Research Article

Long-distance gene flow and recombination shape the evolutionary history of a maize pathogen

Flávia Rogério[1,2], Cock Van Oosterhout[3], Stéphane De Mita[4], Francisco Borja Cuevas-Fernández[1], Pablo García-Rodríguez[1], Sioly Becerra[1], Silvia Gutiérrez-Sánchez[1], Andrés G. Jacquat[5], Wagner Bettiol[6], Guilherme Kenichi Hosaka[7], Sofia B. Ulla[5], Jürg Hiltbrunner[8], Rogelio Santiago[9], Pedro Revilla[9], José S. Dambolena[5], José L. Vicente-Villardón[10], Ivica Buhiniček[11], Serenella A. Sukno[1*], Michael R. Thon[1*]

1 Department of Microbiology and Genetics, Institute for Agribiotechnology Research (CIALE), University of Salamanca, Villamayor, Salamanca, Spain
2 Present Address: Department of Plant Pathology, University of Florida, Gainesville, Florida 32611, USA
3 School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, UK
4 PHIM Plant Health Institute, Univ Montpellier, INRAE, CIRAD, Institut Agro, IRD, Montpellier, France
5 Faculty of Exact, Physical and Natural Science, National University of Córdoba, IMBIV-CONICET-ICTA, Córdoba, Argentina
6 Embrapa Environment, Jaguariúna, São Paulo, Brazil
7 Laboratory of Genetics of Microorganisms "Prof. Joao Lucio de Azevedo", Department of Genetics, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Piracicaba, Brazil
8 Federal Department of Economic Affairs, Agroscope, Centre of Competences Plants and Plant Products, Zurich, Switzerland
9 Misión Biológica de Galicia, Spanish National Research Council (CSIC), Pontevedra, Spain
10 Statistics Department, University of Salamanca, Salamanca, Spain
11 Bc Institute for Breeding and Production of Field Crops, Dugo Selo, Croatia

Corresponding authors: Flávia Rogério (flavia.rogerio@usal.es); Michael R. Thon (mthon@usal.es)

Abstract

The evolutionary history of crop pathogens is shaped by a complex interaction of natural and anthropogenic factors. The fungus *Colletotrichum graminicola* causes maize anthracnose which results in significant yield losses worldwide. We conducted a comprehensive investigation into the evolutionary genomics of *C. graminicola* using a collection of 212 isolates from 17 countries across five continents. Genomic analyses supported the existence of three geographically isolated genetic lineages, with a significant pattern of isolation by distance. We identified two distinct gene flow patterns, driven by short- and long-distance dispersal, likely resulting from the natural spread of the pathogen and the exchange of contaminated seeds. We present evidence of genetic introgression between lineages, suggesting a long history of recombination. We identified significant recombination events coalescing at distinct points in time, with the North American lineage displaying evidence of the most ancient recombination. Demographic modelling has indicated that North America is an intermediate between Brazil, Europe and an ancestral, unsampled source population, which is hypothesised to be Mesoamerican. Our analyses revealed that the global genomic structure of *C. graminicola* is shaped by geographic differentiation driven by long-distance migration and a long history of recombination and introgression. We show historical relationships amongst these lineages, identifying a potential route for fungal spread, with the North American population emerging ancestrally, followed sequentially by the Brazilian and European populations. Our research indicates that the European lineage is more virulent, which has implications for the potential emergence of new outbreaks of maize anthracnose in Europe.

* These are co-senior authors.

Introduction

Examining the processes that drive species evolution over time is a central focus of evolutionary genomics. Population genetics provides valuable insights into patterns of genetic variation within populations, enhancing our comprehension of their evolutionary history and dynamics (Dobzhansky 1973; Gladieux et al. 2011; Casillas and Barbadilla 2017). This knowledge has practical applications for studying how populations of plant pathogens respond to future environmental changes and selective pressure or for determining the role that human activities play in the dispersion of pathogens, which is crucial for effective disease control (Giraud et al. 2008; McDonald and Linde 2002; Everhart et al. 2021).

The genetic composition of pathogen populations is shaped by evolutionary forces and interactions with hosts and local the environmental conditions (Charlesworth 2010). Anthropogenic changes to the environment have a large impact on the interactions between evolutionary forces, which, in turn, have altered the Red Queen Dynamics between parasites and their hosts (Van Oosterhout 2021). Migration encompasses individual or propagule movement, leading to gene flow between isolated gene pools (McDermott and McDonald 1993). Understanding migration patterns is vital for providing information for disease prevention decisions, as it may introduce new genetic variations, including virulence factors and connect adapted populations (McDonald and Linde 2002). Together with gene flow, genetic recombination constitutes a vital evolutionary force driving variation and pathogen adaptation (Stukenbrock and Bataillon 2012; Milgroom 2015; de Vries et al. 2020). Genetic recombination has the potential to accelerate adaptation by the generation of new, potentially adaptive genotypes that are favoured by natural selection or by the combination of adaptive alleles at distinct loci that appear in different genetic backgrounds (Barton 2010). This process offers new substrates for adaptive evolution while simultaneously purging the genome of deleterious mutations accumulated over time (Felsenstein 1974; Taylor et al. 1999; Van Oosterhout 2021).

The fungus *Colletotrichum graminicola* is an important model for studying plant-pathogen interactions, being the pioneer amongst *Colletotrichum* species for genome sequencing (Perfect et al. 1999; O'Connell et al. 2012; Becerra et al. 2023). The genetic structure of *C. graminicola* has been explored globally (Rogério et al. 2023). Population genomics analyses have revealed that the pathogen is structured into three largely isolated populations associated with North America, Europe and Brazil. Genetic exchanges between these populations are facilitated by intra- and intercontinental migration, with evidence of genetic recombination. Nevertheless, the current impact of gene flow and recombination at the genome level, as well as the evolutionary origin of these populations, remains unknown.

Population genomics studies of many crop pathogens have demonstrated that human activities (domestication, trade and migration) play pivotal roles in the emergence and spread of plant diseases (Barrès et al. 2008; Chakraborty

2013; Goss et al. 2014; Kamvar et al. 2015b; McMullan et al. 2018; Tabima et al. 2020; Sotiropoulos et al. 2022). The dissemination of new virulence genes and agrochemical resistance amongst populations across large geographical areas is significantly influenced by gene flow, which plays a central role in driving pathogen evolution (Zhan et al. 2015). Like mutation and recombination, gene flow is a source of new genetic variation, which forms the substrate for selection in host-parasite interactions. By introducing novel alleles from other populations, gene flow indirectly increases the effective population sizes of the pathogen. The increased rates of gene flow in our human-modified environment give present-day pathogens an important co-evolutionary advantage (Van Oosterhout 2021). Understanding and monitoring gene flow is crucial for developing effective strategies to manage and mitigate the potential associated risks. Competition theory posits that a reduction in the competition of related genotypes may lead to the selection of more aggressive strains (Koskella et al. 2006; López-Villavicencio et al. 2007; Zhan and McDonald 2013). In this context, gene flow between genetically divergent lineages can decrease genetic relatedness and potentially promote the emergence of these more aggressive strains. This can pose a risk factor for host resistance breakdown.

Fungi exhibit a range of reproductive strategies involving genetic recombination outside conventional sexual reproduction, such as hyphal fusion (anastomosis) and parasexual cycles (Giraud et al. 2008). While clonality is the predominant mode of reproduction, population genetics studies highlight that recombination plays a key role in the genetic structure of some phytopathogenic fungi (Roca et al. 2005; Meng et al. 2015; Taylor et al. 2015; Ashu et al. 2017; Stukenbrock and Dutheil 2018; Drenth et al. 2019). Notably, recombination is reported to be an important factor in the ecology and evolutionary dynamics of numerous *Colletotrichum* species (Gale 2003; Souza-Paccola et al. 2003; Crouch et al. 2006; Diao et al. 2015; Rogério et al. 2023). However, certain studies have elucidated alternative mechanisms, such as the parasexual cycle, contributing to genetic recombination in the *Colletotrichum* genus (Souza-Paccola et al. 2003; Rosada et al. 2010; Barcelos et al. 2014; Nordzieke et al. 2019).

Gene flow and recombination lead to genomes with admixed ancestry, a process that is commonly referred to as genetic admixture (Beichman et al. 2018). The genomic signature of admixture is characterised by chromosomal segments that originate from distinct genetic backgrounds or populations (Carroll 2013). Closely affiliated with admixture is the term "introgression". Introgression is a process resulting from recombination and gene flow between significantly divergent gene pools, leading to the gradual transfer of genetic material from one pool to the other through admixed individuals (Anderson 1953; Stukenbrock 2016). At the far end of this continuum is horizontal gene transfer, which is the genetic exchange between otherwise reproductively isolated species. In this study, we adopted the term introgression to refer to genetic exchanges between divergent lineages. These genetic exchanges leave distinct signatures in the genome, identified through DNA sequence comparisons, with introgressed regions typically exhibiting high nucleotide similarity, followed by a sudden change in divergence (Gompert and Buerkle 2013; Rosenzweig et al. 2016; Ward and van Oosterhout 2016).

In this study, we focused on unravelling the evolutionary dynamics shaping the genetic structure of *C. graminicola*. Specifically, we addressed key questions:

What drives the genetic structure in *C. graminicola* populations? What role does genetic introgression play in its evolutionary history? What are the sources and historical relationships of its lineages? Our investigation employed a global collection of 212 isolates from 17 countries, enabling a comprehensive analysis of genetic exchanges at the genome scale and determination of the evolutionary origin of *C. graminicola* lineages. By addressing these questions, our study contributes to a deeper understanding of the evolutionary genomics of *C. graminicola*, with implications for disease management.

Methods

Data processing

In total, 212 isolates of *C. graminicola* were obtained from field samples and public culture collections from seventeen countries (Suppl. material 1: table S1). Mycelia from monosporic cultures were incubated in an orbital shaker in potato dextrose broth (PDB), for 3 days at 25 °C and 150 rpm under continuous light. Genomic DNA was extracted using DNeasy Plant Mini Kits (Qiagen Inc., Valencia, California, USA) according to manufacturer's instructions. Ninety-four isolates from Rogério et al. (2023) were included, along with 118 new isolates, sequenced using restriction site-associated DNA sequencing (RAD-seq) or whole-genome sequencing (WGS). A total of 82 new isolates were genotyped with RAD-seq by Floragenex, Inc. (Beaverton, OR, USA). RAD-seq libraries with sample-specific barcode sequences were constructed from DNA digested with the restriction enzyme PstI and then single-end sequenced (1 × 100 bp) in one lane of an Illumina HiSeq 2000 instrument (San Diego, CA; Beaverton, OR, USA). Whole-genome resequencing of 36 isolates was performed by Beijing Genomics Institute (Hong Kong, China) via the DNBSEQ library sequencing platform (2 × 150 bp).

