Means, motive, and opportunity for biological invasions: genetic introgression in a fungal pathogen

FLÁVIA ROGÉRIO,[1,2] COCK VAN OOSTERHOUT,[3] MAISA CIAMPI-GUILLARDI,[1] FERNANDO HENRIQUE CORRER,[4] GUILHERME KENICHI HOSAKA,[4] SANDRINE CROS-ARTEIL,[5] GABRIEL RODRIGUES ALVES MARGARIDO,[4] NELSON S. MASSOLA JÚNIOR,[1] and PIERRE GLADIEUX[5]

[1]Department of Plant Pathology and Nematology, University of São Paulo, Piracicaba, SP, Brazil, [2]Institute for Agribiotechnology Research (CIALE), University of Salamanca, Salamanca, Spain, [3]School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK, [4]Department of Genetics, University of São Paulo, Piracicaba, SP, Brazil, [5]UMR PHIM, University of Montpellier, INRAE, CIRAD, Montpellier SupAgro, Montpellier, France

correspondence
Flávia Rogério
E-mail: flavia.rogerio@usal.es

ABSTRACT

Invasions by fungal plant pathogens pose a significant threat to the health of agriculture ecosystems. Despite limited standing genetic variation, many invasive fungal species can adapt and spread rapidly, resulting in significant losses in crop yields. Here, we report on the population genomics of *Colletotrichum truncatum*, a polyphagous pathogen that can infect more than 460 plant species, and an invasive pathogen on soybean in Brazil. We study the whole-genome sequences of 18 isolates representing 10 fields from two major regions of soybean production. We show that Brazilian *C. truncatum* is subdivided into three phylogenetically distinct lineages that exchange genetic variation through hybridization. Introgression affects 2 to 30% of the nucleotides of genomes and varies widely between the lineages. We find that introgressed regions comprise secreted

protein-encoding genes, suggesting possible co-evolutionary targets for selection in those regions. We highlight the inherent vulnerability of genetically uniform crops in the agro-ecological environment, particularly when faced with pathogens that can take full advantage of the opportunities offered by an increasingly globalized world. Finally, we discuss "The Means, Motive, and Opportunity" of fungal pathogens and how they can become invasive species of crops. We call for more population genomic studies because such analyses can help identify geographic areas and pathogens that pose a risk, thereby helping to inform control strategies to better protect crops in the future.

INTRODUCTION

Understanding the eco-evolutionary and human-associated factors underlying the emergence and spread of fungal plant diseases is essential to the implementation of effective control measures (Hessenauer et al., 2020; Stukenbrock & McDonald, 2008). Population genetics and molecular epidemiology can shed light on both the extrinsic and intrinsic drivers of biological invasions by fungal plant pathogens (Gladieux et al., 2015; Grünwald, McDonald, & Milgroom, 2016). In parallel, molecular ecology and evolution can provide insights into how an invasive species with limited genetic variation can evolve into ecologically successful pest species soon after a founder event.

Fungi can rapidly evolve into devastating invasive species, despite their often low genetic diversity (Ali et al., 2014; De Jonge et al., 2013; Gladieux et al., 2018; Latorre et al., 2020; Stauber, Badet, Prospero, & Croll, 2020). Although this is a characteristic that they share with many other invasive species, some unique aspects of fungal population biology may facilitate rapid evolutionary changes and enhance their invasive potential. Despite regular demographic bottlenecks (e.g.,

during winters), fungal populations generally have periods of huge census size, which can substantial evolutionary changes (Gladieux et al., 2015).

Fungal reproduction is complex and can include both sexual and asexual stages (Alexopoulos, Mims, & Blackwell, 1996). Asexual reproduction can be accomplished through mitotic spores (conidia), which augments the propagule pressure during invasion in non-native habitats. In addition, fungi produce filamentous structures, i.e., hyphae, which provide a second mode of asexual reproduction through mycelial fragmentation. Complementing these asexual modes, during sexual reproduction, compatible individuals may exchange genetic information during plasmogamy and karyogamy between gametes (Alexopoulos et al., 1996). Meiotic spores promote the creation of novel genotypes through recombination, and they serve as dispersal and survival structures (Taylor, Jacobson, & Fisher, 1999). If compatible gametes are derived from genetically diverged lineages, the resulting genetic exchange can lead to genetic introgression.

The large census size of invasive fungal populations enables rapid adaptations to new varieties of resistant plants or antifungals molecules (Barton, 2010; Gladieux et al., 2015). This may be particularly important for fungal populations found on major widespread crops, which due to their vast population sizes, benefit from a high input of novel variation by mutations. These new pathogen genotypes can rapidly spread through genetically uniform host populations. In other words, the high evolvability of some fungi during biological invasions is not necessarily realized through their high standing genetic variation at the point of entry (cf. Fisher's Fundamental Theorem (Price, 1972)). Rather, many fungi are highly potent biological invaders due to input of *de novo* allelic and genotypic variation every generation, which is a consequence of their high potential for gene flow,

recombination and mutation. These drivers of genetic variation play a crucial role in the co-evolutionary arms race between fungal pathogens and their hosts. A sudden increase in these drivers can shift the co-evolutionary balance, and these effects can be particularly severe in an eco-agriculture setting with genetically relatively uniform host plants and animals (Van Oosterhout, 2021).

Multiple populations of the same fungal species can coexist on one host whilst competing for limited resources (Bueno-Sancho et al., 2017; Fournier, Gladieux, & Giraud, 2013; Hartmann, Mcdonald, & Croll, 2018; Hubbard et al., 2015; Persoons et al., 2017; Silva, Várzea, Paulo, & Batista, 2018; Stauber et al., 2020; Thierry et al., 2020; Vieira, Silva, Várzea, Paulo, & Batista, 2018). In the absence of temporal, spatial or habitat barriers, coexistence on the same host may foster genetic exchanges between fungal lineages. If coexisting populations represent previously geographically isolated lineages that have not evolved strong pre- or postzygotic barriers, such introgression can rapidly generate novel genotypic variation. In turn, this can increase the amount of phenotype variation – a phenomenon known as transgressive segregation – which provides more novel substrate for natural selection (Nichols et al., 2015).

Admixture between multiple coexisting populations can also lead to a so-called bridgehead effect, in which highly adapted lineages emerge through recombination among propagules established in an area of first introduction (Bertelsmeier & Keller, 2018; Dutech et al., 2012; Stauber et al., 2020). Ongoing fungal invasions offer a unique opportunity to learn about the ecology and evolution of biotic interactions in human-altered ecosystems (Gladieux et al., 2014; Parker & Gilbert, 2018; Thrall, Hochberg, Burdon, & Bever, 2007; Thrall et al.,

2011), and such studies are important to assess the risks posed by pathogens to crop in agriculture.

Anthracnose, mainly associated with the fungus *Colletotrichum truncatum* (Hyde et al., 2009), is one of the most prominent foliar diseases of soybean. This ascomycete is seed-transmitted and can infect more than 460 plant species, including important crops in the Fabaceae and Solanaceae families (Cannon, Damm, Johnston, & Weir, 2012; Damm, Woudenberg, Cannon, & Crous, 2009; Weidemann, TeBeest, & Cartwright, 1988). In Brazil, the worldwide leader in soybean production, previous population genetic studies showed that *C. truncatum* is a recently introduced invasive species structured into three highly divergent clusters coexisting in soybean fields. This suggests there have been multiple introductions from distinct source populations, which are yet to be identified (Rogério, Gladieux, Massola, & Ciampi-Guillardi, 2019).

Colletotrichum genus are predominantly observed in the asexual phase, and the sexual morph is rarely identified for most species (Barcelos, Pinto, Vaillancourt, & Souza, 2014; Talhinhas & Baroncelli, 2021). Mating system of this group does not follow the usual bipolar self-incompatibility MAT1-1/2 system (Crouch et al., 2014), typically found in ascomycetes. All species that have been studied to date have only the single mating type MAT1-2, however, the loss of one idiomorph does not affected the sexual capabilities of the species, which has led to the development of hypotheses to describe this singular mating behavior (Menat, Banniza, Cabral, Vijayan, & Wei, 2012; Vaillancourt et al., 2000; Wilson, Lelwala, Taylor, Wingfield, & Wingfield, 2021).

C. truncatum is a homothallic species what mean that are capable of sexual reproduce without a mating partner. Although sexual reproduction has not been

described until date to this species, alternative sexual mechanisms are described to other species (Roca, Davide, Mendes-Costa, & Wheals, 2003; Rosada et al., 2010; Souza-Paccola, Fávaro, Casela, & Paccola-Meirelles, 2003; Vaillancourt et al., 2000), and such processes also may be play role in *C. truncatum* reproduction. Some studies reported population genetic signatures of sexual recombination in *C. truncatum* (Diao et al., 2015; Katoch, Sharma, Padder, & Sharma, 2017; (Diao et al., 2015; Katoch, Sharma, Padder, & Sharma, 2017; Rogério, Gladieux, Massola, & Ciampi-Guillardi, 2019) signaling that recombination events may be occurring and playing an important role in the genetic diversity of this pathogen.

