had no changes in synteny and collinearity in the LRAR and neighbouring genomic regions (*e.g.* Figure 3). By contrast, the remaining eight LRARs all showed substantial changes in synteny in the regions with RIP, as well as the regions neighbouring them (*e.g.* Figure 3). The genomic regions neighbouring the LRARs that showed changes in synteny included LRARs numbered 1, and 5-8, that were located respectively on, chromosomes 1, 7, 6, 8 and 10 of the FSP34 assembly (Table S14). The neighbouring regions of LRARs numbered 2 and 3 on chromosomes 2 and 4, respectively, showed no changes in synteny and gene order.

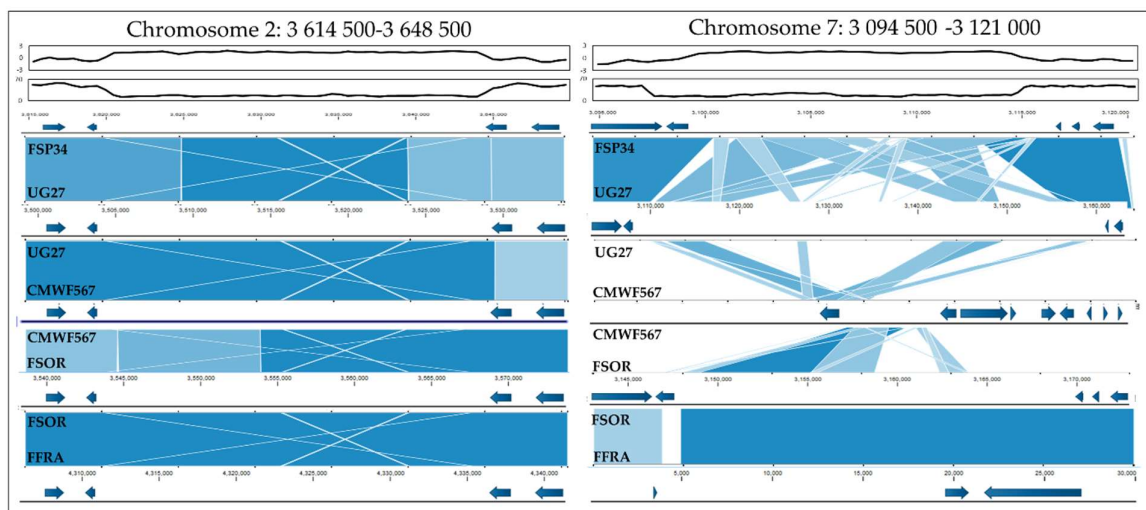


Figure 3. Illustration of changes in RIP, GC content, gene content and nucleotide similarity in two Large RIP affected genomic regions (LRARs) and neighbouring genomic regions of FSP34 relative to homologous regions in other isolates of *F. circinatum*. An LRAR on the 2nd chromosome is indicated on the left, and an LRAR on the 7th chromosome is shown on the right. The top two panels illustrate changes in RIP composite index values (values above 0 indicate RIP), followed by changes in GC content (%) across the particular genomic region. Below them are shown the synteny plots, based on NCBI pairwise alignments (visualized using Kablamo software), for homologous regions for each isolate. The results for all 10 LRAR are illustrated in Figure S and a full list of alignment statistics are listed in Table S13.

3.4. DNA curvature

In order to evaluate the influence of RIP on sequence-dependent DNA curvature, an analysis of the intrinsic curvature distribution was carried out on *F. circinatum* (FSP34) sequences with and without RIP (Table S15, Figure 4). Sequences without RIP had an average intrinsic curvature value of 3.5 degrees per helical turn and varied from 2.7 to 3.95 degrees per helical turn. The regions with RIP had higher average intrinsic curvature values, with an average of 4.55 degrees per helical turn that varied from 3.47 to 4.62 degrees per helical turn. Overall, these sequences showed a non-symmetrical frequency distribution, which is consistent with AT-enriched regions in other organisms [19, 66]. Remarkably, a clear shift towards higher intrinsic curvature values were recorded in *F. circinatum* sequences rich in RIP (Figure 4).