The RAD-seq reads were demultiplexed and quality filtered via the PROCESS_RADTAGS module of the software Stacks v.2.10 (Rochette et al. 2019), using default parameters. The demultiplexed reads were aligned to the *C. graminicola* reference genome M1.001 (NCBI accession number PRJNA900520) with BWA v.0.7.8 (Li and Durbin 2009). For the WGS reads, the sequence quality was checked with FASTQC v.0.11.7 (Babraham Bioinformatics) and low-quality bases with Phred score of less than 30 were trimmed with TRIMMOMATIC v.0.3.8 (Bolger et al. 2014). The alignments were sorted and each bam file was assigned to a read group containing metadata, such as sample name, library and platform information, using SAMTOOLS v. 1.3 (Li et al. 2009). Variant calling was performed with the HaplotypeCaller module of GATK v.3.7, applying hard filters as suggested in GATK's Best Practices (Van der Auwera et al. 2013). Additional filters were applied via VCFTOOLS (Danecek et al. 2011), to retain only biallelic SNPs with a genotype rate of 90% and sites with a minor allele frequency of less than 5% were retained.

Global population structure

Repeated multilocus genotypes (here referred to as clones) were identified via the function MLG.FILTER with a threshold determined via the CUTOFF_PREDICTOR tool on the basis of Euclidean distance via the package POPPR v.2.8.6 (Kamvar et al. 2015). Population genetic analyses were conducted on a clone-corrected dataset

(i.e. only one representative of each repeated multilocus genotype was retained). To investigate population subdivision, first, we used the clustering method implemented in the SNMF software to infer individual ancestry coefficients. The best number of clusters (*K*) was chosen using the cross-entropy criterion, based on the prediction of masked haplotypes to evaluate the error of ancestry estimation and the quality of fit of the model to the data. We ran 10 repetitions for each value of *K* ranging from 1 to 10. To confirm the population differentiation, we employed a principal component analysis (PCA), with *a priori* geographical knowledge from sampling (by continent) and a discriminate analysis of principal components (DAPC), both of which were conducted in R with the package ADEGENET v.2.0.1 (Jombart and Ahmed 2011). The FIND.CLUSTER function was used to identify the most likely grouping in the data on the basis of the Bayesian Information Criterion (BIC) calculated for 1−10 clusters. Additionally, we reconstructed a neighbour network using SPLITSTREE v4 (Steinrücken et al. 2019) to visualise phylogenetic signals based on a distance matrix with default parameters.

To further understand the impact of geographic subdivision on population structure, we conducted isolation-by-distance analyses. We computed pairwise genetic dissimilarity and geographical distance between samples and examined their relationships via regression analysis. The correlation between the matrices of genetic and geographic distance was also observed via a simple Mantel test with 1000 permutations, as implemented in the VEGAN package in R (v.2.6.4). Genome variability was assessed through nucleotide diversity statistics computed via EGGLIB v.3.3.2 (Siol et al. 2022).

Footprints of recombination

To examine the impact of genetic exchanges on tree topologies, we constructed a consensus tree using DENSITREE 2 (Bouckaert 2010). BEAUTI v.1.5.3 was used to generate XML files for each contig. The substitution models were selected using BMODELTEST (Bouckaert and Drummond 2017), which detects the best model and automatically adds it to the XML file. BEAST 2 (v.2.5.1) (Bouckaert et al. 2019) was used to generate phylogenetic trees via Markov Chain Monte Carlo (MCMC) for 1×10^6 generations with a sample frequency of 2,000 and burn-in of 10%.

Recombination was analysed in multiple genome alignments by chromosome, including non-variant sites. We used the module GENOTYPEGVCS in GATK to include genotypes from all positions present in the reference genome and BCFTOOLS v.10.2 (Danecek et al. 2021) to obtain consensus sequences for each chromosome. Recombination events were analysed by the RDP4 software (Martin et al. 2021) via seven independent algorithms (RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMERA, SISCAN and 3 SEQ). Significant events (*P* < 0.05) were considered if they were significant with three or more detection methods (Jouet et al. 2015). Only events for which the software identified the parental sequences (i.e. no "unknowns") were considered. To visualise the significant recombinant blocks identified by RDP4, we used HYBRIDCHECK software (Ward and van Oosterhout 2016) with a step size of 1 and a window size of 500 SNPs. HYBRIDCHECK was also used to date the recombination events. The divergence times between the donor and recombinant sequences were calculated via JC69 mutation correction, assuming a mutation rate of 10^{-8} per generation and a generation time of one year (Kasuga et al. 2002).

Recombination events were also dated through Bayesian Inference in Beast 2. We selected some recombinant blocks detected by hybridcheck and applied a Bayesian coalescent approach for dating. XML files were created using beauti, where the module bmodeltest was used to determine the optimal substitution model under a mutation rate of 10^{-8} and a strict clock rate. We assumed a model of a coalescent constant population with default priors and a chain length of 1×10^7 generations. Each dating analysis was replicated five times. trace was used for visualising posterior parameters. The tree height statistic, which represents the marginal posterior distribution of the age of the root of the entire tree, was considered the divergence time from the common ancestor (tMRCA) for the set of sequences evaluated.

For dating, we assumed a generation time of 1 year. Note that the reported dates are likely an overestimate of the true age of recombination. In other words, the "true" recombination event may be more recent than our inferred estimate. There are two ways dating estimates can be biased, leading to overestimating the true date of introgression. First, there is an overestimation because the descendants of the actual donor and recipient sequences are unlikely to have been sampled.

Consequently, the introgressed regions look more divergent (and thus "older"). Second, even if the direct descendants of the donor sequence were included in our analysis, they may have diverged from the block in the recipient block owing to subsequent recombination. This may have introduced multiple nucleotide polymorphisms into the introgressed region. Given that we assume that divergence is driven by the mutation rate (i.e. a molecular clock), polymorphisms introduced by recombination would make the divergence appear older (see Jouet et al. (2015)). Both processes lead to an overestimation of the true date of introgression.

To characterise genes present in the introgressed regions, we extracted the transcripts and proteins from the corresponding regions in the reference genome (Becerra et al. 2023) using gffread (Pertea and Pertea 2020). For the enrichment analysis the Gene Ontology terms were assigned from re-annotated transcripts using blast2go (Götz et al. 2008) against the SwissProt database and used as input to the topgo R package (Alexa and Rahnenfuhrer 2021). We used signalp v. 5.0 (Almagro Armenteros et al. 2019) to identify secreted proteins.

Demographic inference

We inferred the evolutionary history of *C. graminicola* genetic lineages using Approximate Bayesian Computation (ABC), based on random forest (RF), conducted in the Python package Pyabcranger (Collin et al. 2020). We based our analysis on three genetic lineages (North American, Brazilian and European) treated as independent populations. We used the Python package egglib to perform coalescent simulations and to compute the summary statistics and Pyabcranger to perform model choice and parameter estimation. ABC-RF uses a classification vote system after bagging (i.e. aggregating bootstrap results) the simulated outputs to identify the most likely model. The model with the highest number of votes is deemed to be the most likely model. To shed light on divergence events and colonisation routes, we modelled 22 scenarios of divergence, including asymmetrical migration and bottleneck events, assuming either stepping-stone or single-source models. Models with and without

an unknown population (i.e. the 'ghost' population) as an ancestral population were tested. In our models, every divergence event was immediately followed by a bottleneck; otherwise, a constant population size was maintained for all populations. For the detailed methodology, see Suppl. material 2.

Pathogenicity assays

Pathogenicity assays were performed by *in vivo* leaf blight assays, following Vargas et al. 2012), as described by Rogério et al. (2023). A total of 56 isolates were evaluated with three biological replicates, with three plants inoculated with each fungal strain. The average lesion size for each strain was measured by scanning leaves and quantifying the lesion size with the image processing software Paint. NET v.3.5 (https://www.getpaint.net). Due to the limitations of accommodating all the isolates simultaneously, assays were conducted in batches of 7–11 strains. The reference strain M1.001 was included in all batches to estimate batch effects. We compared the results of strain M1.001 across different batches via analysis of variance (ANOVA), with lesion size as the response variable, batch size as the predictor variable and the null hypothesis of no difference between batches. The highly significant differences ($P < 0.0001$) indicated a clear batch effect. To mitigate these effects, lesion sizes were standardised using the mean and standard deviation of the control strain specific to its batch. ANOVA of standardised data revealed significant differences amongst isolates, so we performed a Sidäk comparison-of-means test. This choice was made because it is exact for independent comparisons and easily translated into a graphical representation, facilitating the identification of significant differences amongst strains.

Additionally, the isolates tested in the virulence assay were divided into groups, based on sampling year and genetic structure. The isolates were classified into three categories, based on genetic group and virulence (less, equal or more virulent than isolate M1.001) and subjected to Fisher's exact test. The isolates were also grouped by year into three categories (before 2012, between 2012 and 2016 and after 2016) and subjected to a Kruskal–Wallis test. The statistical tests and figures were implemented in the R.

Finally, we examined whether there was a positive correlation between pairwise genetic distance between isolates and similarity in their level of virulence. Genetic distance was calculated as the difference in the number of bases in pairwise comparisons. The similarity in virulence level was assessed by comparing the difference between lesion areas (calculated as the lesion area of isolate A minus the lesion area of isolate B). The correlation between pairwise genetic distance and similarity in virulence was analysed using a Mantel test with 1000 permutations, as implemented in the vegan package in R (v.2.6.4).