Here, we use whole-genome sequence and a population genomics approach to characterize the genetic makeup and infer the evolutionary history of *C. truncatum* causing soybean anthracnose in Brazil. We document evidence of extensive introgression between three sympatric lineages that have invaded Brazilian soybean. Our study highlights the risk that Brazilian *C. truncatum* may represent as a bridgehead for future invasions of soybean-producing areas, facilitating admixture between the three lineages (as well as with any unsampled lineages and possible future immigrant lineages). We discuss why fungi have the means to become potent biological invaders of crops, arguing this is due to their ability to rapidly generate novel genetic variation, in combination with their high propagule pressure accomplished through two modes of asexual reproduction. We furthermore discuss why the large biomass of relative uniform crops provides the motive, and the bridgehead populations that enable genetic introgression the opportunity for such biological invasions.

MATERIALS AND METHODS

Fungal isolates, DNA extraction and genome sequencing

We used 18 isolates of *Colletotrichum truncatum* from naturally infected commercial soybean fields from Mato Grosso (MT) and Goiás (GO) states in Brazil. Isolates coexisting in the fields sampled were randomly selected from the three genetic clusters (C1, C2, and C3) previously identified based on the population genetics analysis of microsatellite variation (Rogério, Gladieux, Massola, & Ciampi-Guillardi, 2019) (Table 1). Information about isolation and culture collection can be found in the study aforementioned. Fungal genomic DNA was extracted using Wizard Genomic DNA Purification kit (Promega) from fresh mycelium grown on potato dextrose liquid medium (Difco). Paired-end libraries were prepared and sequenced on Illumina HiSeq2000 (2x150 bp, insert size ~ 500 pb) by Genewiz (South Plainfield, USA). The raw reads were deposited at the NCBI/Genbank under the Sequence Read Archive (SRA) accession numbers SAMN13196067 to SAMN13196084 (Table S1).

Read mapping and SNP calling

Read quality was checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw Illumina reads were trimmed for adapter contamination and bases with a Phred score smaller than 30 were removed using Cutadapt v1.16 software (M. Martin, 2011). Reads were mapped against the reference genome of the previously assembled isolate CMES1059 (Rogério et al., 2020) using BWA-mem v0.7.15 (options -n=5) (Li & Durbin, 2009). Alignments were sorted with Samtools v1.3 (Li et al., 2009), and reads with mapping quality below 30 were removed. Duplicates reads were removed using Picard v2.7 (http://broadinstitute.github.io/picard/). Single nucleotide polymorphisms (SNPs) and indels were called using the HaplotypeCaller module from the Genome Analysis

Toolkit v4.0.12 (GATK) (McKenna et al., 2010), with the option *–emitRefConfidence* GVCF. The gVCF files listing variants were merged using COMBINEGVCFS and genotyped using GENOTYPEGVCFS. Monomorphic sites were included using the argument *include_nonvariantsites*. High confidence SNPs were identified using GATK's VARIANTFILTRATION module, following GATK's best practices (http://www.broadinstitute.org/gatk/guide/best_practices), with parameters: QD < 2.0 (Variant Quality), FS > 60.0 (Phred score Fisher's test), MQ < 40.0 (Mapping Quality), MQRankSum < 12.5 (Mapping Quality of Reference reads vs alternative reads) and ReadPosRankSum < 8.0 (Distance of alternative read from the end of the reads).

Population structure

We used SPLITSTREE v4 (Huson & Bryant, 2006) to visualize relationships between isolates in a phylogenetic network based on pseudoassembled genomic sequences (SNPs in vcf format were converted into a fasta file using the reference sequence as a template) generated from the tables of SNPs with the reference sequence as a template. We also used the pairwise homoplasy index (PHI) test implemented in SPLITSTREE to test the null hypothesis of clonality; recombination is expected to result in exchangeable sites within lineages.

Population structure was analyzed using methods optimized for the analysis of large datasets and which do not assume Hardy-Weinberg equilibrium. We performed principal components (PCA) and discriminant analysis of principal components (DAPC) in R with the package ADEGENET v2.0 (Jombart & Ahmed, 2011), using the FIND.CLUSTERS function. DAPC is a non-model-based method using PCA as a prior step, which provides a description of clusters using discriminant functions. We retained the first 20 principal components. DAPC identifies an optimal number of genetic clusters

that best describes the data by running a k-means clustering algorithm and comparing the different clustering solutions using the Bayesian Information Criterion (BIC). Population structure was also analyzed using the software sNMF (Frichot, Mathieu, Trouillon, Bouchard, & François, 2014) which estimates ancestry coefficients based on sparse non-negative matrix factorization and least squares optimization. We calculated ancestry coefficients for 2 to 10 ancestral populations (*K*) using 100 replicates for each *K*. The preferred number of *K* was chosen using a cross-entropy criterion based on the prediction of masked haplotypes to evaluate the error of ancestry estimation. All clustering analyses were based on biallelic SNPs without missing data.

Diversity and divergence

Polymorphism and divergence statistics were computed in 10 kb non-overlapping windows using the SCIKIT-ALLEL v1.0.2 Python package (Miles & Harding, 2017). Summary statistics were plotted for the 10 largest contig using CIRCOS v0.67 software (Connors et al., 2009).

Recombination analyses

Recombination events were analyzed using the software RDP4 (Martin, Murrell, Golden, Khoosal, & Muhire, 2015) implementing seven independent detection algorithms: RDP (Martin & Rybicki, 2000), GENECONV (Padidam, Sawyer, & Fauquet, 1999), BOOTSCAN (Martin, Posada, Crandall, & Williamson, 2005), MAXCHI (Smith, 1992) CHIMAERA (Posada & Crandall, 2002), SISCAN (Gibbs, Armstrong, & Gibbs, 2000) and 3SEQ (Boni, Posada, & Feldman, 2007). Whole sequences of the ten largest contigs were scanned using the default settings for the window size. Tests were conducted using a critical value $\alpha = 0.05$ and p-values were Bonferroni corrected for

multiple comparisons of sequences. The evidence for a recombination signal was considered to be strong if it was found to be significant with three or more detection methods. Only events for which the software identified the parental sequences (i.e., no 'unknowns') without ambiguous start and end position of the recombination block were considered.

We used POPLDDECAY version v3.4 (Zhang, Dong, Xu, He, & Yang, 2019) to investigate the patterns of linkage disequilibrium decay within *C. truncatum* genetic groups as coefficient of linkage disequilibrium (r^2) (Hill & Robertson, 1968) calculated for all pairs of SNPs less than 300 kb apart. For this, we used biallelic SNPs, excluding missing data and sites with minor allele frequencies below 10%.

Genome scan for signature of genetic exchanges

We used CHROMOPAINTER v0.0.4 (Lawson, Hellenthal, Myers, & Falush, 2012) for probabilistic chromosome painting to infer recent shared ancestry between *C. truncatum* lineages. This method "paints" individuals in "recipient" populations as a combination of segments from "donor" populations, using linkage disequilibrium information for probability computation and assuming that linked alleles are more likely to be exchanged together during recombination events. We ran three separate analyses, each considering one particular *C. truncatum* lineage as a collection of haplotypes to be painted, and all lineages as donors. The recombination scaling constant N_e and emission probabilities (μ) were calculated as averages weighted by contigs' length determined by LDHAT (Auton & McVean, 2007). Estimates of these parameters for each lineage were obtained by running the expectation maximization algorithm with 200 iterations. These analyses were based on biallelic SNPs dataset without missing data.

Fine-scale admixture between *C. truncatum* lineages also was analyzed using the software HYBRIDCHECK (Ward & van Oosterhout, 2016), which uses a sliding window to scan for sudden changes in nucleotide divergence between sequences, thus identifying potential genetic exchanges where nucleotide divergence is significantly lower. The similarities were visualized through a plot employing the primary colors red, green, and blue, using the 100 bp windows based on the proportion of SNPs shared between the pairwise sequences, with a stepwise increment of 1 bp. In cases where all SNPs are shared between just two of the three lineages, the hybrid color is an exact 50% mix of two primary colors. Hence, yellow, purple, and turquoise colors pinpoint regions of possible recent genetic exchange between two sequences. We carried out this analysis on a triplet involving one isolate representative of each lineage: isolates LFN0297 (lineage C1), LFN0318 (lineage C2), and LFN0308 (lineage C3) for the ten largest contigs.