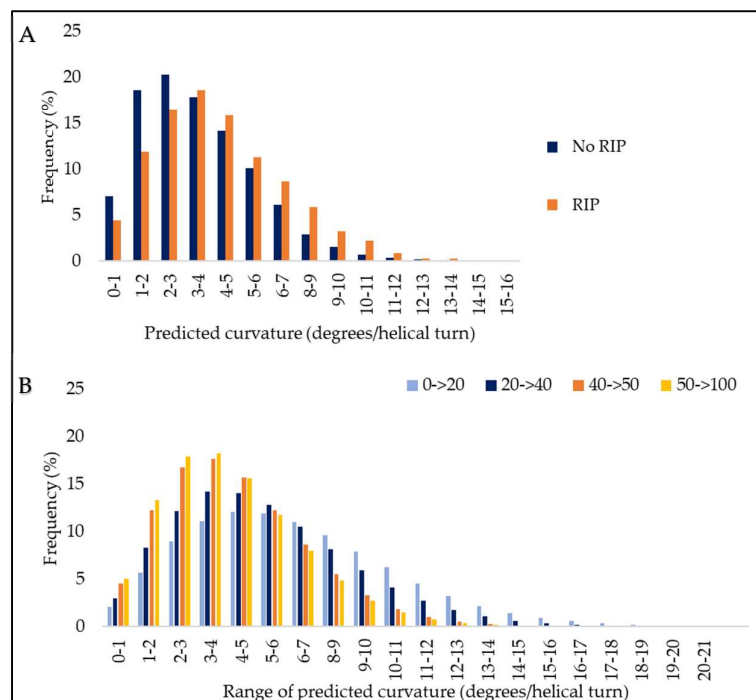


Figure 4. Frequency distribution (%) of the ranges of predicted curvature values (degrees/helical turn) for sequences (each a 1000 bp in size) of FSP34. **(A)** Frequency of predicted curvature values (%) for genomic regions with RIP (orange) and without RIP (blue). **(B)** Frequency distribution of the ranges of predicted curvature values for genomic regions of different GC content ranges for isolate FSP34.

The intrinsic curvature values recorded for the ten LRARs of FSP34 were similar to those recorded for regions affected by RIP (Table S16). Detailed analyses of the change in intrinsic curvature performed on the LRAR located at 1 035 000-1 042 500 bp on chromosome 1 of FSP34 and the regions flanking it, also revealed that high RIP and AT content were associated with increased DNA

curvature (Figure S8). By contrast, the regions outside of the LRAR (with higher GC content) had lower curvature values. The latter was also consistent with results for the DNA curvature analysis of the GC-light isochore components and GC-heavy isochore components of FSP34 (Figure 4). The GC-light isochore had an average intrinsic curvature value of 6.26 degrees per helical turn that varied from 0.006 to 24.31 degrees per helical turn. The GC-heavy isochore component had an average intrinsic curvature value of 4.06 degrees per helical turn and varied from 0.002 to 19.26 degrees per helical turn. The GC-light isochore components had an overall greater frequency of sequences with higher curvature values per helical turn, compared to those of GC-heavy isochore components.

4. Discussion

Intraspecific, genome-wide RIP analyses showed that the five *F. circinatum* genomes investigated harboured extensive evidence of RIP. More specifically, we obtained RIP index values indicative of RIP [10, 20, 59, 60] for the five genomes, and in each genome GC content was also bimodally distributed [2, 10]. The five genome further encoded for the gene thought to underly the RIP pathway, including the two 5-cytosine methyltransferases, DIM-2 (and associated co-factors) and RID [10, 45, 48, 67]. RIP mutations were also more prevalent in the GC-light isochores and accessory sub-genomic compartment of *F. circinatum*. Comparable to other fungi with RIP, the dispensable chromosome, centromeres and telomeres of *F. circinatum* were rich in RIP [9, 10, 64]. Based on RIP mutation content of the five isolates examined, *F. circinatum* belongs to the ‘Moderately RIP Affected’ class of RIP-capable Ascomycetes [64], which also includes other Sordariomycetes such as the closely related *Fusarium pininemorale* and *Fusarium temperatum* [10, 68], and more distantly related fungi such as *Verticillium dahlia* and *Trichoderma reesei* [64].

In this study, we show that RIP can accumulate in both the core and dispensable chromosomes of genetically diverse individuals of *F. circinatum*. Previous studies indicated that RIP extensively affected the dispensable chromosomes of other FFSC fungi, but not that of the *F. circinatum* isolate examined at the time (*i.e.*, FSP34) [10]. The current study showed that this reduced level of RIP might be linked to the dearth of TEs on the 12th chromosome of FSP34, and that RIP is in fact common in this dispensable chromosome of *F. circinatum*. Indeed, the combined actions of TEs and the RIP pathway probably contributed to the isolate-specific differentiation and reduced nucleotide similarity observed for these chromosomes from different isolates of this fungus, although other evolutionary processes such as horizontal gene transfer cannot be excluded. RIP is common in the dispensable chromosomes of fungi, because their repetitive nature provides prime substrates that induce the RIP pathway [4]. For example, the highly repetitive dispensable chromosomes of the Dothideomycetes species, *Zymoseptoria tritici*, is also rich in RIP [9]. As in *F. circinatum*, the dispensable chromosomes of *Z. tritici* are also characterised by many instances of large-scale, isolate-specific sequence differentiation due to the combined actions of TE integration and RIP [9].

The findings presented here demonstrate that RIP affects the core and accessory sub-genomic compartments of *F. circinatum* to a strikingly different extent. The RIP pathway is widely credited for enhancing regional and chromosomal variation [3, 4, 10, 13, 69], but no previous study has set out to quantitatively investigate the extent of RIP beyond gene-level analyses in the accessory sub-genomic compartment of a fungus. Our findings show that large proportions of this sub-genomic compartment of *F. circinatum* were constituted by RIP mutations. The RIP pathway likely represents an important driver of functional novelty in the accessory sub-genomic compartments in isolates of *F. circinatum*. This is because the accessory genomic compartments, dispensable chromosomes, and more variable regions of chromosomes, typically harbour genes contributing to various aspects of fungal lifestyle, including host range and pathogenicity [3, 14, 67].

