Genome compartmentalization predates species divergence in the plant pathogen genus Zymoseptoria

Alice Feurtey1,2, Cécile Lorrain1,2,3*, Daniel Croll4, Christoph Eschenbrenner1,2, Michael Freitag5, Michael Habig1,2, Janine Haueisen1,2, Mareike Möller1,2,5, Klaas Schotanus1,2,6 and Eva H. Stukenbrock1,2

Abstract

Background: Antagonistic co-evolution can drive rapid adaptation in pathogens and shape genome architecture. Comparative genome analyses of several fungal pathogens revealed highly variable genomes, for many species characterized by specific repeat-rich genome compartments with exceptionally high sequence variability. Dynamic genome structure may enable fast adaptation to host genetics. The wheat pathogen Zymoseptoria tritici with its highly variable genome, has emerged as a model organism to study genome evolution of plant pathogens. Here, we compared genomes of Z. tritici isolates and of sister species infecting wild grasses to address the evolution of genome composition and structure.

Results: Using long-read technology, we sequenced and assembled genomes of Z. ardabiliae, Z. brevis, Z. pseudotritici and Z. passerinii, together with two isolates of Z. tritici. We report a high extent of genome collinearity among Zymoseptoria species and high conservation of genomic, transcriptomic and epigenomic signatures of compartmentalization. We identify high gene content variability both within and between species. In addition, such variability is mainly limited to the accessory chromosomes and accessory compartments. Despite strong host specificity and non-overlapping host-range between species, predicted effectors are mainly shared among Zymoseptoria species, yet exhibiting a high level of presence-absence polymorphism within Z. tritici. Using in planta transcriptomic data from Z. tritici, we suggest different roles for the shared orthologs and for the accessory genes during infection of their hosts.

Conclusion: Despite previous reports of high genomic plasticity in Z. tritici, we describe here a high level of conservation in genomic, epigenomic and transcriptomic composition and structure across the genus Zymoseptoria. The compartmentalized genome allows the maintenance of a functional core genome co-occurring with a highly variable accessory genome.

Keywords: Genome evolution, Orphan genes, Effectors, Genome architecture, Accessory genes
Background

Co-evolution between plants and pathogens can drive rapid evolution of genes involved in antagonistic interactions [1]. In filamentous plant pathogens, rapid evolution may be fueled by highly dynamic genome architecture involving repeat-rich compartments such as gene-sparse islands of repetitive DNA and accessory chromosomes [2, 3]. These compartments can show a high plasticity revealed by a high extent of gene and/or chromosome presence-absence variation and structural variants, such as inversions, insertions and deletions [4, 5]. Several plant pathogenic fungi have isolate-specific chromosomes, so-called accessory chromosomes.

Accessory chromosomes are characterized by intra-species presence-absence polymorphism, low gene density, an enrichment of repetitive sequences and, in some species, a different histone methylation pattern [6, 7]. It has been shown that accessory chromosomes encode genes involved in virulence such as in the species Fusarium solani, Fusarium oxysporum and Leptosphaeria maculans [8–11]. Little is known about the evolutionary origin of accessory chromosomes although experimental evidence from the asexual species F. oxysporum shows that accessory chromosomes may be acquired horizontally as chromosomes can be transferred between distinct isolates by hyphal fusion [10]. Through such transfers, virulence determinants may be exchanged between clonal lineages as accessory chromosomes in this species were shown to encode host specific virulence determinants and transcription factors regulating their expression [12].

Genes involved in plant-pathogen interactions may diversify at a higher rate in repeat-rich genome compartments and thereby evolve new virulence specificity faster [3]. These genes encode secreted proteins, so-called effectors [1]. Most known effectors target diverse cellular compartments and molecular pathways, including immune response-related pathways [13, 14]. Genes encoding Carbohydrate-active enzymes (CAZymes) have also been associated to the pathogenic lifestyle of fungal plant pathogens, particularly through their role in plant-cell wall degradation [15]. Thus, some secreted CAZymes may be essential from the early infection stage, like penetration of plant tissue, to later stages such as the necrotrophic phase where the pathogen feeds from dead plant tissue [16]. Likewise, secondary metabolites are known to be involved in plant infection and contribute to virulence and the interaction with other plant-associated microorganisms [17, 18]. Many of these genes can be predicted either according to their composition and known protein domains or through machine learning methods [19]. Thereby, in-depth genome annotations have proven important to predict and compare the content of pathogenicity-related genes in plant pathogens, as well as their genomic localization for example in rapidly evolving genome compartments.

The ascomycete pathogen Zymoseptoria tritici has emerged as a model organism in evolutionary genomics of pathogens. This species originated in the Fertile Crescent during the domestication of its host, wheat [20]. Closely related species of Z. tritici have been collected from wild grasses in the Middle East providing an excellent resource for comparative genome analyses of closely related and recently diverged pathogen species. Comparative analyses of genome organization and gene content within and among Zymoseptoria species have previously revealed a wide distribution of accessory chromosomes and dynamic gene content [21, 22]. The haploid genome of the reference isolate IPO323 comprises thirteen core and eight accessory chromosomes [23]. Some of these accessory chromosomes may encode traits that impact virulence of the fungus, however no gene encoded on an accessory chromosome has so far been described as a virulence or avirulence determinant [24–29]. Interestingly, the accessory chromosomes in Z. tritici show a low transcriptional activity in vitro as well as in planta [30, 31]. This suppression of gene expression correlates with an enrichment of heterochromatin associated with the histone modification H3K27me3 on the accessory chromosomes [6, 32].

In the reference isolate IPO323, the accessory chromosomes comprise more than 11% of the entire genome assembly. To which extent such a high amount of accessory DNA is also found in genomes of other members of the Zymoseptoria genus has so far been unknown due to the lack of high-quality genome assemblies and large-scale population sequencing. Assemblies based on short-read data failed to recover complete sequence of accessory chromosomes and “orphan regions” due to their high repeat content [22]. The asset of genome assemblies based on long-read sequencing was demonstrated in detailed genome comparisons of Z. tritici isolates sequenced with PacBio long-read sequencing [28, 33]. Comparison these Z. tritici high-quality chromosome assemblies revealed the occurrence of “orphan regions” enriched with transposable elements and encoding putative virulence-related genes [28, 34].

In this study, we investigate the genomic architecture and variability among five Zymoseptoria species. Beside presenting a new and significantly improved resource for future genomic studies of these fungal pathogens, we specifically ask: 1) how conserved is the genome architecture among Zymoseptoria species? 2) can we identify accessory compartments in other Zymoseptoria isolates? 3) to which extent does variation in genome architecture reflect variation in gene content?