HYBRIDCHECK and CHROMOPAINTER identify genomic regions of shared ancestry, and such signal can be caused either by genetic introgression, or by incomplete lineage sorting (Durand, Patterson, Reich, & Slatkin, 2011). In order to differentiate between genetic introgression and incomplete lineage sorting, we also used HYBRIDCHECK to estimate the age of recombinant regions. If the genomic region coalescence before the split of the species (or lineages), the signal is consistent with incomplete lineage sorting. However, if the coalescence event is dated after the speciation event (or after the bifurcation of the lineages in the tree), the genetic exchange has occurred after the divergence. In the latter case, the signal is consistent with genetic introgression after hybridization (Jouet, McMullan, & Van Oosterhout, 2015). Recombination blocks were then dated assuming a strict molecular clock with a mutation rate of 10^{-8} per generation, assuming a generation time of one year.

Functional enrichment

To characterize genes present in the introgressed regions between *C. truncatum* lineages, we extracted the transcripts and proteins from the corresponding regions in the reference genome (Rogério et al., 2020) using GFFREAD (Pertea & Pertea, 2020) using the ten largest contigs as reference background. For the enrichment analysis the GO terms were assigned from re-annotated transcripts using BLAST2GO (Conesa et al., 2005) against the NCBI non-redundant database and used as input to the TOPGO R package (Alexa & Rahnenfuhrer 2021). We used SIGNALP v5.0 (Armenteros et al., 2019) to identify secreted proteins.

Demographic inferences

To infer the evolutionary history of the genetic lineages we used the Python package DADI (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009). The method implemented in DADI infers demographic parameters based on a diffusion approximation to the site frequency spectrum (SFS). The Python script EASYSFS.PY (available at https://github.com/isaacovercast/easySFS) was used to convert the VCF file into a three-dimensional joint site frequency spectrum (3D-JSFS). The SFS was folded because no appropriate outgroup was available. We compared twelve demographic models including strict isolation, isolation with migration (asymmetrical migration rates), and isolation with population size changes, with four possible topologies, using the demographic modeling workflow (*dadi_pipeline*) from Portik et al. 2017 (Fig. S1). For each model, we performed four rounds of optimizations; for each round, we ran multiple replicates and used parameter estimates from the best scoring replicate (highest log-likelihood) to seed searches in the following round. We used the default settings in *dadi_pipeline* for

each round (replicates = 10, 20, 30, 40; maxiter = 3, 5, 10, 15; fold = 3, 2, 2, 1), and optimized parameters using the Nelder-Mead method (*optimize_log_fmin*). We used the optimized parameter sets of each replicate to simulate the 3D-JSFS, and the multinomial approach was used to estimate the log-likelihood of the given the model. We assessed the model's goodness-of-fit by maximizing the model likelihood and visual inspection of the residuals between the site frequency spectra generated by the inferred model and the real data. Conventional bootstrapping (100 replicates) was carried out to estimate parameter uncertainties, by dividing data into 100 independent chunks of 1 Mb in which SNPs were sampled with replacement.

RESULTS

Population structure and levels of genetic variation

Read mapping and variant calling for 18 isolates of *C. truncatum* submitted to whole-genome sequencing identified 2,220,191 biallelic Single Nucleotide Polymorphisms (SNPs), distributed across 128 contigs (see Table S1 to sequencing statistics). To assess population subdivision and to visualize relationships among isolates we built a neighbor-net network with SPLITSTREE based on the full set of SNPs. This phylogenetic network revealed three groups, henceforth referred to as "lineages" C1, C2 and C3 (Fig. 1A). Lineage C1 was connected to the rest of the dataset by a long, non-reticulated branch consistent with relatively long-term genetic isolation. Lineage 2 and 3 were connected by branches showing extensive reticulations (looping in the network) indicating a history of recombination or incomplete lineage sorting (Fig. 1A). In a Discriminant Analysis of Principal Component (DAPC) modelling K=2 to K=10 populations, the Bayesian Information Criterion monotonously decreased with increasing K, preventing clear choice of a best supported model, but the composition of

clusters identified at *K*=3 matched what was observed in the neighbor-net network (Fig. 1B and Fig. S2). Likewise clustering by sparse non-negative matrix factorization algorithms, as implemented in the sNMF method, identified $K = 3$ as the best supported model based on cross-entropy. The ancestry coefficients estimated with sNMF revealed essentially the same pattern of population subdivision as the DAPC and neighbor-net network. However, three isolates (LFN0318, LFN0217, and LFN0150) shared ancestry in two clusters, suggesting admixture between lineages (Fig.1C).

Nucleotide diversity was nearly twice as high in C3 than in C2, and it was more than one order of magnitude higher in C2 than in C1 (C3: π=0.0113/bp; C2: π=0.0062/bp; C1: π=0.0002/bp; Table 2; Fig.2). In lineage C2, regions of relatively high nucleotide diversity were interspersed with tracts of low diversity (Fig. S3). Tajima's D values were either close to zero, or they were negative in the three lineages (C1: D=0.008; C2: D=-0.380; C3: D=-0.177; Table 2); a negative value is consistent with population expansion after a recent bottleneck or founder event. Absolute divergence (*dxy*) among lineages was similar between the three pairs of lineages (dxy=0.018/bp between C1 and C2, 0.018/bp between C2 and C3, and 0.015/bp between C1 and C3; Fig. 2).

Footprints of recombination

Recombination was analyzed in the ten largest contigs of each lineage, covering ~30% of the reference assembly. We detected a total of 375 recombined blocks using RDP4 software on the contigs analyzed, with stretches of nucleotide similarity across lineages distributed in a block-like structure (Table S2). Recombination rates differed significantly across lineages; based on the total of 375 recombination events, lineage C3 was found to have received the highest number of recombination events (n=203), followed by C2 (n=148), and with C1 receiving significantly fewer events

(n=24) (Randomization test: $p<10^{-6}$; Table S2). In this analysis, we counted the number of cases in which C1, C2 or C3 was the recombinant in Table S2. Analyses of linkage disequilibrium showed that LD decayed to half of its maximum value in less than 1kb in lineages C2 and C3, while LD decay was markedly slower and more jagged in C1 (Fig. 3). The large number of recombination events would have homogenized the nucleotide diversity and broken up any LD blocks in C2 and C3, resulting in a smooth LD decay. In contrast, the LD decay is more erratic in C1 because the few recombination events have not managed to break-up all LD blocks. We also tested for an alternative explanation; the smaller number of SNPs in C1 could have introduced stochasticity in the LD values as these values are calculated on fewer data. However, subsampling a smaller number of random SNPs from C2 and C3 to give them the same nucleotide diversity as C1 did not produce the erratic pattern of LD-decay. Hence, we conclude that the level of sexual recombination within C1 is lower than that of both other lineages. Finally, the PHI test rejected the null hypothesis of clonality in all three lineages ($P < 0.001$).

Genome scan for signatures of genetic exchanges

Probabilistic chromosome painting revealed genomic regions of shared ancestry between lineages, with shared fragments of size longer than the longest contig in the reference genome (2.37 Mb) (Fig. S4). Regions of shared ancestry were not strictly restricted to the three isolates (LFN0318, LFN0217, and LFN0150) previously detected by sNMF (Fig.1). For lineages C1 and C2, the majority of mutations were assigned to self (i.e., to their cluster of origin), but in some contigs relatively large regions were assigned to other lineages. In lineage C3, mutations tended to have non-zero membership probabilities in multiple clusters. This implies that these polymorphisms are shared across multiple lineages, probably reflecting the extremely high recombination rate

shown by this lineage. However, also in lineage C3, some regions were clearly assigned to lineage C1 and self. For contig66, the isolate LFN0318 – a representative of lineage C2 – shared high genetic similarity with lineage C3, consistent with many recent genetic exchanges between these lineages.

Further analyses using HYBRIDCHECK revealed a mosaic-like genome structure with well-defined blocks of high nucleotide similarity (Fig. S5). For contig66, relatively few short blocks of high similarity were detected between lineage C1 (using LFN0297 as the representative isolate) and the two other lineages (spanning from 1.4 to 1.6 Mb), while large blocks were detected between lineage C2 and C3 (using LFN0318 and LFN0308 as the representative isolates, respectively) (Fig. 4). Assuming that the contigs analyzed are representative of the rest of the genome, the proportion of genome introgression between lineages varied markedly: 2.4% between C1 (LFN0297) and C2 (LFN0318); 12.7% between C1 (LFN0297) and C3 (LFN0308); and 28.7% between C2 (LFN0318), and C3 (LFN0308) (Table 3 and Table S3).

