SAMUEL ALVES DOS SANTOS

GENOMIC AND TRANSCRIPTOMIC ANALYSES OF *Austropuccinia psidii* AND *Ceratocystis fimbriata*, THE CAUSAL AGENTS OF MYRTLE RUST AND CERATOCYSTIS WILT, RESPECTIVELY

Thesis submitted to the Plant Pathology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Acelino Couto Alfenas

Co-adviser: Pedro Marcus Pereira Vidigal

VIÇOSA - MINAS GERAIS
2020

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

S237g
2020

Santos, Samuel Alves dos, 1990-
Genomic and transcriptomic analyses of *Austropuccinia psidii* and *Ceratocystis fimbriata*, the causal agents of myrtle rust and ceratocystis wilt, respectively / Samuel Alves dos Santos. - Viçosa, MG, 2020.
126 f. : il. (algumas color.) ; 29 cm.

Orientador: Acelino Couto Alfenas.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Plantas - Resistência a doenças e pragas. 2. Análise do genoma. 3. Expressão gênica. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. II. Título.

CDD 22 ed. 632.4

SAMUEL ALVES DOS SANTOS

GENOMIC AND TRANSCRIPTOMIC ANALYSES OF *Austropuccinia psidii* AND *Ceratocystis fimbriata*, THE CAUSAL AGENTS OF MYRTLE RUST AND CERATOCYSTIS WILT, RESPECTIVELY

Thesis submitted to the Plant Pathology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: February 21st, 2020.

Assent:

Samuel Alves dos Santos
Author

Acelino Couto Alfenas
Adviser

To my family especially my mom (*in memoriam*),
who once told me that her biggest dream would see
one of her kids became a doctor.

ACKNOWLEDGMENTS

I would like to express my deep acknowledges:

To my whole family, especially to my parents (Aristeu and Maria), brothers (Vando, Izael, and Lucas), and sisters (Vanda, Vanusa, and Edna) who always have given me unconditional love, incentive, and support.

To Prof. Acelino Couto Alfenas who is my principal mentor. For the last five years, I have had the opportunity to learn about Forest Pathology and Science with an excellent professor, researcher, and professional. I am so grateful for all opportunities he has provided me. I will take with me every lesson.

To Dr. Pedro Marcus Pereira Vidigal who introduced me to the bioinformatics world and have helped me in all bioinformatics analyses during my whole PhD. I am so grateful for all bioinformatic training, support, and friendship during this journey.

To Dr. Matt Templeton for all material and intellectual support for carrying out some bioinformatic analysis in New Zealand and especially for the friendship, advice, and all personal support.

To the Universidade Federal de Viçosa and the Plant Pathology Graduate Program for the quality of education.

To the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)" for the financial support and concession of three-years PhD scholarship.

To the "Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)" for the financial support and concession of one-years PhD "sandwich" scholarship, which gave me the opportunity to complete my PhD in Auckland, New Zeland.

To Suzano Papel e Celulose S. A., especially Dr. Reginaldo Mafia and Dr. Edival Zauza for the financial support.

To Zespri International Limited for the financial support granted to Dr. Matt Templeton, which was crucial to sequencing several Ceratocystis genomes.

To the "Núcleo de Análise de Biomoléculas (NuBioMol)" for providing bioinformatics facilities and supporting data analysis.

To the New Zealand Institute for Plant and Food Research (PFR) – site Mt. Albert, Auckland – for providing me all support and computing resources from PowerPlant cluster to execute the majority of bioinformatics analysis.

To "Clonar Resistência a Doenças Florestais LTDA" and all its staff for the support offered for growing all plants used in my experiments. A special thanks to Solange, Paulo, and Eduardo for take care of my plants so well.

To Dr. Lúcio Guimarães for all support and advice during experiments planning, as well as the friendship.

To Márcia Brandão for all support and help with bureaucratic and paperwork stuff, as well as the friendship.

To all friends and team of the Laboratory of Forest Pathology (Patomol), especially to Camilla Concesso, Camila Ribeiro, Fernando Montezano, Mara, Elenice, Ana Cristina, and Paulo for their help during some experiments.

To all friends of the Bioprotection team from PRF (New Zealand), especially to Claudia Lang, Mark Andersen, Jay Jayaraman, and Liz Valentina Lucarelli, and Roshan Khadka for the help, friendship, and good moments.

To all my dear friends that I have made in New Zealand, especially to David Freitas, Valentina Lucarelli, Danilo Nocelli, and Roshan Khadka – thanks for the unforgettable moments.

To my New Zealand hosted – Gay Kelly – who was so lovely and kind to me. Thanks for all support, friendship, and special moments in New Zealand.

To the Brazilian people who work so hard to pay high taxes that allowed me to have a free education in a public University with scholarships in Brazil and abroad.

To all that I forgot to mention, but certainly are important to me, thank you.

“Risk, dedication, focus, goals, persistence, failure, action, sacrifice, habits, and hard work: All come before Success – in the dictionary and in life”
(Jim Kwik)

ABSTRACT

SANTOS, Samuel Alves dos, D.Sc., Universidade Federal de Viçosa, February, 2020. **Genomic and transcriptomic analyses of *Austropuccinia psidii* and *Ceratocystis fimbriata*, the causal agents of myrtle rust and ceratocystis wilt, respectively**. Adviser: Acelino Couto Alfenas. Co-adviser: Pedro Marcus Pereira Vidigal.

Austropuccinia psidii and *Ceratocystis fimbriata* complex, the causal agents of Myrtle rust and Ceratocystis wilt, respectively, have been considered important fungal pathogens especially due to their wide and expanding host range across the world. Whole genome sequencing and RNA-Seq are powerful tools that provide information about pathogens biology, plasticity, evolution, and pathogenicity, which are crucial to a better understanding of host-pathogen interactions. Therefore, in chapter 1 of this Thesis, we present a high-quality assembly and annotation of the haploid nuclear genome and *de novo* transcriptome of *A. psidii* epitype guava (*Psidium guajava*) isolate, which is so far the largest genome among Pucciniales rust fungi. Additionally, in chapter 2, we evaluated the gene expression profile of two contrasting *Eucalyptus grandis* genotypes in resistance level to rust (*A. psidii*) and found a constitutive overexpression of several resistance-related genes in the resistant genotype compared to the susceptible one. In chapter 3, we present *de novo* genome assembly and annotation for the nuclear genome of 21 fungal pathogens of the *C. fimbriata* complex from different host species and geographic region. Comparative genomic analysis revealed that independently on host and geographic location, the majority of genome features showed a high level of similarity among all *Ceratocystis* isolates. Moreover, in order to understand the molecular mechanisms related to *C. fimbriata* pathogenicity on *Eucalyptus* spp., in chapter 4, we evaluated the first transcriptome of *C. fimbriata* LPF1912 isolate during its infection on 16 eucalyptus genotypes. Through comparative genomic analysis with *C. eucalypticola* and *Calonectria pseudoreteaudii*, which also infect eucalyptus, we identified different pathogenicity-related genes among the three eucalyptus pathogens. Finally, our genomic and transcriptomic dataset analyses provide a valuable framework for future studies leading to a better comprehension of the biology, plasticity, evolution, and pathogenicity of *A. psidii* and *C. fimbriata* complex.

Keywords: *Puccinia psidii*. Comparative genomic analysis. Differentially expressed genes. RNA-Seq. OrthoMCL analysis. Plant resistance.

RESUMO

SANTOS, Samuel Alves dos, D.Sc., Universidade Federal de Viçosa, fevereiro de 2020. **Genômina e transcriptômica de *Austropuccinia psidii* and *Ceratocystis fimbriata*, agentes causais da ferrugem das Myrtaceae e murcha de Ceratocystis, respectivamente**. Orientador: Acelino Couto Alfenas. Coorientador: Pedro Marcus Pereira Vidigal.

Austropuccinia psidii e *Ceratocystis fimbriata*, agentes causais da ferrugem das Myrtaceae e murcha de Ceratocystis, respectivamente, tem sido considerados como importantes fitopatógenos principalmente devido a sua ampla e crescente gama de hospedeiros ao redor do mundo. Sequenciamento de genoma completo e RNA-Seq são ferramentas importantes, as quais geram informações sobre a biologia, plasticidade e virulência de patógenos, as quais são fundamentais para uma melhor compreensão das interações patógeno-hospedeiro. Nesse sentido, no capítulo 1 desta Tese, apresentamos uma montagem *de novo* e anotação de alta qualidade para o genoma nuclear haplóide de *A. psidii* isolado de goiabeira (*Psidium guajava*), o qual é ate o momento o maior genoma conhecido dentre os fungos Pucciniales causadores de ferrugem. No capítulo 2, avaliamos o perfil de expressão gênica de dois clones de *Eucalyptus grandis* contrastantes em relação ao nível de resistência a ferrugem (*A. psidii*) e observamos uma superexpressão constitutiva de genes relacionados a resistência no genótipo resistente comparado ao suscetível. Adicionalmente, no capítulo 3, apresentamos uma montagem *de novo* e anotação do genoma nuclear de 21 patógenos do complexo *C. fimbriata*, oriundos de diferentes espécies hospedeiras e regiões geográficas. Análises em genômica comparativa revelaram que independente do hospedeiro ou região, a maioria das características do genoma tiveram alto grau de semelhança dentre os 21 patógenos. Além disso, na busca de um melhor entendimento dos mecanismos moleculares relacionados a patogenicidade de *C. fimbriata* em *Eucalyptus* spp., no capítulo 4, avaliamos o primeiro transcriptoma de *C. fimbriata* LPF1912 durante sua infecção em 16 clones de eucalipto. Por meio de genômica comparativa com *C. eucalypticola* and *Calonectria pseudoreteaudii*, os quais também são patogênicos ao eucalipto, observamos diferente genes relacionados à patogenicidade entre esses três patógenos de eucalipto. Os dados apresentados neste estudo irão servir de base para estudos futuros levando a uma melhor compreensão da biologia, plasticidade, evolução e patogenicidade de *A. psidii* and *C. fimbriata*.

Palavras-chave: *Puccinia psidii*. Genômica comparativa. Genes diferencialmente expressos. RNA-Seq. OrthoMCL análises. Resistência.

SUMMARY

GENERAL INTRODUCTION

Plant diseases, especially those caused by fungal pathogens, always have been threatened the crops and natural vegetation across the world. Among the several fungal plant pathogens, *Austropuccinia psidii* (Winter) Beenken and *Ceratocystis fimbriata* Ellis & Halsted, the causal agents of myrtle rust (MR) and ceratocystis wilt (CW), respectively, have gotten special attention due to mainly their wide and expanding host range and geographic distribution.

Austropuccinia psidii is a biotrophic fungus that was first reported as *Puccinia psidii* in 1884 infecting guava plants (*Psidium guajava*) in São Francisco do Sul, Santa Catarina State, Brazil (Winter 1884). Since then it has spread from South to North America (MacLachlan 1938; Marlatt and Kimbrough 1979; Coutinho et al. 1998; Uchida et al. 2006). The fungus infects several host species within the Myrtaceae family and has been spreading to other continents, including Asia (Kawanishi et al. 2009; Zhuang and Wei 2011; McTaggart et al. 2016; du Plessis et al. 2017), Africa (Roux et al. 2013), and Oceania (Carnegie et al. 2010; Beresford et al. 2018). MR symptoms depend on the susceptibility of the host, however, may range from small pustules with few spores to massive production of powdery bright yellow urediniospores with shoot dieback that may ultimately result in plant death (Alfenas et al. 2009). As susceptible species are dominant components of flora in widespread global areas, *A. psidii* is a constant risk to native biodiversity and disease outbreaks could severely alter the structure, composition, and function of these ecosystems (Pegg et al. 2014; Carnegie et al. 2016; Pegg et al. 2017; Beresford et al. 2018). In Brazil, MR has significant economic importance in *Eucalyptus* spp. plantations (Alfenas et al. 2003; Alfenas et al. 2009), especially in the first year after planting with eucalyptus trees up to 0.5-3.0 m tall (Zauza et al. 2010).

Ceratocystis fimbriata is considered a species complex (de Beer et al. 2014) and also has a wide host range that includes many crops and native tree species around the world, such as sweet potato (*Ipomoea batatas*), eucalyptus (*Eucalyptus* spp.), teak (*Tectona grandis*), rubber tree (*Hevea brasiliensis*), mango (*Mangifera indica*), kiwifruit (*Actinidia* sp.), andiroba (*Carapa guianensis*), aroids (*Colocasia esculenta*), acacia (*Acacia* sp.), pomegranate (*Punica granatum*), ʻōhiʻa (*Metrosideros polymorpha*), among others (Ferreira et al. 1999; Huang et al. 2003; Harrington et al. 2005; Firmino et al. 2012; Valdetaro et al. 2015; Piveta et al. 2016; Galdino et al. 2016; Alam et al. 2017; Barnes et al. 2018; Valdetaro et al. 2019). In Brazil, this fungal pathogen is particularly important in eucalyptus plantations. CW was first detected in Brazil on *Eucalyptus* sp. during the 1990s (Ferreira et al. 1999), since then it has become a serious disease greatly threatening eucalyptus plantations (Oliveira et al. 2015a), affecting

growth and wood quality, as well as causing death of eucalyptus trees in the field (Zauza et al. 2004; Rosado et al. 2010; Mafia et al. 2013; Fernandes et al. 2014). Commonly, the infection begins at the roots or at the base of the stem, with dark brown to black radial streaks of the woody xylem tissue observed as colonization advances and causes cell death in several plant tissues such as ray parenchyma, vascular cambium, phloem and phelloderm (Ferreira et al. 2006). The visible symptoms in susceptible eucalyptus genotype infected by *C. fimbriata* are wilt of the canopy, branch death, and consequently death of the entire tree (Ferreira et al. 2006; Alfenas et al. 2009; Roux and Wingfield 2009; Ferreira et al. 2013).

Over the years, the Brazilian forest companies have been introduced the resistance to these two diseases as a treat in their breeding programs. Thus, in the *Eucalyptus* spp. plantations in Brazil, the major strategies to manage both diseases have been the selection and planting of resistant eucalyptus genotypes (Zauza et al. 2004; Rosado et al. 2010; Miranda et al. 2013; Silva et al. 2013; Santos et al. 2014). Understanding the genetic diversity of the pathogen population, as well as the molecular mechanisms involved in resistance, is crucial to obtain a genotype with robust and durable resistance in the field. Studies to access the genetic diversity of *A. psidii* and *C. fimbriata* populations have been focused mainly in genotyping analysis using microsatellite markers (Ferreira et al. 2010; Zhong et al. 2011; Ferreira et al. 2011; Simpson et al. 2013; Graça et al. 2013; Machado et al. 2015; Oliveira et al. 2015b; Stewart et al. 2018). Advances on the whole genome sequencing (WGS) and transcriptome sequencing (RNA-Seq) technologies have been provided high-quality genome and transcriptome sequences much more rapidly and cheaply (Goodwin et al. 2016). WGS has already been applied in plant pathology and in plant-microbe interaction research to infer the evolution of pathogens, pathogenicity factors, and virulence genes (Cantu et al. 2011). In addition, it also has been employed in comparative analyses between different races or species (Duplessis et al. 2011), as well as the identification of effector proteins (Joly et al. 2010). Moreover, RNA-Seq is an approach to assess transcriptome profiling, which provides expression information of a large number of transcripts and their isoforms (Wang et al. 2009). Recently, this tool has been widely applied in transcriptome analysis in several plant-pathogen interactions (Zhu et al. 2013; Hayden et al. 2014; Liu et al. 2015; Chakraborty et al. 2016; Ye et al. 2017; Hsieh et al. 2018; Tobias et al. 2018), however, it is still unexplored for *Eucalyptus grandis* - *A. psidii* and *E. grandis* - *C. fimbriata* pathosystems.

This study aimed: i) To obtain a high-quality assembly and characterization of the genome structure and gene content of *A. psidii*; ii) To evaluate the transcriptome profiling of *E. grandis*

genotypes inoculated with *A. psidii*; iii) To obtain a high-quality genome assembly of 21 *C. fimbriata* isolates from different host species; and iv) To identify the main pathogenicity-related genes of *C. fimbriata* during its infection on *Eucalyptus* sp. genotypes. Therefore, this Thesis comprises the following four chapters: 1) Genome assembly and annotation of *Austropuccinia psidii* epitype guava (*Psidium guajava*) isolate, the causal agent of myrtle rust; 2) Transcriptome analysis of *Eucalyptus grandis* genotypes reveals constitutive overexpression of resistance-related genes to rust (*Austropuccinia psidii*); 3) Comparative genomic analysis of 22 fungal isolates of the *Ceratocystis fimbriata* complex; and 4) Comparative genomic and transcriptomic analyses reveal different pathogenicity-related genes among three eucalyptus fungal pathogens.

REFERENCES

Alam MW, Gleason ML, Mehboob S, Riaz K, Rehman A (2017) First Report of *Ceratocystis fimbriata* Causing Pomegranate Wilt in Pakistan. Plant Dis 101:251–251. https://doi.org/10.1094/PDIS-06-16-0835-PDN

Alfenas AC, Zauza EA. V, Assis TF (2003) First record of Puccinia psidii on Eucalyptus globulus and E. viminalis in Brazil. Australas Plant Pathol 32:325–326. https://doi.org/10.1071/AP03021

Alfenas AC, Zauza EAV, Mafia RG, Assis TF (2009) Clonagem e doenças do eucalipo, 1st edn. Editora UFV, Viçosa, MG state, Brazil.

Barnes I, Fourie A, Wingfield MJ, Harrington TC, McNew DL, Sugiyama LS, Luiz BC, Heller WP, Keith LM (2018) New ceratocystis species associated with rapid death of metrosideros polymorpha in hawai`i. Persoonia Mol Phylogeny Evol Fungi 40:154–181. https://doi.org/10.3767/persoonia.2018.40.07

Beresford RM, Turner R, Tait A, Paul V, Macara G, Yu ZD, Lima L, Martin R (2018) Predicting the climatic risk of myrtle rust during its first year in New Zealand. New Zeal Plant Prot 71:332–347. https://doi.org/10.30843/nzpp.2018.71.176

Cantu D, Govindarajulu M, Kozik A, Wang M, Chen X, Kojima KK, Jurka J, Michelmore RW, Dubcovsky J (2011) Next Generation Sequencing Provides Rapid Access to the Genome of Puccinia striiformis f. sp. tritici, the Causal Agent of Wheat Stripe Rust. PLoS One 6:e24230. https://doi.org/10.1371/journal.pone.0024230

Carnegie AJ, Kathuria A, Pegg GS, Entwistle P, Nagel M, Giblin FR (2016) Impact of the invasive rust Puccinia psidii (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. Biol Invasions 18:127–144. https://doi.org/10.1007/s10530-015-0996-y

Carnegie AJ, Lidbetter JR, Walker J, Horwood MA, Tesoriero L, Glen M, Priest MJ (2010) Uredo rangelii, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. Australas Plant Pathol 39:463–466. https://doi.org/10.1071/AP10102

Chakraborty S, Britton M, Martínez-García PJ, Dandekar AM (2016) Deep RNA-Seq profile reveals biodiversity, plant–microbe interactions and a large family of NBS-LRR resistance genes in walnut (Juglans regia) tissues. AMB Express 6:12. https://doi.org/10.1186/s13568-016-0182-3

Coutinho TA, Wingfield MJ, Alfenas AC, Crous PW (1998) Eucalyptus rust: A disease with the potential for serious international implications. Plant Dis 82:819–825. https://doi.org/10.1094/PDIS.1998.82.7.819

du Plessis E, McTaggart AR, Granados GM, Wingfield MJ, Roux J, Ali MIM, Pegg GS, Makinson J, Purcell M (2017) First Report of Myrtle Rust Caused by *Austropuccinia psidii* on *Rhodomyrtus tomentosa* (Myrtaceae) from Singapore. Plant Dis 101:1676–1676. https://doi.org/10.1094/PDIS-04-17-0530-PDN

Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, Chiu R, Coutinho PM, Feaue N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr MG, Kodira CD, Kohler A, Kües U, Lindquist EA, Lucas SM, Mago R, Mauceli E, Morin E, Murat C, Pangilinan JL, Park R, Pearson M, Quesneville H, Rouhier N, Sakthikumar S, Salamov AA, Schmutz J, Selles B, Shapiro H, Tanguay P, Tuskan GA, Henrissat B, Van De Peer Y, Rouzé P, Ellis JG, Dodds PN, Schein JE, Zhong S, Hamelin RC, Grigoriev I V., Szabo LJ, Martin F (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci U S A 108:9166–9171. https://doi.org/10.1073/pnas.1019315108

Fernandes B V., Zanuncio AJ V., Furtado EL, Andrade HB (2014) Damage and Loss due to Ceratocystis fimbriata in Eucalyptus wood for charcoal production. BioResources 9:5473–5479.

Ferreira EM, Harrington TC, Thorpe DJ, Alfenas AC (2010) Genetic diversity and interfertility among highly differentiated populations of Ceratocystis fimbriata in Brazil. Plant Pathol 59:721–735. https://doi.org/10.1111/j.1365-3059.2010.02275.x

Ferreira FA, Demuner AMM, Demuner NL, Pigato S (1999) Murcha-de-Ceratocystis em eucalipto no Brasil. Fitopatol Bras 24:p.284.

Ferreira FA, Maffia LA, Barreto RW, Demuner NL, Pigatto S (2006) Symptomatology of Ceratocystis wilt in eucalyptus. Rev Árvore 30:155–162.

Ferreira MA, Harrington TC, Alfenas AC, Mizubuti ESG (2011) Movement of Genotypes of *Ceratocystis fimbriata* Within and Among *Eucalyptus* Plantations in Brazil. Phytopathology 101:1005–1012. https://doi.org/10.1094/PHYTO-01-11-0015

Ferreira MA, Harrington TC, Gongora-Canul CC, Mafia RG, Zauza EA V., Alfenas AC (2013) Spatial-temporal patterns of Ceratocystis wilt in *Eucalyptus* plantations in Brazil. For Pathol 43:153–164. https://doi.org/10.1111/efp.12013

Firmino AC, Tozze Jr HJ, Furtado EL (2012) First report of *Ceratocystis fimbriata* causing wilt in *Tectona grandis* in Brazil. New Dis Reports 25:24. https://doi.org/10.5197/j.2044-0588.2012.025.024

Galdino TV da S, Kumar S, Oliveira LSS, Alfenas AC, Neven LG, Al-Sadi AM, Picanço MC (2016) Mapping Global Potential Risk of Mango Sudden Decline Disease Caused by Ceratocystis fimbriata. PLoS One 11:e0159450. https://doi.org/10.1371/journal.pone.0159450

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: Ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17:333–351.

Graça RN, Ross-Davis AL, Klopfenstein NB, Kim M-S, Peever TL, Cannon PG, Aun CP, Mizubuti ESG, Alfenas AC (2013) Rust disease of eucalypts, caused by *Puccinia psidii* , did not originate via host jump from guava in Brazil. Mol Ecol 22:6033–6047. https://doi.org/10.1111/mec.12545

Harrington TC, Thorpe DJ, Marinho VLA, Furtado EL (2005) First report of black rot of Colocasia esculenta caused by Ceratocystis fimbriata in Brazil. Fitopatol Bras 30:88–89. https://doi.org/10.1590/s0100-41582005000100017

Hayden KJ, Garbelotto M, Knaus BJ, Cronn RC, Rai H, Wright JW (2014) Dual RNA-seq of the plant pathogen Phytophthora ramorum and its tanoak host. Tree Genet Genomes 10:489–502. https://doi.org/10.1007/s11295-014-0698-0

Hsieh J-F, Chuah A, Patel HR, Sandhu KS, Foley WJ, Külheim C (2018) Transcriptome Profiling of *Melaleuca quinquenervia* Challenged by Myrtle Rust Reveals Differences in Defense Responses Among Resistant Individuals. Phytopathology 108:495–509. https://doi.org/10.1094/PHYTO-09-17-0307-R

Huang Q, Zhu YY, Chen HR, Wang YY, Liu YL, Lu WJ, Ruan XY (2003) First Report of Pomegranate Wilt Caused by *Ceratocystis fimbriata* in Yunnan, China. Plant Dis 87:1150–1150. https://doi.org/10.1094/PDIS.2003.87.9.1150B

Joly DL, Feau N, Tanguay P, Hamelin RC (2010) Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (Melampsora spp.). BMC Genomics 11:422. https://doi.org/10.1186/1471-2164-11-422

Kawanishi T, Uematsu S, Kakishima M, Kagiwada S, Hamamoto H, Horie H, Namba S (2009) First report of rust disease on ohia and the causal fungus, Puccinia psidii, in Japan. J Gen Plant Pathol 75:428–431. https://doi.org/10.1007/s10327-009-0202-0

Liu J-J, Sturrock RN, Sniezko RA, Williams H, Benton R, Zamany A (2015) Transcriptome analysis of the white pine blister rust pathogen Cronartium ribicola: de novo assembly, expression profiling, and identification of candidate effectors. BMC Genomics 16:678. https://doi.org/10.1186/s12864-015-1861-1

Machado P da S, Alfenas AC, Alfenas RF, Mohammed CL, Glen M (2015) Microsatellite analysis indicates that Puccinia psidii in Australia is mutating but not recombining. Australas Plant Pathol 44:455–462. https://doi.org/10.1007/s13313-015-0364-5

MacLachlan JD (1938) A rust of the pimento tree in Jamaica. Phytopathology 28:157–170.

Mafia RG, Ferreira MA, Zauza EA V., Silva JF, Colodette JL, Alfenas AC (2013) Impact of Ceratocystis wilt on eucalyptus tree growth and cellulose pulp yield. For Pathol 43:379–385. https://doi.org/10.1111/efp.12041

Marlatt RB, Kimbrough JW (1979) Puccinia psidii on Pimenta dioica in south Florida. Plant Dis 63:510–512.

McTaggart AR, Roux J, Granados GM, Gafur A, Tarrigan M, Santhakumar P, Wingfield MJ (2016) Rust (Puccinia psidii) recorded in Indonesia poses a threat to forests and forestry in South-East Asia. Australas Plant Pathol 45:83–89. https://doi.org/10.1007/s13313-015-0386-z

Miranda AC, de Moraes MLT, Tambarussi EV, Furtado EL, Mori ES, da Silva PHM, Sebbenn AM (2013) Heritability for resistance to Puccinia psidii Winter rust in Eucalyptus grandis Hill ex Maiden in Southwestern Brazil. Tree Genet Genomes 9:321–329. https://doi.org/10.1007/s11295-012-0572-x

Oliveira LSS, Guimarães LMS, Ferreira MA, Nunes AS, Pimenta LVA, Alfenas AC (2015a) Aggressiveness, cultural characteristics and genetic variation of Ceratocystis fimbriata on Eucalyptus spp. For Pathol 45:505–514. https://doi.org/10.1111/efp.12200

Oliveira LSS, Harrington TC, Ferreira MA, Damacena MB, Al-Sadi AM, Al-Mahmooli IHS, Alfenas AC (2015b) Species or Genotypes? Reassessment of Four Recently Described Species of the Ceratocystis Wilt Pathogen, *Ceratocystis fimbriata* , on *Mangifera indica*. Phytopathology 105:1229–1244. https://doi.org/10.1094/PHYTO-03-15-0065-R

Pegg G, Taylor T, Entwistle P, Guymer G, Giblin F, Carnegie A (2017) Impact of Austropuccinia psidii (myrtle rust) on Myrtaceae-rich wet sclerophyll forests in south east Queensland. PLoS One 12:e0188058. https://doi.org/10.1371/journal.pone.0188058

Pegg GS, Giblin FR, McTaggart AR, Guymer GP, Taylor H, Ireland KB, Shivas RG, Perry S (2014) *Puccinia psidii* in Queensland, Australia: disease symptoms, distribution and impact. Plant Pathol 63:1005–1021. https://doi.org/10.1111/ppa.12173

Piveta G, Ferreira M, FB Muniz M, Valdetaro D, Valdebenito-Sanhueza R, Harrington T, Alfenas A (2016) *Ceratocystis fimbriata* on kiwifruit (*Actinidia* spp.) in Brazil. New Zeal J Crop Hortic Sci 44:13–24. https://doi.org/10.1080/01140671.2016.1143020

Rosado CCG, Guimarães LMDS, Titon M, Lau D, Rosse L, Resende MDV De, Alfenas AC (2010) Resistance to Ceratocystis Wilt (Ceratocystis fimbriata) in Parents and Progenies of Eucalyptus grandis x E. urophylla. Silvae Genet 59:99–106. https://doi.org/10.1515/sg-2010-0012

Roux J, Greyling I, Coutinho TA, Verleur M, Wingfeld MJ (2013) The Myrtle rust pathogen, Puccinia psidii, discovered in Africa. IMA Fungus 4:155–159. https://doi.org/10.5598/imafungus.2013.04.01.14

Roux J, Wingfield M (2009) *Ceratocystis* species: emerging pathogens of non-native plantation *Eucalyptus* and *Acacia* species. South For a J For Sci 71:115–120. https://doi.org/10.2989/SF.2009.71.2.5.820

Santos MR, Guimarães LM da S, Resende MDV de, Rosse LN, Zamprogno KC, Alfenas AC (2014) Resistance of Eucalyptus pellita to rust (Puccinia psidii). Crop Breed Appl Biotechnol 14:244–250. https://doi.org/10.1590/1984-70332014v14n4a38

Silva PHM, Miranda AC, Moraes MLT, Furtado EL, Stape JL, Alvares CA, Sentelhas PC, Mori ES, Sebbenn AM (2013) Selecting for rust (Puccinia psidii) resistance in Eucalyptus grandis in São Paulo State, Brazil. For Ecol Manage 303:91–97. https://doi.org/10.1016/J.FORECO.2013.04.002

Simpson MC, Wilken PM, Coetzee MPA, Wingfield MJ, Wingfield BD (2013) Analysis of microsatellite markers in the genome of the plant pathogen Ceratocystis fimbriata. Fungal Biol 117:545–555. https://doi.org/10.1016/j.funbio.2013.06.004

Stewart JE, Ross-Davis AL, Graça RN, Alfenas AC, Peever TL, Hanna JW, Uchida JY, Hauff RD, Kadooka CY, Kim M-S, Cannon PG, Namba S, Simeto S, Pérez CA, Rayamajhi MB, Lodge DJ, Arguedas M, Medel-Ortiz R, López-Ramirez MA, Tennant P, Glen M, Machado PS, McTaggart AR, Carnegie AJ, Klopfenstein NB (2018) Genetic diversity of the myrtle rust pathogen (Austropuccinia psidii) in the Americas and Hawaii: Global implications for invasive threat assessments. For Pathol 48:e12378. https://doi.org/10.1111/efp.12378

Tobias PA, Guest DI, Külheim C, Park RF (2018) De Novo Transcriptome Study Identifies Candidate Genes Involved in Resistance to *Austropuccinia psidii* (Myrtle Rust) in *Syzygium luehmannii* (Riberry). Phytopathology 108:627–640. https://doi.org/10.1094/PHYTO-09-17-0298-R

Uchida J, Zhong S, Killgore E (2006) First Report of a Rust Disease on Ohia Caused by Puccinia psidii in Hawaii . Plant Dis 90:524–524. https://doi.org/10.1094/pd-90-0524c

Valdetaro DCOF, Harrington TC, Oliveira LSS, Guimarães LMS, McNew DL, Pimenta LVA, Gonçalves RC, Schurt DA, Alfenas AC (2019) A host specialized form of Ceratocystis fimbriata causes seed and seedling blight on native Carapa guianensis (andiroba) in Amazonian rainforests. Fungal Biol 123:170–182. https://doi.org/10.1016/j.funbio.2018.12.001

Valdetaro DCOF, Oliveira LSS, Guimarães LMS, Harrington TC, Ferreira MA, Freitas RG, Alfenas AC (2015) Genetic variation, morphology and pathogenicity of Ceratocystis fimbriata on Hevea

brasiliensis in Brazil. Trop Plant Pathol 40:184–192. https://doi.org/10.1007/s40858-015-0036-6

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Winter G (1884) Repertorium. Rabenhorstii fungi europaei et extraeuraopaei. Cent. XXXI et XXXII. Hedwigia 23:164–172.

Ye X, Liu H, Jin Y, Guo M, Huang A, Chen Q, Guo W, Zhang F, Feng L (2017) Transcriptomic Analysis of Calonectria pseudoreteaudii during Various Stages of Eucalyptus Infection. PLoS One 12:e0169598. https://doi.org/10.1371/journal.pone.0169598

Zauza EA V., Alfenas AC, Harrington TC, Mizubuti ES, Silvai JF (2004) Resistance of *Eucalyptus* Clones to *Ceratocystis fimbriata*. Plant Dis 88:758–760. https://doi.org/10.1094/PDIS.2004.88.7.758

Zauza EA V, Couto MMF, Lana VM, Maffia LA, Alfenas AC (2010) Vertical spread of puccinia psidii urediniospores and development of eucalyptus rust at different heights. Australas Plant Pathol 39:141–145. https://doi.org/10.1071/AP09073

Zhong S, Yang B, Puri KD (2011) Characterization of Puccinia psidii isolates in Hawaii using microsatellite DNA markers. J Gen Plant Pathol 77:178–181. https://doi.org/10.1007/s10327-011-0303-4

Zhu Q-H, Stephen S, Kazan K, Jin G, Fan L, Taylor J, Dennis ES, Helliwell CA, Wang M-B (2013) Characterization of the defense transcriptome responsive to Fusarium oxysporum-infection in Arabidopsis using RNA-seq. Gene 512:259–266. https://doi.org/10.1016/J.GENE.2012.10.036

Zhuang JY, Wei SX (2011) Additional materials for the rust flora of Hainan Province, China. Mycosystema 30:853–860.

