University of São Paulo
"Luiz de Queiroz" College of Agriculture

Population genetics of *Colletotrichum truncatum* associated with soybean anthracnose

Flávia Rogério

Thesis presented to obtain the degree of Doctor in Science.
Area: Plant Pathology

Piracicaba
2019

Flávia Rogério
Agronomist

Population genetics of *Colletotrichum truncatum* associated with soybean anthracnose

Advisor:
Prof. Dr. **NELSON SIDNEI MASSOLA JUNIOR**

Thesis presented to obtain the degree of Doctor in Science.
Area: Plant Pathology

Piracicaba
2019

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Rogério, Flávia

Population genetics of *Colletotrichum truncatum* associated with soybean anthracnose / Flávia Rogério. - - Piracicaba, 2019.

78 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Doença da soja 2. Diversidade genética 3. Tipagem de microssatélites multilocus 4. Sequenciamento genômico completo 5. Mistura 6. Linhagem genéticas I. Título

I dedicate with love to
my parents
Maria Cremoneis Becegatto
and João Vitorino Rogério.

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to,

God, its infinite love and mysteries, for allowing me to live so wonderful experiences in my journey.

All my family, especially to my parents, João and Maria, my sister Ana Paula, my brother-in-law Anderson, my nephews Anna Luiza, Giovanna and Pedro for their continuous love and support.

The "Luiz de Queiroz" College of Agriculture - University of São Paulo (ESALQ-USP) and the Graduate Program in Plant Pathology for my education during my master's and doctoral studies.

The National Science and Technology Development Council (CNPq) for providing the scholarship during all my doctorate course (CNPq Process 153958/2016-2).

The National Council for the Improvement of Higher Education (CAPES) for providing the scholarship abroad through Doctoral Program Sandwich Abroad (PDSE) in France from October 2017 to February 2018 (PDSE process 88881.133223/2016-01).

The São Paulo Research Foundation (FAPESP) for financial support to execution this project (FAPESP 2017/09178-8).

My advisor Dr. Nelson Sidnei Massola Júnior and co-advisor Dra. Maisa Ciampi-Guillardi for their generosity, guidance, support, friendship and knowledge, which were essential during my doctoral study.

My abroad-advisor Dr. Pierre Gladieux for his patience, kindness, knowledge and encouragement during my stay at Institut national de la recherche agronomique – INRA, Montpellier – France, in addition to his all support provided to genomic data analyzes.

The staff of Institut national de la recherche agronomique – INRA for their valuable support, specially to Sandrine Cros-Arteil for her support and kindness during my stay in France.

My colleagues from Graduate Program in Plant Pathology at ESALQ and Micology laboratory for all their support, talks, friendship and scientific discussions.

All my friends that I made during my stay in Piracicaba, which I will take their friendship forever.

Lastly, everyone who supported me throughout the course of this journey.

Thank you all!

“Nothing in Biology Makes Sense Except in the Light of Evolution”

(Theodosius Dobzhansky)

CONTENTS

RESUMO 8

ABSTRACT 9

1. GENERAL INTRODUCTION 11

REFERENCES 14

2. MULTIPLE INTRODUCTIONS WITHOUT ADMIXTURE OF *COLLETOTRICHUM TRUNCATUM* ASSOCIATED WITH SOYBEAN ANTHRACNOSE IN BRAZIL 19

ABSTRACT 19

2.1. INTRODUCTION 19
2.2. MATERIALS AND METHODS 22
2.2.1. Population sampling 22
2.2.2. Isolation of *Colletotrichum truncatum* 23
2.2.3. DNA extraction and SSR genotyping 23
2.2.4. Genetic and haplotypic diversity 24
2.2.5. Population differentiation 25
2.2.6. Reproductive mode 26
2.3. RESULTS 26
2.3.1. Summary of microsatellite variability 26
2.3.2. Population differentiation 29
2.3.3. Reproductive mode 33
2.4. DISCUSSION 34

REFERENCES 37

SUPPORTING INFORMATION 43

3. DRAFT GENOME SEQUENCE OF *COLLETOTRICHUM TRUNCATUM* ASSOCIATED WITH SOYBEAN ANTHRACNOSE 49

ABSTRACT 49

REFERENCES 51

4. ADMIXTURE BETWEEN DIVERGENT LINEAGES OF THE FUNGUS *COLLETOTRICHUM TRUNCATUM* ASSOCIATED WITH SOYBEAN ANTHRACNOSE..... 53

ABSTRACT 53

4.1 INTRODUCTION 53
4.2 MATERIALS AND METHODS 55
4.2.1 Sample collection, DNA extraction and sequencing 55
4.2.2 Taxonomic and pathogenicity confirmation 56
4.2.3 Genome sequencing and SNP calling 57
4.2.4 Genealogical relationships and population structure 58
4.2.5 Linkage disequilibrium and recombination 58
4.2.6 Genome scan for genetic exchanges 59
4.2.7 Summary statistics of polymorphism and divergence 60
4.3 RESULTS 60

4.3.1 Genome sequencing and SNP calling 60
4.3.2 Taxonomic and pathogenicity confirmation 60
4.3.3 Genealogical relationships and population structure 63
4.3.4 Linkage disequilibrium and recombination 65
4.3.5 Probabilistic chromosome painting 66
4.3.6 Summary statistics of polymorphism and divergence 68
4.4 DISCUSSION 69
REFERENCES 72
SUPPORTING INFORMATION 78

RESUMO

Genética populacional de *Colletotrichum truncatum* associado à antracnose da soja

A soja é uma das principais culturas agrícolas, com alta relevância econômica global. A grande área sob cultivo de soja no Brasil, incluindo a incorporação de novas áreas nas regiões norte e centro-oeste, principalmente sob monocultura e plantio direto, tem afetado a prevalência e a intensidade das doenças. Entre elas, uma das mais proeminentes é a antracnose, principalmente associada à espécie fúngica *Colletotrichum truncatum*. O conhecimento da estrutura genética das populações de patógenos de plantas pode ser usado para inferir suas histórias de vida e os processos evolutivos que moldam as populações nos agroecossistemas, o que pode ajudar a implementar estratégias eficazes de manejo da doença. No entanto, a estrutura genética das populações de *C. truncatum* associadas à soja permanece desconhecida. Coletamos isolados de *C. truncatum* em 10 áreas, representando duas principais regiões produtoras de soja no Brasil. Utilizamos marcadores microssatélites e sequenciamento do genoma completo para investigar a biologia populacional e a história evolutiva desse importante patógeno. A tipagem de microssatélites multilocus de 237 isolados revelou alta diversidade genética e haplotípica nas populações, bem como baixa diferenciação genética e compartilhamento de haplótipos entre populações e regiões. Além disso, foram detectados três grupos genéticos distintos, coexistindo nas mesmas áreas, sem evidência de mistura entre eles. Isto sugere que as populações *C. truncatum* no Brasil resultaram de pelo menos três eventos fundadores, o que levou á formação das três linhagens genéticas que se espalharam pelo país. No entanto, a composição genética dessas linhagens permanece desconhecida, e sua extrema proximidade geográfica levanta a questão sobre a manutenção de sua integridade genética em face a mistura entre elas. A fim de analisar a história evolutiva das linhagens de *C. truncatum* e investigar a possibilidade de ausência de trocas genéticas entre elas, empregamos uma abordagem genômica populacional. Para isso, produzimos uma versão preliminar do genoma completo de um isolado típico da espécie, o qual foi utilizado como genoma de referência. Dezoito isolados representativos das três linhagens foram submetidos ao sequenciamento completo, alinhados ao genoma de referência, e variantes foram identificados. Nossas análises genômicas populacionais revelaram que a estrutura genética do patógeno é formada por três linhagens profundamente divergentes, com níveis de diversidade consistentes com repetidos eventos de introdução para cada linhagem. Também encontramos evidências de recombinação sexual dentro e entre linhagens, com múltiplos isolados apresentando assinaturas de mistura. Nossas descobertas sugerem um cenário no qual as três linhagens divergiram inicialmente em alopatria antes de experimentar hibridação, após contato secundário. O monitoramento da diversidade do patógeno ao longo do tempo é necessário para revelar se essas linhagens se mantêm geneticamente separadas ou se fundem, o que pode afetar os métodos de controle da doença atualmente empregados.

Palavras-chave: Doença da soja; Diversidade genética; Tipagem de microssatélites multilocus; Sequenciamento genômico completo; Mistura; Linhagem genéticas

ABSTRACT

Population genetics of *Colletotrichum truncatum* associated with soybean anthracnose

The soybean crop is one of the main agricultural crops, with high global economic relevance. The large area under soybean cultivation in Brazil, including the incorporation of new areas in the northern and midwestern regions, mostly under monoculture and non-tillage system, has been affected the prevalence and the intensity of diseases. Among these, one of most prominent is anthracnose, mainly associated with the fungal species *Colletotrichum truncatum.* Knowledge of the genetic structure of plant pathogen populations can be used to infer their life histories and the evolutionary processes that shape populations in the agroecosystems, which can help to implement effective disease management strategies. However, the genetic structure of *C. truncatum* populations associated with soybean remains unknown. We collected *C. truncatum* isolates from 10 sites representing two of main areas of soybean producing in Brazil and used microsatellite markers and whole-genome sequencing to investigate the population biology and evolutionary history of this important pathogen. The multilocus microsatellite typing of 237 isolates revealed high gene and haplotypic diversity within populations, as well low genetic differentiation and sharing of multilocus haplotypes among populations and regions. In addition, three distinct genetic clusters were detected, coexisting in syntopy in the soybean fields, without evidence of admixture between them. Such finding suggesting that Brazilian *C. truncatum* populations resulted from at least three founder events, which led to three genetic lineages that spread throughout the country. However, the genetic makeup of these lineages remains unknow, and their extreme geographic proximity raises the question of the maintenance of their genetic integrity in the face of admixture. In order to gain insights into the evolutionary history of *C. truncatum* lineages and to investigate in more details the possibility of a lack of genetic exchanges between them, we employed a population genomic approach. For that, we produced a draft genome sequence of a typical strain of the species associated with soybean anthracnose, which was used as the reference genome. Eighteen representative *C. truncatum* isolates from the three lineages were submitted to whole genome sequencing, aligned against the reference genome, and variants were identified. Our population genomic analyzes revealed that the genetic structure of *C. truncatum* pathogen causing soybean anthracnose is formed by three deeply divergent lineages with levels of genetic diversity consistent with repeated introduction events for each lineage. We also found evidence for sexual recombination within and between lineages, with multiples isolates displaying signatures of admixture. Our findings support a scenario in which the three lineages initially diverged in allopatry before experiencing hybridization following secondary contact. Monitoring of the pathogen's diversity over time is needed to reveal whether these lineages maintain or fuse, which can impact the disease control methods currently employed.

Keywords: Soybean disease; Genetic diversity; Multilocus microsatellite typing; Whole genome sequencing; Admixture; Genetic lineages

1. GENERAL INTRODUCTION

Global food security is one of the most worrying issues for humanity, and efficient, safe agricultural production is critical to achieving this goal (FAO, 2017). The soybean crop is one of the main agricultural crops, with high global economic relevance. Soybean has risen to one of the top-traded commodities in the world, due mainly the wide versatility of use of its grains and byproducts (Hartman, West, and Herman 2011). In 2017, the world soybean production surpassed 300 million tons, of which about 85% was produced in the Americas, mainly in the United States, Brazil and Argentina (USDA 2018). The soybean planting season of 2017–2018 recorded a yield of 118,9 million tons in Brazil, wherein the largest producer states are Mato Grosso (27.2%), Paraná (16.1%), Rio Grande do Sul (15%) and Goiás (10%) (CONAB, 2018).

The large area under soybean cultivation in Brazil, including the incorporation of new growing areas in the northern and midwestern regions, mostly under monoculture and non-tillage system, has been affected the prevalence and the intensity of diseases. Among these, one of the most prominent is anthracnose, which can lead to total crop loss under favorable weather conditions of high temperature and moisture (Manandhar and Hartman 1999; EMBRAPA 2008). With the emergence of Asian rust, anthracnose disease has been underestimated, although losses of 90 kg/ha of soybean grain were reported for each 1% increment in the anthracnose incidence in commercial soybean crops (Dias, Pinheiro, and Café-Filho 2016). Frequent reports on the increase in the importance of anthracnose in the northern and mid-western regions indicate that chemical control programs for fungal diseases in soybean, that is mainly focused on rust, have not been effective against anthracnose. Fungicides are the main control methods of diseases in the soybean, however, little information on the chemical control of anthracnose is available.

Soybean anthracnose is mainly associated with the fungal species *Colletotrichum truncatum* (Hyde et al. 2009), which can also infect important plant species belonging to Fabaceae and Solanaceae families (Cannon et al. 2012; Damm et al. 2009; Weidemann, TeBeest, and Cartwright 1988). Such pathogen was first described by Andrus and Moore (1935) and it was then classified in the genus *Glomerella* in its sexual state. It has synonyms as *Colletotrichum dematium* f. *truncatum* (Schwein.), *Glomerella truncate*, *Colletotrichum capsica,* and *Vermicularia truncata* (Schwein.) in its basionym (Damm et al. 2009, 2019). The fungus might be native of Asia since the first report of a *Colletotrichum* species causing soybean anthracnose came from Korea in 1917 (Tiffany and Gilman 1954). In Brazil, it was first

observed in Rio Grande do Sul in 1961 and it has been reported more frequently in the Brazilian savanna currently (Araújo et al. 1988). Assuming that the center of origin of a pathogenic species should correspond to the center of genetic diversity of the host species (Stukenbrock and McDonald 2008), *C. truncatum* is hypothesized to have originated in China and established in Brazil as an invasive species (Rogério et al. 2019).

Soybean is susceptible to the *C. truncatum* infection at all stages of development, and the symptoms appear on the stems, pods and leaf petioles as irregularly shaped dark-brown spots, which evolve to pod rot, immature opening of pods and to the premature germination of grains (EMBRAPA 2008). The pathogen is generally seedborne, and contaminated seeds seem to be an efficient mechanism of dispersal over long distances, as related to other soybean pathogens (Rogério et al. 2017; Ciampi et al. 2008; Ramiro et al. 2019). Further, it could also be spread by wind and rain and could survive at least 48 months on infected debris in the soil (Cannon et al. 2012; Ranathunge, Mongkolporn, and Ford 2012). Although there are no commercial soybean cultivars fully resistant to anthracnose, some resistant sources have been identified (Manandhar et al. 1988; Khan and Sinclair 1992; Costa et al. 2009).

Knowledge of the genetic structure of pathogen populations, i.e. the amount and distribution of genetic variation within and among populations, can support definition of effective control measures. Genetic distribution in the populations is influenced by evolutionary forces, such as gene flow, genetic drift, mutation, and selection, which shape the evolutionary history of populations over time (McDonald 1997).Thus, access the genetic structure of fungal pathogens can give information about the evolutionary processes that influenced the populations in the past, besides providing insights into their future evolutionary potential, which could be useful to optimize the management of resistance genes and fungicides in agriculture (McDonald and Linde 2002). It is well known that the best strategy to control diseases is genetic resistance of the host. However, the lack of information about genetic variability and evolutionary history of pathogen populations are limiting factors to the establishment of successful breeding programs (Milgroon et al. 2008).

Population genetic analyses can also be used to make inferences about the predominant mode of reproduction, e.g., to assess the relative contributions of asexual and sexual reproduction in pathogen populations (Liu et al. 1996). Reproduction and mating systems affect how gene diversity is distributed within and among individuals in a population (McDonald and Linde 2002). If sexual recombination occurs sporadically, and even in a low proportion, it still could have an important impact on the population genetic structure by increasing the genotypic

diversity. Although asexual reproduction predominates in the majority of plant-pathogenic fungi, many species undergo regular sexual cycles (Milgroom 1996). *Colletotrichum* has been predominantly observed in the asexual state, therefore, sexual reproduction has already been observed in some species (Kimati and Galli 1970; Vaillancourt and Hanau 1991; Damm et al. 2019). For *C. truncatum*, the identity of the sexual stage is still unclear, since it has not been described either in laboratory conditions or in nature (Damm et al. 2009; Hyde et al. 2009; Katoch et al. 2017). High levels of genetic variability in *C. truncatum* from soybean and other hosts were observed in previous studies using various molecular markers (Ford et al. 2004; Katoch et al. 2017; Ranathunge, Mongkolporn, and Ford 2012; Rogério et al. 2017; Sant'anna et al. 2010; Sharma 2009; Vasconcelos et al. 1994). These diversity levels coupled with population genetic signatures of sexual recombination (Diao et al. 2015; Katoch et al. 2017) suggest that sexual recombination may have been occurring in this species, although it has not yet been detected so far.

Recent technological advances in both high-throughput sequencing and computational tools have made it possible to sequence and analyze whole genomes of many plant pathogens (Grünwald, McDonald, and Milgroom 2016). These technologies continue to advance rapidly, and costs have declined to the point that it is becoming affordable the genotyping thousands of single-nucleotide polymorphisms (SNPs) simultaneously or resequence many individuals within a species (Grünwald, McDonald, and Milgroom 2016). The whole genome sequence of many *Colletotrichum* species belonging to different species complexes have been published in recent years, for instance, *C. graminicola* (O'Connell et al. 2012), *C. higginsianum* (Zampounis et al. 2016) , *C. orbiculare* (Gan et al. 2013), *C. fructicola* (Gan et al. 2013), *C. fiorineae* (Baroncelli, Sreenivasaprasad, et al. 2014), *C. sublineola* (Baroncelli, Sanz-Martin, et al. 2014), among others, giving an impetus to the *Colletotrichum* research field.

The large number of genetic markers now available provides fine-grained details regarding processes of divergence, differentiation, and recombination of pathogen populations, enabling more robust inferences compared to studies based on a limited number of genetic markers (Grünwald, McDonald, and Milgroom 2016). The genomic revolution makes it possible to go far beyond traditional population genetics analyses allowing research of the evolutionary processes that shape the pathogens populations in agriculture environments at chromosome level. Such approach has proved useful to provide information on evolutionary history of invasive species and genetic composition of their populations (Estoup and Guillemaud 2010; Gross, Hosoya, and Queloz 2014). Understanding the genetic diversity and population structure of invasive pathogens in introduced areas is crucial to the revelation of

hidden biological features of an organism, to the reconstruction of the course of invasions and to the establishment of effective control measures (Gross, Hosoya, and Queloz 2014).

There is no information regarding the population structure of *C. truncatum* associated with soybean anthracnose in Brazil. Previous studies, mainly conducted in Asian countries, focused on genetic differences among isolates obtained from hosts such as chili and pepper, using SSR or other markers (Diao et al. 2015; Mahmodi et al. 2014; Ranathunge, Mongkolporn, and Ford 2012; Sharma, Pinnaka, and Shenoy 2014). In Brazil, pioneer works in this pathosystem used RAPD markers in order to detect polymorphism levels in few isolates (Sant'anna et al. 2010; Vasconcelos et al. 1994).

A wide distribution of predominant *C. truncatum* haplotypes associated with anthracnose soybean revealed by our previous study (Rogério et al. 2017), suggest the existence of a highly efficient mechanism of pathogen dispersal over long distances in Brazil, which can greatly impact the genetic structure of pathogen populations. Based on that assumption coupled with the levels of diversity and recombination previously reported for this species, the objective of this study was to investigate the genetic structure of *C. truncatum* populations in the two important soybean production areas in Brazil. To achieve this aim the study consisted of the following steps:

- To access the genetic structure of the *C. truncatum* populations through genotyping a large number of individuals, using microsatellite markers **(chapter 2)**.
- To produce a draft genome of a typical strain of the species associated with soybean anthracnose in order to be used as reference genome **(chapter 3)**.
- To use whole genome sequencing and population genomics analyzes to gain insights into the evolutionary history of *C. truncatum* lineages and to investigate in more details the extension of the lack of genetic exchanges between them **(chapter 4**).

REFERENCES

Andrus, C.F. and Moore, W.D. 1935. *Colletotrichum truncatum* (Schw.), n. comb., on garden and lima beans. Phytopathology 25, 121–125.

Araújo, A. G., Caf´e-Filho, A. C., and Cupertino, F. P. 1988. Antracnose da soja na região geoeconômica do Distrito Federal. Fitopatol. Bras. 13:130.

Baroncelli, R., Sanz-Martin, J. M., Rech, G. E., Sukno, S. A., and Thon, M. R. 2014. Draft Genome Sequence of *Colletotrichum sublineola*, a Destructive Pathogen of Cultivated Sorghum. Genome Announc. 2:e00540-14-e00540-14.

Baroncelli, R., Sreenivasaprasad, S., Sukno, S. A., Thon, M. R., and Eric Holub. 2014. Draft Genome Sequence of *Colletotrichum acutatum* Sensu Lato (*Colletotrichum fioriniae*). Genome Announc. 2:1–2.

Cannon, P. F., Damm, U., Johnston, P. R., and Weir, B. S. 2012. *Colletotrichum* - current status and future directions. Stud. Mycol. 73:181–213.

Ciampi, M. B., Meyer, M. C., Costa, M. J. N., Zala, M., McDonald, B. A., and Ceresini, P. C. 2008. Genetic Structure of Populations of *Rhizoctonia solani* Anastomosis Group-1 IA from Soybean in Brazil. Phytopathology. 98:932–941.

CONAB. 2018. Acompanhamento da safra brasileira de grãos, V.7, Safra 2017/18. Brasília, 158p.