In order to discriminate between incomplete lineage sorting and hybridization after secondary contact, we dated regions of high nucleotide identity between *C. truncatum* lineages detected by HYBRIDCHECK. The age estimates of recombinant blocks along the 10 largest contigs revealed recent introgression events. The most recent hybridization event was dated back to 6,100 years before present, assuming a generation time of one year (Table S4). Variation in the age of introgression events was however extensive, most likely because of a lack of closest-related donors for all recombinant regions detected (see Jouet, McMullan, & Van Oosterhout, 2015). The most recent events were more likely to reflect ongoing genetic exchanges. The older events are more consistent with incomplete lineage sorting, or alternatively, the coalescence time may not

accurately reflect the timing of genetic introgression because the "true" donors have not been sampled.

In the introgression regions, we identified 357 genes between lineage C1 (LFN0297) and lineage C2 (LFN0318), 389 genes between lineage C1 (LFN0297) and lineage C3 (LFN0308), and 584 genes between lineage C2 (LFN0318) and C3 (LFN0308). The introgressed regions include many secreted protein-encoding genes (between 38 and 62) (Table S5), including proteases and hydrolases which are known virulence-associated factors in pathogens (Monod et al., 2002; Soanes, Richards, & J.Talbot, 2007). However, these regions are not significantly enriched for those genes (binomial test $p>0.05$), and hence, we must conclude that genetic exchanges between these lineages are not more likely to involve genomic regions with virulence genes. Gene Ontology (GO) analysis revealed enrichment of GO terms between lineages, and these results are reported in Table S6.

Demographic inferences

To infer the demographic history of the three genetic lineages of *C. truncatum*, we compared three scenarios of isolation with or without migration for the four possible branching orders among lineages, using a diffusion approximation to the SFS implemented using DADI (We compared 3 x 4 = 12 models compared in total). Note that in the context of the DADI analysis, the term "migration" is similar to "genetic introgression" in the recombination analysis. Likelihood ratio tests indicated that the model with trifurcating lineages (topology 1) and asymmetrical migration was the most supported (Fig. S6). These results corroborate our recombination analyses. To convert demographic parameter estimates to physical units, we estimated the ancestral population size (N_{AB1}) (Fig.5). We used the population mutation rate $\theta=4N_{AB1}\mu L$, where μ was

assumed to be approximately 1e-8 per generation (Lynch, 2010) and L was the genome size (~55.1 Mb). This ancestral population size was then used to transform time estimates from DADI (in units of $2N_{AB1}$) into calendar years. Divergence was estimated to have initiated about 960,000 years ago, considering a generation time of a year. The lowest population size was estimated for lineage C1, consistent with a severe bottleneck (nu1=2085 individuals), and/or the least input of genetic variation though recombination or genetic introgression.

The migration rate into C3 (m13=1.05e-6, m23=2.25e-6) was higher than in all other directions, displaying one to two times higher order of magnitude. Furthermore, the rate of migration into C1 (m21 and m31) was lower than in all other directions. This suggests lineage C3 as the higher receptor of migration. These demographic analyses thus support the recombination and introgression analyses reported above.

Discussion

The fungus *Colletotrichum truncatum* is an invasive pathogen on soybean crops in Brazil that causes severe yield losses. We used a population genomics approach to characterize the genetic makeup and infer the evolutionary history of *C. truncatum* using isolates representing two important regions of soybean production in Brazil. We showed that Brazilian *C. truncatum* is subdivided into three genetic lineages. These lineages possess different levels of genetic variation, which could reflect differences in the magnitude of bottlenecks associated with introduction events. A non-exclusive alternative hypothesis is that such differences have been caused by variation in the levels of recombination (i.e., genetic exchanges within a lineage) and/or genetic introgression (genetic exchanges between lineages). All our recombination analyses supported that the lineage C3 with the highest nucleotide diversity was most affected by such genetic

exchanges, and that a larger number of these events also affected more genes. Conversely, C1 with the lowest nucleotide variation and the highest level of linkage disequilibrium was least affected by genetic exchanges. Furthermore, these conclusions are corroborated by our demographic analysis which showed that migration (or gene flow) into C1 (m21 and m32) is lower than into all other directions, and that migration into C3 was the highest. Next, we will discuss the evolutionary genomics of *C. truncatum*, the significance of the Brazilian bridgehead population, and the potential of fungal pathogens to evolve into invasive species.

Evolutionary genomics of C. truncatum

Our clustering analyses supported the existence of three lineages, in agreement with the pattern of population subdivision previously detected based on multilocus microsatellite typing (Rogério, Gladieux, Massola, & Ciampi-Guillardi, 2019). These phylogenetically equidistant lineages are characterized by markedly different levels of genomic standing variation. Genome-wide analyses of variability showed that lineage C1 is genetically depauperate and reproductively more isolated. Lineage C1 was almost free from introgression, whilst between ~10 to 30% of assembled contigs of lineage C2 and C3 comprised recombinant (or introgressed) regions. The low level of genetic variation in C1 is consistent with the low number of recombinant blocks (only 6.4%), but it can also indicate a more recent introduction into the country. It means that C1 lineage may simply not have had the opportunity to engage in many genetic exchanges yet. Such recent invasions could be associated with contaminated or infected soybean seeds imported from the U.S. during the 1960s and 1970s (Arantes and Miranda 1993; Hirimoto and Vello 1986; Wysmierski and Vello 2013). Regions of low nucleotide diversity in this lineage corresponded with negative Tajima's D values, which is

consistent with rapid population expansion after a recent founder event. These observations lend further support to our demographic inference of a recent invader, the limited amount of introgression, and the lower effective population size (*Ne*) of C1 lineage.

By contrast, lineage C3 showed a significantly higher level of genetic diversity, which may have been generated over time by genetic recombination and its much larger *Ne* (note that the large *Ne* estimate may simply be a consequence of the ample genetic variation that has introgressed into this lineage). This lineage may have already been present in Brazil, prior to the introduction of soybean. It is possible that this lineage may have been infecting other host species, such as lima bean and weeds, as proposed by earlier studies (Rogério, Gladieux, Massola, & Ciampi-Guillardi, 2019; Tiffany & Gilman, 1954). If C3 was the first lineage that established itself, it would also have had more opportunity for recombination and genetic introgression than both other lineages, which could have augmented its genetic diversity. Such monopoly effect may also have given this lineage a head-start, both in the co-evolutionary arms race with its host (Van Oosterhout, 2021), as well in the competition with other lineages.

We found evidence of a history of recombination both within and between lineages, applying a combination of integrated approaches. Fine-scale admixture mapping revealed that introgression occurred between the *C. truncatum* lineages coexisting in sympatry despite their relatively deep genomic divergence. The inference of individual ancestry coefficients using probabilistic chromosome painting detected large genomic regions of shared ancestry among the genetic lineages, suggesting relatively recent hybridization. History of recombination and genetic introgression was also supported by the analyzes with HYBRIDCHECK, which detected introgressed blocks that differed markedly in age. By estimating the coalescence time of introgressed regions,

505 we found some events have occurred as recently as 6,100 years ago. This suggests that
506 hybridization between *C. truncatum* lineages is a relatively recent - if not ongoing -
507 process. In this analysis, we used a sexual generation time of a year, which is typical in
508 plant pathogenic fungi from temperate areas. However, considering the climate
509 conditions of Brazil, the generation time may be much shorter than one year, and hence,
510 we may have overestimated the age of introgression events. Furthermore, it is unlikely
511 we sampled the actual parental sequence, which would cause a further overestimation of
512 the age. (when identifying the wrong parental sequence, the SNPs that differentiate the
513 parent and the recipient sequence are assumed to have accumulated since recombination
514 took place, erroneously placing the recombination event further in the past). In other
515 words, hybridization events and genetic exchanges may be considerably more recent than
516 our estimate. This is a potentially systematic bias typical for recombination studies, and
517 this can be corrected for by broader (or more intense) sampling.