CHAPTER 1 - GENOME ASSEMBLY AND ANNOTATION OF *Austropuccinia psidii* EPITYPE GUAVA (*Psidium guajava*) ISOLATE, THE CAUSAL AGENT OF MYRTLE RUST

Samuel A. Santos[1], Pedro M. P. Vidigal[2], Patricia S. Machado[1], Reginaldo G. Mafia[3], Morag Glen[4], Matthew D. Templeton[5] and Acelino C. Alfenas[1*]

[1] Laboratory of Forest Pathology, Department of Plant Pathology, Universidade Federal de Viçosa, Minas Gerais State, Brazil.

[2] Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Minas Gerais State, Brazil.

[3] Centro de Tecnologia, Suzano S.A., Aracruz, Espírito Santo State, Brazil.

[4] Tasmanian Institute of Agriculture, University of Tasmania, Hobart, TAS 7001, Australia.

[5] The New Zealand Institute for Plant and Food Research Limited, Auckland 1142, New Zealand

[*] **Correspondence author.** Acelino C. Alfenas, Laboratory of Forest Pathology, Department of Plant Pathology, Universidade Federal de Viçosa, Minas Gerais State, Brazil; Tel: +55 (31) 98707-2940; E-mail: aalfenas@ufv.br

Abstract Pucciniales fungi are well known to have the largest and complex genomes among Fungi. *Austropuccinia psidii* has a wide and expanding host range and has been a constant threat to many Myrtaceae native species across the world. However, very little is known regarding its genome structure and gene content. In this study, we present a high-quality assembly and annotation of the haploid nuclear genome and *de novo* transcriptome of *A. psidii* epitype guava (*Psidium guajava*) isolate. The genome assembly contains 787 scaffolds with a combined length of 672 Mbp. In addition, the largest and smallest scaffolds have 6,495,440 and 1,144 bp, respectively, with an N50 scaffold length of 1,373,277 bp. About 3.6% of the genome was predicted as a coding sequence comprised of 20,184 protein-coding genes, 29 ribosomal RNAs, and 481 transfer RNAs. Additionally, a high proportion (81.37%) consists of repetitive sequences. Of these, the majority (60.39%) were classified as transposable elements (TEs), which represent about two-fold compared to TEs content observed for other Puccinilaes fungi

such as *Puccinia striiformis* f. sp. *tritici*, *P. sorghi*, *P. triticina*, *Melampsora lini*, and *Hemileia vastatrix*. Furthermore, the assembled transcriptome from *A. psidii* urediniospores contains 46,742 contigs totalizing 71.39 Mbp and showed a high BUSCO completeness (92.2%) with the Basidiomycota. Finally, our genome and transcriptome dataset provide a framework for future studies leading to a better understanding of the biology, plasticity, evolution, and pathogenicity of *A. psidii*.

Keywords: *Puccinia psidii*. Biotrophic pathogen. Rust. RNA-Seq. Whole genome sequencing. Transposable elements.

INTRODUCTION

Myrtle rust caused by *Austropuccinia psidii* (Winter) Beenken is a very important disease around the world, mainly because the fungus is able to infect a wide and expanding host range within the Myrtaceae. The pathogen was first reported as *Puccinia psidii* in 1884 infecting guava plants (*Psidium guajava*) in São Francisco do Sul, Santa Catarina State, Brazil (Winter 1884). Since then it has spread from South to North America (MacLachlan 1938; Marlatt and Kimbrough 1979; Coutinho et al. 1998; Uchida et al. 2006). The fungus has been considered an invasive pathogen and migrated to other continents, including Asia (Kawanishi et al. 2009; Zhuang and Wei 2011), Africa (Roux et al. 2013), and Oceania (Carnegie et al. 2010; Beresford et al. 2018). Recently, *Puccinia psidii* has been reclassified to *Austropuccinia psidii* and placed in the newly circumscribed family Sphaerophragmiaceae (Beenken 2017). The symptoms of the disease depend on the susceptibility of the host, however, may range from small pustules with few spores to massive production of powdery bright yellow urediniospores with shoot dieback that may ultimately result in plant death (Alfenas et al. 2009).

Molecular studies of *A. psidii* started a few years ago with population genetic studies of the pathogen in Hawaii (Zhong et al. 2011) and Brazil (Graça et al. 2013). These studies indicated that *A. psidii* has reproduced clonally in Hawaii and that populations from different hosts in Brazil are derived from different lineages. High throughput sequencing can provide high-quality whole genomic sequences much more rapidly and cheaply than previous sequencing technologies (Goodwin et al. 2016). Whole-genome sequencing has already been applied in plant pathology and in plant-microbe interaction research to infer the evolution of pathogens and the factors of pathogenicity and virulence genes (Cantu et al. 2011), comparative analyses between different races or species (Duplessis et al. 2011) and identification of effector

proteins (Joly et al. 2010). This technology has been used in studies of several rust fungi species including *Puccinia striiformis* f. sp. *tritici* (Cantu et al. 2011), *P. graminis* f. sp. *tritici* (Duplessis et al. 2011), *Melampsora larici populina* (Duplessis et al. 2011), *Melampsora lini* (Nemri et al. 2014), and *Puccinia sorghi* (Rochi et al. 2018).

Rust fungi (Basidiomycota, Pucciniales) are obligate plant pathogens that have the largest and complex genomes among Fungi with a high percent of repetitive sequences, which may difficult the sequencing and assembly (Duplessis et al. 2011; Cantu et al. 2011; Nemri et al. 2014; Tavares et al. 2014; Rochi et al. 2018). The first brief overview of the genome of *A. psidii* for one isolate present in Australia estimated its genome size to be between 103-145 Mb (Tan et al. 2014). Pucciniales species are well known to have the largest fungal genomes with an average genome size estimated at about 305.5 Mb by flow cytometry (Tavares et al. 2014). Recently, a draft of the genome of an *A. psidii* isolate from South Africa estimated its genome in 1.2 Gb (McTaggart et al. 2018). However, the genome structure and gene content of *A. psidii* remains not well characterized.

A complete characterization of the genome of myrtle rust fungus could provide useful information to devise innovative ways to protect crops against them. Therefore, in this study, we present a high-quality assembly and annotation for the nuclear genome and transcriptome of *A. psidii* eptype guava (*Psidium guajava*) isolate.

MATERIAL AND METHODS

Sampling and multiplication of spores

Urediniospores of *A. psidii* VIC42496 isolate were collected from infected leaves of guava (*Psidium guajava*) in Araquari, Santa Catarina, Brazil (Machado et al. 2015). Representative specimens for VIC42496 epitype guava isolate were deposited in the herbarium at the Universidade Federal de Viçosa, Brazil (herbarium code: VIC) and Queensland Plant Pathology Herbarium, Australia (herbarium code: BRIP) (Machado et al. 2015). Collected urediniospores were multiplied in plants of *Sygyzium jambos*. For that, seedlings were cultivated in 2 L pots containing MecPlant® substrate enriched with 26 g of super simple phosphate and 12 g of Osmocote® (NPK, 19-6-10). In each plant, 4-5 healthy young branches were marked with thin strings and a soft-bristle brush was used to inoculate the urediniospores on leaves. After inoculation, plants were incubated for 24 h in a mist irrigation chamber at 25±2°C in the dark, and then they were transferred to a growth chamber at 22 ±2°C with a 12 h light cycle (80 µmol

m^{-2} s^{-1}) (Ruiz et al. 1989). Approximately 12 days after inoculation, urediniospores were collected using an electric spore collector device (Holliday et al. 2013) and stored in an ultra-low temperature (ULT) freezer at -80 °C prior to DNA extraction.

DNA extraction and genome sequencing

For Illumina short reads sequencing, DNA samples were extracted using a Qiagen DNeasy Plant minikit (Qiagen Sciences Inc., Germantown, MD, USA). Two tungsten carbide beads were added to each 1.5 mL microcentrifuge tube containing approximately 100 µg of spores. Spores were disrupted in a Qiagen Tissuelyser for 1 min at 30 Hz frequency and the DNA was extracted according to the kit instructions, except that the 65°C incubation step was extended from 10 min to 1 h. DNA was eluted in TE buffer. Then, paired-end libraries with an average insert size of 500 bp were constructed from each sample of genomic DNA according to the manufacturer's instructions and sequenced using an Illumina HiSeq 2500 platform (Illumina, San Diego, CA) at the Australian Genome Research Facility, Melbourne, Australia (http://www.agrf.org.au/). Higher sequencing coverage was requested. Additionally, to obtain a high-quality assembly, the genome of VIC42496 isolate was also sequenced using Oxford Nanopore Technology (ONT).

Since we found a lot of bacterial contamination on the Illumina sequencing dataset, for ONT long reads sequencing, urediniospores were recovered from -80°C ULT freezer, filtered through a Millipore Nitrocellulose filter membrane with 8 µm pores (Merck Millipore, Burlington, Massachusetts, USA) and washed in sterilized ultrapure water containing Rifampicin antibiotic. This procedure was repeated 3 times. Thereafter, DNA samples were re-extracted following the instructions of a specific protocol for high-quality DNA from Fungi for long read sequencing (dx.doi.org/10.17504/protocols.io.k6qczdw), except that we used only 250 mg of urediospores. Thus, cleaned spores were frozen in liquid Nitrogen and disrupted into a pair of Stainless Steel Grinding Jar 10 mL (Qiagen Sciences Inc., Germantown, MD, USA) using Tissuelyser (Qiagen Sciences Inc., Germantown, MD, USA) for three bursts of 15 seconds at 25 Hz frequency. After the first 15 seconds, the closed grinding jars were immersed into liquid nitrogen for 10 min up to the next 15 seconds disruption cycle, so that the samples remained frozen until complete disruption at the end of the third cycle. A total of 60 µg of DNA in high-quality was obtained. To library preparation, 60 µg of DNA were size-selected (>20 kbp) with a Blue Pippin (Sage Science, Beverly, MA, USA), and processed using the Ligation Sequencing 1D SQK-LSK109 kit (Oxford, Nanopore, Oxford, UK) according to the

manufacturer's instructions. Prepared libraries were sequenced on four different flowcells type R9.4 FLOMIN-106 (Oxford, Nanopore, Oxford, UK) for 72 hours at DNA Link USA, Inc., San Diego, California, USA (http://www.dnalink.com/).

RNA extraction and RNA-Seq

Urediniospores were recovered from -80°C ULT freezer, filtered through a Millipore Nitrocellulose filter membrane with 8 µm pores (Merck Millipore, Burlington, Massachusetts, USA) and washed in sterilized ultrapure water containing Rifampicin antibiotic. This procedure was repeated 3 times. Thereafter, cleaned spores were put in liquid Nitrogen and disrupted using a pair of Stainless Steel Grinding Jar 10 mL (Qiagen Sciences Inc., Germantown, MD, USA) with Tissuelyser (Qiagen Sciences Inc., Germantown, MD, USA) for three bursts of 15 seconds at 25 Hz frequency. After the first 15 seconds, the closed grinding jars were immersed into liquid nitrogen for 10 min up to the next 15 seconds disruption cycle, so that the samples remained frozen until complete disruption at the end of the third cycle. Total RNA was extracted using PureLink® Plant RNA Reagent (ThermoFisher Scientific, Waltham, MA, USA), according to manufacturer's instruction. RNA was suspended in RNase-free water. RNA integrity and quality were assessed using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA).

A total of three RNA libraries (biological replicates) were prepared for RNA-Seq. First, rRNA in total RNA was depleted by Ribo-Zero kit (Illumina, San Diego, CA, USA). The enriched mRNA samples were subjected to Illumina cDNA library construction using TruSeq Stranded mRNA Prep kit (Illumina Inc., San Diego, CA, USA). RNA was purified, fragmented, and primed for cDNA synthesis. The RNA fragments were transcribed into the first-strand cDNA using reverse transcriptase and random hexamers, followed by second-strand cDNA synthesis. These fragments were prepared for sequencing with an end-repair process and the addition of a single 'A' base at the 3' end. Paired-end adapters were ligated to the ends of these 3' adenylated cDNA fragments. Products were purified and enriched through PCR to create the final cDNA library (Ross-Davis et al. 2013). Finally, the three cDNA libraries were sequenced using the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) DNA Link USA, Inc., San Diego, California, USA (http://www.dnalink.com/). A total of $5x10^7$ of 100 bp paired-end (PE100) reads per library were requested.

Raw data quality control

Low-quality Illumina short reads (Q30 threshold) and Illumina adapter sequences were removed using Trimmomatic v. 0.36 (Bolger et al. 2014). Additionally, for RNA-Seq reads, SortMeRNA v. 2.1 (Kopylova et al. 2012) was used to filter raw data and remove reads from rRNA. Quality control checks were performed by FastQC v. 0.11.5 (Andrews 2010) before and after each trimming/filtering step to ensure the quality of the final dataset (clean reads). ONT long reads were filtering by size (≥ 1,000 bp) using SeqKit v. 0.7.0 (Shen et al. 2016).

Genome size estimate, assembly, annotation, and evaluation

Illumina trimmed reads from genomic DNA were also used to estimate the genome size based on k-mer distribution and depth (Tan et al., 2018). First, k-mer counting was performed with Jellyfish v2.2.10 (Marçais and Kingsford 2011). Then, histograms of k-mer frequency distributions of 17-, 25-, and 35-mers were processed by GenomeScope (Vurture et al. 2017).

To get the best genome assembly, filtered ONT long reads were *de novo* assembled by Flye v. 2.5 using default parameters (Kolmogorov et al. 2019). Due to the high heterozygosity (4.2%), to identify and select the haploid genome, the assembled contigs were reduced using Redundans v. 0.14a (Pryszcz and Gabaldón 2016) with default settings and selecting "--noscaffolding" and "--nogapclosing" parameters. Then, haploid contigs were subjected to SSPACE-LongRead v. 1.1 using default parameters (Boetzer and Pirovano 2014). GapFinisher v. 3 was used to gap closing (Kammonen et al. 2019). The assembled scaffolds were polished with 10 interactions of Pilon v.1.22 using trimmed Illumina short reads to correct bases and fix misassemblies (Walker et al. 2014). Additionally, we performed a BLASTn (Altschul et al. 1990) of all assembled scaffolds against a bacteria-virus database to identify and eliminate potential contaminates from the final assembly.

Protein-coding genes were predicted using BRAKER v. 2.1.0 (Stanke et al. 2006; Stanke et al. 2008; Hoff et al. 2016; Hoff et al. 2019). For that, filtered and trimmed RNA-Seq reads from fungal urediniospores of *A. psidii* were mapped to the final genome assembly using STAR v. 2.7.1a by selecting default parameters (Dobin et al. 2013). The obtained BAM files and scaffold sequences of *A. psidii* genome were submitted to the BRAKER pipeline using default parameters. To evaluate gene prediction accuracy, its completeness was assessed using the Benchmarking Universal Single-Copy Orthologs, named BUSCO (Simão et al. 2015). BUSCO

was performed on the amino acid sequences using the Basidiomycota lineage dataset (basidiomycota_odb9).

The predicted coding-protein genes were functionally annotated through similarity searches using BLAST version 2.6.0 (Altschul et al. 1990). In these searches, the encoded proteins were aligned to the sequences of NCBI protein non-redundant database (NCBInr), UniProt Knowledgebase (UniProtKb), using protein BLAST (BLASTp) (cutoff for significant hits: E-value of 1e-10). Furthermore, Gene Ontology (GO) terms were assigned to the proteins using GOanna tool from AgBase (https://agbase.arizona.edu/cgi-bin/tools/GOanna.cgi) (McCarthy et al. 2006) by selecting the following parameters: E-value = 1e-05; matrix BLOSUM62; word size = 3; identity = 30%; and query coverage = 70%. Ribosomal genes were identified by Barrnap v. 0.9 (https://github.com/tseemann/barrnap), using Eukaryota database, and transporter RNA genes were identified by tRNAscan-SE v. 2.0.5 (http://lowelab.ucsc.edu/tRNAscan-SE/) using Eukaryotic database and default options (Chan and Lowe 2019). RepeatMasker v. 4.0.8 (http://www.repeatmasker.org/) was used to identify repetitive sequences using Fungi database of RepBase library v. 20170127 (Jurka et al. 2005). Additionally, RepeatModeler v. 2.0 (http://www.repeatmasker.org/RepeatModeler/) was used to perform a *de novo* transposable element (TE) family identification.

Transcriptome assembly, annotation, and evaluation

Filtered and trimmed RNA-Seq reads from *A. psidii* urediniospores were *de novo* assembled using Trinity version 2.8.6 with default parameters (Grabherr et al. 2011). The assembled transcripts were submitted to BUSCO completeness using the Basidiomycota database (Simão et al. 2015). In addition, the transcript sequences were also functionally annotated through similarity searches by BLAST version 2.6.0 (Altschul et al. 1990) according to UniProtKb and using BLASTx (cutoff for significant hits: E-value of 1e-10).

RESULTS

Assembly stats of the nuclear genome of *A. psidii*

We designed a comprehensive pipeline that allowed us to obtain a high-quality genome and transcriptome assembly through a massive dataset obtained in this study (Fig. 1). The estimated size of the haploid genome of *A. psidii* ranged from 517 to 752 with an average of 634 Mbp

(Fig. 2). In addition, the heterozygosity ranged from 4.06 to 4.26% (Fig. 2). Since DNA was extracted from urediniospores ($n + n$) and the genome showed high heterozygosity, the original assembly contains scaffolds from both haplotypes with a length of 1.56 Gbp (Table 1), which is approximately two-fold the estimated size of the haploid genome.

The final assembly for the haploid genome contains 787 scaffolds, of these 267 (33.9%) were bigger than 1 Mbp (Table 1). The total genome length was 672 Mbp (Table 1). In addition, the largest and shortest scaffolds had 6,495,440 and 1,144 bp, respectively, with an average size of 853,952 bp (Table 1). Furthermore, the genome assembly had an N50 and L50 of 1,373,277 bp and 160, respectively (Table 1). Finally, the GC and N contents in the whole assembled genome were 32.4 and 1.07%, respectively (Table 1).

Annotation of the haploid nuclear genome of *A. psidii*

Genes were predicted and annotated as protein-coding, transfer RNA (tRNA), and ribosomal RNA (rRNA) – four types of rRNA (5S, 5.8S, 18S, and 28S) were found (Table 2). A total of 20,694 genes were predicated, including 20,184 protein-coding, 29 rRNA, and 480 tRNA (Table 2). BUSCO assessment analysis of the predicted protein-coding genes revealed 814 (61.0%) of completeness with 1,335 orthologous genes from Basidiomycota database (Table 2). Of these, 741 (55.5%) genes were single copy, while 73 (5.5%) were duplicated. Additionally, 144 (10.8%) genes were fragmented and 377 (28.2%) were missed (Table 2).

Repetitive sequences were identified and classified as transposable elements (TEs) [Long Interspersed Nuclear Elements (LINE), Long Terminal Repeat Elements (LTR), and DNA transposons], satellites, simple repeats, low complexity repeats, and unclassified (Table 3). A total of 546 Mbp (81.37%) of the haploid nuclear genome consists of repetitive sequences (Table 3). Among the identified and classified elements, the majority was TEs types LTR (48.88%) and DNA transposons (11.16%), followed by simple repeats (0.78%), TEs type LINE (0.35%), satellites (0.29%), and low complexity (0.19%) (Table 3). Furthermore, 19.72% of the repetitive sequences have no classification assigned (Table 3).

Transcriptome assembly of *A. psidii*

The assembled transcriptome from three RNA-Seq libraries of *A. psidii* urediniospores contains 46,742 contigs with a combined length of 71.39 Mbp (Table 4). Most of the contigs (54.9%) were bigger than 1,000 bp. In addition, the largest and shortest contigs had 10,032 and 201 bp,

respectively, with an average size of 1,527 bp (Table 4). The transcriptome assembly had an N50 and L50 of 2,441 bp and 10269, respectively. Furthermore, the GC content in the whole trancriptome was about 39.81% (Table 4).

Austropuccinia psidii transcriptome assembly showed 1,231 (92.2%) of completeness with 1,335 BUSCO groups from the Basidiomycota database (Table 4). Of these, 348 (26.1%) genes were a single copy, while 883 (66.1%) were duplicated. In addition, 65 (4.9%) genes were fragmented and only 39 (2.9%) were missed (Table 4).

DISCUSSION

The first brief overview of *A. psidii* genome for one isolate present in Australia estimated its genome size in 103-145 Mbp (Tan et al. 2014). Nevertheless, Pucciniales species are well known to have the largest fungal genomes with an average genome size estimated at about 305.5 Mb by flow cytometry (Tavares et al. 2014). Recently, a draft assembly for another isolate of *A. psidii* from South Africa estimated its genome in 1.2 Gbp (McTaggart et al. 2018), however, since DNA samples were also extracted from urediniospores ($n + n$) and no reduction step was made during assembly pipeline, is very likely that this draft assembly still contains scaffolds from both haplotypes. In this study, we generated a massive sequencing dataset for the whole-genome and transcriptome of *A. psidii* using Illumina short reads and ONT long reads technologies. In addition, we designed a comprehensive pipeline data analysis, which allowed us to present a high-quality assembly and annotation of the haploid nuclear genome and *de novo* transcriptome of *Austropuccinia psidii* epitype guava (*Psidium guajava*) isolate (Machado et al. 2015). After reduction, the final assembly of the haploid nuclear genome has a total size of about 672 Mbp, which is very similar to the estimated genome size based on kmer depth and distribution (634 Mbp). Furthermore, compared to both previous draft assemblies, our assembly showed a high quality regarding the number of scaffolds, the largest scaffold length, and N50.

A total of 20,694 genes were predicted for the haploid genome of *A. pisidii* VIC42496, which is similar to *Puccinia* species including *P. graminis* f. sp. *tritici* (Duplessis et al. 2011), *P. striiformis* f. sp. *tritici* (Cantu et al. 2011), *P. sorghi* (Rochi et al. 2018), and *P. triticina* (Kiran et al. 2016), but different from other Puccinilles fungi such as *Melampsora larici-populina* (Duplessis et al. 2011), *M. lini* (Nemri et al. 2014), and *Hemileia vastatrix* (Porto et al. 2019). Although BUSCO completeness analysis for the predicted protein-coding genes showed a high percentage of missing genes (28.20%) with the Basidiomycota database, the *de novo* transcriptome assembly for *A. psidii* VIC42496 urediniospores had high completeness

(92.2%) and very low missing (2.9%). This indicates that our transcriptome assembly comprises most of the expected gene set for *A. psidii*. Whole-genome sequencing, assembly, and annotation of rust biotrophic fungal pathogens have been considered a challenge. First, since the fungus does not grow in pure culture (*in vitro* conditions), it is hard to obtain a DNA sample in great quality and quantity for sequencing using different technologies. Second, due to the large genome size, the cost of sequencing is considerably higher (Goodwin et al. 2016). Finally, the high number of repeat-rich regions among its genome becomes the genome assembly as a difficult task (Duplessis et al. 2011; Cantu et al. 2011; Nemri et al. 2014; Tavares et al. 2014; Rochi et al. 2018; Porto et al. 2019).

The assembly of the haploid nuclear genome of *A. psidii* VIC42496 presented here has the largest genome comparing to other Puccinilaes species such as *P. graminis* f. sp. *tritici* (Duplessis et al. 2011), *P. striiformis* f. sp. *tritici* (Cantu et al. 2011), *P. sorghi* (Rochi et al. 2018), *P. triticina* (Kiran et al. 2016), *M. larici-populina* (Duplessis et al. 2011), *M. lini* (Nemri et al. 2014), and *Hemileia vastatrix* (Porto et al. 2019). In fungi, genome size is strongly related to TEs content than the number of genes (Elliott and Gregory 2015). Biotrophic fungal pathogens, especially rust fungi, have been well known by their complex genomes including a high percentage of TEs (Duplessis et al. 2011; Cantu et al. 2011; Nemri et al. 2014; Tavares et al. 2014; Rochi et al. 2018; Porto et al. 2019). However, in the *A. pisidii* VIC42496 genome, we identified that a high proportion (81.37%) consists of repetitive sequences. Of these, about 60.39% was classified as TEs, which represents almost two-fold compared to TEs content observed for other Puccinilaes fungi such as *P. striiformis* f. sp. *tritici* (Cantu et al. 2011), *P. sorghi* (Rochi et al. 2018), *P. triticina* (Kiran et al. 2016), *M. lini* (Nemri et al. 2014), and *Hemileia vastatrix* (Porto et al. 2019).

Furthermore, among these Pucciniales fungal pathogens, *A. psidii* has the widest and expanding host range threating thousands of host species within the Myrtaceae across the world (Pegg et al. 2014; Carnegie et al. 2016; Pegg et al. 2017; Stewart et al. 2018; Berthon et al. 2018; Beresford et al. 2018). The role of the TEs in biology, evolution, and pathogenicity of fungi has been extensively studied over the years. TEs have been especially associated with genome plasticity, evolution, gene content, gene expression regulation, and pathogenicity (Castanera et al. 2016; Krishnan et al. 2018; Razali et al. 2019). Indeed, pathogenic fungi often have more TEs inserted into genes than saprophytic (Muszewska et al. 2019).

CONCLUSIONS

Here, we present a high-quality assembly and annotation of the *de novo* haploid nuclear genome and transcriptome of *A. psidii* epitype guava isolate, which is so far the largest described genome among Pucciniales rust fungi. Our genome and transcriptome dataset provide a framework for future studies leading to a better understanding of the biology, plasticity, evolution, and pathogenicity of *A. psidii*.

Acknowledgments The authors thank to 'Conselho Nacional de Desenvolvimento Cientifico e Tecnológico' (CNPq) and 'Coordenação de Aperfeicoamento de Pessoal de Nivel Superior' (CAPES) for the fellowship granted to first the author. To Clonar Resistência a Doenças Florestais for providing the plant material and the inoculation facilities. To Suzano S.A. for supporting this study. We also thanks to Plant and Food Research for providing all bioinformatics facilities and Núcleo de Análise de Biomoléculas (NuBioMol) for supporting data analysis. NuBioMol is supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), CAPES, CNPq, Financiadora de Estudos e Projetos (Finep) and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).

Table 1 Summary stats of the assembled nuclear genome of *Austropuccinia psidii* epitype guava isolate

	Original assembly[a]	Haploid (reduced)[b]
Total number of scaffolds	3 472	787
scaffolds (≥ 10000 bp)	2 538 (73.1%)	752 (95.6%)
scaffolds (≥ 100000 bp)	1 904 (54.8%)	677 (86.0%)
scaffolds (≥ 1 Mbp)	549 (15.8%)	267 (33.9%)
Total genome length (Mbp)	1 559	672
Shortest scaffold (bp)	1 000	1 144
Largest scaffold (bp)	5 337 186	6 495 440
Mean scaffold size (bp)	449 087	853 952
N50[c] (bp)	1 261 020	1 373 277
L50[d]	404	160
GC content (%)	33.65	32.4
Total N (bp; %)	39 916 314 (2.56)	7 109 501 (1.07)
N's per 100 kbp	2 559	1 058

[a] Flye assembly containing scaffolds from both haplotypes, since DNA was extracted from urediniospores (*n* + *n*).

[b] Reduced assembly containing only haploid scaffolds.

[c] N50 is defined as the sequence length of the shortest scaffold at 50% of the total assembly length.

[d] L50 is defined as the number of scaffolds whose summed length equals 50% of the total assembly length.

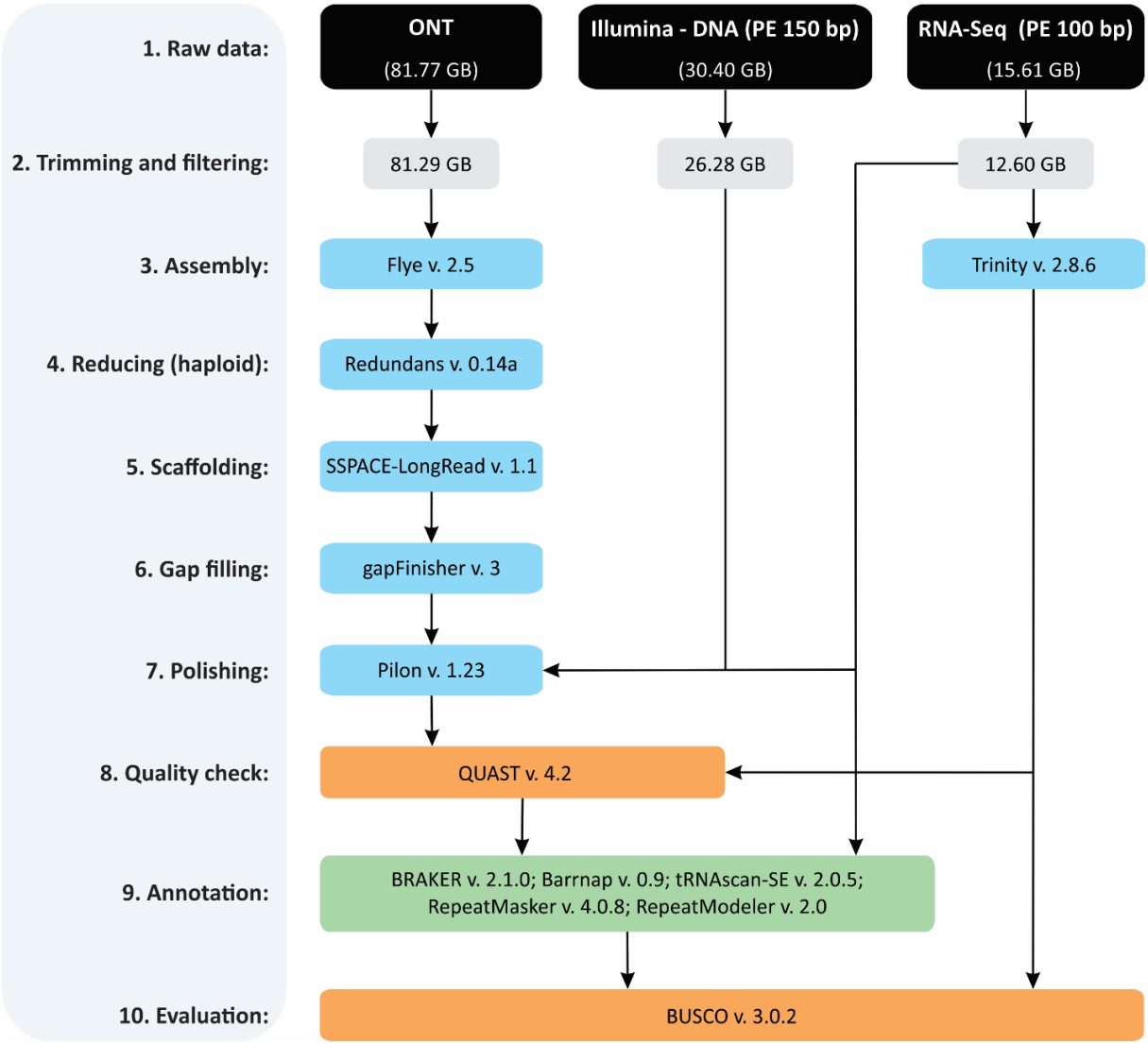

Fig. 1 Overview of the workflow for assembly and annotation of *Austropuccinia psidii* genome and transcriptome. First, DNA and RNA samples extracted from fungal urediniospores were sequenced using Illumina HiSeq and Oxford Nanopore Technology (ONT), as well as RNA-Seq (Illumina NovaSeq 6000). After filtering and trimming ONT long reads were assembled by Flye (Kolmogorov et al. 2019) and Trinity (Grabherr at al. 2011), respectively. Redundans was applied to filtering haploid genome (Pryszcz and Gabaldon 2016). Then, Scaffolding and gap close steps were performed using SSPACE-LongRead (Boetzer and Pirovano 2014) and gapFinisher (Kammonen et al. 2019), respectively. Trimmed Illumina short reads were employed to polishing scaffolds using Pilon (Walker et al. 2014). QUAST was used to check assemblies stats (Gurevich et al. 2013). Protein-coding genes were predicted by BRAKER using BAM files alignment from fungal spores RNA-Seq libraries (Hoff et al. 2019). rRNA and tRNA genes were predicted by Barrnap and tRNAscan-SE (Chan and Lowe 2019), respectively. RepeatMasker v. 4.0.8 and RepeatModeler v. 2.0 were used for repetitive sequences prediction (Jurka et al. 2005). Finally, the accuracy of protein-coding genes prediction and transcriptome assembly was accessed by BUSCO completeness analysis (Simão et al. 2015).

