Complete genome of the Medicago anthracnose fungus, *Colletotrichum destructivum*, reveals a mini-chromosome-like region within a core chromosome

Nicolas Lapalu[1§], Adeline Simon[1§], Antoine Lu[1], Peter-Louis Plaumann[2], Joëlle Amselem[3], Sandrine Pigné[1], Annie Auger[1], Christian Koch[2], Jean-Félix Dallery[1#], Richard J. O'Connell[1#]

[1] Université Paris-Saclay, INRAE, UR BIOGER, 91120 Palaiseau, France

[2] Division of Biochemistry, Department of Biology, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

[3] Université Paris-Saclay, INRAE, URGI, 78000 Versailles, France

[§] These authors contributed equally.

[#] **Corresponding authors:** Richard J. O'Connell, richard.oconnell@inrae.fr; Jean-Félix Dallery jean-felix.dallery@inrae.fr.

Keywords: fungal genomics; accessory chromosome; chromosome rearrangements; segmental duplication; phytopathogenic fungus, *Medicago truncatula*.

Repositories: GEO GSE246592; NCBI BioProject PRJNA1029933.

Abbreviations: AT: acyltransferase domain; AR: accessory region; BDBH: bidirectional best hit; BGC: biosynthetic genes cluster; CAT: conidial anastomosis tubes; CAZyme: carbohydrate active enzyme; CDS: coding sequence; CE: carbohydrate esterase; Chr: chromosome; DNA: deoxyribonucleic acid; GH: glycoside hydrolase; GO; gene ontology; HCT: horizontal chromosome transfer; HPI: hours post-inoculation; KS: ketosynthase domain; LINE: long interspersed nuclear element; LTR: long terminal repeats; MITE: miniature inverted-repeat transposable element; NRPS: non-ribosomal peptide synthetase; PCA: principal component analysis; PCP: peptidyl carrier protein domain; PCR: polymerase chain reaction; PFGE: pulsed-field gel electrophoresis; PKS: polyketide synthase; PL: polysaccharide lyase; RBH: reciprocal best hit; RFP: red fluorescent protein; RNA: ribonucleic acid; SD: segmental duplication; SMKG: secondary metabolism key gene; SMRT: single molecule real time; TE: transposable element; TIR: terminal inverted repeat; TPM: transcript per million.

Impact statement

Colletotrichum is a large genus of fungal phytopathogens that cause major economic losses on a wide range of crop plants throughout the world. These pathogens vary widely in their host specificity and may have either broad or narrow host ranges. Here, we report the first complete genome of the alfalfa (*Medicago sativa*) pathogen, *Colletotrichum destructivum*, which will facilitate the genomic analysis of host adaptation and comparison with other members of the Destructivum clade. We identified a species-specific 1.2 Mb region within chromosome 1 displaying all the hallmarks of fungal accessory chromosomes, which may have arisen through the integration of a mini-chromosome into a core chromosome and could be linked to the pathogenicity of this fungus. We show this region is also a focus for segmental duplications, which may contribute to generating genetic diversity for adaptive evolution. Finally, we report infection by this fungus of the model legume, *Medicago truncatula*, providing a novel pathosystem for studying fungal-plant interactions.

Abstract

Colletotrichum destructivum (*Cd*) is a phytopathogenic fungus causing significant economic losses on forage legume crops (*Medicago* and *Trifolium* species) worldwide. To gain insights into the genetic basis of fungal virulence and host specificity, we sequenced the genome of an isolate from *M. sativa* using long-read (PacBio) technology. The resulting genome assembly has a total length of 51.7 Mb and comprises 10 core chromosomes and two accessory chromosomes, all of which were sequenced from telomere to telomere. A total of 15,631 gene models were predicted, including genes encoding potentially pathogenicity-related proteins such as candidate secreted effectors (484), secondary metabolism key enzymes (110) and carbohydrate-active enzymes (619). Synteny analysis revealed extensive structural rearrangements in the genome of *Cd* relative to the closely-related Brassicaceae pathogen, *C. higginsianum*. In addition, a 1.2 Mb species-specific region was detected within the largest core chromosome of *Cd* that has all the characteristics of fungal accessory chromosomes (transposon-rich, gene-poor, distinct codon usage), providing evidence for exchange between these two genomic compartments. This region was also unique in having undergone extensive intra-chromosomal segmental duplications. Our findings provide insights into the evolution of accessory regions and possible mechanisms for generating genetic diversity in this asexual fungal pathogen.

Data summary

All RNA-seq data were submitted to the NCBI GEO portal under the GEO accession GSE246592. *C. destructivum* genome assembly and annotation are available under the NCBI BioProject PRJNA1029933 with sequence accessions CP137305-CP137317.

Supplementary data (genomic and annotation files, genome browser) are available from the INRAE BIOGER Bioinformatics platform (https://bioinfo.bioger.inrae.fr/). Transposable Elements consensus sequences are also available from the French national data repository, research.data.gouv.fr with doi 10.57745/TOO1JS.

Introduction

The ascomycete fungal pathogen *Colletotrichum destructivum*, causes anthracnose disease on lucerne (alfalfa, *Medicago sativa*) and *Trifolium* species and is responsible for significant economic losses on

these forage legumes [1, 2]. Despite being isolated most frequently from members of the Fabaceae, *C. destructivum* has occasionally been recorded from genera of the Asteraceae (*Helianthus*, *Crupina*), Poaceae (*Phragmites*) and Polygonaceae (*Rumex*) [3, 4]. It has a worldwide distribution that includes the USA, Canada, Argentina, Italy, Netherlands, Greece, Serbia, Morocco, Saudi Arabia, and Korea. *C. destructivum* is a haploid fungus with no known sexual stage [3]. Previous reports of a sexual stage (*Glomerella glycines*) for soybean isolates of *C. destructivum* [5, 6] were based on incorrect identification of the soybean pathogen, which was recently shown to be *C. sojae* [7].

Over the last decade, the application of multi-locus molecular phylogeny approaches has revealed that *C. destructivum* belongs to the Destructivum species complex, which contains 17 accepted taxa [3, 8]. All these plant pathogenic species show distinct host preferences, spanning phylogenetically diverse botanical families. An increasing number of species in the Destructivum complex have now been genome sequenced, namely *C. higginsianum* [9, 10], *C. tanaceti* [11], *C. lentis* [12] and *C. shisoi* [8], which cause disease on Brassicaceae, *Tanacetum* (Asteraceae), *Lens* (Fabaceae) and *Perilla* (Lamiaceae), respectively. The clade therefore provides excellent opportunities for comparative genomic studies on the genetic determinants of host adaptation.

The availability of complete genome sequences is crucial not only for the analysis of large gene clusters, such as secondary metabolism biosynthetic gene clusters, but also for understanding fungal genome evolution. Complete or near-complete genome sequences have enabled the structure and dynamics of accessory mini-chromosomes to be analyzed in several *Colletotrichum* species [9, 13, 14]. The importance of mini-chromosomes for virulence on plant hosts has been demonstrated in several fungal pathogens including *Fusarium oxysporum* f.sp. *lycopersici* [15], *Magnaporthe oryzae* [16], *C. lentis* [12] and *C. higginsianum* [17].

Here, we present the complete genome sequence and gene annotation of *C. destructivum* strain LARS 709, hereafter called *Cd*709, based on long-read sequencing with PacBio Single Molecule, Real-Time (SMRT) Sequel technology. The resulting high-quality chromosome-level assembly allowed us to perform comparative genomics with the close sister species, *C. higginisianum*, highlighting gene content specificity and extensive genomic rearrangements. In particular, the genome showed evidence of multiple segmental duplications, as well as the likely integration of a mini-chromosome into one core chromosome. Although the origin of this integrated region remains to be determined, it displays all the hallmarks of fungal mini-chromosomes. We also show for the first time that *C. destructivum* is pathogenic, and completes its life-cycle, on the model plant *Medicago truncatula*, providing a new tractable pathosystem in which both partners have been genome-sequenced.

Materials and Methods

Fungal and plant materials

The *Colletotrichum destructivum* strains used in this study were originally isolated from *Medicago sativa* in Saudi Arabia (CBS 520.97, LARS 709) and Morocco (CBS 511.97, LARS 202) [2], and are hereafter called *Cd*709 and *Cd*202. The *C. higginsianum* strains used for comparative genome and chromosome analyses were IMI 349063A and MAFF 305635 [10, 17, 18], hereafter called *Ch*63 and *Ch*35, respectively. The fungi were cultured as described previously [18].

Seeds of nine *Medicago truncatula* accessions (Table S1) were provided by the INRAE Centre de Ressources Biologiques Medicago truncatula (UMR 1097, Montpellier, France), while *M. sativa* seeds were purchased from Germ'line SAS (France). *M. truncatula* seeds were first abraded with sandpaper and imbibed with water for 1 h before sowing in seed compost (Floragard Vertriebs-GmbH, Oldenburg, Germany), while *M. sativa* seeds were sown directly in the same compost. All plants were grown in a controlled environment chamber (23°C day, 21°C night, 12-h photoperiod, PPFR 110 µmol·m^{-2}·s^{-1}).

Infection assays and microscopy

To test the susceptibility of *M. truncatula* accessions to *Cd*709, intact plants (17-days-old) were inoculated by first immersing the above-ground parts in a solution of 0.01 % (v/v) Silwet to wet the leaves, then by immersion in a suspension of *C. destructivum* spores (2 x 10^6 ml^{-1}). The inoculated plants were incubated in a humid box inside a controlled environment chamber (25°C, 12-h photoperiod, PPFR 40 µmol m^{-2} s^{-1}). For microscopic examination, pieces of infected tissues were cleared with a 1:3 mixture of chloroform:ethanol for 1h, then with lactophenol for 30 min, before mounting on a microscope slide in 70 % glycerol and imaging with a Leica DM5500 light microscope. Symptoms were recorded at 4 dpi.

Pulsed-field gel electrophoresis (PFGE) and Southern blotting

The plugs containing the conidial protoplasts for PFGE were prepared as previously described [17]. Pulsed-field gel electrophoresis (Bio-rad CHEF-DR II system) was performed using the following conditions: Runtime 260 hours; Switch time 1200 s to 4800 s; 1.5 V / cm; 0.75 x TBE at 8°C. Yeast chromosomal DNA served as size marker (BioRad; 200 kb – 2 Mb).

Southern blotting was conducted using standard protocols [19]. A digoxigenin labeled probe was generated by PCR following the manufacturer's instructions (PCR DIG Probe Synthesis Kit,Roche). The 993 bp probe (*Cd*709 chr1, position 6,711,095 to 6,712,088) was specific to mini-chromosome-like sequences at the right arm of chromosome 1 in *Cd*709. Hybridization was performed in DIG Easy Hyb buffer at 42°C overnight. The membrane was then extensively washed with low and high stringency buffers and subsequently blocked with buffer B2 (1% Blocking powder [Roche] in buffer B1 [100 mM Maleic acid, 150 mM NaCl, pH 7.5]). The blocking solution was then replaced with antibody solution (buffer B2 containing DIG-antibody 1:26,000 (Roche)). The membrane was washed with buffer B1 containing 0.3% Tween20. The membrane was subsequently equilibrated in buffer B3 (100 mM Tris pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$) and developed with chemiluminescence (CDP-Star, Roche).

Genome data, assembly, rearrangements and duplications

The genomic DNA of *Cd*709 was used to prepare a size-selected library (20kb) prior to sequencing with a PacBio Sequel sequencer (kit 2.1, Keygene N.V., Wageningen, The Netherlands) on two SMRT cells, yielding raw data with approximately 224 X genome coverage (1.474.759 reads, N50 10.837 bp). Genome assemblies were generated from several runs of the Hierarchical Genome-Assembly Process version 4 (HGAP4) and Canu [20] assemblers. The draft genome was polished with the Arrow algorithm and the completeness of the assembly was evaluated with BUSCO using the Ascomycota gene set as evidence [21]. The polished assembly was aligned with nucmer against the *Ch*63 and *Ch*35-RFP genomes to visualize chromosome rearrangements. SDDetector [9] was used to detect segmental duplications in combination with Bedtools and BWA-MEM for validation. The *Cd*709 mitochondrial genome was assembled with Organelle_PBA [22] (Table S2).

Transcriptome data and analysis

RNA sequencing was performed on samples of mRNA from undifferentiated mycelium grown axenically and two different stages of plant infection, 48 and 72 h after inoculation, corresponding to the biotrophic and necrotrophic phase, respectively. Mycelium was grown for 3 days in potato dextrose liquid medium (PDB, Difco) at 25°C with shaking (150 rpm) and harvested by filtration. Seedlings of *M. sativa* (8 days old) were inoculated by placing a droplet (10 µl) of *Cd*709 spore suspension (7 x 10^5 spores/ml) onto the surface of each cotyledon and the plants were then incubated as described for *M. truncatula*. Discs of infected cotyledon tissue were harvested using a cork borer (4 mm diameter). After grinding the tissues in liquid nitrogen, total RNA was extracted using the RNeasy plant mini kit (Qiagen). Libraries were then prepared from each sample type using the TruSeq Paired-end Stranded mRNA Kit and sequenced (100 bp reads) using a HiSeq4000 sequencing platform (IntegraGen Genomics, Evry, France). RNA-Seq paired reads were cleaned and trimmed using Trimmomatic [23] and then mapped to the genome assembly of *Cd*709 using STAR [24]. A genome-guided transcript assembly was obtained from mappings with StingTie v1.3.4. Assembled raw transcripts were then filtered based on the TPM distribution per transcript per library.

Genome annotation

Transposable elements (TE) were searched in the *C. destructivum* genome sequence using the REPET package [25, 26]. Consensus sequences identified with the TEdenovo pipeline were classified using the PASTEC tool [27], based on the Wicker hierarchical TE classification system [28], and then manually filtered and corrected. The resulting library of consensus sequences was used to annotate TE copies in the whole genome using the TEannot pipeline.

Protein-coding genes were annotated using the Eugene [29] and FunGAP [30] tools. Predicted genes were filtered out when 10% of their CDS overlapped a Transposable Element predicted by the REPET package. Filtered predicted genes from Eugene and FunGAP were clustered together based on their CDS coordinates (overlap of one base required) with no strand consideration. The Annotation Edit Distance (AED) [31] was computed with transcript and protein evidence for each transcript and the predicted model with the best score was retained at each locus. Mitochondrial genomes were annotated with MFannot [32] and MITOS2 [33]. Results were manually inspected and in case of divergence between the predictions, the longer gene model was retained.

The synteny between *C. destructivum* and *C. higginsianum* proteomes was analysed with SynChro [34] which detects ortholog proteins with Reciprocal Best Hit (RBH), based on 40% similarity and a length ratio of 1.3. Colinear orthologs were then grouped in syntenic blocks, according to a delta threshold = 1 (very stringent mode). Non-syntenic blocks were extracted when 5 or more consecutive non-syntenic genes were found. Proteome similarities with other *Colletotrichum* species were performed with Blast 2.2.28+ and the results filtered with a cut-off of 30% identity and 50% query coverage. Proteome synteny and associated figures were obtained using Clinker [35].

Functional annotation of predicted genes

Functional annotations of genes obtained using Interproscan 5.0 [36] and Blastp (e-value <1e-5) [37] against the NCBI nr databank (September 2019) were then used to perform Gene Ontology [38] annotation with Blast2GO [39]. Carbohydrate active enzymes (CAZymes) were annotated with dbCAN2

[40] launching HMMER, Diamond and Hotpep against dedicated databases. Genes were considered as CAZymes when at least 2 of the three tools provided a positive annotation

Genes encoding potential secreted proteins were predicted with a combination of SignalP v4.1 [41], TargetP v1.1 [42] and TMHMM v2.0 [43] results. The secretome was defined as the union of SignalP and TargetP results and then intersected with TMHMM results (0 or only 1 transmembrane domain). Proteins smaller than 300 amino acids were then extracted and considered as Small Secreted Proteins (SPPs). In parallel, EffectorP v2.0 [44] was applied to the predicted secretome to identify putative effector proteins. Finally, the intersection of EffectorP and SPPs results was retained to establish a list of potential effectors.

To detect secondary metabolism biosynthetic gene clusters (BGCs), predicted genes were submitted to antiSMASH (Antibiotics and Secondary Metabolite Analysis Shell) v5 [45]. Only core biosynthetic genes (commonly known as secondary metabolism key genes, SMKGs) were considered for further analysis. Presence/absence patterns of SMKGs were based on reciprocal best hits with *Ch*63 and *Ch*35, and then manually inspected. Among the newly predicted secondary metabolism key genes (SMKGs), those encoding polyketide synthases (PKS) and non-ribosomal peptide synthases (NRPS) were checked for the presence of the minimal expected set of enzymatic domains, namely KS and AT domains for PKS, and A and PCP domains for NRPS. Terpene synthases and dimethylallyltryptophan synthase (DMATS) genes were manually inspected and retained if they had RNA-seq or protein support. Those *Cd*709 genes not predicted as SMKGs by antiSMASH, but orthologous to a *C. higginsianum* SMKG were also included. For example, antiSMASH failed to annotate six terpene synthase (TS) that are present in both species.