Costa, I.F.D., Balardin, R.S., Medeiros, L.A.M., Lenz, G., Gulart, C.A., Zemolin, C.R. and Silva, T.M.B. (2009) Reacão de germoplasma comercial de soja a *Colletotrichum truncatum*. Trop Plant Pathol 34,47–50

Damm, U., Sato, T., Alizadeh, A., Groenewald, J. Z., and Crous, P. W. 2019. The *Colletotrichum dracaenophilum*, *C. magnum* and *C. orchidearum* species complexes. Stud. Mycol. 92:1–46.

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

Diao, Y., Zhang, C., Xu, J., Lin, D., Liu, L., Mtung, O. G., et al. 2015. Genetic differentiation and recombination among geographic populations of the fungal pathogen *Colletotrichum truncatum* from chili peppers in China. Evol. Appl. 8:108–118.

Dias, M. ., Pinheiro, V. F., and Café-Filho, A. . 2016. Impact of anthracnose on the yield of soybean subjected to chemical control in the north region of Brazil. Summa Phytopathol. 42:18–23.

EMBRAPA. 2008. Tecnologia de Produção de Soja – Região Central do Brasil 2009 e 2010. 261 p. Londrina: Embrapa Soja.

Estoup, A., and Guillemaud, T. 2010. Reconstructing routes of invasion using genetic data: Why, how and so what? Mol. Ecol. 19:4113–4130.

FAO, 2017. FAOSTAT. Food and Agriculture Organization of the United Nations, Rome, Italy. Web. http://www.fao.org/faostat/.

Ford, R. A., Banniza, S. B., Photita, W. C., and Taylor, P. W. J. A. 2004. Morphological and molecular discrimination of *Colletotrichum truncatum* causing anthracnose on lentil in Canada. Australas. Plant Pathol. 33:559–560.

Gan, P., Ikeda, K., Irieda, H., Narusaka, M., O'Connell, R. J., Narusaka, Y., et al. 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol. 197:1236–1249.

Gross, A., Hosoya, T., and Queloz, V. 2014. Population structure of the invasive forest pathogen *Hymenoscyphus pseudoalbidus*. Mol. Ecol. 23:2943–2960.

Grünwald, N. J., McDonald, B. A., and Milgroom, M. G. 2016. Population Genomics of Fungal and Oomycete Pathogens. Annu. Rev. Phytopathol. 54:323–346.

Hartman, G. L., West, E. D., and Herman, T. K. 2011. Crops that feed the World 2. Soybean-worldwide production, use, and constraints caused by pathogens and pests. Food Secur. 3:5–17.

Hyde, K. ., Cai, L., McLenzie, E. H. C., Yang, Y. L., Zhang, J. Z., and Prihastuti, H. 2009. *Colletotrichum* : a catalogue of confusion. Fungal Divers. 39:1–17.

Katoch, A., Sharma, P., Padder, B. A., and Sharma, P. N. 2017. Population Structure of *Colletotrichum truncatum* in Himachal Pradesh and Identification of Broad-Spectrum Resistant Sources in *Capsicum*. Agric. Res. 6:296–303.

Khan, M. and Sinclair, J.B. 1992. Pathogenicity of sclerotia- and nonsclerotia-forming isolates of *Colletotrichum truncatum* on soybean plants and roots. Phytopathology 82, 314–319

Kimati, H., and Galli, F. 1970. *Glomerella cingulata* (STONEM.) SPAULD. ET V. SCHRENK. F. SP. *phaseoli* N.F., FASE ASCÓGENA DO AGENTE CAUSAL DA ANTRACNOSE DO FEIJOEIRO. An. da Esc. Super. Agric. "Luiz Queiroz." 27:437.

Liu, Y.-C., Cortesi, P., Double, M. L., MacDonald, W. L., and Milagroom, M. G. 1996. Diversity and Multilocus Genetic Structure in Population of *Cryphonectria parasitica*. Pythopathology. 86:1344–1351.

Mahmodi, F., Kadir, J. B., Puteh, A., Pourdad, S. S., Nasehi, A., and Soleimani, N. 2014. Genetic Diversity and Differentiation of *Colletotrichum* spp. Isolates Associated with Leguminosae Using Multigene Loci , RAPD and ISSR. Plant Pathol. J. 30:10–24.

Manandhar, J.B. and Hartman, G.L. 1999. Anthracnose. In Compendium of Soybean Disease ed. Hartman, G.L., Sinclair, J.B. and Rupe, J.C. pp. 13–14. Saint Paul: American phytopathological Society.

McDonald, B. A. 1997. The Population Genetics of Fungi : Tools and Techniques. Phytopathology. 87:448–453.

McDonald, B. A., and Linde, C. 2002. PATHOGEN POPULATION GENETICS, EVOLUTIONARY POTENTIAL, AND DURABLE RESISTANCE. Annu. Rev. Phytopathol. 40:349–379.

Milgroom, M. G. 1996. RECOMBINATION AND THE MULTILOCUS STRUCTURE OF FUNGAL POPULATIONS. Annu. Rev. Phytopathol. 34:457–477.

Milgroon, M. G., Sotirovski, K., Spica, D., Davis, J. E., Brewer, M. T., Milev§, M., et al. 2008. Clonal population structure of the chestnut blight fungus in expanding ranges in southeastern Europe. Mol. Ecol. 17:4446–4458.

O'Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M. F., et al. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat. Genet. 44:1060–1065.

Ramiro, J., Ciampi-Guillardi, M., Caldas, D. G. G., de Moraes, M. H. D., Barbieri, M. C. G., Pereira, W. V., et al. 2019. Quick and accurate detection of *Sclerotinia sclerotiorum* and *Phomopsis* spp. in soybean seeds using qPCR and seed-soaking method. J. Phytopathol. :1–10.

Ranathunge, N. P., Mongkolporn, O., and Ford, R. 2012. *Colletotrichum truncatum* Pathosystem on *Capsicum* spp : infection , colonization and defence mechanisms. Australasion Plant Pathol. 41:463–473.

Rogério, F., Ciampi-Guillardi, M., Barbieri, M. C. G., Bragança, C. A. D., Seixas, C. D. S., Almeida, A. M. R., et al. 2017. Phylogeny and variability of *Colletotrichum truncatum* associated with soybean anthracnose in Brazil. J. Appl. Microbiol. 122:402–415.

Rogério, F., Gladieux, P., Massola, N. S., and Ciampi-Guillardi, M. 2019. Multiple Introductions Without Admixture of *Colletotrichum truncatum* Associated with Soybean Anthracnose in Brazil. Phytopathology. 109:681–689.

Sant'anna, J. R., Miyamoto, Cláudia, T., Rosada, L. J., Franco, C. C. S., Kaneshima, E. N., and Castro-Prado, M. A. A. 2010. Genetic relatedness of Brazilian *Colletotrichum truncatum* isolates assessed by vegetative compatibility groups and RAPD analysis. Biol. Res. 43:51–62.

Sharma, G., Pinnaka, A. K., and Shenoy, B. D. 2014. Infra-specific diversity of *Colletotrichum truncatum* associated with chilli anthracnose in India based on microsatellite marker analysis. Arch. Phytopathol. Plant Prot. 47:2509–2523.

Sharma, R. 2009. Genetic differentiation of host limited forms of *Colletotrichum truncatum* from northwestern Himalayas. Arch. Phytopathol. Plant Prot. 42:960–966.

Stukenbrock, E. H., and McDonald, B. A. 2008. The Origins of Plant Pathogens in Agro-Ecosystems. Annu. Rev. Phytopathol. 46:75–100.

Tiffany, A. L. H., and Gilman, J. C. 1954. Species of *Colletotrichum* from Legumes. Mycologia. 46:52–75.

USDA. 2018. United States Department of Agriculture: Foreign Agricultural Service. http://apps.fas.usda.gov/Gats/ExpressQuery1.aspx. Acessed in October, 30th, 2018

Vaillancourt, L. J., and Hanau, R. M. 1991. A Method for Genetic Analysis of *Glomerella graminicola (Colletotrichum graminicola)* from Maize. Pythopathology. 81:530–534.

Vasconcelos, M. J. V. de, Machado, M. A., Almeida, Á. M. R., Henning, A. A., Barros, E. G. de, and Moreira, M. A. 1994. DIFFERENTIATION OF *COLLETOTRICHUM TRUNCATUM* ISOLATES BY RANDOM AMPLIFIED POLYMORPHIC DNA. Fitopatol. Bras. 19:520–523.

Weidemann, G. J., TeBeest, D. O., and Cartwright, R. D. 1988. Host Specificity of *Colletotrichum gloeosporioides* f. sp. *aeschynomene* and *C. truncatum* in the Leguminosae. Phytopathology. 78:986–990.

Zampounis, A., Pigné, S., Dallery, J.-F., Wittenberg, A. H. J., Zhou, S., Schwartz, D. C., et al. 2016. Genome Sequence and Annotation of *Colletotrichum higginsianum*, a Causal Agent of Crucifer Anthracnose Disease. Genome Announc. 4:e00821-16.

2. MULTIPLE INTRODUCTIONS WITHOUT ADMIXTURE OF *COLLETOTRICHUM TRUNCATUM* ASSOCIATED WITH SOYBEAN ANTHRACNOSE IN BRAZIL

ABSTRACT

Knowledge of the population structure, genetic diversity and reproductive mode of plant pathogens can help to implement effective disease management strategies. Anthracnose is one of the most prominent diseases in soybean and is mainly associated with the species *Colletotrichum truncatum*. However, the genetic structure of *C. truncatum* populations associated with soybean remains unknown. We collected *C. truncatum* isolates from 10 sites representing two Brazilian states (Mato Grosso and Goiás) and used 13 highly polymorphic microsatellite markers to investigate the population genetic structure of the pathogen. Analyses revealed high gene and haplotypic diversity within populations, as well low genetic differentiation and sharing of multilocus haplotypes among populations and regions. Bayesian and multivariate analysis revealed the presence of three distinct genetic clusters with at least two coexisting in all locations, and all of them coexisting in 8 locations. We found limited evidence for admixture between clusters, with only two isolates showing non-zero membership with a second cluster. Analyses of linkage disequilibrium rejected the hypothesis of random mating in all clusters, but values of the index of association were low and not consistent with long-term lack of sexual reproduction. Our findings suggest that Brazilian *C. truncatum* populations resulted from at least three founder events that led to three genetic clusters that spread throughout the country, raising questions with respect to the factors allowing their maintenance in syntopy without evidence of admixture between them.

Keywords: Founder events; Genetic lineages; Genetic diversity; Microsatellite markers

2.1. INTRODUCTION

Soybean is one of the main agricultural commodities with high global economic relevance. The extensive area of soybean sown in Brazil, including the incorporation of new growing areas in the northern and central-west regions, characterized by genetically uniform crops and non-tillage system, affect the occurrence and the intensity of diseases. Among these, one of the most prominent is anthracnose, as soybeans are susceptible to infection at all stages of development. Soybean anthracnose is a complex system, with many unresolved questions about the identity of the causal agent, the fungicide management and the need to develop a specific integrated control program (Dias et al. 2016). In Brazil, little information on control of anthracnose is available. Increasing of disease reports in northern and central-west regions indicate that chemical control program for fungal diseases in soybean has not been effective against anthracnose. The disease can be an important limiting factor for soybean production in

areas with favorable weather conditions, such as high humidity and temperature, where reports of significant yield reductions are frequent (Dias et al. 2016).

Soybean (*Glycine max* (Linnaeus) Merrill) originates from East Asia, most likely from China, and would have spread to the rest of the world from this area (Hymowitz 1970). Initially used as forage plant, it was introduced into North America in 1765, achieving a prominent role as a grain crop in the 1920s (Hymowitz 1990). In Brazil, the crop was introduced in 1882 in Bahia state and began to be widely cultivated in Rio Grande do Sul state. Production was boosted in the southern states with the beginning of wheat/soybean succession, and it was expanded to the Central region (Brazilian savanna) in the 1960s and 1970s (EMBRAPA 1996). The soybean cultivars sown in those decades were the outcome of breeding programs for the development of better-adapted cultivars to Brazilian conditions, based on cultivars introduced from the South of the United States (Arantes and Miranda 1993; Hirimoto and Vello 1986; Wysmierski and Vello 2013).

Anthracnose is mainly associated with the fungal species *Colletotrichum truncatum* (Hyde et al. 2009), which can also infect important plant species belonging to *Fabaceae* and *Solanaceae* families (Cannon et al. 2012; Damm et al. 2009; Weidemann et al. 1988). The pathogen might be native of Asia since the first report of a *Colletotrichum* species causing soybean anthracnose in Korea (Tiffany 1950). However, a report of soybean anthracnose in Korea in 1917 was attributed to *C. glycines* (Hemmi 1920 cited in Damm et al. 2019). Lehman and Wolf (1926) studied soybean anthracnose in North Carolina (USA) and noticed an asexual morph associated with the disease symptoms, which was identified as *C. glycines,* based on similarity with the pathogen reported in Korea in 1917. Apparently, this strain is identical to strain CBS 195.32, isolated by Lehman and deposited in the CBS collection in 1932, which was later identified as *C. truncatum* by Damm et al. (2009) based on the previous classification done by Andrus and Moore (1935). In Brazil, soybean anthracnose was first observed in Rio Grande do Sul in 1961 and has been reported more frequently in the Brazilian savanna currently (Araújo et al. 1988).

Assuming that the center of origin of pathogenic species should correspond to the center of genetic diversity of the host species (Stukenbrock and McDonald 2008), *C. truncatum* is hypothesized to have originated in China and established in Brazil as an invasive species. Pathogens can be introduced into new areas by several sources, for instance through the spread of agriculture or globalization of travel and trade, allowing extensive movements of crop species and plant products, which could lead to unintentional transport of fungal pathogens far

from their native location (Desprez-Loustau et al. 2007; Gladieux et al. 2008; Yarwood 1970). Fungal invaders may adapt to the new environment after their introduction, and rapid and drastic evolutionary changes may be facilitated by huge population sizes and mixed mating systems, causing yield losses in agrosystems (Gladieux et al. 2015).

Knowledge of the population structure, genetic diversity and sexual recombination of plant pathogens can support definition of effective disease management strategies, for instance by keeping pathogen population sizes small in order to limit gene diversity (McDonald and Linde 2002). Quantifying the amount of gene diversity in pathogen populations can indicate founder populations, which is predicted to have the highest gene diversity levels (Hallatschek and Nelson 2008). Furthermore, population genetics is a useful approach to infer introduction routes of invasive pathogens and dispersal of fungal propagules (Baroncelli et al. 2015; Ciampi-Guillardi et al. 2014; Leo et al. 2015). The inference of historical patterns of pathogen migration among and within regions can be useful for management and attenuation of negative effects caused by invasive pathogens, for understanding current epidemics, for developing predictive models, and for characterizing the genetic basis of pathogen adaptation (Ali et al. 2014; Grünwald et al. 2012; McDonald and Linde 2002; Milgroom and Fry 1997).

Although asexual reproduction predominates in the majority of plant-pathogenic fungi, many species undergo regular sexual cycles (Milgroom 1996). *Colletotrichum* genus has been predominantly observed in the asexual state. As reported for other *Colletotrichum* species, the identity of the sexual stage is still unclear for *C. truncatum* (Damm et al. 2009; Hyde et al. 2009). Reproduction and mating systems affect how gene diversity is distributed within and among individuals in a population (McDonald and Linde 2002). If sexual recombination occurs sporadically, and even in a low proportion it still could have an important impact on the population genetic structure by increasing the genotypic diversity (Milgroom 1996). High levels of genetic variability in *C. truncatum* from soybean and other hosts were observed in previous studies using various molecular markers (Ford et al. 2004; Katoch et al. 2017; Ranathunge et al. 2009; Rogério et al. 2016; Sant'anna et al. 2010; Sharma 2009; Vasconcelos et al. 1994). Evidences of recombination coupled with high levels of genetic diversity were shown in *C. truncatum* populations infecting chilli from India and China (Diao et al. 2015; Sharma et al. 2014).

There is no information regarding the population structure of *C. truncatum* associated with soybean anthracnose in Brazil. Previous studies, mainly conducted in Asian countries, focused on genetic differences among isolates obtained from hosts such as chili and pepper, using SSR or other markers (ISSR, RAPD and nuclear genes), which revealed high genetic

diversity and possible recombination (Diao et al. 2015; Mahmodi et al. 2014; Ranathunge et al. 2009; Sharma et al. 2014). In Brazil, pioneer works in this pathosystem using RAPD markers also evidenced relevant levels of polymorphism in few soybean isolates (Sant'anna et al. 2010; Vasconcelos et al. 1994).

A wide distribution of predominant *C. truncatum* haplotypes associated with soybean anthracnose was revealed by a previous study, suggesting the existence of a highly efficient mechanism of pathogen dispersal over long distances in Brazil (Rogério et al. 2016). Low differentiation among *C. truncatum* populations infecting Capsicum was found in India, which was attributed to extensive movement of infected material among sampled regions, indicating that neither topography nor distance acts a barrier in differentiating them (Katoch et al. 2017).

The objective of this study was to investigate the population structure of *C. truncatum* in the most important soybean production areas in Brazil using microsatellite markers. Based on the assumption of efficient mechanisms of propagules dissemination of *C. truncatum* over long distances and the levels of diversity and recombination previously reported for this species, we hypothesized that the pathogen should exhibit non-differentiated populations across soybean fields and that pathogen populations should have a recombining population structure.

2.2. MATERIALS AND METHODS

2.2.1. Population sampling

We collected symptomatic plant samples in ten naturally infected commercial soybean fields in Mato Grosso (MT) and Goiás (GO) states (Fig. 1). These regions are located in the Brazilian savanna and represent the most important soybean production areas in Brazil. All samples were collected in 2017 but MT2 sample was collected in 2016. In Goiás, samples were collected in Montevidiu and Rio Verde counties (49 km apart). In Mato Grosso, samples were collected in Lucas do Rio Verde and Sinop counties (148 km apart). Stems, leaves and pods showing typical symptoms of anthracnose were collected from each field. To minimize the chance of sampling recently propagated clones of the pathogen, it was maintained a minimal distance of 10 m between the collected plants.

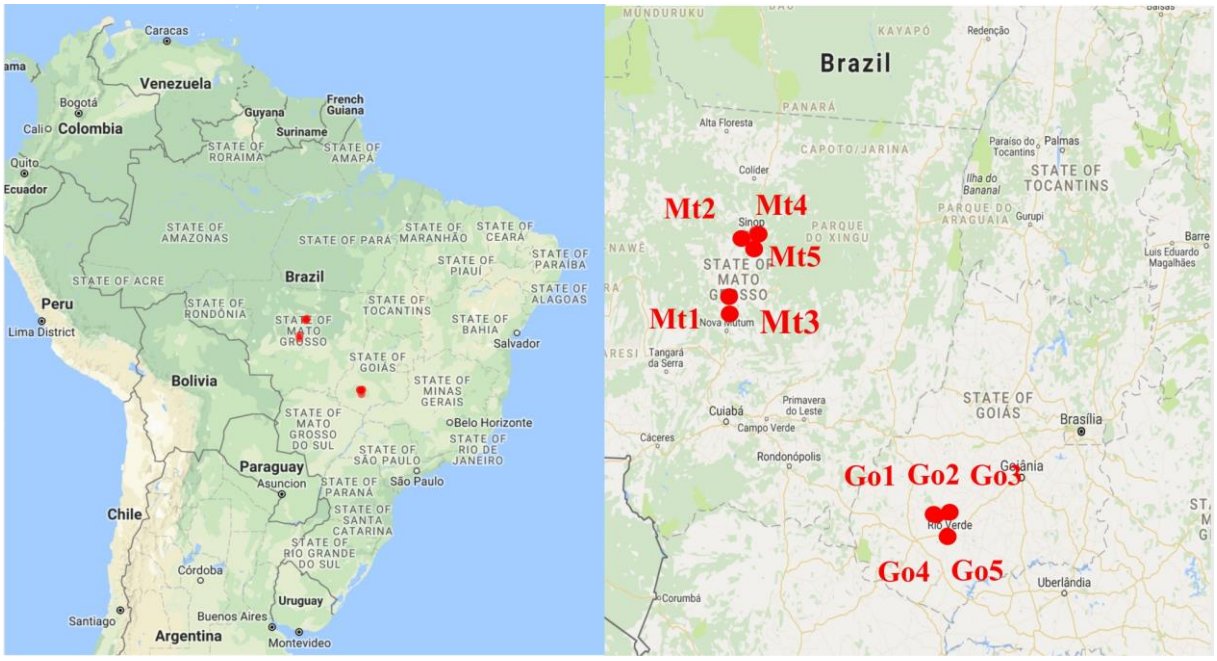


Figure 1. Sampling location of *Colletotrichum truncatum* populations collected from ten soybean fields in Mato Grosso and Goiás States, Brazil.

2.2.2. Isolation of *Colletotrichum truncatum*

Symptomatic stems, leaves and pods were surface disinfected with 70% ethanol for 30 s, then 1% sodium hypochlorite for 1 min, and rinsed in sterile distilled water. Small pieces of leaves and stems were plated onto potato dextrose agar (PDA) medium and incubated at 25°C for 3 days. Pods were kept in Petri dishes onto moistened, sterilized filter paper and checked every day under the microscope to monitor the production of acervuli and conidia. Hyphal tips and acervuli emerging from plated tissues or pods were transferred to PDA medium and cultivated at 25°C. All fungal isolates were single-spored prior to use. Monosporic cultures were preserved on filter paper (Alfenas and Mafia 2007) and in sterile distilled water (Castellani 1939), and kept at 4°C in the Plant Pathology Department of 'Luiz de Queiroz' College of Agriculture, University of São Paulo, Brazil. Each group of fungal isolates collected from the same field was considered as part of the same population.