To answer these questions, we used high-quality assemblies based on long-read sequence data and new gene predictions in two isolates of Z. tritici (Zt05 and Zt10) and one isolate of each of the sister species, Z. ardabiliae, Z. brevis, Z. passerinii, and Z. pseudotritici.
We explore the core and non-core genome architecture of *Zymoseptoria* spp. combining genomic data with transcriptome and histone methylation data and relate this to core and accessory genome compartments. Furthermore, we compare the distribution of orthologous and non-orthologous genes in the *Zymoseptoria* genomes and one additional Dothideomycete species. Our analyses reveal an overall conserved genome architecture characterized by gene-rich core compartments and accessory compartments enriched in species-specific genes. Finally, we report a remarkably high extent of variation in presence-absence of protein coding genes in a eukaryote genome.

**Results**

**De novo assemblies using long-read sequencing for six *Zymoseptoria* spp.**

We sequenced and assembled the genome of the reference isolates of *Z. ardabiliae*, *Z. brevis*, *Z. pseudotritici* and *Z. passerinii* and the genomes of two *Z. tritici* isolates sampled in Denmark and Iran [30]. The obtained contigs were filtered based on base-quality confidence and read depth to ensure high quality of the final assemblies (see Methods). This filter removed a high number of contigs (between 17 and 58% of the total), but little overall length (between 0.4 and 2.6% of the total assemblies), indicating that most of the excluded contigs were of small size (Table S1 and S5). The best assemblies were of the two *Z. tritici* isolates comprising 19 and 30 contigs and the most fragmented was of *Z. passerinii* comprising 103 contigs (Fig. 1). The resulting assembly lengths ranged from 38.1 Mb for *Z. ardabiliae* to 41.6 Mb for *Z. brevis*, which is comparable to the reference assembly length of *Z. tritici* (39.7 Mb) but larger than previous short-read based assemblies (Table 1; previous assemblies ranged from 31.5 Mb for *Z. ardabiliae* to 32.7 Mb for *Z. pseudotritici* [22, 23]. The assembly of the Iranian *Z. tritici* isolate Zt10 has telomeric repeats at the end of all contigs, indicating that each chromosome is completely assembled, comprising six accessory and thirteen core chromosomes. The assemblies for the Danish *Z. tritici* isolate (Zt05), *Z. brevis* (Zb87) and *Z. pseudotritici* (Zp13) contained, respectively, twelve, nine and five fully assembled chromosomes including both core and accessory chromosomes (Figure S1). The assemblies of the *Z. ardabiliae* (Za17) and *Z. passerinii* (Zpa63) genomes included no fully assembled chromosomes, but twelve and ten contigs respectively with telomeres at one of the ends (Table 1; Figure S1).

The transcriptome-based gene predictions for these new assemblies include between 10,528 and 12,386 protein-coding genes (Table S2). This range is consistent with the annotation of the reference genome IPO323 reporting 11,839 protein-coding genes [22]. We used Benchmarking Universal Single-Copy Orthologs (BUSCO) from the lineage dataset *Pezizomycotina* to evaluate the completeness of the assemblies and gene predictions [37]. The proportion of complete BUSCO genes identified in our assemblies were comparable to the one obtained for the reference genome of *Z. tritici*.

---

**Fig. 1** Whole-genome phylogeny of *Zymoseptoria* spp. and basic statistics for the assemblies and gene predictions. **A** Tree based on the distance matrix generated by the software and distances from whole genome sequences were estimated in an alignment-free manner [35]. The same topology was observed in a tree produced from k-mers using the web-based tool CVtree3 [36]. **B** The bar plots represent the number of genes coding for secreted proteins (pink) and non-secreted proteins (grey) for each genome.
The assessment of gene content completeness (see Methods) indicates that, despite more fragmented assemblies of *Z. ardabiliae* and *Z. passerinii*, the genomes are complete in terms of gene content and that the unassembled fragments are more likely to comprise repeats and not protein-coding genes. Based on the whole-genome sequences and the predicted genes we reconstructed the phylogeny of the *Zymoseptoria* genus using the publicly available genome of *Cercospora beticola* as an outgroup [38, 39]. For both trees, the phylogenetic relationship of the *Zymoseptoria* species is in accordance with previously published phylogeny based on seven loci sequenced in multiples isolates (Fig. 1) [21].

Genomes of *Zymoseptoria* spp. comprise accessory chromosomes and compartments but show overall high synteny

Next we addressed the extent of co-linearity of the *Zymoseptoria* genomes. Using coordinates of orthologous genes, we were able to reveal a high extent of synteny conservation among the five *Zymoseptoria* species and between the three isolates of *Z. tritici*, as depicted in Fig. 2 and S2. Based on this high extent of synteny and the prediction of telomeric repeats, we identified the correspondence of chromosomes between the reference genome of *Z. tritici* IPO323 and the other *Zymoseptoria* genomes (Fig. 2 and S2). *Z. brevis* and *Z. pseudotritici* share a near perfect synteny in their core chromosomes, however, when compared to *Z. tritici*, *Z. brevis* and *Z. pseudotritici* have two large-scale inversions comprising roughly ~ 900 kb and ~ 1.2 Mb of chromosomes 2 and 6, respectively (Fig. 2, S2 and S3). Based on the phylogeny in Fig. 1, it is likely that these two events occurred after the divergence of *Z. tritici* from *Z. brevis* and *Z. pseudotritici*. Overall, we observe a higher extent of synteny conservation between *Z. brevis* and *Z. pseudotritici* compared to *Z. tritici* IPO323 (Fig. 2; S2 and S3).

In *Z. tritici*, core and accessory compartments have very distinct genomic features. It was previously shown that hallmarks of accessory regions in the reference isolate IPO323 include lower gene density, lower levels of H3K4me2 methylation and reduced gene expression [6, 31]. In the reference genome of IPO323, compartments with these genomic and epigenomic hallmarks represent either accessory chromosomes or specific regions of the core chromosomes. Here we find that the specific accessory hallmarks including low gene density, low expression, low H3K4me2 methylation and significant enrichment of species-specific genes (see description below) on the non-core contigs are found in genomic compartments throughout the genus (Table S3, Fig. 3 and S4).

In the genome of the reference *Z. tritici* strain, the compartments that exhibit the hallmark of accessory chromosomes includes a particular region of the core chromosome 7 of ~ 0.6 Mb (Fig. 3a) [6].