Codon usage analysis

Codon usage was computed for predicted gene coding sequences (CDS) on each chromosome or region using the EMBOSS tool 'cusp'. The resulting codon usage matrix (i.e. the fraction of each codon in a given amino acid) was subjected to Fisher's exact tests (with a Bonferroni correction for multiple testing) to address the statistical significance of differences between the core and mini-chromosomes. The matrix was also subjected to a Principal Component Analysis (PCA) and the results were projected onto the first two principal components. To analyse the GC percentage of the three letters of each codon, the 'cusp' tool was run individually on each CDS of each chromosome or region and the results were represented as density plots. The corresponding figures were generated using R (v. 4.0.5) and the libraries ggplot2 (v. 3.3.3), cowplot (v.1.1.1) and ggbeeswarm (v. 0.6.0), all available from the CRAN repository (https://cran.r-project.org/).

Results

A novel *Colletotrichum destructivum* - *Medicago truncatula* pathosystem

The cell biology of infection of *Medicago sativa* by *C. destructivum* isolate 709 (*Cd*709) was previously described [2]. Here, we report infection of the model plant *M. truncatula* (barrel medic) by this species. Five out of the nine tested *M. truncatula* accessions, including the genome-sequenced accession ESP074-A [46], were found to be susceptible to *C. destructivum* in two independent infection assays (Fig. 1, Table S1). At 4 days post inoculation (hpi), necrotic water-soaked lesions were visible on the trifoliate leaves of the susceptible accessions (Fig. 1). In contrast, the leaves of resistant accessions

presented only small necrotic flecks or no visible symptoms. The genome-sequenced accession R108-C3, which is widely used for *M. truncatula* functional genomics [47], was resistant to *C. destructivum* in these infection assays.

On cotyledons of the susceptible *M. truncatula* accession ESP155-D, *Cd*709 spores germinated to form melanized appressoria, which by 48 hpi had penetrated host epidermal cells to form bulbous, intracellular biotrophic hyphae that were confined to the first infected cell (Fig. 2a). Thinner necrotrophic hyphae started to emerge from the tips of the biotrophic hyphae at 60 hpi (Fig. 2b), and after 72 hpi the fungus had completed its asexual cycle by producing sporulating structures (acervuli) on the surface of the dead tissues (Fig. 2c). On cotyledons of the resistant accession ESP163-E, appressoria formed abundantly on the leaf surface but penetrated host epidermal cells very infrequently (Fig. 2d, e). Groups of dead epidermal cells underlying the appressoria appeared yellow-brown in colour and had granulated contents, suggesting they had undergone a hypersensitive cell death response. Rarely, small hyphae were visible in epidermal cells beneath appressoria but they developed only a short distance into the dead cells and most remained smaller than the appressorium. Acervuli were never observed on plants of accession ESP163-E.

Genome assembly and structural annotation

Long-read data allowed us to generate a complete genome assembly for *Cd*709, with a total length of 51.75 Mb in which all 12 chromosomes were sequenced from telomere to telomere (Fig. 3), together with the circular mitochondrial genome (34 kb). Annotation of transposable elements revealed a total of 49 consensus sequences, representing all the possible TEs in the *Cd*709 genome. Classification of the TEs (Table S3) showed that the genome contains 18 different families of retrotransposons, including eleven LTR (Long Terminal Repeats) and seven LINE (long interspersed nuclear element), 28 DNA transposons, including 25 TIR (terminal inverted repeat), one helitron and two MITE (Miniature Inverted-Repeat Transposable Elements), as well as three unclassified repeated elements. The library of 49 consensus sequences was then used to annotate TE copies in the *Cd*709 genome. Overall, TEs covered 6.2 % of the genome assembly by length. The Class I LTR Gypsy superfamily was the most abundant in terms of coverage and number of copies, whereas the Class I TIR Tc1-Mariner was the most abundant in terms of full-length copies. Two Gypsy transposons (R172 and G87) resemble the most abundant TE family in *C. higginsianum*, namely the LTR transposon family RLX_R119 [9]. Looking at the distribution of TE families along the chromosomes, we found that the telomeres of all twelve *C. destructivum* chromosomes were associated with a single copy of a TE belonging to the helitron family (G103).

To annotate the protein-coding genes, a genome-guided assembly of RNA-Seq reads provided 16,122, 13,901 and 15,081 transcripts for axenic mycelium, 48 hpi and 72 hpi libraries, respectively (Table S4), with 1.88 TPM, 9.38 TPM and 4.90 TPM as minimum expression levels, respectively (Fig. S1). Assembled transcripts were then used to predict gene models in conjunction with *Colletotrichum* and Ascomycota protein databanks. The results of EuGene and FunGap were combined and filtered to generate the *Cd*709 gene set comprising 15,631 complete gene models, of which 11,853 had transcript support and 15,172 resembled Ascomycota predicted proteins. Features of the gene annotation are summarized in Table S5. The completeness of this annotation was confirmed by comparison to the BUSCO Ascomycota set (1,315 genes), with 1,309 complete genes predicted and only one missing. Functional annotation assigned InterPro entries to 10,298 genes, among which 7,475 had at least one

GO term and 1,105 were potential enzymes (annotated with an Enzyme Code). Based on Blast2GO descriptions, 12,192 predicted genes (78%) had a predicted function, i.e. a description other than "hypothetical protein" (Table S6 tab 'All'). The mitochondrial genome of *Cd*709 was annotated with 29 tRNAs, 2 rRNAs (small and long subunit) and 21 genes.

Plant interaction-related genes

A total of 619 *Cd*709 genes were annotated to encode CAZymes, among which 410 were assigned to the Glycoside Hydrolase (GH), Carbohydrate Esterase (CE) and Polysaccharide Lyase (PL) CAZyme classes (Table S6 tab 'CAZyme'). The proportion of genes in each CAZyme class closely resembled that previously found in *Ch*63 [48], and 98% (400/410) of *Cd*709 CAZyme genes were also detected in the *Ch63* genome. *In silico* analysis of the *Cd*709 secretome revealed a total of 2,608 potential extracellular secreted proteins, including 1,118 small proteins (<300 amino acids). Among these, 484 genes were retained as putative effectors because they were also present among 508 genes identified by EffectorP. Comparing these to the effector repertoire of *Ch*63, a total of 127 putative effectors (26.2%) were unique to *Cd*709, having no Reciprocal Best Blast Hit in *Ch*63 (Table S6 tab 'Predicted effectors'). A total of 110 secondary metabolism key genes (SMKGs) were detected In the *Cd709* genome using the fungal version of antiSMASH and were manually curated. These *C. destructivum* SMKGs were compared to the 105 *C. higginsianum* SMKGs [9]. Overall, 78 % (94 out of 120) of the SMKGs were present in both species (Table S6 tab 'Secondary metabolism', Fig. S2). A total of 17 *C. destructivum* SMKGs, distributed over eight BGCs, were not detected in *C. higginsianum*.

Chromosome structure comparison

Complete chromosome-level assemblies are available for two different *C. higginsianum* strains, namely IMI 349063A (*Ch*63) [9] and MAFF 305635-RFP (*Ch*35-RFP), a transformant of MAFF 305635 (*Ch*35) expressing red fluorescent protein which lacks both mini-chromosomes 11 and 12 [10, 17]. The genetic proximity of *C. destructivum* and *C. higginsianum* allowed us to align assemblies to observe chromosome structural variations. This generated 38 Mb of *C. destructivum* alignments (>10 kb) with each *C. higginsianum* strain, ranging from 88 to 96.7% identity. Thus, *C. destructivum* shared approximately 73.6 % of its total genome length with *C. higginsianum*. At the chromosome scale, alignments revealed that five chromosomes of *C. destructivum* (chr1, 2, 3, 5 and 9) were not involved in any large rearrangements, five others (chr4, 6, 7, 8 and 10) showed inter-chromosomal rearrangements, while the two mini-chromosomes (chr11 and 12) lacked large regions of conserved sequences and appear species-specific (Fig. 4A).

One rearrangement involved chr7 and chr8 of *Cd*709 resulting in chr4 and chr10 of *Ch*63. The break-points in chr7 and chr8 were associated with TEs in *Cd*709 (Fig. S3 A and B). A similar rearrangement was found relative to *Ch*35-RFP, albeit with different break-points in both species that were not associated with TEs (Fig. S3 F and G). A second rearrangement involved chr4 and chr10 of *Cd*709 such that their left and right arms result in chr9 and chr7 of *Ch*63, respectively (Fig. S3 C and D). Interestingly this rearrangement was not found relative to *Ch*35-RFP, suggesting that it is specific to particular *C. higginsianum* strains, as was noted previously [10]. A third inter-chromosomal rearrangement concerned 121 kb at the 5' extremity of *Cd*709 chr6 coming from chr4 and contig_1 of *Ch*63 and *Ch*35-RFP, respectively. In *C. destructivum*, this break-point is surrounded by TEs and non-syntenic regions (Fig. S3 E). Remarkably, a specific rearrangement of 42 kb between chr11 of *Cd*709 and contig 11 of *Ch*35-RFP (Fig. S3 H) corresponds to a region that is absent from the *Ch*63 genome assembly and which

encodes highly variable effectors (having ≤ 90% alignment coverage) and secondary metabolism-related proteins [10]. In addition, several short stretches (2 to 5 kb in length) from chr11 of *Cd*709 were present at the extremities of chromosome 6 in *Ch*63 and the corresponding region of *Ch*35-RFP (contig _9) (Fig. 4).

A notable feature of the *C. destructivum* genome assembly is the unusually large size of chr1 (7.3 Mb), which is 0.9 Mb longer than the largest chromosome in *C. higginsianum* (6.4 Mb). Genome alignments highlighted a near-complete synteny between chr1 of *Cd*709 and chr2 of *Ch*63 except for a 1.2 Mb subtelomeric region (coordinates chr1:6076875-7282542), for which no similarity was found in *C. higginsianum* (Fig. 4). Synteny between the genes of *Cd*709 and those of *Ch*63 was investigated using SynChro. With stringent settings, 400 syntenic blocks were identified based on 12,135 Reciprocal Best Hits. A total of 1,083 genes were found in 47 non-syntenic blocks composed of at least five consecutive *Cd*709-specific genes (Tables S7 & S8). The largest non-syntenic block, corresponding to the 1.2 Mb region specific to Cd709 on chr1, contained 305 genes. Mini-chromosome chr12 contained one non-syntenic block of 170 genes, while chr11 was divided into seven non-syntenic blocks, the largest containing 106 genes. Although only 356/1,083 genes inside non-syntenic blocks could be annotated with a GO term, GO enrichment tests revealed that the *Cd*709-specific genes were enriched in protein kinases, protein phosphorylation activity and secondary metabolism process (Table S9). Likewise, effector genes were found to be enriched in non-syntenic blocks whereas CAZymes were depleted (Table S10).

Validation of the 1.2 Mb non-syntenic region in *C. destructivum* chromosome 1

To verify the large non-syntenic region identified within chr1, we first checked for potential errors in the sequence assembly of this region by manually inspecting long reads spanning the two junctions (Fig. S4). Secondly, to obtain an assembly-independent validation, pulsed-field gel electrophoresis (PFGE) and a Southern hybridization were performed (Fig. 5A, B). A 993 nt probe (coordinates chr1: 6,711,095 to 6,712,088) was designed within the 1Mb non-syntenic region to target a unique locus that avoided TEs (Fig. 5B). This probe is 83.5% identical to the gene CH63R_14488 located on chromosome 11 of *Ch*63 that was used as a hybridization control.

Chromosomes of two *C. destructivum* isolates (*Cd*709 and *Cd*202) and two *C. higginsianum* isolates (*Ch*63 and *Ch*35) were separated by PFGE and analysed by Southern hybridization (Fig. 5C, D). For both *C. destructivum* isolates, the probe hybridized to molecules with high molecular weight that could correspond to the largest chromosome, consistent with a location on chr1 (Fig. 5C, D). The high molecular weight signals were absent in the *C. higginsianum* blots, and instead hybridization signals were detected at a position corresponding to mini-chromosome 11, although these were weak, as expected for a probe with only 83.5% identity to the target. Overall, our findings validate that a non-syntenic region is embedded within chr1 of *C. destructivum*. Hereafter, we refer to the syntenic and non-syntenic portions as chr1A and chr1B, respectively, and their distinct properties were explored further in the following analyses.

Region chr1B shows the characteristic features of fungal accessory chromosomes

In many aspects, the region chr1B of *Cd*709 resembled the mini-chromosomes 11 and 12. All three compartments were more AT-rich than the core genome. Region chr1B was also highly enriched with TEs, having 32.8 % coverage with TE copies by length, similar to chr11 and chr12 (32.3 and 35.1 %,

respectively), whereas the core chromosomes (excluding chr1B) had only 3 to 6.2 % TE coverage (Table 1, Fig. 3, Table S11). Moreover, the distribution of TE families in region 1B and the two mini-chromosomes differed markedly from the core chromosomes in that they were all enriched with LINE retrotransposons (44 %, 19 % and 34 % coverage, respectively), compared to only 7 % in the core genome (Table 3). LINE TEs are also present in *C. higginsianum* on mini-chromosomes 11 and 12, but their expansion was less striking in this species (7 % and 2 % coverage, respectively) (Fig. S5) than in *Cd*709 [9].

Examination of the gene content of region chr1B revealed that, similar to the mini-chromosomes, it was overall depleted in protein-coding genes (2-fold less than the core chromosomes), contained a significantly larger proportion of genes encoding proteins of unknown function (i.e. annotated as hypothetical proteins), and had fewer expressed genes (RNA-seq transcript evidence) compared to the core genome (Table 1). Considering categories of potentially pathogenicity-related genes, no CAZyme genes or SMKGs were detected in either region chr1B or chr 12, although eight SMKGs were present on chr11 (Table S6, Tab 'Secondary metabolism'), all of which had RNA-seq transcript support. Moreover, 38 effectors were found in chr1B and the two mini-chromosomes. Remarkably, 36 of these were absent from *C. higginsianum* (had no RBH in *Ch*63), of which 20 were expressed *in planta* (Table S6 tab 'Predicted effectors). With 15 and 10 effectors respectively, the mini-chromosomes 11 and 12 were significantly enriched in putative effectors compared to the core chromosomes whereas no enrichment was observed for the 13 effectors of the chr1B (Table 1). Remarkably, the most highly expressed effectors during the biotrophic phase (48 hpi), namely CDEST_01870 (chr1B) and CDEST_15472 (chr12), were located on mini-chromosome-like regions. This raises the possibility that genes carried in such regions are important for virulence.

Codon usage in region chr1B and the mini-chromosomes differs from the core chromosomes

Analyses of codon usage were used previously to detect differences between the core and accessory chromosomes or lineage-specific compartments of other plant pathogenic fungi [15, 49, 50]. We therefore computed the codon usage of CDS located on the core chromosomes, mini-chromosomes and the chr1B region of *Cd*709. Based on a principal component analysis, codon usage on the core chromosomes was very homogeneous, whereas that of the mini-chromosomes and region chr1B clustered together and separately from the core chromosomes (Fig. 6A). To illustrate this in greater detail, we plotted the codon usage for each amino acid and for each chromosome or region (Fig. S6, representative examples are given for 3 amino acids in Fig. 6B). For these analyses, we excluded the two amino acids (Trp and Met) that are encoded by a single codon. Based on Fisher's exact tests for each of the remaining 59 codons, almost all the codon usages were different between the core chromosomes on one hand and chr1B, chr11 or chr12 on the other hand. In striking contrast, there were only three differential codon usages between chr1B and chr11 and one between chr1B and chr12. However, chr11 and chr12 were most different from each other with 15 differential codons (Table S12; adjusted $P < 0.001$).

Region chr1B is a hotspot for segmental duplications

The genome of *Cd*709 was inspected for segmental duplications, as described previously for *C. higginsianum* [9]. A total of 48 duplications involving genes were detected on four chromosomes (chr1, chr6, chr11 and chr12). Among them, 12 duplications were larger than 10 kb (Fig. 7) of which only three were inter-chromosomal (all involving chr12). Similar to *C. higginsianum* [9], these inter-

chromosomal duplications were all associated on at least one side with TEs, supporting a potential role of TEs in duplication (Fig. S7). However, in contrast to *C. higginsianum*, these duplications did not take place preferentially near telomeres.

A remarkable feature of region chr1B was that it showed a strong intra-chromosome duplication pattern, with some regions replicated up to three times (Fig. 7). Assembling large duplications can be difficult even with long-read sequences [51]. To check for possible bias during assembly, the eight largest intra-chromosome duplications on chr1B were inspected manually (Table S13). Due to the problem of multiple reads mapping to duplicated regions, we considered only uniquely mapped reads. Consequently, the read-coverage of these eight regions was on average 2-fold lower than the non-duplicated regions. No other regions of chr1B showed a significant decrease in coverage, and the extremities of the SD regions were well-anchored to chr1B. Reads were identified spanning the two smallest duplications, SD1B-2 (10 reads) and SD1B-6 (22 reads), but other duplicated regions were too large (>16 kb) to be spanned by single PacBio reads. Finally, the short-read RNA-seq data used to annotate the genome were also employed to detect mutations within the duplicated genes. Mutations were detected in all the duplicated regions, albeit with support from only few reads in most cases. Taken together, these results support the reliability of the observed duplications in region chr1B.