Abbreviations

RAD-Seq	Restriction site-associated DNA sequencing
WGS	Whole Genome Shotgun Sequencing
NA	North American
EU	European
BR	Brazilian
AMOVA	Analysis of Molecular Variance

Results

Distinct levels of genetic diversity and admixture define *C. graminicola* lineages

We characterised the global genetic structure of *C. graminicola* via a total of 212 isolates, based on 7,207 biallelic SNPs were distributed across 13 chromosomes, with an average of 1.3 SNPs per 10 kb (Suppl. material 3: fig. S1; Suppl. material 1: table S1). After clone correction, 166 unique multilocus genotypes were retained in the dataset (Suppl. material 1: table S2). A neighbour-net network clearly revealed three distinct genetic groups, with isolates largely grouped by their geographic origin: South America, Europe (including Argentina) and North America (Fig. 1A; Suppl. material 3: fig. S2). The observed population subdivision corroborates the conclusion drawn in our previous study
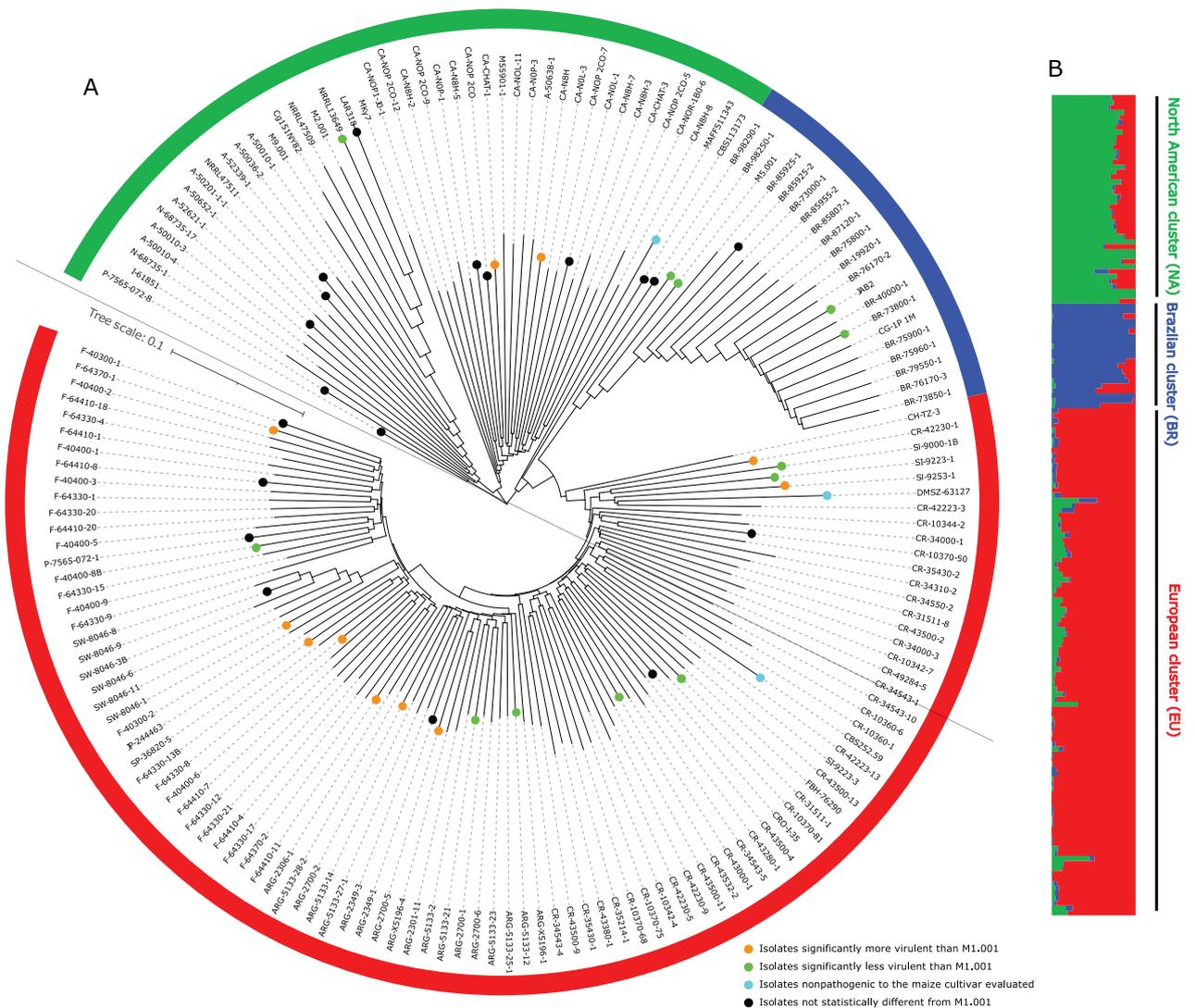
Figure 1. Population subdivision of *Colletotrichum graminicola* based on 7,207 single-nucleotide polymorphisms in the clone-corrected dataset. **A** Neighbour-net network showing relationships between isolates, visualised in a circular tree via ITOL v. 6 (Letunic and Bork 2007). The dots represent isolates whose pathogenicity was essential. Samples SW-8046-2, P-7565-072-6, SW-8046-13, CR-49298-1, SL-9253-4, CRO-I-41, P-7565-072-3, CR-10342-5, F-64330-7 and F-64330-2 were removed from the network because they are clones; **B** individual ancestry proportions in *K* = 3 clusters estimated with SNMF each genotype is represented by a vertical bar.

(Rogério et al. 2023), which was based on a smaller number of isolates (Fig. 1. 108 versus 212 presently). By including more genetic variants, we obtained a better resolution of the genetic diversity distribution, but clustering methods still supported the existence of these groups. These three groups are referred to as Brazilian (BR), European (EU) and North American (NA) groups, respectively.

Population substructuring and admixture were analysed in further detail using the non-parametric method implemented in SNMF. On the basis of the cross-entropy criterion, which was confirmed by the Bayesian Information Criterion, the model with $K = 3$ clusters best captures the population substructure of the whole dataset (Suppl. material 3: fig. S2). Interestingly, 80% of the isolates present evidence of admixture, which suggests genetic exchanges between *C. graminicola* genetic lineages (Fig. 1B; Suppl. material 1: table S4). Some isolates are highly admixed, such as P-7565-072-8, MAFF511343, CBS113173, M5.001, BR-98290-1 and BR-98250-1. These isolates are responsible for the "loops" or reticulate nodes in the neighbour network (Suppl. material 3: fig. S3). In contrast, a small proportion of isolates are completely pure, i.e. 20% of isolates have individual ancestry coefficients that assign them to a single cluster ($q > 0.99$, where q represents the proportion of an individual genome originating from a given ancestral gene pool) and are considered pure (Suppl. material 3: fig. S3, Suppl. material 1: table S4). The Brazilian cluster presented the largest proportion of pure isolates. The North American and Brazilian lineages showed the most ancestry shared with Europe, with isolates showing membership lower than 1% with other clusters (Suppl. material 1: table S4). There was one exception, i.e. the isolate MAFF511343, which originated from a public culture collection. Interestingly, older North American isolates are genetically pure: LAR318 (1990), M2.001 (1975), Cg151NY82 (1982), M9.001 (1990), NRLL13649 (1988) and NRLL47509 (2005). In addition, the NA lineage has the deepest rooting branch and these isolates have the longest terminal branches, indicating a more ancient origin of this group. The European clade encompasses all samples from European countries, one isolate from China and Japan and all isolates from Argentina (see highlighted samples in Suppl. material 3: figs S3, S4). Except for some Portuguese samples (P-7565-072-4, P-7565-072-6, P-7565-072-7, P-7565-072-8), all the isolates were grouped with samples from the same country of origin.

The relationships amongst all the genotypes visualised in a neighbour phylogenetic network (Suppl. material 3: fig. S3) supported a clear worldwide subdivision into three lineages, with long branches separating lineages by geographic origin. We observed several loops (or reticulations) in the network, indicating phylogenetic inconsistencies, where part of the focal sequence resembled that of a second sequence and one part of that sequence was more like a third sequence. This pattern is likely caused by recombination, as evidenced by the pairwise homoplasy index (PHI test), which was also calculated using SPLIT-STREE which detected statistically significant evidence for recombination for the three lineages ($P < 0.001$).

The consensus tree generated by DENSITREE, based on 1291 SNPs on chromosome 1, revealed many different topologies. This representation further corroborated the inferred population subdivision into three groups, with individuals from different lineages consistently separated by consensus branches (Suppl. material 3: fig. S5). We observed that isolates from different continents appear to have exchanged partial genomic sequences, indicating that some of the genomes have

different phylogenetic origins, which is consistent with gene flow and recombination between diverged lineages (i.e. genetic introgression). These events are illustrated by the diagonal lines of terminal branches in the DENSITREE cladogram.

Isolation-by-distance (IBD) analyses revealed a positive relationship between genetic distance and geographic distance. The Mantel test revealed a significant correlation between the genetic and geographic matrices (R^2 = 0.3607, P < 0.001). Regression analyses, including models with logarithmic, square root and quadratic transformations, demonstrated a positive relationship, with geographic distance explaining up to 35.80% of the genetic diversification observed (quadratic regression $F_{2,40142}$ = 11196.59, P < 0.001, R^2 = 0.3580) (Suppl. material 3: fig. S6). The observed pattern of IBD at local and global scales is characterised by genetic dissimilarity increasing rapidly at shorter distances due to increasingly less efficient gene flow. Interestingly, it appears to be plateauing or even declining at larger distances. This lack of correlation at larger distances may result from genetic drift acting independently in distant populations. Consequently, these populations might be genetically more similar due to random processes or shared historical events, rather than continuous gene flow. In addition, long-distance gene flow, mediated by human activities such as trade or agricultural practices, may obscure IBD signals at larger scales by introducing genetic connectivity between distant populations.

Recombinant analysis

Recombination on chromosome 1 was analysed via RDP4, with a focus on fifteen isolates representative of each lineage (North America: LAR318, NRRL13649, I-618151, A-52621-1, CA-CHAT-1; Brazil: BR-85955-2, BR-98290-1, BR-85925-1, M5.001, JAB2; European: F-64330-2, ARG-X5196-1, P-7565-072-8, CR-34543-1, CBS11373). Similar results were found on the other chromosomes (Suppl. material 3: fig. S7). We detected 15 significant recombined blocks with large stretches of nucleotide similarity across lineages (Suppl. material 1: table S5). Amongst the recombination events detected, two (events 2 and 6) were selected for more detailed analysis. We selected these events to illustrate the genomic signature of intercontinental genetic introgression, noting that the other 13 events presented a comparable (albeit less clear) signature. Event 2 involves isolates BR-85955-2 (Brazilian lineage), P-7565-072-8 (European lineage) and I-61851 (North American lineage), henceforth referred to as triplet 1. In addition, event 6 consisted of the isolates M5.001 (Brazilian lineage), CR-34543-1 (European lineage) and A-52621-1 (North American lineage), henceforth referred to as triplet 2. Next, we visualised the recombinant blocks and dated their divergence using HYBRIDCHECK (Fig. 2). The visualisation of nucleotide similarities between sequences through an RBG colour triangle revealed a mosaic-like genome structure in these triplets (Ward and van Oosterhout 2016). Rather than a single recombination block as detected by RDP4, the blocks appear to be fragmented, possibly due to subsequent recombination and/or new mutations. The age estimates of those recombinant blocks varied widely amongst blocks (Suppl. material 1: table S6). We observed younger blocks dating back to 6,100 years before the present and the older block sharing a common ancestor dated to 100,000 years before the present.