| Table 1 Metrics of genome assemblies and annotation |
|----------------------------------------------------|
| Species | *Zymoseptoria tritici* | *Zymoseptoria pseudotritici* | *Zymoseptoria brevis* | *Zymoseptoria ardbiliae* | *Zymoseptoria passerinii* |
| Isolate | Zt05 | Zt10 | Zp13 | Zb87 | Za17 | Zpa63 |
| Origin | Denmark | Iran, Ilam province | Iran, Ardabil province | Iran | Iran, Ardabil province | USA |
| Host | *Triticum aestivum* | *Triticum aestivum* | *Dactylis glomerata* | *Phalaris paradoxa* | *Lolium perenne* | *Hordeum vulgare* |
| Year (of isolation) | 2004 | 2001 | 2004 | 2004 | 2004 | 2004 |
| Contig number | 30 | 19 | 42 | 29 | 50 | 103 |
| Total length (bp) | 41,240,984 | 39,248,105 | 40,312,446 | 41,586,671 | 38,100,668 | 41,398,787 |
| Mean contig size (bp) | 1,374,699 | 2,065,690 | 59,820 | 1,434,023 | 762,013 | 401,930 |
| N50 | 2,454,671 | 2,925,395 | 2,115,121 | 2,744,794 | 1,156,695 | 737,698 |
| L50 | 6 | 5 | 7 | 7 | 11 | 18 |
| Contigs with telomeric repeats on both ends | 12 | 19 | 5 | 9 | 0 | 0 |
| Complete BUSCO genes (%) | 98.4 | 98.7 | 98.5 | 97.0 | 98.2 | 97.5 |
| Number of genes | 12,386 | 11,991 | 11,661 | 11,480 | 11,463 | 10,528 |
| Repeat content (%) | 19.9 | 16.5 | 20.8 | 29.2 | 18.2 | 31.4 |
| Percentage of non-core sequences compared to the IPO323 reference | 12.49 | 8.69 | 4.27 | 4.6 | 3.45 | 6.85 |
| NCBI Biosample | SAMN04494882 | SAMN02981321 | SAMN02981322 | SAMN03294124 | SAMN02981326 | SAMN02981330 |
Fig. 2 Intra- and inter-species synteny conservation in Zymoseptoria genus. **a)** Intra-species synteny between the reference genome of *Z. tritici* IPO323 and the genome of the Iranian *Z. tritici* isolate Zt10. Each color represents a different chromosome as defined in the reference *Z. tritici* IPO323 genome, except for accessory chromosomes, which are in grey. The links represent a subsample of orthologous genes (subsampled 1:2 for the accessory chromosomes and 1:10 for the core chromosomes for clarity of the visual representation). Contigs are ordered according to their synteny to the reference genome IPO323. Telomeric repeats are indicated in orange. **b)** Inter-species synteny between the reference genome of *Z. tritici* and the genome of *Z. brevis* Zb87. The arrows represent the large-scale inversions identified between the genomes of these two species.

Fig. 3 Genome architecture of the reference genome *Z. tritici* IPO323 (a) and *Z. pseudotritici* Zp13 (b). The segments constituting the first circle represents the chromosomes of IPO323 (a) and contigs of Zp13 (b) ordered according to the synteny with the chromosomes of the reference genome. Tracks from the outside to the inside are heatmaps representing respectively: gene density along chromosomes/contigs; gene expression in vitro (TPM); H3K4me2 levels in vitro and species-specific gene density per 100 kb windows. The arrows indicate the location of the region on chromosome 7 (and the corresponding syntenic region in *Z. pseudotritici*) displaying accessory-like genomic and regulatory hallmarks. Telomeric repeats are indicated in orange.
like” because the region exhibits clearly two of the three above-mentioned criteria [6, 31]. The region has low gene expression, low H3K4me2 levels but unlike accessory chromosomes exhibits high gene density. The encoded genes are mostly species-specific. Interestingly, this particularly large accessory region is observed in several Zymoseptoria spp. (Fig. 3b; S4). Based on synteny plots, we recognize ~0.7 Mb of the contig 28 in Z. pseudotritici and ~0.6 Mb of the contig 17 for Z. brevis, corresponding to chromosome 7 of Z. tritici (Fig. 2 and S2) and sharing the same hallmarks of accessory chromosomes (Fig. 3 and S4, Table S3). For the two remaining sister species, the fragmentation of the assembly does not allow the identification of such pattern although we observe a similar tendency with respect to transcription and species-specific gene enrichment on contig 19 (~0.6 Mb) of Z. ardbiliae corresponding to a fragment of chromosome 7 in IPO323 (Figure S5).

We also identified other regions enriched in isolate-specific genes, thus defining orphan loci in the core chromosomes of both Zt10 and Zt05 (Fig. 3 and S4). We observed a region of ~0.2 Mb of contig 1 in the Iranian isolate Zt10 corresponding to the core chromosome 3 in the IPO323 genome with high content of isolate-specific genes (Figure S4; Table S3). We furthermore identified small segments with species-specific genes on the core chromosomes of the wild-grass infecting sister species including a ~0.3 Mb region of the contig 26 in Z. brevis and ~0.1 Mb of contig 30. Overall, we show that genome compartmentalization in core and accessory regions is an ancestral and shared trait among the Zymoseptoria species. This phenomenon generates highly variable compartments and defines loci that deviate from genome averages in terms of gene content, sequence composition and synteny conservation.

**Variable repertoires of effector candidate genes**

To obtain gene annotations for the Zymoseptoria genome assemblies, we established a custom pipeline adapted from Lorrain and co-workers (Figure S5) [40]. Briefly, we use the consensus of three methods to predict gene product localization, then extract secreted proteins to further identify predicted effectors. This detailed functional annotation provided a catalog of predicted gene functions and cellular localizations (Figure S6). For each genome, a large proportion of genes could not be assigned to a protein function. 49.6% of genes in Z. tritici (N = 5953) and up to 71.8% of genes in Z. pseudotritici (N = 8373) lack a predicted function (i.e. proteins of unknown function; Figure S6A). A relatively consistent number of genes are predicted for each functional category among Zymoseptoria spp. (Figure S6A and B). Likewise, the numbers of gene products predicted to belong to the different subcellular localizations are very similar (Figure S6C) across the whole genus, including secreted proteins. The difference between the minimal and maximal gene number for the different categories of subcellular localizations does not exceed 1.6X between species (Figure S6C). Overall, secretomes range from 7% of the genes predicted sin Z. passeriini (N = 828) to 11% of genes in Z. ardbiliae (N = 1328 genes, Figure S6B).

We further investigated the number and distribution of genes predicted to encode proteins with a pathogenicity-related function, such as secondary metabolites, CAZymes and predicted effectors (Figure S1 and S6B). Genes involved in the synthesis of secondary metabolites are typically organized in clusters, with genes participating in the same biosynthetic pathway grouping together at a genomic locus (Shi-Kunne et al. 2019). The number of biosynthetic gene clusters (BGC) ranges from 25 in Z. ardbiliae and Z. passeriini to 33 in the IPO323 reference genome and includes from 305 to 471 predicted genes (Figure S1). The only BGC identified in a non-core contig is a non-ribosomal peptide synthetase BGC found on the contig 38 of Z. brevis which has no orthologous cluster detected in any of the other Zymoseptoria genomes (Figure S1). We identified between 454 and 515 CAZyme genes in the Zymoseptoria species. Both BGCs and CAZymes are almost exclusively found on the core chromosomes (Figure S1). The only exceptions are a CAZyme encoding gene found on chromosome 14 in Z. tritici IPO323 and Zt05, and a CAZyme encoding gene on the putative accessory contig 38 of Z. brevis (Figure S1). These two genes encode for a beta-glucosidase and a carboxylic-ester hydrolase, respectively.