518 Although our analyses identified recent introgression between lineages, a
519 substantiate proportion of the shared ancestry observed between lineages appear to be
520 caused by incomplete lineage sorting (or alternatively, we overestimated the age). In such
521 cases, the recombination blocks pre-date the lineage divergence (Durand et al., 2011;
522 McMullan et al., 2015), which implies that they are shared ancestral polymorphisms, or
523 that the genetic exchanges occurred before the split of the lineages. Consistent with this
524 latter interpretation, the oldest blocks were also the shortest, which implies that they have
525 been eroded by subsequent recombination and mutations, and hence, that these represent
526 genetic exchanges predating the split of the lineages. *Colletotrichum truncatum* genomes
527 therefore appear to be mosaics of distinct gene genealogies with markedly varied
528 coalescence times. Alternatively, we may not have captured all extant lineages, which
529 would have overestimated the introgression events. Future studies with a more

530 comprehensive sampling may be able to shed further light on this. Next, we discuss our
531 finding in the context of fungi as invasive species.
532
533 ***Evolutionary genomics of fungi as invasive species***
534 Fungal reproductive biology is conducive for genetic exchange, and such
535 recombination events could both result from sexual reproduction or parasexual events
536 via hyphal anastomosis. The latter mode has already been described for other
537 *Colletotrichum* species (Roca et al., 2003; Rosada et al., 2010; Souza-Paccola et al.,
538 2003; Vaillancourt et al., 2000).Given that such genetic exchanges were found in all
539 genomes – i.e., no pure genomes were found – introgression is likely to have augmented
540 both the genetic diversity and the fitness of these hybrid genotypes. Therefore, we would
541 conclude that adaptive introgression may have enhanced the evolutionary potential of *C.*
542 *truncatum* during its invasion. However, when we tested this hypothesis, we did not find
543 significant enrichment of secreted protein-encoding genes in the introgressed regions. In
544 hindsight, this may not be surprising; in a coevolutionary arms race, genetic novelty at
545 single virulence gene introduced by recombination could provide a selective advantage
546 that helps the recombinant lineage to establish itself (Van Oosterhout, 2021). Indeed,
547 specific targets are likely to be under positive selection, rather than the total number of
548 introgressed genes (Aguileta, Refrégier, Yockteng, Fournier, & Giraud, 2009). In other
549 words, our study may not have discovered “the smoking gun”, but we have established
550 “the means”, i.e., the large number of secreted protein-encoding genes that are exchanged
551 during genetic introgression, which are possible co-evolutionary targets for selection.
552 This implies that introgression can provide the genetic variation required in a host-
553 parasite arms race, although we do not have the phenotypic data to support this
554 hypothesis.

555 The level of phylogenetic divergence among lineages, coupled with the
556 demographic modeling carried out in this study, enables us to infer that the *C. truncatum*
557 lineages significantly diversified before their joint introduction into Brazil. Based on the
558 ancient signature of some of the recombination events, it is possible that genetic
559 exchanges have occurred during the divergence process. These genetic exchanges would
560 have prevented the accumulation of intrinsic postzygotic barriers, which underpin
561 reproductive isolation in many recently diverged species (Bomblies et al., 2007; Lee et
562 al., 2008; Masly & Presgraves, 2007). Another possibility is that they initially evolved in
563 allopatry, but their later introduction in the same areas may have provided opportunities
564 for secondary contact and hybridization, preventing the accumulation of postzygotic
565 barriers. In summary, the absence of reproductive isolation may have provided ample
566 evolutionary opportunities after secondary contact, allowing for hybridization between
567 diverged lineages.

568 Introduced populations can overcome consequences of low genetic variation from
569 founders, for instance, through the purging of deleterious alleles during bottlenecks
570 events, and via the fixation of *de novo* beneficial mutations from standing variation
571 (Estoup et al., 2018; Frankham, 2005; Schrieber & Lachmuth, 2017). The genetic and
572 environmental homogeneity found in soybean fields is thought to favor a huge census
573 size of invasive fungal populations. Furthermore, without genetic diversity in the host,
574 pathogens can rapidly spread and fix *de novo* evolved adaptations, overcoming new
575 resistant varieties of crops and fungicides applications. Hybridization could be
576 particularly important in this context because admixture could promote adaptation by
577 rapidly creating novel allelic combinations (Hessenauer et al., 2020; McMullan et al.,
578 2015; Nader et al., 2019). We propose that admixture has elevated the amount of
579 genotypic variation generated by genetic introgression. In turn, this could have increased

580 the amount of phenotypic variation due to transgressive segregation, providing novel
581 substrate for natural selection (Nichols et al., 2015).

582 Bridgehead population may enhance the adapted invasive potential of species by
583 enabling genetic exchanges between diverged lineages in the areas of first introduction.
584 In this scenario, bridgehead populations may acquire new traits increasing the probability
585 of successful establishment and further spread relative to native population (Bertelsmeier
586 & Keller, 2018). We hypothesize that the admixture between *C. truncatum* lineages may
587 lead to a bridgehead population, producing a highly adapted invasive population.
588 Although the bridgehead effect has been proposed as a potential explanation for many
589 successful biological invasions (Gau, Merz, Falloon, & Brunner, 2013; Leduc et al.,
590 2015; van Boheemen et al., 2017) there is currently no clear empirical support for this
591 hypothesis (but see Simon et al., 2011 for an example of an invasive fish species). To
592 contain biological invasions, vigilance for invasive bridgehead populations is needed
593 since they have the potential to generate new introductions (Bertelsmeier & Keller, 2018)
594 and increase the adaptive evolutionary potential through genetic reassortment during
595 hybridization.

596 Our study reinforces the practical applications of population genomics in preventing
597 or curtailing, pathogen dissemination by supporting early interventions to limit economic
598 damage (Stam et al., 2021). The Brazilian *C. truncatum* may represent a risk as a
599 bridgehead for future invasions of soybean-producing areas. Our data highlights the
600 inherent vulnerability of genetically uniform crops in the agro-ecological environment,
601 particularly when faced with pathogens that can take full advantage of the opportunities
602 offered by an increasingly globalized word. Some fungi have "The Means, Motive and
603 Opportunity" to become invasive pathogens of crops. Many fungal pathogens possess the
604 means in the form of a high propagule pressure through two modes of asexual

reproduction, as well as the ability to rapidly generate novel genotypic variation, particularly through genetic introgression. Some fungi also have a motive, given the large biomass of genetically near-uniform crops that are their natural host plants. Few species also have the opportunity in the form of bridgehead populations that enable the genetic exchange that fuel the co-evolutionary arms race. Invasive crop pathogens, like *C. truncatum*, have "The Means, Motive and Opportunity" to pose the greatest risk to future food security. Population genomics can help identify pathogens that pose such risk, thereby helping to inform control strategies to better protect crops in the future.

Acknowledgements

The authors are grateful for the financial support given by São Paulo Research Foundation (FAPESP, Grand/Award Number:2017/09178-8), National Science and Technology Development Council (CNPq, Grand/Award Number: 153958/2016-2), and National Council for the Improvement of Higher Education (CAPES/PDSE, Grand/Award Number: 88881.133223/2016-01, PROEX/CAPES, Grand/Award Number: 330002037002P3).

References

Aguileta, G., Refrégier, G., Yockteng, R., Fournier, E., & Giraud, T. (2009). Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution*, *9*(4), 656–670. doi: 10.1016/j.meegid.2009.03.010

Alexopoulos, C., Mims, C., & Blackwell, M. (1996). *Introductory Mycology*. New York: John Wiley and Sons.

Ali, S., Gladieux, P., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., … de Vallavieille-Pope, C. (2014). Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen *Puccinia striiformis* f.sp. *tritici*. *PLoS Pathogens*, *10*(1). doi: 10.1371/journal.ppat.1003903

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., … Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. doi: 10.1038/s41587-019-0036-z

Auton, A., & McVean, G. (2007). Recombination rate estimation in the presence of hotspots. *Genome Research*, *17*, 1219–1227. doi: 10.1101/gr.6386707

Barcelos, Q. L., Pinto, J. M. A., Vaillancourt, L. J., & Souza, E. A. (2014). Characterization of Glomerella strains recovered from anthracnose lesions on

common bean plants in Brazil. *PLoS ONE*, *9*(3). doi: 10.1371/journal.pone.0090910

Barton, N. (2010). Understanding adaptation in large populations. *PLoS Genetics*, *6*(6), 1–3. doi: 10.1371/journal.pgen.1000987

Bertelsmeier, C., & Keller, L. (2018). Bridgehead Effects and Role of Adaptive Evolution in Invasive Populations. *Trends in Ecology and Evolution*, *33*(7), 527–534. doi: 10.1016/j.tree.2018.04.014

Bomblies, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangl, J. L., & Weigel, D. (2007). Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biology*, *5*(9), 1962–1972. doi: 10.1371/journal.pbio.0050236

Boni, M. F., Posada, D., & Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, *176*(2), 1035–1047. doi: 10.1534/genetics.106.068874

Bueno-Sancho, V., Persoons, A., Hubbard, A., Cabrera-Quio, L. E., Lewis, C. M., Corredor-Moreno, P., … Saunders, D. G. O. (2017). Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects Seasonal Variation and Host Specificity. *Genome Biology and Evolution*, *9*(12), 3282–3296. doi: 10.1093/gbe/evx241

Cannon, P. F., Damm, U., Johnston, P. R., & Weir, B. S. (2012). *Colletotrichum* - current status and future directions. *Studies in Mycology*, *73*, 181–213. doi: 10.3114/sim0014

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676. doi: 10.1093/bioinformatics/bti610

Connors, J., Krzywinski, M., Schein, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos : An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645.