2.2.3. DNA extraction and SSR genotyping

Fungal isolates were grown for 7 days on PDA at 25°C. Microscopic observations were carried out to certify that the morphological features of the isolates corresponded to *C. truncatum*. Mycelium plugs were collected from the media, ground in sterilized sand, and fungal genomic DNA was extracted using Wizard Genomic DNA Purification Kit (Promega)

according to the manufacturer's recommendations. SSR genotyping was carried out using a set of 13 loci (SSR1, SSR5, SSR10, SSR17, SSR23, SSR29, SSR34, SSR42, SSR44, SSR53, SSR55, SSR56 and SSR59) developed by Ranathunge et al. (2009). Loci were individually amplified by PCR as previously described by the authors. Forward primers were labeled with the fluorescent dyes 6-FAM, VIC®, PET® and NED®. Amplifications were performed in a 15 µl volume containing 25 ng of template DNA, 0.8 mM of fluorescently labeled forward and common reverse primers, 20 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl, 0.15 mM of each dNTP, and 1 U of Taq DNA polymerase in a Techne® thermal cycler (TC-512; Techne Inc., Burlington, NJ) using the following program: 95°C for 2 min, then 30 cycles of 95°C for 1 min, 60°C for 30 s, 72°C for 1 min, and a final extension step of 72°C for 10 min. PCR products were visualized on 2.0% agarose gel and amplicon sizes were initially estimated against known molecular size standards (1 Kb plus ladder, Invitrogen). Amplified microsatellite fragments were electrophoretically separated on an automated DNA sequencer ABI 3500 (Applied Biosystems, Foster City, CA) using GS-600 Liz (Applied Biosystems) as an internal size standard. Raw data were collected, scored, and exported as a fragment size for simple sequence repeats (SSRs) using Genemarker v.1.191 (SoftGenetics). To avoid any error bias, the genotype of each individual was confirmed by two independent repeats, revealing that the amplification patterns were highly reproducible. Statistical binning of the alleles into fragment size categories consistent with the repeat unit increments was performed with FlexiBin software (B. Amos, Cambridge University, UK).

2.2.4. Genetic and haplotypic diversity

The GenAlex v.6.5 software (Peakall and Smouse 2006) was used to identify multilocus microsatellite haplotypes (MLHs). Only one allele was amplified per locus, as expected for a haploid organism. GenAlEx was also used to generate a clone-corrected data set by detecting isolates with identical MLHs within each population and retaining only one representative sample per MLH in the data set. Unbiased gene diversity (Nei 1978) was calculated for each population using the *poppr* package for R (Kamvar et al. 2014). Allelic richness (A_R) was estimated using a rarefaction method (El Mouadik and Petit 1996), which enables comparisons among populations of unequal sample sizes by calculating the expected number of alleles in a subsample of *n* genes corresponding to the sample size of the smallest population, implemented in ADZE (Szpiech and Rosenberg 2008). We estimated haplotypic

diversity as the number of distinct haplotypes/sample size in each fungal population. Clonal fraction was estimated per population as 1 - haplotypic diversity.

2.2.5. Population differentiation

The distribution of genetic variation within and among *C. truncatum* populations was estimated by analysis of molecular variance (AMOVA, Excoffier et al. 1992) based on weighted average *F*-statistic over loci (Weir and Cockerham 1984) and pairwise comparisons of F_{ST}. Analyses were performed in Arlequin v.3.5 software (Excoffier and Lischer 2010), with significance testing by using 1,000 permutations. Genetic differentiation between population pairs was considered significant when *P* value was less than the 5% nominal level.

We used three methods to examine population subdivision. First, we used the model-based Bayesian clustering approach implemented in the STRUCTURE v.2.3 software (Pritchard et al. 2000) to access the number of genetic clusters and estimate admixture. The software uses multilocus haplotype data to define a set of clusters that maximizes linkage equilibrium and to probabilistically assign haplotypes to clusters, without considering information on their geographic origin (Falush et al. 2003). STRUCTURE was run using the admixture ancestry model-based clustering method, allowing mixed ancestry among individuals from different populations (*K*) with correlated allele frequencies among populations. Ten independent runs were performed using 500,000 Markov chain steps after a burn-in period of 100,000 steps, assuming 1 to 10 subpopulations. Second, to confirm the pattern of population subdivision inferred using STRUCTURE, we used an alternative, non-model-based, clustering method: the discriminant analysis of principal components (DAPC). This multivariate method is based on a discriminant analysis on data transformed after a principal component analysis, which seeks to maximize the intergroup component of variation (Jombart et al. 2010). DAPC was carried out using the Adegenet package 1.3-1 (Jombart and Ahmed 2011) in R environment. We retained the first 30 principal components. Ten separate runs of *K*-means were performed and the mean Bayesian information criterion (BIC) value at each value of *K* from 1 to 10 was plotted (Jombart et al. 2010). Third, to visualize ancestry relationships among multilocus haplotypes and populations, while taking into account the possibility of recombination, we built a phylogenetic network based on a matrix of Euclidean distances obtained using the Neighbor-Net method in the SplitsTree software (http://www.splitstree.org/). Splitstree allows visualizing conflicting phylogenetic signals,

indicated by the presence of reticulations in the network and caused by recombination or incomplete lineage sorting.

2.2.6. Reproductive mode

To assess the predominant reproduction strategy of the pathogen, the nonrandom association of alleles between pairs of loci was measured by using the index of association (I_A), a multilocus linkage disequilibrium estimate based on the variance of pairwise distances between individuals, using clone-corrected dataset in the R package poppr. Because the magnitude of I_A depends on the number of loci studied, we also estimated r_D, a measure that is independent of the number of loci (Agapow and Burt 2001). A r_D value close to zero indicates random mating, while significant gametic disequilibrium – and thus higher r_D values – is expected in asexual or inbreeding populations. To examine whether the observed values deviated significantly from the null hypothesis of linkage equilibrium among loci, the observed variance was compared with the expected variance under random mating, which was obtained after reshuffling the alleles within each clone-corrected population data set over 1,000 permutations.

2.3. RESULTS

2.3.1. Summary of microsatellite variability

Ten populations were collected from individual fields, and a total of 237 *C. truncatum* fungal strains were isolated across the sampled soybean fields (Table 1). The sample size ranged from 3 to 53 isolates per population. All SSR markers were polymorphic and the dataset showed only 0.9% of missing data. A total of 108 alleles were found across populations and the number of alleles per locus ranged from 4 (SSR56 locus) to 13 (SSR59 locus), with an average of 8.3 alleles per locus. The haplotype accumulation curve tended to plateau as the number of loci increased, confirming the importance of using all 13 loci for discriminating MLHs (Fig. 2). A total of 129 unique MLHs were detected within the 237 isolates, representing an overall clonal fraction of 46% (Table 2). For each population, the number of distinct MLHs ranged from 3 to 35 and was highly dependent on population sample size (R2 = 0.95, Fig. 3). Population GO3, due to its small sample size (N = 3), was excluded from genetic diversity analyses. Averaged estimates of allelic richness (AR) and Nei's gene diversity (H) were 8.24 and 0.63, respectively.

The GO1 population displayed the highest allelic richness and gene diversity values (4.4 and 0.76, respectively), while the MT3 population showed the smallest ones (2.48 and 0.21, Table 2).

Table 1. Populations of soybean-infecting *Colletotrichum truncatum* evaluated in this study.

State	County	Population	N	Geographical coordinates	Cultivar	Sampling year
Goiás	Montividiu	GO1	15	17°25'45.3" S - 51°10'14.2" W	Bonus	2017
	Montividiu	GO2	13	17°24'43.1" S - 50°57'29.0" W	7739 Monsoy	2017
	Montividiu	GO3*	3	17°25'50.9" S - 51°01'10.9" W	7739 Monsoy	2017
	Rio Verde	GO4	34	17°27'36.2" S - 51°07'10.3" W	PP7200 Macro seed	2017
	Rio Verde	GO5	53	17°45'51.5" S - 51°02'06.9" W	Nidera 5909	2017
Mato Grosso	Lucas do Rio Verde	MT1	13	13°10'10.1" S - 56°04'08.0" W	Y-70 Pioneer	2017
	Sinop	MT2	20	13°18'46.7" S - 56°02'33.4" W	8766 Monsoy	2016
	Lucas do Rio Verde	MT3	28	13°24'39.4" S - 56°04'04.0" W	8372 Monsoy	2017
	Sinop	MT4	24	11°55'22.8" S - 55°37'00.0" W	7709 Nidera	2017
	Sinop	MT5	34	11°58'58.5" S - 55°30'28.6" W	7709 Nidera	2017

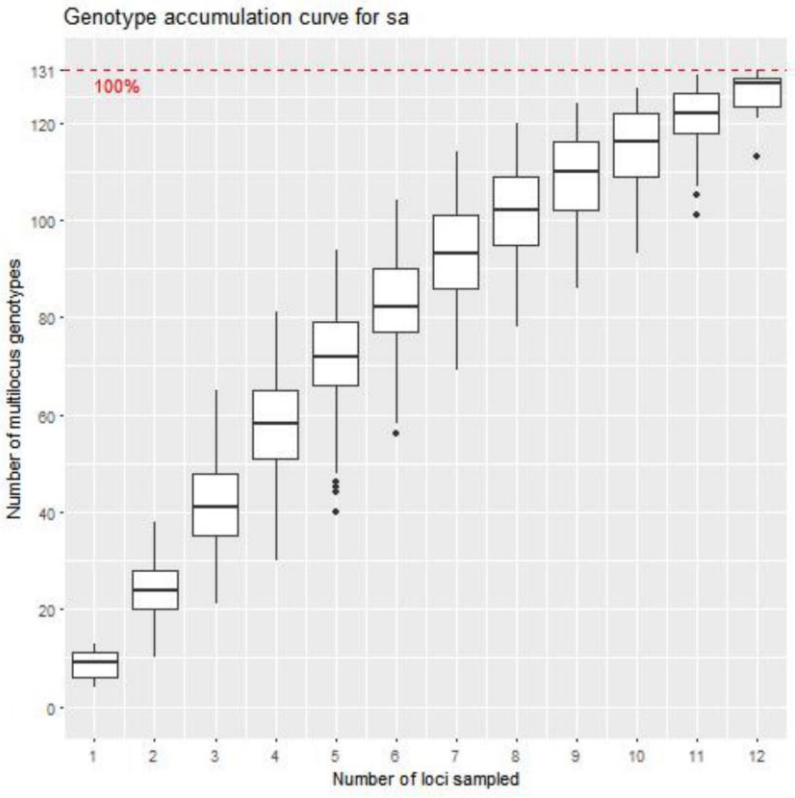


Figure 2. Haplotype accumulation curve for 13 microsatellite loci in ten *Colletotrichum truncatum* populations. Number of observed multilocus haplotypes (MLHs) are denoted by the vertical axis.

Table 2. Estimates of genetic and haplotypic diversity among *Colletotrichum truncatum* populations associated with soybean anthracnose in Brazil.

Population	N[a]	MLHs[b]	Genotypic diversity[c]	Clonal fraction[d]	*H*[e]	N_a[f]	A_R[g]	r_D[h] (*P*)***
GO1	15	13	0.87	0.13	0.76	4.8 (0.3)	4.40	0.33
GO2	13	11	0.85	0.15	0.62	3.9 (0.3)	3.77	0.60
GO3	3	3	1.00	0.00	N/A	N/A	N/A	N/A
GO4	34	18	0.53	0.47	0.45	4.7 (0.5)	3.85	0.47
GO5	53	35	0.66	0.34	0.61	5.7 (0.4)	2.13	0.54
MT1	13	13	1.00	0.00	0.68	3.8 (0.2)	3.58	0.28
MT2	20	16	0.80	0.20	0.59	4.6 (0.3)	3.84	0.44
MT3	28	14	0.50	0.50	0.21	2.9 (0.4)	2.48	0.40
MT4	24	19	0.79	0.21	0.66	5.2 (0.4)	4.12	0.45
MT5	34	24	0.71	0.29	0.64	5.2 (0.4)	3.92	0.31
Total	237	129	0.54	0.46	0.63	4.4 (0.1)	-	-

N/A= Small sample size
[a] Sample size
[b] Number of distinct multilocus haplotypes (MLHs)
[c] Number of distinct genotypes/sample size or fraction of nonrepeated genotypes in the sample
[d] 1 - haplotypic diversity
[e] Nei's unbiased gene diversity (Nei 1978)
[f] Number of distinct alleles averaged across loci and standard deviation in parenthesis
[g] Allelic richness estimated by rarefaction to N = 11 (Szpiech et al. 2008)
[h] Multilocus linkage disequilibrium (Agapow and Burt 2001).
****P* =0.001 for all r_D values (associated probability estimated after 1,000 randomizations)

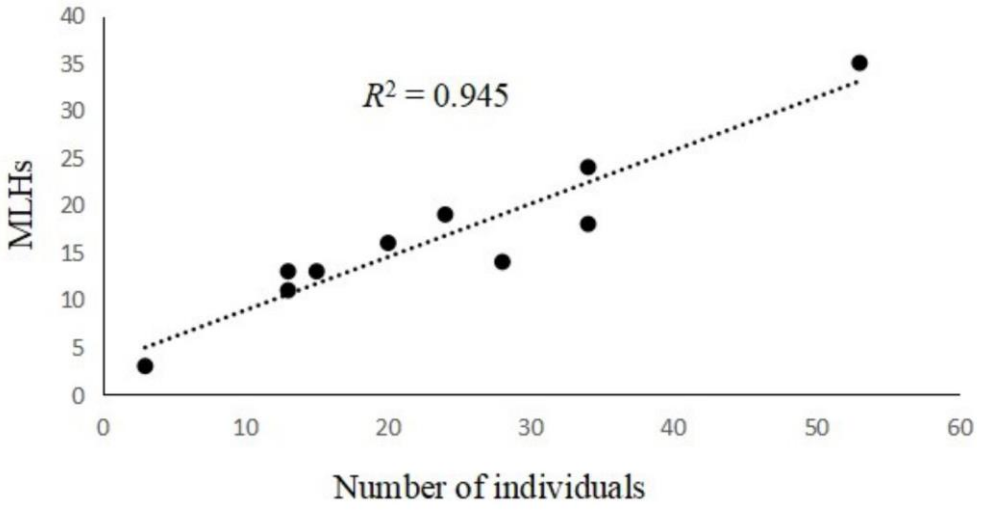


Figure 2. Correlation between sample size and multilocus microsatellite haplotypes (MLHs) in *Colletotrichum truncatum* populations from Brazil.

2.3.2. Population differentiation

Hierarchical AMOVA was performed using a clone-corrected data set in which a single representative of each MLH was kept in each population. When defining samples from different regions as two distinct groups, related to their state of origin, low to moderate differentiation was found (F_{ST} = 0.08; P = 0.05), with 91.51% of genetic diversity distributed within populations and only 3.69% between states. The populations from Mato Grosso were significantly differentiated from other populations in 18 out of 35 pairwise comparisons. The highest differentiation was between MT1 and MT3 (F_{ST} = 0.334, P = 0) and the smallest significant value was between MT5 and GO4 (F_{ST} = 0.07, P = 0.018, Table 3). Population MT3 was significantly differentiated from all populations; F_{ST} estimates between this population and others tended to be the highest ones. The populations sampled in Goiás were not significantly differentiated from each other, and they showed significant differentiation only in comparisons with populations from Mato Grosso.

Patterns of shared MLHs between populations were consistent with estimates of population differentiation. Of 129 MLHs detected, 25 were shared both within the same population, within the same state and even between different states. Results showed five cases of haplotypes sharing among the sampled states, always involving some GO populations (sampled in Goiás in 2017) and the MT2 population (sampled in Mato Grosso in 2016, Table S1). The MT3 population shared haplotypes only within other populations from Mato Grosso state.

We used the clustering method implemented in STRUCTURE, assuming a model with admixture and correlated allele frequencies, to investigate the number of genetic clusters represented in our dataset. Patterns of clustering and likelihood values were highly similar across repeats at each K value, indicating convergence of the MCMC toward single clustering solutions. The model with K = 3 had both the highest likelihood [Pr(K)] and Evanno's ΔK, suggesting this model captures the most salient features of population subdivision in this system (Fig. 4B). At K = 2 and K = 3, differentiation between clusters was clear, with only limited admixture among groups (only 2 haplotypes with maximum membership proportion in a single cluster below 98%, Fig. 4A). The three clusters coexisted in eight locations; in the remaining two locations, only two clusters were found (Fig. 5). K > 4 did not reveal any further biologically relevant structure and mostly produced genotypes with intermediate membership probabilities, thus only the membership coefficients for runs with $2 \leq K \leq 6$ are presented. A

similar pattern of clustering was obtained with the non-parametric DAPC analysis and the network obtained using the Neighbor-Net method (Fig. 4C and Fig. 6).

When AMOVA was performed by grouping haplotypes according to the inferred clusters, 51 % of genetic diversity was distributed among clusters and only 10.3% among geographical regions within clusters. However, when we partitioned the variation according to geographical origin, nested within the inferred clusters, there was no difference between regions and 78.26% of the diversity was found among clusters within regions. The highest pairwise differentiation was observed among clusters within regions (Table 4), which indicates that the genetic variation was mostly distributed between clusters and not between geographic regions within clusters.

Genetic diversity was estimated for both samples (Table 2) and inferred genetic clusters (Table 5). In the clone-corrected dataset of 166 MLHs, 131 unique haplotypes were detected across the three clusters, not associated with the geographic origin of isolates. The first cluster (C1), composed of 81 haplotypes, encompassed the highest proportion of repeated haplotypes (clonal fraction = 0.37) and was the least diverse (H = 0.24; Na = 3.85; AR = 2.51). The second cluster (C2) showed 26 isolates with intermediate level of diversity (clonal fraction = 0.19; H = 0.34; N_a = 2.54; A_R = 2.54). The 59 isolates assigned to the third cluster (C3) were all unique MLHs, making this cluster the most diverse one (H = 0.70; N_a = 5.46; A_R = 6.65).

Table 3. Pairwise population differentiation among *Colletotrichum truncatum* associated to soybean in Brazil.

Populations	Populations									
	MT1	MT2	MT3	MT4	MT5	GO5	GO4	GO3	GO2	GO1
MT1	-									
MT2	0.087*	-								
MT3	0.334*	0.205*	-							
MT4	0.006	0.042	0.182*	-						
MT5	0.038	0.042	0.173*	-0.007	-					
GO5	0.157*	0.009	0.161*	0.086*	0.079*	-				
GO4	0.155*	-0.003	0.164*	0.075*	0.067*	0.016	-			
GO3	0	-0.089	0.298*	-0.077	-0.052	0.056	0.048	-		
GO2	0.148*	-0.031	0.163*	0.062	0.059	0.003	0.041	0.044	-	
GO1	0.085*	0.037	0.282*	0.045	0.030	0.036	0.018	0.126	0.053	-

F_{ST} values are presented below the diagonal
*Significant genetic differentiation ($P \leq 0.05$) in pairwise comparisons based on 1.000 permutation

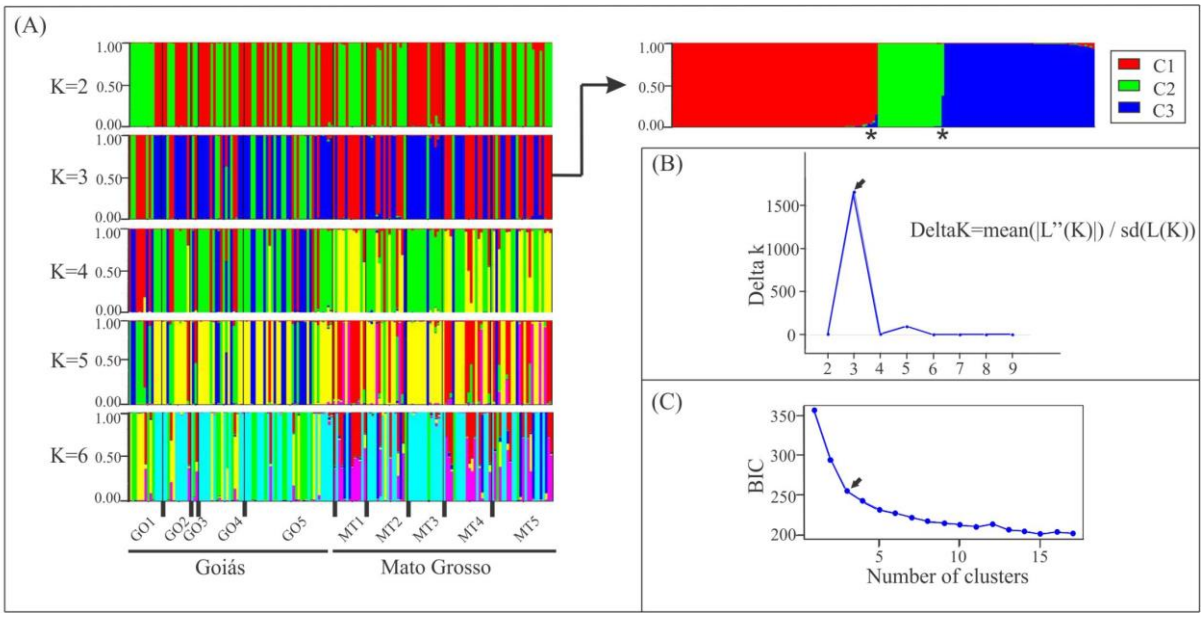


Figure 4. Population subdivision of *Colletotrichum truncatum* associated with soybean anthracnose in Brazil. (A) STRUCTURE Barplots showing patterns of clustering in the data. Haplotypes with maximum membership proportion to a single cluster below 98% are denoted by asterisks. (B) The most likely number of clusters (*K*) based on Evanno's Δ*K* method. (C) Bayesian information criteria (BIC) indicating the most probable number of genetic groups by DAPC analysis.