In contrast to the high conservation of CAZyme and BGC gene content among the Zymoseptoria genomes, we find that predicted effector genes exhibit a large variation in gene numbers between genomes (Figure S6B). In fact, the predicted effector gene repertoire in Z. ardbiliae (N = 637) is three times higher compared to Z. brevis (N = 206). Interestingly, the three Z. tritici isolates also vary considerably in their predicted effector repertoires. The reference isolate IPO323 has a reduced set of predicted effector genes (N = 274) compared to Zt05 and Zt10 that encode approximately 30% more predicted effector genes (N = 417 and N = 403, respectively, Figure S6B). Despite the high variability, the predicted effector genes are mostly located on core chromosomes and none of the five Zymoseptoria species have more than ten predicted effector genes located on accessory chromosomes (Figure S1).

The accessory genes of Z. tritici are shared with the closely related wild-grass infecting species

To further characterize variation in gene content among the five Zymoseptoria species, we identified orthologous
genes (i.e. orthogroups) from the gene predictions. We categorized 22,341 gene orthogroups identified in the seven Zymoseptoria genomes and in C. beticola according to their distribution among fungal genomes (Fig. 4a). The core orthogroups, which are genes present in all eight genomes, represent around 30% of all orthogroups ($N = 6698$). The genus-specific orthogroups, shared between several Zymoseptoria spp. but not found in the C. beticola genome, represent 45% of the orthogroups ($N = 9955$; ranging from 2066 to 3212 per species). Among the genus-specific orthogroups, 1100 are found in all Zymoseptoria genomes (Fig. 4a), whereas all others show presence-absence polymorphisms within the genus. A total of 2476 species-specific orthogroups (ranging from 552 to 1191 per species) are found only in individual species. Among the species-specific genes, 205 orthogroups ($N_{genes} = 414$ to 562) are found in all three Z. tritici genomes while the isolate-specific genes in Z. tritici represent 391 (Zt10) to 792 (IPO323) genes.

Comparing the three Z. tritici isolates independently from the other species, we observe extensive gene presence-absence polymorphisms between the three isolates: 1540 orthogroups are identified in only two strains and 2522 are found in only one (Fig. 4b). The number of genes showing presence-absence variation is striking compared to the 10,098 core genes in Z. tritici as these genes comprise almost 30% of all predicted genes. Interestingly, we show that the number of orthogroups detected as isolate-specific is much larger when the comparison includes only members of the same species than when the other species are included (1035, 849 and 638 vs 792, 659 and 391 genes for IPO323, Zt05 and Zt10 respectively; Fig. 4a and b). This indicates that a large part of the accessory gene content in Z. tritici is shared among the sister species, and highlights the importance of including sister species when establishing core and accessory gene content.

Interestingly, we show that predicted effectors are enriched among the genus-specific genes but not among the species-specific or isolate-specific gene categories (with the exception of Z. ardabiliae). Fifty-six percent of predicted effectors in Z. ardabiliae and up to 78% of predicted effectors in Z. pseudotritici are shared with at least one of the other five Zymoseptoria species (Fig. 4c). Indeed, 427 predicted effector orthogroups are found in at least two genomes. However, only 47 (10% of the total

---

**Fig. 4 Orthogroups and functional gene categories in Zymoseptoria spp. genomes.**

A. Orthogroups shared by the reference Z. tritici genome, our new Zymoseptoria assemblies and the outgroup genome of C. beticola. Only intersects higher than 100 are displayed on the upset plot. The doughnut plot summarizes the number of orthologs grouped by larger categories: specific to some isolates, to a species or shared by all. The colored bars under the upset plot link each intersect to its corresponding category in the doughnut plot.

B. Venn diagram representing the genes shared by the three isolates of Z. tritici. The only gene category found to be overrepresented in any of the specificity categories - other than unknown function genes - are predicted effector genes. Predicted effectors genes are overrepresented in the genus-specific genes and in Z. ardabiliae specific genes (** represent Fisher exact test p-value < 0.05)
predicted effector orthogroups \( N = 474 \) are found in all seven Zymoseptoria genomes. Among the predicted effectors shared by Z. tritici, and at least one other Zymoseptoria species, 32\% \( (N = 112 \text{ of } 352) \) are present in all three Z. tritici isolates while 68\% \( (N = 240 \text{ of } 352) \) show presence-absence polymorphisms in at least one of the three isolates. These results indicate that the majority of these shared predicted effectors are actually accessory (i.e. presence-absence polymorphism) in Z. tritici.

Among in planta differentially expressed genes, species-specific are more expressed than core genes
Finally, we addressed the functional relevance of accessory and orphan genes in Z. tritici by analyzing gene expression patterns. We used previously published in planta expression data of three Z. tritici isolates [30]. The expression profiling was obtained from four subsequent infection stages including infection establishment (stage A), biotrophic colonization (stage B), the transition from biotrophic to necrotrophic phase (stage C) and necrotrophic colonization (stage D) [30]. We sorted in planta expression data into two different infection phases: the biotrophic phase and the necrotrophic phase, a separation supported by principal component analysis of normalized DESeq2 counts (Figure S7). Furthermore, we distinguished gene expression of the above-defined categories (core genes, genus-specific, species-specific, and isolate-specific). We compared expression levels by mapping RNA-seq reads to the genomes of IPO323, Zt05 and Zt10, using normalized read mappings to transcript per million. We tested differences among gene categories using pairwise comparisons with a Kruskal-Wallis test (Fig. 5). Overall, we find that gene expression of the species-specific and isolate-specific genes is significantly lower in IPO323 and Zt10, but not in Zt05 (Kruskal-Wallis p-value < 0.05). Species-specific and isolate-specific gene median expression ranges from 3.2 to 5.6 TPM in IPO323 and Zt10 while median expression of core genes is 12.1 and 10.9, respectively. The Zt05 expression profile does not follow the same trend: the core genes are the lowest expressed gene category (8.9 median TPM), while genus-; species- and isolate-specific genes showed higher transcription levels (12.0; 14.4 and 13.5 median TPM respectively, Kruskal-Wallis p-value < 0.05).

---

**Fig. 5** In planta expression of genes belonging to different specificity levels in the Zymoseptoria pangeneome. The boxplots represent the expression levels in both biotrophic and necrotrophic phase in transcript per million (TPM) for a) the whole transcriptome of Z. tritici isolates and b) in planta differentially expressed genes identified by DESeq2. Comparisons are performed by Kruskal-Wallis test, different letters represent p-value < 0.05.
In contrast, we observe a significantly higher expression of the species-specific and isolate-specific genes for all three isolates (Table S4; Kruskal-Wallis p-value < 0.05) when comparing the expression of genes that are differentially expressed (DEGs; DESeq2 p-adjusted < 0.05) between the biotrophic and necrotrophic phases. Species-specific and isolate-specific DEGs are higher expressed in planta than the core and genus-specific genes (Fig. 5b; Kruskal-Wallis p-value < 0.05). The expression patterns of DEGs with different levels of specificity present a consistent pattern in all three isolates (Table S4). Overall, this comparison reveals a potential functional relevance of accessory genes, which are up-regulated during infection of Z. tritici.