RIP mutations resulting from RIP leakage into regions outside TEs/repeats (*i.e.*, RIP mutations acquired beyond the bounds of the intended targets), most likely contributed to the changes in synteny and collinearity observed in regions neighbouring LRARs in *F. circinatum*. These changes could also reflect the actions of TE integration, which in turn induced RIP. Also, many of the effector genes previously found to be upregulated during pathogenesis on pine [50] were located on genomic regions in close proximity to LRARs and/or within regions with extensive RIP. These putative effector

genes varied in their predicted protein functional domain, and this variation was often isolate-specific. This situation is thus reminiscent from what was observed in quantitative RIP analyses in *N. crassa*, which showed that RIP can leak up to 40 Kb from its intended target [7]. RIP leakage is also known from other RIP capable fungi and is thought to contribute to the diversification of genes underlying host-pathogen interactions [8, 69]. The results presented here are also consistent with studies showing that RIP is acquired in an isolate-specific manner [8, 69, 70]. A recent study of *Fusarium xylarioides* (a close relative of *F. circinatum*) reported RIP-driven diversification were more prevalent amongst putative effector genes than for any other type of gene classification [71]. Together these data suggest that RIP is an important source of variation in gene regions underlying important biological traits, pathogen development and host-pathogen interactions in *F. circinatum* and other fungi.

Apart from driving variation in specific gene regions, RIP may also be an important source of genome-wide variation in populations of *F. circinatum*. Large chromosomal regions rich in RIP (*i.e.*, LRARs) occurred across chromosomes where they had patchy genomic distributions and were frequently characterized by reduced nucleotide similarity among isolates, as well as changes in gene order and synteny. The latter often also extended into regions neighbouring such RIP-dense areas of the *F. circinatum* genome. However, many of these long stretches of RIP affected regions occurred in homologous positions, which suggests that these regions, and LRARs specifically, were maintained over evolutionary time in different *F. circinatum* strains. Similar RIP distribution patterns have been reported in Dothideomycetes such as *L. maculans* [8] and *Z. tritici* [17, 72] and in Sordariomycetes such as *Neurospora tetrasperma* [73, 74] and *V. dahlia* [4]. In *L. maculans*, extensive RIP was shown to drive isolate-specific diversification [8, 70]. Similarly, species-specific RIP directed sequence diversification was recorded in the FFSC and in different isolates of *N. tetrasperma* [74, 75]. In the same way, the cumulative influence of RIP thus may drive the independent divergence of *F. circinatum* sequences, thereby contributing to genetic variation of different strains or populations of *F. circinatum*.

Most of the RIP in the *F. circinatum* genome is accommodated in its GC-light isochore. The current study provides the first quantitative examination of RIP in this isochore of any fungus, and the findings support the view that the RIP drives the development of the GC-light isochore in RIP-capable fungi [2]. These results presented are in full agreement with reports where surrogate methods were employed (*i.e.*, those analysing changes in genome-wide GC content distribution) for investigating the cumulative influence of RIP [2]. Taken together, our findings suggest that RIP drives localized GC depletion but that it also has a quantifiable influence on the overall genome composition of a fungus in that it underpins the formation and potentially the maintenance of distinct sub-genomic environments within *F. circinatum*.

RIP is particularly prominent in genomic regions that correspond to the telomeres and centromeres of the *F. circinatum* chromosomes. Detailed investigation of LRARs showed that the location of these regions coincided with a consistent increase in frequency of telomeric repeat sequences in the predicted telomeres and centromeres of the fungus. These analyses also revealed the presence of interstitial telomeric motifs in the *F. circinatum* chromosomes that were also enriched for RIP. Interstitial telomeric motifs are regions enriched for telomeric repeats occurring in locations other than the telomeres and centromeres of chromosomes and are typically found in genomic regions that have undergone double stranded DNA breaks [56]. Such breaks can occur due to aberrant recombination between TEs [76, 77], a process in which telomeric repeat sequences are thought to be important signatures for directing DNA repair mechanisms [56]. This is consistent with the fact that RIP was prevalent in regions of the *F. circinatum* genome containing high densities of TEs [10]. In other words, regions with high TE density may be more vulnerable to double stranded DNA breaks, causing increased incorporation of telomeric repeat sequences by the telomerase enzyme, whilst the TEs induce RIP.