To gain insight into the possible origin of region chr1B, we examined conservation of the 300 genes contained within this region in the genomes of 23 other *Colletotrichum* species (Table S14). As expected, given that *C. destructivum* and *C. higginsianum* belong to the same species complex [3], the total proteome of *Cd*709 showed greatest similarity to that of *Ch*63 (14,372 conserved proteins). Surprisingly however, the chr1B proteome shared most conserved proteins with a phylogenetically distant species, namely *C. truncatum* (217 protein matches, compared to only 134 matches in *C. higginsianum*) [52]. Almost half of the genes shared with *C. truncatum* were involved in segmental duplications within the *Cd*709 chr1B. Remarkably, the region triplicated in SD1B-1, SD1B-3 and SD1B-7 was also found in a large duplicated region represented by two contigs within the *C. truncatum* genome assembly (Fig. S8) [53], which may be located on a mini-chromosome due to their low GC content (49.0%, compared to 51.2% in the longer contigs of *C. truncatum*). Other genes located within the *Cd*709 SD1B-1 duplications had Blast matches that were mostly restricted to *C. incanum, C. spaethianum* and *C. tofieldiae* (Spaethianum species complex), *C. salicis* and *C. nymphaeae* (Acutatum species complex), *C. fructicola* (Gloeosporioides species complex), *C. sublineola* (Graminicola species complex) and *C. orchidophilum*, which vary in their phylogenetic distance from *C. destructivum* [52]. The absence of these gene sequences from the *C. higginsianum* genome was confirmed by Tblastn searches against the NCBI wgs Colletotrichum database (266 genomes).

Examination of the gene content in duplicated regions of chr1B gave few clues to their possible role in the host interaction or the advantage for the fungus to maintain multiple mutated copies of these genes. One gene duplicated four times (CDEST_01898, CDEST_01949, CDEST_02058 and CDEST_02116) encoded a major facilitator superfamily transporter. The five genes duplicated between SD1B-2 and SD1B-6 comprised four FAD-binding domain-containing proteins and a patatin-like serine hydrolase.

Discussion

In this study, we present a chromosome-level reference assembly of the *C. destructivum* genome, a phytopathogen causing anthracnose disease principally on species of *Medicago* and *Trifolium* (Fabaceae). Among other members of the Destructivum species complex, which currently contains 17

recognised species [3], the genomes of *C. lentis, C. tanaceti* and *C. shisoi* were sequenced previously but the resulting assemblies were highly fragmented, containing 2980, 5242 and 36,350 contigs, respectively [8, 11, 12]. Using PacBio long-read sequencing, we were able to generate a gapless assembly of the *Cd*709 genome which, together with that of *Ch*63 [9], provides a second complete genome within the Destructivum species complex, facilitating future comparative genomic analyses within this important group of plant pathogens.

Alignment of the *Cd*709 genome assembly with those of *C. higginsianum* strains *Ch*63 and *Ch*35 revealed large-scale chromosome rearrangements between the two closely-related species. Some of these rearrangements were potentially mediated by recombination between homologous regions containing TEs, which flanked one or both of the breakpoints. Similar TE-mediated chromosome rearrangements were previously reported at the intra-species level in *C. higginsianum* [10]. Our analysis of synteny between the genomes of *Cd*709 and *Ch*63 also revealed the presence of a 1.2 Mb species-specific region within Chr1 of *Cd*709, which we called Chr1B. This 'accessory region' (AR) displays many of the hallmarks that characterize fungal mini-chromosomes, or 'accessory chromosomes', in that it is AT-rich, transposon-rich, gene-poor and has a distinct codon usage [50, 54–56]. In all these respects, Chr1B resembles the mini-chromosomes Chr11 and Chr12 but is strikingly different from the rest of Chr1 and other core chromosomes of *Cd*709. The TE enrichment observed in Chr1B and both mini-chromosomes is largely caused by the specific expansion of LINE and TIR elements in these compartments, unlike the core chromosomes where the Gipsy TE family predominates.

Using PFGE and Southern hybridization with a probe specific to Chr1B, we were able to confirm that this AR is carried not only on Chr1 of *Cd*709 but also on the largest chromosome of *Cd*202, despite the widely-separated geographical origins of these two isolates (Saudi Arabia and Morocco, respectively). Analysis of a larger collection of *C. destructivum* isolates is now needed to determine the extent to which Chr1B is conserved within this pathogen species. To the best of our knowledge, the presence of an AR embedded within a core chromosome has not been found in any genome-sequenced *Colletotrichum* species, although there are precedents in other plant pathogenic fungi. For example, isolates of the T race of *Cochliobolus heterostrophus* harbor an AR of about 1.2 Mb distributed between two core chromosomes that contains the *Tox1* locus producing the T-toxin polyketide [57, 58]. In *Verticillium dahliae*, Chr3 and Chr4 each harbor two ARs of ~300 kb [59], while in *Fusarium poae* a 204 kb block with AR characteristics is inserted near one telomere of Chr3 [56]. However, it should be noted that in these two examples the inserted AR blocks are 4- to 6-fold smaller than Chr1B of *Cd*709.

Our working hypothesis is that the AR Chr1B arose by the integration of a mini-chromosome into a core chromosome of *C. destructivum*, but the mechanism by which this occurred is unclear. Despite its subtelomeric position of Chr1B, its integration is unlikely to have resulted from the telomeric fusion of a mini-chromosome with a core chromosome because it is flanked on both sides by portions of Chr1, both of which are highly syntenic to Chr2 of *C. higginsianum*. In other fungi, chromosome breakage-fusion-bridge (BFB) cycles have been invoked not only in the creation of accessory chromosomes from core chromosomes [60], but also in their reintegration into core chromosomes [61].

A distinguishing feature of the Chr1B AR is that it has undergone extensive region-specific segmental duplications. Some inter-chromosomal SDs in *Cd*709 were associated with TEs at one or both of their borders, as we found previously in *Ch*63 [9], but there was little evidence that the region-specific SDs in Chr1B were mediated by TEs. Among fungal pathogens, SDs can play important roles in generating

genetic diversity and novel gene functions, either at the level of expression or coding sequence [62, 63]. A recent study on *Fusarium* strains infecting banana also highlighted the importance of SDs in driving the evolution of ARs and the effector genes contained within them [64]. Although the *C. destructivum* genome contains a complete Mat1-2-1 mating-type locus (Table S6, Tab MAT1-2-1), and should therefore be capable of sexual reproduction, this has never been observed [65], [3]. In this context, segmental duplication may therefore provide an important mechanism for generating genetic diversity for host adaptation in this essentially asexual pathogen.

A remarkable finding was that some segmentally duplicated blocks of genes within Chr1B of *C. destructivum* are conserved and syntenic with duplicated regions in the genome of *C. truncatum*, a species that is phylogenetically very distant [52]. Given that these two taxa diverged ~60 million years ago [66], soon after speciation in Colletotrichum, these SDs may be very ancient and have been selectively retained in some species and lost in others. Alternatively, these duplicated regions may have been acquired by horizontal transfer of a mini-chromosome (HCT) from another species to a common ancestor, or through independent transfers to *C. destructivum* and *C. truncatum*. HCT would be consistent with the distinct codon bias in Chr1B and the taxonomic incongruity of many genes within this region. The horizontal transfer of a mini-chromosome between vegetatively incompatible biotypes of *C. gloeosporioides* was shown experimentally [67, 68], and it is well-documented that genetic material can be exchanged following fusion between conidial anastomosis tubes of the same, or even different, *Colletotrichum* species [69–71].

Chr1B contains a variety of genes with potential roles in fungal virulence, some of which were expressed during infection. These include genes encoding 13 candidate secreted effector proteins, 8 protein kinases, 5 major facilitator superfamily membrane transporters, 5 heterokaryon incompatibility (HET) proteins and 8 putative transcription factors (TFs) (Table S6). It is interesting to note that, similar to Chr1B, the accessory 'pathogenicity chromosome' of *Fusarium oxysporum* f.sp. *lycopersici* is enriched not only with effectors genes but also with genes encoding protein kinases, membrane transporters, HET proteins and TFs, of which one TF was shown to regulate the expression of plant-induced effector genes [72],[73]. TFs were also found to be enriched in the four lineage-specific ARs of *V. dahliae* [59]. Overall, the gene content of Chr1B suggests that it may contribute to *C. destructivum* pathogenicity. This was demonstrated experimentally for ARs in two other members of the Destructivum species complex, namely Chr11 of *C. higginsianum* (isolate *Ch*35) which was essential for virulence on *A. thaliana* [17], and Chr11 of *C. lentis*, which was required for virulence on lentil [12]. In the case of *Cd*709, it is noteworthy that the three most highly expressed and plant-induced effector genes are all located in ARs, namely CDEST_01870 on Chr1B, CDEST_15404 on Chr11 and CDEST_15472 on Chr12. These and other pathogenicity-related genes carried within these genomic compartments will provide interesting candidates for future functional analysis.

Finally, we show here that *Cd*709 can complete its life cycle not only on its original host, *M. sativa*, but also on the widely-studied model legume, *M. truncatula*. Until now, the only other *Colletotrichum* species known to attack *M. truncatula* was *C. trifolii*, which belongs to the phylogenetically distant Orbiculare species complex and uses a different infection process where the biotrophic phase extends to many host cells [74, 75]. With complete genome assemblies and high-quality gene annotations available for both partners, together with abundant genetic tools and resources on the plant side, the *C. destructivum* - *M. truncatula* interaction could provide a tractable new model pathosystem for studying hemibiotrophic fungal interactions with Fabaceae hosts. Our identification of susceptible and

resistant *M. truncatula* accessions also raises the possibility that natural variation among accessions could be exploited to analyse the genetic basis of resistance to *C. destructivum* [76].

Conflicts of interest:

The authors declare that there are no conflicts of interest.

Funding:

This work was partly supported by funding from the Agence Nationale de la Recherche (ERA-CAPS grant ANR-17-CAPS-0004-01) to R.J.O. The BIOGER unit benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). The Funders had no role in the study design, data analysis, data interpretation or decision to publish.

Author contributions:

Conceptualization: NL, AS, CK, JFD, RJO; Investigation: PLP, SP, AA, RJO; Formal analysis: NL, AS, AL, JA, JFD; Visualization: NL, AS, PLP, JA, CK, JFD, RJO; Writing – Original Draft: NL, AS, PLP, JA, CK, JFD, RJO; Writing – Review & Editing: NL, AS, CK, JFD, RJO; Supervision: NL, CK, RJO; Funding acquisition: RJO.

Acknowledgements:

We are grateful to the following bioinformatics platforms and partners for providing help and/or computing and/or storage resources: Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, doi: 10.15454/1.5572369328961167E12), CATI BARIC (https://www.cesgo.org/catibaric/), and INRAE-LIPME Bioinfo (Sébastien Carrère).We also thank Dr Alexander Wittenberg (KeyGene N.V.,The Netherlands) for help with sequencing, and the INRAE Centre de Ressources Biologiques *Medicago truncatula* for providing seeds.

References

1. **Frayssinet S**. Colletotrichum destructivum: a new lucerne pathogen in Argentina. *Australasian Plant Disease Notes 2008 3:1* 2008;3:68–68.

2. **Latunde-Dada AO, Bailey JA, Lucas JA**. Infection process of Colletotrichum destructivum O'Gara from lucerne (Medicago sativa L.). *European Journal of Plant Pathology* 1997;103:35–41.

3. **Damm U, O'Connell RJ, Groenewald JZ, Crous PW**. The Colletotrichum destructivum species complex – hemibiotrophic pathogens of forage and field crops. *Studies in Mycology* 2014;79:49–84.

4. **Sun HY, Liang Y**. First report of anthracnose on sunflower caused by colletotrichum destructivum in China. *Plant Disease* 2018;102:245.

5. **Manandhar JB**. Colletotrichum destructivum, the Anamorph of Glomerella glycines. *Phytopathology* 1986;76:282.

6. **Tiffany LH, Gilman JC**. Species of Colletotrichum from Legumes. *Mycologia* 1954;46:52–75.

7. **Damm U, Sato T, Alizadeh A, Groenewald JZ, Crous P**. The Colletotrichum dracaenophilum, C. magnum and C. Orchidearum Species Complexes. *Studies in mycology*;92. Epub ahead of print 2019. DOI: 10.1016/J.SIMYCO.2018.04.001.

8. **Gan P, Tsushima A, Hiroyama R, Narusaka M, Takano Y, *et al.*** Colletotrichum shisoi sp. nov., an anthracnose pathogen of Perilla frutescens in Japan: molecular phylogenetic, morphological and genomic evidence. *Scientific Reports*;9. Epub ahead of print 1 December 2019. DOI: 10.1038/s41598-019-50076-5.

9. **Dallery J-F, Lapalu N, Zampounis A, Pigné S, Luyten I, *et al.*** Gapless genome assembly of Colletotrichum higginsianum reveals chromosome structure and association of transposable elements with secondary metabolite gene clusters. *BMC Genomics* 2017;18:667.

10. **Tsushima A, Gan P, Kumakura N, Narusaka M, Takano Y, *et al.*** Genomic Plasticity Mediated by Transposable Elements in the Plant Pathogenic Fungus Colletotrichum higginsianum. *Genome biology and evolution* 2019;11:1487–1500.

11. **Lelwala R V., Korhonen PK, Young ND, Scott JB, Ades PK, *et al.*** Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in Colletotrichum tanaceti. *PLOS ONE* 2019;14:e0212248.

12. **Bhadauria V, MacLachlan R, Pozniak C, Cohen-Skalie A, Li L, *et al.*** Genetic map-guided genome assembly reveals a virulence-governing minichromosome in the lentil anthracnose pathogen Colletotrichum lentis. *The New phytologist* 2019;221:431–445.

13. **Wang Haoming, Huang Rong, Ren Jingyi, Tang Lihua, Huang Suiping, *et al.*** The evolution of mini-chromosomes in the fungal genus Colletotrichum. *mBio* 2023;14:e00629-23.

14. **Plaumann P-L, Koch C**. The Many Questions about Mini Chromosomes in Colletotrichum spp. *Plants*;9. Epub ahead of print 2020. DOI: 10.3390/plants9050641.

15. **Ma LJ, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, *et al.*** Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature 2010 464:7287* 2010;464:367–373.

16. **Liu S, Lin G, Ramachandran SR, Daza LC, Cruppe G, *et al.*** Rapid mini-chromosome divergence among fungal isolates causing wheat blast outbreaks in Bangladesh and Zambia. *New Phytologist*;n/a. Epub ahead of print 20 November 2023. DOI: 10.1111/nph.19402.

17. **Plaumann P-L, Schmidpeter J, Dahl M, Taher L, Koch C**. A Dispensable Chromosome Is Required for Virulence in the Hemibiotrophic Plant Pathogen Colletotrichum higginsianum. *Frontiers in microbiology* 2018;9:1005.

18. **O'Connell R, Herbert C, Sreenivasaprasad S, Khatib M, Esquerré-Tugayé M-T, *et al.*** A novel Arabidopsis-Colletotrichum pathosystem for the molecular dissection of plant-fungal interactions. *Molecular plant-microbe interactions : MPMI* 2004;17:272–82.

19. **Stiehler F, Steinborn M, Scholz S, Dey D, Weber APM, *et al.*** Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics* 2021;36:5291–5298.

20. **Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, *et al.*** Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome research* 2017;27:722–736.

21. **Seppey M, Manni M, Zdobnov EM**. BUSCO: Assessing genome assembly and annotation completeness. In: *Methods in Molecular Biology*. Humana Press Inc.; 2019. pp. 227–245.

22. **Soorni A, Haak D, Zaitlin D, Bombarely A**. Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC genomics* 2017;18:49.

23. **Bolger AM, Lohse M, Usadel B**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 2014;30:2114–20.

24. **Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, *et al.*** STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.

25. **Flutre T, Duprat E, Feuillet C, Quesneville H**. Considering transposable element diversification in de novo annotation approaches. *PloS one* 2011;6:e16526.

26. **Amselem J, Lebrun M-H, Quesneville H**. Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC genomics* 2015;16:141.

27. **Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, *et al.*** PASTEC: An Automatic Transposable Element Classification Tool. *PLoS ONE* 2014;9:e91929.

28. **Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, *et al.*** A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 2007;8:973–982.

29. **Sallet E, Gouzy J, Schiex T**. EuGene: An Automated Integrative Gene Finder for Eukaryotes and Prokaryotes. Humana, New York, NY; 2019. pp. 97–120.

30. **Min B, Grigoriev I V, Choi I-G**. FunGAP: Fungal Genome Annotation Pipeline using evidence-based gene model evaluation. *Bioinformatics* 2017;33:2936–2937.