We performed an additional recombination analysis to further investigate the variation in introgression age at the intracontinental scale. Isolates from the
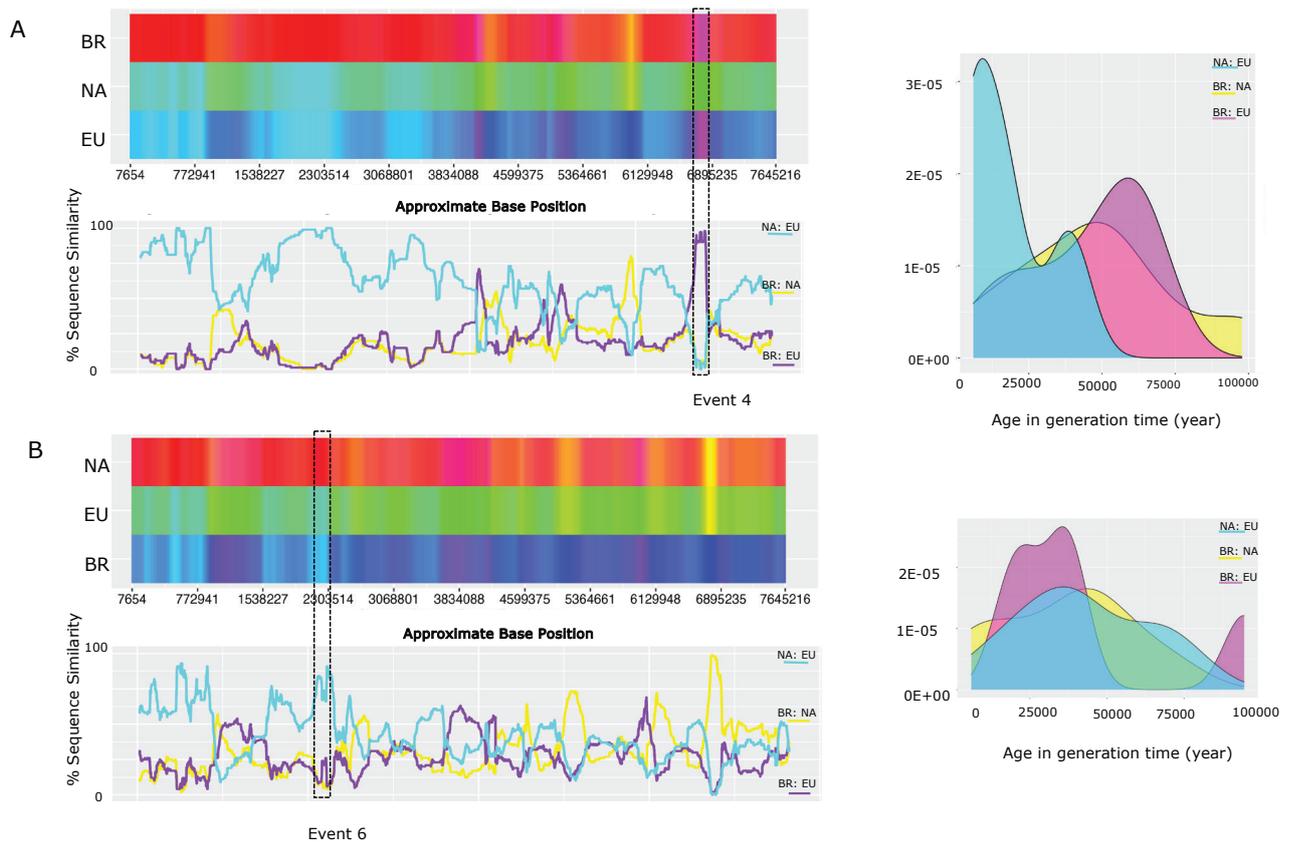
Figure 2. Recombination analysis of chromosome 1. **A** Left: Sequence similarity visualised through an RBG colour triangle by HYBRIDCHECK (involving the isolates BR-85955-2, P-7565-072-8 and I-61851, respectively). The significant recombination event detected by the rpd4 software is enclosed in a dashed box; right: age distribution of recombinant blocks detected by HYBRIDCHECK; **B** similar analyses of the BR, EU and NA lineages using the isolates M5.001, CR-34543-1 and A-52621-1, respectively.

same lineage were organised into three further triplets: triplet 3 – M5.001:BR-98290-1:BR-85925-1 (Brazilian lineage), triplet 4 – F-64330-2:F-64330-7:F-64330-13B (European lineage) and triplet 5 – NRRL13649:I-61851:CA-CHAT-1 (North America lineage) (Suppl. material 3: fig. S8). The recombinant blocks detected within each lineage were dated using HYBRIDCHECK and a Bayesian coalescent approach implemented in BEAST (Suppl. material 1: tables S6, S7). The results indicated that the number of recombinant blocks and their age estimates differed significantly between lineages (Kruskal–Wallis test, H = 33.5, d.f. = 2, $P < 0.001$). The age estimates provided by HYBRIDCHECK revealed that the blocks involving Brazilian and European isolates were younger, i.e. 25,000 and 30,000 years, respectively (Suppl. material 3: fig. S7). In contrast, the North American isolates presented a greater number of blocks, with a much older signature of recombination, with some blocks dating nearly 100,000 years (Suppl. material 1: table S6 and Suppl. material 3: fig. S9).

To apply the Bayesian dating method, we selected three recombination blocks within each triplet identified by HYBRIDCHECK (for the blocks selected, refer to Suppl. material 1: table S6, highlighted in yellow). These blocks were aligned against five isolates from each lineage to determine whether they were conserved in the isolates (Suppl. material 1: table S7). Interestingly, these blocks were present in all the isolates, displaying a pairwise identity greater than 99%. These sequences were subsequently used to estimate their age using BEAST. The dating results were consistent with the HYBRIDCHECK algorithm dating re-

sults, although there were differences in the order of magnitude of dates between the methods (Suppl. material 1: tables S6, S7). The age estimates for the North American (NA) blocks were more than seven times older than those for blocks from Brazil, ranging from 90,000 to 1,048,860 years (Suppl. material 1: table S7), which was a highly significant difference (Kruskal-Wallis test: H = 32.37; d.f. = 2; P < 0.001). This suggests that the NA lineage has experienced more ancient introgression than the other two lineages. We note that, although the age of introgression may be overestimated, for the reasons explained in the Materials and Methods, the relative difference in age estimates is noteworthy and likely to represent genuine differences in the history of introgression between the three geographic clusters.

Additionally, we characterised genes present in the introgressed regions; however, transcripts were recovered only for a few events (Suppl. material 1: table S11). Enrichment analysis for Gene Ontology terms did not identify significant terms within the recombinant blocks. SIGNALP analysis predicted that several proteins correspond to genes encoding secreted proteins (see Suppl. material 1: table S12). However, these regions do not show a significant enrichment for known virulence-associated genes. Therefore, we conclude that genetic exchanges are not likely to occur preferentially in genomic regions associated with virulence.

Unravelling population history

We analysed the evolutionary history of the isolates, assuming three genetic lineages, North American (NA), Brazilian (BR) and European (EU), as independent populations. In total, we analysed 22 demographic models (Suppl. material 3: fig. S9) using Approximate Bayesian Computation (ABC) based on random forest (RF), conducted in the Python package PYABCRANGER (Collin et al. 2020). The demographic models tested were organised into four independent comparisons to determine the most likely ancestral population (Suppl. material 3: figs S10–S12). Overall, our demographic modelling analysis strongly indicates a significant contribution of an unsampled population as the ancestral source shaping the complex evolutionary history of the lineages investigated. However, the North American population appears to have emerged first, followed by the Brazilian and European populations (Fig. 3). North America seems to have acted as an intermediate between the unknown source population and the rest of the world.

Virulence analysis

We detected significant differences in virulence amongst the fifty-three isolates evaluated in an analysis of variance test (ANOVA, P < 0.0001). Sidäk's comparison-of-means test revealed that, compared with M1.001, 16 isolates presented significantly greater virulence, whereas 14 isolates presented significantly reduced virulence compared to the reference strain M1.001 (Fig. 1A; Suppl. material 3: fig. S13; Suppl. material 1: table S13). The isolates CBS252.59, MAFF511343 and DMSZ-63127 were non-pathogenic to the maize cultivar evaluated.

The isolates were also classified according to genetic lineage (NA, EU and BR) and sampling year. Differences in virulence between lineages were not statistically significant, according to a Kruskal-Wallis test, although greater variabil-
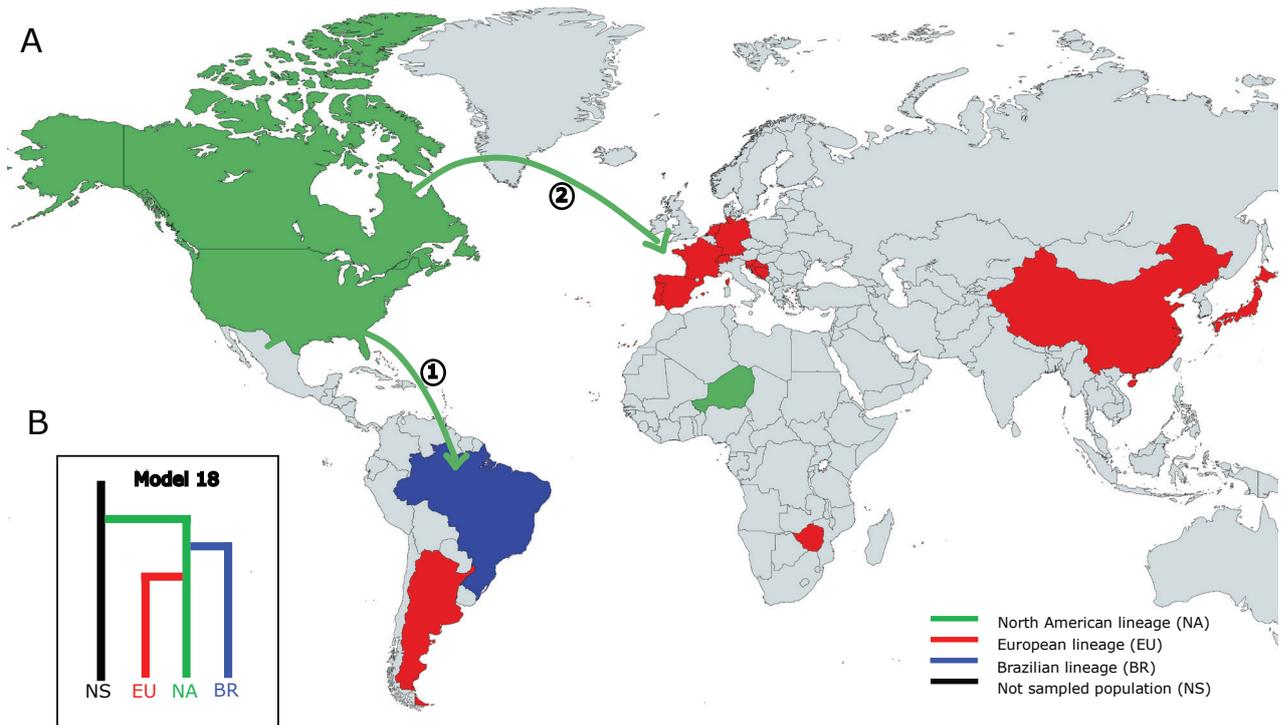
Figure 3. Worldwide sampling of *Colletotrichum graminicola*. **A** Map showing a likely colonisation route of *C. graminicola* reconstructed by Approximate Bayesian Computation (ABC) using 208 isolates from 17 countries. Different colours indicate distinct lineages and arrows indicate divergence events: **(1)** first divergence event from NA to BR; **(2)** second divergence event from NA to EU; **B** the most strongly supported demographic model was selected based on random forest population voting (see Suppl. material 2 for details).