Discussion
In this study we present a new resource of high-quality whole genome assemblies and gene annotations for the fungal grass pathogens Z. ardbiliae, Z. brevis, Z. passerinii, Z. pseudotritici, and two isolates of the wheat pathogen Z. tritici. This new dataset provides a valuable resource for detailed analyses of genome architecture and evolutionary trajectories in this group of plant pathogens. Here, we conduct some detailed comparative analyses of genome architecture and show a considerable extent of variation in sequence composition during the recent evolution of the Zymoseptoria lineages. We show that genome compartmentalization and accessory chromosomes represent shared ancestral traits among these pathogen species.

We identify extensive presence-absence variation of protein coding genes in genomes of the five Zymoseptoria species consistent with the variable gene repertoire already reported for one of the species, Z. tritici [28]. Furthermore, the different species, share a particular genomic architecture that comprises specific accessory genome compartments. In spite of this variation, we observe an overall conserved synteny of the core chromosomes. In the Zymoseptoria genomes, we observe gene-dense, actively transcribed and H3K4me2-enriched compartments associated with most of the core chromosomes. These compartments are clearly distinguishable from gene-sparse, non-transcribed and H3K4me2-deprived compartments. Based on previous analyses of accessory chromosomes in Z. tritici, we here consider this pattern as a specific hallmark of accessory genome compartments in the genus Zymoseptoria beyond only in Z. tritici [6]. We hypothesize that these compartments likely represent accessory chromosomes in the different Zymoseptoria species.

We also identify accessory signatures in core chromosomes, including the previously described right arm of chromosome 7 [6]. Although this region has not been reported to share the same extent of presence-absence polymorphism as the accessory chromosomes, a considerably smaller chromosome 7 was reported in a single Z. tritici isolate originating from Yemen [33]. Here we show that the region homologous to chromosome 7 in the other Zymoseptoria species also exhibits accessory compartment hallmarks. Our results support the occurrence of a past chromosome fusion, but hereby show that it very likely occurred prior to the divergence of the species (estimated to date tens of thousands of years [21]). The specific genomic and epigenetic features have remained stable through speciation and evolutionary time.

In this study, we confirm previously reported genome comparisons showing that gene content in Z. tritici is highly variable [28]. We further extended the identification of orthologs throughout the whole Zymoseptoria genus. Thereby, we show that more than 25% of the genes identified as isolate-specific in a comparison including only Z. tritici isolates are actually present in the wild-grass infecting sister species. This observation suggests that a large proportion of the accessory genome of Z. tritici is not specific to this species. Instead, the accessory genome content is shared among Zymoseptoria species. The proportion of accessory Z. tritici genes shared with other Zymoseptoria species was found to be the highest in the Iranian isolate, which is the only isolate sympatric with the four sister-species. A likely explanation for this observation would be inter-specific gene flow, which would allow the different wild species to exchange genes with sympatric Z. tritici isolates. This new finding is consistent with recent findings from population genomic data studies revealing extensive introgression between Zymoseptoria species [41, 42]. Our observation opens new perspectives for further analysis to understand how inter-specific gene flow has affected the evolution of the accessory genome of Z. tritici.

The genes with predicted functions and, in particular, functions related to pathogenicity are largely shared among species in the Zymoseptoria genus. Although the lifestyles of the wild-grass infecting Zymoseptoria are poorly understood, the species share major features of their lifestyles. Thus, as expected, we find similar CAZymes and BGC contents across the genomes studied here. In Zymoseptoria, most of the predicted effectors are shared among all species, although they show presence-absence variation. In the Botrytis genus (Dothideomycetes), sister species infecting different hosts share effectors with confirmed functions [43]. Likewise, in Microbotryum, a fungal plant pathogen including several species specialized on different hosts a large set of genes encoding conserved and shared secreted proteins was identified. This repertoire of effectors is hypothesized to include traits relevant for pathogenicity among Microbotryum lineages [44]. We hypothesize that the different specificity levels reflect functional differences in the effector repertoire of Zymoseptoria. Predicted effector
genes conserved across the *Zymoseptoria* genus are likely core pathogenicity factors potentially targeting key plant defense mechanisms common to all of the grass hosts [45]. Variation in the composition of effector genes in plant pathogen genomes, including the presence of species-specific and isolate-specific may reflect different host specificities and rapid evolution of these genes [45]. Here we find that only a fraction of the genus-specific predicted effector genes is shared among all *Zymoseptoria* species; the majority shows presence-absence polymorphisms suggesting that a variable effector repertoire is an ancestral trait in these plant pathogens. *Z. ardbiliae* has been isolated from leaves of distantly related grass species in Iran, including *Lolium* spp., *Elymus repens*, and *Dactylis glomerata* and potentially resulting in a broader species-specific effector repertoire of *Z. ardbiliae* compared to the other species [21].

Consistent with a previous study [28], we found that *Z. tritici* core genes are generally more expressed compared to accessory genes *in planta*. Core genes are more likely to encode essential functions, which could explain higher expression pattern during infection. Differentially expressed genes that are specifically induced during the course of the infection are very likely to have functions essential to pathogenicity of the fungus. Interestingly, here we show that within the genes differentially expressed between the biotrophic and the necrotrophic phases of infection, isolate- and species-specific genes have higher expression levels than core genes. These isolate- and species-specific genes could be functionally important and regulate functions linked to infection success in the biotrophic phase or to leaf colonization in the necrotrophic phase. Since these genes show presence-absence polymorphisms in the genus and in the *Z. tritici* species, they could represent a reservoir for possible adaptations to host species, host cultivars or local environments.

**Conclusion**

We investigated the genomic architecture in a genus of plant pathogens, including the economically relevant wheat pathogen *Z. tritici*. Comparing genome content and genome structure, we identified a large shared predicted effector repertoire characterized by inter- and intraspecies presence-absence polymorphisms. Major features of genomic, transcriptomic and epigenetic compartmentalization, distinguishing accessory and core compartments, were shared among wheat and wild-grass infecting *Zymoseptoria* species. We conclude that compartmentalization of genomes is an ancestral trait in the *Zymoseptoria* genus.

**Methods**

**Fungal material, DNA extraction and sequencing**

Details regarding the individual *Zymoseptoria* isolates can be found in Table 1. For genomic data we used the three *Z. tritici* isolates IPO323 (reference), Zt05 and Zt10, one *Z. ardbiliae* isolate (Za17), one *Z. brevis* isolate (Zb87), one *Z. passerinii* isolate (Zpa63), and the *Z. pseudotritici* isolate (Zp13). For transcriptomic and epigenomic data we used *Z. tritici* Zt09 (IPO323 ΔChr18) a derivate of the reference isolate IPO323 deleted with the chromosome 18 [31].