The results of this study suggest that the RIP pathway may have substantial influence on the genomic composition of *F. circinatum* and may bring about a range of functionally important genetic changes, thereby influencing chromosomal architecture, epigenetic regulation and higher order

chromatin organization. Other fungi that have GC-light isochores enriched for RIP show unique 5-cytosine methylation patterns [10, 57, 78]. These distribution patterns, which correspond with the RIP driven AT-enriched genomic regions, have been suggested to reflect the “mosaic pattern” of DNA methylation, which differ from that observed in most plant and animal lineages [78]. Additionally, RIP-associated methylation of histones and cytosines are important processes for directing heterochromatin formation [20, 46], which guides heterochromatic silencing and suppressed recombination of RIP targeted regions. In this way, the RIP pathway may have key roles in the evolution of genome architecture and the epigenetic landscape, to ultimately contribute to higher order chromatin structure of *F. circinatum* chromosomes.

Finally, our study showed that the RIP pathway may influence the physical properties of the targeted DNA. The *F. circinatum* sequences with RIP displayed an overall shift to higher curvature values compared to those of regions lacking these mutations, which was also true for its extensively RIP affected GC-light isochore. Such increased curvature was also observed across the length of LRARs. Changes in DNA curvature often correlates with GC depleted regions in some organisms [66, 79–81]. Sequence-dependent DNA curvature changes, particularly in regions biased towards AT nucleotides, are thought to increase the flexibility of the DNA, thereby making it more amenable to certain DNA-protein interactions [79]. For example, the increased curvature of the AT-rich centromeric regions of *Saccharomyces cerevisiae* is essential for facilitating centromere complex formation and assembly in this fungus [79]. Additionally, DNA curvature and AT-rich DNAs play key roles in chromatin organisation and higher order chromatin structure, because the intrinsic curvature and flexibility of DNA determine the ease that DNA would wrap around the nucleosome octamers and the cellular energy needed for the process [22]. Although other DNA-protein interactions can also alter the spatial properties of genomic regions [22, 24], our findings suggest that sequence-dependent curvature and sequence composition may contribute to chromatin structure and three-dimensional organisation of the *F. circinatum* genome.

5. Conclusions

Intraspecific evaluation of the genome-wide occurrence of RIP has provided valuable insights on the manner that RIP influences the genetic composition and genome architecture of *F. circinatum*. RIP mutations in *F. circinatum* were likely caused by induction of the RIP pathway due to TE integration and intragenomic gene duplication, as well as the leakage of the pathway’s effects to regions outside of repeated areas. RIP further leads to the formation and possibly maintenance of the GC-light isochore of *F. circinatum*, and its cumulative effects on nucleotide base composition likely affect the physical properties of the targeted regions. Despite being a prominent feature of all the *F. circinatum* genomes examined, the higher occurrence and extent of RIP in the various accessory genomic compartments, together with the concomitant changes in synteny and collinearity, highlights the role of RIP in driving intraspecific genome variation in this fungus. The fact that the impact of this pathway also occurred in genes underlying host-pathogen interactions further suggest RIP may be a key driver of pathogen development of this economically important fungus.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/, Figure S1: Phylogenetic relationships of *F. circinatum* strain investigated in this study; Figure S2: RIP analyses of effector gene data set; Figure S3: Genomic location of effector genes investigated in this study in relation to the genetic features of the chromosome assemblies of *F. circinatum* (FSP34); Figure S4: Frequency distribution of the GC content of the *F. circinatum* strains investigated in this study; Figure S5: Changes in nucleotide frequencies per GC content range; Figure S6: Graphical representation of the genetic features of the core 11 *Fusarium circinatum* chromosome assemblies for each isolate investigated in this study; Figure S7: Illustration of changes in RIP, GC content, gene content and nucleotide similarity in FSP34 Large RIP affected genomic regions (LRARs) and neighbouring genomic regions in comparison to that of homologous regions in other isolates of *F. circinatum*; Figure S8: Summary of physical features of the LRAR at locus 1 035 000–1 042 500 bp on chromosome 1 of FSP34, relative to its RIP mutation and GC content. Table S1: Summary of alignment statistics and the sizes of *F. circinatum* chromosome assemblies investigated in this study; Table S2: Summary of the BUSCO analyses of *F. circinatum* isolates investigated in this study; Table S3: Table S3: The genomic location of the *F. circinatum* RIP and RIP associated DNA methylation associated genes; Table S4: Name and genomic locations of *F. circinatum* sequences

with similarity to that of the FSP34 query set of effector genes; Table S5: InterProscan results of the genes sharing nucleotide similarity to that of the effector gene data set of FSP34; Table S6: RIP statistics of the coding genomic compartment of the *F. circinatum* isolates investigated; Table S7: RIP statistics of the GC-light and GC heavy isochore components of the sequenced isolates of *F. circinatum* isolates investigated in this study; Table S8: RIP statistics of the core and accessory genomic compartments of the *F. circinatum* strain investigated in this study; Table S9: Correlation coefficients calculated for percentage RIP, count of Transposable elements (TEs) and size (bp) for the dispensable chromosome assemblies of *F. circinatum* strains investigated in this study; Table S10: List of genomic regions corresponding to the telomeric regions in *F. circinatum* isolates investigated in this study; Table S11: Summary of the RIP statistics for the putative genomic locations of the centromeric regions of *F. circinatum* isolates investigated in this study; Table S12: Nearest genes to *F. circinatum* centromeres and synteny with *Fusarium fujikuroi* and *Fusarium verticillioides*; Table S13: List of Large RIP Affected Regions (LRARs) calculated for each genome assembly of *F. circinatum* strains investigated in this study; Table S14: List of NCBI nucleotide alignment statistics using ten FSP34 Large RIP affected Regions (LRARs) as query sequence; Table S15: List of RIP statistics and predicted sequence-dependent curvature of genomic regions of FSP34 that had RIP or were not affected by RIP; Table S16: RIP statistics and predicted sequence-dependent curvature of Large RIP Affected Regions (LRARs) investigated in this study.