ity was observed in the European group (Suppl. material 3: fig. S13a). We used the strain M1.001, which was originally collected from North America in 1972 during the U.S. outbreak and is highly virulent (O'Connell et al. 2012; Rogério et al. 2023), as a control (or baseline) to classify the isolates into three categories (less, equal or more virulent than the isolate M1.001; see Suppl. material 1: table S10). When the isolates were cross-tabulated into these categories, the percentages of strains in the different lineages were statistically significant (Fisher's exact test, $P < 0.05$). All the isolates from the Brazilian lineage (BR) are less than or equally virulent to the M1.001 strain is, while 41.2% of the European lineage (EU) and 20.0% of the North American lineage (NA) are more virulent. These results indicate a relationship between genetic structure and virulence, particularly in strains that are now more virulent than M1.001, especially in the European lineage. When the isolates were grouped by year (before 2012, between 2012 and 2016 and after 2016), a significant difference was observed between groups (Kruskal–Wallis test, H = 9.02, d.f. = 2, $P = 0.01$). The isolates from 2012–2016 presented the highest virulence, suggesting that, within the three genetic groups, the relative level of virulence can change over time (Suppl. material 3: fig. S13b).

The Mantel test performed between pairwise genetic distance and pairwise similarity in virulence revealed no correlation between these variables ($R = -0.006257$, $P = 0.47053$). Interestingly, the isolates SW-8046-1 and SW-8046-2 exhibited (nearly) identical genotypes. Yet, they displayed contrasting virulence values (Fig. 1A and Suppl. material 3: fig. S14), indicating that minor genetic differences can have a large effect on virulence or that variations in virulence may be attributed to factors other than genetic diversity.

Discussion

A global collection of *Colletotrichum graminicola* field isolates from 17 countries allowed a comprehensive investigation into the evolutionary history of this important maize pathogen. We generated and analysed a broad dataset with a higher resolution than our previous study (Rogério et al. 2023). By including nearly twice the number of isolates, we revealed a consistent global structure, indicating that the three geographically near-isolated populations were restricted and largely defined by their continental origin. We also observed positive correlation in the isolation-by-distance (IBD) analysis at local and global scales. Recombination analysis revealed compelling evidence of genetic introgression between the three geographic groups, suggesting a long history of recombination, possibly through sexual reproduction. The North American lineage displayed the most ancient evidence of recombination. Demographic modelling supported these findings, indicating that North America is an intermediate between the source population and the global distribution, potentially originating from an unsampled population, which is hypothesised to be Mesoamerican. We propose that the global genetic structure of *C. graminicola* is shaped by geographic differentiation driven by long-distance migration and genetic recombination events occurring at distinct points in time.

These lineages exhibit different levels of overall genetic diversity, as indicated by estimated genomic indices, suggesting independent evolutionary trajectories (Suppl. material 1: table S3). The North American lineage possesses the highest estimates of genetic variation. Many factors, including demographic effects can increase genetic diversity. We formulated four (non-exclusive) hypotheses explaining the greater level of diversity in the NA lineage: 1) NA has a larger contemporary effective population size or (2) NA is the ancestral (source) population and other populations experienced a bottleneck when they emerged; (3) The NA is experiencing recurrent incoming gene flow and/or (4) The NA has higher levels of (sexual or parasexual) recombination, limiting the consequences of background selection. Below, we examine these hypotheses in more detail.

Natural gene flow in fungi tends to be mediated through the transport of spores. Members of the genus *Colletotrichum* typically produce asexual spores (conidia) in acervuli immersed in mucilaginous masses disseminated through water splashes, accounting for short-range dispersal (Madden 1997). In contrast, sexual spores (ascospores) are associated with medium- to long-distance dissemination, as they can be ejected upwards and dispersed by air currents, reaching longer distances (Taylor et al. 1999; Alaniz et al. 2019). Sexual reproduction in *C. graminicola* has the potential (even sporadically) to contribute to medium- to long-distance gene flow (Rogério et al. 2023), given the prevailing assumption that asexual reproduction serves as the primary reproductive strategy for short-distance fungal dissemination (Drenth et al. 2019).

Human-mediated transport is a major driver of the long-distance dissemination of plant pathogens, acting as an artificial gene flow mechanism in scenarios where fungal spores may not naturally reach (Finlay 2002). The movement of contaminated seeds and infected plant material, driven by global human trade, can facilitate the migration of pathogens. This artificial gene flow may promote the movement of individuals and their alleles from one population to another, allowing genetic exchange between samples separated continentally,

particularly amongst countries with strong trading ties. On a global scale, human-mediated migration is primarily responsible for the geographic differentiation of numerous populations of plant pathogens (Linde et al. 2002; Stukenbrock et al. 2006; Zaffarano et al. 2006; Barrès et al. 2008 Chakraborty 2013; Goss et al. 2014; Kamvar et al. 2015; McMullan et al. 2018; Tabima et al. 2020; Sotiropoulos et al. 2022).

We detected migration between Argentina and Europe, where isolates from Argentina were grouped within the European lineage. This suggests some form of genetic exchange between these geographically distant regions, facilitated by long-distance gene flow. The role of contaminated seeds as sources of inoculum, promoting the movement of pathogens, has been well established in this pathosystem. In this study, we proposed the hypothesis that winter nurseries in South America, which are frequently employed in breeding programmes, might serve as a potential source for the introduction of the pathogen into new regions. Furthermore, there seems to be a recurring pattern of maize germplasm movement between international breeding companies and experimental stations in Argentina. A case in point is the experimental station "La Josefina S.A", in Mercedes, Argentina, which provided winter nursey service from 1985–2014, collaborating with numerous (10–15 companies/institutes) maize breeding programmes, including European (Murad 2023, personal communication). This station routinely handles maize seeds from international companies working with germplasm from Europe, the USA and Argentina. Consequently, it regularly receives and returns seeds to the original breeding programmes, potentially contributing to the long-distance dissemination of the pathogen.

We identified signatures of genetic introgression at the genome-wide level between lineages, supported by statistically significant recombination events. The examination of nucleotide similarities between sequences uncovered a mosaic-like genome structure, in which different segments exhibit distinct genetic ancestries. The recombination analysis performed with RDP4 revealed a few large recombinant blocks, which contrasts with the results of HYBRIDCHECK which displayed numerous smaller, fragmented blocks. Through visualising nucleotide similarities, we observed a pattern where smaller blocks appeared to be later fragmented by recurrent recombination. Multiple mutations may have accumulated after ancient recombination (i.e. before divergence), giving rise to the observed pattern of fragmented blocks. In essence, these small regions might be remnants of a common ancient introgression event, where subsequent recombination fragmented the originally larger block.

Importantly, subsequent recombination may also increase the genetic divergence between the donor and recipient sequences, making the introgressed regions appear more divergent and hence "older". We are, therefore, cautious in our interpretation of the actual estimated divergence time dates. Nevertheless, we are confident in our interpretation that the introgression events observed in the North American lineage date further back in time than those in the other two lineages (i.e. Europe and Brazil). Our recombination analysis detected events that took place as recently as 6100 years ago, as well as potentially very old events (> 100,000 years ago). These more recent occurrences are likely indicative of genetic exchanges that have occurred in a relatively recent timeframe if not an ongoing process. Overall, our introgression investigation revealed a prolonged history of genetic exchange.

Sexual reproduction has been observed under laboratory conditions, suggesting an absence of prezygotic reproductive barriers between genetically divergent strains (Rogério et al. 2023) as observed in other plant pathogenic ascomycetes (Gac and Giraud 2008). Although the sexual state of this species has not been documented in nature, our data suggest a significant contribution of genetic recombination resulting from either sexual recombination or other non-meiotic recombination mechanisms, such as parasexual events via hyphal anastomosis, as previously described in other *Colletotrichum* species (Vaillancourt et al. 2000; Souza-Paccola et al. 2003; Roca et al. 2003; Rosada et al. 2010; Nordzieke 2022).

We propose that *C. graminicola* populations may have undergone intense recombination before genetic divergence. Repeated sexual recombination cycles over time could have led to the development of a mosaic-like genome structure, as evidenced by our analyses. Ancient hybridisation events have been observed in the wheat powdery mildew pathogen *Blumeria graminis* f. sp. *tritici*, characterised by admixture between divergent strains resulting in a mosaic genome. Chromosomal segments inherited from both parental sources exhibit fragmented segments, likely produced by backcrossing (Menardo et al. 2016; Sotiropoulos et al. 2022). A similar mosaic-like genome structure has been observed in the oomycete *Albugo candida* (McMullan et al. 2015).

Regarding the evolutionary origin of these lineages, our demographic inference indicates that the North American lineage emerged before the Brazilian and European populations. Age estimates of recombinant blocks detected within lineages further support this finding. Dating the blocks involving isolates from North America revealed that this lineage is older compared to those blocks from other lineages. Moreover, the deeper branches in the phylogenetic tree and the higher level of genetic diversity are all consistent with the North American lineage predating the European and Brazilian lineages.

North America seems to act as an intermediate between the source population and the rest of the world, suggesting a potential colonisation route of the fungus. Our results strongly indicate a significant contribution of an unsampled population as the ancestral source. Given the widely accepted hypothesis that maize domestication originated in Mexico (Matsuoka et al. 2002; Sawers and Sanchez Leon 2011) and that the centre of origin of a pathogenic species usually corresponds to that of its host, the unsampled population might indeed represent a Mesoamerican population. However, our sampling did not include isolates from Mesoamerica, a region that could be linked to the ancestral population. To date, attempts to obtain isolates of *C. graminicola* from Mexico have been unsuccessful and personal communications with researchers have revealed that it is not a disease of major importance in the region.

Recent studies have indicated that maize migrated out of Mexico in two waves at different time points (Wade 2023; Yang et al. 2023). These findings suggest that two (or more) ancestral lineages following the maize dispersal route may have contributed to the genetic diversity of this fungal species. After domestication, maize spread along the Pacific coast, initially reaching Central America (Panama) and then extending through South America, particularly the Andean Region of Peru and potentially Brazil (Kistler et al. 2018). Maize then appears to have moved northwards, reaching south-western North America before reaching Europe (Yang et al. 2023). In Europe, maize was initially introduced into Spain from the Caribbean area in 1494 and then spread through the Mediterra-

nean area, Africa and Asia. Additionally, it was also introduced through the European Atlantic coast from North America (Revilla et al. 1998; Revilla et al. 2022). However, owing to the lack of comprehensive fungal sampling from South America, this occurrence may have originated from the Mesoamerican *C. graminicola* lineage first, given the historical trajectory of the maize movement.