Crouch, J., O'Connell, R., Gan, P., Buiate, E., Torres, M. F., Beirn, L., … Vaillancourt, L. (2014). The Genomics of Colletotrichum. In D. R., L.-P. A., & Kole C. (Eds.), *Genomics of Plant-Associated Fungi: Monocot Pathogens* (pp. 123–140). doi: 10.1007/978-3-662-44053-7

Damm, U., Woudenberg, J. H. C., Cannon, P. F., & Crous, P. W. (2009). Colletotrichum species with curved conidia from herbaceous hosts. *Fungal Diversity*, *39*, 45–87.

De Jonge, R., Bolton, M. D., Kombrink, A., Van Den Berg, G. C. M., Yadeta, K. A., & Thomma, B. P. H. J. (2013). Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Research*, *23*(8), 1271–1282. doi: 10.1101/gr.152660.112

Diao, Y., Zhang, C., Xu, J., Lin, D., Liu, L., Mtung'e, O. G., & Liu, X. (2015). Genetic differentiation and recombination among geographic populations of the fungal pathogen Colletotrichum truncatum from chili peppers in China. *Evolutionary Applications*, *8*(1), 108–118. doi: 10.1111/eva.12233

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252. doi: 10.1093/molbev/msr048

Dutech, C., BarrÈs, B., Bridier, J., Robin, C., Milgroom, M. G., & RavignÉ, V. (2012). The chestnut blight fungus world tour: Successive introduction events from diverse origins in an invasive plant fungal pathogen. *Molecular Ecology*, *21*(16), 3931–3946. doi: 10.1111/j.1365-294X.2012.05575.x

691 Estoup, A., Ravign, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2018). Is
692 There A Genetic Paradox of Biological Invasion. *Annual Review of Ecology,*
693 *Evolution, and Systematics*, *47*, 51–72.
694 Fournier, E., Gladieux, P., & Giraud, T. (2013). The "Dr Jekyll and Mr Hyde fungus":
695 Noble rot versus gray mold symptoms of *Botrytis cinerea* on grapes. *Evolutionary*
696 *Applications*, *6*(6), 960–969. doi: 10.1111/eva.12079
697 Frankham, R. (2005). Resolving the genetic paradox in invasive species. *Heredity*,
698 *94*(4), 385. doi: 10.1038/sj.hdy.6800634
699 Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and
700 Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, *196*(4), 973–
701 983. doi: 10.1534/GENETICS.113.160572
702 Gau, R. D., Merz, U., Falloon, R. E., & Brunner, P. C. (2013). Global Genetics and
703 Invasion History of the Potato Powdery Scab Pathogen, *Spongospora subterranea*
704 f.sp. *subterranea*. *PLoS ONE*, *8*(6). doi: 10.1371/journal.pone.0067944
705 Gibbs, M. J., Armstrong, J. S., & Gibbs, A. J. (2000). Sister-scanning: A Monte Carlo
706 procedure for assessing signals in rebombinant sequences. *Bioinformatics*, *16*(7),
707 573–582. doi: 10.1093/bioinformatics/16.7.573
708 Gladieux, P., Feurtey, A., Hood, M. E., Snirc, A., Clavel, J., Dutech, C., … Giraud, T.
709 (2015). The population biology of fungal invasions. *Molecular Ecology*, *24*(9),
710 1969–1986. doi: 10.1111/mec.13028
711 Gladieux, Pierre, Ravel, S., Rieux, A., Cros-Arteil, S., Adreit, H., Milazzo, J., …
712 Tharreau, D. (2018). Coexistence of multiple endemic and pandemic lineages of
713 the rice blast pathogen. *MBio*. doi: 10.1128/mBio.01806-17
714 Gladieux, Pierre, Ropars, J., Badouin, H., Branca, A., Aguileta, G., De Vienne, D. M.,
715 … Giraud, T. (2014). Fungal evolutionary genomics provides insight into the
716 mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology*, *23*(4),
717 753–773. doi: 10.1111/mec.12631
718 Grünwald, N. J., McDonald, B. A., & Milgroom, M. G. (2016). Population Genomics
719 of Fungal and Oomycete Pathogens. *Annual Review of Phytopathology*, *54*, 323–
720 346. doi: 10.1146/annurev-phyto-080614-115913
721 Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009).
722 Inferring the Joint Demographic History of Multiple Populations from
723 Multidimensional SNP Frequency Data. *PLoS Genetics*, *5*(10), e1000695.
724 doi:10.1371/journal.pgen.1000695. doi: 10.1371/journal.pgen.1000695
725 Hartmann, F. E., Mcdonald, B. A., & Croll, D. (2018). Genome-wide evidence for
726 divergent selection between populations of a major agricultural pathogen.
727 *Molecular Ecology*, 2725–2741. doi: 10.1111/mec.14711
728 Hessenauer, P., Feau, N., Gill, U., Schwessinger, B., Brar, G. S., & Hamelin, R. C.
729 (2020). Evolution and Adaptation of Forest and Crop Pathogens in the
730 Anthropocene. *Phytopathology*. doi: 10.1094/phyto-08-20-0358-fi
731 Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations.
732 *Theoretical and Applied Genetics*, *38*(6), 226–231. doi: 10.1007/bf01245622
733 Hubbard, A., Lewis, C. M., Yoshida, K., Ramirez-Gonzalez, R. H., de Vallavieille-
734 Pope, C., Thomas, J., … Saunders, D. G. O. (2015). Field pathogenomics reveals
735 the emergence of a diverse wheat yellow rust population. *Genome Biology*, *16*(1),
736 1–15. doi: 10.1186/s13059-015-0590-8
737 Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in
738 evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. doi:
739 10.1093/molbev/msj030
740 Hyde, K. ., Cai, L., McLenzie, E. H. C., Yang, Y. L., Zhang, J. Z., & Prihastuti, H.

(2009). Colletotrichum: a catalogue of confusion. *Fungal Diversity*, *39*, 1–17.
Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. doi: 10.1093/bioinformatics/btr521
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*(12), 3077–3092. doi: 10.1111/mec.13213
Katoch, A., Sharma, P., Padder, B. A., & Sharma, P. N. (2017). Population Structure of *Colletotrichum truncatum* in Himachal Pradesh and Identification of Broad-Spectrum Resistant Sources in *Capsicum*. *Agricultural Research*, *6*(3), 296–303. doi: 10.1007/s40003-017-0261-y
Latorre, S., Reyes-Avila, C. S., Malmgren, A., Win, J., Kamoun, S., & Burbano, H. (2020). Recently expanded clonal lineages of the rice blast fungus display distinct patterns of presence/absence of effector genes. *BMC Biology*, 1–15. doi: 10.1101/2020.01.09.900308
Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, *8*(1), 11–17. doi: 10.1371/journal.pgen.1002453
Leduc, A., Traoré, Y. N., Boyer, K., Magne, M., Grygiel, P., Juhasz, C. C., … Pruvost, O. (2015). Bridgehead invasion of a monomorphic plant pathogenic bacterium: Xanthomonas citri pv. citri, an emerging citrus pathogen in Mali and Burkina Faso. *Environmental Microbiology*, *17*(11), 4429–4442. doi: 10.1111/1462-2920.12876
Lee, H. Y., Chou, J. Y., Cheong, L., Chang, N. H., Yang, S. Y., & Leu, J. Y. (2008). Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility between Two Yeast Species. *Cell*, *135*(6), 1065–1073. doi: 10.1016/j.cell.2008.10.047
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. doi: 10.1093/bioinformatics/btp324
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi: 10.1093/bioinformatics/btp352
Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, *26*(8), 345–352. doi: 10.1016/j.tig.2010.05.003
Martin, D. P., Posada, D., Crandall, K. A., & Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses*, *21*(1), 98–102. doi: 10.1089/aid.2005.21.98
Martin, D., & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. Bioinformatics. *Bioinformatics*, *16*(6), 562–563. Retrieved from www.uct.ac.za/
Martin, Darren P, Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1). doi: 10.1093/ve/vev003
Martin, M. (2011). Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.Journal*, *17*, 10–12. doi: https://doi.org/10.14806/ej.17.1.200
Masly, J. P., & Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS Biology*, *5*(9), 1890–1898. doi:

10.1371/journal.pbio.0050243
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePri, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. doi: 10.1101/gr.107524.110.20
McMullan, M., Gardiner, A., Bailey, K., Kemen, E., Ward, B. J., Cevik, V., … Jones, J. D. G. (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*, *2015*(4), 1–24. doi: 10.7554/eLife.04550
Menat, J., Banniza, S., Cabral, A. L., Vijayan, P., & Wei, Y. (2012). Glomerella truncata : another Glomerella species with an atypical mating system . *Mycologia*, *104*(3), 641–649. doi: 10.3852/10-265
Miles A. Harding. (n.d.). Scikit-allel - Explore and analyse genetic variation. In., 1. 2.0 edn. Retrieved from https://github.com/cggh/scikit-allel:
Monod, M., Capoccia, S., Léchenne, B., Zaugg, C., Holdom, M., & Jousson, O. (2002). Secreted proteases from pathogenic fungi. *International Journal of Medical Microbiology*, *292*(5–6), 405–419. doi: 10.1078/1438-4221-00223
Nader, J. L., Mathers, T. C., Ward, B. J., Pachebat, J. A., Swain, M. T., Robinson, G., … Tyler, K. M. (2019). Evolutionary genomics of anthroponosis in *Cryptosporidium*. *Nature Microbiology*, *doi.org/10*. doi: 10.1038/s41564-019-0377-x
Nichols, P., Genner, M. J., van Oosterhout, C., Smith, A., Parsons, P., Sungani, H., … Joyce, D. A. (2015). Secondary contact seeds phenotypic novelty in cichlid fishes. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798). doi: 10.1098/rspb.2014.2272
Padidam, M., Sawyer, S., & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology*, *265*(2), 218–225. doi: 10.1006/viro.1999.0056
Parker, I. M., & Gilbert, G. S. (2018). Density-dependent disease, life-history trade-offs, and the effect of leaf pathogens on a suite of co-occurring close relatives. *Journal of Ecology*, *106*(5), 1829–1838. doi: 10.1111/1365-2745.13024
Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., & Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, *26*(7), 1902–1918. doi: 10.1111/mec.13980
Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research*, *9*. doi: 10.12688/f1000research.23297.2
Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., … Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, *26*(19), 5245–5263. doi: 10.1111/mec.14266
Posada, D., & Crandall, K. A. (2002). Evaluation of methods for detecting recombination from DNA sequences: Empirical data. *Molecular Biology and Evolution*, *19*(5), 708–717. doi: 10.1093/oxfordjournals.molbev.a004129
Price, G. R. (1972). Fisher's "fundamental theorm" made clear. *Annals of Human Genetics*, *36*, 129–140.
Roca, M. G., Davide, L. C., Mendes-Costa, M. C., & Wheals, A. (2003). Conidial anastomosis tubes in *Colletotrichum*. *Fungal Genetics and Biology*, *40*(2), 138–145. doi: 10.1016/S1087-1845(03)00088-4
Rogério, F., Boufleur, T. R., Ciampi-Guillardi, M., Sukno, S. A., Thon, M. R., Massola,

N. S., & Baroncelli, R. (2020). Genome sequence resources of Colletotrichum truncatum, C. plurivorum, C. Musicola, and ,C. Sojae: Four species pathogenic to soybean (Glycine max). *Phytopathology*, *110*(9), 1497–1499. doi: 10.1094/PHYTO-03-20-0102-A

Rogério, F., Gladieux, P., Massola, N. S., & Ciampi-Guillardi, M. (2019). Multiple Introductions Without Admixture of Colletotrichum truncatumAssociated with Soybean Anthracnose in Brazil. *Phytopathology*, *109*(49), 681–689. doi: 10.1094/phyto-08-18-0321-r

Rosada, L. J., Franco, C. C. S., Sant'anna, J. R., Kaneshima, E. N., Gonçalves-Vidigal, M. C., & Castro-Prado, M. A. A. (2010). Parasexuality in Race 65 *Colletotrichum lindemuthianum* Isolates. *Journal of Eukaryotic*, *57*(4), 383–384. doi: 10.1111/j.1550-7408.2010.00486.x

Schrieber, K., & Lachmuth, S. (2017). The genetic paradox of invasions revisited: The potential role of inbreeding × environment interactions in invasion success. *Biological Reviews*, *92*(2), 939–952. doi: 10.1111/brv.12263

Silva, D. N., Várzea, V., Paulo, O. S., & Batista, D. (2018). Population genomic footprints of host adaptation, introgression and recombination in coffee leaf rust. *Molecular Plant Pathology*, *19*(7), 1742–1753. doi: 10.1111/mpp.12657

Simon, A., Britton, R., Gozlan, R., Van Oosterhout, C., Volckaert, F. A. M., & Hanfling, B. (2011). Invasive Cyprinid Fish in Europe Originate from the Single Introduction of an Admixed Source Population Followed by a Complex Pattern of Spread. *PLoS ONE*, *6*(6), e18560. doi:10.1371/journal.pone.0018560. doi: 10.1371/journal.pone.0018560

Smith, J. M. (1992). Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, *34*(2), 126–129. doi: 10.1007/BF00182389

Soanes, D. M., Richards, T. A., & J.Talbot, N. (2007). Insights from Sequencing Fungal and Oomycete Genomes: What Can We Learn about Plant Disease and the Evolution of Pathogenicity? *The Plant Cell*, *19*, 3318–3326. doi: 10.1105/tpc.107.056663

Souza-Paccola, E. A., Fávaro, L. C. L., Casela, C. R., & Paccola-Meirelles, L. D. (2003). Genetic Recombination in *Colletotrichum sublineolum*. *Journal of Phytopathology*, *151*(6), 329–334. doi: 10.1046/j.1439-0434.2003.00727.x

Stam, R., Gladieux, P., Vinatzer, B. A., Goss, E. M., Potnis, N., Candresse, T., & Brewer, M. T. (2021). Population Genomic- And phylogenomic-Enabled Advances to Increase Insight into Pathogen Biology and Epidemiology. *Phytopathology*, *111*(1), 8–11. doi: 10.1094/PHYTO-11-20-0528-FI

Stauber, L., Badet, T., Prospero, S., & Croll, D. (2020). Emergence and diversification of a highly invasive chestnut pathogen lineage across south-eastern Europe. *BioRxiv*. doi: 10.1101/2020.02.15.950170

Stukenbrock, E. H., & McDonald, B. A. (2008). The Origins of Plant Pathogens in Agro-Ecosystems. *Annual Review of Phytopathology*, *46*, 75–100. doi: 10.1146/annurev.phyto.010708.154114

Talhinhas, P., & Baroncelli, R. (2021). *Colletotrichum* species and complexes: geographic distribution, host range and conservation status. In *Fungal Diversity*. Springer Netherlands. doi: 10.1007/s13225-021-00491-9

Taylor, J., Jacobson, D., & Fisher, M. (1999). THE E VOLUTION OF A SEXUAL F UNGI : Reproduction, specification and classification. *Annual Review of Phytopathology*, *37*, 197–246.

Thierry, M., Milazzo, J., Adreit, H., Ravel, S., Borron, S., Sella, V., … Gladieux, P. (2020). Ecological Differentiation and Incipient Speciation in the Fungal Pathogen

Causing Rice Blast. *BioRxiv*. doi: 10.1101/2020.06.02.129296
Thrall, P. H., Hochberg, M. E., Burdon, J. J., & Bever, J. D. (2007). Coevolution of symbiotic mutualists and parasites in a community context. *Trends in Ecology and Evolution*, *22*(3), 120–126. doi: 10.1016/j.tree.2006.11.007
Thrall, P. H., Oakeshott, J. G., Fitt, G., Southerton, S., Burdon, J. J., Sheppard, A., … Ford Denison, R. (2011). Evolution in agriculture: The application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. *Evolutionary Applications*, *4*(2), 200–215. doi: 10.1111/j.1752-4571.2010.00179.x
Tiffany, L., & Gilman, J. (1954). Species of Colletotrichum from Legumes. *Mycologia*, *46*(1), 52–75.
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*(3), 430. doi: 10.2307/3761501
van Boheemen, L. A., Lombaert, E., Nurkowski, K. A., Gauffre, B., Rieseberg, L. H., & Hodgins, K. A. (2017). Multiple introductions, admixture and bridgehead invasion characterize the introduction history of *Ambrosia artemisiifolia* in Europe and Australia. *Molecular Ecology*, *26*(20), 5421–5434. doi: 10.1111/mec.14293
Van Oosterhout, C. (2021). Mitigating the threat of emerging infectious diseases; a coevolutionary perspective. *Virulence*, *12*(1), 1288–1295. doi: 10.1080/21505594.2021.1920741
Vieira, A., Silva, D. N., Várzea, V., Paulo, O. S., & Batista, D. (2018). Novel insights on colonization routes and evolutionary potential of *Colletotrichum kahawae*, a severe pathogen of <i>Coffea arabica. *Molecular Plant Pathology*, *19*(11), 2488–2501. doi: 10.1111/mpp.12726
Ward, B. J., & van Oosterhout, C. (2016). HYBRIDCHECK: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Molecular Ecology Resources*, *16*(2), 534–539. doi: 10.1111/1755-0998.12469
Weidemann, G. J., TeBeest, D. O., & Cartwright, R. D. (1988). Host Specificity of *Colletotrichum gloeosporioides* f. sp. *aeschynomene* and *C. truncatum* in the Leguminosae. *Phytopathology*, *78*, 986–990.
Wilson, A. M., Lelwala, R. V., Taylor, P. W. J., Wingfield, M. J., & Wingfield, B. D. (2021). Unique patterns of mating pheromone presence and absence could result in the ambiguous sexual behaviors of Colletotrichum species. *G3: Genes, Genomes, Genetics*, *11*(9). doi: 10.1093/g3journal/jkab187
Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, *35*(10), 1786–1788. doi: 10.1093/bioinformatics/bty875