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

Funding: This research was funded by the University of Pretoria, the Department of Science and Innovation (DSI) and the National Research Foundation (NRF) via the Centre of Excellence in Plant Health Biotechnology and the South African Research Chairs Initiative (SARChI) Chair in Fungal Genomics. The APC was partially funded by the University of Pretoria.

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.

References

1. Raffaele, S. and S. Kamoun. "Genome evolution in filamentous plant pathogens: Why bigger can be better." *Nature Reviews Microbiology* 10 (2012): 417-30.
2. Testa, A. C., R. P. Oliver and J. K. Hane. "Occultercut: A comprehensive survey of AT-rich regions in fungal genomes." *Genome Biology and Evolution* 8 (2016): 2044-64.
3. Croll, D. and B. A. McDonald. "The accessory genome as a cradle for adaptive evolution in pathogens." *PLoS Pathogens* 8 (2012): e1002608.
4. Bertazzoni, S., A. H. Williams, D. A. Jones, R. A. Syme, K.-C. Tan and J. K. Hane. "Accessories make the outfit: Accessory chromosomes and other dispensable DNA regions in plant-pathogenic fungi." *Molecular Plant-Microbe Interactions* 31 (2018): 779-88.
5. Selker, E. U., E. B. Cambareri, B. C. Jensen and K. R. Haack. "Rearrangement of duplicated DNA in specialized cells of *Neurospora*." *Cell* 51 (1987): 741-52.
6. Cambareri, E. B., B. C. Jensen, E. Schabtach and E. U. Selker. "Repeat-Induced GC to AT mutations in *Neurospora*." *Science* 244 (1989): 1571-75.
7. Ireland, J. T., A. T. Hagemann and E. U. Selker. "High frequency Repeat-Induced Point mutation (RIP) is not associated with efficient recombination in *Neurospora*." *Genetics* 138 (1994): 1093-103.

8. Rouxel, T., J. Grandaubert, J. K. Hane, C. Hoede, A. P. Van de Wouw, A. Couloux, V. Dominguez, V. Anthouard, P. Bally and S. Bourras. "Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations." *Nature Communications* 2 (2011): 1-10.
9. Croll, D., M. Zala and B. A. McDonald. "Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen." *PLoS Genet* 9 (2013): e1003567.
10. van Wyk, S., B. D. Wingfield, L. De Vos, N. A. van der Merwe, Q. C. Santana and E. T. Steenkamp. "Repeat-Induced Point mutations drive divergence between *Fusarium circinatum* and its close relatives." *Pathogens* 8 (2019): 298.
11. Parlange, F., G. Daverdin, I. Fudal, M. L. Kuhn, M. H. Balesdent, F. Blaise, B. Grezes-Besset and T. Rouxel. "*Leptosphaeria maculans* avirulence gene avrIm4-7 confers a dual recognition specificity by the rlm4 and rlm7 resistance genes of oilseed rape, and circumvents rlm4-mediated recognition through a single amino acid change." *Molecular Microbiology* 71 (2009): 851-63.
12. Dong, S., S. Raffaele and S. Kamoun. "The two-speed genomes of filamentous pathogens: Waltz with plants." *Current Opinion in Genetics & Development* 35 (2015): 57-65.
13. Waalwijk, C., A. Vanheule, K. Audenaert, H. Zhang, S. Warris, H. van de Geest and T. van der Lee. "*Fusarium* in the age of genomics." *Tropical Plant Pathology* 42 (2017): 184-89.
14. Fouché, S., C. Plissonneau and D. Croll. "The birth and death of effectors in rapidly evolving filamentous pathogen genomes." *Current Opinion in Microbiology* 46 (2018): 34-42.
15. Rouxel, T., J. Grandaubert, J. K. Hane, C. Hoede, A. P. van de Wouw, A. Couloux, V. Dominguez, V. Anthouard, P. Bally, S. Bourras, *et al.* "Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations." *Nat Commun* 2 (2011): 202.
16. Meerupati, T., K.-M. Andersson, E. Friman, D. Kumar, A. Tunlid and D. Åhrén. "Genomic mechanisms accounting for the adaptation to parasitism in nematode-trapping fungi." *PLoS Genetics* 9 (2013): e1003909.
17. Faino, L., M. F. Seidl, X. Shi-Kunne, M. Pauper, G. C. van den Berg, A. H. Wittenberg and B. P. Thomma. "Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen." *Genome Research* 26 (2016): 1091-100.
18. Kanzi, A. M., E. T. Steenkamp, N. A. Van der Merwe and B. D. Wingfield. "The mating system of the eucalyptus canker pathogen *Chrysosporthe austroafricana* and closely related species." *Fungal Genetics and Biology* 123 (2019): 41-52.
19. Gabrielian, A., K. Vlahovicek and S. Pongor. "Distribution of sequence-dependent curvature in genomic DNA sequences." *FEBS Letters* 406 (1997): 69-74.
20. Lewis, Z. A., S. Honda, T. K. Khlafallah, J. K. Jeffress, M. Freitag, F. Mohn, D. Schübeler and E. U. Selker. "Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*." *Genome Research* 19 (2009): 427-37.
21. Grewal, S. I. and S. Jia. "Heterochromatin revisited." *Nature Reviews Genetics* 8 (2007): 35-46.
22. Richmond, T. J. and C. A. Davey. "The structure of DNA in the nucleosome core." *Nature* 423 (2003): 145-50.
23. Kiyama, R. and E. N. Trifonov. "What positions nucleosomes?—a model." *FEBS Letters* 523 (2002): 7-11.
24. Snyder, U. K., J. F. Thompson and A. Landy. "Phasing of protein-induced DNA bends in a recombination complex." *Nature* 341 (1989): 255.
25. Radic, M. Z., K. Lundgren and B. A. Hamkalo. "Curvature of mouse satellite DNA and condensation of heterochromatin." *Cell* 50 (1987): 1101-08.
26. Rimsky, S. "Structure of the histone-like protein h-ns and its role in regulation and genome superstructure." *Current Opinion in Microbiology* 7 (2004): 109-14.