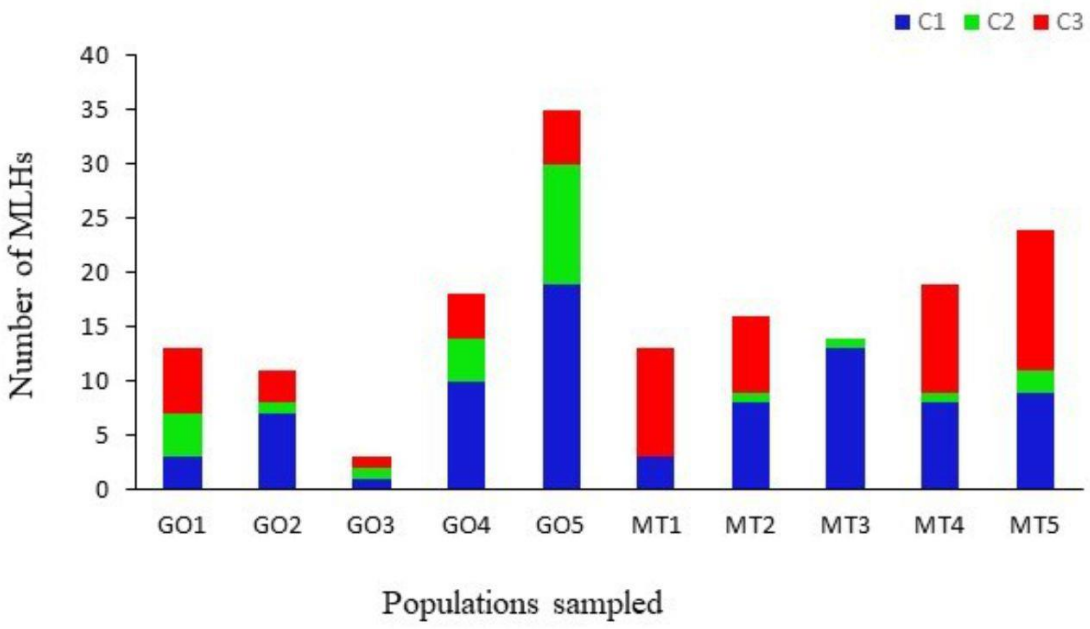


Figure 5. Coexistence of genetic clusters (C1, C2 and C3) in each *Colletotrichum truncatum* population.

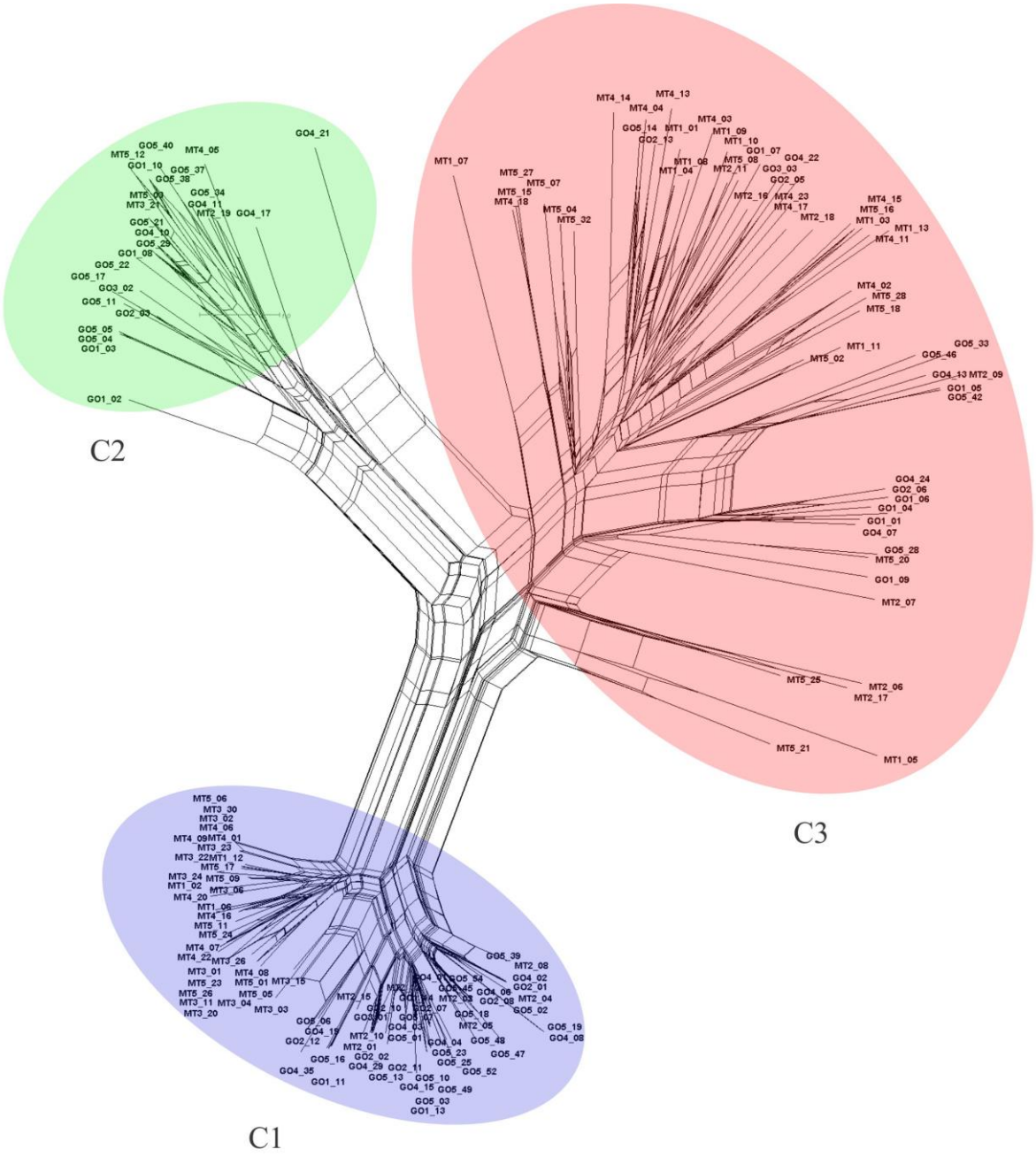


Figure 6. Phylogenetic network using the Neighbor-Net methods of *Colletotrichum truncatum* isolates used in this study. Reticulation indicates likely occurrence of recombination.

Table 4. Pairwise population differentiation among genetic clusters (C1, C2, C3) and geographic regions (GO, MT) of *Colletotrichum truncatum* associated with soybean in Brazil.

Populations	Populations					
	C1 - GO	C1 - MT	C2 - GO	C2 - MT	C3 - GO	C3 - MT
C1 - GO	-					
C1 - MT	0.314	-				
C2 - GO	0.741	0.709	-			
C2 - MT	0.802	0.782	0.163	-		
C3 - GO	0.666	0.649	0.541	0.560	-	
C3 - MT	0.570	0.557	0.459	0.480	0.158	-

F_{ST} values are presented below the diagonal
All pairwise comparisons based on 1.000 permutations had significant genetic differentiation ($P \leq 0.05$)

Table 5. Estimates of genetic and genotypic diversity of the genetic clusters of *Colletotrichum truncatum* associated to soybean in Brazil.

Clusters	Code	N[a]	MLHs[b]	Genotypic diversity[c]	Clonal fraction[d]	H^e	N_a[f]	A_R[g]	r_D[h] (*P*)
Cluster 1	C1	81	51	0.63	0.37	0.24	3.85 (0.66)	2.51	0.03 (0.003)
Cluster 2	C2	26	21	0.81	0.19	0.34	2.54 (0.31)	2.54	0.08 (0.001)
Cluster 3	C3	59	59	1.00	0.00	0.61	5.46 (0.72)	4.65	0.07 (0.001)
Total	-	166	131	0.79	0.21	0.70	3.95 (0.39)	8.31	0.39 (0.001)

[a] Sample size
[b] Number of distinct multilocus haplotypes (MLHs)
[c] Number of distinct genotypes/sample size or fraction of nonrepeated genotypes in the sample
[d] 1 - haplotypic diversity
[e] Nei's unbiased gene diversity (Nei 1978)
[f] Number of distinct alleles averaged across loci and standard deviation in parenthesis
[g] Allelic richness estimated by rarefaction to N = 21 (Szpiech et al. 2008)
[h] Multilocus linkage disequilibrium (Agapow and Burt 2001).
P = associated probability estimated after 1,000 randomizations

2.3.3. Reproductive mode

Levels of multilocus linkage disequilibrium r_D were low and consistent with recombination within clusters (r_D ranging from 0.03 to 0.07), but the null hypothesis ($r_D = 0$) for a random association of alleles between pairs of loci was rejected for all clusters (Table 5). Multilocus linkage disequilibrium was higher within populations (r_D ranging from 0.28 to 0.60; Table 2) than within clusters, consistent with the coexistence of multiple clusters within populations (Fig. 4).

2.4. DISCUSSION

This study represents the first investigation of *C. truncatum* population structure associated with soybean anthracnose in Brazil. We used SSRs markers developed for the synonym species *C. capsici* (Damm et al. 2009) that have already proved to be highly efficient in the characterization of *C. truncatum* populations (Diao et al. 2015). This set of markers showed sufficient power to discriminate haplotypes and detect polymorphisms in this species. Soybean anthracnose is a complex system, with many unresolved issues. Our previous study, based on *C. truncatum* isolates obtained from different soybean-producing regions of Brazil, showed moderate genetic variability and sharing of haplotypes among areas (Rogério et al. 2016). The current study has focused on a population-level distribution of genetic diversity across pathogen populations sampled from central-west (Mato Grosso and Goiás States), the main soybean producing areas in Brazil. We tested the hypotheses that geographical populations of *C. truncatum* are genetically homogeneous (hence not subdivided or differentiated) and that these populations have a recombining population structure.

The *C. truncatum* isolates obtained in this study showed relatively high levels of genetic variability, which would be unexpected for a recently introduced pathogen. However, multiple introductions of genetically diverse plant pathogens could increase the level of genetic variation of local populations (Ciampi-Guillardi et al. 2014; Leo et al. 2015). *C. truncatum* populations were also characterized by moderate to high haplotypic diversity and low to moderate clonal fractions, with some MLHs shared among areas separated by more than 800 km. The high genetic diversity observed in Brazilian populations of *C. truncatum* is in agreement with previous reports for populations of *C. truncatum* isolated from chili in China and India, using the same set of markers (Diao et al. 2015; Sharma et al. 2014). High levels of genetic variability in *C. truncatum* were also observed in previous studies using different molecular markers (Ford et al. 2004; Katoch et al. 2017; Mahmodi et al. 2013; Rogério et al. 2016; Sant'anna et al. 2010; Sharma 2009; Vasconcelos et al. 1994).

Despite the fact that low population differentiation was detected between regions, clustering and molecular variance analyses indicated the existence of three strongly differentiated genetic clusters within each region. This finding is consistent with a possible history of introduction of *C. truncatum* from at least three distinct source populations, leading to three genetic clusters that spread throughout the country. The majority of soybean cultivars currently sown in Brazil are derived of cultivars introduced from U.S. that were used in breeding programs (Wysmierski and Vello 2013). As *C. truncatum* is generally seed-borne

(Manandhar and Hartman 1999), it could be easily introduced in Brazil by infected seeds from U.S. Our previous work corroborates the hypothesis of three distinct source populations (Rogério et al. 2017). Sequencing data from a historical collection of *C. truncatum* isolates from different soybean-producing regions, revealed three genetic clades. A first clade encompassed *C. truncatum* strains isolated from soybean in U.S, which grouped together with the most widely distributed haplotype within that collection. Interestingly, this clade included strain CBS19532, which is apparently the same one regarded as the ex-holotype of *Glomerella glycines* (ATCC 1936) by Lehman and Wolf in 1926 (Damm et al. 2019). This strain also corresponds to drawings and description of the asexual morph of *G. glycines* (*C. glycines*), later classified as *C. truncatum* by Damm et al. (2009). It is possible that this strain has a connection with the *C. glycines* Hori, previously identified at Korea by Hemmi in 1917, as described by Tiffany (1950) as the first causal agent of anthracnose disease in soybean. A second clade enclosed *C. truncatum* sequences from lima bean in Brazil and U.S. that is also a host susceptible to the species, as some other Fabaceae (Carvalho et al. 2015; Damm et al. 2009). When soybeans were introduced in Brazil and became widely planted, other hosts as lima bean might have already established in the cropping system (Silva et al. 2017), and due to broad host range of the fungus, that lineage might be able to maintain itself on this new host. This hypothesis had already been considered for the widespread occurrence of *C. truncatum*, originally from lima beans, on soybeans in the United States (Tiffany 1950). The third clade included strains from hosts as weeds and other leguminous plants. The clade that tends to be exclusively associated with soybean tends to be the less variable, while the cluster associated with soybean and weeds was the most variable. These observations suggest that future studies on the origin of invasive *C. truncatum* populations will need to cover multiple geographic regions and multiple hosts.

Although the three clusters detected in Brazilian *C. truncatum* were found to coexist in the majority of populations, admixture between clusters was not detected, indicating the existence of barriers to gene flow between them. In contrast with our initial hypothesis of geographical populations of *C. truncatum* being genetically homogeneous due to intense movement of infected material among fields, our results do not indicate a strong impact of intra-regional gene flow on patterns of genetic structure. Our findings do suggest, however, that haplotype flows have played a greater role in structuring populations mainly in the maintained the genetic distribution of the three clusters within fields. Differentiation between regions (Mato Grosso and Goiás) was low to moderate, a substantial proportion of MLHs were shared among populations across regions and consecutive years, and levels of variability in the two regions

were similar, suggesting extensive transportation of the pathogen across regions, possibly via contaminated seed (Ciampi et al. 2008). The fact that populations from Mato Grosso showed higher differentiation among themselves than observed in Goiás suggests a greater impact of founder events in the former region. This observation is not consistent with strict host-tracking of the pathogen during the spread of soybean cultivation in Brazil, which began in the south and extended towards the Cerrado, reaching Mato Grosso first (in the late 80's) and only later in Goiás (Fearnside 2001).

Colletotrichum species have been predominantly observed in the asexual or vegetative state. Only a small number of species has been associated with the teleomorph, *Glomerella*, and in general, these forms are rarely observed in nature, with a few notable exceptions (Barcelos et al. 2014; Damm, et al. 2012). The genus *Colletotrichum* has an unusual mating system and to date all the strains sampled possess only one idiomorph (Barcelos et al. 2014; Chen et al. 2002; Crouch et al. 2008). For *C. truncatum*, the identity of the sexual stage is still unclear, since it has not been described either in laboratory conditions or in nature (Damm et al. 2009; Hyde et al. 2009; Katoch et al. 2017). A recent study found population genetic evidence for sexual recombination in *C. truncatum* populations infecting chili from China (Diao et al. 2015), which supported our initial hypothesis that *C. truncatum* infecting soybean in Brazil may be recombinant in structure. However, this hypothesis was not fully corroborated by our population genetic analyses, as estimates of multilocus linkage disequilibrium were low in all three clusters, but the hypothesis of random association of alleles was still rejected. The limited evidence for admixture between clusters is also consistent with lack of recombination resulted from sexual reproduction. The moderate to high levels of haplotypic diversity detected in clonal pathogens could be generated by mechanisms such as extensive chromosome rearrangements mediated by transposable elements (Stukenbrock 2016). A population genomic approach could make it possible to resolve the reproductive mode of the three Brazilian clusters of *C. truncatum*, by giving access to the mating types, their frequencies, and genomic rates of recombination.

This study is the first to use SSR markers and analyze the population structure of *C. truncatum* isolates infecting soybean in Brazil. The results obtained here provided novel insights about the pathogen introduction in Brazil and their implication in distribution of genetic variation across populations from two important areas of soybean production. The current genetic structure of the pathogen reflects possible multiple recent introduction events, represented by three genetic groups widely distributed across soybean fields, raising questions

with respect to the factors allowing their maintenance in syntopy without evidence of admixture between them. These observations can have a direct impact on disease management strategies. Although there is currently no breeding program aiming anthracnose resistance for soybean, future breeding efforts should take into account the current population structure and the genetic diversity levels of the pathogen, for instance by using isolates that are representative of the genetic variability of the species in Brazil for screening resistant varieties. The possible lack of sexual reproduction is also important from the plant pathology point of view, as it may contribute to prevent admixture between clusters and the emergence of recombinant haplotypes with increased virulence. Population genomics studies using higher resolution molecular markers such as single nucleotide polymorphisms (SNPs) may improve the discrimination of clonal lineages, the detection of recombination events in *C. truncatum* populations in Brazil and can provide further insight into the history and population biology of this important plant pathogen.

REFERENCES

Agapow, P.M., and Burt, A. 2001. Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1: 101-102.

Ali, S., Gladieux, P., Leconte, M., Gautier, A., Justesen, A.F., Hovmoller, M.S., Enjalbert, J., and 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 Pathog. 10: e1003903

Alfenas, A.C., and Mafia, R.G. 2007. Armazenamento de Microorganismos em Cultura com Ênfase em Fungos Fitopatogênicos. Pages 91-102 in: Métodos em Fitopatologia ed. Gonçalves, R.C., Alfenas, A.C. and Mafia, R.G. ed. UFV, Viçosa.

Andrus, C.F., and Moore, W.D. 1935. *Colletotrichum truncatum* (Schw.), n. comb., on garden and lima beans. Phytopathology. 25: 21–125.

Arantes, N.E., Miranda and M.A.C. 1993. Melhoramento Genético e cultivares de soja para o cerrado da região sudeste do Brazil. Pages 225-266 in: Cultura da soja nos cerrados. Arantes, N.E., and Souza, P.I.M. ed. Potafos, Piracicaba.

Araújo, A.G., Café-Filho, A.C., and Cupertino, F.P. 1988. Antracnose da soja na região geoeconômica do Distrito Federal. Fitopatol. Bras. 13: 130.

Barcelos, Q.L., Pinto, J.M.A., Vaillancourt, L.J., and Souza. E.A. 2014. Characterization of *Glomerella* Strains Recovered from Anthracnose Lesions on Common Bean Plants in Brazil. Plos One 9(3): e90910.

Baroncelli, R., Zapparata, A., Sarrocco, S., Sukno, S.A., Lane, C.R., Thon, M.R., Vannacci, G. Holub, E., Screenivasaprasad, S. 2015. Molecular diversity of anthracnose pathogen populations associated with UK strawberry production suggests multiple introductions of three different *Colletotrichum* species. PLoS ONE 10: e0129140.

Cannon, P.F., Damm, U., Johnston, P.R. and Weir, B.S. 2012. *Colletotrichum* – current status and future directions. Stud. Mycol. 73: 181–213.

Carvalho, E.M.S., Beserra, J.E.A., Barguil, B.M., 2015. Lima Bean Diseases. In: Lopes ACA, Ferreira ASA, Gomes RLF (eds) *Phaseolus lunatus*: Diversity, Growth and Production. Pages 113–13. Nova Science Publishers, New York.

Castellani, A. 1939. Viability of some pathogenic fungi in distilled water. Am. J. Trop. Med. Hyg. 24: 270–276.

Chen, F., Goodwin, P.H., Khan, A., and Hsiang, T. 2002. Population structure and mating-type genes of *Colletotrichum graminicola* from *Agrostis palustris.* Can. J. Microbiol. 48: 427-436.

Ciampi, M.B., Meyer, M.C., Costa, M.J.N., Zala, M., McDonald, B.A., and Ceresini, P.C. 2008. Genetic structure of populations of *Rhizoctonia solani* anastomosis group-1 IA from soybean in Brazil. Phytopathology 98: 932-941.

Ciampi-Guillardi, M., Baldauf, C., Souza, A.P., Silva-Junior, G.J., and Amorim, L. 2014. Recent introduction and recombination in *Colletotrichum acutatum* populations associated with citrus postbloom fruit drop epidemics in São Paulo, Brazil. Phytopathology 104: 769-778.

Crouch, J.A., Glasheen, B.M., Giunta, M.A., Clarke, B.B., and Hillman, B.I. 2008. The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an ‘‘asexual’’ pathogen. Fungal Gen. Biol. 45: 190-206.

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

Damm, U., Cannon, P.F., Woudenberg, J.H.C., and Crous, P.W. 2012. The *Colletotrichum acutatum* species complex. Stud. Mycol. 73: 37-113.

Damm, U., Sato, T., Alizadeh, A., Groenewald, J.Z., and Crous, P.W. 2019. The *Colletotrichum dracaenophilum, C. magnum* and *C. orchidearum* species complexes. Stud. Mycol. 92: 1–46.

Desprez-Loustau M.L., Robin, C. Buee, M., Courtecuisse, R., Garbaye, J., Suffert, F., Sache, I., and Rizzo, D.M. 2007. The fungal dimension of biological invasions. Trends Ecol. Evol. 22: 472–480.