Data accessibility

DNA sequences: Short Read Archive Accession in Table S1 (Supporting information)

Author Contributions

M.C.G., and N.S.M.J conceived and designed the research; F.R., and M.C.G collected the samples; S.C.A. obtained genomic data; F.H.C., G.K.H., and G.R.A.M. performed genetic analysis; F.R., C.V.O., and P.G. analyzed the data and wrote the manuscript.

Figure captions

Figure 1. Population subdivision of *Colletotrichum truncatum*. (A) Neighbor-Net networks showing relationships between isolates identified on the basis of the full set of SNPs without missing data. The groups revealed are referred to as lineage C1, C2, and C3. (B) Scatterplot from discriminant analysis of principal components (DAPC). (C) Individual ancestry coefficients estimated using sNMF. Each isolate is represented by a thick vertical line in the most probably number of groups (K=3), and bar colors represent each lineage. Asterisks indicate admixture isolates.

Figure 2. Box plots of the average populations pairwise nucleotide changes per site in 100-kb windows within (nucleotide diversity (π)) and between lineages (*dxy*).

Figure 3. Linkage disequilibrium (LD) decay plots of three genetic lineages of *Colletotrichum truncatum* (C1, C2, and C3). LD measured was calculated for all pairs of SNPs less than 300 kb distance apart.

Figure 4. Scan for signature of genetic exchanges between *C. truncatum* lineages. The isolates LFN0297 (lineage C1), LFN0318 (lineage C2), and LFN0308 (lineage C3) were used as representative of each lineage. The sequence similarity along contig 66 among isolates visualized through RBG color triangular in the software HYBRIDCHECK. Areas where two sequences have the same color (yellow, purple or turquoise) are indicative of two lineages sharing the same polymorphisms. The bottom panel shows the linear plot of the proportion of SNPs shared between the three pairwise comparisons.

Figure 5. The best support demographic model (isolation with asymmetrical migration) using the three-dimensional site frequency spectrum (3D-SFS) between *Colletotrichum truncatum* lineages (C1, C2, and C3). The following parameters were estimated: nu1: size of C1 after split; nu2: size of C2 after split; nu3: size of C3 after split; m12: asymmetrical migration between C1 and C2; m21: asymmetrical migration between C2 and C1; m23: asymmetrical migration between C2 and C3; m32:

asymmetrical migration between C3 and C2.; m13: asymmetrical migration between C1 and C3; m31: asymmetrical migration between C3 and C1; T1: scaled time between the split and the size change (in units of 2*Na generations). Thick arrows indicate higher migration rates.

Supplemental figure captions

Figure S1. Visual representation of the 3D demographic models fitted.

Figure S2. (A) Bayesian information criteria (BIC) indicating the most probable number of genetic groups by discriminant analysis of principal components analysis (DAPC). (B) Cross-entropy as a function of the number of clusters K modeled in sNMF analyses of population subdivision.

Figure S3. Circos plot showing genome-wide diversity of the ten largest contigs of *Colletotrichum truncatum* genomes. Nucleotide diversity (π) per SNP and Tajima's D were calculated in sliding windows of 10 kb.

Figure S4. Probabilistic chromosome painting bar plots of the ten largest contigs among three *Colletotrichum truncatum* lineages (C1, C2, and C3).

Figure S5. Variation in sequence similarity between *Colletotrichum truncatum* lineages (C1, C2, and C3) on the ten largest contigs. The isolates LFN0297 (lineage C1), LFN0318 (lineage C2), and LFN0308 (lineage C3) were used as representative of each lineage and visualized as mosaic-like genome structure through RBG color triangular in the software HYBRIDCHECK.

Figure S6. Demographic models tested using the three-dimensional site frequency spectrum (3D-SFS) between *Colletotrichum truncatum* lineages (C1, C2, and C3). Data, models and residuals are presented using a heatmap. The best model fitted is in bold. The following parameters were estimated: nu1: size of C1 after split; nu2: size of C2 after split; nu3: size of C3 after split; m12: asymmetrical migration between C1 and C2; m21: asymmetrical migration between C2 and C1; m23: asymmetrical migration between C2 and C3; m32: asymmetrical migration between C3 and C2.; m13: asymmetrical migration between C1 and C3; m31: asymmetrical migration between C3 and C1; T1: scaled time

between the split and the size change (in units of 2*Na generations). LL = log likelihood of the model. AIC: Akaike information criterion.

Table 1. *Colletotrichum truncatum* isolates used in this study

Cluster	Isolate	State, city	Decimal coordinates	Cultivar	Sampling year
C2	LFN0217	Mato Grosso, Sinop	-11.923, -55.616	Nidera 7709	2017
C3	LFN0225	Mato Grosso, Sinop	-11.923, -55.616	Nidera 7709	2017
C2	LFN0248	Mato Grosso, Sinop	-11.982, -55.507	Nidera 7709	2017
C1	LFN0262	Mato Grosso, Sinop	-11.982, -55.507	Nidera 7709	2017
C3	LFN0268	Mato Grosso, Sinop	-11.982, -55.507	Nidera 7709	2017
C3	LFN0150	Mato Grosso, Lucas do Rio Verde	-13.169, -56.068	Y-70 Pioneer	2017
C1	LFN0169	Mato Grosso, Sinop	-13.312, -56.042	Monsoy 8766	2016
C1	LFN0185	Mato Grosso, Lucas do Rio Verde	-13.410, -56.067	Monsoy 8372	2017
C2	LFN0205	Mato Grosso, Lucas do Rio Verde	-13.410, -56.067	Monsoy 8372	2017
C2	LFN0288	Goiás, Montividiu	-17.411, -50.950	Monsoy 7739	2017
C3	LFN0291	Goiás, Montividiu	-17.411, -50.950	Monsoy 7739	2017
C1	LFN0297	Goiás, Montividiu	-17.411, -50.950	Monsoy 7739	2017
C3	LFN0308	Goiás, Rio Verde	-17.460, -51.119	Macroseed PP7200	2017
C1	LFN0309	Goiás, Rio Verde	-17.460, -51.119	Macroseed PP7200	2017
C2	LFN0318	Goiás, Rio Verde	-17.460, -51.119	Macroseed PP7200	2017
C1	LFN0346	Goiás, Rio Verde	-17.764, -51.035	Nidera 5909	2017
C2	LFN0349	Goiás, Rio Verde	-17.764, -51.035	Nidera 5909	2017
C1	LFN0360	Goiás, Rio Verde	-17.764, -51.035	Nidera 5909	2017

Table 2. Summary of genomic diversity within *Colletotrichum truncatum* lineages in nonoverlapping 100kb windows

Lineage	**N[a]**	**π[b]**	**D[c]**
C1	7	0.0002 (0.0003) *	0.04 (1.118)

C2	6	0.0062 (0.0041)	-0.380 (0.858)
C3	5	0.0113 (0.0056)	-0.180 (0.623)

[a] Sample size
[b] Nucleotide diversity per base pair, for all contigs larger than 10kb
[c] Tajima's neutrality statistic, for all contigs larger than 10kb
* Standard deviation

Table 3. Percentage of genetic introgression between lineages of *Colletotrichum truncatum*

	Pairwise lineages		
Contig	**C1 x C2**	**C1 x C3**	**C2 x C3**
tig55	2.70	13.56	11.17
tig66	1.88	1.31	68.43
tig97	-	-	67.74

tig164	-	36.00	-
tig209	1.99	20.08	6.74
tig284	-	0.22	19.95
tig332	4.38	-	12.55
tig526	0.03	1.19	78.54
tig70486	4.03	16.16	3.66
tig70488	0.22	2.23	11.87
total	2.43	12.75	28.72

1083