27. Wingfield, M., A. Hammerbacher, R. Ganley, E. Steenkamp, T. Gordon, B. Wingfield and T. Coutinho. "Pitch canker caused by *Fusarium circinatum*-a growing threat to pine plantations and forests worldwide." *Australasian Plant Pathology* 37 (2008): 319-34.
28. Geiser, D. M., A. Al-Hatmi, T. Aoki, T. Arie, V. Balmas, I. Barnes, G. C. Bergstrom, M. Bhattacharyya, C. L. Blomquist and R. Bowden. "Phylogenomic analysis of a 55.1 kb 19-gene dataset resolves a monophyletic *Fusarium* that includes the *Fusarium solani* species complex." *Phytopathology* (2020):
29. Wingfield, B. D., E. T. Steenkamp, Q. C. Santana, M. Coetzee, S. Bam, I. Barnes, C. W. Beukes, W. Yin Chan, L. De Vos and G. Fourie. "First fungal genome sequence from africa: A preliminary analysis." *South African Journal of Science* 108 (2012): 01-09.
30. Katoh, K., J. Rozewicki and K. D. Yamada. "Mafft online service: Multiple sequence alignment, interactive sequence choice and visualization." *Briefings in Bioinformatics* 20 (2019): 1160-66.
31. Darriba, D., G. L. Taboada, R. Doallo and D. Posada. "Jmodeltest 2: More models, new heuristics and parallel computing." *Nature Methods* 9 (2012): 772-72.
32. Kumar, S., G. Stecher, D. Peterson and K. Tamura. "MEGA-cc: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis." *Bioinformatics* 28 (2012): 2685-86.
33. Möller, E., G. Bahnweg, H. Sandermann and H. Geiger. "A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues." *Nucleic Acids Research* 20 (1992): 6115.
34. Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones and I. Birol. "Abyss: A parallel assembler for short read sequence data." *Genome Research* 19 (2009): 1117-23.
35. Boetzer, M., C. Henkel, H. Jansen, D. Butler and W. Pirovano. "Scaffolding pre-assembled contigs using SSPACE". *Bioinformatics* 27: 578–579. 2011,
36. Boetzer, M. and W. Pirovano. "Toward almost closed genomes with gapfiller." *Genome Biology* 13 (2012): R56.
37. Harris, R. S. "Improved pairwise alignment of genomic DNA." (2007): Thesis, Pennsylvania State University, Pennsylvania, PA, USA, 2007
38. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz and C. Duran. "Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data." *Bioinformatics* 28 (2012): 1647-49.
39. Herron, D. A., M. J. Wingfield, B. D. Wingfield, C. Rodas, S. Marinowitz and E. T. Steenkamp. "Novel taxa in the *Fusarium fujikuroi* species complex from *pinus* spp." *Studies in Mycology* 80 (2015): 131-50.
40. Santana, Q. C., M. P. A. Coetzee, B. D. Wingfield, M. J. Wingfield and E. T. Steenkamp. "Nursery-linked plantation outbreaks and evidence for multiple introductions of the pitch canker pathogen *Fusarium circinatum* into South Africa." *Plant Pathology* 65 (2016): 357-68.
41. Wikler, K. and T. R. Gordon. "An initial assessment of genetic relationships among populations of *Fusarium circinatum* in different parts of the world." *Canadian Journal of Botany* 78 (2000): 709-17.
42. Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov. "BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs." *Bioinformatics* 31 (2015): 3210-12.
43. Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva and E. M. Zdobnov. "BUSCO applications from quality assessments to gene prediction and phylogenomics." *Molecular Biology and Evolution* 35 (2017): 543-48.
44. Hoff, K. J. and M. Stanke. "Webaugustus—a web service for training augustus and predicting genes in Eukaryotes." *Nucleic Acids Research* 41 (2013): W123-W28.