31. **Eilbeck K, Moore B, Holt C, Yandell M**. Quantitative measures for the management and comparison of annotated genomes. *BMC bioinformatics* 2009;10:67.

32. **Valach M, Burger G, Gray MW, Lang BF**. Widespread occurrence of organelle genome-encoded 5S rRNAs including permuted molecules. *Nucleic acids research* 2014;42:13764–77.

33. **Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, *et al.*** MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 2013;69:313–319.

34. **Drillon G, Carbone A, Fischer G**. SynChro: A Fast and Easy Tool to Reconstruct and Visualize Synteny Blocks along Eukaryotic Chromosomes. *PLoS ONE* 2014;9:e92621.

35. **Gilchrist CLM, Chooi YH**. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics (Oxford, England)* 2021;37:2473–2475.

36. **Jones P, Binns D, Chang H-Y, Fraser M, Li W, *et al.*** InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;30:1236.

37. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, *et al.*** BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.

38. **Gene Ontology Consortium**. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* 2004;32:258D – 261.

39. **Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, *et al.*** High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 2008;36:3420.

40. **Zhang H, Yohe T, Huang L, Entwistle S, Wu P, *et al.*** dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 2018;46:W95–W101.

41. **Petersen TN, Brunak S, von Heijne G, Nielsen H**. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 2011;8:785–786.

42. **Emanuelsson O, Nielsen H, Brunak S, von Heijne G**. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *Journal of Molecular Biology* 2000;300:1005–1016.

43. **Krogh A, Larsson B, von Heijne G, Sonnhammer ELL**. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* 2001;305:567–580.

44. **Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM**. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular plant pathology* 2018;19:2094–2110.

45. **Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, *et al.*** antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* 2019;47:W81–W87.

46. **Zhou P, Silverstein KAT, Ramaraj T, Guhlin J, Denny R, *et al.*** Exploring structural variation and gene family architecture with De Novo assemblies of 15 Medicago genomes. *BMC Genomics* 2017;18:261.

47. **Moll KM, Zhou P, Ramaraj T, Fajardo D, Devitt NP, *et al.*** Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, Medicago truncatula. *BMC Genomics* 2017;18:578.

48. **O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, *et al.*** Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses. *Nature Genetics* 2012;44:1060–1065.

49. **Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, *et al.*** The Genome of Nectria haematococca: Contribution of Supernumerary Chromosomes to Gene Expansion. *PLOS Genetics* 2009;5:e1000618.

50. **Hu J, Chen C, Peever T, Dang H, Lawrence C, *et al.*** Genomic characterization of the conditionally dispensable chromosome in Alternaria arborescens provides evidence for horizontal gene transfer. *BMC genomics* 2012;13:1–13.

51. **Vollger MR, Dishuck PC, Sorensen M, Welch AE, Dang V, *et al.*** Long-read sequence and assembly of segmental duplications. *Nature Methods* 2019;16:88–94.

52. **Liu F, Ma ZY, Hou LW, Diao YZ, Wu WP, *et al.*** Updating species diversity of Colletotrichum, with a phylogenomic overview. *Studies in Mycology*;101.

668 53. **Rogério F, Boufleur TR, Ciampi-Guillardi M, Sukno SA, Thon MR, *et al.*** Genome Sequence
669 Resources of Colletotrichum truncatum, C. plurivorum, C. musicola, and C. sojae: Four Species
670 Pathogenic to Soybean (Glycine max). *Phytopathology* 2020;110:1497–1499.

671 54. **Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Win J, *et al.*** Genomic rearrangements
672 generate hypervariable mini-chromosomes in host-specific lineages of the blast fungus. *bioRxiv*
673 2020;2020.01.10.901983.

674 55. **Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Malmgren A, *et al.*** Genomic
675 rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast
676 fungus. *PLOS Genetics* 2021;17:e1009386.

677 56. **Vanheule A, Audenaert K, Warris S, van de Geest H, Schijlen E, *et al.*** Living apart together:
678 Crosstalk between the core and supernumerary genomes in a fungal plant pathogen. *BMC*
679 *Genomics* 2016;17:1–18.

680 57. **Condon BJ, Leng Y, Wu D, Bushley KE, Ohm RA, *et al.*** Comparative Genome Structure,
681 Secondary Metabolite, and Effector Coding Capacity across Cochliobolus Pathogens. *PLoS*
682 *Genetics* 2013;9:e1003233.

683 58. **Yang G, Turgeon BG, Yoder OC, Bronson CR, Yoder OC, *et al.*** Toxin-deficient mutants from a
684 toxin-sensitive transformant of Cochliobolus heterostrophus. *Genetics* 1994;137:751–7.

685 59. **Klosterman SJ, Subbarao K V., Kang S, Veronese P, Gold SE, *et al.*** Comparative Genomics Yields
686 Insights into Niche Adaptation of Plant Vascular Wilt Pathogens. *PLOS Pathogens*
687 2011;7:e1002137.

688 60. **Croll D, Zala M, McDonald BA**. Breakage-fusion-bridge Cycles and Large Insertions Contribute to
689 the Rapid Evolution of Accessory Chromosomes in a Fungal Pathogen. *PLOS Genetics*
690 2013;9:e1003567.

691 61. **Bertazzoni S, Williams AH, Jones DA, Syme RA, Tan K-C, *et al.*** Accessories Make the Outfit:
692 Accessory Chromosomes and Other Dispensable DNA Regions in Plant-Pathogenic Fungi.
693 *Molecular Plant-Microbe Interactions* 2018;31:779–788.

694 62. **Francis A, Ghosh S, Tyagi K, Prakasam V, Rani M, *et al.*** Evolution of pathogenicity-associated
695 genes in Rhizoctonia solani AG1-IA by genome duplication and transposon-mediated gene
696 function alterations. *BMC Biology* 2023;21:1–19.

697 63. **Fraser JA, Huang JC, Pukkila-Worley R, Alspaugh JA, Mitchell TG, *et al.*** Chromosomal
698 translocation and segmental duplication in Cryptococcus neoformans. *Eukaryotic Cell*
699 2005;4:401–406.

700 64. **van Westerhoven A, Aguilera-Galvez C, Nakasato-Tagami G, Shi-Kunne X, Martinez E, *et al.***
701 Segmental Duplications Drive the Evolution of Accessory Regions in a Major Crop Pathogen.
702 *bioRxiv* 2023;2023.06.07.544053.

703 65. **Wilson AM, Lelwala R V., Taylor PWJ, Wingfield MJ, Wingfield BD**. Unique patterns of mating
704 pheromone presence and absence could result in the ambiguous sexual behaviors of
705 Colletotrichum species. *G3 Genes|Genomes|Genetics*;11. Epub ahead of print 6 September
706 2021. DOI: 10.1093/G3JOURNAL/JKAB187.

66. **Bhunjun CS, Phukhamsakda C, Jeewon R, Promputtha I, Hyde KD**. Integrating Different Lines of Evidence to Establish a Novel Ascomycete Genus and Family (Anastomitrabeculia, Anastomitrabeculiaceae) in Pleosporales. *Journal of Fungi 2021, Vol 7, Page 94* 2021;7:94.

67. **Manners JM, He C**. Slow-growing heterokaryons as potential intermediates in supernumerary chromosome transfer between biotypes of Colletotrichum gloeosporioides. *Mycological Progress* 2011;10:383–388.

68. **He C, Rusu AG, Poplawski AM, Irwin JAG, Manners JM**. Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus Colletotrichum gloeosporioides. *Genetics* 1998;150:1459–1466.

69. **Roca MG, Davide LC, Davide LMC, Mendes-Costa MC, Schwan RF, *et al.*** Conidial anastomosis fusion between Colletotrichum species. *Mycological research* 2004;108:1320–1326.

70. **Ishikawa FH, Souza EA, Shoji JY, Connolly L, Freitag M, *et al.*** Heterokaryon incompatibility is suppressed following conidial anastomosis tube fusion in a fungal plant pathogen. *PloS one*;7. Epub ahead of print 2 February 2012. DOI: 10.1371/JOURNAL.PONE.0031175.

71. **Mehta N, Baghela A**. Quorum sensing-mediated inter-specific conidial anastomosis tube fusion between Colletotrichum gloeosporioides and C. siamense. *IMA fungus*;12. Epub ahead of print 1 December 2021. DOI: 10.1186/S43008-021-00058-Y.

72. **Schmidt SM, Houterman PM, Schreiver I, Ma L, Amyotte S, *et al.*** MITEs in the promoters of effector genes allow prediction of novel virulence genes in Fusarium oxysporum. *BMC genomics* 2013;14:1–21.

73. **van der Does HC, Fokkens L, Yang A, Schmidt SM, Langereis L, *et al.*** Transcription Factors Encoded on Core and Accessory Chromosomes of Fusarium oxysporum Induce Expression of Effector Genes. *PLOS Genetics* 2016;12:e1006401.

74. **Mould MJR, Boland GJ, Robb J**. Ultrastructure of the Colletotrichum trifolii-Medicago sativa pathosystem. I. Pre-penetration events. *Physiological and Molecular Plant Pathology* 1991;38:179–194.

75. **Damm U, Cannon PF, Liu F, Barreto RW, Guatimosim E, *et al.*** The Colletotrichum orbiculare species complex: Important pathogens of field crops and weeds. *Fungal Diversity* 2013;61:29–59.

76. **C A-T, BB W, MS O, S D, H Z, *et al.*** Identification and Characterization of Nucleotide-Binding Site-Leucine-Rich Repeat Genes in the Model Plant Medicago Truncatula. *Plant physiology*;146. Epub ahead of print 2008. DOI: 10.1104/PP.107.104588.

Table 1: Characteristics of *C. destructivum* core and mini chromosomes.

	***C. destructivum* chromosomes**			
	1-10 (except 1B)	**1B region**	**11**	**12**
Total length	48 456 982 bp	1 205 667 bp	1 275 594 bp	812 569 bp
G+C content	54.7 %	52.3 %	48.7 %	50.2 %
Number of protein-coding genes	14882	300	278	171
Proportion of genes by length	61.7 %	30.9 % ***	32.3 % ***	26.8 % ***
Proportion of genes with unknown function	21.3 %	42.0 % ***	28.4 % *	32.2 % *
Proportion of genes with RNA support	77.0 %	52.0 % ***	46.4 % ***	59.6 %*
Proportion of CAZyme genes	4.1 %	0.0 % ***	1.4 %	1.2 %
Proportion of effector genes	3.0 %	4.3 %	5.4 % *	5.8 % *
Proportion of SMKG	0.7 %	0.0 %	2.9 % **	0.0 %
Proportion of TE by length	4.4 %	32.8 % ***	32.3 % ***	35.1 % ***

Asterisks indicate that the data for chromosomes 1B, 11 or 12 differ significantly from the core chromosomes (Fisher's exact test, *** $P<0.001$; ** $P<0.01$; * $P<0.05$)

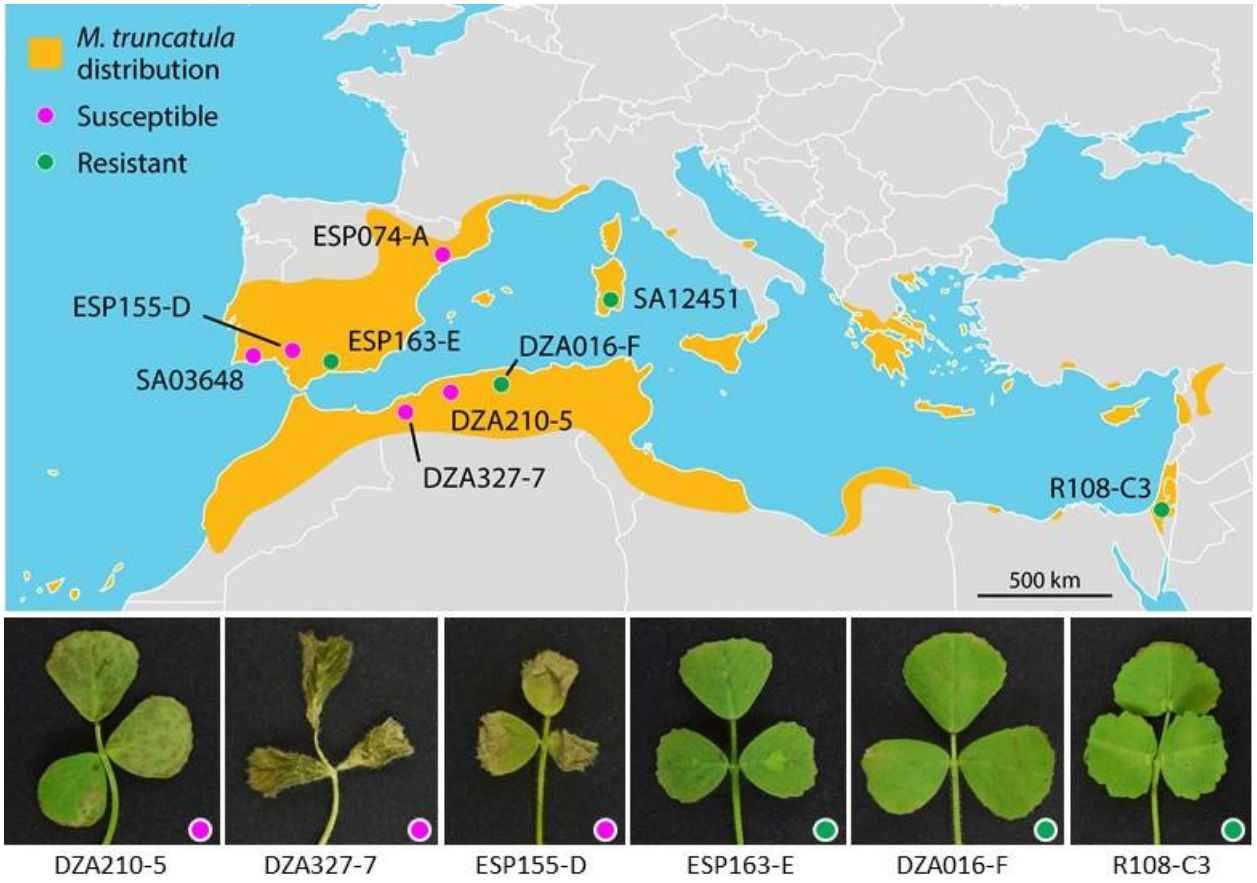


Figure 1: *Medicago truncatula* accessions used in this study and their infection phenotypes with *C. destructivum* LARS 709. **Upper panel:** Geographical distribution of *M. truncatula* in the Mediterranean area according to GBIF (2019) and collection locations of the nine ecotypes used in this study. **Lower panel:** Symptoms produced on the trifoliate leaves of six *M. truncatula* accessions at 4 days post inoculation with spore suspension of *C. destructivum* LARS 709. Leaves of the susceptible accession DZA210-5 showed large necrotic lesions, while those of DZA327-7 and ESP155-D were completely necrotic. Leaves of the resistant accessions ESP163-E, DZA016-F and R108-C3 showed small necrotic flecks or no visible symptoms. Note that R108-C3 is considered to be *Medicago truncatula* ssp. *tricycla*.