Establishing the origin of emerging infectious diseases is both challenging and important. Identifying the ancestral population reveals the standing genetic variation that could introgress into the outbreak lineages. We encourage future studies to comprehensively survey across the Mesoamerica Region, as well as South America, which would further advance our understanding of how this pathogen has adapted and spread to other parts of the world. A genomic comparison between the genetic variation in the source population and population worldwide would enable the examination of genes and genomic regions that experienced strong selection in invasive strains. Such an analysis could shed light on the role of recombination and hybridisation, which become more powerful when there is a greater presence of ancestral variation.

C. graminicola shares several characteristics with another significant maize pathogen, *Setosphaeria turcica*, which is the cause of northern corn leaf blight. Evidence of sexual recombination in *S. turcica* has been reported in a population genetic study by Borchardt et al. (1998). However, the spread of *S. turcica* in Europe results primarily from the expansion of maize cultivation rather than from rapid adaptation. In this instance, *S. turcica*'s spread seems to be due to clonal reproduction (Vidal-Villarejo et al. 2023), in contrast to *C. graminicola* which is supposed to reproduce sexually. The origin of *S. turcica* appears to be in Mexico, from where it later migrated to North America and then to Europe, following a path similar to that of *C. graminicola* (Borchardt et al. 1998; Miedaner and Garbelotto 2024). Breeding for disease resistance becomes more challenging when dealing with pathogens that undergo recombination. Plant breeders must continuously monitor pathogen populations and develop new resistance genes to keep up with the evolving pathogen strains (McDonald and Linde 2002). In *C. graminicola*, genetic diversity is further enhanced by the existence of three major populations and evidence of migration between them. Increased genetic variation resulting from sexual recombination can drive the rapid evolution of pathogen populations. Breeding strategies may need to incorporate multiple resistance genes or use gene rotation to mitigate the risk of resistance breakdown. Addressing sexual recombination in pathogens requires a proactive and continuous strategy for breeding disease-resistant plants. This underscores the importance of understanding pathogen population genetics and implementing robust breeding strategies to ensure long-term crop protection.

Finally, we observed great variability in virulence among the isolates tested, suggesting potential fungal adaptive evolution. A significant, albeit subtle, relationship was found between genetic groups and virulence levels, with the European group exhibiting increased virulence compared with the strain collected during the U.S. outbreak. This level of virulence changed over time, raising concerns about potential new outbreaks of maize anthracnose, particularly in European cultivation. We found that two pairs of isolates with nearly identical genotypes presented marked differences in virulence, emphasising that adaptive evolution may be attributed to factors other than genetic diversity, such as environmental influences or epigenetic modifications. Additionally, it is well

known that minor variations in single genes such as effector encoding genes can have marked effects on virulence. Further studies using these genotypes with contrasting virulence phenotypes may be used to elucidate the genetic basis of virulence in pathogens, contributing to a better understanding of plant diseases and the development of durable control strategies.

Conclusion

We have observed two different gene flow patterns, which are likely the result of the natural spread of the pathogen and the exchange of contaminated seeds. Our findings indicate genetic introgression between lineages, revealing a long history of recombination. We have identified significant recombination events occurring at specific points in time, with the North American lineage showing evidence of the most ancient recombination. According to demographic modelling, North America is positioned between Brazil, Europe and an unsampled ancestral source population, which is believed to be Mesoamerican. Our analyses have shown that the global genomic structure of *C. graminicola* is influenced by geographic differentiation driven by long-distance migration and a long history of recombination and introgression.

Acknowledgements

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Adherence to national and international regulations

All the fungal strains used in this study have been legally obtained, respecting the Convention on Biological Diversity (Rio Convention).

Funding

Author contributions

S.A.S., M.R.T. and F.R. designed the study and designed the project; S.A.S., F.R. and S.G.-S. performed fungal isolation and DNA extraction; F.R., M.R.T., C.V.O., S.D.M., G.K.H. and J.L.V.V. performed the analyses; F.R., M.R.T., C.V.O., S.D.M. and S.A.S. wrote the manuscript; F.R., P.G.R., S.B., S.G.S., F.B.C.F. and S.A.S. performed the pathogenic characterisation; A.G.J., W.B., S.B.U., J.H., R.S., P.R., J.S.D., J.L.V. and I.B. collected the samples. All the authors reviewed the manuscript.

Author ORCIDs

Flávia Rogério https://orcid.org/0000-0001-7801-5112
Cock Van Oosterhout https://orcid.org/0000-0002-5653-738X
Stéphane De Mita https://orcid.org/0000-0003-2752-865X
Francisco Borja Cuevas-Fernández https://orcid.org/0000-0002-6543-5094
Pablo García-Rodríguez https://orcid.org/0000-0001-9730-8412
Sioly Becerra https://orcid.org/0000-0003-2564-1982
Silvia Gutiérrez-Sánchez https://orcid.org/0009-0002-4389-4801
Andrés G. Jacquat https://orcid.org/0000-0001-6811-9604
Wagner Bettiol https://orcid.org/0000-0001-7724-6445
Guilherme Kenichi Hosaka https://orcid.org/0000-0001-8353-1138
Rogelio Santiago https://orcid.org/0000-0001-5036-3975
Pedro Revilla https://orcid.org/0000-0002-4826-6073
José S. Dambolena https://orcid.org/0000-0001-8690-6564
José L. Vicente-Villardón https://orcid.org/0000-0003-1416-6813
Ivica Buhiniček https://orcid.org/0000-0002-7743-8954
Serenella A. Sukno https://orcid.org/0000-0003-3248-6490
Michael R. Thon https://orcid.org/0000-0002-7225-7003

Data availability

The raw Illumina RAD-seq and WGS reads were submitted to GenBank and their accession numbers are listed in Suppl. material 1: table S2.

References

Alaniz S, Cuozzo V, Martínez V, Stadnik MJ, Mondino P (2019) Ascospore infection and *Colletotrichum* species causing *Glomerella* leaf spot of apple in Uruguay. Plant Pathology Journal 35: 100–111. https://doi.org/10.5423/PPJ.OA.07.2018.0145

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology 37: 420–423. https://doi.org/10.1038/s41587-019-0036-z

Anderson E (1953) Introgressive hybridization. Biological Reviews 28: 280–307. https://doi.org/10.1111/j.1469-185X.1953.tb01379.x

Ashu EE, Hagen F, Chowdhary A, Meis JF, Xu J (2017) Global population genetic analysis of *Aspergillus fumigatus*. mSphere 2: e00019-17. https://doi.org/10.1128/mSphere.00019-17

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K V., Altshuler D, Gabriel S, DePristo MA (2013) From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics 43: 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Barcelos QL, Pinto JMA, Vaillancourt LJ, Souza EA (2014) Characterization of *Glomerella* strains recovered from anthracnose lesions on common bean plants in Brazil. PLOS ONE 9: e90910. https://doi.org/10.1371/journal.pone.0090910

Barrès B, Halkett F, Dutech C, Andrieux A, Pinon J, Frey P (2008) Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. Infection, Genetics and Evolution 8: 577–587. https://doi.org/10.1016/j.meegid.2008.04.005

Barton NH (2010) Mutation and the evolution of recombination. Philosophical Transactions of the Royal Society B, Biological Sciences 365: 1281–1294. https://doi.org/10.1098/rstb.2009.0320

Becerra S, Baroncelli R, Boufleur TR, Sukno SA, Thon MR (2023) Chromosome-level analysis of the *Colletotrichum graminicola* genome reveals the unique characteristics of core and minichromosomes. Frontiers in Microbiology 14: 1129319. https://doi.org/10.3389/fmicb.2023.1129319

Beichman AC, Huerta-Sanchez E, Lohmueller KE (2018) Using genomic data to infer historic population dynamics of nonmodel organisms. Annual Review of Ecology, Evolution, and Systematics 49: 433–456. https://doi.org/10.1146/annurev-ecolsys-110617-062431

Borchardt DS, Welz HG, Geiger HH (1998) Genetic structure of *Setosphaeria turcica* populations in tropical and temperate climates. Ecology and Population Biology 88: 322–329. https://doi.org/10.1094/PHYTO.1998.88.4.322

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, Du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Computational Biology 15: e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Bouckaert RR (2010) DensiTree: Making sense of sets of phylogenetic trees. Bioinformatics 26: 1372–1373. https://doi.org/10.1093/bioinformatics/btq110

Bouckaert RR, Drummond AJ (2017) bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evolutionary Biology 17: 1–11. https://doi.org/10.1186/s12862-017-0890-6

Carroll D (2013) Brenner's Encyclopedia of Genetics Genetic Recombination. Second Edi. In: Maloy S, Hughes K (Eds) Academic Press, San Diego, 277–280. https://doi.org/10.1016/B978-0-12-374984-0.00627-6

Casillas S, Barbadilla A (2017) Molecular population genetics. Genetics 205: 1003–1035. https://doi.org/10.1534/genetics.116.196493

Chakraborty S (2013) Migrate or evolve: Options for plant pathogens under climate change. Global Change Biology 19: 1985–2000. https://doi.org/10.1111/gcb.12205

Charlesworth B (2010) Molecular population genomics: A short history. Genetics Research 92: 397–411. https://doi.org/10.1017/S0016672310000522

Collin F-D, Estoup A, Marin J-M, Raynal L (2020) Bringing ABC inference to the machine learning realm: AbcRanger, an optimized random forests library for ABC Bringing ABC

inference to the machine learning realm: AbcRanger, an optimized random forests library for ABC Introduction: challenges for ABC from Population Genetics. In: JOBIM 2020. https://hal.science/hal-02910067v2

Crouch JA, Clarke BB, Hillman BI (2006) Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. Phytopathology 96: 46–60. https://doi.org/10.1094/PHYTO-96-0046

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM (2021) Twelve years of SAMtools and BCFtools. GigaScience 10: giab008. https://doi.org/10.1093/gigascience/giab008

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Diao Y, Zhang C, Xu J, Lin D, Liu L, Mtung'e OG, Liu X (2015) Genetic differentiation and recombination among geographic populations of the fungal pathogen *Colletotrichum truncatum* from chili peppers in China. Evolutionary Applications 8: 108–118. https://doi.org/10.1111/eva.12233

Dobzhansky T (1973) Nothing in biology makes sense except in the light of evolution. American Biology Teacher 35: 125–129. https://doi.org/10.1093/oxfordjournals.jhered.a108767