Diao, Y., Zhang, C., Xu, J., Lin, D., Liu, L., Mtung'e, O.G., and Liu, X. 2015. Genetic differentiation and recombination among geographic populations of the fungal pathogen *Colletotrichum truncatum* from chili peppers in China. Evol. Appl. 8: 108-118.

Dias, M.D., Pinheiro, V.F., and Café-Filho, A.C. 2016. Impact of anthracnose on the yield of soybean subjected to chemical control in the north region of Brazil. Summa Phytopathol. 42: 18-23.

El Mousadik, A., and Petit, R. J. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theor. Appl. Genet. 92: 832-839.

EMBRAPA, 1996. Cultura da soja. edited by Ministério da Agricultura Pecuária e Abastecimento: Empresa Brasileira de Pesquisa Agropecuária - Embrapa Soja, Embrapa Cerrados, Embrapa Agropecuária Oeste, Brazil.

Excoffier, L., Smouse, P., and Quattro, J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131: 479-491.

Excoffier, L., and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10: 564-567.

Falush, D., Stephens, M., and Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Fearnside, P.M. 2001. Soybean cultivation as a threat to the environment in Brazil. Env. Conserv. 28: 23–38.

Ford, R., Banninza, C., Photita, W., and Taylor, P.W.J. 2004. Morphological and molecular discrimination of *Colletotrichum truncatum* causing anthracnose on lentil in Canada. Australas. Plant Pathol. 33: 559–569.

Gladieux, P., Feurtey, A., Hood, M.E., Snirc, A., Clavel, J., Dutech, C., Roy, and M., Giraud, T. 2015. The population biology of fungal invasions. Mol. Ecol. 24: 1969–1986.

Gladieux, P., Zhang, X-G., Afoufa-Bastien, D., Sanhueza, R.M.V., Sbaghi, M., Cam, B.L. 2008. On the origin and spread of the scab disease of apple: out of central Asia. PLoS ONE 3(1): e1455.

Grünwald, N.J., Garbelotto, M., Goss, E.M., Heungens, K., and Prospero, S. 2012. Emergence of the sudden oak death pathogen *Phytophthora ramorum*. Trends Microbiol. 20: 131-138.

Hallatschek, O., Nelson, D.R. 2008. Gene surfing in expanding populations. Theor Popul Biol 73: 158–170.

Hirimoto, D.M., and N.A. Vello. 1986. The genetic base of Brazilian soybean cultivars. Braz. J. Genet. 9: 295–306.

Hyde, K.D., Cai, L., McKenzie, E.H.C., Yang, Y.L., Zhang, J.Z., and Prihastuti, H. 2009. *Colletotrichum*, a catalogue of confusion. Fungal Divers. 29: 1–17.

Hymowitz, T. 1970. On the domestication of the soybean. Econ. Bot. 24: 408-421.

Hymowitz, T. 1990. Soybeans: The success story. Pages. 159-163. In: Advances in new crops. J. Janick and J.E. Simon, eds. Timber Press, Portland.

Jombart, T., and Ahmed I. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27: 3070-3071.

Jombart, T., Devillard, S., and Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11: 95.

Kamvar, Z.N., Tabima, J.F., and Grunwald, N.J. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2: e281.

Katoch, A., Sharma, P., Padder, B.A., and Sharma, P.N. 2017. Population Structure of *Colletotrichum truncatum* in Himachal Pradesh and Identification of Broad-Spectrum Resistant Sources in Capsicum. Agric. Res. 6: 96–303.

Lehman, S.G., and Wolf, F.A. 1926. Soybean anthracnose. J. Environ. Agric. Res. 33: 381–390.

Leo, A.E., Ford, R., Linde, C.C. 2015. Genetic homogeneity of a recently introduced pathogen of chickpea, *Ascochyta rabiei*, to Australia. Biol Invasions 17: 609–623.

Mahmodi, F., Kadir, J., and Puteh, A., Pourdad, S.S., Nasehi, A., Soleimani, N. 2014. Genetic Diversity and Differentiation of *Colletotrichum* spp. Isolates Associated with Leguminosae Using Multigene Loci, RAPD and ISSR. Plant Pathol. J. 30: 10–24.

Manandhar, J.B., and Hartman, G.L. 1999. Anthracnose. Pages 13-14 in: Compendium of Soybean Diseases. G. L. Hartman, J. B. Sinclair, and J. C. Rupe, eds. American Phytopathological Society, St. Paul, MN.

McDonald, B.A., and Linde, C. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. 40: 349–79.

Milgroom, M.G. 1996. Recombination and the multilocus structure of fungal populations. Annu. Rev. Phytopathol. 34: 457-477.

Milgroom, M.G., and Fry, W.E. 1997. Contributions of population genetics to plant disease epidemiology and management. Adv. Bot. Res. 24: 1-30.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a number of individuals. Genetics 89: 538-590.

Peakall, R., and Smouse, P.E. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6: 288-295.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Ranathunge, N.P., Ford, R., and Taylor, P. W. J. 2009. Development and optimization of sequence-tagged microsatellite site markers to detect genetic diversity within *Colletotrichum capsici*, a causal agent of chilli pepper anthracnose disease. Mol. Ecol. Resour. 9: 1175-1179.

Rogério, F., Ciampi–Guillardi, M., Barbieri, M.C., Bragança, C.A.D., Seixas, C.D.S., Almeida, A.M.R., and Massola Junior, N.S. 2016. Phylogeny and variability of *Colletotrichum truncatum* associated to soybean anthracnose in Brazil. J. Appl. Microbiol. 122: 402-415.

Sant'anna, J.R., Miyamoto, C.T., Rosada, L.J., Franco, C.C.S., Kaneshima, E.N. and Castro-Prado, M.A.A. 2010. Genetic relatedness of Brazilian *Colletotrichum truncatum* isolates assessed by vegetative compatibility groups and RAPD analysis. Biol. Res. 43: 51–62.

Sharma, G., Pinnaka, A.K., and Shenoy, B.D. 2014. Infra-specific diversity of *Colletotrichum truncatum* associated with chilli anthracnose in India based on microsatellite marker analysis. Arch. Phytopathology Plant Protect 40: 2509–2523.

Sharma, R. 2009. Genetic differentiation of host limited forms of *Colletotrichum truncatum* from northwestern Himalayas. Arch Phytopathology Plant Protect 42: 960–966.

Silva, R.N.O., Burle, M.L., Pádua, J.G., Lopes, A.C.A., Gomes, R.L.F., and Castillo, J.M. 2017. Phenotypic diversity in lima bean landraces cultivated in Brazil, using the Ward-MLM strategy. Chilean J. Agric. Res. 77: 35-40.

Stukenbrock, E.H. 2016 The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology 106: 104–112.

Stukenbrock, E.H., and McDonald, B.A. 2008. The origins of plant pathogens in agro-ecosystems. Annu Rev Phytopathol. 46: 75–100.

Szpiech, Z.A., Jakobsson, M., and Rosenberg, N.A. 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24: 2498–2504.

Tiffany, L.H. 1950. The anthracnose complex on soybeans. Retrospective Theses and Dissertations. Iowa State University. UMI Number: DP12218. 76p.

Vasconcelos, M.J.V., Machado, M.A., Almeida, A.M.R., Henning, A.A., Barros, E.G. and Moreira, M.A. 1994. Differentiation of *Colletotrichum truncatum* isolates by random amplified polymorphic DNA. Fitopatol. Bras. 19: 520–523.

Weidemann, G.J., Tebeest, D.O. and Cartwright, R.D. 1988. Host specificity of *Colletotrichum gloeosporioides* f. sp. *aeschynoemene* and *C. truncatum* in the Leguminosae. Phytopathology 78: 986–990.

Weir, B.S. and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Wysmierski, P.T., and Vello, N.A. 2013. The genetic base of Brazilian soybean cultivars: evolution over time and breeding implications. Gen. Mol. Biol. 36: 547-555.

Yarwood, C.E .1970. Man-Made Plant Diseases. Science 168: 218–220

SUPPORTING INFORMATION

Table S1. Repeated matching multilocus haplotypes (MLHs) among *Colletotrichum truncatum* populations isolated from soybean in Brazil, based on 13 SSR loci.

MLH	Sample	Pop	SSR01	SSR05	SSR10	SSR17	SSR23	SSR29	SSR34	SSR42	SSR44	SSR53	SSR55	SSR56	SSR59
MLH01	GO5-22	GO5	122	138	127	151	97	199	148	149	173	134	139	148	125
	GO5-43	GO5	122	138	127	151	97	199	148	149	173	134	139	148	125
	GO5-50	GO5	122	138	127	151	97	199	148	149	173	134	139	148	125
MLH02	GO5-11	GO5	122	138	127	151	97	199	148	149	173	134	139	154	125
	GO5-35	GO5	122	138	127	151	97	199	148	149	173	134	139	154	125
	MT2-19	MT2	122	138	127	151	97	199	148	149	173	134	139	154	125
MLH03	GO2-03	GO2	122	138	127	151	97	203	172	175	173	134	139	148	131
	GO2-04	GO2	122	138	127	151	97	203	172	175	173	134	139	148	131
MLH04	GO1-10	GO1	122	138	127	153	97	203	172	175	173	134	139	148	125
	GO5-04	GO5	122	138	127	153	97	203	172	175	173	134	139	148	125
	GO5-09	GO5	122	138	127	153	97	203	172	175	173	134	139	148	125
	GO5-15	GO5	122	138	127	153	97	203	172	175	173	134	139	148	125
	GO5-32	GO5	122	138	127	153	97	203	172	175	173	134	139	148	125
MLH05	MT3-21	MT3	122	140	127	151	97	195	148	149	173	132	139	154	131
	MT5-12	MT5	122	140	127	151	97	195	148	149	173	132	139	154	131
	MT5-14	MT5	122	140	127	151	97	195	148	149	173	132	139	154	131
	MT5-19	MT5	122	140	127	151	97	195	148	149	173	132	139	154	131
MLH06	GO1-02	GO1	122	140	127	151	97	195	148	149	173	134	139	148	125
	GO1-12	GO1	122	140	127	151	97	195	148	149	173	134	139	148	125
MLH07	GO1-03	GO1	122	140	127	151	97	195	148	149	173	134	139	154	131
	GO4-10	GO4	122	140	127	151	97	195	148	149	173	134	139	154	131
	GO5-05	GO5	122	140	127	151	97	195	148	149	173	134	139	154	131
MLH08	GO5-13	GO5	136	134	155	167	105	201	172	173	179	136	155	154	167
	GO5-41	GO5	136	134	155	167	105	201	172	173	179	136	155	154	167

MLH09	GO1-11	GO1	142	134	155	167	107	201	172	173	179	136	155	154	209
	GO1-15	GO1	142	134	155	167	107	201	172	173	179	136	155	154	209
	GO4-19	GO4	142	134	155	167	107	201	172	173	179	136	155	154	209
	GO5-16	GO5	142	134	155	167	107	201	172	173	179	136	155	154	209
MHL10	MT3-11	MT3	144	134	155	167	107	201	172	173	179	134	153	154	188
	MT3-16	MT3	144	134	155	167	107	201	172	173	179	134	153	154	188
	MT4-07	MT4	144	134	155	167	107	201	172	173	179	134	153	154	188
	MT4-10	MT4	144	134	155	167	107	201	172	173	179	134	153	154	188
	MT4-12	MT4	144	134	155	167	107	201	172	173	179	134	153	154	188
	MT4-25	MT4	144	134	155	167	107	201	172	173	179	134	153	154	188
	MT5-26	MT5	144	134	155	167	107	201	172	173	179	134	153	154	188
MHL11	GO4-08	GO4	154	134	155	165	105	201	172	173	175	136	155	154	182
	GO5-19	GO5	154	134	155	165	105	201	172	173	175	136	155	154	182
MHL12	MT1-12	MT1	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT3-06	MT3	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT3-10	MT3	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT3-12	MT3	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT3-18	MT3	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT3-27	MT3	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT4-01	MT4	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT4-19	MT4	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT5-09	MT5	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT5-13	MT5	154	134	155	167	105	201	172	173	179	134	153	154	167
	MT5-34	MT5	154	134	155	167	105	201	172	173	179	134	153	154	167
MHL13	MT3-02	MT3	154	134	155	167	105	201	172	173	179	134	153	154	194
	MT3-05	MT3	154	134	155	167	105	201	172	173	179	134	153	154	194
	MT3-08	MT3	154	134	155	167	105	201	172	173	179	134	153	154	194
	MT3-13	MT3	154	134	155	167	105	201	172	173	179	134	153	154	194

	MT3-14	MT3	154	134	155	167	105	201	172	173	179	134	153	154	194
	MT5-17	MT5	154	134	155	167	105	201	172	173	179	134	153	154	194
	MT5-22	MT5	154	134	155	167	105	201	172	173	179	134	153	154	194
MHL14	MT3-23	MT3	154	134	155	167	105	201	172	173	179	134	157	154	167
	MT3-28	MT3	154	134	155	167	105	201	172	173	179	134	157	154	167
	MT5-06	MT5	154	134	155	167	105	201	172	173	179	134	157	154	167
	MT5-10	MT5	154	134	155	167	105	201	172	173	179	134	157	154	167
MHL15	GO5-45	GO5	154	134	155	167	105	201	172	173	179	136	155	154	131
	GO5-53	GO5	154	134	155	167	105	201	172	173	179	136	155	154	131
	MT2-02	MT2	154	134	155	167	105	201	172	173	179	136	155	154	131
	MT2-03	MT2	154	134	155	167	105	201	172	173	179	136	155	154	131
	MT2-14	MT2	154	134	155	167	105	201	172	173	179	136	155	154	131
MHL16	GO2-11	GO2	154	134	155	167	105	201	172	173	179	136	155	154	167
	GO4-03	GO4	154	134	155	167	105	201	172	173	179	136	155	154	167
	GO4-12	GO4	154	134	155	167	105	201	172	173	179	136	155	154	167
	GO5-01	GO5	154	134	155	167	105	201	172	173	179	136	155	154	167
	GO5-20	GO5	154	134	155	167	105	201	172	173	179	136	155	154	167
	GO5-31	GO5	154	134	155	167	105	201	172	173	179	136	155	154	167
	GO5-36	GO5	154	134	155	167	105	201	172	173	179	136	155	154	167
	MT2-12	MT2	154	134	155	167	105	201	172	173	179	136	155	154	167
	MT2-13	MT2	154	134	155	167	105	201	172	173	179	136	155	154	167
	MT2-20	MT2	154	134	155	167	105	201	172	173	179	136	155	154	167
MHL17	GO2-07	GO2	154	134	155	167	105	201	172	173	179	136	155	154	185
	GO4-01	GO4	154	134	155	167	105	201	172	173	179	136	155	154	185
	GO4-05	GO4	154	134	155	167	105	201	172	173	179	136	155	154	185
	GO4-27	GO4	154	134	155	167	105	201	172	173	179	136	155	154	185
	GO5-03	GO5	154	134	155	167	105	201	172	173	179	136	155	154	185
	GO5-30	GO5	154	134	155	167	105	201	172	173	179	136	155	154	185

	GO5-44	GO5	154	134	155	167	105	201	172	173	179	136	155	154	185
	GO5-55	GO5	154	134	155	167	105	201	172	173	179	136	155	154	185
MHL18	GO1-14	GO1	154	134	155	167	105	201	172	173	179	136	155	154	188
	GO5-07	GO5	154	134	155	167	105	201	172	173	179	136	155	154	188
	GO5-27	GO5	154	134	155	167	105	201	172	173	179	136	155	154	188
MHL19	GO2-08	GO2	154	134	155	167	105	201	172	173	179	136	155	154	194
	GO4-06	GO4	154	134	155	167	105	201	172	173	179	136	155	154	194
	GO4-09	GO4	154	134	155	167	105	201	172	173	179	136	155	154	194
	GO4-18	GO4	154	134	155	167	105	201	172	173	179	136	155	154	194
	GO4-32	GO4	154	134	155	167	105	201	172	173	179	136	155	154	194
	GO5-18	GO5	154	134	155	167	105	201	172	173	179	136	155	154	194
	GO5-51	GO5	154	134	155	167	105	201	172	173	179	136	155	154	194
	MT2-05	MT2	154	134	155	167	105	201	172	173	179	136	155	154	194
MHL20	GO4-04	GO4	154	134	155	167	105	201	172	173	179	136	157	154	185
	GO4-28	GO4	154	134	155	167	105	201	172	173	179	136	157	154	185
MHL21	MT1-06	MT1	154	134	155	167	107	201	172	173	179	134	153	154	167
	MT4-16	MT4	154	134	155	167	107	201	172	173	179	134	153	154	167
	MT4-21	MT4	154	134	155	167	107	201	172	173	179	134	153	154	167
	MT5-24	MT5	154	134	155	167	107	201	172	173	179	134	153	154	167
	MT5-29	MT5	154	134	155	167	107	201	172	173	179	134	153	154	167
MHL22	GO2-02	GO2	154	134	155	167	107	201	172	173	179	136	155	154	167
	GO2-09	GO2	154	134	155	167	107	201	172	173	179	136	155	154	167
	GO3-01	GO3	154	134	155	167	107	201	172	173	179	136	155	154	167
	GO4-29	GO4	154	134	155	167	107	201	172	173	179	136	155	154	167
	GO4-34	GO4	154	134	155	167	107	201	172	173	179	136	155	154	167
	MT2-10	MT2	154	134	155	167	107	201	172	173	179	136	155	154	167
MHL23	MT3-04	MT3	168	134	155	167	105	201	172	173	179	134	153	154	182
	MT3-17	MT3	168	134	155	167	105	201	172	173	179	134	153	154	182

	MT3-19	MT3	168	134	155	167	105	201	172	173	179	134	153	154	182
	MT3-25	MT3	168	134	155	167	105	201	172	173	179	134	153	154	182
	MT4-08	MT4	168	134	155	167	105	201	172	173	179	134	153	154	182
	MT5-01	MT5	168	134	155	167	105	201	172	173	179	134	153	154	182
	MT5-30	MT5	168	134	155	167	105	201	172	173	179	134	153	154	182
	MT5-31	MT5	168	134	155	167	105	201	172	173	179	134	153	154	182
	MT5-33	MT5	168	134	155	167	105	201	172	173	179	134	153	154	182
MHL24	GO2-01	GO2	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-02	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-14	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-16	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-20	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-23	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-25	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-26	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-30	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO4-33	GO4	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO5-02	GO5	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO5-24	GO5	168	134	155	167	105	201	172	173	179	136	155	154	182
	GO5-26	GO5	168	134	155	167	105	201	172	173	179	136	155	154	182
MHL25	MT3-01	MT3	168	134	155	167	107	201	172	173	179	134	153	154	182
	MT3-07	MT3	168	134	155	167	107	201	172	173	179	134	153	154	182
	MT5-23	MT5	168	134	155	167	107	201	172	173	179	134	153	154	182

3. DRAFT GENOME SEQUENCE OF *COLLETOTRICHUM TRUNCATUM* ASSOCIATED WITH SOYBEAN ANTHRACNOSE

ABSTRACT

Advances in high-throughput sequencing and computational tools have made it possible to sequence and analyze whole genomes of many plant pathogens. Soybean anthracnose is one of the most important diseases in this crop, associated mainly with the fungus *Colletotrichum truncatum*, which also can infect plants belonging to Fabaceae and Solanaceae families. Here we report the draft genome sequence of a typical isolate of the *C. truncatum* (strain CMES1059) associated with soybean anthracnose in Brazil. This genome represents a new resource useful for further research into the biology, ecology, and evolution of this important pathogen.

Keywords: *De novo* assembly; Long reads; Short reads; Fungal

Colletotrichum is a large genus of plant pathogenic fungi that cause economically important diseases on numerous crops worldwide (Crouch et al. 2014). Soybean anthracnose is mainly associated with the fungal species *Colletotrichum truncatum* (Hyde et al. 2009), which can also infect important plant species belonging to families Fabaceae and Solanaceae (Damm et al. 2009). Recent advances in high-throughput sequencing and computational tools have made it possible to sequence and analyze whole genomes of many plant pathogens (Grünwald, McDonald, and Milgroom 2016). These technologies continue to advance rapidly, and costs have been declined, resulting in an increasing number of pathogenic fungi sequenced (O'Connell et al. 2012). Whole genome sequences of many *Colletotrichum* species belonging to different species complexes have been published in recent years, for instance, *C. graminicola* (O'Connell et al. 2012), *C. higginsianum* (Zampounis et al. 2016) , *C. orbiculare* (Gan et al. 2013), *C. fructicola* (Gan et al. 2013), *C. fiorineae* (Baroncelli et al. 2014a), *C. sublineola* (Baroncelli et al. 2014b), *C. orchidophilum* (Baroncelli et al. 2018) among others, giving an impetus to the *Colletotrichum* research field.

The *C. truncatum* strain CMES1059 (provided by Embrapa Soja, Londrina, Paraná State, Brazil) was isolated from Brazilian soybean seeds and previously characterized by molecular and pathogenic study (Rogério et al. 2017). This strain represents a typical isolate of the species associated with soybean anthracnose, thus was chosen to be whole-genome sequenced.