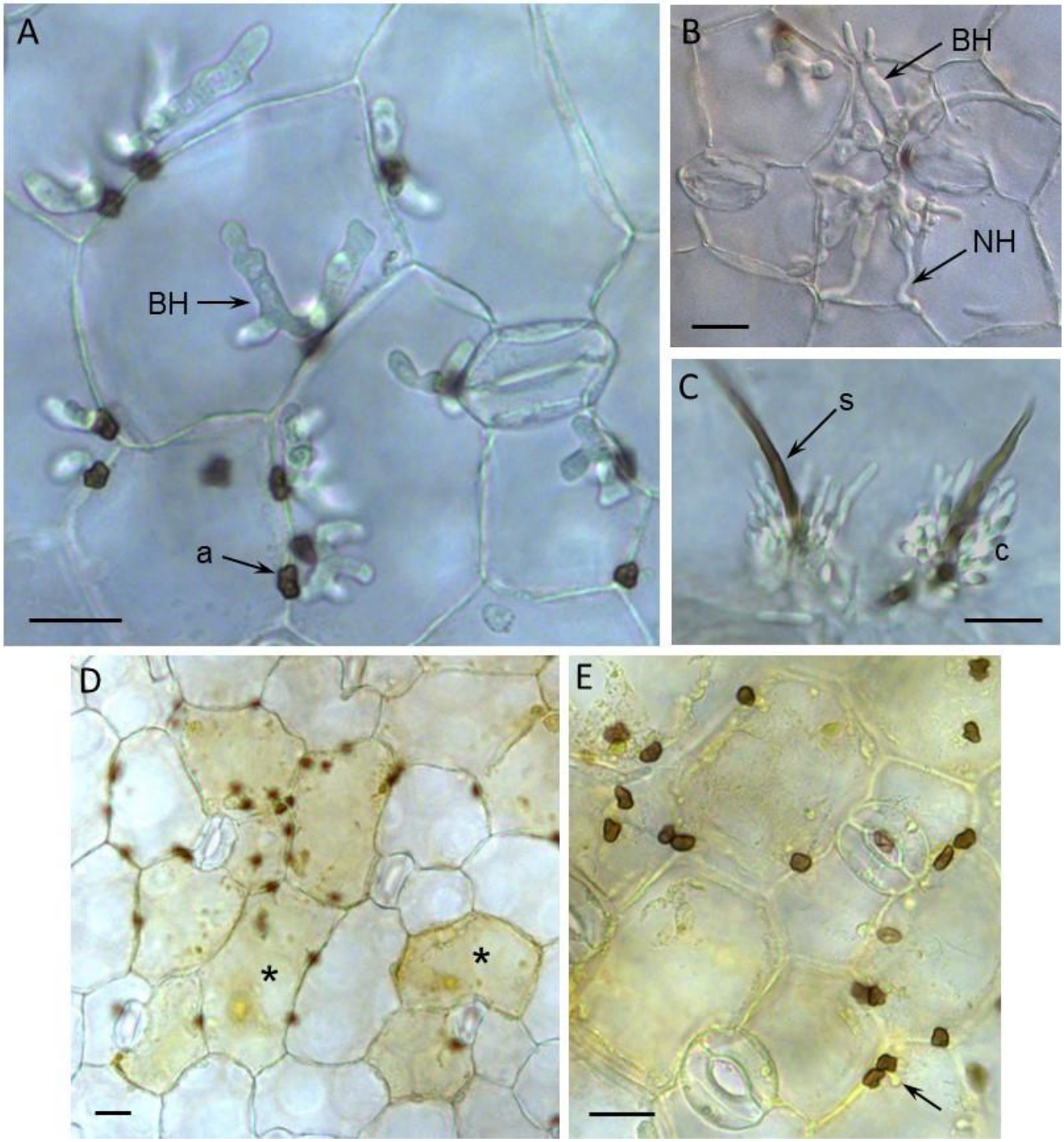


Figure 2: Microscopic analysis of *C. destructivum* LARS 709 infecting cotyledon tissues of *M. truncatula*. (A-C) Susceptible accession ESP155-D. At 48 hpi **(A)**, melanized appressoria (a) had formed on the plant surface and penetrated epidermal cells to form bulbous biotrophic hyphae (BH). At 60 hpi **(B)**, thin necrotrophic hyphae (NH) developed from the tips of biotrophic hyphae. At 72 hpi **(C)**, acervuli erupted from the plant surface, consisting of a melanized, hair-like seta (s) and a mass of conidia (c). **(D,E)** Resistant accession ESP163-E. At 72 hpi, few appressoria had penetrated cotyledon epidermal cells, and groups of cells underlying the appressoria were pigmented yellowish brown with granular contents. Any hyphae visible inside epidermal cells were typically smaller than the appressorium. Scale bars = 20 µm.

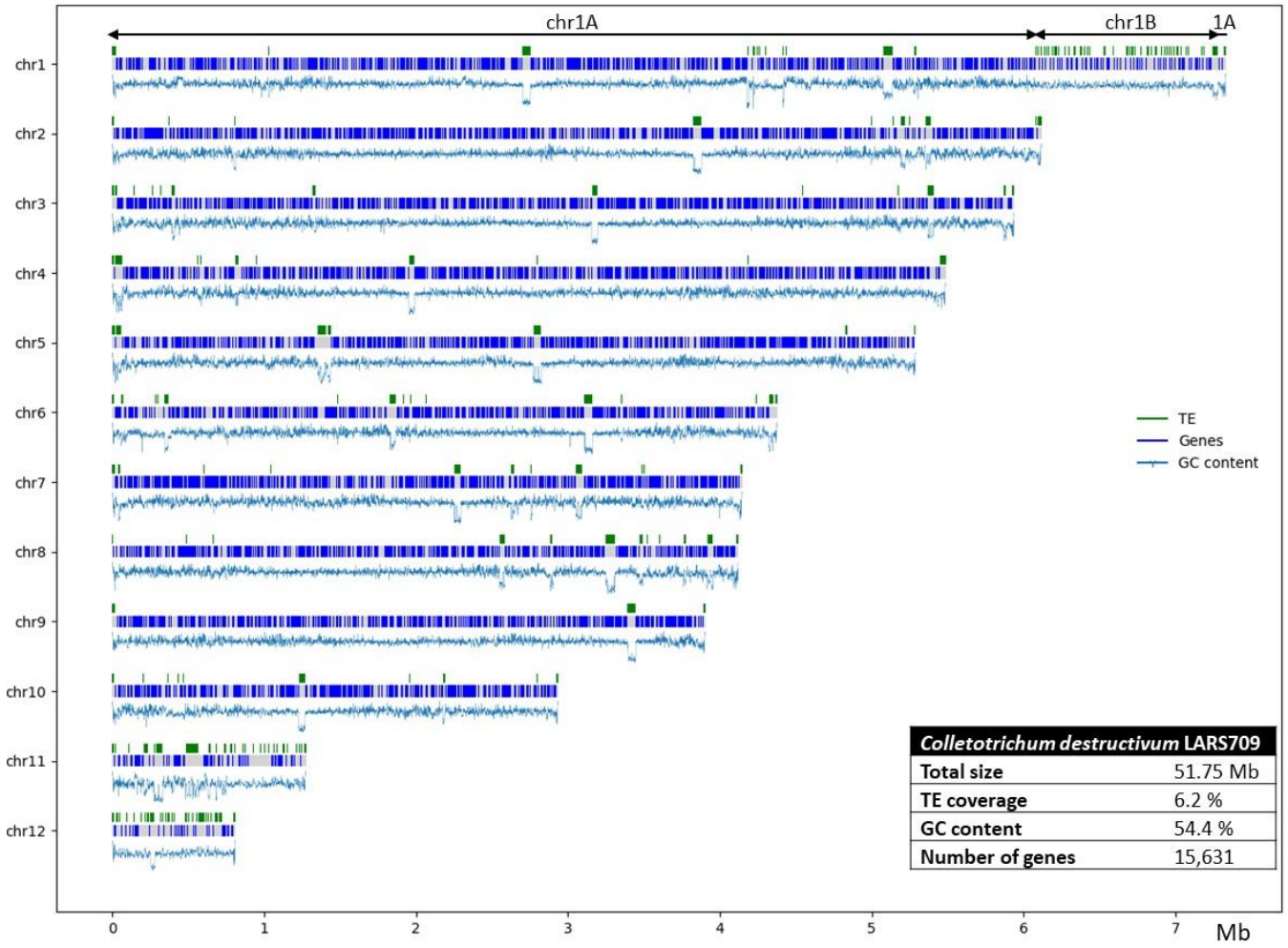


Figure 3: Schematic representation of the 12 chromosomes of *C. destructivum* isolate 709. The distribution of genes and transposable elements (TE) across each chromosome are shown together with the corresponding genome statistics (inset table).

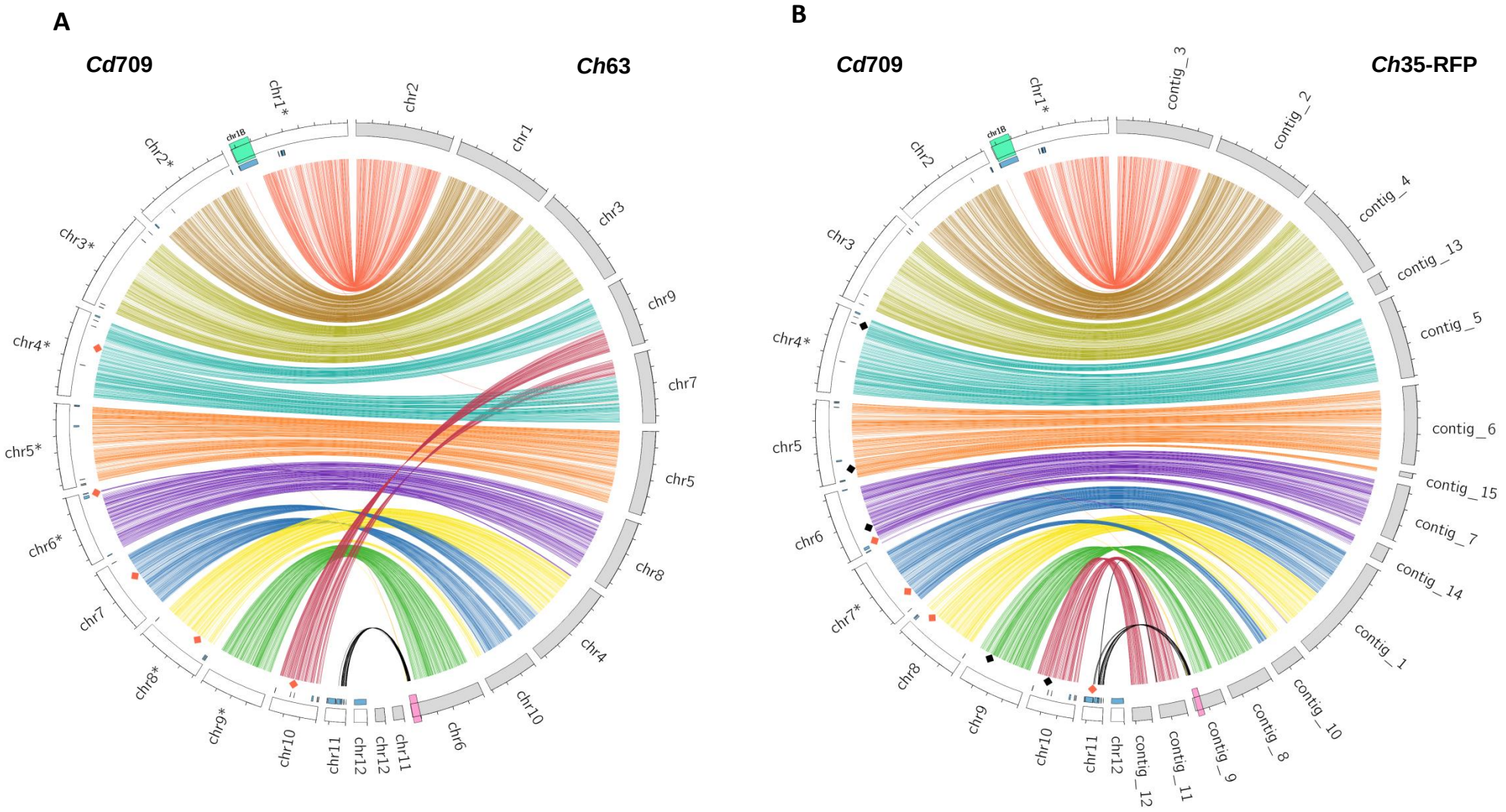


Figure 4: **Whole-genome alignments between *C. destructivum* LARS 709 (*Cd*709) and two *C. higginsianum* strains.** Chromosomes of *Cd*709 (white bars) were aligned with **(A)** the chromosomes of *C. higginsianum* IMI 349063 (*Ch*63, grey bars) or **(B)** the contigs of *C. higginsianum* MAFF 304535-RFP (*Ch*35-RFP, grey bars). Syntenic regions (length >10 kb and percent identity > 88%) were linked together using coloured arcs specific for each chromosome in the *Cd*709 genome assembly. Red diamonds indicate interchromosomal rearrangements. Black diamonds indicate chromosome breakpoints associated with separate contigs in the *Ch*35-RFP assembly only. The blue track indicates gene blocks that are unique to *Cd*709. Note that region chr1B of *Cd*709 (highlighted in green) has no alignments in either of the *C. higginsianum* isolates. The black arcs linking chr11 of *Cd*709 to the 3' end of chr6/contig_9 in *C. higginsianum* (highlighted in pink) indicate regions with strong sequence similarity that are smaller than 10 Kb. Asterisks indicate where chromosome sequences were reverse-complemented for better visualization. Tick mark spacing = 1 Mb

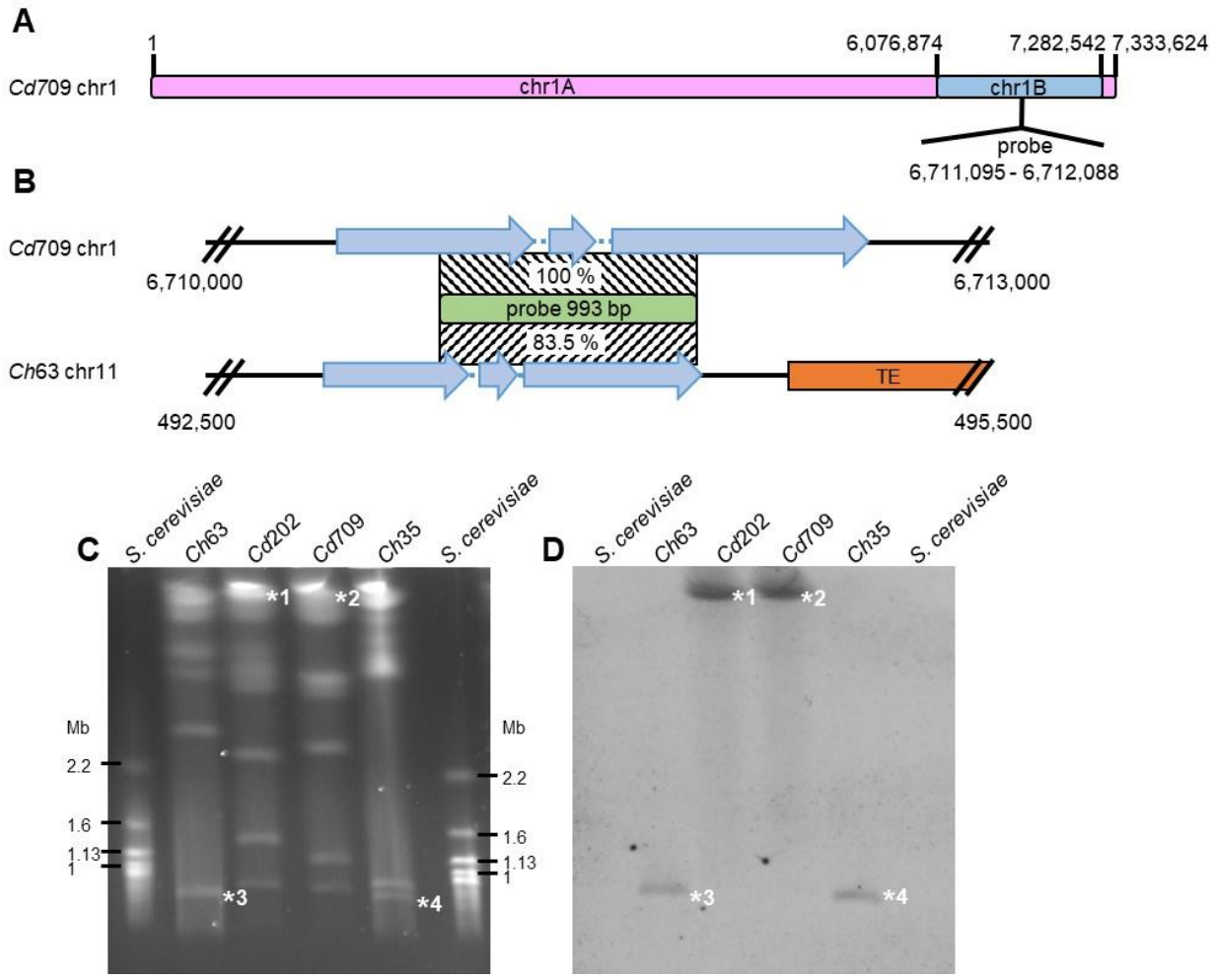


Figure 5: Chromosome 1 of *Colletotrichum destructivum* has a bipartite structure. (A) Scheme of the structure of *Cd*709 chromosome 1. The probe is specific to the mini-chromosome-like part of the chromosome (chr1B). **(B)** Detailed scheme of the regions targeted by the 993 bp DIG-labelled probe in *Cd*709 and in *Ch*63 (chr11: 493,380 to 494,373). Patterned boxes indicate sequence identity of the target regions to the probe. **(C)** Pulsed-field gel electrophoresis of chromosomal DNA from *C. destructivum* isolates LARS 202 (*Cd*202) and LARS 709 (*Cd*709) compared to *C. higginsianum* isolates IMI349063 (*Ch*63) and MAFF305635 (*Ch*35). **(D)** Southern hybridisation. Numerals 1 to 4 indicate signals corresponding to chromosomes displayed in **(C)**.