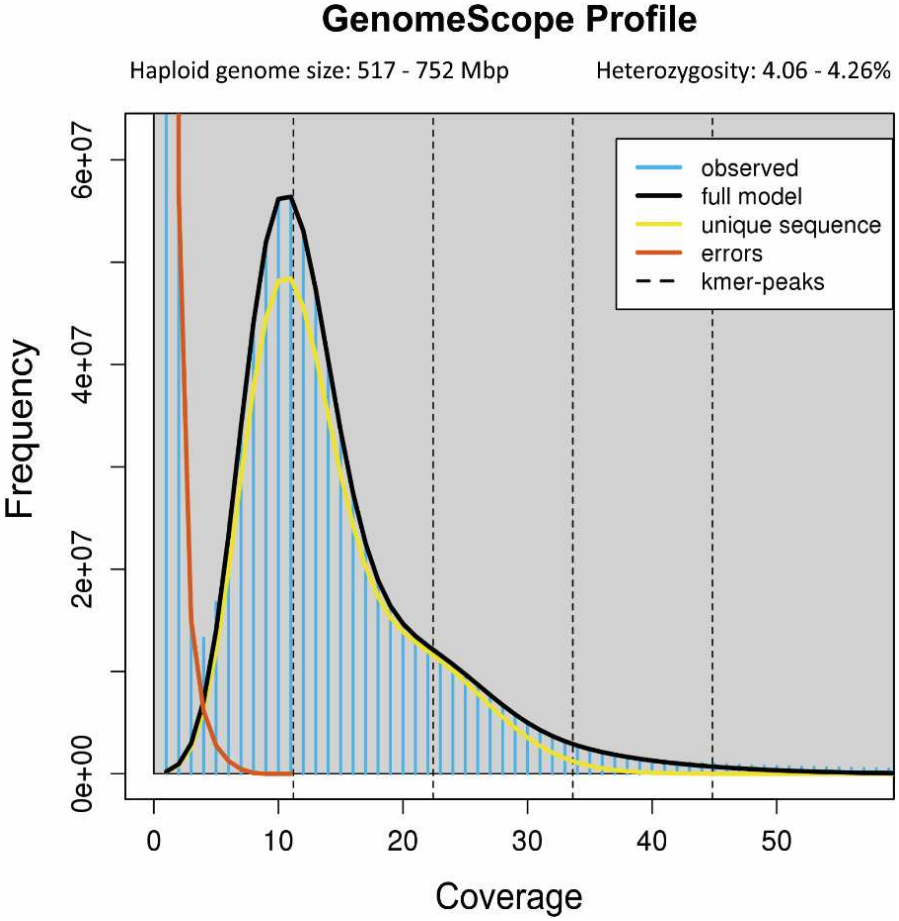

Fig. 2 Haploid genome size estimation based on the k-mers frequency distributions and depth. k-mer counting was performed by Jellyfish v2.2.10 (Marcais et al. 2011) and the generated histograms of 17-, 25-, and 35-mers were processed by GenomeScope (Vurture et al. 2017).

Table 2 Gene prediction for the haploid nuclear genome of *Austropuccinia psidii* epitype guava isolate

Protein-coding genes	20 184
CDS[a] (Mbp)	24.10
Ribossomal RNA (rRNA) genes	
5S	10
5.8S	6
18S	5
28S	8
Tranfer RNA (tRNA) genes	481
Total number of genes	20 694
GC content in CDS (%)	42.67
BUSCO[b] assessment (%)	
Complete	61
Complete and syngle-copy	55.50
Complete and duplicate	5.50
Fragmented	10.80
Missing	28.20

[a] CDS: coding DNA sequence.

[b] Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment using 1,335 conserved genes of Basidiomycota database (basidiomycota_odb9).

Table 3 Overview of predicted repetitive sequences for the haploid nuclear genome of *Austropuccinia psidii* epitype guava isolate

Type[a]	Number	Length (bp)	% of sequence
TEs[b]			
Retroelements			
LINE	1 643	2 348 751	0.35
LTR	154 720	328 530 318	48.88
DNA transposons	66 078	75 002 999	11.16
Satellites	4 459	1 963 612	0.29
Simple repeats	61 374	5 247 221	0.78
Low complexity	16 285	1 259 855	0.19
Unclassified	110 960	132 503 060	19.72
Total	415 519	546 855 816	81.37

[a] Repetitive sequences were predicted by RepeatMasker v. 4.0.8 and RepeatModeler v. 2.0 (Jurka et al. 2005).
[b] Transposable elements (TEs): Retroelements [Long Interspersed Nuclear Elements (LINE), Long Terminal Repeat Elements (LTR)] and DNA transposons.

Table 4 Summary stats of the assembled transcriptome of *Austropuccinia psidii* epitype guava isolate

Total number of contigs	46 742
contigs (≥ 1000 bp)	25 669 (54.9%)
contigs (≥ 10000 bp)	2
Total length (Mbp)	71.39
Largest contig (bp)	10 032
Shortest contig (bp)	201
Mean contig size (bp)	1 527
N50[a] (bp)	2 441
L50[b]	10 269
GC content (%)	39.81
BUSCO[c] assessment (%)	
Complete	92.2
Complete and syngle-copy	26.1
Complete and duplicate	66.1
Fragmented	4.9
Missing	2.9

[a] N50 is defined as the sequence length of the shortest contig at 50% of the total assembly length.

[b] L50 is defined as the number of contigs whose summed length equals 50% of the total assembly length.

[c] Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment using 1,335 conserved genes of the Basidiomycota database (basidiomycota_odb9).

REFERENCES

Alfenas AC, Zauza EAV, Mafia RG, Assis TF (2009) Clonagem e doenças do eucalipo, 1st edn. Editora UFV, Viçosa, MG state, Brazil.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 18 Jul 2019.

BEENKEN L (2017) Austropuccinia: a new genus name for the myrtle rust Puccinia psidii placed within the redefined family Sphaerophragmiaceae (Pucciniales). Phytotaxa 297:53–61. https://doi.org/10.11646/phytotaxa.297.1.5

Beresford RM, Turner R, Tait A, Paul V, Macara G, Yu ZD, Lima L, Martin R (2018) Predicting the climatic risk of myrtle rust during its first year in New Zealand. New Zeal Plant Prot 71:332–347. https://doi.org/10.30843/nzpp.2018.71.176

Berthon K, Esperon-Rodriguez M, Beaumont LJ, Carnegie AJ, Leishman MR (2018) Assessment and prioritisation of plant species at risk from myrtle rust (Austropuccinia psidii) under current and future climates in Australia. Biol Conserv 218:154–162.

https://doi.org/10.1016/j.biocon.2017.11.035

Boetzer M, Pirovano W (2014) SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 15:211. https://doi.org/10.1186/1471-2105-15-211

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Cantu D, Govindarajulu M, Kozik A, Wang M, Chen X, Kojima KK, Jurka J, Michelmore RW, Dubcovsky J (2011) Next Generation Sequencing Provides Rapid Access to the Genome of Puccinia striiformis f. sp. tritici, the Causal Agent of Wheat Stripe Rust. PLoS One 6:e24230. https://doi.org/10.1371/journal.pone.0024230

Carnegie AJ, Kathuria A, Pegg GS, Entwistle P, Nagel M, Giblin FR (2016) Impact of the invasive rust Puccinia psidii (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. Biol Invasions 18:127–144. https://doi.org/10.1007/s10530-015-0996-y

Carnegie AJ, Lidbetter JR, Walker J, Horwood MA, Tesoriero L, Glen M, Priest MJ (2010) Uredo rangelii , a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. Australas Plant Pathol 39:463–466. https://doi.org/10.1071/AP10102

Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, Grimwood J, Pérez G, Pisabarro AG, Grigoriev I V., Stajich JE, Ramírez L (2016) Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. PLoS Genet 12:e1006108. https://doi.org/10.1371/journal.pgen.1006108

Chan PP, Lowe TM (2019) tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

Coutinho TA, Wingfield MJ, Alfenas AC, Crous PW (1998) Eucalyptus rust: A disease with the potential for serious international implications. Plant Dis 82:819–825. https://doi.org/10.1094/PDIS.1998.82.7.819

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Duplessis S, Cuomo CA, Lin YC, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, Chiu R, Coutinho PM, Feaue N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr MG, Kodira CD, Kohler A, Kües U, Lindquist EA, Lucas SM, Mago R, Mauceli E, Morin E, Murat C, Pangilinan JL, Park R, Pearson M, Quesneville H, Rouhier N, Sakthikumar S, Salamov AA, Schmutz J, Selles B, Shapiro H, Tanguay P, Tuskan GA, Henrissat B, Van De Peer Y, Rouzé P, Ellis JG, Dodds PN, Schein JE, Zhong S, Hamelin RC, Grigoriev I V., Szabo LJ, Martin F (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci U S A 108:9166–9171. https://doi.org/10.1073/pnas.1019315108

Elliott TA, Gregory TR (2015) What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos Trans R Soc B Biol Sci 370:20140331. https://doi.org/10.1098/rstb.2014.0331

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: Ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17:333–351.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome

assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Graça RN, Ross-Davis AL, Klopfenstein NB, Kim M-S, Peever TL, Cannon PG, Aun CP, Mizubuti ESG, Alfenas AC (2013) Rust disease of eucalypts, caused by *Puccinia psidii* , did not originate via host jump from guava in Brazil. Mol Ecol 22:6033–6047. https://doi.org/10.1111/mec.12545

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2016) BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769. https://doi.org/10.1093/bioinformatics/btv661

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-genome annotation with BRAKER. In: Methods in Molecular Biology. Humana Press Inc., pp 65–95.

Holliday JL, Jones SA, Simpson JA, Glen M, Edwards J, Robinson A, Burgman MA (2013) A novel spore collection device for sampling exposure pathways: A case study of Puccinia psidii. Plant Dis 97:828–834. https://doi.org/10.1094/PDIS-06-12-0565-RE

Joly DL, Feau N, Tanguay P, Hamelin RC (2010) Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (Melampsora spp.). BMC Genomics 11:422. https://doi.org/10.1186/1471-2164-11-422

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467. https://doi.org/10.1159/000084979

Kammonen JI, Smolander O-P, Paulin L, Pereira PAB, Laine P, Koskinen P, Jernvall J, Auvinen P (2019) gapFinisher: A reliable gap filling pipeline for SSPACE-LongRead scaffolder output. PLoS One 14:e0216885. https://doi.org/10.1371/journal.pone.0216885

Kawanishi T, Uematsu S, Kakishima M, Kagiwada S, Hamamoto H, Horie H, Namba S (2009) First report of rust disease on ohia and the causal fungus, Puccinia psidii, in Japan. J Gen Plant Pathol 75:428–431. https://doi.org/10.1007/s10327-009-0202-0

Kiran K, Rawal HC, Dubey H, Jaswal R, Devanna BN, Gupta DK, Bhardwaj SC, Prasad P, Pal D, Chhuneja P, Balasubramanian P, Kumar J, Swami M, Solanke AU, Gaikwad K, Singh NK, Sharma TR (2016) Draft genome of the wheat rust pathogen (Puccinia triticina) unravels genome-wide structural variations during evolution. Genome Biol Evol 8:2702–2721. https://doi.org/10.1093/gbe/evw197

Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8

Kopylova E, Noé L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Krishnan P, Meile L, Plissonneau C, Ma X, Hartmann FE, Croll D, McDonald BA, Sánchez-Vallet A (2018) Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. BMC Biol 16:78. https://doi.org/10.1186/s12915-018-0543-2

Machado P da S, Glen M, Pereira OL, Silva AA, Alfenas AC (2015) Epitypification of Puccinia psidii, Causal Agent of Guava Rust. Trop Plant Pathol 40:5–12. https://doi.org/10.1007/s40858-014-0002-8

MacLachlan JD (1938) A rust of the pimento tree in Jamaica. Phytopathology 28:157–170.

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences

of k-mers. Bioinformatics 27:764–70. https://doi.org/10.1093/bioinformatics/btr011

Marlatt RB, Kimbrough JW (1979) Puccinia psidii on Pimenta dioica in south Florida. Plant Dis 63:510–512.

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, Luthe DS, Bridges SM, Burgess SC (2006) AgBase: A functional genomics resource for agriculture. BMC Genomics 7:229. https://doi.org/10.1186/1471-2164-7-229

McTaggart AR, Duong TA, Le VQ, Shuey LS, Smidt W, Naidoo S, Wingfield MJ, Wingfield BD (2018) Chromium sequencing: the doors open for genomics of obligate plant pathogens. Biotechniques 65:253–257. https://doi.org/10.2144/btn-2018-0019

Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K (2019) Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci Rep 9:4307. https://doi.org/10.1038/s41598-019-40965-0

Nemri A, Saunders DGO, Anderson C, Upadhyaya NM, Win J, Lawrence GJ, Jones DA, Kamoun S, Ellis JG, Dodds PN (2014) The genome sequence and effector complement of the flax rust pathogen Melampsora lini. Front Plant Sci 5:98. https://doi.org/10.3389/fpls.2014.00098

Pegg G, Taylor T, Entwistle P, Guymer G, Giblin F, Carnegie A (2017) Impact of Austropuccinia psidii (myrtle rust) on Myrtaceae-rich wet sclerophyll forests in south east Queensland. PLoS One 12:e0188058. https://doi.org/10.1371/journal.pone.0188058

Pegg GS, Giblin FR, McTaggart AR, Guymer GP, Taylor H, Ireland KB, Shivas RG, Perry S (2014) *Puccinia psidii* in Queensland, Australia: disease symptoms, distribution and impact. Plant Pathol 63:1005–1021. https://doi.org/10.1111/ppa.12173

Porto BN, Caixeta ET, Mathioni SM, Vidigal PMP, Zambolim L, Zambolim EM, Donofrio N, Polson SW, Maia TA, Chen C, Adetunji M, Kingham B, Dalio RJD, Resende MLV de (2019) Genome sequencing and transcript analysis of Hemileia vastatrix reveal expression dynamics of candidate effectors dependent on host compatibility. PLoS One 14:e0215598. https://doi.org/10.1371/journal.pone.0215598

Pryszcz LP, Gabaldón T (2016) Redundans: An assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 44:e113. https://doi.org/10.1093/nar/gkw294

Razali NM, Cheah BH, Nadarajah K (2019) Transposable elements adaptive role in genome plasticity, pathogenicity and evolution in fungal phytopathogens. Int. J. Mol. Sci. 20:3597.

Rochi L, Diéguez MJ, Burguener G, Darino MA, Pergolesi MF, Ingala LR, Cuyeu AR, Turjanski A, Kreff ED, Sacco F (2018) Characterization and comparative analysis of the genome of Puccinia sorghi Schwein, the causal agent of maize common rust. Fungal Genet Biol 112:31–39. https://doi.org/10.1016/j.fgb.2016.10.001

Ross-Davis AL, Stewart JE, Hanna JW, Kim M-S, Knaus BJ, Cronn R, Rai H, Richardson BA, McDonald GI, Klopfenstein NB (2013) Transcriptome of an Armillaria root disease pathogen reveals candidate genes involved in host substrate utilization at the host-pathogen interface. For Pathol 43:468–477. https://doi.org/10.1111/efp.12056

Roux J, Greyling I, Coutinho TA, Verleur M, Wingfield MJ (2013) The Myrtle rust pathogen, Puccinia psidii, discovered in Africa. IMA Fungus 4:155–159. https://doi.org/10.5598/imafungus.2013.04.01.14

Ruiz RAR, Alfenas AC, Ferreira FA, Vale FXR (1989) Influência da temperatura, do tempo de molhamento foliar, fotoperíodo e da intensidade de luz sobre a infecção de Puccinia psidii em

eucalipto. Fitopatol Bras 14:55–61.

Shen W, Le S, Li Y, Hu F (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One 11:e0163962. https://doi.org/10.1371/journal.pone.0163962

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637–644. https://doi.org/10.1093/bioinformatics/btn013

Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62

Stewart JE, Ross-Davis AL, Graça RN, Alfenas AC, Peever TL, Hanna JW, Uchida JY, Hauff RD, Kadooka CY, Kim M-S, Cannon PG, Namba S, Simeto S, Pérez CA, Rayamajhi MB, Lodge DJ, Arguedas M, Medel-Ortiz R, López-Ramirez MA, Tennant P, Glen M, Machado PS, McTaggart AR, Carnegie AJ, Klopfenstein NB (2018) Genetic diversity of the myrtle rust pathogen (Austropuccinia psidii) in the Americas and Hawaii: Global implications for invasive threat assessments. For Pathol 48:e12378. https://doi.org/10.1111/efp.12378

Tan M-K, Collins D, Chen Z, Englezou A, Wilkins MR (2014) A brief overview of the size and composition of the myrtle rust genome and its taxonomic status. Mycology 5:52–63. https://doi.org/10.1080/21501203.2014.919967

Tan MH, Austin CM, Hammer MP, Lee YP, Croft LJ, Gan HM (2018) Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (Amphiprion ocellaris) genome assembly. Gigascience 7:1–6. https://doi.org/10.1093/gigascience/gix137

Tavares S, Ramos AP, Pires AS, Azinheira HG, Caldeirinha P, Link T, Abranches R, Silva M do C, Voegele RT, Loureiro J, Talhinhas P (2014) Genome size analyses of Pucciniales reveal the largest fungal genomes. Front Plant Sci 5:422. https://doi.org/10.3389/fpls.2014.00422

Uchida J, Zhong S, Killgore E (2006) First Report of a Rust Disease on Ohia Caused by Puccinia psidii in Hawaii . Plant Dis 90:524–524. https://doi.org/10.1094/pd-90-0524c

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC (2017) GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33:2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963

Winter G (1884) Repertorium. Rabenhorstii fungi europaei et extraeuraopaei. Cent. XXXI et XXXII. Hedwigia 23:164–172.

Zhong S, Yang B, Puri KD (2011) Characterization of Puccinia psidii isolates in Hawaii using microsatellite DNA markers. J Gen Plant Pathol 77:178–181. https://doi.org/10.1007/s10327-011-0303-4

Zhuang JY, Wei SX (2011) Additional materials for the rust flora of Hainan Province, China. Mycosystema 30:853–860.

CHAPTER 2 - TRANSCRIPTOME ANALYSIS OF *Eucalyptus grandis* GENOTYPES REVEALS CONSTITUTIVE OVEREXPRESSION OF RESISTANCE-RELATED GENES TO RUST (*Austropuccinia psidii*)

Samuel A. Santos[1,4], Pedro M. P. Vidigal[2], Lúcio M. S. Guimarães[1], Reginaldo G. Mafia[3], Matthew D. Templeton[4], Acelino C. Alfenas[1,*]

[1] Laboratory of Forest Pathology, Department of Plant Pathology, Universidade Federal de Viçosa, Minas Gerais State, Brazil.

[2] Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Minas Gerais State, Brazil.

[3] Centro de Tecnologia, Suzano S.A., Aracruz, Espírito Santo State, Brazil.

[4] The New Zealand Institute for Plant and Food Research Limited, Auckland 1142, New Zealand.

* Corresponding author. Department of Plant Pathology, Instituto de Biotecnologia Aplicada à agropecuária-BIOAGRO, Universidade Federal de Viçosa, Av. P.H. Rolfs s/n, Campus Universitário, 36570-900, Viçosa, MG, Brazil. Tel.: +55 31 3612 2428. E-mail address: *aalfenas@ufv.br* (*A.C., Alfenas*).

Authors Orcid-ID

Samuel A. Santos: 0000-0002-4192-0829

Pedro M. P. Vidigal: 0000-0002-5116-9856

Lúcio M. S. Guimarães: 0000-0003-1397-8086

Reginaldo G. Mafia: 0000-0003-2408-5531

Matthew D. Templeton: 0000-0003-0192-9041

Acelino C. Alfenas: 0000-0001-7776-3362

Obs: This chapter has been published as an original article and still under revision at Plant Molecular Biology. The full publication can be accessed through the following link: https://doi.org/10.1007/s11103-020-01030-x

Key message A resistant *E. grandis* genotype showed a constitutive overexpression of genes related to resistance to myrtle rust caused by *A. psidii*.

Abstract Myrtle rust caused by *Austropuccinia psidii* is considered one of the most important fungal diseases affecting *Eucalyptus* spp. plantations in Brazil. Although the selection and planting of resistant eucalypt genotypes have been the major strategies to manage the disease in Brazil, the molecular mechanisms involved in resistance are still unclear. In this study, we evaluated the gene expression profile of two contrasting *Eucalyptus grandis* genotypes in resistance level to rust by RNA-Seq. The two genotypes showed a very different background gene expression level even without *A. psidii* infection. The resistant genotype had a constitutive overexpression of a large number of protein-coding genes compared to the susceptible genotype. These genes were mainly associated with signal transduction, photosynthesis, regulation and response to salicylic acid (SA), and protein kinase leucine-rich receptors (PK-LRR). PK-LRR and SA mediated disease resistance are well known to be effective against obligate biotroph pathogens, such as *A. psidii*. In addition, at 24 hours after infection, the susceptible genotype was able to activate some response, however, several resistance-related proteins had their expression level reduced with *A. psidii* infection. Here, we present the first analysis of *E. grandis* genotypes transcriptomes infected by *A. psidii* and it reveals a constitutive overexpression of several resistance-related genes in the resistant genotype compared to the susceptible one. Our findings have the potential to be used as candidate molecular markers for resistance to myrtle rust.

Keywords: Myrtle rust. RNA-Seq. DEGs. Protein kinase. Leucine-rich receptors.

INTRODUCTION

Eucalypt (*Eucalyptus* spp. L'Hér) is the most commonly grown forest tree in Brazil (IBÁ 2019). Eucalypt plantations support a multi-billion-dollar industry based on cellulose pulp and paper, charcoal for the steel industry and solid products (Grattapaglia and Kirst 2008). As the planted areas have expanded plus the anticipated impact climate changes, several pathogens have threatened the eucalypt plantations throughout the country. Eucalypt rust caused by *Austropuccinia psidii* (Winter) Beenken (Beenken 2017) is one of the most important fungal diseases affecting *Eucalyptus* spp. in Brazil (Alfenas et al. 2003; Alfenas et al. 2009). *A. psidii* infects mainly young shoots and coppice (Alfenas et al. 2009), which are common in 1-year old

eucalypt trees up to 3.0 m in height (Zauza et al. 2010b). The symptoms depend on the susceptibility of the eucalypt genotype but can range from small uredinias with few urediniospores to massive production of powdery bright yellow spores and shoot dieback, which may ultimately lead in plant death (Alfenas et al. 2009).

Eucalyptus grandis is one of the most commonly planted eucalypt species worldwide (Harwood 2011). Although most *E. grandis* genotypes are susceptible to rust (Zauza et al. 2010a; Silva et al. 2013), selection and planting of resistant eucalypt clones have been the major strategies for managing the disease in Brazil (Miranda et al. 2013; Silva et al. 2013; Santos et al. 2014). A study of the inheritance of rust resistance in full-sibling families of *E. grandis* found that most of the resistance phenotypic variation is controlled by a major locus, named *Ppr1* (*Puccinia psidii* resistance gene 1) (Junghans et al. 2003). This locus was positioned in the reference genetic map for *Eucalyptus* on linkage group 3 and validated in two unrelated families, which suggest that *Ppr1* is important on the rust resistance segregation (Mamani et al. 2010). However, as already stated by Junghans et al. (2003) there was evidence of a more complex pattern of inheritance. At least in one case, even if the family was derived from a genotype that carried *Ppr1*, a high number of susceptible plants was observed which suggests that the segregation of resistance did not fit a simple Mendelian model (Junghans et al. 2003).

In recent years, further studies have mapped several quantitative trait loci (QTLs) in *Eucalyptus* spp. and *Corymbia* spp., indicating that the genetic control of rust resistance is complex and may depend on the many other minor-effect genes (Rosado et al. 2010; Lima et al. 2011; Alves et al. 2012; Butler et al. 2016; Butler et al. 2019). However, information about the time course of expression and level, as well as the relationship among these genes remains unknown. RNA-Seq is an approach to assess transcriptome profiling via high throughput sequencing technologies, which provides expression information of a large number of transcripts and their isoforms (Wang et al. 2009). The technique has been widely applied in transcriptome analysis in several plant-pathogen interactions, including in response to *A. psidii* (Zhu et al. 2013; Hayden et al. 2014; Liu et al. 2015; Chakraborty et al. 2016; Ye et al. 2017; Hsieh et al. 2018; Tobias et al. 2018), however, it is still unexplored for *Eucalyptus-A. psidii* pathosystem.

In the *E. grandis-A. psidii* interaction, the primary mycelia and haustoria are most observed at 24 hours after infection (hai) (Xavier et al. 2001). The haustorium is a specialized structure of rust fungi that produces and delivers effector proteins to the host cytoplasm where they are thought to promote the infection (Garnica et al. 2014; Lorrain et al. 2019). Thus, 24

hai seems to be a crucial phase of interaction between *E. grandis* and *A. psidii*. Here, in order to better comprehend the molecular mechanisms of resistance to rust, we analyze the transcriptome profiling of two *E. grandis* genotypes (resistant and susceptible) at 24 hours after *A. psidii* inoculation using RNA-Seq.

MATERIAL AND METHODS

Experimental condition and samples collect

In this study, a rust resistant (CLR385 = G21) and susceptible (CLR220 = G45) *E. grandis* genotypes were employed (Junghans et al. 2003). Both eucalypt genotypes were obtained from the Breeding Program of the Suzano S.A., Itapetininga, São Paulo, Brazil. They originated from the same species (*E. grandis*) and the same population. Over the years, these two *E. grandis* genotypes have been used in studies involving histological, phenotypical, molecular, and physiological responses to *A. psidii* (Xavier et al. 2001; Junghans et al. 2003; Xavier et al., 2015; Santos et al., 2019). Eucalypt plants were clonally propagated by cutting. At 60 days of age, the cuttings were transplanted into 2 L pots containing MecPlant® substrate enriched with 26 g of super simple phosphate and 12 g of Osmocote® (19-6-10). During transplanting, each plant per pot received 100 mL of mono-ammonium phosphate solution per pot (P and N at 52% and 12%, respectively; Vale Fertilizantes S.A., Uberaba, MG, Brazil). Plants were cultivated in a greenhouse with an average temperature of ~25 °C (± 5 °C) and natural light (~12:12 h photoperiod) for 30 days, then they were moved to the inoculation chamber.

The experiment was conducted in randomized blocks with five replicates and four treatments as following: resistant genotype non-infected (R-NI), resistant genotype infected (R-I), susceptible genotype non-infected (S-NI), and susceptible genotype infected (S-I). The UFV-2 isolate of *A. psidii* (race 1) (Xavier et al. 2015) was used as inoculation. Fungal spores were multiplied in plants of *Sygygium jambos* as described by Ruiz et al. (Ruiz et al. 1989). An inoculum suspension (2×10^4 urediniospores mL^{-1}), prepared with sterilized water and Tween 20® (0.05%), was sprayed evenly on both leaf surfaces, using an electric compressor Jet Master 1/3 HP (Schulz, Joinville, SC, Brazil). To evaluate the efficiency of the inoculation, three *S. jambos* plants were randomly distributed among the eucalypt plants. Non-infected control plants from both eucalypt genotypes were inoculated only with sterilized water plus Tween 20® (0.05%). After inoculation, plants were incubated in a mist irrigation chamber, in the dark. After 24 h of incubation, plants were maintained in a growth chamber at 22 °C, with a 12 h

photoperiod and a light intensity of 130 µmol photons m^{-2} s^{-1} (Ruiz et al. 1989), until the end of the experiment.

The experiment was repeated three times. In each experiment, leaf samples from the apical part (with two internodes) of three branches per plant were collected at 24 hai. The samples were immediately frozen in liquid nitrogen and then stored in an ultra low temperature (ULT) freezer at -80 °C until use. Samples that showed the strongest symptoms of infection on the plants of susceptible genotype, were kept for further analysis.

RNA samples preparation

Samples collected at 24 hai were recovered from -80°C ULT freezer, put in liquid nitrogen and disrupted using a pair of Stainless Steel Grinding Jar 10 mL (Qiagen Sciences Inc., Germantown, MD, USA) with Tissuelyser (Qiagen Sciences Inc.) for three bursts of 15 seconds at 25 Hz frequency. After the first 15 seconds, the closed grinding jars were immersed into liquid nitrogen for 10 min up to the next disruption cycle, so that the samples remained frozen until complete disruption. Total RNA was extracted using PureLink® Plant RNA Reagent (ThermoFisher Scientific, Waltham, MA, USA), according to manufacturer's instruction. RNA was suspended in RNase-free water. Before sequencing, the RNA samples were subjected to quality control (QC) test using a NanoDrop 2000c spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) and Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). QC results can be found in Supplementary File 1.

RNA library preparation and transcriptome sequencing

Three replicates per treatment were sequenced, giving 12 RNA libraries. The rRNA in total RNA was depleted by Ribo-Zero kit (Illumina Inc., San Diego, CA, USA). The enriched mRNA samples were subjected to Illumina cDNA library construction using TruSeq Stranded mRNA Prep kit (Illumina Inc.) according to the manufacturer's instructions. The cDNA libraries were sequenced using the Illumina NovaSeq 6000 platform (Illumina Inc.) at DNA Link USA, Inc., San Diego, California, USA (http://www.dnalink.com/). A total of approximately $5x10^7$ of 100 bp paired-end (PE100) reads per library were obtained.

Raw data quality control

SortMeRNA v. 2.1 (Kopylova et al. 2012) was used to filter raw data and remove reads from rRNA. Low-quality reads (Q20 threshold) and Illumina adapter sequences were removed using Trimmomatic v. 0.36 (Bolger et al. 2014). Quality control checks were performed by FastQC v. 0.11.5 (Andrews 2010) before and after each filtering step to ensure the quality of the final dataset (clean reads).

Reads mapping to the *E. grandis* genome and quantification of gene expression levels

The *E. grandis* genome and its GFF (general feature format) annotation file were downloaded from the GenBank database (Accession Number AUSX00000000.1, downloaded on July 15th, 2019) (Myburg et al. 2014; Bartholomé et al. 2015) and used in all subsequent analysis. GFF file was converted to GTF (gene transfer format) using gffread tool of Cufflinks v. 2.2.1 (Trapnell et al. 2010). These were selected to indexing the *E. grandis* genome using STAR v. 2.7.1a (Dobin et al. 2013). Cleaned RNA-Seq reads were aligned to the indexed genome using STAR and reads count per gene were obtained by selecting '**--**quantMode' and '--twopassMode' functions (Dobin et al. 2013).

Analysis of differentially expressed genes

The unstranded reads count per gene from all RNA-seq libraries were compiled into a single count table (Supplementary File 2), which was used as input for differential gene expression analysis in R v. 3.6.1 (https://cran.r-project.org/). To analyze the background expression of the two eucalypt genotypes we first compared resistant vs susceptible non-infected (R-NI vs S-NI). To evaluate the effect of pathogen infection in both genotypes we compared resistant infected vs resistant non-infected (R-I vs R-NI) and susceptible infected vs susceptible non-infected (S-I vs S-NI). Differentially expressed genes (DEGs) analysis was performed using DESeq2 R package v. 1.26.0 (Love et al. 2014). The P value for false discovery rate was adjusted (padj) using the Hochberg and Benjamini test. DEGs with padj $\leq$ 0.0001 were considered as significant.

GO enrichment analyses of DEGs

Gene Ontology (GO) enrichment analysis of DEGs was performed using DAVID database v. 6.8. DAVID Gene Functional Annotation Tool (Huang et al. 2007; Huang et al. 2009), allowing the classification of a large number of genes into biological process (BP), cellular component (CC) and molecular functions (MF) categories. A list with official gene names (Locus ID), based on GenBank annotation (Accession number AUSX00000000.1), of DEGs was submitted to DAVID (https://david.ncifcrf.gov/summary.jsp) to functionally annotation, selecting *E. grandis* as species. P values were adjusted (padj) by Benjamini and Hochberg test. GO terms with a padj ≤ 0.05 and gene count ≥ 2 were considered as significantly enriched.

Clustering analysis of DEGs and *E. grandis* resistance-related proteins

The Markov clustering program OrthoMCL v. 2.0 (Li et al. 2003) was used to identify orthologous genes between DEGs (protein-coding genes) and *E. grandis* resistance-related genes according to the Plant Resistance Genes database (PRGdb) v. 3.0 (Osuna-Cruz et al. 2018). Thus, amino acid sequences of proteins encoded by DEGs and *E. grandis* PRGdb were submitted to reciprocal similarity searches using BLASTp tool of BLAST v. 2.9.0 (Altschul et al. 1990), considering an E-value of 1e-10 as a threshold for the significant alignments. Putative orthologues were then clustered with OrthoMCL using an inflation value of 1.5 and a similarity of 50%.

Single Nucleotide Polymorphisms (SNP) analysis

Cleaned RNA-Seq reads from each eucalypt genotype were mapped to the indexed *E. grandis* reference genome (GenBank accession number AUSX00000000.1, downloaded on July 15th, 2019) using STAR v. 2.7.1a with default parameters (Dobin et al. 2013). The Sequence Alignment Map files were processed using Picard v. 2.18.27 (https://github.com/broadinstitute/picard/) to produce Binary Alignment Map files, containing ordered and deduplicated data. A flag identifying each eucalypt genotype ("susceptible" and "resistant") was added to each mapping file. Sequence variants were called using FreeBayes v. 1.2.0 (https://github.com/ekg/freebayes) (Garrison and Marth 2012) setting the ploidy as diploid and considering a minimum mapping quality of 30, minimum base quality of 30, and minimum coverage of 30 reads at every position in the reference genome. After variant calling,

SNPs were filtered using vcftools v. 0.16.15 (https://vcftools.github.io/index.html) and AWK shell scripts. We focused only on loci mapped for both genotypes and the effects of polymorphisms were analyzed and annotated using Ensembl Variant Effect Predictor v. 99.2 (https://github.com/Ensembl/ensembl-vep), considering the *E. grandis* genome Annotation Release 101.

PCA, heat map, and volcano plots analyses

Principal Components Analysis (PCA), heat map and volcano plots were generated in R v. 3.6.1 using the plotPCA, heatmap.2, and plot functions, respectively, which are available in the gplots CRAN library package (https://cran.r-project.org/). PCA and heat map analyses were performed using rlog-transformed values of read count per gene data from all gene set and the 1,000 most variable genes, respectively (Supplementary File 2). The subset containing the most variable genes was obtained using the topVarGenes function from geneFilter library (Love et al. 2015). Volcano plots were created using padj and log2FoldChange values of the significant DEGs from each comparison.

Ppr1 cluster genes analysis

Puccinia psidii resistance gene 1 (Ppr1) for resistance to *A. psidii* (myrtle rust) has been mapped on the Eucalyptus reference map (Mamani et al. 2010). EMBRA125, one of the flanking markers of *Ppr1* locus is closely linked to marker ePT_640786 (Kullan et al. 2012). This marker sequence has an estimated position of 52,900,000 bp on chromosome 3 of the *E. grandis* reference genome sequence and overlaps with the position of supercluster C-3 (Christie et al. 2016). We evaluated the expression profile and SNPs of genes located at a 5 Mbp window (2.5 Mbp up and downstream) at the *Ppr1* locus.