Single-spore isolate was cultured on potato dextrose agar (PDA) and incubated for 7 days at 25°C. Mycelium plugs were taken from the colony margin, transferred to potato dextrose broth (PDB) and incubated with shaking at 180 rpm for 3 days. The genomic DNA was extracted by incubating 200 mg of fresh mycelium in 300 µl of lysis buffer (SDS 10% - 100nM, TE 1X – pH 8.0, proteinase K - 20 mg/mL, NaCl - 5M), following by treatment with phenol/chloroform/isoamyl alcohol 25:24:1, precipitation overnight in isopropanol and rinsed in 70% ethanol. The nucleic acid extract was treated with RNaseA (0.2 mg/mL) and purified by another round of phenol/chloroform/isoamyl alcohol 25:24:1 treatment. The DNA quality and quantity were measured in a NanoDrop® spectrophotometer (Thermo Scientific, Wilmington, DE) and Qubit® dsDNA HS Assay Kits (Applied Biosystems), respectively. The fragment size distribution, was checked using Genomic DNA ScreenTape® (Agilent, USA).

The DNA was sequenced by a combination of PacBio long reads and Illumina short reads data. Long reads were generated by Pacbio Technology (Eid et al. 2009), using two Single Molecule Real-Time cell (SMRT) (P5-C3 chemistry), sequenced on the PacBio Sequel (Pacific Biosciences of California, Inc.). A PacBio large-insert library (15-20kb) was prepared using PacBio large insert library prep kit performed by Duke Center for Genomic and Computational Biology (Durham, USA). Short reads were sequenced by Illumina HiSeq2000 sequencing machine (paired end 2x150bp, insert size ~ 500pb). Library was constructed following Illumina's standard protocol performed by commercial company Genewiz (South Plainfield, USA). Illumina reads were trimmed with Trim Galore 0.3.6. in order to remove low quality base. To correct the high error rate from PacBio sequencing, reads were corrected through Illumina reads using LoRDEC 0.9 (Salmela and Rivals 2014). *De novo* assembly was performed using CANU assembler (Koren et al. 2014). We performed manual curation in order to check regions very similar (duplicated), and unbalanced coverage sequences, using Geneious® 7.1.5 (Kearse et al. 2012). The mitochondrial sequences were identified by performing BLASTn (Altschul et al. 1990) searches of the contigs in a custom database of mitochondrial genomes of other fungi, resulting in the identification and removal of them. The nuclear genome was annotated using the MAKER3 pipeline (Cantarel et al. 2008).

The draft nuclear genome of *C. truncatum* generated consists of 128 sequence scaffolds with a total size of 56.2 Mb (N_{50}= 831,709 bp), largest scaffold length of 2.71 Mb, and average coverage about 124X. The estimated genome size and GC content of 50.1% was comparable to most of the *Colletotrichum* strains sequenced. Overall, 15.154 protein-coding gene models were predicted.

In this study we report the first draft genome sequence of a strain belonging to *C. truncatum* isolated from soybean, which was used as reference genome to subsequent population genomics analyses described in next chapter.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–10.

Baroncelli, R., Sreenivasaprasad, S., Sukno, S. A., Thon, M. R., and Eric Holub. 2014. Draft Genome Sequence of *Colletotrichum acutatum* Sensu Lato (*Colletotrichum fioriniae*). Genome Announc. 2:1–2.

Baroncelli, R., Sukno, S., Sarrocco, S., Cafà, G., Le Floch, G., and Thon, M. R. 2018. Whole-genome sequence of the orchid anthracnose pathogen *Colletotrichum orchidophilum*. Mol. Plant-Microbe Interact. 31:979–981.

Cantarel, B., Robb, S., Parra, G., Ross, E., Moore, B., Holt, C., et al. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18:188–196.

Crouch, J., O'Connell, R., Gan, P., Buiate, E., Torres, M. F., Beirn, L., et al. 2014. The Genomics of Colletotrichum. In *Genomics of Plant-Associated Fungi: Monocot Pathogens*, eds. Dean R., Lichens-Park A., and Kole C. , p. 123–140.

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

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. 2009. Single Polymerase Molecules. Exch. Organ. Behav. Teach. J. :133–138.

Gan, P., Ikeda, K., Irieda, H., Narusaka, M., O'Connell, R. J., Narusaka, Y., et al. 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol. 197:1236–1249.

Grünwald, N. J., McDonald, B. A., and Milgroom, M. G. 2016. Population Genomics of Fungal and Oomycete Pathogens. Annu. Rev. Phytopathol. 54:323–346.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2014. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736.

O'Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M. F., et al. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat. Genet. 44:1060–1065 Available at: http://dx.doi.org/10.1038/ng.2372.

Rogério, F., Ciampi-Guillardi, M., Barbieri, M. C. G., Bragança, C. A. D., Seixas, C. D. S., Almeida, A. M. R., et al. 2017. Phylogeny and variability of *Colletotrichum truncatum* associated with soybean anthracnose in Brazil. J. Appl. Microbiol. 122:402–415.

Salmela, L., and Rivals, E. 2014. LoRDEC: Accurate and efficient long read error correction. Bioinformatics. 30:3506–3514.

Zampounis, A., Pigné, S., Dallery, J.-F., Wittenberg, A. H. J., Zhou, S., Schwartz, D. C., et al. 2016. Genome Sequence and Annotation of *Colletotrichum higginsianum*, a Causal Agent of Crucifer Anthracnose Disease. Genome Announc. 4:e00821-16.

4. ADMIXTURE BETWEEN DIVERGENT LINEAGES OF THE FUNGUS *COLLETOTRICHUM TRUNCATUM* ASSOCIATED WITH SOYBEAN ANTHRACNOSE

ABSTRACT

Whole-genome sequencing can provide fine-grained details regarding processes of divergence, differentiation, and recombination in fungal species. Furthermore, population genomics can provide insights into the evolutionary history of invasive species and the genetic composition of their populations. The fungus *Colletotrichum truncatum* associated with soybean anthracnose is an invasive species recently introduced in Brazil, structured into three genetic clusters, coexisting in syntopy in the soybean fields. However, the genetic makeup of the three lineages remains unknown, and their extreme geographic proximity raises the question of the maintenance of their genetic integrity in the face of admixture. Here we used whole genome sequencing and population genomics analyzes to gain insights into the evolutionary history of *C. truncatum* lineages and to investigate in details the possibility of a lack of genetic exchanges between them. Eighteen representative *C. truncatum* isolates from the three lineages were submitted to whole genome sequencing, aligned against the reference genome, and variants were identified. Our analyses revealed that the genetic structure of *C. truncatum* pathogen causing soybean anthracnose is formed by three deeply divergent lineages with levels of genetic diversity consistent with repeated introduction events for each lineage. We also found evidence for sexual recombination within and between lineages, with multiples isolates displaying signatures of admixture. Our findings support a scenario in which the three lineages initially diverged in allopatry before experiencing hybridization following secondary contact. Monitoring of the pathogen's diversity over time is needed to reveal whether the lineages maintain or fuse, which can impact the disease control methods currently employed

Keywords: Population genomics; Speciation, SNPs; Recombination; Gene flow; Disease emergence

4.1 INTRODUCTION

Recent technological advances in both high-throughput sequencing and computational tools have made it possible to sequence and analyze whole genomes of many plant pathogens. These technologies continue to advance rapidly, and costs have declined to the point that it is becoming affordable to genotype thousands of single-nucleotide polymorphisms (SNPs) simultaneously or resequence many individuals within a species (Grünwald, McDonald, and Milgroom 2016). The large number of genetic markers now available in genomic scale provides

fine-grained details regarding processes of divergence, differentiation, and recombination of the pathogen populations, enabling more robust inferences compared to studies based on a limited number of genetic markers (Grünwald, McDonald, and Milgroom 2016; Luikart et al. 2003; Brumfield et al. 2003; Helyar et al. 2011).

The genomic revolution makes it possible to go far beyond traditional population genetics analyses, allowing the research of the evolutionary processes shaping pathogen populations in agriculture environments at chromosome level. Such approach has proved useful to provide information on evolutionary history of invasive species and genetic composition of their populations. (Estoup and Guillemaud 2010; Gross, Hosoya, and Queloz 2014). Understanding the genetic diversity and population structure of invasive pathogens in introduced areas is crucial to reveal hidden biological features of an organism, to the reconstruction of the course of invasions and to the establishment of effective control measures (Gross, Hosoya, and Queloz 2014).

Soybean anthracnose is one of the most important diseases of this crop, associated mainly with the fungus *Colletotrichum truncatum* (Hyde et al. 2009; Damm et al. 2009). Our previous study suggested that the fungus is an invasive species recently introduced in Brazil, in which distinct introduction events were responsible by three strongly differentiated genetic clusters, coexisting in syntopy in the same areas (Rogério et al. 2019). The existence of a highly efficient mechanism of pathogen dispersal over long distances, possibly via contaminated seed, seems to play a crucial role in structuring populations of the pathogen, mainly in the maintenance of the genetic clusters distribution within the fields (Rogério et al. 2017).

The genetic variability of invading populations depends on the history of their original populations and on the historical and demographical features of their introduction (Estoup and Guillemaud 2010). The levels of genetic diversity observed in *C. truncatum* are much higher than would be generally expected for an organism that has been recently founded, as questioned by the genetic paradox of invasion (Blischak et al. 2018; Rogério et al. 2019). Such finding raises questions about how the pathogen genome has evolved following the introduction. The process of introduction into a new location, as well as spatial expansion from the point of introduction, often imposes a transitory reduction in population size (i.e., a demographic bottleneck), which has the potential to reduce genetic variation (Dlugosch et al. 2015; Dlugosch and Parker 2008). In agricultural systems, many plant pathogenic fungi have genetic structures consistent with having experienced a population bottleneck due to founder (Banke and McDonald 2005; Milgroon et al. 2008; Linde, Zala, and McDonald 2009). However, introduced

populations can overcome the various consequences of low genetic variation from its founders, through the purge of deleterious alleles accumulated following bottlenecks events, or via the fixation of *de novo* beneficial mutations from standing variation, allowing rapid adaptation (Blischak et al. 2018; Schrieber and Lachmuth 2017; Frankham 2005).

Although *Colletotrichum* is considered an asexual fungus (Crouch et al. 2014), sexual reproduction has been observed in some species (Kimati and Galli 1970; Vaillancourt and Hanau 1991; Damm et al. 2019). Alternative sexual mechanisms such as parasexual reproduction, hyphal anastomosis process, and cryptic sexual cycle are described as mechanisms responsible to create genetic recombination in this genus (Souza-Paccola et al. 2003; Roca et al. 2003; Rosada et al. 2010; Vaillancourt et al. 2000). Some studies reported population genetic signatures of sexual recombination in *C. truncatum* (Diao et al. 2015; Katoch et al. 2017), signaling that recombination events may be occurring and playing an important role in the genetic structure of the pathogen. The presence of three genetic lineages of *C. truncatum* widely spread in soybean fields coexisting in syntopy has been raised questions about the genetic makeup of these lineages, their reproductive strategy, and their ability to maintain it in the face of gene flow (Rogério et al. 2019).

The increased importance of soybean anthracnose in some regions of Brazil, together with the theoretical background obtained from our previous studies on this pathosystem have been driving further genetic investigations of this important plant pathogen. This study has focused on population genomic analyses using SNP markers, in order to better understand the distribution of genetic variation of *C. truncatum* lineages previously detected based on SSR markers. We aimed to gain insights into the history of divergence of the three lineages, as well as to investigate the possibility of a lack of genetic exchanges between them.

4.2 MATERIALS AND METHODS

4.2.1 Sample collection, DNA extraction and sequencing

We used eighteen isolates of *Colletotrichum truncatum* from naturally infected commercial soybean fields from Mato Grosso (MT) and Goiás (GO) states in Brazil in this study. These isolates were selected based on genetic population analyses derived by SSRs markers employed in chapter 2. According to the genetic structure found, eighteen isolates representative of the three genetic clusters were selected and submitted to whole genome sequencing (Table 1). Fungal genomic DNA was extracted using Wizard Genomic DNA

Purification kit (Promega) from fresh mycelium grown on potato dextrose liquid medium (Difco). Sequencing libraries were prepared and Illumina HiSeq2000 sequencing (paired end 2x150bp, insert size ~ 500pb), performed at commercial company Genewiz (South Plainfield, USA).

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

Cluster	Isolate	Population	State	County	Cultivar	Sampling year
C1	MT2-05	MT2	Mato Grosso	Sinop	Y-70 Pioner	2017
C1	MT3-01	MT3	Mato Grosso	Lucas do Rio Verde	7709 Nidera	2017
C1	MT5-26	MT5	Mato Grosso	Sinop	7709 Nidera	2017
C1	GO4-08	GO4	Goiás	Rio Verde	7739 Monsoy	2017
C1	GO5-25	GO5	Goiás	Rio Verde	PP7200Macro	2017
C1	GO2-12	GO2	Goiás	Montividiu	Nidera 5909	2017
C2	MT3-21	MT3	Mato Grosso	Lucas do Rio Verde	8372 Nidera	2017
C2	MT4-05	MT4	Mato Grosso	Sinop	7709 Nidera	2017
C2	MT5-12	MT5	Mato Grosso	Sinop	7709 Nidera	2017
C2	GO4-17	GO4	Goiás	Rio Verde	PP7200Macro	2017
C2	GO5-11	GO5	Goiás	Rio Verde	Nidera 5909	2017
C2	GO2-03	GO2	Goiás	Montividiu	7739 Monsoy	2017
C3	MT1-01	MT1	Mato Grosso	Lucas do Rio Verde	8766 Monsoy	2016
C3	MT4-13	MT4	Mato Grosso	Sinop	8372 Nidera	2017
C3	MT5-32	MT5	Mato Grosso	Sinop	7709 Nidera	2017
C3	GO2-06	GO2	Goiás	Montividiu	PP7200Macro	2017
C3	GO4-07	GO4	Goiás	Rio Verde	Nidera 5909	2017
C3	GO5-14	GO5	Goiás	Rio Verde	7739 Monsoy	2017

4.2.2 Taxonomic and pathogenicity confirmation

In order to confirm the identity and genetic distance among isolates employed in this study and the strain used as genome reference, we performed a Bayesian phylogenetic inference, using conserved genes as described for *Colletotrichum* genus identification (Damm et al. 2009). For this propose, *de novo* assemblies of the Illumina reads were performed for each isolate using SPAdes assembler (Bankevich et al. 2012). Partial sequences of theglyceraldehyde-3-phosphate dehydrogenase (GAPDH), actin (ACT), chitin synthase 1 (CHS-1), beta-tubulin (Tub2) and histone3 (HIS3) genes were retrieved from draft genomes using BLASTn (Altschul et al. 1990). Phylogenetic analysis was performed with sequences of 41 isolates, encompassing the main epitypes of *Colletotrichum* species, and *C. truncatum* isolates used in this study, as employed by Rogério et al. (2017). Sequences were aligned using the software MEGA (Tamura

et al. 2007) and Mesquite program was used to concatenate the multiple alignments (http://www.mesquiteproject.org). The nucleotide substitution model was estimated using MrModeltest ver. 3.7 (Nylander 2004) for each locus individually. Bayesian inference was used for phylogenetic reconstruction based on multilocus alignment. The analysis was performed twice on the program MrBayes ver. 3.2 (Ronquist and Huelsenbeck 2003) using algorithm MCMC (Markov Chain Monte Carlo) to generate phylogenetic trees with a posterior probability. Four MCMC chains were executed simultaneously with $1x10^7$ generations. Trees were randomly sampled every 1,000 generations, and 25% of them were discarded as burn-in. FigTree ver. 3.1 software (http://tree.bio.ed.ac.uk/) was used to edit the trees.

The studied isolates were inoculated in the cultivar IPRO 8579 in order to confirm their pathogenicity to the soybean. Pre-germinated seeds were inoculated with conidial suspensions ($1x10^6$ conidia/mL) of each isolate and incubated into a plant growth chamber at 25°C, in photoperiod of 12 h. The seedlings were evaluated for the presence of symptoms 120 h after inoculation, according to Hartman et al.(2015). The average values were submitted to the analysis of variance (ANOVA) and compared by the Tukey test with $P \leq 0.05$.

4.2.3 Genome sequencing and SNP calling

Raw Illumina reads were trimmed for adapter contamination and sequencing quality using cutadapt 1.16 software. Reads were mapped against the draft genome of strain CMES1059 previously generated as described in chapter 3, using BWA mem v0.7.15 (Li and Durbin 2009). Alignments were sorted with Samtools v1.3 (Li et al. 2009), and reads with mapping quality below 30 were removed. Duplicates were removed with Picard tools v2.7 (http://broadinstitute.github.io/picard/). The Genome Analyses Toolkit v4.0.12 (GATK) (McKenn et al. 2009) was used to calling single nucleotide polymorphisms (SNPs) and indels, according to the GATK best practices (http://www.broadinstitute.org/gatk/guide/best_practices) using HaplotypeCaller, with the option –emitRefConfidence GVCF. The gvcf 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 with the following 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 ALT reads) and ReadPosRankSum < 8.0 (Distance of ALT read from the end of the reads).

4.2.4 Genealogical relationships and population structure

Total evidence genealogy across samples was inferred with RAxML from pseudoassembled genomic sequences, i.e., tables of total SNPs obtained converted into a fasta file, using the reference sequence as a template with Python script. The maximum likelihood genealogy was estimated assuming the GTRGAMMA model of sequence evolution. Bootstrap confidence levels were determined using 100 replicates. We based the analysis on a dataset combining the full set of SNPs and monomorphic sites, rather than just SNPs, to increase topological and branch lengths accuracy (Leaché et al. 2015). Sequence Read Archive (SRA) of *C. gloeosporioides* (SRR892046) obtained from NCBI/GenBank database were mapped against genome reference and included as outgroup.

The population structure was analyzed using principal components analysis (PCA) and discriminant analysis of principal components (DAPC) using the R package Adegenet version 2.0 (Jombart and Ahmed 2011). 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. This method identifies an optimal number of genetic clusters that best describe the data by running a *K* means clustering algorithm and comparing the different clustering solutions using the Bayesian Information Criterion (BIC). Such analysis was based on biallelic SNPs without missing data. In order to explore the clusters defined by DAPC, the estimation of individual admixture coefficients was performed using the software sNMF, (Frichot et al. 2014), which is 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.

4.2.5 Linkage disequilibrium and recombination

In order to access the genomic impact of recombination in the genetic lineages, we analyzed the patterns of linkage disequilibrium (LD), i.e., the tendency of different alleles to occur together in a nonrandom manner. For that, we calculated the coefficient of linkage disequilibrium (r^2) (Hill and Robertson 1968) using Vcftools (Danecek et al. 2011). Were used biallelic SNPs data, excluding missing data and sites with minor allele frequencies below 10%, in order to minimize the dependence of r^2 on minor allele frequency. We based on LD decay

plot function to visualize the rate at which LD declines with physical distance (Flint-Garcia, Thornsberry, and Buckler 2003). For all lineages, we calculated r^2 values for all pairs of SNPs less than 500 kb apart and averaged LD values in distance classes of 200 pb.

The estimation of recombination rates was implemented in the program LDHAT version 2.2 (McVean and Auton 2011), which estimate population recombination rates based on the LD between SNPs using Hudson's composite likelihood method, under the crossing-over model. The method infers the population-scaled recombination rate (ρ) relates to the actual recombination frequency by the equation $\rho = 2N_e r$ for haploid individuals, where N_e is the effective population size and *r* is the per site rate of recombination per generation across the region (McVean and Auton 2011). Only biallelic variants were retained in analysis and employed by program INTERVAL to create inputs based on VCF files, which were split by the ten largest scaffolds for each lineage using custom python scripts. Singletons and sites with missing data were excluded. Reversible-jump Markov Chain Monte Carlo scheme was executed for 5e6 iterations, with a block penalty of 10, samples taken every 5000 iterations and a burn-in phase of 5e5. For each scaffold COMPLETE program was used to generate likelihood look up tables for INTERVAL, using the population-scaled mutation rate estimated as total nucleotide diversity (π per pb estimated on 10kb nonoverlapping windows, as described in 4.2.7 Summary statistics of polymorphism and divergence) and the population-scaled recombination rate ranging from 0 to 100 with an increment of 0.5. Results of INTERVAL were summarized using STAT program, discarding the first 100 samples as burn-in.

The null hypothesis of no recombination was tested using the PHI test implemented in SplitsTree software (http://www.splitstree.org/) for each lineage, which is based on the expectation that sites are exchangeable if there is no recombination in the data.

4.2.6 Genome scan for genetic exchanges

We used Chromopainter program version 0.0.4 (Lawson et al. 2012) for probabilistic chromosome painting, in order to scan the genomes for the exchange of mutations between genetic lineages. This method "paints" individuals in "recipient" populations as a combination of segments from "donor" populations, using linkage information for probability computation and assuming that linked alleles are more likely to be exchanged together during recombination events. We employed each isolate as a single haplotype to be painted, and all lineages as donor, using biallelic SNPs data set without missing data. We initially ran the model using increments of 50 expectation maximization iterations, starting at 10 iterations, and we examined the

convergence of parameter estimates to determine how many iterations to use. The recombination scaling constant N_e and emission probabilities (μ) were estimated for each lineage running the expectation maximization algorithm with 200 iterations for each lineage and scaffolds. Estimates of N_e and μ were then calculated as averages weighted by scaffolds length. These parameters and recombination rates the per scaffold were determined with LDhat program, which were then used to paint the scaffold of each lineage using 200 expectation maximization iterations.

4.2.7 Summary statistics of polymorphism and divergence

Genome variability was assessed using EggLib (De Mita et al. 2012). We computed the average number of nucleotide differences (π), as an estimate of nucleotide diversity, as well as Tajima's D (Tajima 1989), in 10 kb non-overlapping sliding windows for each genetic cluster. To produce input files, VCF files were converted with custom python scripts to pseudo-alignments in fasta format, where the reference sequence was substituted with the variant nucleotides for each strain. The genomic-wise statistics were plotted along of 10 largest contigs through Circos 0.67 software (Connors et al. 2009).