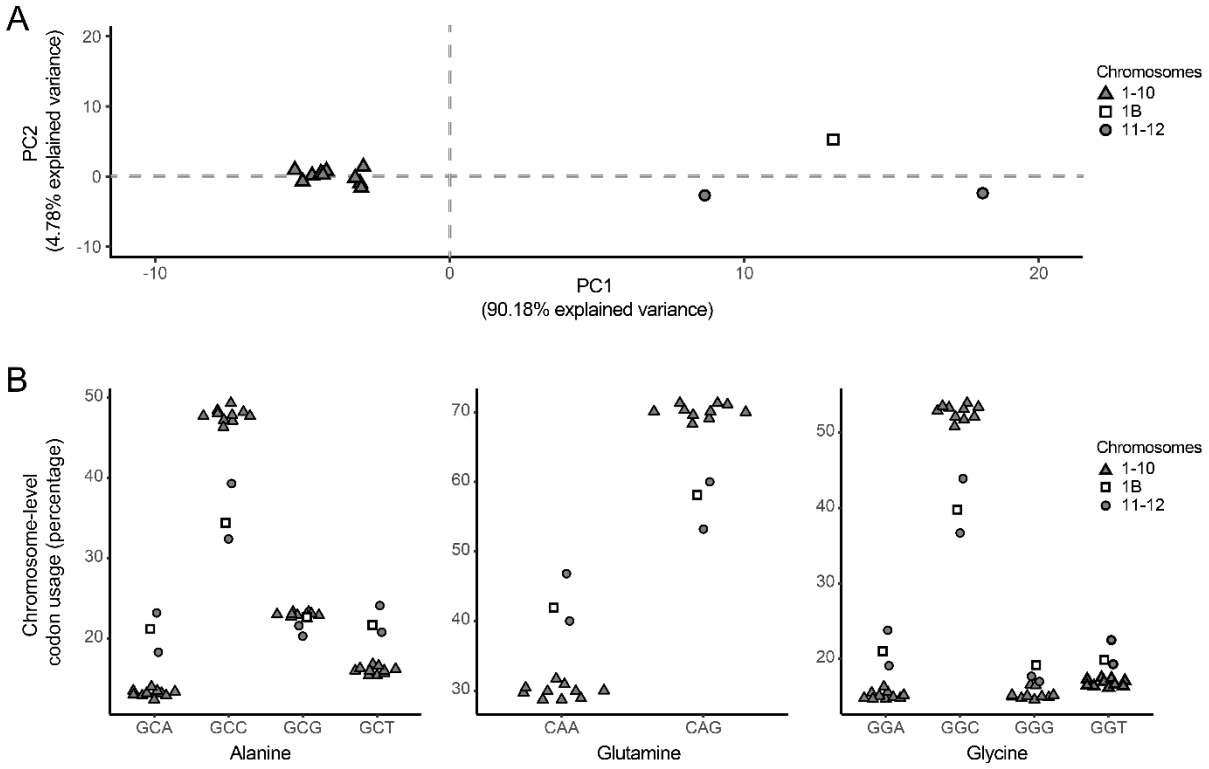


Figure 6: Codon usage bias in the core and mini-chromosomes of *C. destructivum*. (A) Principal component analysis (PCA) of codon usage for all amino acids on each chromosome. The region chr1B was considered separately from the rest of chr1. The first two axes accounted for 95% of the variance. **(B)** Plots showing codon usage bias for three amino acids (Alanine, Glutamine, Glycine) in genes located on core chromosomes (1 to 10 excluding region 1B), mini-chromosomes 11 and 12 and region 1B. Codon usage on chr11, chr12 and region chr1B differed significantly from that on core chromosomes (Fisher's exact test, $P < 0.001$) for the 10 codons presented except GCG (all comparisons) and GGG (chr12 vs core). Other amino acids are displayed in Fig. S6. The significance is reported for all the codons in the Table S12.

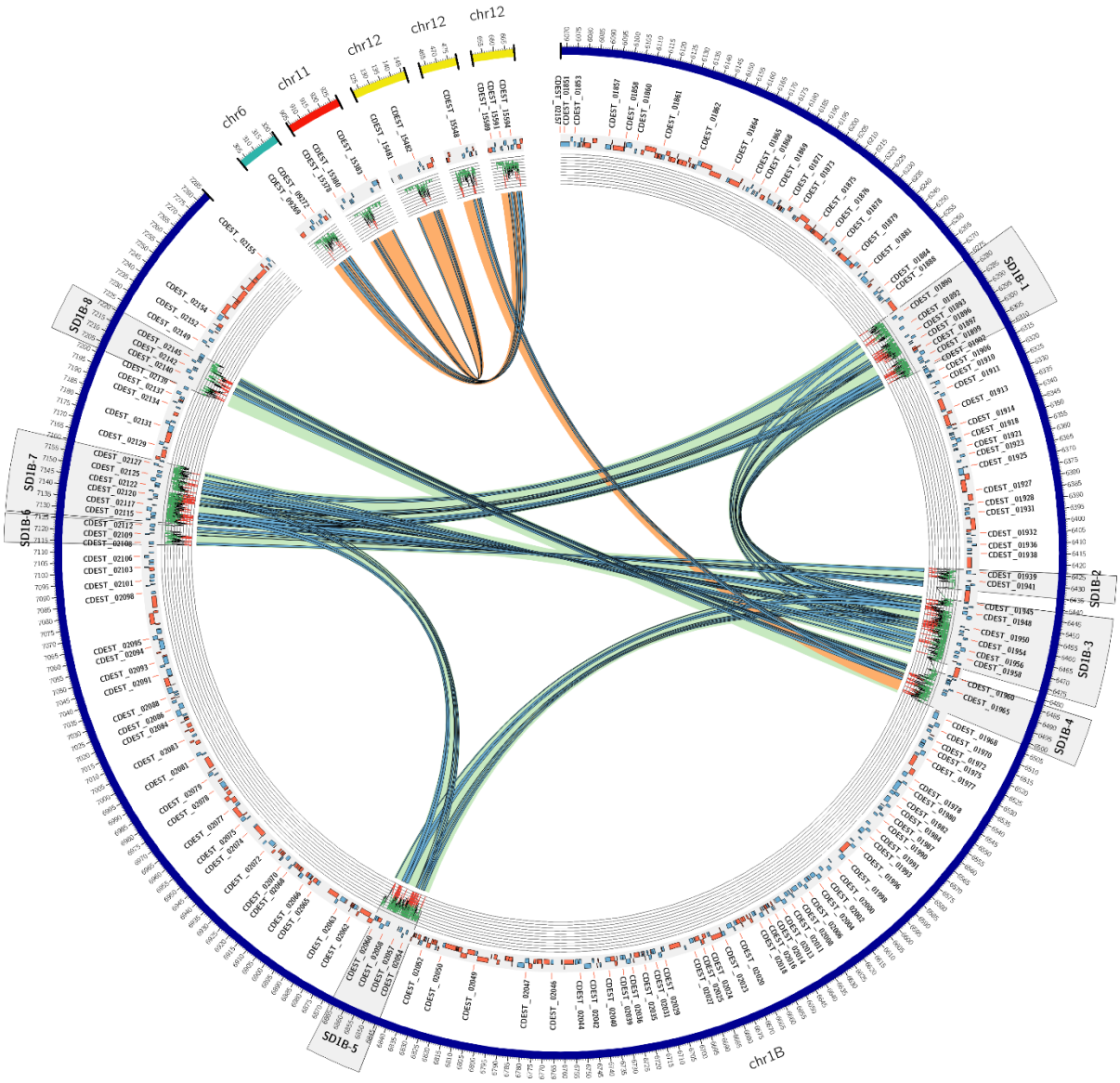


Figure 7: Circos plot showing *C. destructivum* segmental duplications larger than 10kb found with SDDetector. The green and red tracks represent genes and transposable elements, respectively. Light grey arcs show duplications and those involving genes were highlighted with blue arcs.

Table S1: List of *Medicago truncatula* accessions used in this study and their infection phenotypes with *C. destructivum* LARS 709.

Accession ID	Hapmap ID	Country of origin	Infection phenotype
DZA016-F	HM048	Algeria	Resistant
DZA210-5	HM083	Algeria	Susceptible
DZA327-7	HM011	Algeria	Susceptible
ESP074-A	HM056	Spain	Susceptible
ESP155-D	HM057	Spain	Susceptible
ESP163-E	HM058	Spain	Resistant
R108-C3*	HM029	Israel	Resistant
SA03648	HM068	Portugal	Susceptible
SA12451	HM075	Italy	Resistant

* R108-C3 (R108) was derived by *in vitro* selection from the natural accession 108-1 (Hoffmann et al. 1997) and is now considered to be *Medicago truncatula* ssp. *tricycla*.

Table S2: *Colletotrichum* mitochondrial genomes used as reference for assembly of the *C. destructivum* mitochondrial genome with Organelle_PBA

Species/Strain	Size (bp)	Accession number
C. acutatum KC05	30892	NC_027280.1/KR349346.1
C. fioriniae	30020	NC_030052.1/KU375885.1
C. graminicola M1.001	39649	MT: NW_007361658.1/CM001021.1
C. lindemuthianum	36957	NC_023540.1/KF953885.1
C. lupini CBS 119142	36554	NC_029213.1/KT918406.1
C. salicis	33950	NC_035496.1/KY774449.1
C. tamarilloi	30824	NC_029706.1/KU196965.1
C. lindemuthianum isolate 89 A2 2-3	37446	MF595869.1
C. lindemuthianum isolate 83.501	37440	MF595868.1
C. aenigma XY15	57252	KX885105.1
C. gloeosporioides LQ33	55169	KX885104.1
C. siamense YT02	53317	KX885103.1
C. siamense SQ01	54645	KX885102.1
C. siamense LQ22	58666	KX885101.1
C. siamense ZH01	52671	KX885100.1
C. siamense ZH03	54658	KX885099.1
C. siamense ZH02	54679	KX885098.1
C. fructicola	56051	KX034082.1

Table S3: Classification of transposable element (TE) consensus sequences identified in the *C. destructivum* genome. Sequences were classified according to [28]. The number and mean length of the different consensus sequences (i.e. non-redundant sequences) per TE superfamily are indicated. The total number of copies in the genome, as well as the number of complete copies, and the corresponding genome coverages were computed. LTR: long terminal repeat, LINE: long interspersed element, TIR: terminal inverted repeat, MITE: miniature inverted-repeat transposable element.

Order	Superfamily	Wicker Code	Number of consensus	Length consensus in bases	Number of copies	Coverage in Kb (%)	Number of complete copies	Coverage by complete copies in Kb (%)
Class I (retrotransposons)								
LTR	Copia	RLC	4	5970 (+/- 1864)	153	301 (0.58%)	39	245 (0.47%)
	Gypsy	RLG	7	10762 (+/- 5480)	719	1489 (2.88%)	106	699 (1.35%)
LINE	I	RII	6	5968 (+/- 881)	289	508 (0.98%)	61	350 (0.68%)
	Other	RIX	1	2598	12	7 (0.01%)	1	3 (0.01%)
Class II (DNA transposons)								
TIR	Tc1-Mariner	DTT	18	1877 (+/-15)	401	355 (0.69%)	136	255 (0.49%)
	hAT	DTA	3	2545 (+/- 921)	67	109 (0.21%)	28	83 (0.16%)
	PiggyBac	DTB	2	2246 (+/- 50)	60	52 (0.10%)	18	41 (0.08%)
	PIF-Harbinger	DTH	1	3061	30	22 (0.04%)	5	15 (0.03%)
	Other	DTX	1	721	7	3 (0.01%)	4	3 (0.01%)
Helitron		DHX	1	11678	56	280 (0.54%)	22	257 (0.50%)
MITE		DXX-MITE	2	614 (+/- 302)	26	7 (0.01%)	5	4 (0.01%)
Uncharacterized TEs			3	858 (+/- 342)	95	44 (0.08%)	32	29 (0.06%)
Total			***49***		***1915***	***3177 (6.14%)***	***457***	***1983 (3.83%)***

Table S4: Parameters and metrics of RNA-Seq transcriptome assemblies.

	PDB	48 hpi	72 hpi
Uniquely mapped reads	93%	2.5%	6.5%
Mean coverage depth (no. of reads)	82	16	33
StringTie no. of reads junction (-j)	10	3	5
TPM threshold	1.88	9.38	4.90
No. of transcripts	16122	13901	15081
No. of genes	15209	13496	14338
Mean transcript length	1469	1169	1482
Standard deviation	1255	952	1148
Median	1134	883	1202
Minimum	150	150	150
Maximum	23615	13099	13082
Mean read coverage	129	21	47
Standard deviation	763	83	258
Median	16	6	12
Minimum	4	3	4
Maximum	38440	3168	24873
Mean TPM	54	63	58
Standard deviation	319	243	317
Median	6	18	15
Minimum	1.88	9.38	4.90
Maximum	16091	9184	30486

RNA-seq libraries were prepared from total RNA isolated from the following samples:

PDB = mycelium grown in potato dextrose broth

48 hpi = infected *Medicago sativa* cotyledons at 48 h post inoculation

72 hpi = infected *Medicago sativa* cotyledons at 72 h post inoculation

TPM = transcripts per million

Table S5: Genome annotation statistics for *C. destructivum* isolate LARS 709.

Annotation statistics	
Coverage by Transposable Elements	6.2%
Number of predicted genes	15 631
Number of genes with RNA-Seq support	11 853
Number of genes with protein support	15 172
Average gene length (bp)	1976
Average exon length (bp)	649
Average number of exons per gene	2.79
Average intron length (bp)	88
Average number of introns per gene	1.79
Number of mono-exon genes	4055
Average CDS length (bp)	1378
Annotation completeness (BUSCO)*	
Complete proteins	1309 (99.54%)
Fragmented proteins	5 (0.38%)
Missing proteins	1 (0.08%)
Number of predicted CAZymes	619
Number of predicted effector proteins	484
Number of predicted secondary metabolism key genes (SMKGs)	110

* BUSCO = Benchmarking Universal Single-Copy Orthologs

Table S7: Summary of the results of predicting syntenic blocks in *C. destructivum* and *C. higginsianum* using SynChro (delta parameter = 1). The number of syntenic blocks generated from Reciprocal Best-Hits and numbers of associated genes are shown.

	Fungal Genome	
	C. destructivum	*C. higginsianum*
No. of Reciprocal Best-Hits	12135	
Similarity (%)	93.9	
No. of syntenic blocks (SB)	400	
No. of genes in syntenic blocks	14311	13921
Proportion of genome in SB (%)	88.0	91.6

Table S8: List of non-syntenic blocks identified using Synchro with at least 5 consecutive genes

Chromosome	Start	End	Length (bp)	No. of genes	Block ID
chr1	4178820	4203721	24901	5	block_0
chr1	4210554	4340246	129692	32	block_1
chr1	4348070	4378433	30363	12	block_2
chr1	4397823	4426916	29093	9	block_3
chr1	4568340	4599600	31260	13	block_4
chr1	6066423	7282629	1216206	305	block_5
chr1	7307926	7321922	13996	6	block_6
chr2	12042	52349	40307	11	block_7
chr2	4605099	4613936	8837	5	block_8
chr2	5997899	6047905	50006	15	block_9
chr2	6062466	6093962	31496	8	block_10
chr3	46529	55509	8980	6	block_11
chr3	723104	737102	13998	6	block_12
chr3	5641079	5662083	21004	12	block_13
chr3	5871897	5895410	23513	7	block_14
chr4	13059	104775	91716	15	block_15
chr4	352058	363077	11019	7	block_16
chr4	752022	769918	17896	8	block_17
chr4	3379899	3389434	9535	5	block_18
chr5	21675	79289	57614	9	block_19
chr5	89463	100537	11074	5	block_20
chr5	1346906	1442665	95759	11	block_21
chr5	4889101	4902562	13461	5	block_22
chr5	5176186	5186138	9952	5	block_23
chr5	5238579	5277212	38633	17	block_24
chr6	13368	75899	62531	18	block_25
chr6	200524	370735	170211	46	block_26
chr6	4320111	4364960	44849	5	block_27
chr7	4117709	4131681	13972	6	block_28
chr8	7311	37105	29794	11	block_29
chr8	3909912	3964560	54648	6	block_30
chr8	4012267	4028388	16121	7	block_31
chr8	4074802	4104606	29804	11	block_32
chr10	11605	31856	20251	6	block_33
chr10	92306	124025	31719	15	block_34
chr10	373355	477333	103978	34	block_35
chr10	1628812	1639928	11116	5	block_36
chr10	1876452	1892426	15974	7	block_37
chr10	2898367	2924158	25791	14	block_38
chr11	12418	20573	8155	6	block_39
chr11	139951	178181	38230	9	block_40
chr11	271755	357039	85284	9	block_41
chr11	378813	604554	225741	38	block_42
chr11	663865	691604	27739	7	block_43
chr11	697412	1175606	478194	106	block_44
chr11	1223231	1263891	40660	7	block_45
chr12	15082	800244	785162	171	block_46

Table S9: Gene ontology enrichment tables of *C. destructivum*-specific genes in blocks that are non-syntenic with *Ch*63 detected with SynChro. No enrichments in Cellular Component (CC) were detected with the topGO R library.

a) Molecular Function

GO.ID	Term	Annotated	Significant	Expected	p-value
GO:0004672	protein kinase activity	187	28	9.51	1.90E-07
GO:0016773	phosphotransferase activity	230	28	11.7	1.30E-05
GO:0046914	transition metal ion binding	992	78	50.45	2.60E-05
GO:0005506	iron ion binding	297	31	15.11	9.10E-05
GO:0016301	kinase activity	257	28	13.07	9.60E-05
GO:0031177	phosphopantetheine binding	57	11	2.9	0.00011
GO:0072341	modified amino acid binding	57	11	2.9	0.00011
GO:0033218	amide binding	65	11	3.31	0.00037
GO:0016772	transferase activity	331	31	16.84	0.00063
GO:0005488	binding	3713	216	188.85	0.00068
GO:0020037	heme binding	301	28	15.31	0.00129
GO:0046906	tetrapyrrole binding	301	28	15.31	0.00129
GO:0043167	ion binding	2217	138	112.76	0.00155
GO:0003700	DNA-binding transcription factor activity	55	9	2.8	0.00163
GO:0046872	metal ion binding	1166	80	59.3	0.00189
GO:0043169	cation binding	1173	80	59.66	0.00225
GO:0140110	transcription regulator activity	77	10	3.92	0.00529
GO:0004834	tryptophan synthase activity	3	2	0.15	0.00748
GO:0008270	zinc ion binding	635	46	32.3	0.00784
GO:0140096	catalytic activity, acting on a protein	542	40	27.57	0.00968

b) Biological Process

GO.ID	Term	Annotated	Significant	Expected	p-value
GO:0006468	protein phosphorylation	184	28	7.98	3.00E-09
GO:0016310	phosphorylation	262	28	11.37	5.90E-06
GO:0009403	toxin biosynthetic process	63	12	2.73	1.20E-05
GO:0009404	toxin metabolic process	63	12	2.73	1.20E-05
GO:0043385	mycotoxin metabolic process	63	12	2.73	1.20E-05
GO:0043386	mycotoxin biosynthetic process	63	12	2.73	1.20E-05
GO:0044550	secondary metabolite biosynthetic process	65	12	2.82	1.70E-05
GO:0019748	secondary metabolic process	69	12	2.99	3.20E-05
GO:0008152	metabolic process	3257	163	141.3	0.00019
GO:0055114	oxidation-reduction process	1053	67	45.68	0.00028
GO:0006464	cellular protein modification process	346	28	15.01	0.00083
GO:0036211	protein modification process	346	28	15.01	0.00083
GO:0043412	macromolecule modification	385	28	16.7	0.0041

Table S10: Functional enrichment test (Fisher's exact test). Complete lists of *C. destructivum* Secondary metabolism key genes (SMKGs), and genes encoding effectors and carbohydrate-active enzymes (CAZYmes) were used to detect enrichments among these functional categories in the 1083 *C. destructivum*-specific genes in non-syntenic blocks detected using SynChro.