RESULTS

Symptoms of *A. psidii* infection

In all experiments, the *S. jambos* plants showed typical symptoms of rust, confirming the efficiency of the inoculation. At eight days after inoculation, uredinia with abundant urediniospores production, typical of eucalypt rust, were observed on the leaves of infected plants of the susceptible eucalypt genotype (Fig. 1A). However, inoculated plants of resistant

genotype did not show any symptoms (Fig. 1A). Furthermore, control plants of both eucalypt genotypes remained asymptomatic.

Transcriptome sequencing and read mapping to *E. grandis* genome

After the trimming of low-quality reads and filtering to remove those that mapped to rRNA, high-quality clean data was obtained for each sample (Table 1). For all RNA-Seq libraries, more than 93.66% of reads had a high-quality score (Phred value > 30) (Table 1). GC content was very similar among the samples and ranges from 50.5 to 51.0%. The proportion of clean reads that mapped to the *E. grandis* genome of each sample was greater than 91.35% (Table 1).

Read count per gene data: distribution and clustering

The set of samples from resistant and susceptible eucalypt genotypes were separated in distinct clusters in the principal component analysis plot (Fig. 1B). The same was observed for the treatments S-NI and S-I, indicating that they have distinct transcriptome profiles. However, there was an overlap between R-NI and R-I (Fig. 1B), indicating that resistant genotype transcriptome was not strongly affected by pathogen infection.

A clustering heatmap analysis of the 1,000 most altered genes also showed a contrasting gene expression profile between samples from resistant and susceptible genotypes (Fig. 1C). R-NI and R-I treatments had a very similar profile of gene expression; however, S-NI and S-I showed a different profile (Fig. 1C).

Differentially expressed genes

DEGs were classified as protein-coding, long non-coding RNA (lncRNA), pseudogenes, and miscellaneous RNAs (miscRNAs). In the R-NI vs S-NI, a total of 4,709 DEGs, including 2,390 up-regulated (1,992 proteins, 252 lncRNAs, 141 pseudogenes, and five miscRNAs) and 2,319 down-regulated (1,874 proteins, 282 lncRNAs, 155 pseudogenes, and eight miscRNAs) genes were identified (Figs. 2A-B, Supplementary File 3: Tables S1 and S2).

In the R-I vs R-NI, only two significantly DEGs were observed (Figs. 2A and C, Supplementary File 3: Table S3). However, in the S-I vs S-NI, 1,484 up-regulated (1,261 proteins, 129 lncRNAs, 85 pseudogenes, and nine miscRNAs) and 1,271 down-regulated (1,061 proteins, 135 lncRNAs, 72 pseudogenes, and three miscRNAs) genes, totalizing 2,755 DEGs identified (Figs. 2A and D, Supplementary File 3: Tables S4 and S5).

A total of 1,163 DEGs were common in both contrasts (R-NI vs S-NI and S-I vs S-NI), which of 636 (55%) were up and 527 (45%) were down-regulated (Fig. 2E). R-NI vs S-NI showed 3546 exclusive DEGs with 1,754 (49%) up and 1,792 (51%) down-regulated (Fig. 2E). On the other hand, S-NI vs S-I had 1,592 exclusive DEGs, which 848 (53%) were up and 744 (47%) were down-regulated (Fig. 2E).

GO enrichment analysis of DEGs

Among up-regulated DEGs, in the R-NI vs S-NI, the significantly enriched BP were those involved in metabolic process, lipid metabolic process, protein-chromophore linkage, signal transduction, photosynthesis, salicylic acid (SA) mediated signaling, and SA responsive genes (Table 2). CC category was mainly enriched in integral component of membrane, photosystem I, photosystem II, and chloroplast thylakoid membrane (Table 2). Furthermore, enriched terms in MF were protein kinase activity, iron ion binding, catalytic activity, ADP binding, chlorophyll-binding, acid phosphatase activity, and iron ion transmembrane transporter activity (Table 2).

In the S-I vs S-NI comparison, BP category was enriched in metabolic process and defense response (Table 2). CC was enriched only for integral component of membrane. In addition, the significantly enriched terms observed for the MF category were heme-binding, transferase activity, monooxygenase activity, calcium ion binding, ADP binding, amino acid biding, and ribonuclease activity (Table 2).

Down-regulated DEGs from R-NI vs S-NI were significantly enriched as metabolic process in BP, cytoplasm in CC, transferase and hydrolase activities in the MF category (Table 3). For the S-I vs S-NI comparison, GO terms were mostly enriched as membrane in CC, zinc ion binding, oxidoreductase, protein kinase and ADP binding in MF category (Table 3).

DEGs related to resistance

Among up-regulated DEGs, 62 orthologous clusters related to resistance were conserved with 401 proteins from R-NI vs S-NI and only 267 from S-I vs S-NI (Fig. 3A). A total of 58 clusters with 69 proteins were unique to R-NI vs S-NI (Fig. 3A). Of these, 46 (67%) are proteins kinases (PK), 13 (19%) are resistance-related proteins (R proteins), six (9%) are uncharacterized (U), three (4%) have other functions, and one (1%) is polygalacturonase inhibitors (Table 4). On the other hand, S-I vs S-NI had only 34 exclusive clusters comprising 37 resistance-related proteins

(Fig. 3A) with 29 (78%) PK, four (11%) R proteins, two (5.5%) enzymes, and two (5.5%) have other functions (Table 5).

For the down-regulated DEGs, 46 clusters were conserved, with 237 proteins from R-NI vs S-NI and 165 from S-I vs S-NI (Fig. 3B). A total of 55 orthologous clusters with 64 proteins were exclusive to R-NI vs S-NI (Fig. 3B), containing 42 (66%) PK, 10 (16%) R proteins, six (9%) have other functions, four (6%) U, and two (3%) polygalacturonase inhibitor (Supplementary File 4: Table S6). Furthermore, in the S-I vs S-NI, 24 exclusive clusters with 25 proteins were identified (Fig. 3B), comprising 13 (52%) PK, 6 (24%) R proteins, three (12%) have other functions, one (4%) transcription factor, one (4%) enzyme, and one (4%) U (Table 5).

SNP analysis

Among the 326,407 SNPs identified in coding sequence regions for both resistant and susceptible eucalypt genotypes, 75,545 are synonymous substitutions, 56,526 are non-synonymous substitutions and 194,336 are other sequence variants, such as those located at regulatory regions and introns (Table 6, Supplementary File 5: Table S7). Of these, 222,667 (68.2%) were exclusive to the susceptible genotype, 36,722 (11.3%) were shared by both genotypes, and 67,018 (20.5%) were exclusive to the resistant genotype (Table 6, Supplementary File 5: Table S7). For both genotypes, the majority of SNPs were assigned as synonymous and heterozygous. However, in this study, we focused on non-synonymous and homozygous SNPs located at coding sequences of proteins related to resistance such as PK and R proteins, and that distinguish resistant and susceptible genotypes (Supplementary File 6: Table S8).

Among the up-regulated DEGs in R-NI compared to S-NI, 379 nonsynonymous and homozygous SNPs were observed in 178 transcripts (Supplementary File 6: Table S9). Interestingly, 22 SNPs were identified in 16 PK with serine-, proline- or leucine-rich domains, while the other 27 were observed in nine R proteins (Supplementary File 6: Table S9). For down-regulated DEGs in R-NI, 361 SNPs were found in 190 transcripts. Of these, 23 SNPs were associated with eight PK and the other five were identified in four R proteins (Supplementary File 6: Table S9).

In the susceptible genotype, 145 and 225 nonsynonymous and homozygous SNPs were identified in 70 and 115 transcripts for the up- and down-regulated DEGs in S-I compared to S-NI, respectively (Supplementary File 6: Table S9). Among up-regulated DEGs, 23 and six

SNPs were associated with 10 PK and three R proteins, respectively (Supplementary File 6: Table S9). Furthermore, for the down-regulated DEGs in S-I, 28 and 26 SNPs were identified in 10 PK and 14 R proteins, respectively (Supplementary File 6: Table S9).

DEGs and SNP analyses in *Ppr1* cluster genes

Among 101 transcripts identified in a 5 Mbp window at the *Ppr1* locus, seven transcripts were assigned as LRR kinase and R proteins (Supplementary File 7: Table S10). Of these, two R proteins were significant up- and down-regulated in both R-NI vs S-NI and S-I vs S-NI comparisons (Supplementary File 7: Table S11). However, none of these genes was identified as differentially expressed in R-I compared to R-NI at 24 h after *A. psidii* inoculation. Moreover, non-synonymous substitutions were identified in six transcripts related to resistance from both genotypes (Supplementary File 7: Table S12). More specifically, the resistant genotype showed nine heterozygous and two homozygous totalizing 11 exclusive non-synonymous substitutions (Table 7).

DISCUSSION

In this study, we obtained high-quality data from 12 deep-sequenced RNA-Seq libraries of two *E. grandis* genotypes: one resistant and another susceptible to rust (*A. psidii*). Our data had a high consistency among replicates and treatments as judged by PCA and heatmap clustering analyses. Additionally, at the end of the experiments, eucalypt plants of the susceptible genotype showed typical symptoms of rust infection, however, the resistant genotype remained asymptomatic, confirming previous results (Xavier et al. 2001; Junghans et al. 2003; Santos et al. 2019).

The two *E. grandis* genotypes showed a very different background of gene expression level in the absence of *A. psidii* infection. The resistant genotype constitutively overexpressed a large number of protein-coding genes compared to the susceptible one. Interestingly, GO enrichment analysis of these proteins revealed that terms involved in the plant defense against pathogen infection such as signal transduction, photosynthesis, regulation and response to SA, photosystems I and II, PK activity, and acid phosphatase activity were significantly enriched (Table 2). Moreover, an OrthoMCL clustering analysis found 58 orthologous clusters containing 69 resistance-related proteins. Of these, the majority are PK with leucine or proline-rich receptors and R proteins but also polygalacturonase inhibitors (Table 4). During pathogen-

host interactions, rust fungi such as *A. psidii* produced a specialized structure called haustoria that secret into the host cells a wide range of effector proteins (Garnica et al. 2014; Lorrain et al. 2019). Fungal effectors play a crucial role in pathogenicity (Lo Presti et al. 2015; Uhse and Djamei 2018), modulating the transcription and suppression of genes induced in response to pathogen infection (Ahmed et al. 2018). Pathogen recognition and signaling are crucial steps in the plant immune system (Jones and Dangl 2006; Nobori and Tsuda 2019). Immune sensors such as PK containing leucine-rich receptors (LRR) are able to recognize a wide broad of pathogen effectors and activate defense responses (Zhao et al. 2005; Afzal et al. 2008; Padmanabhan et al. 2009; DeYoung et al. 2012; Hohmann et al. 2017; Palm-Forster et al. 2017; Kourelis and van der Hoorn 2018; Chakraborty et al. 2019). After direct or indirect recognition (Van der Biezen and Jones 1998; Dangl and Jones 2001), signal transduction via phosphorylation cascades within the plant cell is initiated (Zhao et al. 2005; Monaghan and Zipfel 2012; Bigeard et al. 2015). LRR proteins activation induce SA dependent responses (Jones and Dangl 2006). SA is a signal molecule in local defenses and in systemic acquired resistance (Durner et al. 1997). Increases in the endogenous levels of SA contributes to the expression of several R proteins and consequently activation of disease resistance (Dorey et al. 1997; Shah 2003). LRR and SA mediated disease resistance are effective against obligate biotrophic pathogens (Glazebrook 2005; Jones and Dangl 2006; Naidoo et al. 2014) such as *A. psidii*. In addition, plant pathogen polygalacturonase genes are often expressed during early infection stages and have an important role in triggering plant defense responses (De Lorenzo et al. 2001; Di et al. 2006).

Photosynthesis has a crucial role in plant defense. ATP, NADPH, and carbohydrates generated by photosynthesis are utilized for the synthesis of many important compounds, such as primary metabolites, SA, and antimicrobial compounds (Lu and Yao 2018). Our results show that photosynthetic pathways of the resistant genotype plants are more transcriptionally active than the susceptible one. Enhanced carbon fixation pathways increase the demand for ATP and NADPH, which consequently leads to positive feedback on the electron transport chain (Fernández-Marín et al. 2020). Noteworthy, several genes associated with these energetic molecules were up-regulated in R-NI compared to S-NI such as XP_010033434.1, XP_010033435.1, XP_018719250.1, XP_018721923.1, XP_010048745.1, XP_010054116.1, XP_018731779.1, XP_010065570.1, XP_018732584.1, XP_010067502.1 (Supplementary File 3: Table S1). Additionally, in the resistant plants, an increase in the gene expression level of enzymes present in both the reduction [ribulose bisphosphate carboxylase (XP_010043499.1),

phosphoglycerate kinase (XP_010060861.1), glyceraldehyde-3-phosphate dehydrogenase (XP_018716342.1, XP_010048829.1)] and regeneration [triose phosphate isomerase (XP_010057365.1)] phases of the Calvin cycle was observed (Supplementary File 3: Table S1).

At 24 h after *A. psidii* inoculation, the resistant genotype did not significantly change its gene expression level, even for 101 transcripts related to resistance in a 5 Mbp window at the *Ppr1* locus. Typically, plant constitutive defense mechanisms are well known as physical barriers such as the plant cell wall and cuticle (Freeman and Beattie 2008; Hématy et al. 2009). Plants often refrain from producing toxic compounds and expression of R proteins until pathogens are detected due to the high energy cost involved (Jones and Dangl 2006; Freeman and Beattie 2008; Nobori and Tsuda 2019). However, our results show that the effective resistance to *A. psidii* may be related to the constitutive overexpression of several PK with LRR, R proteins, SA response, and higher photosynthesis level. Constitutive expression of LRR and R proteins have been largely associated with plant resistance to pathogens (Sarowar et al. 2006; Budak et al. 2006; Maschietto et al. 2016). In addition, overexpression of PK containing LRR and R proteins were previously associated to be involved in the defense response of two Myrtaceae native species from Australia infected by *A. psidii* (Hsieh et al. 2018; Tobias et al. 2018). LRR genes were also identified as differentially expressed in two *E. grandis* clones with different resistance level to canker disease caused by *Chrysoporthe austroafricana* (Christie et al. 2016). Another explanation for a little significant alteration on the gene expression level in the R-I compared to R-NI (only two DEGs) could be related to a very early response. Our transcriptome analysis was done at 24 hours after inoculation. Response time is considered as a crucial key for plant resistance, especially to biotrophic pathogens such as rust fungi. In some cases, the response is very quickly, such as in barley that can activate a resistance related gene (*Rpg1*) at 4 min after inoculation of *Puccinia graminis* f. sp. *tritici* (Nirmala et al. 2011). However, in wheat the response time to *P. triticina* ranged from 4 to 12 h (Yadav et al. 2016; Shen et al. 2017), reaching 24 h after inoculation of *P. striiformis* f. sp. *tritici* (Dobon et al. 2016).

Among the down-regulated DEGs from R-NI vs S-NI, 55 orthologous clusters with 64 proteins were found to been related to *E. grandis* resistance (Supplementary File 4: Table S6). Since these proteins were down-regulated in R-NI, so they were up-regulated in S-NI, therefore, comprise the constitutively overexpressed proteins related to defense mechanisms from susceptible genotype. The main difference between the constitutively overexpressed proteins from both *E. grandis* genotypes is regarding the rich protein residues coupled to the kinase

domain. For the susceptible genotype, the majority are cysteine-rich receptors (CRR) (Supplementary File 4: Table S6), while for the resistant one had LRR (Table 4). Although LRR is considered to be the largest family of plant receptor kinases in *Arabidopsis* (Gou et al. 2010; Chakraborty et al. 2019) and also is the major class of genes linked to resistance to rust (Butler et al. 2016), differences on efficiency of pathogen recognition by LRR and CRR are still unknown.

The susceptible genotype was able to activate some response at 24 h after *A. psidii* inoculation. Among the up-regulated DEGs of S-I compared to S-NI, defense response was found as significantly enriched biological process (Table 2). Additionally, 34 orthologous clusters containing 37 proteins were found related to resistance (Fig. 3A). However, the susceptible genotype may not have the required background for effective defense, since other crucial biological processes such as signal transduction and photosynthesis (Jones and Dangl 2006; Lu and Yao 2018; Nobori and Tsuda 2019) were not found as enriched (Table 5). In our study, a total of 24 orthologous clusters with 25 resistance-related proteins had their expression level reduced with *A. psidii* infection (Fig. 3B). The majority are PK that plays an essential role in signal transduction and plant immunity (Tena et al. 2011; Tang et al. 2017; Chakraborty et al. 2019), including two proteins (XP_010039878.2 and XP_010042765.1) related to leaf rust resistance (Table 5) and one R protein (XP_018725762.1) located in the *Ppr1* cluster genes (Supplementary File 7: Table S11). Additionally, one transcription factor (XP_018722094.1: WRKY transcription factor 19) also had the expression level reduced. Transcription factors (TF) such as WRKY family genes are very important for plant immunity system, especially for induce the expression of defense-related genes (Ryu et al. 2006; Eulgem and Somssich 2007; Naoumkina et al. 2008; Pandey and Somssich 2009; Amorim et al. 2017). Functional studies on TFs that mediate defense responses in *Eucalyptus* are limited (Naidoo et al. 2014). However, WRKY genes have enhanced the resistance to a broad-spectrum of fungal pathogens in several plant species such as Arabidopsis (Zheng et al. 2006; Lai et al. 2008; Zhao et al. 2018), rice (Shimono et al. 2007; Inoue et al. 2013), barley (Gao et al. 2018), soybean (Cui et al. 2019), and wheat (Li et al. 2020). TF genes frequently have been targeted by pathogen effectors (Ding and Redkar 2018). In this context, it can be speculated that the down-regulated genes associated to plant defense such as PK, R, and TF genes in the susceptible genotype infected by *A. psidii* may be targeted for effector proteins, however, further investigations have to be done to prove this point.

Previous studies have mapped several QTLs in *Eucalyptus* spp. and *Corymbia* spp., indicating that the genetic control of rust resistance is complex and may depend on the many other minor-effect genes (Rosado et al. 2010; Lima et al. 2011; Alves et al. 2012; Butler et al. 2016; Butler et al. 2019). However, the identity and characterization of these involved genes were unknown yet. Here, we presented the first transcriptome analysis of *E. grandis* genotypes infected by *A. psidii*. Variant calling analysis identified a set of SNPs as non-synonymous substitutions in transcripts encoded by genes located at region of *Ppr1* locus and in other transcripts encoded by resistance and defense-related genes at other genomic regions. Some of those SNPs were identified in LRR kinase protein domain. Previous studies have demonstrated that even a single amino acid substitution alters the effectively of recognition by the LRR receptor (Warren et al. 1998; Tao et al. 2000; Halterman and Wise 2004; Stirnweis et al. 2014). Therefore, those SNPs are candidate markers to distinguish resistant from susceptible genotypes and these genes are possibly involved in resistance phenotype.

CONCLUSIONS

The analysis of transcriptome profiles using RNA-Seq showed that an *E. grandis* genotype resistant to *A. psidii* infection constitutively overexpress a set of genes related to resistance when compared to a susceptible genotype. Such genes are mainly involved in pathogen recognition, signaling, and photosynthesis, which represent a crucial process to an effective resistance response. At 24 hai, the susceptible genotype was able to activate defense response, however, not enough to stop the fungal infection. The DEGs identified in our study will be particularly useful as targets for investigating the molecular mechanisms involved *E. grandis* resistance to rust (*A. psidii*). Our findings also have the potential to be used as candidate molecular markers for resistance to myrtle rust, through their evaluation in populations of contrasting *Eucalyptus* genotypes.

Acknowledgments The authors thank to Clonar Resistência a Doenças Florestais for providing the plant material and the inoculation facilities. We also thanks to Plant and Food Research for providing all bioinformatics facilities and Núcleo de Análise de Biomoléculas (NuBioMol) for supporting data analysis. NuBioMol is supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), CAPES, CNPq, Financiadora de Estudos e Projetos (Finep) and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).

Author contributions All the authors conceived and designed the experiments; S.A.S. conducted the experiments; S.A.S. and P.M.P.V. performed the data analysis; S.A.S. wrote the manuscript text; all the authors revised and contributed to the final version of the manuscript; M.D.T. and A.C.A. supervised the research.

Funding This work was supported by CNPq (Conselho Nacional de Desenvolvimento Cientıfico e Tecnologico), CAPES (Coordenação de Aperfeicoamento de Pessoal de Nivel Superior), and Suzano S.A.

Data availability The sequencing raw data from 12 RNA-Seq libraries were deposited on the Sequence Read Archive from NCBI under SRA accession: PRJNA588626. Data are available here: https://www.ncbi.nlm.nih.gov/sra/PRJNA588626 and will be made publicly accessible after the publication of the manuscript. Additionally, the processed data including all supplementary files are temporally available here: https://figshare.com/s/8caab8398ff2002df311; and will be publish under the following DOI: https://dx.doi.org/10.6084/m9.figshare.12061626.

Table 1 Summary of sequencing and mapping of RNA-Seq data from *Eucalyptus grandis* genotypes during infection by *Austropuccinia psidii*, the causal agent of eucalyptus rust

Treatment	S-NI			S-I			R-NI			R-I		
Replicate	1	2	3	1	2	3	1	2	3	1	2	3
Raw reads	30,341,745	28,732,272	33,078,941	26,429,273	29,835,907	28,067,978	28,123,920	27,631,847	29,979,049	25,413,710	28,910,093	26,876,501
Clean reads[a]	26,416,779	25,292,197	28,944,372	23,360,786	26,280,504	23,900,632	24,972,655	24,215,016	26,176,938	22,301,332	25,509,393	23,198,646
Clean bases[b] (Gbp)	2.67	2.55	2.92	2.36	2.65	2.41	2.52	2.45	2.64	2.25	2.58	2.34
Reads uniquely mapped	24,183,894 (91.55%)	23,103,488 (91.35%)	26,570,514 (91.80%)	21,432,765 (91.75%)	24,050,897 (91.52%)	21,843,933 (91.39%)	23,133,075 (92.63%)	22,377,727 (92.41%)	24,122,469 (92.15%)	20,556,128 (92.17%)	23,492,634 (92.09%)	21,325,693 (91.93%)
Q30[c] (%)	94.33	93.87	94.33	94.73	94.47	94.61	94.52	94.79	94.06	94.48	94.70	93.66
GC content (%)	50.50	50.50	50.50	50.50	50.50	50.50	50.50	50.50	50.50	51.00	50.50	50.50

[a] Clean reads: Reads from sequencing after filtering low-quality and reads from rRNA. [b] Clean bases: The number of clean reads is multiplied by the length and converted to Giga base pairs (Gbp). [c] Q30: Average of the percentage of bases from both paired-end raw reads with a Phred value >30. **Note:** R-NI = resistant genotype non-infected; S-NI = susceptible genotype non-infected; R-I = resistant genotype infected; S-I = susceptible genotype infected.

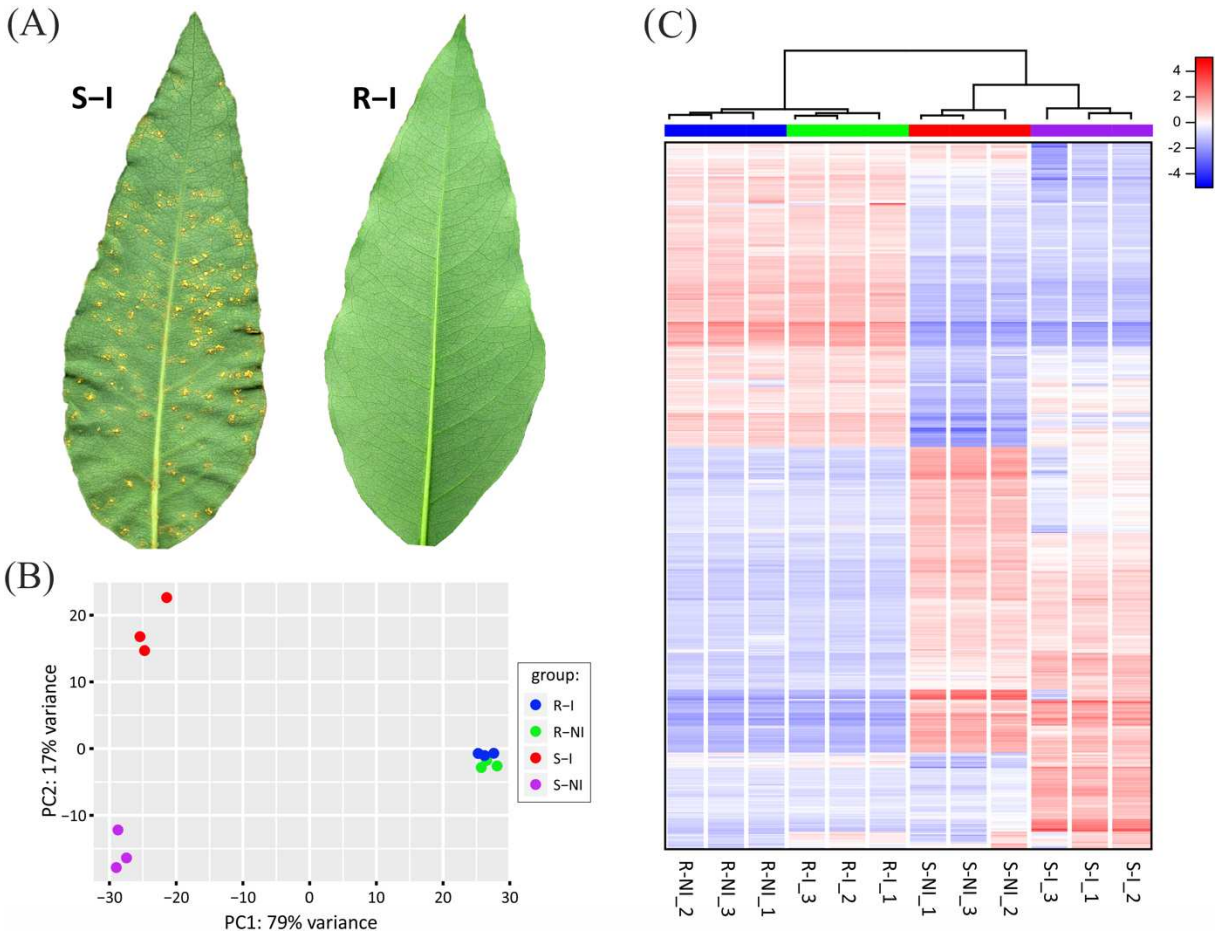

Fig. 1 Symptoms of rust infection and gene clustering. **a** Leaf of susceptible *Eucalyptus grandis* genotype showing uredinia containing urediniospores of *Austropuccinia psidii* at eight days after inoculation, while resistant genotype remains without any symptoms. **b** PCA (principal components analysis) plot using rlog-transformed values from reads count per gene from each RNA-Seq library. **c** Heat map plot of relative rlog-transformed values of the 1,000 most variable genes across samples. The subset containing the most variable genes can be found in Supplementary File 8: Table S13. **Note:** R-NI = resistant genotype non-infected; S-NI = susceptible genotype non-infected; R-I = resistant genotype infected; S-I = susceptible genotype infected.

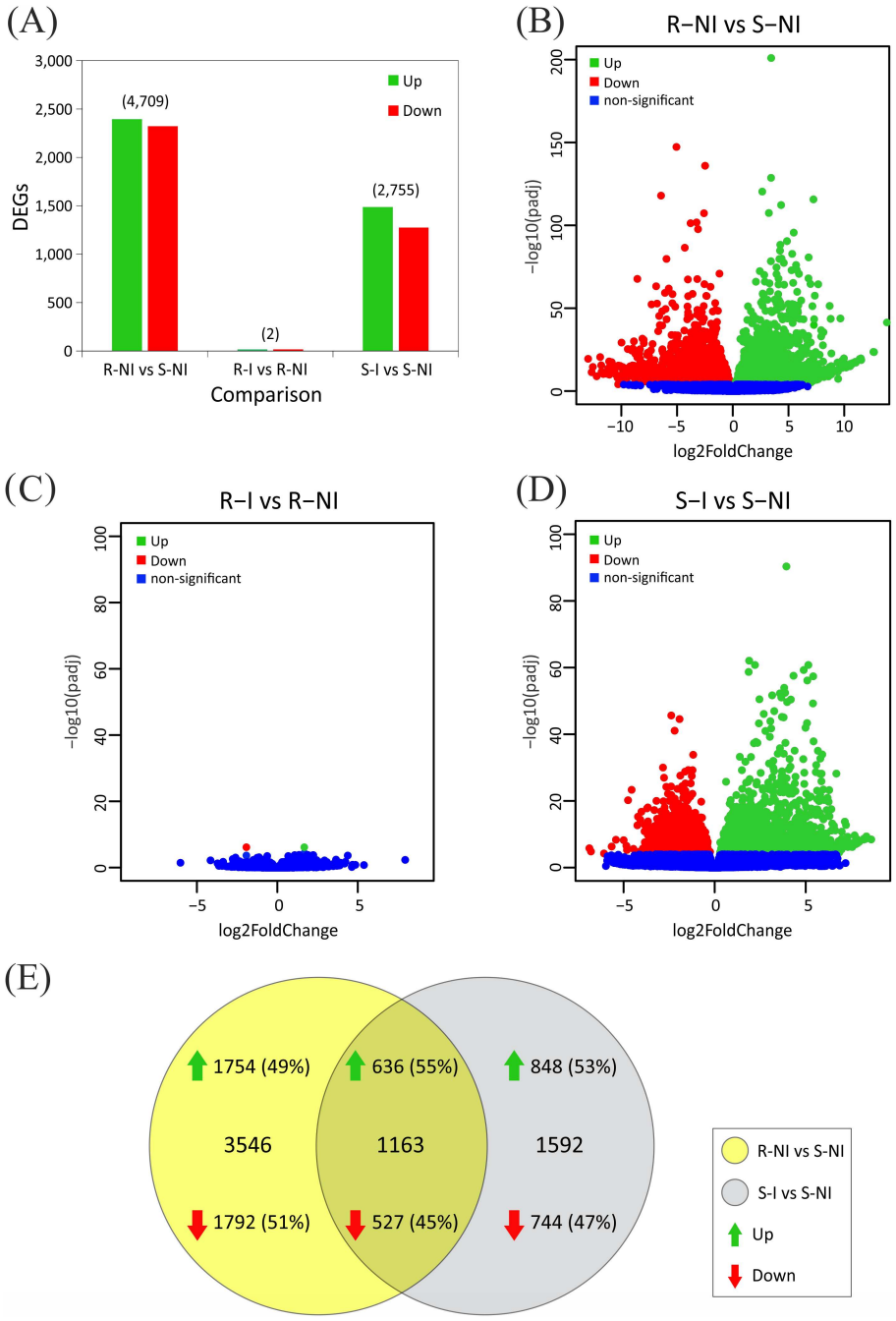

Fig. 2 Differentially expressed genes (DEGs) of resistant and susceptible *Eucalyptus grandis* genotypes at 24 hours after inoculation with *Austropuccinia psidii*. **a** Number of significant (padj < 0.0001) DEGs in each comparison. **b-d** Volcano plots of DEGs from R-NI vs S-NI, R-I vs R-NI, and S-I vs S-NI, respectively. **e** Venn diagram showing unique and shared DEGs. **Note:** R-NI = resistant genotype non-infected; S-NI = susceptible genotype non-infected; R-I = resistant genotype infected; S-I = susceptible genotype infected.

Table 2 GO significantly enriched analysis of the up-regulated DEGs of *Eucalyptus grandis* genotypes at 24 hours after inoculation with *Austropuccinia psidii*

	Category[a]	Term	Description	Gene count	padj[b]
R-NI vs S-NI	BP	GO:0008152	Metabolic process	31	2.4E-03
	BP	GO:0006629	Lipid metabolic process	9	3.5E-02
	BP	GO:0018298	Protein-chromophore linkage	8	1.7E-04
	BP	GO:0007165	Signal transduction	8	3.9E-02
	BP	GO:0009765	Photosynthesis, light harvesting	7	5.5E-05
	BP	GO:2000031	Regulation of salicylic acid mediated signaling pathway	3	2.3E-02
	BP	GO:0071446	Cellular response to salicylic acid stimulus	3	2.3E-02
	CC	GO:0016021	Integral component of membrane	189	8.3E-04
	CC	GO:0009522	Photosystem I	8	2.1E-04
	CC	GO:0009523	Photosystem II	7	4.3E-04
	CC	GO:0009535	Chloroplast thylakoid membrane	7	6.9E-03
	MF	GO:0004672	Protein kinase activity	29	9.0E-03
	MF	GO:0005506	Iron ion binding	24	3.8E-02
	MF	GO:0003824	Catalytic activity	20	4.8E-02
	MF	GO:0043531	ADP binding	13	2.4E-04
	MF	GO:0016168	Chlorophyll binding	8	2.0E-04
	MF	GO:0003993	Acid phosphatase activity	6	2.8E-02
	MF	GO:0005381	Iron ion transmembrane transporter activity	3	2.8E-02
S-I vs S-NI	BP	GO:0008152	Metabolic process	37	2.3E-08
	BP	GO:0006952	Defense response	11	6.5E-03
	CC	GO:0016021	Integral component of membrane	121	1.2E-02
	MF	GO:0020037	Heme binding	21	2.2E-02
	MF	GO:0016758	Transferase activity, transferring hexosyl groups	19	2.1E-04
	MF	GO:0004497	Monooxygenase activity	16	2.3E-02
	MF	GO:0005509	Calcium ion binding	13	3.5E-02
	MF	GO:0043531	ADP binding	11	2.2E-04
	MF	GO:0016597	Amino acid binding	6	1.0E-02
	MF	GO:0004540	Ribonuclease activity	3	4.7E-02

[a] Gene Ontology (GO) terms in biological process (BP), cellular component (CC), and molecular function (MF) categories. [b] P values were adjusted (padj) using the Hochberg and Benjamini tests and GO terms with a padj ≤ 0.05 and gene count ≥ 2 were considered as significantly enriched. **Note:** R-NI = resistant genotype non-infected; S-NI = susceptible genotype non-infected; R-I = resistant genotype infected; S-I = susceptible genotype infected.