4.3 RESULTS

4.3.1 Genome sequencing and SNP calling

In this study were used eighteen isolates of *C. truncatum* obtained from infected soybean fields, sequenced by Illumina HiSeq2000, whose reads were aligned against the reference genome of strain CMES1059. Mean sequencing depth ranged from 55X to 100X with paired-end reads. We obtained a total of 2.220,191 high-confidence, bi-allelic SNP loci, without missing data, distributed over 128 scaffolds.

4.3.2 Taxonomic and pathogenicity confirmation

The Bayesian phylogenetic inference based on the concatenation of sequences of five genes from eighteen isolates and of the strain employed as reference, confirmed the identity of *C. truncatum* strains used in this study. Based on their position in the cladogram, all the

sequenced isolates belong to *C. truncatum* complex (Fig. 1). All inoculated isolates were pathogenic to the soybean seedlings, with differences in the aggressiveness among them (Table 1S and Fig. 1S).

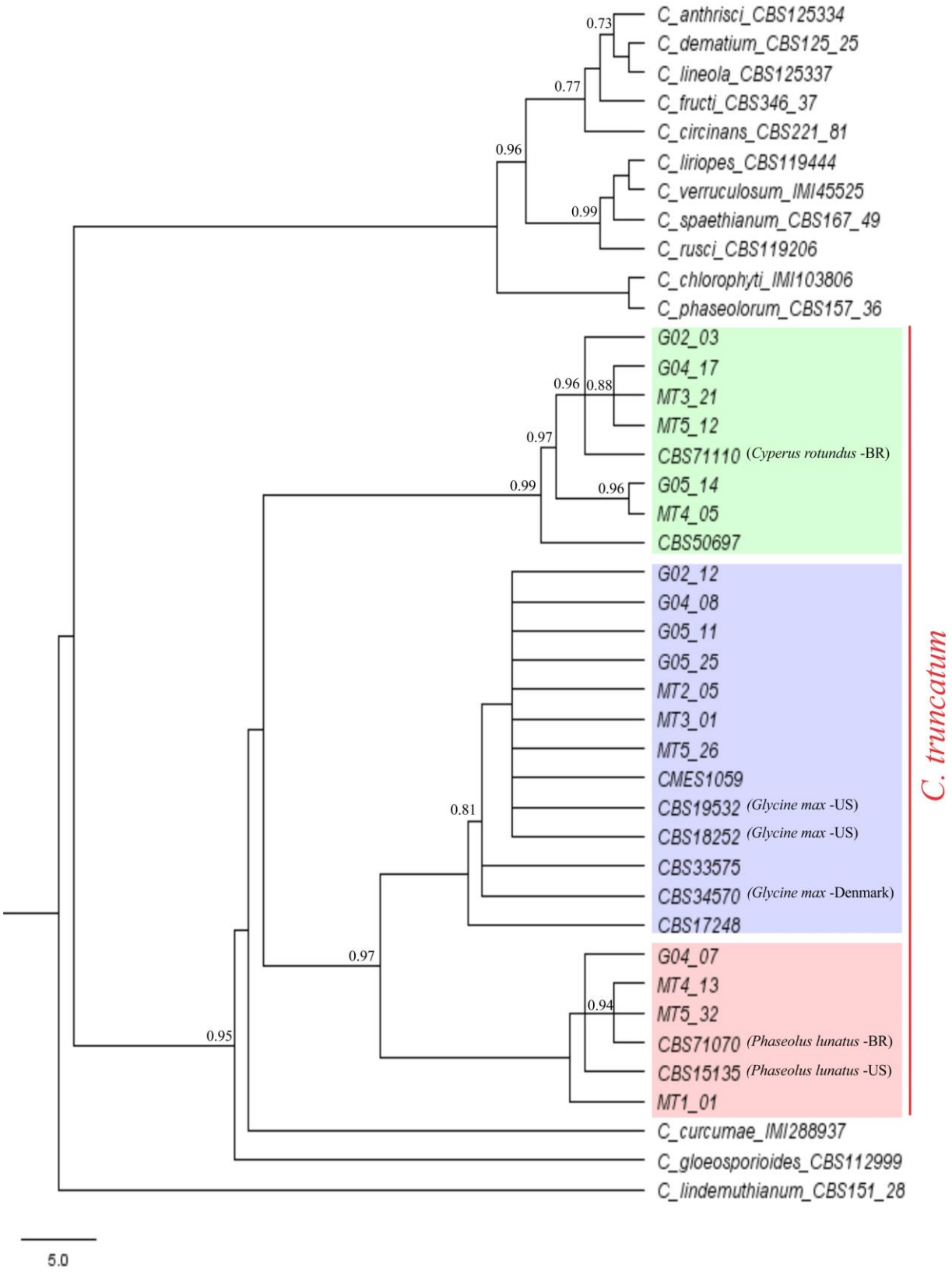


Figure 1. Bayesian inference phylogenetic tree of 41 *Colletotrichum truncatum* isolates. The tree was reconstructed using concatenated data from ACT, CHS-1, GAPDH, HIS3 and TUB2 sequences. Posteriori probability values below 1 are represented in the branches.

4.3.3 Genealogical relationships and population structure

The maximum-likelihood phylogenetic analysis revealed the presence of three divergent and well supported groups with high bootstrap support in most of the branches (Fig. 2A). According to DAPC analysis, the biggest decrease of BIC was reached from *K*=2 to *K*=3, which suggest that this model captures the most salient features of the population analyzed (Fig. 2B). The Ancestry coefficients estimated using the sNMF method showed a similar pattern of population structure, which *K*=3 had the lowest cross-entropy value, indicating the most probable number of genetic groups in the data. Furthermore, signatures of admixture were detected among the isolates, GO4-17, MT4-05 and MT1-01, which belong to the lineages C2 and C3 respectively, shaping 16.6% of admixture in the isolates analyzed (Fig. 2C).

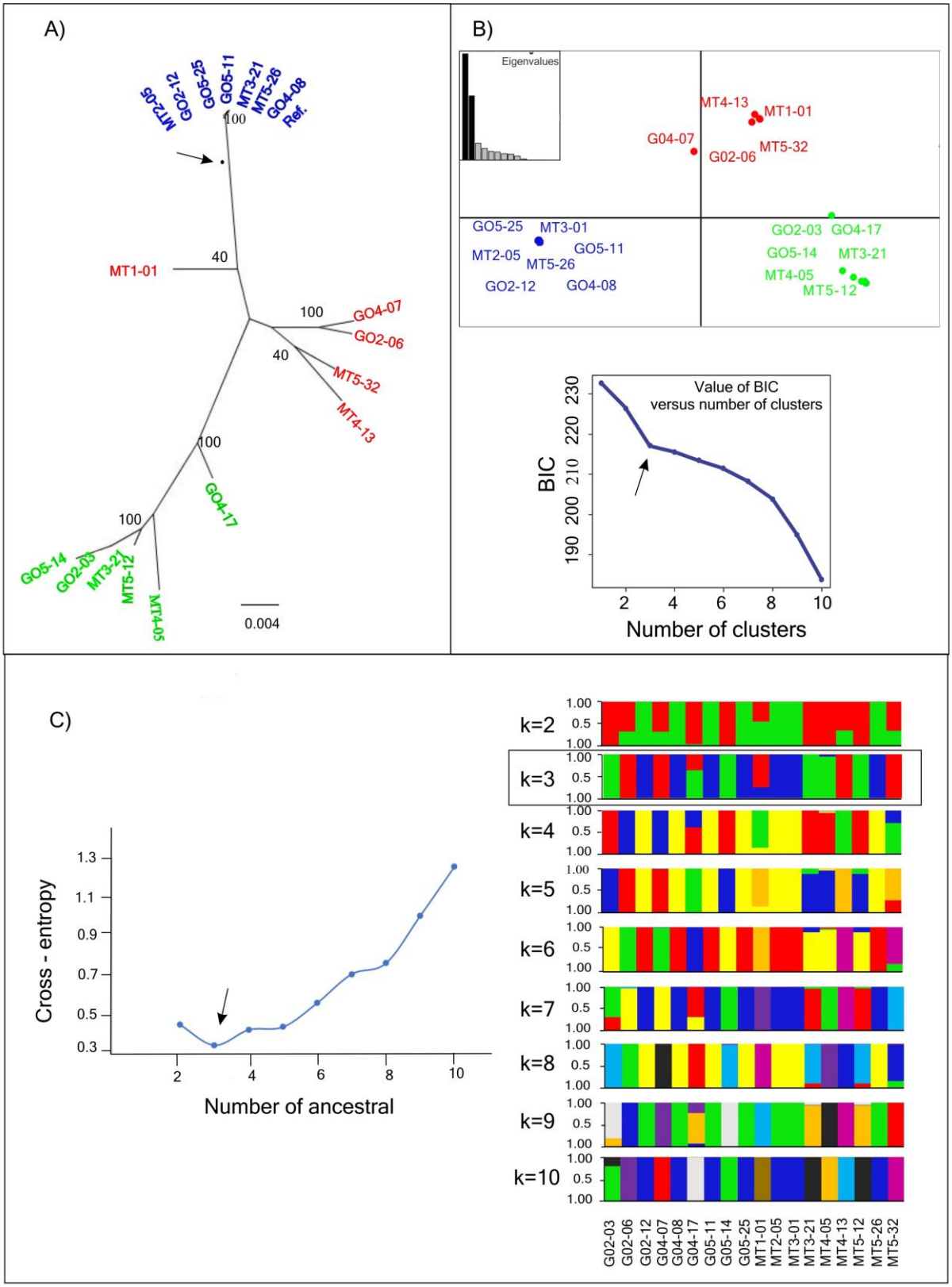


Figure 2. Population subdivision of *Colletotrichum truncatum*. (A) Maximum likelihood tree based on full data set of SNPs with 100 bootstrap replicates. The arrow indicates the position of the outgroup in the phylogenetic tree. (B) Scatter plot from DAPC analysis and Bayesian information criteria (BIC). (C) Individual ancestry coefficients estimated using sNMF and cross-entropy criterion, indicating the most probable number of genetic groups.

4.3.4 Linkage disequilibrium and recombination

Analyses of linkage disequilibrium based on LD decay in nonoverlapping sliding windows of 500 kb showed that the LD decay pattern was in the order of 100bp (Fig. 4). For lineage C1, the LD decay occurred rapidly, displaying an interesting pattern of variation in the r^2 values along of the windows. In contrast, for lineages C2 and C3 the LD decay was very smoothly and the r^2 values remained constant.

The PHI test for recombination rejected the null hypothesis of no recombination for the three lineages (P=0.0000). Estimate of recombination rates varied across scaffolds and lineages, with lineage C3 showing the highest, and C2 the lowest recombination rates on average across scaffolds (Table 2).

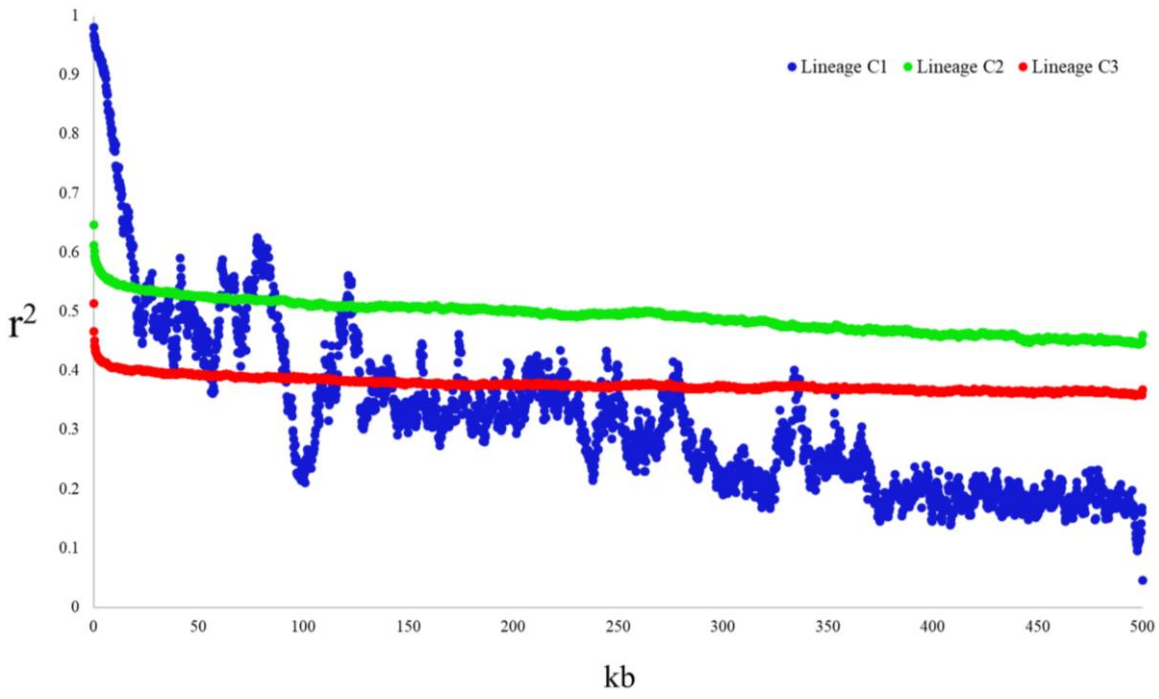


Figure 3. Linkage disequilibrium (LD) decay plots of genetic lineages of *Colletotrichum truncatum*. LD, measured as was calculated for all pairs of SNPs less than 500kb distance apart, and LD values were averaged in distance classes of 200pb.

Table 2. Estimates of recombination rates (ρ) obtained using the software LDHat for the ten largest scaffolds in *Colletotrichum truncatum* genetic lineages.

Scaffolds	Length (pb)	N° of crossovers/Mbp/generation Lineages		
		C1	C2	C3
Scaffold 00332	1184263	9.97	0.11	200.97
Scaffold 00284	1288737	60.67	16.97	135.96
Scaffold 70488	1313461	0.32	6.35	100.77
Scaffold 00225	1434454	1.79	6.35	98.91
Scaffold 00209	1434772	1.94	0.45	0.06
Scaffold 00164	1560415	19.54	1.48	35.07
Scaffold 00066	1726189	26.98	2.34	8.86
Scaffold 00097	1727678	8.75	17.09	29.23
Scaffold 00055	1989181	133.97	11.60	5.53
Scaffold 70486	2707168	44.12	0.07	14.34
Mean	_	30.81	6.28	62.97

4.3.5 Probabilistic chromosome painting

We scanned the genomes to detected exchanges of mutations between lineages at the chromosome level, taking the ten largest scaffolds as potential full chromosomes. Using this approach, we detected genetic exchanges among the genetic groups in the scaffolds analyzed, although not strictly restricted to the three isolates previously identified as admixed (Fig. 5). For lineages C1 and C2, the majority of mutations were assigned to self (i.e. to their cluster of origin), with some scaffolds in isolate GO4-17 of lineage C2 assigned to lineage C3. For lineage C3, mutations tended to have non-zero membership probabilities in multiple clusters, although some regions were clearly assigned to lineage C1 and self.

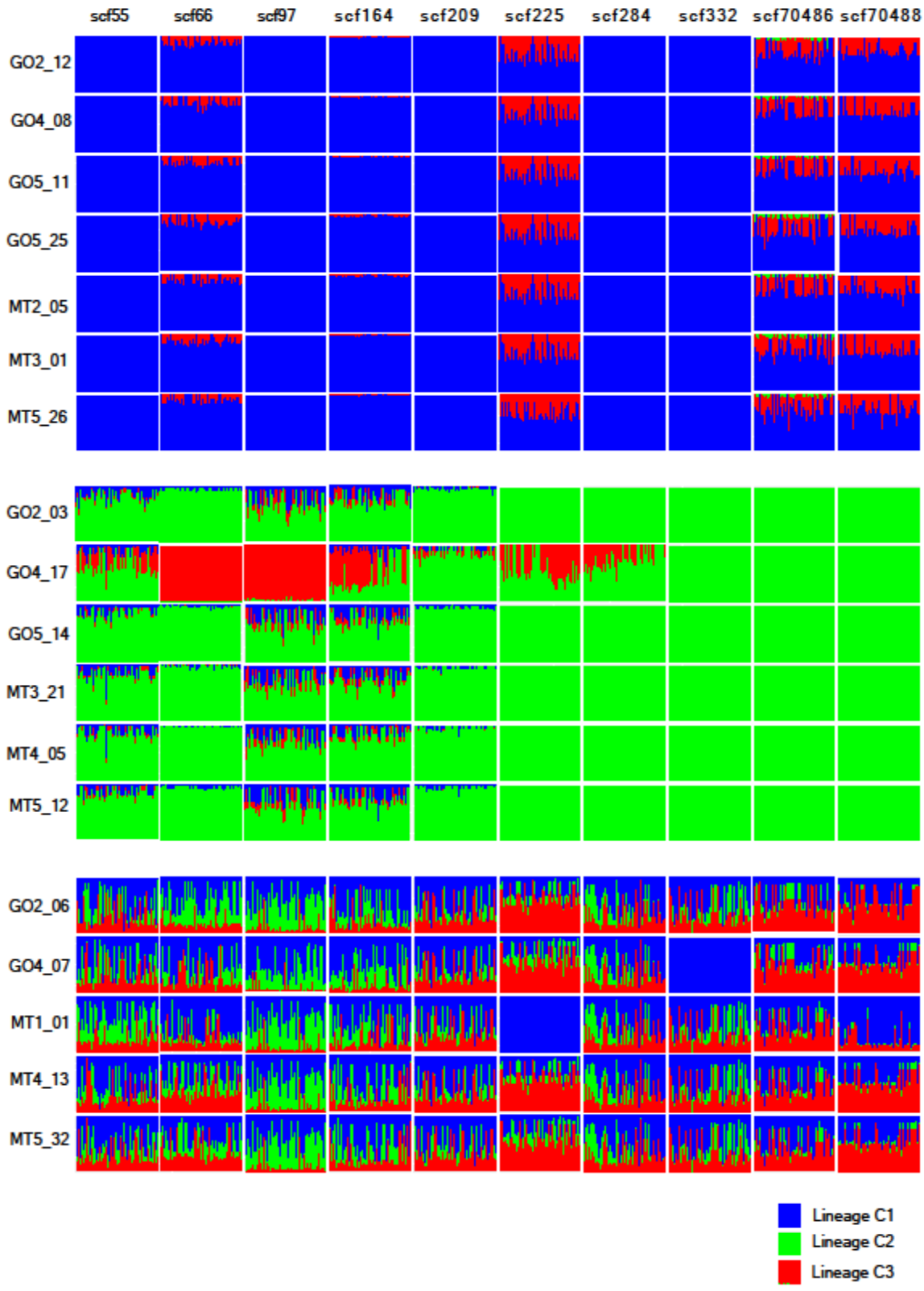


Figure 4. Probabilistic chromosome painting scatter plots of ten largest scaffolds from eighteen *Colletotrichum truncatum* genomes analyzed in this study.

4.3.6 Summary statistics of polymorphism and divergence

To better understand genomic diversity in the three *C. truncatum* genetic lineages, we calculated population genetic statistics for each one, including nucleotide diversity and Tajima's D. Genome-wide calculations revealed that C3 had higher diversity than others lineages (Table 3). When diversity was plotted along each scaffold, the same pattern of variation was clearly evidenced, as shown in Figure 3. The lineage C2 showed a pattern with high diversity regions interspersed with tracts of low diversity. In contrast, the lineage C1 revealed in both stats values ranging close to zero, with lower values spread in the plots, evidencing its lower genetic diversity. Tajima's D values were low and positive in all lineages.

Table 3. Summary of genomic diversity in genetic lineages of *Colletotrichum truncatum* in nonoverlapping 100kb windows.