Long read assemblies of the *Z. tritici* isolates Zt05 and Zt10 were described and published previously [30]. For DNA extraction and long read sequencing cultures of *Z. pseudotritici*, *Z. ardbiliae*, *Z. brevis* and *Z. passerinii* were maintained in liquid YMS medium (4 g/L yeast extract, 4 g/L malt extract, 4 g/L sucrose) at 200 rpm and 18°C. DNA extraction was conducted as previously described [26]. PacBio SMRTbell libraries were prepared using DNA extracted from single cells based on a CTAB extraction protocol [30, 46]. The libraries were size selected with an 8-kb cutoff on a BluePippin system (Sage Science).

After selection, the average fragment length was 15 kb. Sequencing of the isolates Za17, Zb87, and Zp13 was run on a PacBio RS II instrument at the Functional Genomics Center, Zurich, Switzerland. Sequencing of the Zpa63 isolate was performed at the Max Planck-Genome-Centre, Cologne, Germany.

**Genome assembly, and repeat and gene predictions**

For each isolate, we assembled the genome de novo using SMRT Analysis software v.5 (Pacific Bioscience) with two sets of parameters: default parameters and “fungal” parameters. We chose the best assemblies generated by comparison of all assembly statistics produced by the software Quast such as the number of finished contigs, the size of the assembly and the N50 as compared to the genome size (Table S1; S5). In order to exclude poor quality contigs from the raw assemblies, we filtered out the contigs with less than 1.5X and more than 2X median read coverage as these might be unreliable from lack of data or because they contain only repeated DNA [34]. This filter removed a high number of contigs, i.e., between 58 and 17% of contigs, but a only a small quantity of base pairs as compared to the genome size (Table S1; S5). In order to identify the number of fully assembled chromosomes or chromosome arms, we investigated the presence of telomeric repeats (“CCCTAA”) in the assembled contigs using bowtie2 and recorded the presence of more than six repeats at the contig extremities [48–50]. We reported for each contig the number of such blocks of telomeric repeats and considered a contig flanked by these repeats on both sides to be a fully assembled chromosome.

We next used the REPET package to annotate the repeat regions of *Z. ardbiliae* Za17, *Z. brevis* Zb87, *Z.
pseudotriticzi Zp13, Z. passerinii Zpa63, and the three Z. tritici isolates IPO323, Zt05 and Zt10 (https://urgi.ver sailles.inra.fr/Tools/REPET [51, 52]). For each genome, we annotated the repetitive regions as follows: we first identified repetitive elements in each genome using TEdenovo following the developer’s recommendations and default parameters. The library of identified consensus repeats was then used to annotate the respective genomes using TEannot with default parameters.

We used previously published RNA sequencing data to increase the quality of the gene prediction and combined three distinct methodologies [22, 30, 32]. As a first approach, we used GeneMark-ES for an ab initio prediction using the option “--fungus” [53]. Our second and third approaches both used RNA-seq data. For this, we first trimmed the reads using Trimmomatic [54]. We mapped the filtered and trimmed reads to the newly assembled genomes using hisat2 [50] and used the BRAKER1 pipeline to predict genes for each genome using the fungus flag [55]. BRAKER applies GeneMark-ET and Augustus to create the first step of gene predictions based on spliced alignments and to produce a final gene prediction based on the best prediction of the first set [56, 57]. For our third approach, the RNA-seq reads were separately assembled into gene transcripts using Trinity [58]. These were aligned using PASA and EVIDence Modeler to produce consensus gene models from the two independent predictions and the de novo assembled transcripts [59]. Gene counts, length and other summary statistics presented in Table S1 and S3 were obtained using GenomeTools [60] and custom scripts (https://gitlab.gwdg.de/alice.feurtey/genome_architecture_zymoseptoria).

The predicted gene sequences were the basis for an evaluation of the completeness of the assembly and gene prediction by the program BUSCO v.3 [37]. We used this method with the lineage dataset Pezizomycotina. The predicted genes were also used to create a phylogeny with the online implementation of CVtree3, using kmer sizes of 6 and 7 as recommend for fungi [36]. We generated a second tree with the whole assemblies, estimating a distance matrix using the andi software [35]. We predicted orthologs between the newly assembled genomes, the reference Z. tritici genome and the reference Cercospora beticola genome, a related Dothideomycete, which we used as outgroup to identify genes with orthologs restricted to the Zymoseptoria genus [21]. For this, we used the software PoFF [39, 61] which takes into account synteny information in the analyses of similarity inferred by the program Proteinortho [39]. These orthogroups were used to visualize synteny between genomes using Circos [62].

The whole-genome assemblies were used to create a matrix distance with the software andi, from which we generated a tree [35]. A second tree was generated from the gene prediction with the online implementation of CVtree3 [36].

Functional annotations
We used several tools to predict the putative functions for the gene models. First, we used the eggnog-mapper which provide COG, GO and KEGG annotations [63]. The online resource dbCAN2 was run to identify carbohydrate-active enzymes (CAZymes) [64]. Finally, for each genome, we used Antismash v3 (fungal version) to detect biosynthetic gene clusters (Figure S1 [65]).

Additionally, we designed a pipeline to predict protein cell localization and to identify effector candidates. The pipeline for effector prediction is outlined in Figure S1 and includes the software DeepLoc [66], SignalP [67], TargetP [68], phobius [69] and TMHMM [70, 71], which predict the cellular location, the peptide signals and whether proteins are transmembrane. Effector candidates were identified with EffectorP v2 which uses both a new machine learning approach and more complete databases to improve effector prediction compared to the previous version [19]. The pipeline also includes software which are specifically targeted to annotate plant pathogenic functions, namely the program ApoplastP [72] and LOCALIZER [73]. We wrote wrappers scripts, which run the software and create consensus between the different prediction tools providing one command line from the user. These scripts are available at https://gitlab.gwdg.de/alice.feurtey/genome_architecture_zymoseptoria. Briefly, we gathered outputs of several software to predict the cellular location, transmembrane domain and secretion and created a consensus based on the different output to prevent the pitfalls of any one of these methods. From this consensus, we extracted the gene products predicted to be secreted and without a transmembrane domain. The comparisons of genes functions repartition were done by combining predictions of COG categories, secondary metabolite genes with pathogenicity-related gene functional categories such as CAZymes and effector predictions.

Gene expression analyses
To update expression profiles on the new genome assemblies and new gene predictions of the three Z. tritici isolates IPO323, Zt05 and Zt10, we used previously generated RNA-seq data from in planta and in vitro growth [30, 32]. The in planta RNA-seq data was obtained from infected leaves at four different stages corresponding to early and late biotrophic and necrotrophic stages of the three Z. tritici isolates [30]. Strand-specific RNA-libraries were sequenced using Illumina HISEq2500, with 100pb single-end reads for a total read number ranging from 89.5 to 147.5 million reads per sample. This data was previously analyzed [30], using
gene predictions generated from an Illumina-based assembly [22]. The reads were here mapped on the new assemblies of Zt05 and Zt10 and the reference genome of IPO323 after trimming. We used the DESeq2 R package to determine differential gene expression during in planta infection, considering only two infection stages: biotrophic and necrotrophic [74]. Gene expression was assessed as Transcript per Million (TPM). Briefly, TPM is calculated by normalizing read counts with coding region length resulting in the number of reads per kilobase (RPK). RPK total counts per sample are then divided by 1 million to generate a “per million” scaling factor. We calculated the coding region length of each gene with GenomicFeatures R package using the function called “exonsBy” [75]. For gene expression analyses, we further filtered our gene predictions to remove any predicted transposasases and other TE-related annotations based on the EggnoG mapper annotations.