Table 3 GO significantly enriched analysis of the down-regulated DEGs of *Eucalyptus grandis* genotypes at 24 hours after inoculation with *Austropuccinia psidii*

	Category[a]	Term	Description	Gene count	padj[b]
R-NI vs S-NI	BP	GO:0008152	Metabolic process	31	7.2E-03
	CC	GO:0005737	Cytoplasm	20	6.2E-03
	MF	GO:0016758	Transferase activity, transferring hexosyl groups	17	2.2E-02
	MF	GO:0016279	Protein-lysine N-methyltransferase activity	3	7.9E-03
	MF	GO:0016810	Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	3	4.8E-02
S-I vs S-NI	CC	GO:0016020	Membrane	7	2.5E-02
	MF	GO:0008270	Zinc ion binding	28	4.4E-02
	MF	GO:0016491	Oxidoreductase activity	25	3.8E-04
	MF	GO:0004672	Protein kinase activity	17	7.6E-03
	MF	GO:0043531	ADP binding	6	3.7E-02

[a] Gene Ontology (GO) terms in biological process (BP), cellular component (CC), and molecular function (MF) categories. [b] P values were adjusted (padj) using the Hochberg and Benjamini tests and GO terms with a padj ≤ 0.05 and gene count ≥ 2 were considered as significantly enriched. **Note:** R-NI = resistant genotype non-infected; S-NI = susceptible genotype non-infected; R-I = resistant genotype infected; S-I = susceptible genotype infected.

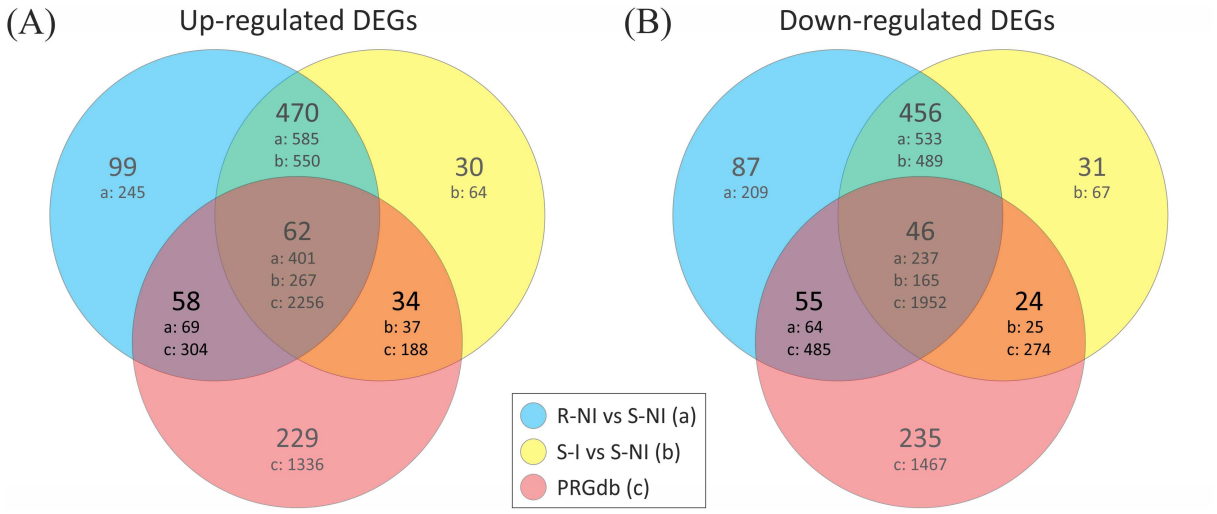

Fig. 3 OrthoMCL clustering analysis of differentially expressed genes (DEGs) of *Eucalyptus grandis* genotypes at 24 hours after inoculation with *Austropuccinia psidii* with *E. grandis* resistance-related proteins from Plant Resistance Genes database (PRGdb). **a-b** Venn diagrams showing unique and shared orthologous clusters and number of proteins among DEGs and PRGdb. **Note:** R-NI = resistant genotype non-infected; S-NI = susceptible genotype non-infected; R-I = resistant genotype infected; S-I = susceptible genotype infected.

Table 4 Constitutive overexpressed resistance-related proteins in the resistant *Eucalyptus grandis* genotype comparing to the susceptible

OrthoMCL cluster	Protein ID[a]	Protein name	Log_2FC[b]
1	XP_018720519.1	L-type lectin-domain containing receptor kinase IV.1	6,17
	XP_010037812.1	L-type lectin-domain containing receptor kinase IV.1	4,91
	XP_018721127.1	L-type lectin-domain containing receptor kinase IV.1	2,19
	XP_010036273.1	Lectin-domain containing receptor kinase VI.4	1,54
2	XP_018721785.1	Plant intracellular Ras-group-related LRR protein 4-like	6,81
	XP_018721791.1	Disease resistance protein LAZ5	6,32
	XP_018729851.1	TMV resistance protein N isoform X2	4,25
3	XP_010051086.1	Protein MKS1	4,50
	XP_018722534.1	Uncharacterized protein LOC108956903	2,13
	XP_018726611.1	Uncharacterized protein LOC104437579 isoform X1	1,04
4	XP_018727105.1	Proline-rich receptor-like protein kinase PERK1 isoform X3	7,24
	XP_018727098.1	Proline-rich receptor-like protein kinase PERK3 isoform X2	0,87
5	XP_018729839.1	Probable disease resistance protein At4g19060	4,02
	XP_010055455.1	Probable disease resistance protein At5g45490	1,74
6	XP_010036338.1	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1	2,59
	XP_018722557.1	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1	1,76
7	XP_018726992.1	TMV resistance protein N isoform X2	1,45
	XP_018722228.1	TMV resistance protein N	1,12
8	XP_018728602.1	Disease resistance protein RPM1-like	10,67
9	XP_010046438.1	Lysm domain receptor-like kinase 4	10,26
10	XP_010024187.1	Polygalacturonase inhibitor-like	7,00
11	XP_010051304.1	Toll/interleukin-1 receptor-like protein	6,65
12	XP_018728742.1	Disease resistance protein RPM1	6,59
13	XP_018730739.1	Leucine-rich repeat-containing protein 40-like	6,52
14	XP_018722313.1	Protein kinase APK1B, chloroplastic-like	6,51
15	XP_018730912.1	LRR receptor-like serine/threonine-protein kinase GSO2 isoform X2	6,28
16	XP_010058955.1	Stress response protein nst1 isoform X1	5,11
17	XP_018719939.1	Disease resistance protein At4g27190 isoform X1	4,96
18	XP_018723657.1	Uncharacterized protein LOC104426303	4,80
19	XP_018727691.1	Uncharacterized protein LOC104441500 isoform X1	4,70
20	XP_018726959.1	Proline-rich receptor-like protein kinase PERK3	4,57
21	XP_018731295.1	Probable leucine-rich repeat receptor-like serine/threonine-protein kinase At3g14840	4,30
22	XP_018721940.1	Disease resistance protein LAZ5-like	4,11
23	XP_018731627.1	Mitogen-activated protein kinase 7	3,78
24	XP_018728922.1	Toll/interleukin-1 receptor-like protein	3,74
25	XP_010057722.2	Putative disease resistance protein At4g19050	3,64
26	XP_018720140.1	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 isoform X2	3,37
27	XP_018727431.1	Uncharacterized protein LOC104441110 isoform X1	3,19
28	XP_018725686.1	Uncharacterized protein LOC104439442	2,99
29	XP_018716780.1	Serine/threonine-protein kinase-like protein CCR4 isoform X1	2,77
30	XP_010034802.1	Pollen receptor-like kinase 3	2,75
31	XP_018720614.1	Rust resistance kinase Lr10-like isoform X1	2,58
32	XP_018719055.1	Disease resistance protein TAO1	2,31
33	XP_010050650.2	CBL-interacting serine/threonine-protein kinase 24	2,29
34	XP_010030026.1	ABC transporter G family member 11	1,84
35	XP_010036366.1	CDPK-related kinase 1	1,60
36	XP_010040027.1	Disease resistance protein RPP13 isoform X2	1,60
37	XP_010051382.1	Leucine-rich repeat protein soc-2 homolog	1,58
38	XP_010056223.1	Disease resistance protein RPM1	1,52
39	XP_010055829.1	Probable inactive receptor kinase At4g23740	1,51
40	XP_010049433.2	Proline-rich receptor-like protein kinase PERK3 isoform X1	1,46
41	XP_010036335.1	Somatic embryogenesis receptor kinase 1	1,42
42	XP_010060769.1	Receptor-like serine/threonine-protein kinase At1g78530	1,35
43	XP_010045511.1	Leucine-rich repeat receptor-like serine/threonine/tyrosine-protein kinase SOBIR1	1,26
44	XP_010041713.1	Protein kinase and PP2C-like domain-containing protein	1,21
45	XP_010051912.1	Probable LRR receptor-like serine/threonine-protein kinase At4g20940	1,19
46	XP_010063079.2	Probable serine/threonine-protein kinase WNK11 isoform X1	1,18
47	XP_010062002.1	Calcium-dependent protein kinase 11 isoform X2	1,16
48	XP_010052291.1	Putative leucine-rich repeat receptor-like serine/threonine-protein kinase At2g14440	1,08
49	XP_010052182.1	CBL-interacting serine/threonine-protein kinase 12	0,90
50	XP_018730706.1	Probable receptor-like protein kinase At5g47070	0,86
51	XP_018718480.1	Phototropin-2	0,86
52	XP_018717263.1	Serine/threonine-protein kinase HT1	0,82
53	XP_010042572.1	CBL-interacting protein kinase 2	0,79
54	XP_018727101.1	Proline-rich receptor-like protein kinase PERK3	0,73
55	XP_010056802.1	PTI1-like tyrosine-protein kinase 1 isoform X1	0,59
56	XP_010032521.1	SNF1-related protein kinase catalytic subunit alpha KIN10 isoform X1	0,47
57	XP_010033439.1	Mitogen-activated protein kinase 7	0,43
58	XP_010024632.1	Serine/threonine-protein kinase CDL1	0,37

[a] NCBI nomenclature. [b] Log_2 of fold-change (FC) value from R-NI vs S-NI comparison. **Note:** R-NI = resistant genotype non-infected; S-NI = susceptible genotype non-infected.

Table 5 Resistance-related proteins with expression level affected positively or negatively by *Austropuccinia psidii* infection in the susceptible *Eucalyptus grandis* genotype

	OrthoMCL cluster	Protein ID[a]	Protein name	Log_2FC[b]
Up-regulated DEGs	1	XP_018732206.1	Serine/threonine-protein kinase/endoribonuclease IRE1a	4,55
		XP_010060235.2	Serine/threonine-protein kinase/endoribonuclease IRE1a isoform X2	4,02
	2	XP_018726292.1	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5	2,14
		XP_010048193.1	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5	1,12
	3	XP_010049411.1	Serine/threonine-protein kinase At5g01020	1,77
		XP_010049350.1	Serine/threonine-protein kinase At5g01020	1,49
	4	XP_018715809.1	Toll/interleukin-1 receptor-like protein	6,97
	5	XP_010026327.1	TMV resistance protein N-like	6,56
	6	XP_010057335.1	Disease resistance protein LAZ5	6,26
	7	XP_018730120.1	Toll/interleukin-1 receptor-like protein	5,73
	8	XP_010046528.1	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 2.7	5,58
	9	XP_010053967.1	Probably inactive leucine-rich repeat receptor-like protein kinase At5g48380	5,34
	10	XP_010053969.1	Probably inactive leucine-rich repeat receptor-like protein kinase At5g48380	5,19
	11	XP_018721024.1	LRR receptor-like serine/threonine-protein kinase GSO2 isoform X2	3,72
	12	XP_018726962.1	Proline-rich receptor-like protein kinase PERK3 isoform X2	3,11
	13	XP_010063715.1	MDIS1-interacting receptor like kinase 2	2,87
	14	XP_018718033.1	Serine/threonine-protein kinase CDL1	2,65
	15	XP_010051745.1	G-type lectin S-receptor-like serine/threonine-protein kinase LECRK2	2,56
	16	XP_018716214.1	Plant intracellular Ras-group-related LRR protein 5-like	2,33
	17	XP_010027357.1	G-type lectin S-receptor-like serine/threonine-protein kinase LECRK4	2,26
	18	XP_010070641.2	Putative disease resistance RPP13-like protein 2 isoform X1	2,16
	19	XP_010045165.2	Putative inactive disease susceptibility protein LOV1	2,04
	20	XP_010064986.1	Mitogen-activated protein kinase kinase kinase 1 isoform X1	1,86
	21	XP_010054953.1	Histidine--trna ligase, cytoplasmic	1,63
	22	XP_010045108.1	Protein kinase 2B, chloroplastic	1,53
	23	XP_010042885.2	U-box domain-containing protein 33-like	1,30
	24	XP_010063769.1	leucine-rich repeat receptor-like tyrosine-protein kinase PXC3	0,98
	25	XP_018715250.1	Probable inactive receptor-like protein kinase At3g56050	0,90
	26	XP_010066552.1	G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	0,84
	27	XP_010063879.2	Serine decarboxylase	0,66
	28	XP_010062525.1	U-box domain-containing protein 35	0,66
	29	XP_010066742.1	Leucine-rich repeat receptor-like serine/threonine-protein kinase At2g14440	0,64
	30	XP_010047911.1	Cyclin-dependent kinase F-1	0,57
	31	XP_010043234.1	Probable receptor-like protein kinase At3g55450	0,57
	32	XP_010037082.1	Probably inactive leucine-rich repeat receptor-like protein kinase At5g48380	0,45
	33	XP_010047526.1	LOW QUALITY PROTEIN: mitogen-activated protein kinase kinase 2-like	0,39
	34	XP_010030065.1	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1	0,39
Down-regulated DEGs	1	XP_010029904.1	Jacalin-related lectin 4	-2,96
		XP_010029918.1	Myrosinase-binding protein 2 isoform X1	-1,43
	2	XP_018719919.1	Disease resistance protein RLM3	-3,34
	3	XP_010042765.1	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 2.1	-3,10
	4	XP_018716675.1	Protein scribble homolog	-3,01
	5	XP_010042977.1	Plant intracellular Ras-group-related LRR protein 4-like	-2,96
	6	XP_018729655.1	L-type lectin-domain containing receptor kinase IX.1 isoform X2	-2,47
	7	XP_010039878.2	LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-like 1.1	-2,35
	8	XP_010064640.2	Wall-associated receptor kinase 3	-2,17
	9	XP_018724713.1	TMV resistance protein N-like	-2,09
	10	XP_018717720.1	Probable receptor-like protein kinase At5g24010	-1,86
	11	XP_018726301.1	Probable L-type lectin-domain containing receptor kinase S.5 isoform X1	-1,74
	12	XP_010060724.1	TMV resistance protein N isoform X1	-1,53
	13	XP_010024993.1	Probable serine/threonine-protein kinase NAK isoform X1	-1,53
	14	XP_018722094.1	Probable WRKY transcription factor 19	-1,45
	15	XP_018725380.1	Receptor-like protein kinase 5 isoform X2	-1,43
	16	XP_010046666.1	Probable leucine-rich repeat receptor-like protein kinase At1g35710 isoform X1	-1,38
	17	XP_018724419.1	LOW QUALITY PROTEIN: lysm domain receptor-like kinase 3	-0,91
	18	XP_018719366.1	Disease resistance RPP13-like protein 4 isoform X2	-0,81
	19	XP_010059914.1	Probable serine/threonine-protein kinase At1g01540	-0,67
	20	XP_010070081.1	Shaggy-related protein kinase alpha	-0,65
	21	XP_010042581.1	Probable inactive receptor-like protein kinase At3g56050, partial	-0,65
	22	XP_010057375.1	Probable LRR receptor-like serine/threonine-protein kinase At1g05700	-0,58
	23	XP_018725483.1	Uncharacterized protein LOC104438755	-0,49
	24	XP_018719688.1	Probable protein phosphatase 2C 21 isoform X3	-0,49

[a] NCBI nomenclature. [b] Log_2 of fold-change (FC) value from S-I vs S-NI comparison. **Note:** S-NI = susceptible genotype non-infected; S-I = susceptible genotype infected.

Table 6 Summary of Single Nucleotide Polymorphisms (SNP) identified in RNA-Seq data of two *Eucalyptus grandis* genotypes resistant and susceptible to rust (*Austropuccinia psidii*)

SNP category	Susceptible genotype[a]	Resistant genotype[a]	Synonymous variants	Non-synonymous variants	Other variants[b]	Total of polymorphic loci
exclusive	0/1	0/0	51413	37894	127155	216462
exclusive	1/1	0/0	1253	1018	3934	6205
exclusive	0/0	0/1	12717	10728	36685	60130
exclusive	0/0	1/1	1440	1100	4348	6888
shared	0/1	1/1	6215	3953	15290	25458
shared	1/1	0/1	2507	1833	6924	11264

[a] 0/0 is identical to the *E. grandis* reference genome; 0/1 and 1/1 contain one and two different alleles, respectively, compared to the *E. grandis* reference genome. [b] Other variants include those located at regulatory regions and introns.

Table 7 Heterozygous and homozygous non-synonymous substitutions identified exclusively in the *Eucalyptus grandis* genotype resistant to rust (*Austropuccinia psidii*)

Transcript ID	Protein ID	Chromosome	Position	Reference genome	Allele[a]	Susceptible[b]	Resistant[c]	Transcript position	CDS position	Protein position	Amino acids	Codons
XM_018870217.1	XP_018725762.1	NW_010092440.1	50564259	A	C	0/0[b]	0/1	3067	2869	957	Y/D	Tac/Gac
		NW_010092440.1	50564284	T	A	0/0	1/1	3042	2844	948	R/S	agA/agT
XM_010050616.2	XP_010048918.1	NW_010092440.1	52278292	T	C	0/0	0/1	1224	1065	355	I/M	atA/atG
		NW_010092440.1	52280572	C	A	0/0	0/1	366	207	69	K/N	aaG/aaT
		NW_010092440.1	52280902	C	T	0/0	0/1	199	40	14	V/I	Gtt/Att
XM_018870229.1	XP_018725774.1	NW_010092440.1	52648034	T	C	0/0	1/1	3966	3149	1050	Q/R	cAg/cGg
XM_010050814.2	XP_010049116.1	NW_010092440.1	55117514	C	G	0/0	0/1	2936	2684	895	G/A	gGc/gCc
		NW_010092440.1	55117518	G	A	0/0	0/1	2932	2680	894	L/F	Ctc/Ttc
		NW_010092440.1	55117530	C	T	0/0	0/1	2920	2668	890	A/T	Gcc/Acc
		NW_010092440.1	55118781	C	T	0/0	0/1	2092	1840	614	D/N	Gac/Aac
		NW_010092440.1	55119669	T	G	0/0	0/1	1204	952	318	S/R	Agc/Cgc

[a] Single Nucleotide Polymorphisms identified in RNA-Seq data. [b] 0/0 is identical to the *E. grandis* reference genome. [c] 0/1 and 1/1 contain one and two different alleles, respectively, compared to the *E. grandis* reference genome.

REFERENCES

Afzal AJ, Wood AJ, Lightfoot DA (2008) Plant Receptor-Like Serine Threonine Kinases: Roles in Signaling and Plant Defense. Mol Plant-Microbe Interact 21:507–517. https://doi.org/10.1094/MPMI-21-5-0507

Ahmed MB, Santos KCG dos, Sanchez IB, Petre B, Lorrain C, Plourde MB, Duplessis S, Desgagné-Penix I, Germain H (2018) A rust fungal effector binds plant DNA and modulates transcription. Sci Rep 8:14718. https://doi.org/10.1038/s41598-018-32825-0

Alfenas AC, Zauza EA. V, Assis TF (2003) First record of *Puccinia psidii* on *Eucalyptus globulus* and *E. viminalis* in Brazil. Australas Plant Pathol 32:325–326. https://doi.org/10.1071/AP03021

Alfenas AC, Zauza EAV, Mafia RG, Assis TF (2009) Clonagem e doenças do eucalipo, 1st edn. Editora UFV, Viçosa, MG state, Brazil

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Alves AA, Rosado CCG, Faria DA, Guimarães LM da S, Lau D, Brommonschenkel SH, Grattapaglia D, Alfenas AC (2012) Genetic mapping provides evidence for the role of additive and non-additive QTLs in the response of inter-specific hybrids of *Eucalyptus* to *Puccinia psidii* rust infection. Euphytica 183:27–38. https://doi.org/10.1007/s10681-011-0455-5

Amorim L, Santos R, Neto J, Guida-Santos M, Crovella S, Benko-Iseppon A (2017) Transcription Factors Involved in Plant Resistance to Pathogens. Curr Protein Pept Sci 18:335–351. https://doi.org/10.2174/1389203717666160619185308

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 18 Jul 2019

Bartholomé J, Mandrou E, Mabiala A, Jenkins J, Nabihoudine I, Klopp C, Schmutz J, Plomion C, Gion J-M (2015) High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. New Phytol 206:1283–1296. https://doi.org/10.1111/nph.13150

Beenken L (2017) Austropuccinia: a new genus name for the myrtle rust *Puccinia psidii* placed within the redefined family Sphaerophragmiaceae (Pucciniales). Phytotaxa 297:53–61. https://doi.org/10.11646/phytotaxa.297.1.5

Bigeard J, Colcombet J, Hirt H (2015) Signaling Mechanisms in Pattern-Triggered Immunity (PTI). Mol Plant 8:521–539. https://doi.org/10.1016/J.MOLP.2014.12.022

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Budak H, Su S, Ergen N (2006) Revealing constitutively expressed resistance genes in Agrostis species using PCR-based motif-directed RNA fingerprinting. Genet Res 88:165–175. https://doi.org/10.1017/S0016672307008518

Butler JB, Freeman JS, Vaillancourt RE, Potts BM, Glen M, Lee DJ, Pegg GS (2016) Evidence for different QTL underlying the immune and hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*. Tree Genet Genomes 12:39. https://doi.org/10.1007/s11295-016-0987-x

Butler JB, Potts BM, Vaillancourt RE, Lee DJ, Pegg GS, Freeman JS (2019) Independent QTL underlie resistance to the native pathogen Quambalaria pitereka and the exotic pathogen *Austropuccinia*

psidii in *Corymbia*. Tree Genet Genomes 15:72. https://doi.org/10.1007/s11295-019-1378-x

Chakraborty S, Britton M, Martínez-García PJ, Dandekar AM (2016) Deep RNA-Seq profile reveals biodiversity, plant–microbe interactions and a large family of NBS-LRR resistance genes in walnut (*Juglans regia*) tissues. AMB Express 6:12. https://doi.org/10.1186/s13568-016-0182-3

Chakraborty S, Nguyen B, Wasti SD, Xu G (2019) Plant Leucine-Rich Repeat Receptor Kinase (LRR-RK): Structure, Ligand Perception, and Activation Mechanism. Molecules 24:3081. https://doi.org/10.3390/molecules24173081

Christie N, Tobias PA, Naidoo S, Külheim C (2016) The Eucalyptus grandis NBS-LRR gene family: Physical clustering and expression hotspots. Front Plant Sci 6:1–16. https://doi.org/10.3389/fpls.2015.01238

Cui X, Yan Q, Gan S, Xue D, Wang H, Xing H, Zhao J, Guo N (2019) GmWRKY40, a member of the WRKY transcription factor genes identified from Glycine max L., enhanced the resistance to *Phytophthora sojae*. BMC Plant Biol 19:1–15. https://doi.org/10.1186/s12870-019-2132-0

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature 411:826–833. https://doi.org/10.1038/35081161

De Lorenzo G, D'Ovidio R, Cervone F (2001) The Role of Polygalacturonase-Inhibiting Proteins (PGIPs) in Defense Against Pathogenic Fungi. Annu Rev Phytopathol 39:313–335. https://doi.org/10.1146/annurev.phyto.39.1.313

DeYoung BJ, Qi D, Kim S-H, Burke TP, Innes RW (2012) Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. Cell Microbiol 14:1071–1084. https://doi.org/10.1111/j.1462-5822.2012.01779.x

Di C, Zhang M, Xu S, Cheng T, An L (2006) Role of polygalacturonase-inhibiting protein in plant defense. Crit. Rev. Microbiol. 32:91–100. https://doi.org/10.1080/10408410600709834

Ding P, Redkar A (2018) Pathogens Suppress Host Transcription Factors for Rampant Proliferation. Trends Plant Sci 23:950–953. https://doi.org/10.1016/j.tplants.2018.08.010

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Dobon A, Bunting DCE, Cabrera-Quio LE, Uauy C, Saunders DGO (2016) The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. BMC Genomics 17:380. https://doi.org/10.1186/s12864-016-2684-4

Dorey S, Baillieul F, Pierrel M-A, Saindrenan P, Fritig B, Kauffmann S (1997) Spatial and Temporal Induction of Cell Death, Defense Genes, and Accumulation of Salicylic Acid in Tobacco Leaves Reacting Hypersensitively to a Fungal Glycoprotein Elicitor. Mol Plant-Microbe Interact 10:646–655. https://doi.org/10.1094/MPMI.1997.10.5.646

Durner J, Shah J, Klessig DF (1997) Salicylic acid and disease resistance in plants. Trends Plant Sci 2:266–274. https://doi.org/10.1016/S1360-1385(97)86349-2

Eulgem T, Somssich IE (2007) Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol 10:366–371. https://doi.org/10.1016/j.pbi.2007.04.020

Fernández-Marín B, Gulías J, Figueroa CM, Iñiguez C, Clemente-Moreno MJ, Nunes-Nesi A, Fernie AR, Cavieres LA, Bravo LA, García-Plazaola JI, Gago J (2020) How do vascular plants perform photosynthesis in extreme environments? An integrative ecophysiological and biochemical story. Plant J 101:979–1000. https://doi.org/10.1111/tpj.14694

Freeman BC, Beattie GA (2008) An Overview of Plant Defenses against Pathogens and Herbivores. Plant Heal Instr. https://doi.org/10.1094/PHI-I-2008-0226-01

Gao J, Bi W, Li H, Wu J, Yu X, Liu D, Wang X (2018) WRKY transcription factors associated with NPR1-mediated acquired resistance in barley are potential resources to improve wheat resistance to *Puccinia triticina*. Front Plant Sci 871:1–14. https://doi.org/10.3389/fpls.2018.01486

Garnica DP, Nemri A, Upadhyaya NM, Rathjen JP, Dodds PN (2014) The Ins and Outs of Rust Haustoria. PLoS Pathog 10:e1004329. https://doi.org/10.1371/journal.ppat.1004329

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv Prepr arXiv12073907 Jul:1–9

Glazebrook J (2005) Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. Annu Rev Phytopathol 43:205–227. https://doi.org/10.1146/annurev.phyto.43.040204.135923

Gou X, He K, Yang H, Yuan T, Lin H, Clouse SD, Li J (2010) Genome-wide cloning and sequence analysis of leucine-rich repeat receptor-like protein kinase genes in Arabidopsis thaliana. BMC Genomics 11:19. https://doi.org/10.1186/1471-2164-11-19

Grattapaglia D, Kirst M (2008) *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytol 179:911–929. https://doi.org/10.1111/j.1469-8137.2008.02503.x

Halterman DA, Wise RP (2004) A single-amino acid substitution in the sixth leucine-rich repeat of barley MLA6 and MLA13 alleviates dependence on RAR1 for disease resistance signaling. Plant J 38:215–226. https://doi.org/10.1111/j.1365-313X.2004.02032.x

Harwood C (2011) New introductions - doing it right. In: Walker (ed) Developing a Eucalypt Resource: Learning from Australia and elsewere. Wood Technology Research Centre, University of Canterbury, Christchurch, New Zealand, pp 125–136

Hayden KJ, Garbelotto M, Knaus BJ, Cronn RC, Rai H, Wright JW (2014) Dual RNA-seq of the plant pathogen *Phytophthora ramorum* and its tanoak host. Tree Genet Genomes 10:489–502. https://doi.org/10.1007/s11295-014-0698-0

Hématy K, Cherk C, Somerville S (2009) Host-pathogen warfare at the plant cell wall. Curr. Opin. Plant Biol. 12:406–413. https://doi.org/10.1016/j.pbi.2009.06.007

Hohmann U, Lau K, Hothorn M (2017) The Structural Basis of Ligand Perception and Signal Activation by Receptor Kinases. Annu Rev Plant Biol 68:109–137. https://doi.org/10.1146/annurev-arplant-042916-040957

Hsieh J-F, Chuah A, Patel HR, Sandhu KS, Foley WJ, Külheim C (2018) Transcriptome Profiling of *Melaleuca quinquenervia* Challenged by Myrtle Rust Reveals Differences in Defense Responses Among Resistant Individuals. Phytopathology 108:495–509. https://doi.org/10.1094/PHYTO-09-17-0307-R

Huang D, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 8:R183. https://doi.org/10.1186/gb-2007-8-9-r183

Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13. https://doi.org/10.1093/nar/gkn923

IBÁ – Indústria Brasileira de Árvores (2019) Report 2019. Sao Paulo, Brazil.

https://iba.org/datafiles/publicacoes/relatorios/iba-relatorioanual2019.pdf. Accessed 01 Nov 2019

Inoue H, Hayashi N, Matsushita A, Xinqiong L, Nakayama A, Sugano S, Jiang CJ, Takatsuji H (2013) Blast resistance of CC-NB-LRR protein Pb1 is mediated by WRKY45 through protein-protein interaction. Proc Natl Acad Sci U S A 110:9577–9582. https://doi.org/10.1073/pnas.1222155110

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444:323–329. https://doi.org/10.1038/nature05286

Junghans DT, Alfenas AC, Brommonschenkel SH, Oda S, Mello EJ, Grattapaglia D (2003) Resistance to rust (*Puccinia psidii* Winter) in Eucalyptus: mode of inheritance and mapping of a major gene with RAPD markers. Theor Appl Genet 108:175–180. https://doi.org/10.1007/s00122-003-1415-9

Kopylova E, Noé L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Kourelis J, van der Hoorn RAL (2018) Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. Plant Cell 30:285–299. https://doi.org/10.1105/tpc.17.00579

Kullan ARK, van Dyk MM, Hefer CA, Jones N, Kanzler A, Myburg AA (2012) Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of *Eucalyptus grandis* and *E. urophylla*. BMC Genet 13:1–12. https://doi.org/10.1186/1471-2156-13-60

Lai Z, Vinod K, Zheng Z, Fan B, Chen Z (2008) Roles of Arabidopsis WRKY3 and WRKY4 transcription factors in plant responses to pathogens. BMC Plant Biol 8:1–13. https://doi.org/10.1186/1471-2229-8-68

Li H, Wu J, Shang X, Geng M, Gao J, Zhao S, Yu X, Liu D, Kang Z, Wang X, Wang X (2020) WRKY transcription factors shared by BTH-induced resistance and NPR1-mediated acquired resistance improve broad-spectrum disease resistance in wheat. Mol Plant-Microbe Interact 33:433–443. https://doi.org/10.1094/MPMI-09-19-0257-R

Li L, Stoeckert CJ, Roos DS, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189. https://doi.org/10.1101/gr.1224503

Lima BM, Teixeira JE, Gazaffi R, Garcia AA, Grattapaglia D, Valle RK, Camargo LE (2011) Identification of a novel QTL contributing to rust resistance in Eucalyptus. BMC Proc 5:P32. https://doi.org/10.1186/1753-6561-5-S7-P32

Liu J-J, Sturrock RN, Sniezko RA, Williams H, Benton R, Zamany A (2015) Transcriptome analysis of the white pine blister rust pathogen Cronartium ribicola: de novo assembly, expression profiling, and identification of candidate effectors. BMC Genomics 16:678. https://doi.org/10.1186/s12864-015-1861-1

Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R (2015) Fungal Effectors and Plant Susceptibility. Annu Rev Plant Biol 66:513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Lorrain C, Gonçalves dos Santos KC, Germain H, Hecker A, Duplessis S (2019) Advances in understanding obligate biotrophy in rust fungi. New Phytol 222:1190–1206. https://doi.org/10.1111/nph.15641

Love MI, Anders S, Kim V, Huber W (2015) RNA-Seq workflow: gene-level exploratory analysis and differential expression. F1000Research 4:1070. https://doi.org/10.12688/f1000research.7035.1

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Lu Y, Yao J (2018) Chloroplasts at the Crossroad of Photosynthesis, Pathogen Infection and Plant Defense. Int J Mol Sci 19:3900. https://doi.org/10.3390/ijms19123900

Mamani EMC, Bueno NW, Faria DA, Guimarães LMS, Lau D, Alfenas AC, Grattapaglia D (2010) Positioning of the major locus for *Puccinia psidii* rust resistance (Ppr1) on the Eucalyptus reference map and its validation across unrelated pedigrees. Tree Genet Genomes 6:953–962. https://doi.org/10.1007/s11295-010-0304-z

Maschietto V, Lanubile A, Leonardis S De, Marocco A, Paciolla C (2016) Constitutive expression of pathogenesis-related proteins and antioxydant enzyme activities triggers maize resistance towards *Fusarium verticillioides*. J Plant Physiol 200:53–61. https://doi.org/10.1016/J.JPLPH.2016.06.006