Lineage	**N[a]**	**K[b]**	**He[c]**	**π[d]**	**D[e]**
C1	8	7.21	2.80	0.0003 (0.000777) *	0.0003 (0.000889)
C2	6	4.50	3.61	0.0092 (0.006963)	0.0085 (0.006574)
C3	5	3.30	4.32	0.0157 (0.008624)	0.0152 (0.008322)

[a] Sample size
[b] Number of haplotypes
[c] Haplotypic diversity
[d] Nucleotide diversity per base pair
[e] Tajima's neutrality statistic.
* Standard Deviation

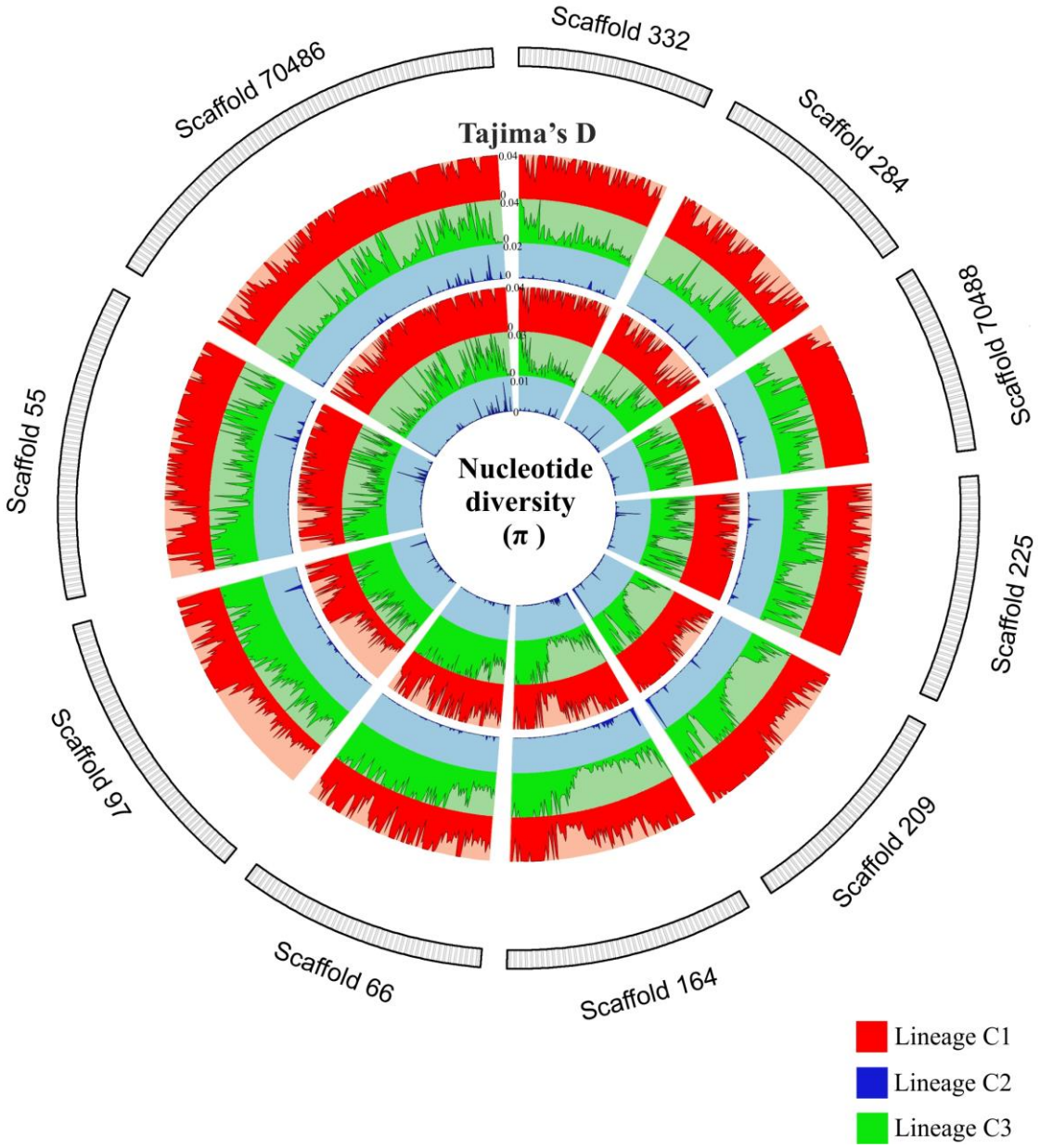


Figure 5. Circos plot of genome variability from the ten largest scaffolds of *Colletotrichum truncatum* genomes. Nucleotide diversity (π) per SNP and Tajima's D calculated in sliding windows of 10kb.

4.4 DISCUSSION

We used a population genomics approach to better understand the distribution of genetic variation within and between *C. truncatum* lineages associated with soybean anthracnose in Brazil. The isolates evaluated in this study were highly pathogenic to soybean in experimental conditions, raising concerns about their potential to cause high losses in soybean producing fields. The re-emergence of the disease in some production areas coupled

with the ubiquitous presence of the pathogen, which is mainly disseminated via seed, have highlighted the need for further investigations about its population biology and evolutionary history. Here we report population genomic analyzes of whole genome sequencing data for eighteen *C. truncatum* isolates collected in two of the main soybeans producing areas in Brazil. Our analyses revealed that the three coexisting genetic lineages of *C. truncatum* were more divergent than what was expected based on multiple gene genealogies. We also found evidence for a history of recombination within and between lineages, with multiples isolates displaying signatures of admixture.

Clustering analyzes clearly supported the existence of three genetic groups in Brazilian *C. truncatum* populations, in agreement with the pattern of population subdivision previously detected based on multilocus microsatellite typing (Rogério et al. 2019). However, our admixture analyses using genomic data revealed that three isolates that appeared pure with microsatellites data had mixed ancestry in two clusters. The inference of individual ancestry coefficients using probabilistic chromosome painting showed that genetic fragments of size longer than our longest scaffolds (2.37 Mb) have been exchanged between clusters, suggesting relatively recent genetic exchanges. Our results also illustrate the power offered by whole genome single-nucleotide polymorphisms regarding analyses of species borders and permeability to gene flow, compared to SSR markers.

Sexual reproduction can contribute to the survival and establishment of invasive fungal populations in novel environment, despite of a limited genetic diversity of founding propagules (Desprez-Loustau et al. 2007; Gladieux et al. 2015). Although sexual reproduction has not been described for *C. truncatum*, either in nature or in laboratory conditions, previous studies reported population genetic signatures of sexual recombination (Diao et al. 2015; Katoch et al. 2017). For some species of the genus, such as *C. sublineolum, C. lindemuthianum,* and *C. graminicola*, the occurrence of parasexual reproduction, hyphal anastomosis process, and cryptic sexual cycle are described as mechanisms responsible to create genetic recombination (Souza-Paccola et al. 2003; Roca et al. 2003; Rosada et al. 2010; Vaillancourt et al. 2000). Our previous inference of the reproductive mode based on microsatellite variation revealed deviation from the null hypothesis of random mating in all three genetic clusters. However, the high rate of unique multilocus microsatellite haplotypes and the presence of reticulations in the phylogenetic network were not consistent with a long history of asexual reproduction (Rogério et al. 2019). Our genomic analyzes confirmed the hypothesis of sexual reproduction, rejecting the null hypothesis of strict clonality. Furthermore, the sharp pattern of LD decay observed in

all three lineages, together with the high rates of recombination estimated, are consistent with a history of relatively recent – if not ongoing – recombination in this species.

The pattern of linkage disequilibrium across a genome reflects the population's history, which may be affected by many evolutionary forces (Stukenbrock and Dutheil 2018; Talas and McDonald 2015; Slatkin 2016). The overall recombination rates estimated in our study may have been significantly impacted by changes in effective population size associated with introduction bottlenecks. Lineage C1 has lowest levels of diversity, suggesting strong bottleneck, which may also have impacted LD patterns, relative to other lineages.

The level of phylogenetic divergence between lineages uncovered in our study suggests that the three lineages were experiencing speciation before their joint introduction in Brazil. The divergence found is too high to have occurred strictly after their introduction since the introduction is relatively recent. The lineages may have evolved independently in allopatry, prior to their introduction in Brazil, without evolving complete reproductive isolation, as reported to allopatric fungal species (Gac and Giraud 2008). Their later introduction in the same areas may have provided opportunities for secondary contact and hybridization. The finding that gene flow does not seem widespread along the genome suggests that partial extrinsic or intrinsic reproductive barriers may currently contribute to restrict gene flow among lineages and somehow maintain the observed pattern of population subdivision.

The distinct diversity levels between clusters can help us to understand how introductions may have affected the current genetic structure of the pathogen. The lower diversity of lineage C1 may suggest a recent introduction, possibly associated with contaminated soybean seeds imported from the U.S. during 1960s and 1970s (Arantes and Miranda 1993; Hirimoto and Vello 1986; Wysmierski and Vello 2013). Regions of low nucleotide diversity in this group corresponded with low Tajima's D values, which could suggest an excess of rare alleles (Tajima 1989). These alleles have probably emerged from a small founder population that may have undergone repeated bottlenecks while spreading geographically in the country. In contrast, the isolates from lineage C3 showed five times higher genetic diversity than isolates from lineage C1 and C2. Indeed, it is much higher than would be generally expected for an organism that has been recently founded, as questioned by genetic paradox of invasion (Blischak et al. 2018). The greater variability observed in this lineage may be explained by: (i) a higher number of founding individuals and more diverse founding sources, followed by admixture; (ii) more pronounced demographic expansion; (iii) the possibility that this lineage may have already been present in Brazil prior to the introduction of soybean, for instance infecting other hosts, such as lima bean, as proposed by earlier studies

(Tiffany and Gilman 1954; Rogério et al. 2019). However, hypotheses about differences in levels of variability across lineages have to be further investigated, for instance fitting demographic models to the genomic data available, in order to shed light on lineage divergence times or number and diversity of propagules introduced in Brazil.

Monitoring of the genetic makeup of *C. truncatum* populations should reveal whether lineages merge through repeated genetic admixture, becoming one genetically homogeneous population, or whether hybridization ceases and result in complete genetic isolation between the lineages. Regardless of the outcome, the monitoring of populations has important implications for disease management since diverse genetic makeups can entail in different arsenals of the pathogen to overcome disease control measures employed.

This study represents the first genetic investigation employing a population approach in this pathosystem. Such approach allowed us to access the current distribution of genetic variation of *C. truncatum* populations across two important areas of soybean production in Brazil, and to propose hypotheses to explain the sources of genetic variation found. The use of the multilocus microsatellite typing schemes had enabled the characterization of a large number of individuals, providing us an overview of the population structure of pathogen and its impact on disease management. The production of the first draft genome of the *C. truncatum* associated with soybean anthracnose represents significant advance towards a better understanding of the population biology and the evolutionary history of this important plant pathogenic fungus. The population genomics analyses have provided insights regarding the evolutionary history of the species, as well as the genetic divergence of the lineages introduced in the country and their sexual recombination potential, what it will serve as a basis for future demographic and evolutionary investigations in this pathosystem. Our results reveal that the *C. truncatum* solution to the paradox of invasion involves repeated introduction from multiple highly divergent sources, and the maintenance of sexual reproduction within and between divergent clusters contribute for that.

REFERENCES

Nylander, J.A.A. (2004) MrModeltest Version 2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University, Uppsala.

EMBRAPA. 1996. Cultura da soja. edited by Ministério da Agricultura Pecuária e Abastecimento. Empresa Brasileira de Pesquisa Agropecuária- Embrapa Soja, Embrapa Cerrados, Embrapa Agropecuária Oeste, Brazil.

Arantes, N. E. and Miranda, M. A. C. 1993. Melhoramento genético e cultivares de soja para o cerrado da região sudeste do Brazil. Pages 225-266 in: Cultura da Soja nos Cerrados. N. E. Arantes and P. I. M. Souza ed, eds. Potafos, Piracicaba

Hirimoto, D. M., and Vello, N. A. 1986. The genetic base of Brazilian soybean cultivars. Braz. J. Genet. 9:295-306.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–10.

Auton, A., and McVean, G. a T. 2007. Recombination rate estimation in the presence of hotspots. Genome Res. 17:1219–1227 Available at: file:///Users/richardeveritt/Documents/Mendeley Desktop/Auton, McVean/2007/Auton, McVean - 2007 - Recombination rate estimation in the presence of hotspots - Genome Research.pdf.

Banke, S., and McDonald, B. A. 2005. Migration patterns among global populations of the pathogenic fungus *Mycosphaerella graminicola*. Mol. Ecol. 14:1881–1896.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 19:455–477.

Blischak, P. D., Mabry, M. E., Conant, G. C., and Pires, J. C. 2018. Is There A Genetic Paradox of Biological Invasion. Annu. Rev. Ecol. Evol. Syst. 47:51–72.

Brumfield, R. T., Beerli, P., Nickerson, D. A., and Edwards, S. V. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. 18:249–256.

Connors, J., Krzywinski, M., Schein, J., Gascoyne, R., Horsman, D., Jones, S. J., et al. 2009. Circos : An information aesthetic for comparative genomics. Genome Res. 19:1639–1645.

Crouch, J., O'Connell, R., Gan, P., Buiate, E., Torres, M. F., Beirn, L., et al. 2014. The Genomics of Colletotrichum. In *Genomics of Plant-Associated Fungi: Monocot Pathogens*, eds. Dean R., Lichens-Park A., and Kole C. , p. 123–140.

Damm, U., Sato, T., Alizadeh, A., Groenewald, J. Z., and Crous, P. W. 2019. The *Colletotrichum dracaenophilum*, *C. magnum* and *C. orchidearum* species complexes. Stud. Mycol. 92:1–46.

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

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158.

Desprez-Loustau, M. L., Robin, C., Buée, M., Courtecuisse, R., Garbaye, J., Suffert, F., et al. 2007. The fungal dimension of biological invasions. Trends Ecol. Evol. 22:472–480.

Diao, Y., Zhang, C., Xu, J., Lin, D., Liu, L., Mtung, O. G., et al. 2015. Genetic differentiation and recombination among geographic populations of the fungal pathogen *Colletotrichum truncatum* from chili peppers in China. Evol. Appl. 8:108–118.

Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., and Gillette, H. D. 2015. The devil is in the details: Genetic variation in introduced populations and its contributions to invasion. Mol. Ecol. 24:2095–2111.

Dlugosch, K. M., and Parker, I. M. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol. Ecol. 17:431–49.

Estoup, A., and Guillemaud, T. 2010. Reconstructing routes of invasion using genetic data: Why, how and so what? Mol. Ecol. 19:4113–4130.

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. 2003. Structure of Linkage Disequilibrium in Plants. Annu. Rev. Plant Biol. 54:357–374.

Frankham, R. 2005. Resolving the genetic paradox in invasive species. Heredity. 94:385.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. 2014. A short manual for sNMF: a program to estimate ancestry coefficients (command-line version). Genetics. 196:973–983.

Gac, M. Le, and Giraud, T. 2008. Existence of a pattern of reproductive character displacement in Homobasidiomycota but not in Ascomycota. J. Evol. Biol. 21:761–772.

Gladieux, P., Feurtey, A., Hood, M. E., Snirc, A., Clavel, J., Dutech, C., et al. 2015. The population biology of fungal invasions. Mol. Ecol. 24:1969–1986.

Gross, A., Hosoya, T., and Queloz, V. 2014. Population structure of the invasive forest pathogen *Hymenoscyphus pseudoalbidus*. Mol. Ecol. 23:2943–2960.

Grünwald, N. J., McDonald, B. A., and Milgroom, M. G. 2016. Population Genomics of Fungal and Oomycete Pathogens. Annu. Rev. Phytopathol. 54:323–346.

Hartman, G. L., Bowen, C. R., Haudenshield, J. S., Fox, C. M., Cary, T. R., and Diers, B. W. 2015. Evaluation of disease and pest damage on soybean cultivars released from 1923 through 2008 under field conditions in central Illinois. Agron. J. 107:2373–2380.

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., et al. 2011. Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. Mol. Ecol. Resour. 11:123–136.

Hill, W. G., and Robertson, A. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226–231.

Hyde, K. ., Cai, L., McLenzie, E. H. C., Yang, Y. L., Zhang, J. Z., and Prihastuti, H. 2009. *Colletotrichum* : a catalogue of confusion. Fungal Divers. 39:1–17.

Jombart, T., and Ahmed, I. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics. 27:3070–3071.

Katoch, A., Sharma, P., Padder, B. A., and Sharma, P. N. 2017. Population Structure of *Colletotrichum truncatum* in Himachal Pradesh and Identification of Broad-Spectrum Resistant Sources in *Capsicum*. Agric. Res. 6:296–303.

Kimati, H., and Galli, F. 1970. *Glomerella cingulata* (STONEM.) SPAULD. ET V. SCHRENK. F. SP. *phaseoli* N.F., FASE ASCÓGENA DO AGENTE CAUSAL DA ANTRACNOSE DO FEIJOEIRO. An. da Esc. Super. Agric. “Luiz Queiroz.” 27:437.

Lawson, D. J., Hellenthal, G., Myers, S., and Falush, D. 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8:11–17.

Leaché, A. D., Banbury, B. L., Felsenstein, J., De Oca, A. N. M., and Stamatakis, A. 2015. Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64:1032–1047.

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–2079.

Linde, C. C., Zala, M., and McDonald, B. A. 2009. Molecular Phylogenetics and Evolution Molecular evidence for recent founder populations and human-mediated migration in the barley scald pathogen *Rhynchosporium secalis*. Mol. Phylogenet. Evol. 51:454–464.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., and Taberlet, P. 2003. The power and promise of population genomics: From genotyping to genome typing. Nat. Rev. Genet. 4:981–994.

McKenn, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. 2009. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Milgroon, M. G., Sotirovski, K., Spica, D., Davis, J. E., Brewer, M. T., Milev§, M., et al. 2008. Clonal population structure of the chestnut blight fungus in expanding ranges in southeastern Europe. Mol. Ecol. 17:4446–4458.

De Mita, S., Siol, M., Schuster, S., Stajich, J., Block, D., Boulez, K., et al. 2012. EggLib: processing, analysis and simulation tools for population genetics and genomics. BMC Genet. 13:27.

Roca, M. G., Davide, L. C., Mendes-Costa, M. C., and Wheals, A. 2003. Conidial anastomosis tubes in *Colletotrichum*. Fungal Genet. Biol. 40:138–145.

Rogério, F., Ciampi-Guillardi, M., Barbieri, M. C. G., Bragança, C. A. D., Seixas, C. D. S., Almeida, A. M. R., et al. 2017. Phylogeny and variability of *Colletotrichum truncatum* associated with soybean anthracnose in Brazil. J. Appl. Microbiol. 122:402–415.

Rogério, F., Gladieux, P., Massola, N. S., and Ciampi-Guillardi, M. 2019. Multiple Introductions Without Admixture of *Colletotrichum truncatum* Associated with Soybean Anthracnose in Brazil. Phytopathology. 109:681–689.

Ronquist, F., and Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Rosada, L. J., Franco, C. C. S., Sant'anna, J. R., Kaneshima, E. N., Gonçalves-Vidigal, M. C., and Castro-Prado, M. A. A. 2010. Parasexuality in Race 65 *Colletotrichum lindemuthianum* Isolates. J. Eukaryot. 57:383–384.

Schrieber, K., and Lachmuth, S. 2017. The genetic paradox of invasions revisited: The potential role of inbreeding × environment interactions in invasion success. Biol. Rev. 92:939–952.

Slatkin, M. 2016. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9:477–485.

Souza-Paccola, E. A., Fávaro, L. C. L., Casela, C. R., and Paccola-Meirelles, L. D. 2003. Genetic Recombination in *Colletotrichum sublineolum*. J. Phytopathol. 151:329–334.

Stukenbrock, E. H., and Dutheil, J. Y. 2018. Fine-Scale Recombination Maps of Fungal Plant Pathogens Reveal Dynamic Recombination Landscapes and Intragenic Hotspots. Genetics. 208:1209–1229.

Tajima, F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics. 123:585–595.

Talas, F., and McDonald, B. A. 2015. Genome-wide analysis of *Fusarium graminearum* field populations reveals hotspots of recombination. BMC Genomics. 16:1–12.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24:1596–1599.

Tiffany, A. L. H., and Gilman, J. C. 1954. Species of *Colletotrichum* from Legumes. Myco. 46:52–75.

Vaillancourt, L. J., and Hanau, R. M. 1991. A Method for Genetic Analysis of *Glomerella graminicola (Colletotrichum graminicola)* from Maize. Pythopathology. 81:530–534.

Vaillancourt, L., Wang, J., Hanau, R., Rollins, J., and Du, M. 2000. Genetic Analysis of Cross Fertility between Two Self-Sterile Strains of *Glomerella graminicola*. Mycologia. 92:430.

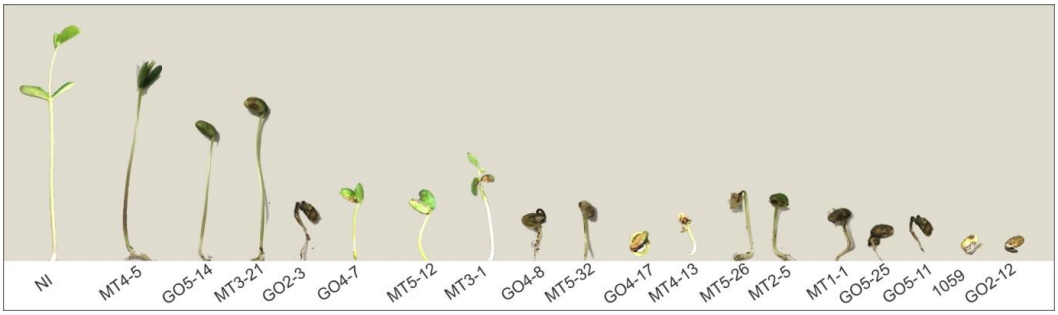
Wysmierski, P. T., and Vello, N. A. 2013. The genetic base of Brazilian soybean cultivars: Evolution over time and breeding implications. Genet. Mol. Biol. 36:547–555.

SUPPORTING INFORMATION

Table S1. Pathogenicity of *Colletotrichum truncatum* isolates used in this study on soybean 'IPRO 8579'.

Isolated	Symptoms[1]
Control	0 B
MT4_05	0 B
GO5_14	1.67 AB
MT3_21	1.67 AB
GO2_03	2 AB
GO4_07	2 AB
MT5_12	2.33 AB
MT3_01	3 AB
GO4_08	3.33 AB
MT5_32	3.33 Ab
GO4_17	3.67 AB
MT4_13	3.67 AB
MT5_26	3.67 AB
MT2_05	4 AB
MT1_01	4.33 AB
GO5_25	4.33 AB
GO5_11	4.33 AB
GO2_12	5 AB
CMES1059	5 AB
Mean	1.96

[1] Symptoms according Hartman et al. (2015). Means followed by the same letter are not significantly different (P< 0.05) according to the Tukey test applied over transformed data (X+1)^0.5.



Fi**gure 1S**. Symptoms of anthracnose in soybean 'IPRO 8579' seedlings inoculated with eighteen *Colletotrichum truncatum* isolates used in this study.