ChiP-sequencing and data analysis
Z. ardabiliae (Za17) and Z. pseudotritici (Zp13) cells were grown in liquid YMS medium for 2 days at 18°C until an OD₆₀₀ of ~ 1 was reached. Chromatin immunoprecipitation and library preparation were performed as previously described [76]. We sequenced two biological and two technical replicates per isolate and used antibodies against the euchromatin histone mark H3K4me2 (#07–030, Merck Millipore). Sequencing was performed at the OSU Center for Genome Research and Biocomputing (Oregon State University, Corvallis, USA) on an Illumina HiSeq2000 to obtain 50-nt reads. The data was quality-filtered using the FastX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), mapping was performed using bowtie2 [77] and peaks were called using HOMER [78]. Peaks were called individually for each replicate, but only peaks that were detected in all replicates were considered and merged for further analysis. Merging of peaks and genome wide sequence coverage with enriched regions was assessed using bedtools [49].

Additional file 1: Table S1. Summary metrics of assemblies filtering process in function of contig coverage quality, presence of telomeric repeats and mitochondrial DNA.

Additional file 2: Table S2. Summary of gene expression in Z. tritici isolates during infection stages.

Additional file 3: Table S3. Enrichment of species-specific and isolate-specific genes per chromosome/contig.

Additional file 4: Table S4. Summary metrics of Z. tritici IPO323, Zt10 and Zt05 expressed genes during in planta infection stages: biotrophic versus necrotrophic.

Additional file 5: Table S5. Summary metrics of Zymoseptoria genome sequencing coverage, assembly size and median coverage per contig.

Additional file 6: Figure S1. Plant-associated genes compartmentalization along the chromosomes. The first track represents core (dark grey) and accessory (light grey) chromosomes/contigs. Telomeric repeats are indicated in orange. Circles from outside to inside represent the position of predicted effector genes (blue), biosynthetic gene clusters (BGC, green) and CAZymes (yellow) per 100 kb windows, respectively.

Additional file 7: Figure S2. A) Intra-species synteny between the reference genome of Z. tritici and the genome of the Z. tritici isolate Zt05. Each color represents a different chromosome as based on the reference Z. tritici genome, except for accessory chromosomes, which are in grey. The connecting lines represent orthologs between each genome. The track between the chromosomes and connecting lines are predicted effector genes. Telomeric repeats are indicated in orange. B) Inter-species synteny between reference genome of Z. tritici and Z. pseudotritici. The arrows represent the large-scale inversions identified between the genomes of these two species.

Additional file 8: Figure S3. Inter-species synteny between genomes of Z. pseudotritici and Z. brevis. A) High level of inter-species synteny at the whole genome level between Z. pseudotritici (dark grey) and Z. brevis (blue). The arrows indicate the contigs identified as carrying rearrangement with Z. tritici but not between Z. pseudotritici and Z. tritici. B – D) Zoom-in on the contigs highlighted with the arrow in part A, showing the details of the pair-wise synteny of both Z.pseudotritici (Zp13) and Z. brevis (Zb87) with the Ztritici reference (IPO323) chromosomes.

Additional file 9: Figure S4. Genome architecture in A) Z. tritici Zt05, B) Z. tritici Zt10. C) Z. brevis Zb87 and D) Z. ardabiliae Za17. Circles from the outside to the inside represent respectively: gene density along chromosomes/contigs; gene expression in vitro (TPM); H3K4me2 distribution in vitro (only for Za17) and species-specific gene density per 100 kb windows. Contigs are ordered based on synteny with chromosomes of the reference strain IPO323. Telomeric repeats are indicated in orange.

Additional file 10: Figure S5. Simplified diagram of the pipeline used to predict the functions and subcellular localization of gene model products.

Additional file 11: Figure S6. Functional gene categories in Zymoseptoria spp. genomes. A) The number of genes in Eggmapper COG categories in addition to Effectors and CAZymes. B) Pathogenicity-related genes of interest: secreted proteins, predicted effectors, secreted CAZymes and non-secreted CAZymes. C) Subcellular localization of predicted gene products.

Additional file 12: Figure S7. Principal component analysis of DESeq2 rlog transformed expression data. Principal component analysis (PCA) of Zymoseptoria tritici IPO323, Zt05 and Zt110 transcripts levels measured in biotrophic (red) and necrotrophic (blue) stages using RNA-sequencing. Reads detected per transcript (counts) were normalized using the size factor method used by DESeq2 package [73]. The PCA plot places biological replicates (four replicates per stage) along the two first PC axes explaining 58% to 65% of the variance (x-axis) and 15% to 25% (y-axis) of the variance within samples, respectively.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1186/s12864-020-06871-w.

Abbreviations
BGC: Biosynthetic gene cluster; BUSCO: Benchmarking universal single-copy orthologies; CAZymes: Genes encoding carbohydrate-active enzymes; COG: Clusters of orthologous groups; DEG: Differentially expressed gene; GO: Gene ontology; H3K27me3: Histone H3, 27th lysine residue trimethylation; H3K4me2: Histone H3, 4th lysine residue di-methylation; KEGG: Kyoto encyclopedia of genes and genomes; PacBio: Pacific biosciences; RPK: Reads per kilobase; TPM: Transcripts per million

Acknowledgements
Not applicable.

Authors’ contributions
AF, CL and MM performed data and results analyses. EHS, AF and CL contributed to the design and implementation of the research. AF, CL, CE and DC performed genome assemblies. MH, JH, MM, MF and KS performed
the experimental procedures. All authors contributed to the writing of the manuscript. All authors read and approved the manuscript.

**Authors’ information**
Present Address: Pathogen Evolutionary Ecology, ETH Zurich, 8092 Zurich, Switzerland.

Present Address: Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland.

**Funding**
CL is funded by the Institut national de la recherche agronomique (INRA) in the framework of a “Contrat Jeune Scientifique” and by the Labex ARBRE (Lab of Excellence ARBRE). Genome research in the group of EHS is funded by the DFG Priority Program SPP1819 and the Canadian Institute for Advanced Research (CIFAR). AF is funded by the DFG Priority Program SPP1819.