Drenth A, McTaggart AR, Wingfield BD (2019) Fungal clones win the battle, but recombination wins the war. IMA Fungus 10: 1–6. https://doi.org/10.1186/s43008-019-0020-8

Everhart S, Gambhir N, Stam R (2021) Population genomics of filamentous plant pathogens – A brief overview of research questions, approaches, and pitfalls. Phytopathology 111: 12–22. https://doi.org/10.1094/PHYTO-11-20-0527-FI

Felsenstein J (1974) The evolutionary advantage of recombination. Genetics 78: 737–756. https://doi.org/10.1093/genetics/78.2.737

Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. Science 296: 1061–1063. https://doi.org/10.1126/science.1070710

Gac M Le, Giraud T (2008) Existence of a pattern of reproductive character displacement in *Homobasidiomycota* but not in *Ascomycota*. Journal of Evolutionary Biology 21: 761–772. https://doi.org/10.1111/j.1420-9101.2008.01511.x

Gale LR (2003) A Population Genetic Approach to Variation in *Colletotrichum graminicola*, the Causal Agent of Sorghum Anthracnose. In: Leslie JF (Ed.) Sorghum and Millets Diseases. Iowa State Press, Ames, Iowa, 191–199. https://doi.org/10.1002/9780470384923.ch34

Giraud T, Enjalbert J, Fournier E, Delmotte F, Dutech C (2008) Population genetics of fungal diseases of plants. Parasite 15: 449–454. https://doi.org/10.1051/parasite/2008153449

Gladieux P, Byrnes EJ, Aguileta G, Fisher MC, Heitman J, Giraud T (2011) Epidemiology and Evolution of Fungal Pathogens in Plants and Animals. In: Tibayrenc M (Ed.) Genetics and Evolution of Infectious Diseases. Elsevier, London, 59–132. https://doi.org/10.1016/B978-0-12-384890-1.00004-2

Gompert Z, Buerkle CA (2013) Analyses of genetic ancestry enable key insights for molecular ecology. Molecular Ecology 22: 5278–5294. https://doi.org/10.1111/mec.12488

Goss EM, Tabima JF, Cooke DEL, Restrepo S, Frye WE, Forbes GA, Fieland VJ, Cardenas M, Grünwald NJ (2014) The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. Proceedings of the National Academy of Sciences 111: 8791–8796. https://doi.org/10.1073/pnas.1401884111

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36: 3420–3435. https://doi.org/10.1093/nar/gkn176

Jouet A, McMullan M, Van Oosterhout C (2015) The effects of recombination, mutation and selection on the evolution of the Rp1 resistance genes in grasses. Molecular Ecology 24: 3077–3092. https://doi.org/10.1111/mec.13213

Kamvar ZN, Larsen MM, Kanaskie AM, Hansen EM, Grünwald NJ (2015) Spatial and temporal analysis of populations of the sudden oak death pathogen in Oregon forests. Phytopathology 105: 982–989. https://doi.org/10.1094/PHYTO-12-14-0350-FI

Kasuga T, White TJ, Taylor JW (2002) Estimation of nucleotide substitution rates in eurotiomycete fungi [2]. Molecular Biology and Evolution 19: 2318–2324. https://doi.org/10.1093/oxfordjournals.molbev.a004056

Kistler L, Yoshi Maezumi S, De Souza JG, Przelomska NAS, Costa FM, Smith O, Loiselle H, Ramos-Madrigal J, Wales N, Ribeiro ER, Morrison RR, Grimaldo C, Prous AP, Arriaza B, Thomas P, Gilbert M, De Oliveira Freitas F, Allaby RG (2018) Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. Science 362: 1309–1313. https://doi.org/10.1126/science.aav0207

Koskella B, Giraud T, Hood ME (2006) Pathogen relatedness affects the prevalence of within-host competition. American Naturalist 168: 121–126. https://doi.org/10.1086/505770

Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127–128. https://doi.org/10.1093/bioinformatics/btl529

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Linde CC, Zhan J, Mcdonald BA (2002) Population structure of *Mycosphaerella graminicola*: From lesions to continents. Phytopathology 92: 946–955. https://doi.org/10.1094/PHYTO.2002.92.9.946

López-Villavicencio M, Jonot O, Coantic A, Hood ME, Enjalbert J, Giraud T (2007) Multiple infections by the Anther Smut pathogen are frequent and involve related strains. PLoS Pathogens 3: 1710–1715. https://doi.org/10.1371/journal.ppat.0030176

Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B (2021) RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evolution 7: veaa087. https://doi.org/10.1093/ve/veaa087

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez JG, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences 99: 6080–6084. https://doi.org/10.1073/pnas.052125199

McDermott JM, McDonald BA (1993) Gene flow in plant pathosystems. Annual Review of Phytopathology 31: 353–373. https://doi.org/10.1146/annurev.py.31.090193.002033

McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. Annual Review of Phytopathology 40: 349–379. https://doi.org/10.1146/annurev.phyto.40.120501.101443

McMullan M, Gardiner A, Bailey K, Kemen E, Ward BJ, Cevik V, Robert-Seilaniantz A, Schultz-Larsen T, Balmuth A, Holub E, van Oosterhout C, Jones JDG (2015) Evidence for suppression of immunity as a driver for genomic introgressions and host range expansion in races of *Albugo candida*, a generalist parasite. eLife 4: e04550. https://doi.org/10.7554/eLife.04550

McMullan M, Rafiqi M, Kaithakottil G, Clavijo BJ, Bilham L, Orton E, Percival-Alwyn L, Ward BJ, Edwards A, Saunders DGO, Garcia Accinelli G, Wright J, Verweij W, Koutsovoulos G, Yoshida K, Hosoya T, Williamson L, Jennings P, Ioos R, Husson C, Hietala AM, Vivian-Smith A, Solheim H, Maclean D, Fosker C, Hall N, Brown JKM, Swarbreck D, Blaxter M, Downie JA, Clark MD (2018) The ash dieback invasion of Europe was founded by two genetically divergent individuals. Nature Ecology and Evolution 2: 1000−1008. https://doi.org/10.1038/s41559-018-0548-9

Menardo F, Praz CR, Wyder S, Ben-David R, Bourras S, Matsumae H, McNally KE, Parlange F, Riba A, Roffler S, Schaefer LK, Shimizu KK, Valenti L, Zbinden H, Wicker T, Keller B (2016) Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. Nature Genetics 48: 201−205. https://doi.org/10.1038/ng.3485

Meng JW, Zhu W, He MH, Wu EJ, Duan GH, Xie YK, Jin YJ, Yang LN, Shang LP, Zhan J (2015) Population genetic analysis reveals cryptic sex in the phytopathogenic fungus *Alternaria alternata*. Scientific Reports 5: 1−10. https://doi.org/10.1038/srep18250

Miedaner T, Garbelotto MM (2024) Human-mediated migration of plants, their pathogens and parasites. Journal of Plant Pathology 106: 301−325. https://doi.org/10.1007/s42161-024-01589-0

Milgroom MG (2015) American Phytopathological Society Population Biology of Plant Pathogens: Genetics, Ecology, and Evolution. St. Paul, Minnesota, 399 pp. https://doi.org/10.1007/978-1-61779-501-5_20

Nordzieke DE (2022) Hyphal fusions enable efficient nutrient distribution in *Colletotrichum graminicola* conidiation and symptom development on maize. Microorganisms 10: 1146. https://doi.org/10.3390/microorganisms10061146

Nordzieke DE, Sanken A, Antelo L, Raschke A, Deising HB, Pöggeler S (2019) Specialized infection strategies of falcate and oval conidia of *Colletotrichum graminicola*. Fungal Genetics and Biology 133: 103276. https://doi.org/10.1016/j.fgb.2019.103276

O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiate EA, Epstein L, Alkan N, Altmüller J, Alvarado-Balderrama L, Bauser CA, Becker C, Birren BW, Chen Z, Choi J, Crouch JA, Duvick JP, Farman MA, Gan P, Heiman D, Henrissat B, Howard RJ, Kabbage M, Koch C, Kracher B, Kubo Y, Law AD, Lebrun MH, Lee YH, Miyara I, Moore N, Neumann U, Nordström K, Panaccione DG, Panstruga R, Place M, Proctor RH, Prusky D, Rech G, Reinhardt R, Rollins JA, Rounsley S, Schardl CL, Schwartz DC, Shenoy N, Shirasu K, Sikhakolli UR, Stüber K, Sukno SA, Sweigard JA, Takano Y, Takahara H, Trail F, Van Der Does HC, Voll LM, Will I, Young S, Zeng Q, Zhang J, Zhou S, Dickman MB, Schulze-Lefert P, Ver Loren Van Themaat E, Ma LJ, Vaillancourt LJ (2012) Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nature Genetics 44: 1060−1065. https://doi.org/10.1038/ng.2372

Van Oosterhout C (2021) Mitigating the threat of emerging infectious diseases; a coevolutionary perspective. Virulence 12: 1288−1295. https://doi.org/10.1080/21505594.2021.1920741

Perfect SE, Hughes HB, O'Connell RJ, Green JR (1999) *Colletotrichum*: A model genus for studies on pathology and fungal-plant interactions. Fungal Genetics and Biology 27: 186−198. https://doi.org/10.1006/fgbi.1999.1143

Pertea G, Pertea M (2020) GFF Utilities: GffRead and GffCompare. F1000Research 9: 304. https://doi.org/10.12688/f1000research.23297.2

Revilla P, Soengas P, Malvar RA, Cartea ME, Ordás A (1998) Isozyme variation and historical relationships among the maize races of Spain. Maydica 43.