	In genome	**In non-syntenic blocks**	**Odds ratio**	***p*-value**
Total No. of genes	15631	1083		
No. of SMKGs	123	14	1.6	8.16E-02
No. of effector genes	484	49	1.5	1.62E-02
No. of CAZyme genes	619	14	0.3	1.62E-06

Table S11: Characteristics of each assembled chromosome of *C. destructivum*.

Chr	Size (bp)	No. genes	GC (%)	TE content %)	Telomeres
1	7333624	2172	54.60	8.7	2
1A	*6127957*	*1872*	*55.10*	*4.0*	*2*
1B	*1205667*	*300*	*52.30*	*32.8*	*NA*
2	6118321	1884	54.69	3.8	2
3	5939657	1874	55.06	4.1	2
4	5487742	1666	55.09	3.9	2
5	5290227	1573	54.75	4.3	2
6	4379568	1360	54.01	5.5	2
7	4147912	1312	54.54	5.0	2
8	4124671	1252	53.98	6.2	2
9	3904609	1167	55.26	3.0	2
10	2936318	922	54.23	4.6	2
11	1275594	278	51.32	32.3	2
12	812569	171	50.22	35.1	2

TE = transposable element

Table S13: List of chromosome locations of the eight largest segmentally duplicated areas present in region chr1B of chromosome 1.

Code	chr	Start	End	Duplication	Length (bp)	Pairing
SD1B-1	chr1B	6275896	6310837	chr1:6275896-6309509 // chr1:7126431-7156805	28087	SD1B-1 // SD1B-7
				chr1:6275896-6310837 // chr1:6440881-6474186	30727	SD1B-1 // SD1B-3
SD1B-2	chr1B	6423690	6434928	chr1:6423690-6434928 // chr1:7113913-7125518	11316	SD1B-2 // SD1B-6
SD1B-3	chr1B	6440881	6476743	chr1:6275896-6310837 // chr1:6440881-6474186	30727	SD1B-3 // SD1B-1
				chr1:6459979-6476743 // chr1:7145173-7159363	14169	SD1B-3 // SD1B-7
				chr1:6442213-6459984 // chr1:7126431-7143900	16220	SD1B-3 // SD1B-7
				chr1:6442842-6464026 // chr1:6844058-6866689	19709	SD1B-3 // SD1B-5
SD1B-4	chr1B	6481996	6501507	chr1:6481996-6501507 // chr1:7201724-7220918	16885	SD1B-4 // SD1B-8
SD1B-5	chr1B	6844058	6866689	chr1:6846262-6862419 // chr1:7129446-7147001	15244	SD1B-5 // SD1B-7
				chr1:6442842-6464026 // chr1:6844058-6866689	19709	SD1B-5 // SD1B-3
SD1B-6	chr1B	7113913	7125518	chr1:6423690-6434928 // chr1:7113913-7125518	11316	SD1B-6 // SD1B-2
SD1B-7	chr1B	7126431	7159363	chr1:6275896-6309509 // chr1:7126431-7156805	28087	SD1B-7 // SD1B-1
				chr1:6846262-6862419 // chr1:7129446-7147001	15244	SD1B-7 // SD1B-5
				chr1:6442213-6459984 // chr1:7126431-7143900	16220	SD1B-7 // SD1B-3
				chr1:6459979-6476743 // chr1:7145173-7159363	14169	SD1B-7 // SD1B-3
SD1B-8	chr1B	7201724	7220918	chr1:6481996-6501507 // chr1:7201724-7220918	16885	SD1B-8 // SD1B-4

Table S14: Full Proteome similarity with other *Colletotrichum* species.

	***Cd*709 total proteome (15631)**		***Cd*709 Chr1B Proteome (300)**	
Species	Retrieved	%*	Retrieved	%*
C. chlorophyti	11098	71.0	85	28.3
C. fioriniae	12339	78.9	108	36.0
C. fructicola	11852	75.8	165	55.0
C. gloeosporioides	11666	74.6	84	28.0
C. graminicola	11798	75.5	93	31.0
C. higginsianum	14372	91.9	134	44.7
C. incanum	12700	81.2	173	57.7
C. musicola	12370	79.1	136	45.3
C. nymphaeae	12443	79.6	177	59.0
C. orbiculare	12091	77.4	92	30.7
C. orchidophilum	12381	79.2	203	67.7
C. plurivorum	12352	79.0	121	40.3
C. salicis	12171	77.9	164	54.7
C. shisoi	12644	80.9	92	30.7
C. sidae	11447	73.2	89	29.7
C. simmondsii	12361	79.1	103	34.3
C. sojae	12362	79.1	120	40.0
C. spinosum	11520	73.7	88	29.3
C. sublineola	12107	77.5	202	67.3
C. tanaceti	12183	77.9	94	31.3
C. tofieldiae	12210	78.1	142	47.3
C. trifolii	11481	73.5	84	28.0
C. truncatum	12561	80.4	217	72.3
All species	15081	96.4	278	92.6

* % of completeness of the whole set of protein compared to *Cd*709 (Blastp cut-offs 30% identity, 50% coverage).

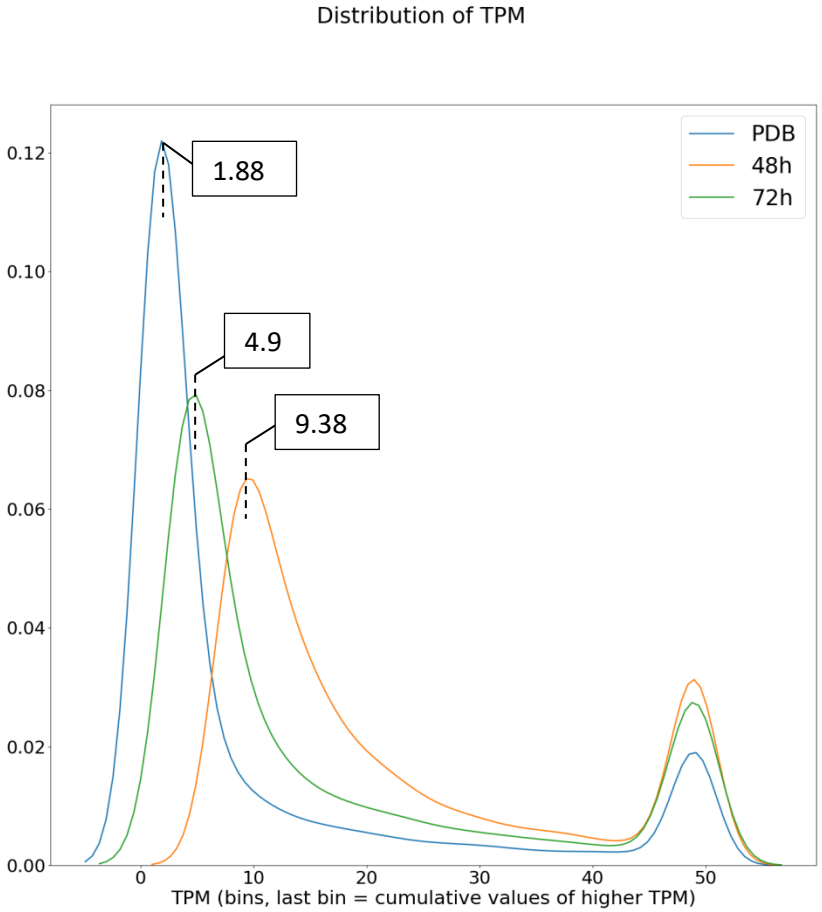


Figure S1: Distribution of TPM of the three assembled RNA-Seq libraries. Thresholds used to filter out assembled transcripts lacking sufficient coverage with RNA-Seq reads are indicated for each of the tested conditions, namely mycelia grown in potato dextrose broth (PDB) and infected *Medicago sativa* cotyledons at 48 h and 72 h post-inoculation.

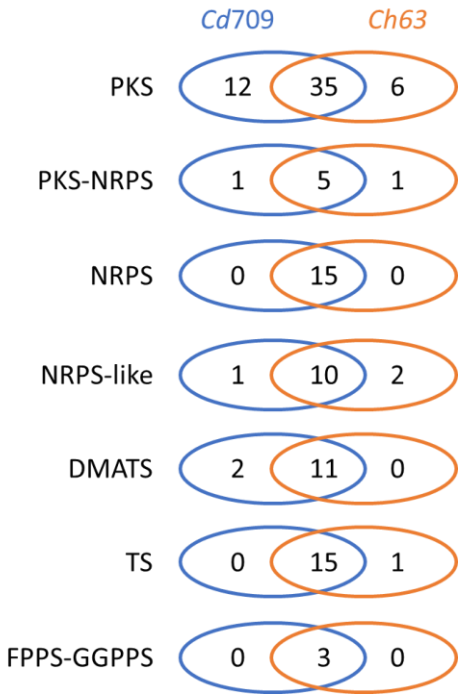
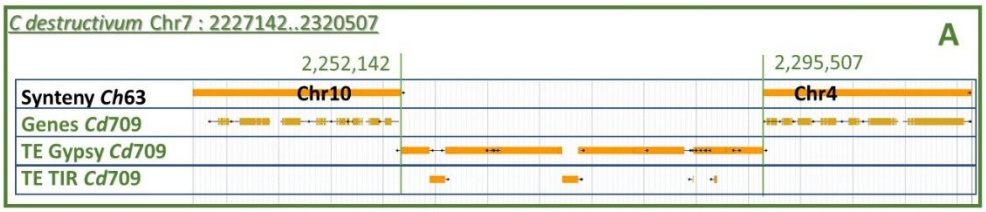
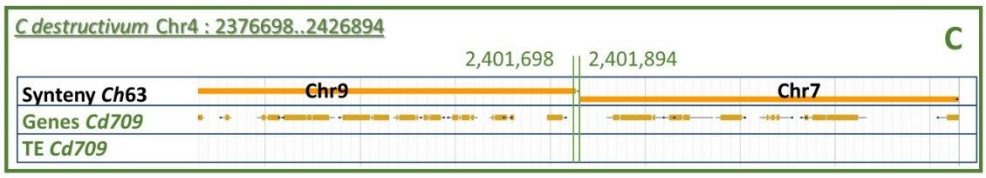
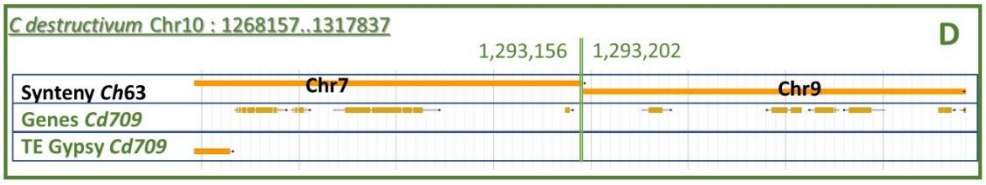
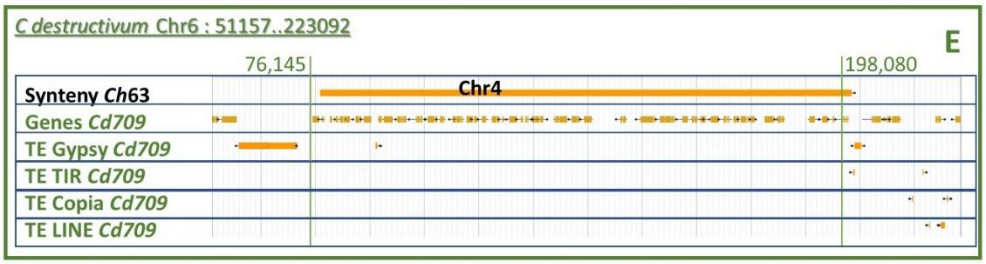


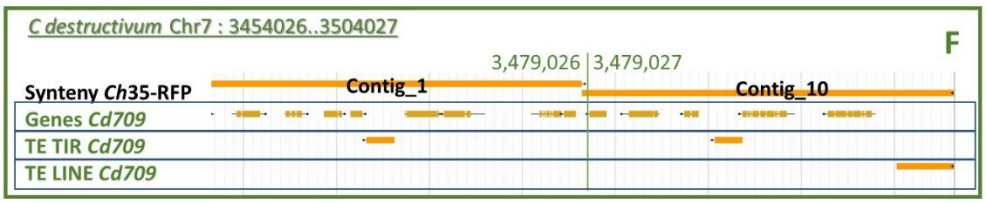
Figure S2: Number of genes encoding secondary metabolism key genes (SMKG) in *C. destructivum* (*Cd*709) and *C. higginsianum* (*Ch*63) genomes. Genes were those predicted as biosynthetic SMKG by Dallery et al. (2017) for *Ch*63 or by antiSMASH for *Cd*709. *Cd*709 genes not predicted as SMKG by antiSMASH, but orthologous to a *Ch*63 SMKG were included. PKS: Polyketide Synthase, NRPS: Nonribosomal Peptide Synthetase, DMATS: Dimethylallyltryptophan synthase, TS: Terpene Synthase, FPPS: Farnesyl pyrophosphate sythase, GGPPS: Geranylgeranyl pyrophosphate synthase

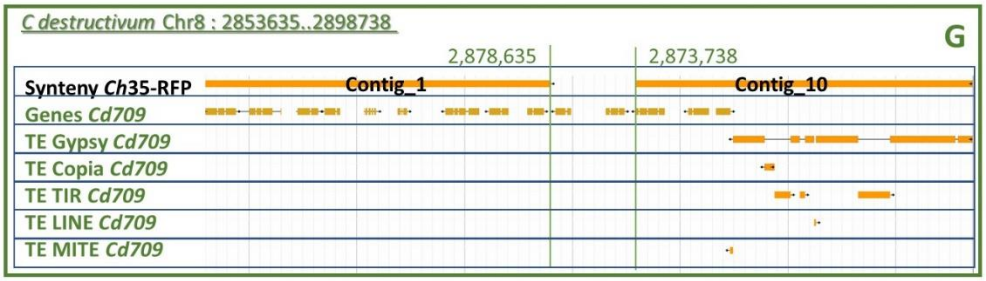












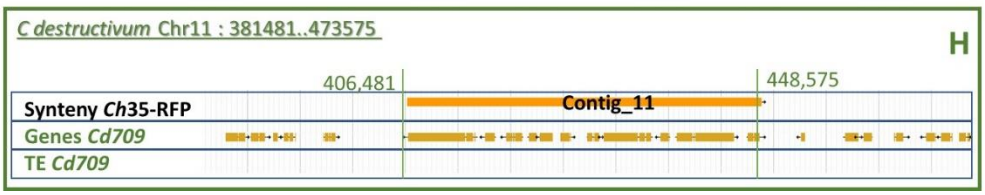


Figure S3: Genomic context of the large-scale rearrangements found between the chromosomes of *C. destructivum* (*Cd*709) and *C. higginsianum* IMI 349063 (*Ch*63) or MAFF 305635 (*Ch*35-RFP). A region of 25 kb surrounding each rearrangement zone is depicted. The synteny of each *C. destructivum* chromosome region with *C. higginsianum* IMI 349063 or MAFF 305635 is displayed, as well as predicted genes and transposable elements (TE).