45. Li, W.-C., C.-H. Huang, C.-L. Chen, Y.-C. Chuang, S.-Y. Tung and T.-F. Wang. "Trichoderma reesei complete genome sequence, Repeat-Induced Point mutation, and partitioning of cazy gene clusters." *Biotechnology for Biofuels* 10 (2017): 170.
46. Gladyshev, E. "Repeat-Induced Point mutation and other genome defense mechanisms in fungi." *The Fungal Kingdom* (2017): 687-99.
47. Gessaman, J. D. and E. U. Selker. "Induction of H3K9ME3 and DNA methylation by tethered heterochromatin factors in *Neurospora crassa*." *Proceedings of the National Academy of Sciences* 114 (2017): E9598-E607.
48. He, C., Z. Zhang, B. Li and S. Tian. "The pattern and function of DNA methylation in fungal plant pathogens." *Microorganisms* 8 (2020): 227.
49. Lewis, Z. A., K. K. Adhvaryu, S. Honda, A. L. Shiver, M. Knip, R. Sack and E. U. Selker. "DNA methylation and normal chromosome behavior in *Neurospora* depend on five components of a histone methyltransferase complex, DCDC." *PLoS Genet* 6 (2010): e1001196.
50. Visser, E. A., J. L. Wegrzyn, E. T. Steenkamp, A. A. Myburg and S. Naidoo. "Dual ma-seq analysis of the pine-*Fusarium circinatum* interaction in resistant (*Pinus tecunumanii*) and susceptible (*Pinus patula*) hosts." *Microorganisms* 7 (2019): 315.
51. Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell and G. Nuka. "Interproscan 5: Genome-scale protein function classification." *Bioinformatics* 30 (2014): 1236-40.
52. Ozer, E. A., J. P. Allen and A. R. Hauser. "Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt." *BMC Genomics* 15 (2014): 737.
53. Quesneville, H., C. M. Bergman, O. Andrieu, D. Autard, D. Nouaud, M. Ashburner and D. Anxolabehere. "Combined evidence annotation of transposable elements in genome sequences." *PLoS Computational Biology* (2005): e22.
54. Flutre, T., E. Duprat, C. Feuillet and H. Quesneville. "Considering transposable element diversification in *de novo* annotation approaches." *PloS One* 6 (2011): e16526.
55. Van Wyk, S., B. D. Wingfield, L. De Vos, Q. C. Santana, N. A. Van der Merwe and E. T. Steenkamp. "Multiple independent origins for a subtelomeric locus associated with growth rate in *Fusarium circinatum*." *IMA Fungus* 9 (2018): 27-36.
56. Bolzán, A. D. and M. S. Bianchi. "Telomeres, interstitial telomeric repeat sequences, and chromosomal aberrations." *Mutation Research/Reviews in Mutation Research* 612 (2006): 189-214.
57. Wiemann, P., C. M. Sieber, K. W. Von Bargen, L. Studt, E.-M. Niehaus, J. J. Espino, K. Huß, C. B. Michielse, S. Albermann and D. Wagner. "Deciphering the cryptic genome: Genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites." *PLoS Pathogens* 9 (2013): e1003475.
58. van Wyk, S., C. H. Harrison, B. D. Wingfield, L. De Vos, N. A. van Der Merwe and E. T. Steenkamp. "The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations." *PeerJ* 7 (2019): e7447.
59. Margolin, B. S., P. W. Garrett-Engele, J. N. Stevens, D. Y. Fritz, C. Garrett-Engele, R. L. Metzenberg and E. U. Selker. "A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element inactivated by Repeat-Induced Point mutation." *Genetics* 149 (1998): 1787-97.
60. Selker, E. U., N. A. Tountas, S. H. Cross, B. S. Margolin, J. G. Murphy, A. P. Bird and M. Freitag. "The methylated component of the *Neurospora crassa* genome." *Nature* 422 (2003): 893-97.
61. Wintersinger, J. A. and J. D. Wasmuth. "Kablammo: An interactive, web-based blast results visualizer." *Bioinformatics* 31 (2014): 1305-06.