Revilla P, Alves ML, Andelković V, Balconi C, Dinis I, Mendes-Moreira P, Redaelli R, Ruiz de Galarreta JI, Vaz Patto MC, Žilić S, Malvar RA (2022) Traditional foods from maize (*Zea mays* L.) in Europe. Frontiers in Nutrition 8: 683399. https://doi.org/10.3389/fnut.2021.683399

Roca MG, Read ND, Wheals AE (2005) Conidial anastomosis tubes in filamentous fungi. FEMS Microbiology Letters 249: 191–198. https://doi.org/10.1016/j.femsle.2005.06.048

Roca MG, Davide LC, Mendes-Costa MC, Wheals A (2003) Conidial anastomosis tubes in *Colletotrichum*. Fungal Genetics and Biology 40: 138–145. https://doi.org/10.1016/S1087-1845(03)00088-4

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Molecular Ecology 28: 4737–4754. https://doi.org/10.1111/mec.15253

Rogério F, Van Oosterhout C, Ciampi-Guillardi M, Correr FH, Hosaka GH, Cros-Arteil S, Margarido GRA, Massola Júnior NS, Gladieux P (2022) Means, motive, and opportunity for biological invasions: genetic introgression in a fungal pathogen. Molecular Ecology 32: 2428–2442. https://doi.org/10.1111/mec.16366

Rogério F, Baroncelli R, Cuevas-Fernández FB, Becerra S, Crouch JA, Bettiol W, Azcárate-Peril MA, Malapi-Wight M, Ortega V, Betran J, Tenuta A, Dambolena JS, Esker PD, Revilla P, Jackson-Ziems TA, Hiltbrunner J, Munkvold G, Buhiniček I, Vicente-Villardón JL, Sukno SA, Thon MR (2023) Population genomics provide insights into the global genetic structure of *Colletotrichum graminicola*, the causal agent of maize Anthracnose. mBio 14: e02878-22. https://doi.org/10.1128/mbio.02878-22

Rosada LJ, Franco CCS, Sant'anna JR, Kaneshima EN, Gonçalves-Vidigal MC, Castro-Prado MAA (2010) Parasexuality in Race 65 *Colletotrichum lindemuthianum* Isolates. Journal of Eukaryotic 57: 383–384. https://doi.org/10.1111/j.1550-7408.2010.00486.x

Rosenzweig BK, Pease JB, Besansky NJ, Hahn MW (2016) Powerful methods for detecting introgressed regions from population genomic data. Molecular Ecology 25: 2387–2397. https://doi.org/10.1111/mec.13610

Sawers RJH, Sanchez Leon NL (2011) Origins of maize: A further paradox resolved. Frontiers in Genetics 2: 53. https://doi.org/10.3389/fgene.2011.00053

Siol M, Coudoux T, Ravel S, De Mita S (2022) EggLib 3: A python package for population genetics and genomics. Molecular Ecology Resources 22: 3176–3187. https://doi.org/10.1111/1755-0998.13672

Sotiropoulos AG, Arango-Isaza E, Ban T, Barbieri C, Bourras S, Cowger C, Czembor PC, Ben-David R, Dinoor A, Ellwood SR, Graf J, Hatta K, Helguera M, Sánchez-Martín J, McDonald BA, Morgounov AI, Müller MC, Shamanin V, Shimizu KK, Yoshihira T, Zbinden H, Keller B, Wicker T (2022) Global genomic analyses of wheat powdery mildew reveal association of pathogen spread with historical human migration and trade. Nature Communications 13: 1–14. https://doi.org/10.1038/s41467-022-31975-0

Souza-Paccola EA, Fávaro LCL, Casela CR, Paccola-Meirelles LD (2003) Genetic recombination in *Colletotrichum sublineolum*. Journal of Phytopathology 151: 329–334. https://doi.org/10.1046/j.1439-0434.2003.00727.x

Steinrücken M, Kamm J, Spence JP, Song YS (2019) Inference of complex population histories using whole-genome sequences from multiple populations. Proceedings

of the National Academy of Sciences 116: 17115–17120. https://doi.org/10.1073/pnas.1905060116

Stukenbrock EH (2016) The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology 106: 104–112. https://doi.org/10.1094/PHYTO-08-15-0184-RVW

Stukenbrock EH, Bataillon T (2012) A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. PLOS Pathogens 8: 1–4. https://doi.org/10.1371/journal.ppat.1002893

Stukenbrock EH, Dutheil JY (2018) Fine-scale recombination maps of fungal plant pathogens reveal dynamic recombination landscapes and intragenic hotspots. Genetics 208: 1209–1229. https://doi.org/10.1534/genetics.117.300502

Stukenbrock EH, Banke S, McDonald BA (2006) Global migration patterns in the fungal wheat pathogen *Phaeosphaeria nodorum*. Molecular Ecology 15: 2895–2904. https://doi.org/10.1111/j.1365-294X.2006.02986.x

Tabima JF, Søndreli KL, Keriö S, Feau N, Sakalidis ML, Hamelin RC, LeBoldus JM (2020) Population genomic analyses reveal connectivity via human-mediated transport across populus plantations in north america and an undescribed subpopulation of *Sphaerulina musiva*. Molecular Plant-Microbe Interactions 33: 189–199. https://doi.org/10.1094/MPMI-05-19-0131-R

Taylor J, Jacobson D, Fisher M (1999) THE EVOLUTION OF ASEXUAL FUNGI: Reproduction, specification and classification. Annual Review of Phytopathology 37: 197–246. https://doi.org/10.1146/annurev.phyto.37.1.197

Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE (2015) Clonal reproduction in fungi. Proceedings of the National Academy of Sciences 112: 8901–8908. https://doi.org/10.1073/pnas.1503159112

Vaillancourt L, Wang J, Hanau R, Rollins J, Du M (2000) Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. Mycologia 92: 430–435. https://doi.org/10.2307/3761501

Vargas WA, Sanz Martín JM, Rech GE, Rivera LP, Benito EP, Díaz-Mínguez JM, Thon MR, Sukno SA (2012) Plant defense mechanisms aAe activated during biotrophic and necrotrophic development of *Colletotricum graminicola* in maize. Plant Physiology 158: 1342–1358. https://doi.org/10.1104/pp.111.190397

Vidal-Villarejo M, Freund F, Hanekamp H, Von Tiedemann A, Schmid K (2023) Population genomic evidence for a repeated introduction and rapid expansion of the fungal maize pathogen *Setosphaeria turcica* in Europe. Genome Biology and Evolution 15: evad130. https://doi.org/10.1093/gbe/evad130

de Vries S, Stukenbrock EH, Rose LE (2020) Rapid evolution in plant−microbe interactions − an evolutionary genomics perspective. New Phytologist 226: 1256–1262. https://doi.org/10.1111/nph.16458

Wade L (2023) Maize has an unexpected wild ancestor. Science 382. https://doi.org/10.1126/science.adn2113

Ward BJ, van Oosterhout C (2016) HYBRIDCHECK: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. Molecular Ecology Resources 16: 534–539. https://doi.org/10.1111/1755-0998.12469

Yang N, Wang Y, Liu X, Jin M, Vallebueno-Estrada M, Calfee E, Chen L, Dilkes BP, Gui S, Fan X, Harper TK, Kennett DJ, Li W, Lu Y, Ding J, Chen Z, Luo J, Mambakkam S, Menon M, Snodgrass S, Veller C, Wu S, Wu S, Zhuo L, Xiao Y, Yang X, Stitzer MC, Runcie D, Yan J, Ross-Ibarr J (2023) Two teosintes made modern maize. Science 382. https://doi.org/10.1126/science.adg8940

Zaffarano PL, McDonald BA, Zala M, Linde CC (2006) Global hierarchical gene diversity analysis suggests the fertile crescent is not the center of origin of the barley scald pathogen *Rhynchosporium secalis*. Phytopathology 96: 941−950. https://doi.org/10.1094/PHYTO-96-0941

Zhan J, McDonald BA (2013) Experimental measures of pathogen competition and relative fitness. Annual Review of Phytopathology 51: 131−153. https://doi.org/10.1146/annurev-phyto-082712-102302

Zhan J, Thrall PH, Papaïx J, Xie L, Burdon JJ (2015) Playing on a pathogen's weakness: Using evolution to guide sustainable plant disease control strategies. Annual Review of Phytopathology 53: 19−43. https://doi.org/10.1146/annurev-phyto-080614-120040

Supplementary material 1

Supplementary tables

Authors: Flávia Rogério

Data type: xlsx

Explanation note: **table S1:** Isolates of *Colletotrichum graminicola* used in this study; **table S2:** Repeated multilocus genotypes removed from the dataset with 212 Colletotrichum graminicola isolates; **table S3:** Summary of genomic diversity within Colletotrichum graminicola lineages; **table S4:** Membership proportion of each isolate for *Colletotrichum graminicola* genetic lineages considering number of cluster (*K*) equal to 3; **table S5:** List of the recombination blocks detected in the RDP4 software under chromossome 1. Table shows the assigned event number, the recombinant and minor parent, the predicted start and end positions each block relative to the referenced contig, and the p values for the different recombination detection methods; **table S6:** Estimated dates of introgression events from triplet1 (BR-85955-2; P-7565-072-8; I-61851), triplet 2 (M5.001; CR-34543-1; A-52621-1), triplet 3 (M5.001: BR-98290-1: BR-85925-1) , triplet 4 (F-64330-2: F-64330-7: F-64330-13), and triplet 5 (NRRL13649: I-61851: CA-CHAT-1) obtained by HYBRIDCHECK algorithm dating; **table S7:** Estimated dates of recombinant blocks detected by HYBRIDCHECK using BEAST software; **table S8:** Description of summary statistics employed on ABC demographic inference; **table S9:** Description of the demographic models tested; **table S10:** Classification of isolate in each genetic group into three categories based on their virulence compared to the control isolate M1.001. Values represent the percentage of isolates in each category; **table S11:** Funtional annotation of genes in the introgression regions detected between Colletotrichum graminicola lineages; **table S11:** Funtional annotation of genes in the introgression regions detected between Colletotrichum graminicola lineages; **table S13:** Untransformed virulence data.

Link: https://doi.org/10.3897/imafungus.16.138888.suppl1

Supplementary material 2

Supplementary methods

Authors: Flávia Rogério

Data type: docx

Link: https://doi.org/10.3897/imafungus.16.138888.suppl2

Supplementary material 3

Supplementary figures

Authors: Flávia Rogério

Data type: docx

Explanation note: **figure S1:** Manhattan plot of the distribution of the SNPs along the chromosomes; **figure S2:** Population subdivision of *Colletotrichum graminicola*; **figure S3:** Neighbor-net network showing relationships between isolates of *Colletotrichum graminicola* identified based on the clone-correct dataset; **figure S4:** Population subdivision on European lineage of *Colletotrichum graminicola*; **figure S5:** Densitree cloudogram based on 203 samples across chromosomes; **figure S6:** Isolation-by-distance plot showing the correlation between pairwise genetic dissimilarity and geographical distance (in km); **figure S7:** Recombination analysis across chromosomes for isolates M5.001:CR-34543-1:A-52621- and BR-85955-2:P-7565-072-8: I-61851; **figure S8:** Recombination analysis of chromosome 1 for isolates M5.001:BR-98290-1:BR-85925-1 (triplet 3), F-64330-2:F-64330-7:F-64330-13 (triplet 4) and NRRL13649:I-61851:CA-CHAT-1 (triplet 5); **figure S9:** Age distribution of recombinant blocks between lineages; **figure S10:** Demographic models tested; **figure S11:** The fit of the data used in the ABC analysis; **figure S12:** Results of model choice from ABC analysis; **figure S13:** Pathogenic characterization of *Colletotrichum graminicola* isolates; **figure S14:** Bar plot showing the distribution of the virulence for the three genetic lineages.

Link: https://doi.org/10.3897/imafungus.16.138888.suppl3