Miranda AC, de Moraes MLT, Tambarussi EV, Furtado EL, Mori ES, da Silva PHM, Sebbenn AM (2013) Heritability for resistance to *Puccinia psidii* Winter rust in *Eucalyptus grandis* Hill ex Maiden in Southwestern Brazil. Tree Genet Genomes 9:321–329. https://doi.org/10.1007/s11295-012-0572-x

Monaghan J, Zipfel C (2012) Plant pattern recognition receptor complexes at the plasma membrane. Curr Opin Plant Biol 15:349–357. https://doi.org/10.1016/j.pbi.2012.05.006

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, Goodstein DM, Dubchak I, Poliakov A, Mizrachi E, Kullan ARK, Hussey SG, Pinard D, van der Merwe K, Singh P, van Jaarsveld I, Silva-Junior OB, Togawa RC, Pappas MR, Faria DA, Sansaloni CP, Petroli CD, Yang X, Ranjan P, Tschaplinski TJ, Ye C-Y, Li T, Sterck L, Vanneste K, Murat F, Soler M, Clemente HS, Saidi N, Cassan-Wang H, Dunand C, Hefer CA, Bornberg-Bauer E, Kersting AR, Vining K, Amarasinghe V, Ranik M, Naithani S, Elser J, Boyd AE, Liston A, Spatafora JW, Dharmwardhana P, Raja R, Sullivan C, Romanel E, Alves-Ferreira M, Külheim C, Foley W, Carocha V, Paiva J, Kudrna D, Brommonschenkel SH, Pasquali G, Byrne M, Rigault P, Tibbits J, Spokevicius A, Jones RC, Steane DA, Vaillancourt RE, Potts BM, Joubert F, Barry K, Pappas GJ, Strauss SH, Jaiswal P, Grima-Pettenati J, Salse J, Van de Peer Y, Rokhsar DS, Schmutz J (2014) The genome of *Eucalyptus grandis*. Nature 510:356–362. https://doi.org/10.1038/nature13308

Naidoo S, Külheim C, Zwart L, Mangwanda R, Oates CN, Visser EA, Wilken FE, Mamni TB, Myburg AA (2014) Uncovering the defence responses of eucalyptus to pests and pathogens in the genomics age. Tree Physiol 34:931–943. https://doi.org/10.1093/treephys/tpu075

Naoumkina MA, He X, Dixon RA (2008) Elicitor-induced transcription factors for metabolic reprogramming of secondary metabolism in *Medicago truncatula*. BMC Plant Biol 8:1–14. https://doi.org/10.1186/1471-2229-8-132

Nirmala J, Drader T, Lawrence PK, Yin C, Hulbert S, Steber CM, Steffenson BJ, Szabo LJ, von Wettstein D, Kleinhofs A (2011) Concerted action of two avirulent spore effectors activates Reaction to *Puccinia graminis* 1 (Rpg1)-mediated cereal stem rust resistance. Proc Natl Acad Sci 108:14676–14681. https://doi.org/10.1073/pnas.1111771108

Nobori T, Tsuda K (2019) The plant immune system in heterogeneous environments. Curr Opin Plant Biol 50:58–66. https://doi.org/10.1016/J.PBI.2019.02.003

Osuna-Cruz CM, Paytuvi-Gallart A, Di Donato A, Sundesha V, Andolfo G, Aiese Cigliano R, Sanseverino W, Ercolano MR (2018) PRGdb 3.0: a comprehensive platform for prediction and analysis of plant disease resistance genes. Nucleic Acids Res 46:D1197–D1201. https://doi.org/10.1093/nar/gkx1119

Padmanabhan M, Cournoyer P, Dinesh-Kumar SP (2009) The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. Cell Microbiol 11:191–198. https://doi.org/10.1111/j.1462-5822.2008.01260.x

Palm-Forster MAT, Eschen-Lippold L, Uhrig J, Scheel D, Lee J (2017) A novel family of proline/serine-rich proteins, which are phospho-targets of stress-related mitogen-activated protein kinases, differentially regulates growth and pathogen defense in Arabidopsis thaliana. Plant Mol Biol 95:123–140. https://doi.org/10.1007/s11103-017-0641-5

Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. Plant Physiol 150:1648–1655. https://doi.org/10.1104/pp.109.138990

Rosado TB, Tomaz RS, Ribeiro Junior MF, Rosado AM, Guimarães LM da S, Araújo EF de, Alfenas AC, Cruz CD (2010) Detection of QTL associated with rust resistance using IBD-based methodologies in exogamic *Eucalyptus* spp. populations. Crop Breed Appl Biotechnol 10:321–328. https://doi.org/10.1590/S1984-70332010000400006

Ruiz RAR, Alfenas AC, Ferreira FA, Vale FXR (1989) Influência da temperatura, do tempo de molhamento foliar, fotoperíodo e da intensidade de luz sobre a infecção de *Puccinia psidii* em eucalipto. Fitopatol Bras 14:55–61

Ryu HS, Han M, Lee SK, Cho J Il, Ryoo N, Heu S, Lee YH, Bhoo SH, Wang GL, Hahn TR, Jeon JS (2006) A comprehensive expression analysis of the WRKY gene superfamily in rice plants during defense response. Plant Cell Rep 25:836–847. https://doi.org/10.1007/s00299-006-0138-1

Santos MR, Guimarães LM da S, Resende MDV de, Rosse LN, Zamprogno KC, Alfenas AC (2014) Resistance of *Eucalyptus pellita* to rust (*Puccinia psidii*). Crop Breed Appl Biotechnol 14:244–250. https://doi.org/10.1590/1984-70332014v14n4a38

Santos SA dos, Tuffi-Santos LD, Tanaka FAO, Sant'Anna-Santos BF, Rodrigues F de Á, Alfenas AC (2019) Carfentrazone-ethyl and glyphosate drift inhibits uredinial formation of *Austropuccinia psidii* on *Eucalyptus grandis* leaves. Pest Manag Sci 75:53–62. https://doi.org/10.1002/ps.5163

Sarowar S, Kim YJ, Kim EN, Kim KD, Choi JY, Hyung NI, Shin JS (2006) Constitutive expression of two pathogenesis-related genes in tomato plants enhanced resistance to oomycete pathogen *Phytophthora capsici*. Plant Cell Tissue Organ Cult 86:7–14. https://doi.org/10.1007/s11240-006-9090-6

Shah J (2003) The salicylic acid loop in plant defense. Curr Opin Plant Biol 6:365–371. https://doi.org/10.1016/S1369-5266(03)00058-X

Shen Y, Liu N, Li C, Wang X, Xu X, Chen W, Xing G, Zheng W (2017) The early response during the interaction of fungal phytopathogen and host plant. Open Biol 7:170057. https://doi.org/10.1098/rsob.170057

Shimono M, Sugano S, Nakayama A, Jiang CJ, Ono K, Toki S, Takatsuji H (2007) Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. Plant Cell 19:2064–2076. https://doi.org/10.1105/tpc.106.046250

Silva PHM, Miranda AC, Moraes MLT, Furtado EL, Stape JL, Alvares CA, Sentelhas PC, Mori ES, Sebbenn AM (2013) Selecting for rust (*Puccinia psidii*) resistance in *Eucalyptus grandis* in São Paulo State, Brazil. For Ecol Manage 303:91–97. https://doi.org/10.1016/J.FORECO.2013.04.002

Stirnweis D, Milani SD, Jordan T, Keller B, Brunne S (2014) Substitutions of two amino acids in the nucleotide-binding site domain of a resistance protein enhance the hypersensitive response and enlarge the PM3F resistance spectrum in wheat. Mol Plant-Microbe Interact 27:265–276. https://doi.org/10.1094/MPMI-10-13-0297-FI

Tang D, Wang G, Zhou J-M (2017) Receptor Kinases in Plant-Pathogen Interactions: More Than Pattern Recognition. Plant Cell 29:618–637. https://doi.org/10.1105/tpc.16.00891

Tao Y, Yuan F, Leister RT, Ausubel FM, Katagiri F (2000) Mutational analysis of the Arabidopsis nucleotide binding site-leucine-rich repeat resistance gene RPS2. Plant Cell 12:2541–2554. https://doi.org/10.1105/tpc.12.12.2541

Tena G, Boudsocq M, Sheen J (2011) Protein kinase signaling networks in plant innate immunity. Curr Opin Plant Biol 14:519–529. https://doi.org/10.1016/J.PBI.2011.05.006

Tobias PA, Guest DI, Külheim C, Park RF (2018) De Novo Transcriptome Study Identifies Candidate Genes Involved in Resistance to *Austropuccinia psidii* (Myrtle Rust) in *Syzygium luehmannii* (Riberry). Phytopathology 108:627–640. https://doi.org/10.1094/PHYTO-09-17-0298-R

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515. https://doi.org/10.1038/nbt.1621

Uhse S, Djamei A (2018) Effectors of plant-colonizing fungi and beyond. PLOS Pathog 14:e1006992. https://doi.org/10.1371/journal.ppat.1006992

Van der Biezen EA, Jones JD (1998) Plant disease-resistance proteins and the gene-for-gene concept. Trends Biochem Sci 23:454–456. https://doi.org/10.1016/s0968-0004(98)01311-5

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Warren RF, Henk A, Mowery P, Holub E, Innes RW (1998) A mutation within the leucine-rich repeat domain of the arabidopsis disease resistance gene RPS5 partially suppresses multiple bacterial and downy mildew resistance genes. Plant Cell 10:1439–1452. https://doi.org/10.1105/tpc.10.9.1439

Xavier AA, Alfenas AC, Matsuoka K, Hodges CS (2001) Infection of resistant and susceptible *Eucalyptus grandis* genotypes by urediniospores of *Puccinia psidii*. Australas Plant Pathol 30:277–281. https://doi.org/10.1071/AP01038

Xavier AA, da Silva AC, da Silva Guimarães LM, Matsuoka K, Hodges CS, Alfenas AC (2015) Infection process of *Puccinia psidii* in *Eucalyptus grandis* leaves of different ages. Trop Plant Pathol 40:318–325. https://doi.org/10.1007/s40858-015-0043-7

Yadav IS, Sharma A, Kaur S, Nahar N, Bhardwaj SC, Sharma TR, Chhuneja P (2016) Comparative Temporal Transcriptome Profiling of Wheat near Isogenic Line Carrying Lr57 under Compatible and Incompatible Interactions. Front Plant Sci 7:1943. https://doi.org/10.3389/fpls.2016.01943

Ye X, Liu H, Jin Y, Guo M, Huang A, Chen Q, Guo W, Zhang F, Feng L (2017) Transcriptomic Analysis of *Calonectria pseudoreteaudii* during Various Stages of Eucalyptus Infection. PLoS One 12:e0169598. https://doi.org/10.1371/journal.pone.0169598

Zauza EA V., Alfenas AC, Old K, Couto MMF, Graça RN, Maffia LA (2010a) Myrtaceae species resistance to rust caused by *Puccinia psidii*. Australas Plant Pathol 39:406–411. https://doi.org/10.1071/AP10077

Zauza EA V., Couto MMF, Lana VM, Maffia LA, Alfenas AC (2010b) Vertical spread of *Puccinia psidii* urediniospores and development of eucalyptus rust at different heights. Australas Plant Pathol 39:141–145. https://doi.org/10.1071/AP09073

Zhao J, Davis LC, Verpoorte R (2005) Elicitor signal transduction leading to production of plant secondary metabolites. Biotechnol Adv 23:283–333.

https://doi.org/10.1016/j.biotechadv.2005.01.003

Zhao J, Zhang X, Guo R, Wang Y, Guo C, Li Z, Chen Z, Gao H, Wang X (2018) Over-expression of a grape WRKY transcription factor gene, VlWRKY48, in *Arabidopsis thaliana* increases disease resistance and drought stress tolerance. Plant Cell Tissue Organ Cult 132:359–370. https://doi.org/10.1007/s11240-017-1335-z

Zheng Z, Qamar SA, Chen Z, Mengiste T (2006) Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. Plant J 48:592–605. https://doi.org/10.1111/j.1365-313X.2006.02901.x

Zhu Q-H, Stephen S, Kazan K, Jin G, Fan L, Taylor J, Dennis ES, Helliwell CA, Wang M-B (2013) Characterization of the defense transcriptome responsive to *Fusarium oxysporum*-infection in Arabidopsis using RNA-seq. Gene 512:259–266. https://doi.org/10.1016/J.GENE.2012.10.036

CHAPTER 3 - COMPARATIVE GENOMIC ANALYSIS OF 22 FUNGAL PATHOGENS OF THE *Ceratocystis fimbriata* COMPLEX

Samuel A. Santos[1], Pedro M. P. Vidigal[2], Christopher A. Smith[3], Camilla C. Ferreira[1], Mark Andersen[3], Thomas C. Harrington[4], Acelino C. Alfenas[1], and Matthew D. Templeton[3*]

[1]Laboratory of Forest Pathology, Department of Plant Pathology, Universidade Federal de Viçosa, Minas Gerais State, Brazil.
[2]Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Minas Gerais State, Brazil.
[3]The New Zealand Institute for Plant and Food Research Limited, Auckland 1142, New Zealand
[4]Department of Plant Pathology, Iowa State University, Ames, Iowa 50010, USA.

*** Corresponding author.** The New Zealand Institute for Plant & Food Research Limited, Mt Albert site, 120 Mt Albert Road, Sandringham, Auckland 1025, New Zealand. E-mail address: matt.templeton@plantandfood.co.nz (M.D., Templeton).

Abstract In this study, we present an assembly and annotation for the nuclear genome of 21 fungal pathogens of the *Ceratocystis fimbriata* complex from different host species and geographic location. The average size of the assembled genomes was 30.4 Mbp, with 7,551 protein-coding, 111 rRNAs, and 366 tRNAs genes. Furthermore, all 21 genomes showed a high score of BUSCO completeness (> 90%), which indicates the high-quality genome assemblies and high accuracy of the gene annotation. In general, independently on host and geographic location, the majority of genome features revealed a high level of similarity among the *Ceratocysstis* fungal isolates. We found that 21 assembled genomes plus *C. fimbriata* type isolate shared a set of 6,141 conserved orthologous clusters genes, which represents 81.33% of the average number of protein-coding genes. Moreover, our high-quality dataset may be useful in future genomic studies involving this important fungal pathogen group.

Keywords: Ceratocystis wilt. Rapid ʻōhiʻa death. Whole-genome sequencing. OrthoMCL analysis. Genomic analysis.

INTRODUCTION

The *Ceratocystis fimbriata* complex (de Beer et al. 2014) comprises many important pathogens that cause wilt, cankers, and rot roots on agricultural and forestry crops, as well as on natural woody ecosystems around the world. There is a limited morphological variation among the fungi of the complex (Webster and Butler 1967), which difficult the species delimitation (Harrington et al. 2014; Fourie et al. 2015). Currently, these fungal pathogens have been grouped in four geographics clades: Latin American clade (LAC) (Engelbrecht and Harrington 2005); the North American clade (NAC) (Johnson et al. 2005); the African clade (AFC) (Heath et al. 2009; Mbenoun et al. 2014); and the Asian-Australian clade (AAC) (Thorpe et al. 2005; Johnson et al. 2005; Li et al. 2017).

Fungal pathogens from the *C. fimbriata* complex have been considered as a continuous threat to crops and native vegetation across the world. In Brazil, this fungus has been causing significant economic losses in different crops, especially in *Eucalyptus* spp. plantations (Zauza et al. 2004; Guimarães et al. 2010; Mafia et al. 2013), mango (Oliveira et al. 2015) and kiwifruit (*Actinidia* sp.) orchards (Piveta et al. 2016; Ferreira et al. 2017). However, in other parts of the globe such as Hawai, the pathogen has killed hundreds of thousands of *Metrosideros polymorpha*, which is the most common and widespread native tree species of Hawai (Keith et al. 2015; Mortenson et al. 2016; Barnes et al. 2018). Due to its wide host range and economic importance, this pathogen group has been extensively studied, especially using molecular markers such as microsatellite to access the genetic variability on the pathogen population (Steimel et al. 2004; Nkuekam et al. 2009; Ferreira et al. 2010; Rizatto et al. 2010; Simpson et al. 2013; Oliveira et al. 2015; Li et al. 2016).

The advance on the genome sequencing technologies has been providing high-quality whole genomic sequences much more rapidly and cheaply (Goodwin et al. 2016). Recently, comparative genomic analysis based on the whole-genome features has been extensively applied in several fungal groups, providing significant insights in their biology, taxonomy, and pathogenicity (Yang et al. 2017; Choi and Kim 2017; Muñoz et al. 2018). In this context, a complete characterization of the genome of fungal pathogens from *C. fimbriata* complex, obtained from different host species and geographic locations, can provide crucial information to a better understanding of their biology and pathogenicity.

In this study, we present a high-quality assembly and annotation of the nuclear genome for 21 isolates of the *C. fimbriata* complex from different host species and geographic locations.

In addition, we performed an OrthoMCL clustering analysis, providing a minimal quorum of orthologous conserved genes among them, as well as the specific genes for each fungal isolate.

MATERIAL AND METHODS

Fungal isolates

A total of 21 fungal isolates of the *C. fimbriata* complex was sequenced. These fungal pathogens were initially isolated from different host species and geographic regions (Table 1). The pure cultures were obtained from three collections: Laboratory of Forest Pathology (LPF) of the Universidade Federal de Vicosa, Brazil; Collection of T.C. Harrington of the Iowa State University, USA (C); and Plant and Food Research Collection, New Zealand. Additionally, the genome of *C. fimbriata* type isolate (Accession Number: APWK00000000.3) (Wingfield et al. 2019), isolated from sweet potato, was downloaded from GenBank and used on the OrthoMCL clustering analysis.

DNA extraction and genome sequencing

DNA was extracted using Wizard Genomic DNA Purification Kit® (Promega, Madison, United States) from mycelia collected from a pure single spore culture, grown in Petri plate, containing 2% malt extract agar (MEA, 20 g L^{-1} Agar, 20 g L^{-1} malt extract) at 25°C for 20 days. The fungal mycelia were added in a microcentrifuge tube of 2 mL containing two tungsten carbide beads, then it was disrupted in a Qiagen Tissuelyser for 2 min at 30 Hz frequency and the DNA was extracted according to the kit instructions.

For sequencing, pair-end libraries with an average insert size of 350 bp were constructed from each sample of the genomic DNA of 21 fungal isolates according to the manufacturer's instructions (Illumina, San Diego, CA). The sequencing was performed using an Illumina MiSeq or Illumina NovaSeq 6000 platforms (Table 1), at the Australian Genome Research Facility, Melbourne, Australia (http://www.agrf.org.au/) and GenOne Biotechnologies, Rio de Janeiro, Brazil (http://www.genone.com.br/). The genomes were sequenced with high coverage, ranging from 130 to 280-fold (Table 1).

Genome assembly and evaluation

Low-quality reads (Q30 threshold) and Illumina adapter sequences were removed using Trimmomatic v. 0.36 (Bolger et al. 2014). Quality control checks were performed by FastQC v. 0.11.5 (Andrews 2010) before and after the trimming step to ensure the quality of the final dataset (clean reads). Trimmed reads were *de novo* assembled testing the following assemblers using default parameters: A5-Miseq v. 20160825 (Coil et al. 2015); SPAdes v. 3.13.0 (Bankevich et al. 2012); Platanus v. 1.2.4 (Kajitani et al. 2014); and SOAPdenovo2 v. 2.04 (Luo et al. 2012).

For each isolate, the four assemblies were evaluated using QUAST v. 4.2 regarding the following metrics: total size, number of contigs, length of the largest contig, and N50 (Gurevich et al. 2013). The best draft assembly was submitted to Redundans v. 0.14a for scaffolding and gap close using default parameters (Pryszcz and Gabaldón 2016). Thereafter, Pilon v. 1.23 was used for polishing to reduce misassembles (Walker et al. 2014). To identify and eliminate potential contaminates from the final assembly, a BLASTn of all assembled scaffolds against a bacteria-virus database was performed (Altschul et al. 1990), considering an E-value of 1e-25. Finally, to get the final assembly of the nuclear genome, the scaffold containing the mitochondrial genome was removed.

Genome size estimation

Trimmed reads were used to estimate the genome size based on k-mer distribution and depth as described by Tan et al. (2018). First, k-mer counting was performed with jellyfish v2.2.10 (Marçais and Kingsford 2011). Then, histograms of 17-, 21-, and 25-kmers frequency distributions were processed by GenomeScope (Vurture et al. 2017).

Genome annotation and evaluation

For the LPF1912 isolate, protein-coding genes were predicted using BRAKER pipeline v. 2.1.0 (Stanke et al. 2006; Stanke et al. 2008; Hoff et al. 2016; Hoff et al. 2019). Trimmed reads of 12 RNA-Seq libraries from eucalyptus infected by *C. fimbriata* LPF1912 were mapped to the scaffold sequences of the LPF1912 nuclear genome using STAR v. 2.7.1a by selecting default parameters (Dobin et al. 2013). Thereafter, the obtained BAM files and scaffold sequences of the LPF1912 nuclear genome were submitted to BRAKER v. 2.1.0 using default parameters. For the other isolates, which do not have available RNA-seq data, protein-coding genes were

predicted using AUGUSTUS version 3.3 (Stanke and Morgenstern 2005), considering the trained set of parameters for *C. fimbriata* LPF1912. To evaluate gene prediction accuracy, the completeness of the gene prediction was assessed using the Benchmarking Universal Single-Copy Orthologs, named BUSCO (Simão et al. 2015). BUSCO was performed on all predicted genes, making use of the Ascomycota lineage dataset (ascomycota_odb9).

The predicted coding-protein genes were functionally annotated through similarity searches using BLAST version 2.6.0 (Altschul et al. 1990). In these searches, the encoded proteins were aligned to the sequences of NCBI protein non-redundant database (NCBInr), UniProt Knowledgebase (UniProtKb), using protein BLAST (BLASTp) (cutoff for significant hits: E-value of 1e-10). Furthermore, gene ontology (GO) terms were assigned to the proteins using GOanna tool from AgBase (https://agbase.arizona.edu/cgi-bin/tools/GOanna.cgi) (McCarthy et al. 2006) by selecting the following parameters: E-value = 1e-05; matrix BLOSUM62; word size = 3; identity = 30%; and query coverage = 70%. Ribosomal genes were identified by Barrnap v. 0.9 (https://github.com/tseemann/barrnap), using Eukaryota database, and transporter RNA genes were identified by tRNAscan-SE v. 2.0.5 (http://lowelab.ucsc.edu/tRNAscan-SE/) using Eukaryotic database and default options (Chan and Lowe 2019). RepeatMasker version open-4.0.8 (http://www.repeatmasker.org/) was used to identify repetitive sequences using Fungi database of RepBase library (v. 20170127) (Jurka et al. 2005). Additionally, RepeatModeler v. 2.0 (http://www.repeatmasker.org/RepeatModeler/) was used to perform a *de novo* transposable element (TE) family identification and modeling.

OrthoMCL clustering analysis

The Markov clustering program OrthoMCL v. 2.0 (Li et al. 2003) was used to identify orthologous genes between all sequenced fungal isolates and *C. fimbriata* type isolate. Thus, amino acid sequences of proteins encoding genes were submitted to reciprocal similarity searches using BLASTp tool of BLAST v. 2.9.0 (Altschul et al. 1990), considering an E-value of 1e-10 as a threshold for the significant alignments. Putative orthologous were then clustered with OrthoMCL using an inflation value of 1.5 and a similarity of 50%.

RESULTS

Genome assembly stats

The estimated genome size ranged from 26.1 to 31.2 Mbp among the 21 isolates of the *C. fimbriata* complex, with the majority close to the average (29 Mbp) (Table 2). A5Miseq assembler was the most effective software for the majority of genomes, followed by SPAdes and Platanus (Table 2). In addition, the total size of the assembled genome ranged from 27.9 to 31.8 with an average of 30.3 Mbp and was similar to the estimated size for the majority of isolates (Table 2).

Among the 21 *Ceratocytis* genomes, the number of assembled scaffolds ranged from 346 to 1,693 with an average of 764 (Table 2). The minimum size of the largest scaffold was observed for CFNZ isolate (167,813 bp), while the maximum size was found for LPF1922 (1,229,531 bp) (Table 2). Additionally, N50 ranged from 30,693 to 260,845 bp among the assembled genomes with an average size of 151,948 bp (Table 2). Finally, the sequencing coverage of the assemblies ranged from 37 to 199-fold with an average of 107-fold among all analysed isolates (Table 2).

Genome annotation

Genes were predicted and annotated as protein-coding, ribosomal RNA (rRNA), and transfer RNA (tRNA) (Table 3). For most isolates, the number of protein-coding genes was close to the average (7,551), except for LPF1560, C1714, C4118, C4186, and CFNZ (Table 3). All isolates showed a BUSCO completeness higher than 90% and most of the complete orthologous genes were classified as a single copy (Table 3).

Four types of rRNA were found among the isolates and 5S and 5.8S were the one with higher number of copies (Table 3). The majority of *Ceratocystis* isolates showed a single copy of 18S and 28S, except LPF1912, LPF1413, LPF1922, C1714, C4184, and C4191 isolates (Table 3). Additionally, the number of tRNAs was similar among the majority of isolates with an average of 366, except for C4118 isolate that had 318 (Table 3).

For all isolates, the G+C content in the coding DNA sequence (CDS) regions was higher than in the whole-genome (Fig. 1A). The average in the whole-genome and CDS were 48.26 and 52.65%, respectively. In the CDS regions, the G+C content was similar for all isolates (Fig. 1A). However, in the whole-genome, C1714 and C4186 isolates had G+C content higher than others (Fig. 1A).

The proportion of repetitive sequences in the whole-genome ranged from 9.86 (C1714 isolate) to 17.18% (LPF1912 isolate) with an average of 14.59% among all isolates (Fig. 1B). Repetitive sequences were identified and classified as transposable elements (TEs) [Short Interspersed Nuclear Elements (SINEs), Long Interspersed Nuclear Elements (LINE), Long Terminal Repeat Elements (LTR), and DNA transposons], small RNA, satellites, simple repeats, low complexity repeats, and unclassified (Table 4).

Retroelements type SINEs were found only for LPF1912, LPF1400, LPF1427, LPF1470, LPF1674, LPF1688, C1714, C4186, C4191, and CFNZ isolates. However, LINEs, LTRs, and DNA transposons were observed for all isolates (Table 4). The number of LINEs ranged from 782 (LPF1560 isolate) to 1,915 (LPF1912 isolate) with an average of 1,424, while LTRs ranged from 471 (C4186 isolate) to 1,527 (LPF1415 isolate) with an average of 1,098 among all isolates (Table 4). Additionally, the number of DNA transposons found ranged from 682 (C1714 isolate) to 2,452 (LPF1912 isolate) with an average of 1,881 (Table 4).

Small RNAs were found only in LPF1400, LPF1413, and LPF1415 isolates (Table 4). However, satellite regions were annotated for most isolates, except LPF1912, LPF1674, LPF1560, and C4184 (Table 4). Simple and Low complexity repeats were abundant in all *Ceratocystis* nuclear genome (Table 4). For all isolates, a large proportion of the repetitive sequences was unclassified (Table 4).

OrthoMCL clustering analysis of protein-coding genes

The 22 Ceratocytis genomes shared a total of 6,141 conserved orthologous clusters genes with the majority (5,931) as a single-copy (Fig. 2). C4186, C1714, and LPF1560 isolates had 8, 7, and 3 unique clusters with 18, 15, and 7 protein-coding genes, respectively (Fig. 2; Table 5). LPF1912, LPF1668, C4191, and *C. fimbriata* (type) showed two unique clusters with 7, 4, 4, and 4 genes, respectively (Fig. 2; Table 5). In addition, CFNZ, LPF1413, LPF1674, LPF1922, and C4118 had only one unique cluster (Fig. 2; Table 5). Finally, LPF1400, LPF1415, LPF1427, LPF1458, LPF1470, LPF1673, LPF1682, LPF1688, LPF1701, and C4184 had no unique orthologous clusters (Fig. 2).

The two *C. fimbriata* isolates from sweet potato (*Ipomoea batatas*) shared a high number of exclusive orthologous clusters genes (Fig. 2; Table 6). On the other hand, the 13 isolates from kiwifruit (*Actinidia* sp.) shared three clusters. Furthermore, the three isolates from ʻōhiʻa (*Metrosideros polymorpha*) shared only one exclusive single-copy cluster (Fig. 2; Table 6).

Functional annotation of unique and shared OrthoMCL clusters genes

Unique orthologous genes had function associated with enzymes such as endonuclease, chitinase, chromatin biding and organization, vegetative incompatibility, and retrotransposons (Table 5).

OrthoMCL exclusive clusters shared between the two sweet-potato isolates were associated mainly with enzymatic functions such as phosphatase, endonuclease, phosphodiesterase, alginate lyase, and β-1,3-glucanase (Table 6). Exclusive proteins shared by kiwifruit isolates had an uncharacterized function. Finally, the exclusive cluster shared among the three isolates from ōhi‘a had no-hit with UniProtKB (Table 6).

DISCUSSION

In this study, we present a high-quality assembly and annotation of 21 nuclear genome of *C. fimbriata* complex, isolated from different host species and geographic location. This pathogen group is well known to cause serious wilt diseases in crops and native tree species around the world (Zauza et al. 2004; Guimarães et al. 2010; Mafia et al. 2013; Keith et al. 2015; Mortenson et al. 2016; Ferreira et al. 2017; Barnes et al. 2018). Additionally, we used the genome sequence of *C. fimbriata* type isolate (Wingfield et al. 2019) as a reference in the OrthoMCL clustering analysis. Among the 21 fungal isolates, the average size of the assembled genome was 30.4 Mbp, with 7,551 protein-coding, 111 rRNAs, and 366 tRNAs genes. All 21 isolates showed a high score of BUSCO completeness (> 90%), which indicates that gene annotation was accurate. Furthermore, the stats of genome assembly and annotation were similar to other *Ceratocystis* species, such as *C. manginecans* (van der Nest et al. 2014), *C. eucalypticola* (Wingfield et al. 2015), *C. cacaofunesta* (Molano et al. 2018), *C. fimbriata* type isolate (Wingfield et al. 2019), and *C. albifundus* (van der Nest et al. 2019).

The majority of genome features were similar among the isolates, except for predicted rRNA genes and transposable elements (TEs). Internal transcribed sequences (ITS) of rRNA has been largely used as a DNA barcode marker for fungi (Schoch et al. 2012). ITS also has been applied to the *C. fimbriata* complex, especially in taxonomic studies. Many *Ceratocystis* new species have been proposed based on the phylogenetic species concept proposed by Harrington and Rizzo (1999) or using variation on the ITS sequence (Barnes et al. 2003; van Wyk et al. 2007; van Wyk et al. 2009; van Wyk et al. 2010; van Wyk et al. 2011a; van Wyk et al. 2011b; van Wyk et al. 2012; Li et al. 2017; Barnes et al. 2018). As already reported, there is

an intraspecific and intragenomic variability on the ITS sequence which has caused taxonomic problems (Harrington et al. 2014). In addition, our results revealed that some isolates have multiple copies of 5.8, 18, and 28 rRNA subunits, which are localized in different positions of the nuclear genome (Table 3). Thus, comparing polymorphism on the ITS sequence in these specific isolates may lead to wrong interpretation.

OrthoMCL clustering analysis revealed high similarity among *Ceratocystis* pathogens, independently on host species or geographic location. A total of 6,141 orthologous clusters genes were conserved for all 22 *Ceratocystis* genomes, which represents 81.33% of the average number of protein-coding genes. The close relationship between fungal pathogens from *C. fimbriata* complex have been reported over the years in several studies (Harrington et al. 2014; Fourie et al. 2015; Oliveira et al. 2015), leading to grouping these fungal pathogens into the *C. fimbriata* complex (de Beer et al. 2014). Even though, some isolates had unique clusters of genes, especially C4186, C1714, LPF1560, C4191, and LPF1912 isolates. Although the majority has an uncharacterized function, some of these genes were associated with enzymatic function and transposable elements (TEs). Some enzymes such as polygalacturonase has been considered as one of the most important enzyme family responsible for the depolymerization of cell walls (Quoc and Chau 2017), and have been considered as a virulence factor to many fungal pathogens (Shieh et al. 1997; Oeser et al. 2002; Mori et al. 2008; Ye et al. 2017). Moreover, the role of TEs in biology, evolution, and pathogenicity of fungi has been extensively studied. In fact, pathogenic fungi often have more TEs inserted into genes than saprophytic (Muszewska et al. 2019). TEs have been especially associated with genome plasticity, evolution, gene content, gene expression regulation, and pathogenicity (Castanera et al. 2016; Krishnan et al. 2018; Razali et al. 2019).

C4186 and C4191 isolates were recently found causing a devastating disease in Hawai`i called rapid ʻōhiʻa death (Barnes et al. 2018). They represent two distinct lineages and were described as new species of *Ceratocystis* – *C. lukuohia* and *C. huliohia*, respectively (Barnes et al. 2018). In addition, C1714 has been known in Hawai`i since the 1970s (Uchida and Aragaki 1979), but it was classified as a new species (*C. uchidae*) just in 2017 (Li et al. 2017). On the other hand, LPF1560 and LPF1912 were collected in Brazil and both were maintained as *C. fimbriata*. LPF1560 was found causing seed rot and seedling blight on a native *Carapa guianensis* f.sp. carapa (andiroba) in Amazonian rainforests, Acre state, Brazil (Valdetaro et al. 2019), while LPF1912 is well known to causes wilt on *Eucalyptus* spp. plantations since the 1990s (Ferreira et al. 1999). We believe that further comparative genomic analysis may be

helpful to a better understanding of the host specialization processes involving the *C. fimbriata* complex pathogens.