**Availability of data and materials**
The datasets generated and analyzed during the current study are available in the NCBI Short Read Archive (https://www.ncbi.nlm.nih.gov/bioproject/?term=) under the BioProject accession numbers: PRJNA638602, PRJNA638603 (Zt05), PRJNA638553 (Zb87), PRJNA638515 (Zp21), PRJNA638382 (Za17), PRJNA4114407 (Zt05 and Zt10). All the data supporting the findings of this study are openly available at https://doi.org/10.5281/zenodo.3820378. The assembled genomes can be found at https://doi.org/10.5281/zenodo.3820378. The functional annotation pipeline, additional scripts and command lines used to create the results presented in this manuscript can be found at https://gitlab.gwdg.de/alice.feurtey/genome_architecture_zymoseptoria. In planta RNA-seq datasets for Z. tritici IPO323, Zt05, Zt10 from [30] are available at the accession number NCBI Gene Expression Omnibus GSE106136 and in vitro RNA-seq data for Za17, Zp31 and Zb87 from [22] are available at the accession number NCBI Bioprojects PRJNA277173, PRJNA277174, and PRJNA277175. The genome sequence of the reference isolate IPO323 available at: http://genome.jgi.doe.gov/Mycgr3/Mycgr3.home.html. The ChIP-seq datasets for the genome reference IPO323 from [6] are available under the accession number PRJNA286790.

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
Pr. Eva Stukenbrock is a member of the editorial board for the BMC Genomics journal.

**Author details**
1Environmental Genomics, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany.
2Environmental Genomics, Christian-Albrechts University of Kiel, 24118 Kiel, Germany.
3INRA Centre Grand Est – Nancy, UMR 1136 INRA/Univérsite de Lorraine Interactions Arbrés/Microorganismes, 54280 Champenoux, France.
4Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland.
5Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA.
6Department of Molecular Genetics and Microbiology, Duke University, Duke University Medical Center, Durham, NC 27710, USA.

Received: 13 December 2019 Accepted: 26 June 2020
Published online: 26 August 2020

**References**
1. Cook DE, Mesarch CH, Thomma BPHJ. Understanding plant immunity as a surveillance system to detect invasion. Annu Rev Phytopathol. 2015;53:541–63.
2. Rouxel T, Gaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, et al. Effector diversification within compartments of the Leptosphaeria maculans genome affected by repeat-induced point mutations. Nat Commun. 2011;2:202.
3. Dong S, Raaffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opin Genet Dev. 2015;35:57–65.
4. Todd RT, Winkl TD, Forche A, Selmecki A. Genome plasticity in Candida albicans is driven by long repeat sequences. Elife. 2019;8.
5. Mehbabi R, Mirzadi Gohari A, Kema GHI. Karyotype variability in plant-pathogenic Fungi. Annu Rev Phytopathol. 2017;55:483–503.
6. Schotanus K, Seyer JL, Connolly LR, Grandaubert J, Happei P, Smith KM, et al. Histone modifications rather than the novel regional centralomers of Zymoseptoria tritici distinguish core and accessory chromosomes. Epigenetics. 2015;8:41.
7. Fokkens L, Shaii S, Connolly LR, Stam R, Schmidt SM, Smith KM, et al. The multi-speed genome of Fusarium oxysporum reveals associations of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events. bioRxiv. 2018;465070.
8. Miao VP, Covert SF, Vaneternet HD. A Fungal Gene for Antibiotic Resistance on a Dispersable (“B”) Chromosome. Science. 1991;254.
9. Temporini E, VanEtten H. An analysis of the phylogenetic distribution of the pea pathogenicity genes of Nectria haematococca MPJ1 supports the hypothesis of their origin by horizontal transfer and uncovers a potentially new pathogen of garden pea. Nectosmospora bonensis. Curr Genet. 2004;46:29–36.
10. Ma L-Y, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature. 2010;464:367–73.
11. Balcsics M-H, Fudal L, Olivier B, Bally P, Grandaubert J, Eber J, et al. The dispensable chromosome of Leptosphaeria maculans shelters an effector gene confriming avrulence Brassica rapa. New Phytol. 2013;198:879–98.
12. van der Does HC, Rep M. Adaptation to the host environment by plant-pathogenic Fungi. Annu Rev Phytopathol. 2017;55:427–50.
13. Lorrain C, Petre B, Duplessis S. Show me the way: rust effector targets in heterologous plant systems. Curr Opin Microbiol. 2018;46:19–25. https://doi.org/10.1016/j.mib.2018.01.016.
14. Toruño TY, Stergiopoulos L, Caoster G. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. Annu Rev Phytopathol. 2016;54:419–41.
15. Zhao Z, Liu H, Wang C, Xu J-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 2013;14:274.
16. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. Annu Rev Plant Biol. 2015;66:513–45. https://doi.org/10.1146/annurev-arplant-043014-114623.
17. Snelders NC, Kettles GJ, Rudd JJ, Thomma BPHJ. Plant pathogen effector proteins as manipulators of host microbiomes? Mol Plant Pathol. 2018;19:257–9. https://doi.org/10.1111/mpp.12628.
18. Shi-Kunne X, Jové R de P, Depotter JRL, Ebert MK, Seidl MF, Thomma BPHJ. In silico prediction and characterisation of secondary metabolite clusters in the plant pathogenic fungus Verticillium dahliae. FEMS Microbiol Lett 2019; 366. doihttps://doi.org/10.1093/femsec/fnd881.
19. Sperchneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol Plant Pathol. 2018;19:2094–110.
20. Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. Origin and domestication of the fungal wheat pathogen Mycosphaerella graminicola via sympatric speciation. Mol Biol Evol. 2007;24:398–411.
21. Stukenbrock EH, Quaedvlieg W, Javan-Nikkhah M, Zalo M, Crous PW, McDonald BA. Zymoseptoria ardeliziae and Z. pseudotectica, two progenitor species of the septoria tritici leaf blotch fungus Z. tritici (synonym: Mycosphaerella graminicola). Mycologia. 2012;104:397–407.
22. Grandaubert J, Bhattacharya A, Stukenbrock EH. RNA-seq-based gene annotation and comparative genomics of four fungal grass pathogens in the genus Zymoseptoria identify novel orphan genes and species-specific invasions of transposable elements. G3 Genes, Genomes Genet. 2015;5:1523–33.
23. Goodwin SB, Ben M'Barek S, Dhillon B, Wittergen AHI, Crane CF, Hane JK, et al. Finished genome of the fungal wheat pathogen Mycosphaerella graminicola reveals Dispensene structure, chromosome plasticity, and stealth pathogenesis. PLoS Genet. 2011;7:e1002070.
24. Habig M, Quade J, Stukenbrock EH. Forward genetics approach reveals host genotype-dependent importance of accessory chromosomes in the fungal wheat pathogen Zymoseptoria tritici. MBio. 2017;8:e001919–7.
73. Sperschneider J, Catanzariti A-M, DeBoer K, Petre B, Gardiner DM, Singh KB, et al. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep. 2017;7:44598.

74. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

75. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9:e1003118.

76. Soyer JL, Möller M, Schotanus K, Connolly LR, Galazka JM, Freitag M, et al. Chromatin analyses of Zymoseptoria tritici: methods for chromatin immunoprecipitation followed by high-throughput sequencing (ChiP-seq). Fungal Genet Biol. 2015;79:63–70.

77. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9:357–9.

78. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89.

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