62. Vlahovic`ek, K., L. Kajan and S. Pongor. "DNA analysis servers: Plot. It, bend. It, model. It and is." *Nucleic Acids Research* 31 (2003): 3686-87.
63. Wingfield, B. D., M. Liu, H. D. Nguyen, F. A. Lane, S. W. Morgan, L. De Vos, P. M. Wilken, T. A. Duong, J. Aylward and M. P. Coetzee. "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*, *Grosmannia galeiformis*, *Quambalaria eucalypti*, and *Teratosphaeria destructans*." *IMA Fungus* 9 (2018): 401-18.
64. van Wyk, S., B. D. Wingfield, L. De Vos, N. A. van Der Merwe and E. T. Steenkamp.. "Genome-wide analyses of Repeat-Induced Point (RIP) mutations in the Ascomycota " *Frontiers in Microbiology* (2021): Accepted.
65. De Vos, L., E. T. Steenkamp, S. H. Martin, Q. C. Santana, G. Fourie, N. A. van der Merwe, M. J. Wingfield and B. D. Wingfield. "Genome-wide macrosynteny among *Fusarium* species in the *Gibberella fujikuroi* complex revealed by amplified fragment length polymorphisms." *PloS One* 9 (2014): e114682.
66. Smircich, P., D. Forteza, N. M. El-Sayed and B. Garat. "Genomic analysis of sequence-dependent DNA curvature in *Leishmania*." *PloS One* 8 (2013): e63068.
67. Clutterbuck, A. J. "Genomic evidence of Repeat-Induced Point mutation (RIP) in filamentous Ascomycetes." *Fungal Genetics and Biology* 48 (2011): 306-26.
68. Wingfield, B. D., D. K. Berger, E. T. Steenkamp, H.-J. Lim, T. A. Duong, B. H. Bluhm, Z. W. De Beer, L. De Vos, G. Fourie and K. Naidoo. "Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntia decipiens* and *Ophiostoma ips*." *IMA Fungus* 8 (2017): 385-96.
69. Coleman, J. J., S. D. Rounsley, M. Rodriguez-Carres, A. Kuo, C. C. Wasmann, J. Grimwood, J. Schmutz, M. Taga, G. J. White and S. Zhou. "The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion." *PLoS Genet* 5 (2009): e1000618.
70. Fudal, I., S. Ross, H. Brun, A.-L. Besnard, M. Ermel, M.-L. Kuhn, M.-H. Balesdent and T. Rouxel. "Repeat-Induced Point mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*." *Molecular Plant-Microbe Interactions* 22 (2009): 932-41.
71. Peck, L., R. Nowell, J. Flood, M. Ryan and T. Barraclough. "Historical genomics reveals the evolutionary mechanisms behind multiple outbreaks of the host-specific coffee wilt pathogen *Fusarium xylarioides*." *bioRxiv* (2020): doi: <https://doi.org/10.1101/2020.08.07.241695>
72. Testa, A., R. Oliver and J. Hane. "Overview of genomic and bioinformatic resources for *Zymoseptoria tritici*." *Fungal Genetics and Biology* 79 (2015): 13-16.
73. Ellison, C. E., J. E. Stajich, D. J. Jacobson, D. O. Natvig, A. Lapidus, B. Foster, A. Aerts, R. Riley, E. A. Lindquist and I. V. Grigoriev. "Massive changes in genome architecture accompany the transition to self-fertility in the filamentous fungus *Neurospora tetrasperma*." *Genetics* 189 (2011): 55-69.
74. Sun, Y., J. Svedberg, M. Hiltunen, P. Corcoran and H. Johannesson. "Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*." *Nature Communications* 8 (2017): 1-8.
75. Amyotte, S. G., X. Tan, K. Pennerman, M. del Mar Jimenez-Gasco, S. J. Klosterman, L.-J. Ma, K. F. Dobinson and P. Veronese. "Transposable elements in phytopathogenic *Verticillium* spp.: Insights into genome evolution and inter-and intra-specific diversification." *BMC Genomics* 13 (2012): 314.
76. Daboussi, M.-J., J.-M. Daviere, S. Graziani and T. Langin. "Evolution of the *fof1* transposons in the genus *Fusarium*: Discontinuous distribution and epigenetic inactivation." *Molecular Biology and Evolution* 19 (2002): 510-20.
77. Spanu, P. D., J. C. Abbott, J. Amselem, T. A. Burgis, D. M. Soanes, K. Stüber, E. V. L. van Themaat, J. K. Brown, S. A. Butcher and S. J. Gurr. "Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism." *Science* 330 (2010): 1543-46.

78. Suzuki, M. M. and A. Bird. "DNA methylation landscapes: Provocative insights from epigenomics." *Nature Reviews Genetics* 9 (2008): 465-76.
79. Bechert, T., S. Diekmann, S. Heck, U. Fleig and J. H. Hegemann. "All 16 centromere dnas from *Saccharomyces cerevisiae* show DNA curvature." *Nucleic Acids Research* 27 (1999): 1444-49.
80. Gordon, B. R., Y. Li, A. Cote, M. T. Weirauch, P. Ding, T. R. Hughes, W. W. Navarre, B. Xia and J. Liu. "Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins." *Proceedings of the National Academy of Sciences* 108 (2011): 10690-95.
81. Jahani, S. and S. K. Setarehdan. "Centromere and length detection in artificially straightened highly curved human chromosomes." *International Journal of Biological Engineering* 2 (2012): 56-61.