CONCLUSIONS

In this study, we present *de novo* genome assembly and annotation for the nuclear genome of 21 fungal pathogens of the *Ceratocystis fimbriata* complex. All assemblies exhibited a higher level of genome quality and BUSCO completeness compared to other *Ceratocystis* genomes. Regardless of host and geographic location, the majority of genome features showed a high level of similarity among all isolates. We found that 21 *Ceratocystis* assembled genomes and *C. fimbriata* type isolate shared a set of 6,141 conserved orthologous clusters genes, which represents 81.33% of the average number of protein-coding genes. Moreover, we provide a high-quality dataset that may be useful in future genomic studies involving this important fungal pathogen group.

Acknowledgments The authors thank to 'Conselho Nacional de Desenvolvimento Cientifico e Tecnológico' (CNPq) and 'Coordenação de Aperfeicoamento de Pessoal de Nivel Superior' (CAPES) for the fellowship granted to first the author. To Suzano S.A. for supporting this study. We also thanks to Plant and Food Research for providing all bioinformatics facilities and Núcleo de Análise de Biomoléculas (NuBioMol) for supporting data analysis. NuBioMol is supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), CAPES, CNPq, Financiadora de Estudos e Projetos (Finep) and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).

Table 1 List of *Ceratocystis fimbriata* complex isolates used in this study

Isolate code	Identity	Isolated from	Location	Illumina platform	Data[e] (Gb)
LPF1912[a]	*C. fimbriata*	*Eucalyptus* sp.	Bahia State, Brazil	Miseq	5.91
LPF1400[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	NovaSeq 6000	4.12
LPF1413[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	Miseq	6.29
LPF1415[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	NovaSeq 6000	4.23
LPF1427[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	NovaSeq 6000	4.34
LPF1458[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	NovaSeq 6000	4.09
LPF1470[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	NovaSeq 6000	4.63
LPF1668[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	Miseq	7.58
LPF1673[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	NovaSeq 6000	4.37
LPF1674[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	NovaSeq 6000	4.69
LPF1682[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	NovaSeq 6000	4.66
LPF1688[a]	*C. fimbriata*	*Actinidia* sp.	Fraiburgo, SC State, Brazil	NovaSeq 6000	4.02
LPF1701[a]	*C. fimbriata*	*Actinidia* sp.	São Joaquim, SC State, Brazil	NovaSeq 6000	4.49
LPF1922[a]	*C. fimbriata*	*Actinidia* sp.	Farroupilha, RS State, Brazil	Miseq	9.11
LPF1560[a]	*C. fimbriata*	*Carapa guianensis*	Rio Branco, AC, Brazil	NovaSeq 6000	8.87
C1714[b]	*C. uchidae*	*Colocasia esculenta*	Oahu, Hawai`i, USA	Miseq	6.87
C4118[b]	*C. fimbriata*	*Syngonium* sp.	Hawai`i nursery, USA	Miseq	7.32
C4184[b]	*C. lukuohia*	*Metrosideros polymorpha*	Leilani Estates, Hawai`i, USA	Miseq	6.31
C4186[b]	*C. lukuohia*	*M. polymorpha*	South Kona, Hawai`I, USA	Miseq	6.22
C4191[b]	*C. huliohia*	*M. polymorpha*	Hawai`I, USA	Miseq	6.35
CFNZ[c]	*C. fimbriata*	*Ipomoea batatas*	New Zealand	Miseq	8.91
CBS114723[d]	*C. fimbriata* (type)	*I. batatas*	North Carolina, USA	-	-

[a] Pure culture collection of the Laboratory of Forest Pathology of the Universidade Federal de Vicosa, Brazil.

[b] Pure culture collection of T.C. Harrington, Iowa State University, Iowa, USA.

[c] Pure culture collection of The New Zealand Institute for Plant and Food Research, Auckland, New Zealand.

[d] Genome sequence downloaded from GenBank (Accession Number: APWK00000000.3) (Wingfield et al. 2019).

[e] Amount of raw data (paired-end reads) generated by sequencing.

Table 2 Summary stats of the assembled nuclear genome for 21 isolates of the *Ceratocystis fimbriata* complex

Isolate code	Estimated size[a] (Mbp)	Assembler[b]	Size[c] (Mbp)	Number of scaffolds[d]	Largest scaffold (bp)	N50[e] (bp)	Assembly coverage[f]
LPF1912	31.2	A5Miseq	31.6	467	922 786	230 304	86
LPF1400	30.1	A5Miseq	30.4	695	639 079	165 709	83
LPF1413	29.0	A5Miseq	31.5	598	1 181 466	216 652	133
LPF1415	30.6	A5Miseq	31.0	860	977 131	146 028	106
LPF1427	29.8	A5Miseq	30.4	791	631 135	160 008	106
LPF1458	30.6	A5Miseq	31.1	873	793 222	140 717	100
LPF1470	29.8	A5Miseq	30.3	813	616 834	160 367	118
LPF1668	28.9	A5Miseq	31.8	661	901 347	165 784	141
LPF1673	29.8	A5Miseq	30.3	781	571 911	158 207	109
LPF1674	30.7	A5Miseq	31.2	891	615 496	123 308	114
LPF1682	29.7	A5Miseq	30.3	873	578 166	147 968	121
LPF1688	30.0	A5Miseq	30.5	821	576 252	147 866	89
LPF1701	29.9	A5Miseq	30.3	839	632 759	148 027	113
LPF1922	29.3	Platanus	31.0	441	1 229 531	260 845	40
LPF1560	28.6	Platanus	28.9	693	482 818	90 556	199
C1714	26.1	A5Miseq	27.9	879	389 213	80 060	166
C4118	27.4	SPAdes	28.9	486	388 407	111 853	37
C4184	26.7	A5Miseq	30.5	931	637 241	128 196	131
C4186	27.9	SPAdes	28.0	346	624 682	214 104	43
C4191	29.9	A5Miseq	31.2	604	624 239	163 647	42
CFNZ	28.1	SPAdes	29.2	1693	167 813	30 693	169
Mean	29.2	-	30.3	764	675 311	151 948	107

[a] Mean estimated genome size based on the k-mers frequency distributions of 17-, 21-, and 25-mers.

[b] Software that showed the best results for each isolate.

[c] Combined length of all assembled scaffolds bigger than 1000 bp.

[d] Assembled scaffolds bigger than 1000 bp.

[e] N50 is defined as the sequence length of the shortest scaffold at 50% of the total genome length.

[f] Mean coverage of all assembled scaffolds.

Table 3 Gene prediction for the nuclear genome of 21 isolates of the *Ceratocystis fimbriata* complex

Isolate code	Protein-coding genes	CDS[a] (Mbp)	BUSCO[b]	Number of rRNA genes				Number of tRNA genes	Total number of genes
				5S	5.8S	18S	28S		
LPF1912	7 742	12.5	C: 91.7% [S: 91.4%, D: 0.3%], F: 4.5%, M: 3.8%	111	7	3	5	376	8 244
LPF1400	7 461	12.0	C: 91.1% [S: 91.0%, D: 0.1%], F: 4.9%, M: 4.0%	115	1	1	1	376	7 955
LPF1413	7 695	12.4	C: 91.3% [S: 91.2%, D: 0.1%], F: 4.7%, M: 4.0%	111	9	3	3	374	8 195
LPF1415	7 622	12.2	C: 91.2% [S: 91.1%, D: 0.1%], F: 4.7%, M: 4.1%	110	1	1	1	374	8 109
LPF1427	7 461	12.0	C: 91.1% [S: 91.0%, D: 0.1%], F: 4.9%, M: 4.0%	114	1	1	1	376	7 954
LPF1458	7 635	12.2	C: 91.3% [S: 91.2%, D: 0.1%], F: 4.6%, M: 4.1%	109	1	1	1	373	8 120
LPF1470	7 459	12.0	C: 91.2% [S: 91.1%, D: 0.1%], F: 4.8%, M: 4.0%	108	2	1	1	370	7 941
LPF1668	7 791	12.6	C: 91.3% [S: 91.2%, D: 0.1%], F: 4.6%, M: 4.1%	109	3	1	1	375	8 280
LPF1673	7 463	12.0	C: 91.1% [S: 91.0%, D: 0.1%], F: 4.9%, M: 4.0%	112	1	1	1	376	7 954
LPF1674	7 656	12.2	C: 91.3% [S: 91.2%, D: 0.1%], F: 4.7%, M: 4.0%	106	2	1	1	370	8 136
LPF1682	7 458	12.0	C: 91.0% [S: 90.9%, D: 0.1%], F: 4.9%, M: 4.1%	115	1	1	1	376	7 952
LPF1688	7 515	12.1	C: 91.1% [S: 91.0%, D: 0.1%], F: 4.9%, M: 4.0%	109	1	1	1	375	8 002
LPF1701	7 448	11.9	C: 91.1% [S: 91.0%, D: 0.1%], F: 4.9%, M: 4.0%	115	1	1	1	375	7 941
LPF1922	7 562	12.2	C: 91.2% [S: 91.1%, D: 0.1%], F: 4.8%, M: 4.0%	119	4	6	2	390	8 083
LPF1560	7 282	11.9	C: 91.2% [S: 91.1%, D: 0.1%], F: 4.8%, M: 4.0%	98	1	1	1	344	7 727
C1714	7 146	11.6	C: 91.0% [S: 90.9%, D: 0.1%], F: 5.0%, M: 4.0%	91	11	5	1	342	7 596
C4118	7 277	11.9	C: 91.3% [S: 91.2%, D: 0.1%], F: 4.8%, M: 3.9%	87	1	1	1	318	7 685
C4184	7 581	12.2	C: 91.3% [S: 91.2%, D: 0.1%], F: 4.8%, M: 3.9%	99	7	3	1	366	8 057
C4186	7 235	11.8	C: 91.5% [S: 91.4%, D: 0.1%], F: 4.5%, M: 4.0%	90	1	1	1	356	7 684
C4191	7 658	12.3	C: 91.3% [S: 91.2%, D: 0.1%], F: 4.8%, M: 3.9%	101	4	2	2	371	8 138
CFNZ	7 272	11.8	C: 90.2% [S: 90.1%, D: 0.1%], F: 5.3%, M: 4.5%	74	1	1	1	336	7 685
Mean	7 551	12.1	C: 91.2% [S: 91.1%, D: 0.1%], F: 4.8%, M: 4.0%	105	3	2	1	366	7 973

[a] CDS, transcribed sequencing.

[b] Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment using 3,725 conserved genes of Sordariomyceta lineage dataset (sordariomyceta_odb9). C: complete; S: single-copy; D: duplicate; F: fragmented; M: missing.

Table 4 Overview of repetitive sequences annotation for the nuclear genome of 21 isolates of the *Ceratocystis fimbriata* complex

Isolate code	Retroelements[a]			DNA transposons	Small RNA	Satellites	Simple repeats	Low complexity	Unclassified	Total (Mbp)
	SINE	LINE	LTR							
LPF1912	52 (4328 bp)	1915 (805412 bp)	1514 (1252590 bp)	2452 (1037633 bp)	0	0	11599 (487076 bp)	2575 (136964 bp)	4028 (1706823 bp)	5.43
LPF1400	54 (4892 bp)	1821 (631064 bp)	1300 (954322 bp)	2075 (850982 bp)	40 (8183 bp)	49 (11042 bp)	11990 (509773 bp)	2553 (135290 bp)	3598 (1319271 bp)	4.42
LPF1413	0	1645 (709978 bp)	1414 (945497 bp)	1351 (529288 bp)	21 (28036 bp)	93 (21422 bp)	11611 (494340 bp)	2525 (134879 bp)	4999 (2377741 bp)	5.24
LPF1415	51 (4673 bp)	1771 (574358 bp)	1527 (1145403 bp)	2039 (883377 bp)	84 (17346 bp)	78 (15555 bp)	11684 (497140 bp)	2643 (139798 bp)	4205 (1780293 bp)	5.06
LPF1427	17 (1683 bp)	1414 (585955 bp)	633 (573724 bp)	1828 (745776 bp)	0	142 (26555 bp)	11807 (498376 bp)	2617 (138232 bp)	4962 (1796125 bp)	4.37
LPF1458	0	1628 (573583 bp)	1273 (913328 bp)	2358 (934205 bp)	0	105 (24460 bp)	11800 (502369 bp)	2603 (137341 bp)	4524 (1892927 bp)	4.98
LPF1470	48 (4821 bp)	1471 (571390 bp)	918 (755477 bp)	2122 (855585 bp)	0	97 (21215 bp)	11898 (504459 bp)	2609 (137654 bp)	4239 (1547493 bp)	4.40
LPF1668	0	1660 (712586 bp)	1513 (1332194 bp)	2139 (944866 bp)	0	91 (20995 bp)	11852 (498565 bp)	2562 (135360 bp)	4628 (1788171 bp)	5.43
LPF1673	0	1602 (622584 bp)	1091 (822724 bp)	2174 (964782 bp)	0	99 (22545 bp)	11875 (502207 bp)	2559 (135786 bp)	4082 (1502064 bp)	4.57
LPF1674	57 (5045 bp)	1453 (565680 bp)	1467 (1180569 bp)	2408 (1015179 bp)	0	0	11936 (504666 bp)	2599 (137363 bp)	4902 (1835589 bp)	5.24
LPF1682	0	1603 (596496 bp)	1421 (1131649 bp)	2114 (834808 bp)	0	137 (26585 bp)	11908 (504472 bp)	2547 (135187 bp)	3553 (1117261 bp)	4.35
LPF1688	52 (5182 bp)	1685 (600855 bp)	869 (603075 bp)	1901 (833523 bp)	0	114 (26069 bp)	11911 (504062 bp)	2507 (132013 bp)	4625 (1851896 bp)	4.56
LPF1701	0	1160 (464540 bp)	1178 (883542 bp)	2314 (966515 bp)	0	101 (22022 bp)	11888 (504212 bp)	2511 (133266 bp)	4742 (1489435 bp)	4.46
LPF1922	0	1249 (608079 bp)	1194 (995147 bp)	1966 (882500 bp)	0	65 (15934 bp)	11819 (497481 bp)	2660 (141414 bp)	4502 (1546593 bp)	4.68
LPF1560	0	782 (392451 bp)	906 (615143 bp)	1628 (524026 bp)	0	0	11981 (509035 bp)	2554 (137162 bp)	3975 (1384386 bp)	3.56
C1714	31 (2023 bp)	1145 (524193 bp)	617 (202119 bp)	686 (276605 bp)	0	83 (18677 bp)	9987 (423885 bp)	2163 (112285 bp)	4161 (1193749 bp)	2.75
C4118	0	986 (473719 bp)	547 (563376 bp)	1542 (602482 bp)	0	89 (18173 bp)	12057 (507368 bp)	2502 (131910 bp)	4013 (1447138 bp)	3.74
C4184	0	1041 (560501 bp)	1115 (801125 bp)	2109 (888360 bp)	0	0	12256 (510163 bp)	2611 (137329 bp)	4965 (1593513 bp)	4.49
C4186	179 (25505 bp)	1329 (507188 bp)	471 (281648 bp)	971 (447400 bp)	0	80 (18839 bp)	10020 (425928 bp)	2181 (112901 bp)	3458 (991036 bp)	2.81
C4191	88 (12947 bp)	1729 (685505 bp)	926 (1111779 bp)	1858 (787332 bp)	0	89 (19231 bp)	12429 (518442 bp)	2559 (134041 bp)	4927 (2032774 bp)	5.30
CFNZ	70 (7081 bp)	821 (323180 bp)	1163 (1102133 bp)	1474 (553523 bp)	0	87 (17253 bp)	11081 (473251 bp)	2329 (124321 bp)	3904 (1092171 bp)	3.69
Mean	33 (3723 bp)	1424 (575681 bp)	1098 (865074 bp)	1881 (778988 bp)	7 (2551 bp)	76 (16503 bp)	11685 (494156 bp)	2522 (133357 bp)	4333 (1585069 bp)	4.46

[a] SINEs: Short interspersed nuclear elements; LINEs: Long Interspersed Nuclear Elements; LTRs: Long Terminal Repeat Elements.

Table 5 Functional annotation of OrthoMCL unique orthologous clusters for each *Ceratocystis* isolate

Isolate code	OrthoMCL cluster	Gene ID	Protein name[a]	UniProt ID
C4186	1: cluster8339	g1588; g7018; g7194; g7229	Uncharacterized protein	W9NNH0
	2: cluster8630	g1584; g2014	Chromatin organization modifier	A0A179F0E8
	3: cluster8631	g1585; g6652	APH domain-containing protein	A0A2C5WSZ5
	4: cluster8632	g604; g4063	Chromo domain-containing protein	A0A0B1P9N0
	5: cluster8633	g4569; g7040	Uncharacterized protein	A0A0F8CNH7
	6: cluster8634	g4571; g7160	CCHC-type domain-containing protein	A0A0F8D7K9
	7: cluster8635	g6020; g6655	HTH CENPB-type domain-containing protein	A0A4Q2UZR0
	8: cluster8636	g6967; g6970	Uncharacterized protein	A0A2C5X0M9
C1714	1: cluster8448	g5905; g6704; g7035	no hit	no hit
	2: cluster8645	g1310; g6234	no hit	no hit
	3: cluster8646	g3267; g7051	Putative double-stranded RNA/RNA-DNA hybrid binding protein	A0A0F8BRS4
	4: cluster8647	g5365; g6927	Patatin	A0A0F8DI55
	5: cluster8648	g7100; g7103	Vegetative incompatibility protein	A0A2C5XIB7
	6: cluster8649	g7109; g7118	DDE-type integrase/transposase/recombinase	A0A535FP05
	7: cluster8652	g6381; g7070	no hit	no hit
LPF1560	1: cluster8443	g7092; g7134; g7277	Endonuclease	A0A0F8BJD4
	2: cluster8640	g6869; g6879	no hit	no hit
	3: cluster8641	g6926; g7061	Retrotransposon gag protein	A0A0F8AW62
LPF1912	1: cluster8251	g451; g1284; g3535; g3536; g7397	Minor spike protein	A0A0M7AVB0
	2: cluster8654	g1183; g2366	no hit	no hit
C4191	1: cluster8637	g7570; g7610	Uncharacterized protein	A0A0F8AZY0
	2: cluster8638	g7576; g7656	Retrotransposon gag protein	A0A0F8AW62
LPF1413	1: cluster8639	g7671; g7673	no hit	no hit
LPF1674	1: cluster8644	g2447; g7525	Chitinase 1	A0A0F8CQV0
LPF1922	1: cluster8650	g7457; g7463	no hit	no hit
C4118	1: cluster8629	g5674; g6649	Retrotransposable element Tf2	A0A0F8D048
CFNZ	1: cluster8450	g5675; g6576; g6966	Retrotransposon gag protein	A0A0F8BIR3

[a] Automatic annotation by BLASTp against of NCBI protein non-redundant - UniProt Knowledgebase (UniProtKb).

Table 6 Functional annotation of OrthoMCL exclusive orthologous clusters shared among *Ceratocystis* fungal isolates from sweet-potato, kiwifruit, and ōhi'a

Host	OrthoMCL cluster	Gene ID	Protein name[d]	UniProt ID
sweet-potato[a]	1: cluster9117	cfnz|g6685; cfim|PHH51181.1	Vegetative incompatibility protein	A0A2C5WZV7
	2: cluster9143	cfnz|g6902; cfim|PHH53731.1	Alkaline phosphatase	A0A0F8B1R4
	3: cluster9155	cfnz|g6833; cfim|PHH49489.1	Uncharacterized protein	A0A2C5W2R3
	4: cluster9164	cfnz|g6953; cfim|PHH49151.1	Uncharacterized protein	A0A2C5WU27
	5: cluster9165	cfnz|g6574; cfim|PHH49265.1	Endonuclease	A0A0F8D8U4
	6: cluster9166	cfnz|g4609; cfim|PHH49375.1	Uncharacterized protein	A0A2C5WS90
	7: cluster9167	cfnz|g2152; cfim|PHH49965.1	Uncharacterized protein	A0A2C5WUU6
	8: cluster9168	cfnz|g6688; cfim|PHH50089.1	Uncharacterized protein	A0A2C5W995
	9: cluster9169	cfnz|g3778; cfim|PHH50198.1	1-phosphatidylinositol phosphodiesterase	A0A2C5WMP0
	10: cluster9170	cfnz|g6184; cfim|PHH50561.1	Uncharacterized protein	A0A2C5WY34
	11: cluster9171	cfnz|g3915; cfim|PHH51280.1	Meiotically up-regulated protein	A0A0F8CUG9
	12: cluster9172	cfnz|g5822; cfim|PHH51450.1	Uncharacterized protein	A0A2C5X0F5
	13: cluster9173	cfnz|g6630; cfim|PHH51984.1	Alginate lyase	A0A0F8B1G9
	14: cluster9174	cfnz|g538; cfim|PHH53454.1	Uncharacterized protein	A0A2C5X743
	15: cluster9175	cfnz|g4394; cfim|PHH53838.1	Uncharacterized protein	A0A2C5X1Z7
	16: cluster9176	cfnz|g947; cfim|PHH54698.1	Putative endo-1,3(4)-beta-glucanase	A0A2C5X4G3
	17: cluster9177	cfnz|g7203; cfim|PHH54723.1	Uncharacterized protein	A0A2C5XCH0
	18: cluster9178	cfnz|g6943; cfim|PHH55059.1	Uncharacterized protein	A0A2C5X1Y9
	19: cluster9179	cfnz|g2093; cfim|PHH55535.1	Uncharacterized protein	A0A2C5XF87
	20: cluster9180	cfnz|g6781; cfim|PHH55758.1	Uncharacterized protein	A0A2C5X8V3
	21: cluster9181	cfnz|g5554; cfim|PHH56195.1	Quinic acid utilization activator	A0A2C5XIY2
kiwifruit[b]	1: cluster7776	c400|g3773; c413|g3280; c415|g6366; c427|g5553; c458|g3960; c470|g5578; c668|g7054; c673|g4986; c674|g6527; c682|g5653; c688|g5063; c701|g5677; c922|g6183	Uncharacterized protein	A0A0F8D7W3
	2: cluster7777	c400|g1125; c413|g2372; c415|g1452; c427|g683; c458|g1398; c470|g895; c668|g1335; c673|g2162; c674|g866; c682|g917; c688|g328; c701|g965; c922|g2432	Uncharacterized protein	A0A2C5WVM7
	3: cluster7780	c400|g4775; c413|g3242; c415|g6203; c427|g5331; c458|g5872; c470|g3692; c668|g5133; c673|g5443; c674|g4858; c682|g5378; c688|g4658; c701|g4536; c922|g5678	Uncharacterized protein	A0A2C5XF46
ʻōhiʻa[c]	1: cluster8502	c184|g1402; c186|g678; c191|g2936	no hit	no hit

[a] cfnz: CFNZ and cfim: *Ceratocystis fimbriata* type isolate.

[b] c400: LPF1400, c413: LPF1413, c415: LPF1415, c427: LPF1427, c458: LPF1458, c470: LPF1470, c668: LPF1668, c673: LPF1673, c674: LPF1674, c682: LPF1682, c688: LPF1688, c701: LPF1701, and c922: LPF1922.

[c] c184: C4184, c186: C4186, and c191: C4191.

[d] Automatic annotation by BLASTp against of NCBI protein non-redundant - UniProt Knowledgebase (UniProtKb).

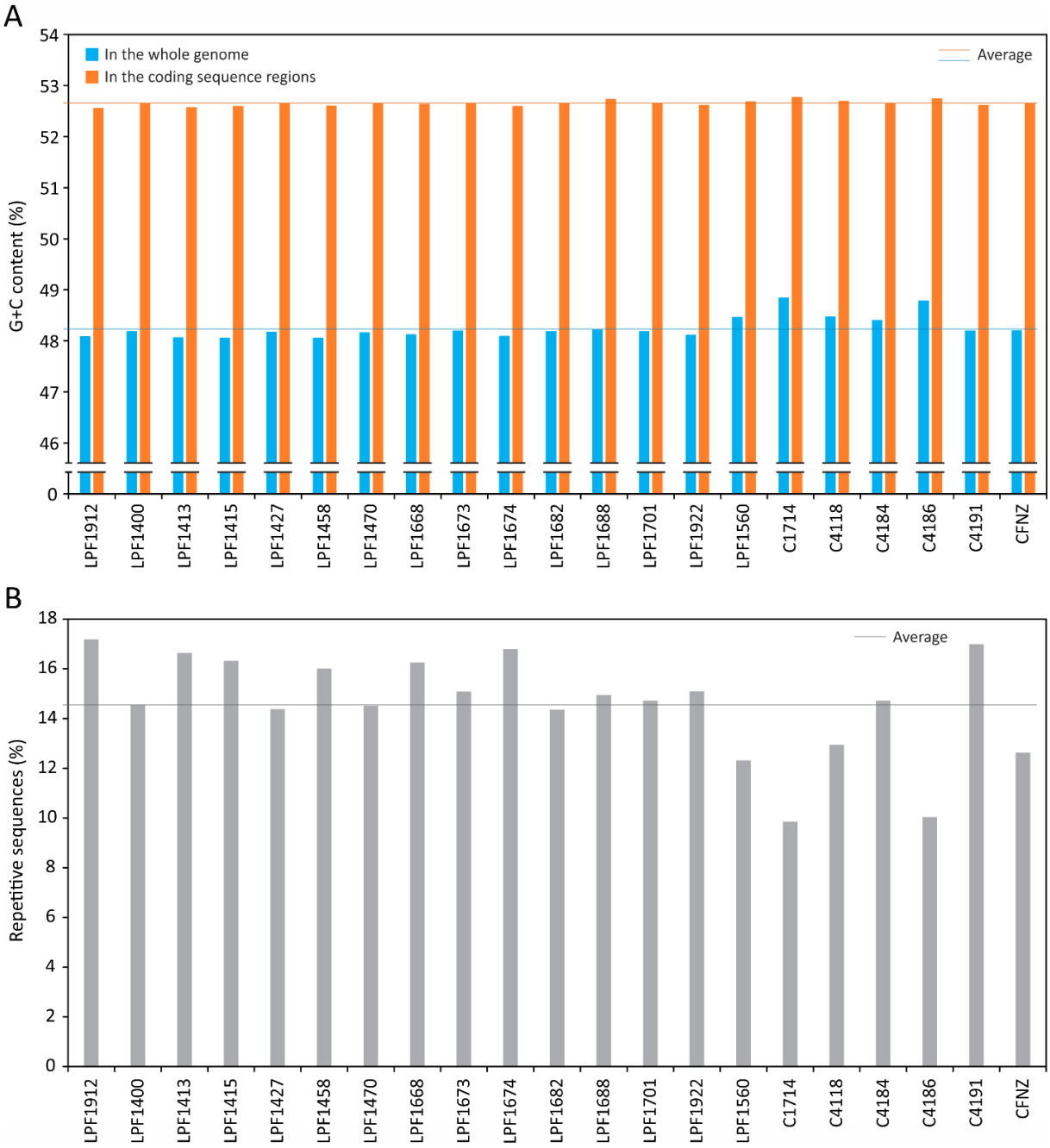

Fig. 1 Genome annotation of the assembled nuclear genome for 21 pathogens of the *Ceratocystis fimbriata* complex. **A.** G+C content in the whole genome and coding-sequence regions with an average of 48.28 and 52.68%, respectively. **B.** The proportion of total repetitive sequences in the whole nuclear genome.

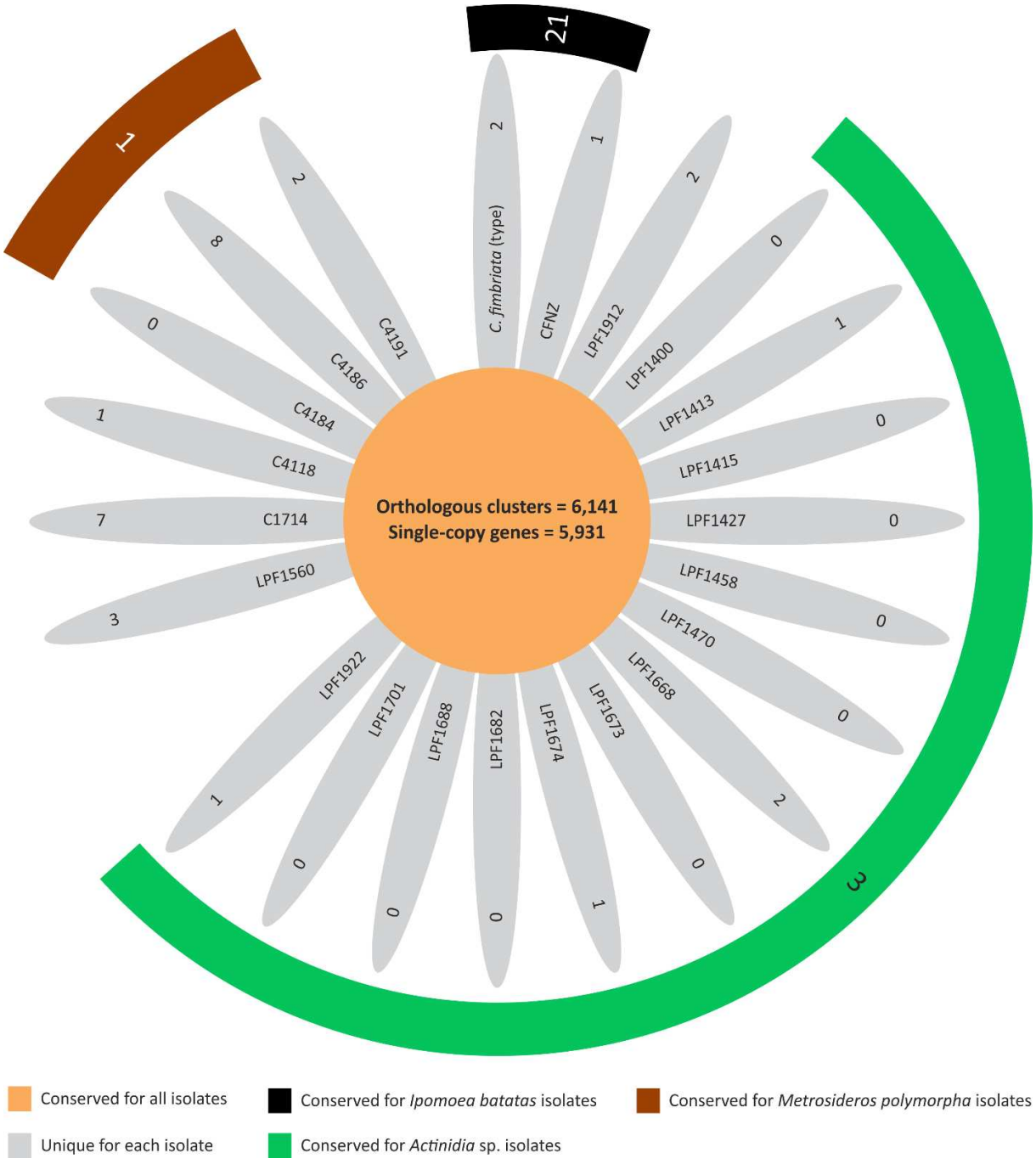

Fig. 2 OrthoMCL clustering analysis showing unique and conserved orthologous clusters genes among 22 pathogens of the *Ceratocystis fimbriata* complex.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 18 Jul 2019.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Barnes I, Fourie A, Wingfield MJ, Harrington TC, McNew DL, Sugiyama LS, Luiz BC, Heller WP, Keith LM (2018) New ceratocystis species associated with rapid death of metrosideros polymorpha in hawai`i. Persoonia Mol Phylogeny Evol Fungi 40:154–181. https://doi.org/10.3767/persoonia.2018.40.07

Barnes I, Roux J, Wingfield BD, Dudzinski MJ, Old KM, Wingfield MJ (2003) Ceratocystis pirilliformis, a new species from Eucalyptus nitens in Australia. Mycologia 95:865–871. https://doi.org/10.1080/15572536.2004.11833046

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, Grimwood J, Pérez G, Pisabarro AG, Grigoriev I V., Stajich JE, Ramírez L (2016) Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. PLoS Genet 12:e1006108. https://doi.org/10.1371/journal.pgen.1006108

Chan PP, Lowe TM (2019) tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

Choi JJ, Kim SH (2017) A genome Tree of Life for the Fungi kingdom. Proc Natl Acad Sci U S A 114:9391–9396. https://doi.org/10.1073/pnas.1711939114

Coil D, Jospin G, Darling AE (2015) A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31:587–589. https://doi.org/10.1093/bioinformatics/btu661

de Beer ZW, Duong TA, Barnes I, Wingfield BD, Wingfield MJ (2014) Redefining Ceratocystis and allied genera. Stud Mycol 79:187–219. https://doi.org/10.1016/j.simyco.2014.10.001

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Engelbrecht CJB, Harrington TC (2005) Intersterility, morphology and taxonomy of Ceratocystis fimbriata on sweet potato, cacao and sycamore. Mycologia 97:57–69. https://doi.org/10.3852/mycologia.97.1.57

Ferreira EM, Harrington TC, Thorpe DJ, Alfenas AC (2010) Genetic diversity and interfertility among highly differentiated populations of Ceratocystis fimbriata in Brazil. Plant Pathol 59:721–735. https://doi.org/10.1111/j.1365-3059.2010.02275.x

Ferreira FA, Demuner AMM, Demuner NL, Pigato S (1999) Murcha-de-Ceratocystis em eucalipto no Brasil. Fitopatol Bras 24:p.284.