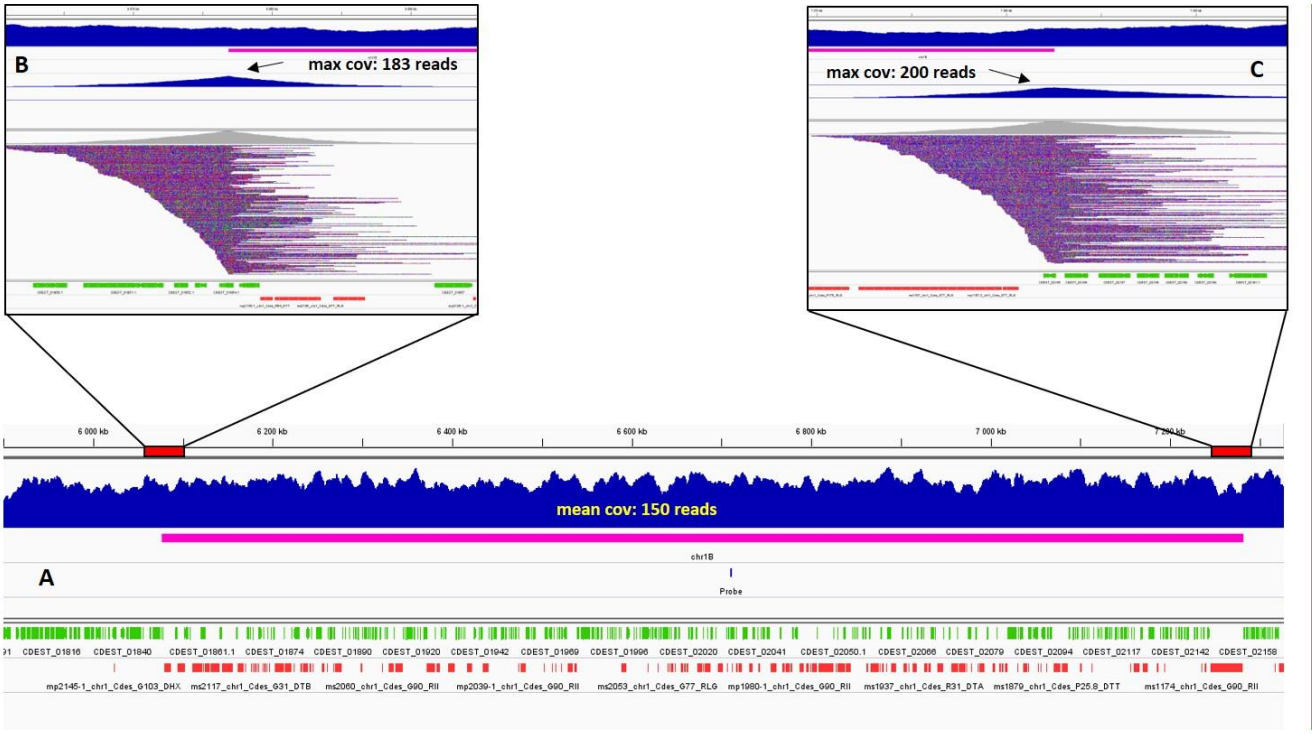


Figure S4: Screenshot showing long read coverage of chr1B and its boundaries (a). The chr1B region is shown in pink. Genes and transposable elements are in green and red, respectively. The read coverage is shown in blue. Long reads spanning the last base of 5' (b) and 3' (c) of the chr1B region were extracted from mapping to verify the support for each junction.

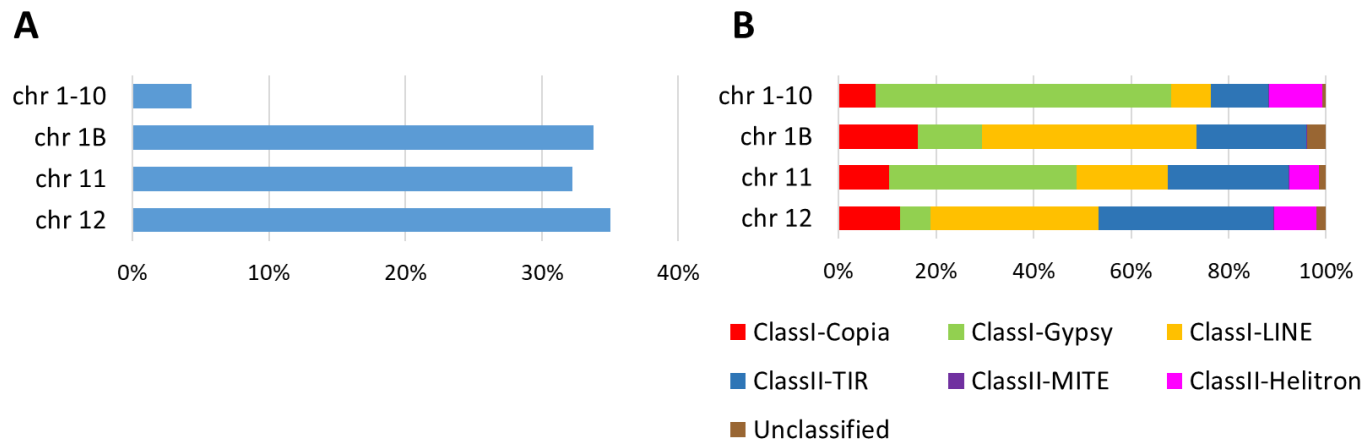


Figure S5: Transposable element (TE) content of the core chromosomes (1-10, excluding 1B), mini-chromosomes 11 and 12 and region 1B of *C. destructivum* (*Cd*709). (A) Histogram showing TE percent coverage by length for each chromosome compartment. **(B)** Histogram showing percent coverage by length of the identified TE orders or superfamilies for each chromosome compartment.

.

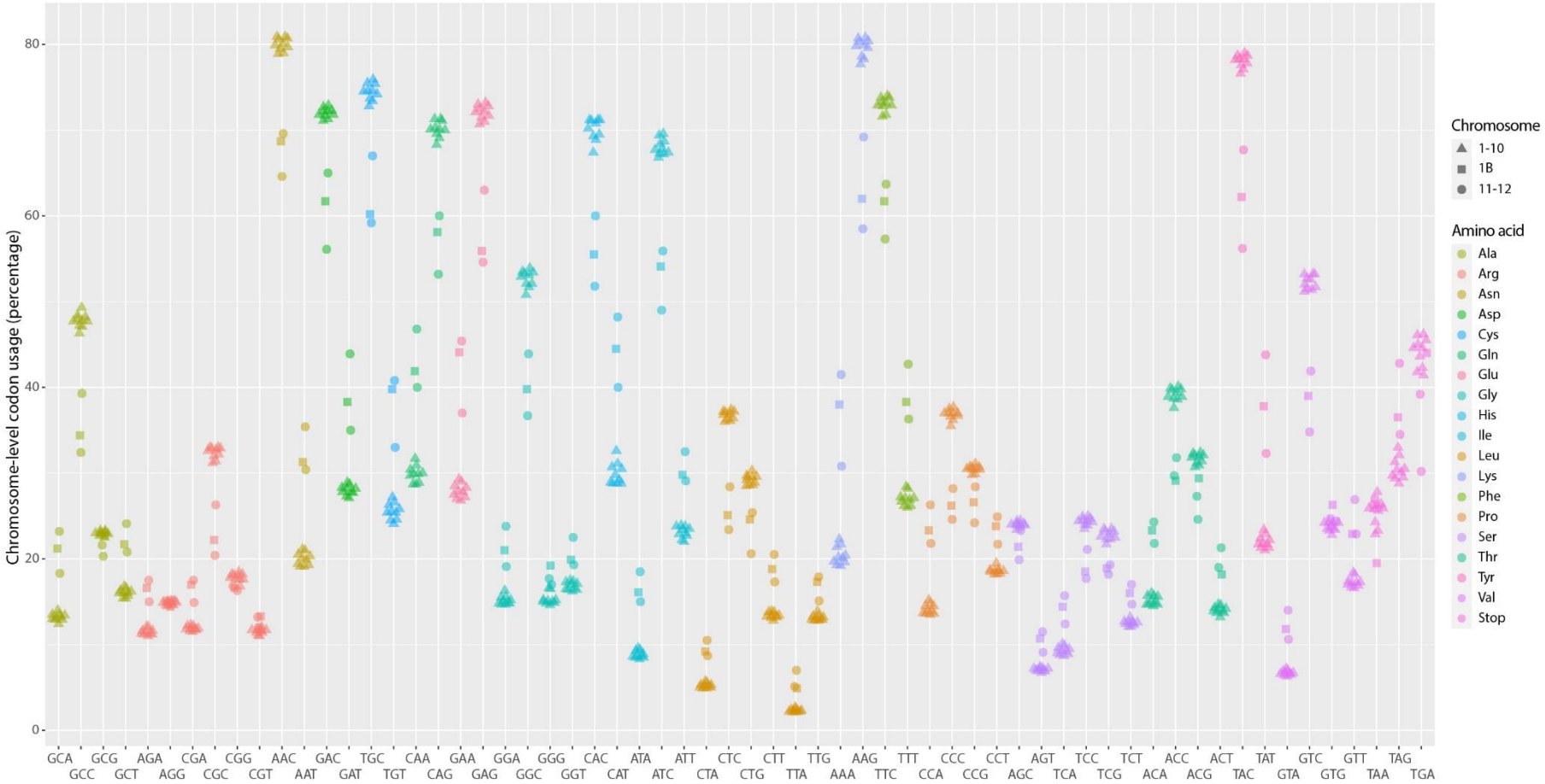


Figure S6: Codon usage bias in core chromosomes compared to mini chromosomes of *C. destructivum* LARS709. The 18 amino acids and their 59 corresponding codons are represented.

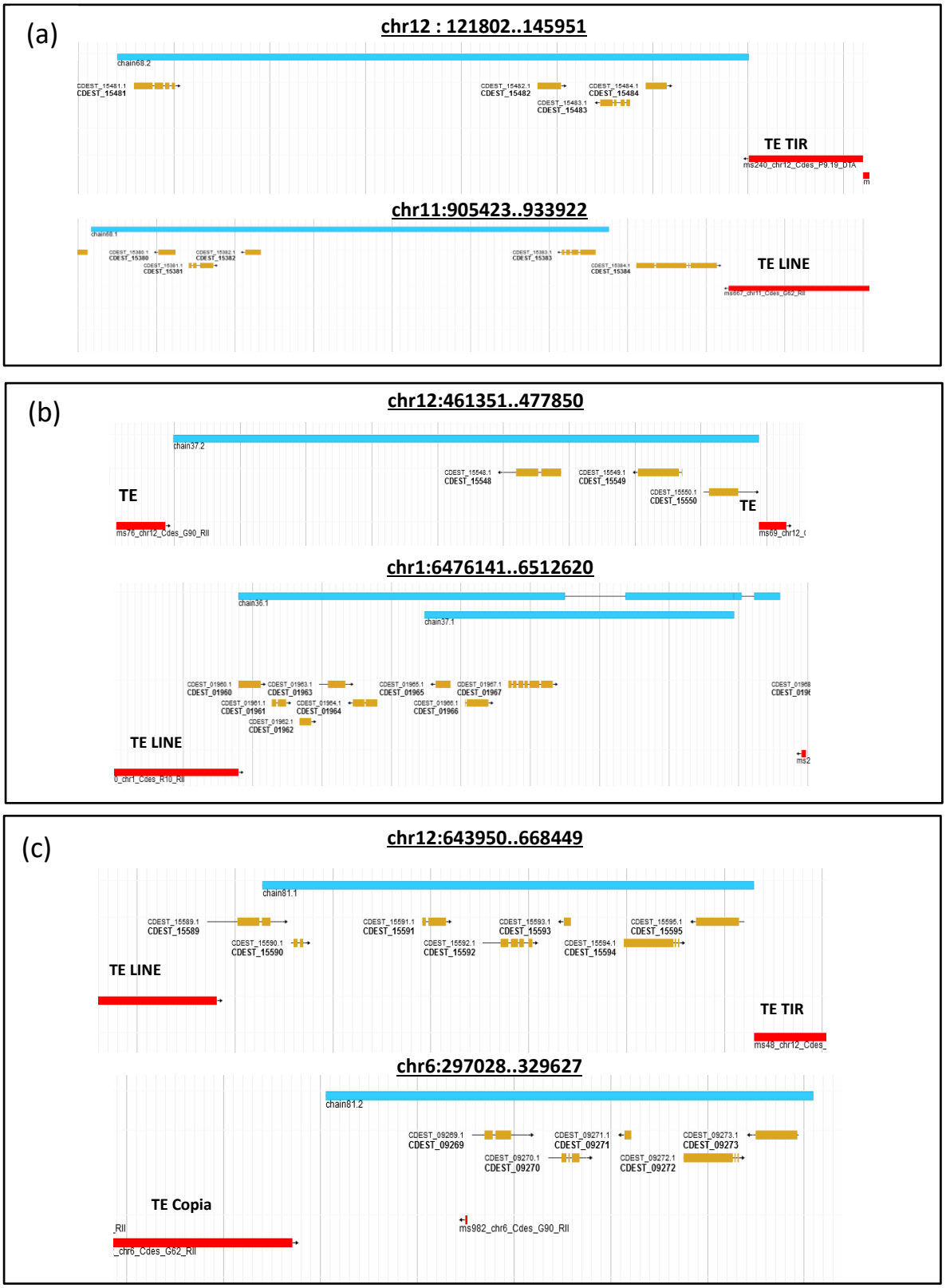


Figure S7: The inter-chromosomal segmental duplications regions detected in *C. destructivum* (*Cd*709) and their surrounding genomic regions rich in Transposable Elements (TE).

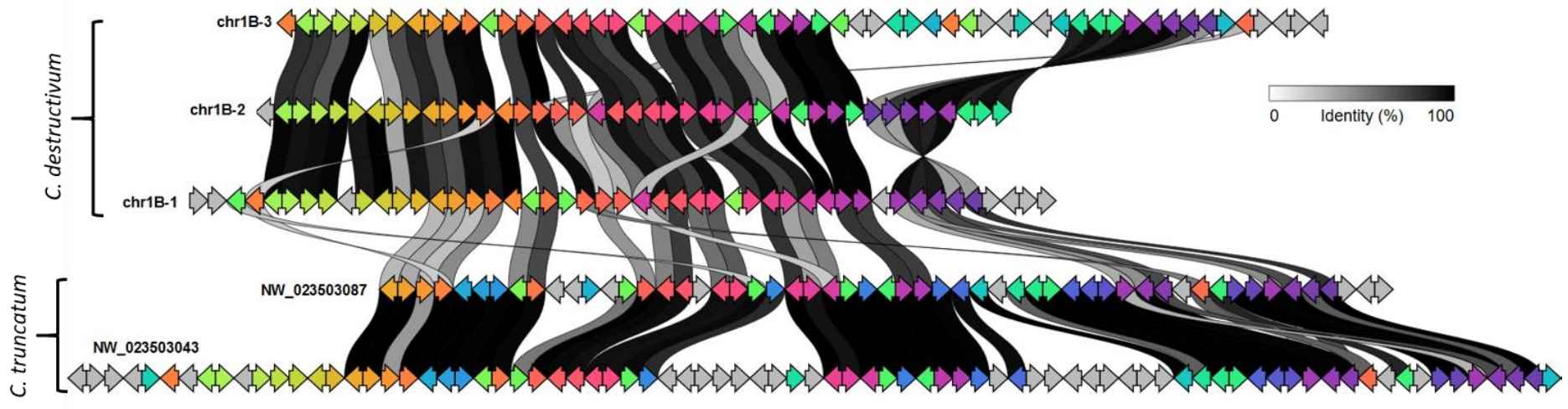


Figure S8: Synteny of segmentally duplicated regions in *C. destructivum* chr1B and *C. truncatum* CMES1059 contigs.

The duplicated regions chr1B-1 (chr1: 6242820..6372217), chr1B-2 (chr1: 6381352..6493327) and chr1B-3 (chr1: 7087658..7243846) were aligned at the protein level against the *C. truncatum* contigs NW_02350387 (218796..352904) and NW_023503043 (1580..217244). The chr1B-1 region contains the duplication SD1B-1 and adjacent genes (CDEST_1880 to CDEST_1926), the chr1B-2 region contains SD1B-2, SD1B-3 and SD1B-4 (CDEST_1927 to CDEST_1967) and the chr1B-3 contains SD1B-6, SD1B-7 and SD1B-8 (CDEST_2098 to CDEST_2154). The synteny map generated by Clinker shows a strong intra-species similarity and more divergence between the two species.