Ferreira MA, Harrington TC, Piveta G, Alfenas AC (2017) Genetic variability suggests that three populations of Ceratocystis fimbriata are responsible for the Ceratocystis wilt epidemic on kiwifruit in Brazil. Trop Plant Pathol 42:86–95. https://doi.org/10.1007/s40858-017-0131-y

Fourie A, Wingfield MJ, Wingfield BD, Barnes I (2015) Molecular markers delimit cryptic species in Ceratocystis sensu stricto. Mycol Prog 14:1020. https://doi.org/10.1007/s11557-014-1020-0

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: Ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17:333–351.

Guimarães LM da S, Titon M, Lau D, Rosse LN, Oliveira LSS, Rosado CCG, Christo GGO, Alfenas AC (2010) Eucalyptus pellita as a source of resistance to rust, ceratocystis wilt and leaf blight. Crop Breed Appl Biotechnol 10:124–131.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Harrington TC, Kazmi MR, Al-Sadi AM, Ismail SI (2014) Intraspecific and intragenomic variability of ITS rDNA sequences reveals taxonomic problems in Ceratocystis fimbriata sensu stricto. Mycologia 106:224–242. https://doi.org/10.3852/106.2.224

Harrington TC, Rizzo DM (1999) Defining species in the fungi. In: Worrall JJ (ed) Structure and Dynamics of Fungal Populations. Kluwer Academic, Dordrecht, the Netherlands, pp 43–71.

Heath RN, Wingfield MJ, Wingfield BD, Meke G, Mbaga A, Roux J (2009) Ceratocystis species on Acacia mearnsii and Eucalyptus spp. in eastern and southern Africa including six new species. Fungal Divers 34:41–67.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2016) BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32:767–769. https://doi.org/10.1093/bioinformatics/btv661

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-genome annotation with BRAKER. In: Methods in Molecular Biology. Humana Press Inc., pp 65–95.

Johnson JA, Harrington TC, Engelbrecht CJB (2005) Phylogeny and taxonomy of the North American clade of the Ceratocystis fimbriata complex. Mycologia 97:1067–1092. https://doi.org/10.3852/mycologia.97.5.1067

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467. https://doi.org/10.1159/000084979

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res 24:1384–1395. https://doi.org/10.1101/gr.170720.113

Keith LM, Hughes RF, Sugiyama LS, Heller WP, Bushe BC, Friday JB (2015) First report of ceratocystis wilt on ‘Ōhi’a (Metrosideros polymorpha). Plant Dis. 99:1276.

Krishnan P, Meile L, Plissonneau C, Ma X, Hartmann FE, Croll D, McDonald BA, Sánchez-Vallet A (2018) Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. BMC Biol 16:78. https://doi.org/10.1186/s12915-018-0543-2

Li L, Stoeckert CJ, Roos DS, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–2189. https://doi.org/10.1101/gr.1224503

Li Q, Harrington TC, McNew D, Li J (2017) Ceratocystis uchidae, a new species on Araceae in Hawaii and Fiji. Mycoscience 58:398–412. https://doi.org/10.1016/j.myc.2017.06.001

Li Q, Harrington TC, McNew D, Li J, Huang Q, Somasekhara YM, Alfenas AC (2016) Genetic Bottlenecks for Two Populations of *Ceratocystis fimbriata* on Sweet Potato and Pomegranate in China. Plant Dis 100:2266–2274. https://doi.org/10.1094/PDIS-03-16-0409-RE

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18

Mafia RG, Ferreira MA, Zauza EA V., Silva JF, Colodette JL, Alfenas AC (2013) Impact of Ceratocystis wilt on eucalyptus tree growth and cellulose pulp yield. For Pathol 43:379–385. https://doi.org/10.1111/efp.12041

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764–70. https://doi.org/10.1093/bioinformatics/btr011

Mbenoun M, Wingfield MJ, Begoude Boyogueno AD, Wingfield BD, Roux J (2014) Molecular phylogenetic analyses reveal three new Ceratocystis species and provide evidence for geographic differentiation of the genus in Africa. Mycol Prog 13:219–240. https://doi.org/10.1007/s11557-013-0907-5

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, Luthe DS, Bridges SM, Burgess SC (2006) AgBase: A functional genomics resource for agriculture. BMC Genomics 7:229. https://doi.org/10.1186/1471-2164-7-229

Molano EPL, Cabrera OG, Jose J, do Nascimento LC, Carazzolle MF, Teixeira PJPL, Alvarez JC, Tiburcio RA, Tokimatu Filho PM, de Lima GMA, Guido RVC, Corrêa TLR, Leme AFP, Mieczkowski P, Pereira GAG (2018) Ceratocystis cacaofunesta genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). BMC Genomics 19:58. https://doi.org/10.1186/s12864-018-4440-4

Mori T, Jung H-Y, Maejima K, Hirata H, Himeno M, Hamamoto H, Namba S (2008) Magnaporthe oryzae endopolygalacturonase homolog correlates with density-dependent conidial germination. FEMS Microbiol Lett 280:182–188. https://doi.org/10.1111/j.1574-6968.2008.01062.x

Mortenson LA, Flint Hughes R, Friday JB, Keith LM, Barbosa JM, Friday NJ, Liu Z, Sowards TG (2016) Assessing spatial distribution, stand impacts and rate of Ceratocystis fimbriata induced ‘ōhi‘a (Metrosideros polymorpha) mortality in a tropical wet forest, Hawai‘i Island, USA. For Ecol Manage 377:83–92. https://doi.org/10.1016/j.foreco.2016.06.026

Muñoz JF, McEwen JG, Clay OK, Cuomo CA (2018) Genome analysis reveals evolutionary mechanisms of adaptation in systemic dimorphic fungi. Sci Rep 8:4473. https://doi.org/10.1038/s41598-018-22816-6

Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K (2019) Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci Rep 9:4307. https://doi.org/10.1038/s41598-019-40965-0

Nkuekam GK, Barnes I, Wingfield MJ, Roux J (2009) Distribution and population diversity of Ceratocystis pirilliformis in South Africa. Mycologia 101:17–25. https://doi.org/10.3852/07-171

Oeser B, Heidrich PM, Müller U, Tudzynski P, Tenberge KB (2002) Polygalacturonase is a pathogenicity factor in the Claviceps purpurea/rye interaction. Fungal Genet Biol 36:176–186.

https://doi.org/10.1016/s1087-1845(02)00020-8

Oliveira LSS, Harrington TC, Ferreira MA, Damacena MB, Al-Sadi AM, Al-Mahmooli IHS, Alfenas AC (2015) Species or Genotypes? Reassessment of Four Recently Described Species of the Ceratocystis Wilt Pathogen, *Ceratocystis fimbriata* , on *Mangifera indica*. Phytopathology 105:1229–1244. https://doi.org/10.1094/PHYTO-03-15-0065-R

Piveta G, Ferreira M, FB Muniz M, Valdetaro D, Valdebenito-Sanhueza R, Harrington T, Alfenas A (2016) *Ceratocystis fimbriata* on kiwifruit (*Actinidia* spp.) in Brazil. New Zeal J Crop Hortic Sci 44:13–24. https://doi.org/10.1080/01140671.2016.1143020

Pryszcz LP, Gabaldón T (2016) Redundans: An assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 44:e113. https://doi.org/10.1093/nar/gkw294

Quoc NB, Chau NNB (2017) The Role of Cell Wall Degrading Enzymes in Pathogenesis of Magnaporthe oryzae. Curr Protein Pept Sci 18:1019–1034. https://doi.org/10.2174/1389203717666160813164955

Razali NM, Cheah BH, Nadarajah K (2019) Transposable elements adaptive role in genome plasticity, pathogenicity and evolution in fungal phytopathogens. Int. J. Mol. Sci. 20:3597.

Rizatto S, de Araújo Batista CE, Bajay MM, Sigrist MS, Ito MF, Monteiro M, Cavallari MM, Pinheiro JB, Zucchi MI (2010) A new set of microsatellite markers for the genetic characterization of Ceratocystis fimbriata, an economically important plant pathogen. Conserv Genet Resour 2:55–58. https://doi.org/10.1007/s12686-009-9144-2

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An KD, Bai FY, Barreto RW, Begerow D, Bergeron MJ, Blackwell M, Boekhout T, Bogale M, Boonyuen N, Burgaz AR, Buyck B, Cai L, Cai Q, Cardinali G, Chaverri P, Coppins BJ, Crespo A, Cubas P, Cummings C, Damm U, de Beer ZW, de Hoog GS, Del-Prado R, Dentinger B, Diéguez-Uribeondo J, Divakar PK, Douglas B, Dueñas M, Duong TA, Eberhardt U, Edwards JE, Elshahed MS, Fliegerova K, Furtado M, García MA, Ge ZW, Griffith GW, Griffiths K, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Guo LD, Hagen F, Hambleton S, Hamelin RC, Hansen K, Harrold P, Heller G, Herrera C, Hirayama K, Hirooka Y, Ho HM, Hoffmann K, Hofstetter V, Högnabba F, Hollingsworth PM, Hong SB, Hosaka K, Houbraken J, Hughes K, Huhtinen S, Hyde KD, James T, Johnson EM, Johnson JE, Johnston PR, Jones EBG, Kelly LJ, Kirk PM, Knapp DG, Kõljalg U, Kovács GM, Kurtzman CP, Landvik S, Leavitt SD, Liggenstoffer AS, Liimatainen K, Lombard L, Luangsa-ard JJ, Lumbsch HT, Maganti H, Maharachchikumbura SSN, Martin MP, May TW, McTaggart AR, Methven AS, Meyer W, Moncalvo JM, Mongkolsamrit S, Nagy LG, Nilsson RH, Niskanen T, Nyilasi I, Okada G, Okane I, Olariaga I, Otte J, Papp T, Park D, Petkovits T, Pino-Bodas R, Quaedvlieg W, Raja HA, Redecker D, Rintoul TL, Ruibal C, Sarmiento-Ramírez JM, Schmitt I, Schüßler A, Shearer C, Sotome K, Stefani FOP, Stenroos S, Stielow B, Stockinger H, Suetrong S, Suh SO, Sung GH, Suzuki M, Tanaka K, Tedersoo L, Telleria MT, Tretter E, Untereiner WA, Urbina H, Vágvölgyi C, Vialle A, Vu TD, Walther G, Wang QM, Wang Y, Weir BS, Weiß M, White MM, Xu J, Yahr R, Yang ZL, Yurkov A, Zamora JC, Zhang N, Zhuang WY, Schindel D (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A 109:6241–6246. https://doi.org/10.1073/pnas.1117018109

Shieh MT, Brown RL, Whitehead MP, Cary JW, Cotty PJ, Cleveland TE, Dean RA (1997) Molecular genetic evidence for the involvement of a specific polygalacturonase, P2c, in the invasion and spread of Aspergillus flavus in cotton bolls. Appl Environ Microbiol 63:3548–3552.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics

31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simpson MC, Wilken PM, Coetzee MPA, Wingfield MJ, Wingfield BD (2013) Analysis of microsatellite markers in the genome of the plant pathogen Ceratocystis fimbriata. Fungal Biol 117:545–555. https://doi.org/10.1016/j.funbio.2013.06.004

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637–644. https://doi.org/10.1093/bioinformatics/btn013

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33:W465–W467. https://doi.org/10.1093/nar/gki458

Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62

Steimel J, Engelbrecht CJB, Harrington TC (2004) Development and characterization of microsatellite markers for the fungus Ceratocystis fimbriata. Mol Ecol Notes 4:215–218. https://doi.org/10.1111/j.1471-8286.2004.00621.x

Tan MH, Austin CM, Hammer MP, Lee YP, Croft LJ, Gan HM (2018) Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (Amphiprion ocellaris) genome assembly. Gigascience 7:1–6. https://doi.org/10.1093/gigascience/gix137

Thorpe DJ, Harrington TC, Uchida JY (2005) Pathogenicity, internal transcribed spacer-rDNA variation, and human dispersal of Ceratocystis fimbriata on the family Araceae. Phytopathology 95:316–323. https://doi.org/10.1094/PHYTO-95-0316

Uchida JY, Aragaki M (1979) Ceratocystis blight of Syngonium podophyllum. Plant Dis Report 63:1053–1056.

Valdetaro DCOF, Harrington TC, Oliveira LSS, Guimarães LMS, McNew DL, Pimenta LVA, Gonçalves RC, Schurt DA, Alfenas AC (2019) A host specialized form of Ceratocystis fimbriata causes seed and seedling blight on native Carapa guianensis (andiroba) in Amazonian rainforests. Fungal Biol 123:170–182. https://doi.org/10.1016/j.funbio.2018.12.001

van der Nest MA, Bihon W, De Vos L, Naidoo K, Roodt D, Rubagotti E, Slippers B, Steenkamp ET, Wilken PM, Wilson A, Wingfield MJ, Wingfield BD (2014) Draft genome sequences of Diplodia sapinea, Ceratocystis manginecans, and Ceratocystis moniliformis. IMA Fungus 5:135–140. https://doi.org/10.5598/imafungus.2014.05.01.13

van der Nest MA, Steenkamp ET, Roodt D, Soal NC, Palmer M, Chan WY, Wilken PM, Duong TA, Naidoo K, Santana QC, Trollip C, De Vos L, van Wyk S, McTaggart AR, Wingfield MJ, Wingfield BD (2019) Genomic analysis of the aggressive tree pathogen Ceratocystis albifundus. Fungal Biol 123:351–363. https://doi.org/10.1016/j.funbio.2019.02.002

van Wyk M, Al Adawi AO, Khan IA, Deadman ML, Al Jahwari AA, Wingfield BD, Ploetz R, Wingfield MJ (2007) Ceratocystis manginecans sp. nov., causal agent of a destructive mango wilt disease in Oman and Pakistan. Fungal Divers 27:213–230.

van Wyk M, Roux J, Nkuekam GK, Wingfield BD, Wingfield MJ (2012) Ceratocystis eucalypticola sp. nov. from Eucalyptus in South Africa and comparison to global isolates from this tree. IMA Fungus 3:45–58. https://doi.org/10.5598/imafungus.2012.03.01.06

van Wyk M, Wingfield BD, Al-Adawi AO, Rossetto CJ, Ito MF, Wingfield MJ (2011a) Two new Ceratocystis species associated with mango disease in Brazil. Mycotaxon 117:381–404.

https://doi.org/10.5248/117.381

van Wyk M, Wingfield BD, Marin M, Wingfield MJ (2010) New Ceratocystis species infecting coffee, cacao, citrus and native trees in Colombia. Fungal Divers 40:103–117. https://doi.org/10.1007/s13225-009-0005-9

van Wyk M, Wingfield BD, Mohali S, Wingfield MJ (2009) Ceratocystis fimbriatomima, a new species in the C. fimbriata sensu lato complex isolated from Eucalyptus trees in Venezuela. Fungal Divers 34:175–185.

van Wyk M, Wingfield BD, Wingfield MJ (2011b) Four new Ceratocystis spp. associated with wounds on Eucalyptus, Schizolobium and Terminalia trees in Ecuador. Fungal Divers 46:111–131. https://doi.org/10.1007/s13225-010-0051-3

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC (2017) GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33:2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963

Webster RK, Butler EE (1967) A morphological and biological concept of the species Ceratocystis fimbriata. Can J Bot 45:1457–1468. https://doi.org/10.1139/b67-149

Wingfield BD, Barnes I, Wilhelm de Beer Z, De Vos L, Duong TA, Kanzi AM, Naidoo K, Nguyen HDT, Santana QC, Sayari M, Seifert KA, Steenkamp ET, Trollip C, van der Merwe NA, van der Nest MA, Markus Wilken P, Wingfield MJ (2015) IMA Genome-F 5: Draft genome sequences of Ceratocystis eucalypticola, Chrysoporthe cubensis, C. deuterocubensis, Davidsoniella virescens, Fusarium temperatum,Graphilbum fragrans, Penicillium nordicum, and Thielaviopsis musarum. IMA Fungus 6:493–506. https://doi.org/10.5598/imafungus.2015.06.02.13

Wingfield BD, Fourie A, Simpson MC, Bushula-Njah VS, Aylward J, Barnes I, Coetzee MPA, Dreyer LL, Duong TA, Geiser DM, Roets F, Steenkamp ET, van der Nest MA, van Heerden CJ, Wingfield MJ (2019) IMA Genome-F 11: Draft genome sequences of Fusarium xylarioides, Teratosphaeria gauchensis and T. zuluensis and genome annotation for Ceratocystis fimbriata. IMA Fungus 10:13. https://doi.org/10.1186/s43008-019-0013-7

Yang Y, Ye Q, Li K, Li Z, Bo X, Li Z, Xu Y, Wang S, Wang P, Chen H, Wang J (2017) Genomics and Comparative Genomic Analyses Provide Insight into the Taxonomy and Pathogenic Potential of Novel Emmonsia Pathogens. Front Cell Infect Microbiol 7:105. https://doi.org/10.3389/fcimb.2017.00105

Ye X, Liu H, Jin Y, Guo M, Huang A, Chen Q, Guo W, Zhang F, Feng L (2017) Transcriptomic Analysis of Calonectria pseudoreteaudii during Various Stages of Eucalyptus Infection. PLoS One 12:e0169598. https://doi.org/10.1371/journal.pone.0169598

Zauza EA V., Alfenas AC, Harrington TC, Mizubuti ES, Silvai JF (2004) Resistance of *Eucalyptus* Clones to *Ceratocystis fimbriata*. Plant Dis 88:758–760. https://doi.org/10.1094/PDIS.2004.88.7.758

CHAPTER 4 - COMPARATIVE GENOMIC AND TRANSCRIPTOMIC ANALYSES REVEAL DIFFERENT PATHOGENICITY-RELATED GENES AMONG THREE EUCALYPTUS FUNGAL PATHOGENS

Samuel A. Santos[a,c], Pedro M. P. Vidigal[b], Amali Thrimawithana[c], Blanca M. L. Betancourth[a], Lúcio M. S. Guimarães[a], Matthew D. Templeton[c], Acelino C. Alfenas[a,*]

[a] Laboratory of Forest Pathology, Department of Plant Pathology, Universidade Federal de Viçosa, Minas Gerais state, Brazil.

[b] Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Minas Gerais State, Brazil.

[c] The New Zealand Institute for Plant and Food Research Limited, Auckland 1142, New Zealand.

* Corresponding author. Department of Plant Pathology, Instituto de Biotecnologia Aplicada à agropecuária-BIOAGRO, Universidade Federal de Viçosa, Av. P.H. Rolfs s/n, Campus Universitário, 36570-900, Viçosa, MG, Brazil. Tel.: +55 31 3612 2428. E-mail address: *aalfenas@ufv.br* (*A.C., Alfenas*).

Obs: This chapter has been published as an original article at Fungal Genetics and Biology. The full publication can be accessed through the following link: https://doi.org/10.1016/j.fgb.2019.103332

Abstract *Ceratocystis fimbriata* is an important plant pathogen known to cause Ceratocystis Wilt (CW), a prevalent fungal disease known to affect *Eucalyptus* spp. plantations in Brazil. To better understand the molecular mechanisms related to pathogenicity in eucalyptus, we generated a high-quality assembly and annotation of the *Ce. fimbriata* LPF1912 isolate (LPF1912) genome, as well as the first transcriptome of LPF1912 from 16 eucalyptus clones at three infection incubation periods (12, 18, and 24h). The LPF1912 genome assembly contains 805 scaffolds, totaling 31.8 Mb, with 43% of the genome estimated to be coding sequence comprised of 7,390 protein-coding genes of which 626 (8.5%) were classified as secreted proteins, 120 ribosomal RNAs, and 532 transfer RNAs. Comparative genomic analysis among

three eucalyptus fungal pathogens (*Ce. fimbriata, Ce. eucalypticola*, and *Calonectria pseudoreteaudii*), showed high similarity in the proteome (21.81%) and secretome (52.01%) of LPF1912 and *Ce. eucalypticola*. GO annotation of pathogenicity-related genes of LPF1912 and *Ce. eucalypticola,* revealed enrichment in cell wall degrading enzymes (CWDEs), and lipid/cutin metabolism for *Ca. pseudoreteaudii*. Additionally, a transcriptome analysis between resistant and susceptible eucalyptus clones to CW infection indicated that a majority (11) of LPF1912 differentially expressed genes had GO terms associated with enzymatic functions, such as the polygalacturonase gene family, confirming the crucial role of CWDEs for *Ce. fimbriata* pathogenicity. Finally, our genomic and transcriptomic analysis approach provides a better understanding of the mechanisms involved in *Ce. fimbriata* pathogenesis, as well as a framework for further studies.

Keywords: Ceratocystis Wilt. *Ceratocystis fimbriata*. *Ceratocystis eucalypticola*. *Calonectria pseudoreteaudii*. RNA-Seq.

INTRODUCTION

Eucalyptus (*Eucalyptus* spp. and hybrids) is considered an important forest crop around the world and the most commonly grown forest tree in Brazil (IBÁ, 2019). Eucalyptus plantations support a multi-billion industry based on cellulose pulp and paper, charcoal for the steel industry, and solid products (Grattapaglia and Kirst, 2008). As the planted areas have expanded and considering climate change, fungal diseases such as Ceratocystis Wilt (CW) and Calonectria Leaf Blight (CLB) have threatened the eucalyptus plantations. In Brazil, CW is caused by *Ceratocystis fimbriata* (Ferreira et al., 1999), however, in South Africa, the same disease has been associated with *Ceratocystis eucalypticola* (van Wyk et al., 2012). *Ce. eucalypticola* was established as a new species within *Ce. fimbriata* complex (van Wyk et al., 2012), though its taxonomy remains open to debate (Fourie et al., 2015). In addition, CLB is caused by different *Calonectria* species including *Ca. pteridis* complex and *Ca. candelabra* complex, most frequent in Brazil (Alfenas et al., 2015), as well as *Ca. pseudoreteaudii*, which is the most widely distributed and aggressive species in China (Chen et al., 2013, 2011; Ye et al., 2017).

Both diseases have caused significant economic losses but show different symptoms on the eucalyptus tree. For CW, commonly the infection begins at the roots or at the base of the stem, with dark brown to black radial streaks of the woody xylem tissue observed as

colonization advances and causes cell death in several tissues types such as ray parenchyma, vascular cambium, phloem and phelloderm (Ferreira et al., 2006). The visible symptoms in susceptible eucalyptus genotype infected by *Ce. fimbriata* are wilt of the canopy, branch death, and consequently death of the entire tree (Alfenas et al., 2009; Ferreira et al., 2006, 2013; Roux and Wingfield, 2009). On the other hand, CLB infects eucalyptus leaves and branches causing leaf and shoot blight, minicankers on the branches, and defoliation (Chen et al., 2011).

As CW is a lethal disease in eucalyptus (Baker et al., 2003; Ferreira et al., 1999; Guimarães et al., 2010; Mafia et al., 2013; Zauza et al., 2004), the selection and planting of resistant eucalyptus genotypes have been the major strategies to manage the disease in Brazil (Rosado et al., 2010; Zauza et al., 2004). However, resistance to CW is a variable trait that is dependent on the pathogen isolate (Oliveira et al., 2015). Due to the high genetic variability found in the *Ce. fimbriata* population (Ferreira et al., 2010; Oliveira et al., 2015), it is difficult to obtain a plant genotype with robust resistance in the field to the pathogen population. The demand for new strategies to control this disease requires an improved understanding of the molecular mechanisms that evolved during *Ce. fimbriata* interaction with eucalyptus plants.

Whole genome sequencing (WGS) and RNA-Seq are powerful tools that can be applied to study the mechanisms involved in plant-pathogen interactions. WGS has already been used to infer the evolution of pathogens, identify pathogenicity factors, virulence genes (Cantu et al., 2011), and candidate effector proteins (Joly et al., 2010). Additionally, transcriptome sequencing by RNA-Seq has provided fundamental information that has led to gene discovery, quantification of gene expression levels (Huynh et al., 2015; Kim et al., 2014), secretome analysis, and prediction of fungal candidate effectors (Bruce et al., 2014; Liu et al., 2015; Meinhardt et al., 2014). Thus, a complete characterization of *Ce. fimbriata* genome and a comparative genomic analysis with different eucalyptus fungal pathogens can provide crucial information to a better understanding of its biology. Furthermore, a contrasting analysis of differentially expressed genes during *Ce. fimbriata* infection on plants from resistant and susceptible eucalyptus genotypes may reveal which genes are related to *Ce. fimbriata* pathogenicity in eucalyptus. Such information may be useful to the breeding programs, leading to the development of eucalyptus resistant genotypes to CW.

In order to understand the molecular mechanisms related to *Ce. fimbriata* pathogenicity on *Eucalyptus* spp., we present (i) a high-quality assembly and annotation of *Ce. fimbriata* LPF1912 genome; (ii) analysis of the transcriptome of *Ce. fimbriata* LPF1912 during its interaction with different eucalyptus genotypes; and (iii) identification of *Ce. fimbriata*

LPF1912 pathogenicity-related genes through comparisons with the genomes of *Ceratocystis eucalypticola* (Wingfield et al., 2015) and *Calonectria pseudoreteaudii* (Ye et al., 2017).

MATERIAL AND METHODS

Fungal isolate and DNA extraction

To characterize the whole genome we used the LPF1912 isolate (other codes: C1451 and SBS-1) of *Ceratocystis fimbriata* sensu stricto. This isolate was collected from infected eucalyptus trees in Bahia State, Brazil, the same area where the disease was first described (Ferreira et al., 1999). This isolate has shown to be highly aggressive in several studies (Baker et al., 2003; Guimarães et al., 2010; Zauza et al., 2004) and is frequently used in resistance screening experiments (Rosado et al., 2010). DNA was extracted using Wizard Genomic DNA Purification Kit® (Promega, Madison, United States) from mycelia collected from a pure single spore culture of LPF1912, grown in Petri plate, containing 2% malt extract agar (MEA, 20 g L^{-1} Agar, 20 g L^{-1} malt extract) at 25°C for 20 days. The mycelia were added in a microcentrifuge tube of 2 mL containing two tungsten carbide beads. Mycelia were disrupted in a Qiagen Tissuelyser for 2 min at 30 Hz frequency and the DNA was extracted according to the kit instructions.

Transcriptome experiment and RNA samples preparation

To evaluate the differentially expressed genes of *Ceratocystis fimbriata* during its interaction with eucalyptus plants, 16 clones of *Eucalyptus* [eight resistant (resistant *pool*) and eight susceptible (susceptible *pool*) to *Ce. fimbriata*] were used. The experiment was conducted in randomized blocks with five replicates per clone. Thus, 40 eucalyptus plants formed each eucalyptus *pool* (resistant and susceptible).

The eucalyptus plants were inoculated with the LPF1912 isolate of *Ce. fimbriata* following the protocol as described by Zauza et al. (2004). Samples of the stem, from the region next to the inoculation point, were collected at 12, 18 and 24 h after inoculation, and immediately put in liquid nitrogen and then stored at -80°C. The samples were mechanically ground into a fine powder in liquid nitrogen, and the total RNA was isolated using a RNeasy Plant Mini Kit (QIAGEN, Texas, USA) in accordance with the manufacturer's guidelines. The RNA was resuspended in RNase-free water, and RNA integrity and quality were assessed using

an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). After treated with DNase (RQ1 RNase free DNase Promega®), the total RNA samples were sent to sequencing.

Genome and transcriptome sequencing

For genome sequencing, a paired-end library with an average insert size of 300 bp was prepared from approximately 5 µg of DNA and sequenced using an Illumina MiSeq platform according to the manufacturer's instructions (Illumina, San Diego, CA) at the Australian Genome Research Facility, Melbourne, Australia (http://www.agrf.org.au/). High sequencing coverage of 200× was requested.

For RNA sequencing, the rRNA in total RNA was depleted by Ribo-Zero kit (Illumina, San Diego, CA, USA). The enriched mRNA samples were subjected to Illumina cDNA library construction using TruSeq stranded mRNA (Microbe) kit (Illumina, San Diego, CA, USA). The RNA was purified, fragmented, and primed for cDNA synthesis. The RNA fragments were transcribed into first strand cDNA using reverse transcriptase and random hexamers, followed by second strand cDNA synthesis. These fragments were prepared for sequencing with an end-repair process and the addition of a single 'A' base at the 3' end. Paired-end adapters were ligated to the ends of these 3' adenylated cDNA fragments. Products were purified and enriched with PCR to create the final cDNA library (Ross-Davis et al., 2013). A total of six cDNA libraries were obtained and sequenced by RNA-Seq using Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA) at the Central Laboratory of High Performance Technologies of the University of Campinas, Sao Paulo, Brazil (https://www.lactad.unicamp.br/en).

Genome assembly, size estimate, annotation, and evaluation

The sequenced reads were subjected to quality control checks using FastQC version 0.11.5 (Andrews, 2010) and trimmed for quality (Q20 threshold) using AfterQC (Chen et al., 2017). The trimmed reads were *de novo* assembled into scaffolds, testing different k-mers, and using different assemblers, such as A5-Miseq (Coil et al., 2015), SPAdes (Bankevich et al., 2012), CLC Genomics Workbench, and SOAPdenovo2 (Luo et al., 2012). We performed a BLASTn of all assembled scaffolds against a bacteria-virus database to identify and eliminate potential contaminates.

Trimmed reads were also used to estimate the genome size based on k-mer distribution and depth as described by Tan et al. (2018). First, k-mer counting was performed with jellyfish

v2.2.10 (Marçais and Kingsford, 2011). Histograms of k-mer frequency distributions of 17-, 21-, and 25-mers were processed by GenomeScope (Vurture et al., 2017).

Protein-coding genes were predicted using AUGUSTUS version 3.3 (Stanke and Morgenstern, 2005), which was trained using reference sequences of genomes of *Ceratocystis* species available in GenBank (Accession Numbers: APWK00000000.3; LBBL00000000.1; LJOA00000000.1; MAOA00000000.2; and PEJQ00000000.1). To evaluate gene prediction accuracy, the completeness of the gene prediction was assessed using the Benchmarking Universal Single-Copy Orthologs, named BUSCO (Simão et al., 2015). BUSCO was performed on all predicted genes, making use of the Ascomycota lineage dataset (ascomycota_odb9).

The predicted genes were functionally annotated through similarity searches using BLAST version 2.6.0 (Altschul et al., 1990). In these searches, the encoded proteins were aligned to the sequences of NCBI protein non-redundant database (NCBInr), UniProt Knowledgebase (UniProtKb), using protein BLAST (BLASTp) (cutoff for significant hits: E-value of 1e-10). Putative gene functions were classified according to the Eukaryotic Orthologous Groups (KOG) database using Reverse Position-Specific BLAST (RPS-BLAST), Additionally, gene ontology (GO) terms were assigned to the proteins using Blast2GO (Conesa et al., 2005). Ribosomal genes were identified by Barrnap version 0.9 (https://github.com/tseemann/barrnap), using Eukaryota database, and transporter RNA genes were identified using tRNAscan-SE version 2.0 (Chan and Lowe, 2019). Additionally, RepeatMasker version open-4.0.8 (http://www.repeatmasker.org/) was used to identify repetitive sequences using Fungi database of RepBase library (v20170127) (Jurka et al., 2005).

Secretome prediction

Isolation of the secreted proteins from the predicted proteome was through a three-step process. First, the proteins containing signal peptides were selected by SignalP version 4.0 (Petersen et al., 2011) (cutoff for discrimination score: D-Score = 0.45). Proteins flagged as related to a secretory pathway signal peptide were selected by TargetP version 1.0 (Emanuelsson et al., 2000). Second, the proteins without transmembrane domains were selected by TMHMM version 2.0 (Krogh et al., 2001). Finally, GPI (glycosylphosphatidylinositol)-anchor proteins, which are surface proteins of the fungus cell, were identified by PredGPI (http://gpcr.biocomp.unibo.it/predgpi) (Pierleoni et al., 2008) and removed from the final set of the secreted proteins.

Differentially expressed genes (DEGs) of *Ceratocystis fimbriata* LPF1912 during its interaction with eucalyptus plants

Low quality sequenced reads were removed using AfterQC (Chen et al., 2017). Filtered reads were mapped to the transcript sequences of *Eucalyptus grandis* version 2.0 (Bartholomé et al., 2015; Myburg et al., 2014), available on Phytozome database (http://phytozome.jgi.doe.gov), using Bowtie2 (Langmead and Salzberg, 2012) by selecting the very-sensitive and local settings to remove the eucalyptus sequences. Unmapped reads were mapped to the transcript sequences of predicted genes of *Ce. fimbriata* LPF1912 using the same settings. Expression frequencies were calculated using Kallisto (Bray et al., 2016).

To evaluate the pathogenicity-related genes of *Ce. fimbriata* LPF1912 in eucalyptus we performed the following comparisons: resistant *pool* at 12 h vs resistant *pool* at 18 h after inoculation (R-12h vs R-18h); resistant *pool* at 12 h vs resistant *pool* at 24 h after inoculation (R-12h vs R-24h); resistant *pool* at 18 h vs resistant *pool* at 24 h after inoculation (R-18h vs R-24h); susceptible *pool* at 12 h vs susceptible *pool* at 18 h after inoculation (S-12h vs S-18h); susceptible *pool* at 12 h vs susceptible *pool* at 24 h after inoculation (S-12h vs S-24h); and susceptible *pool* at 18 h vs susceptible *pool* at 24 h after inoculation (S-18h vs S-24h). The DEGs were identified and quantified using DESeq2 R package (Love et al., 2014). For controlling the false discovery rate, P values were adjusted (padj) using the Hochberg and Benjamini tests. DEGs with padj ≤ 0.05 were considered as significant. The mapped reads also were *de novo* assembled using Trinity version 2.4.0 (Grabherr et al., 2011).

Prediction of pathogenicity-related proteins

To predict putative pathogenicity-related proteins, similarity searches were performed against Pathogen-Host Interaction (PHI) database (Winnenburg et al., 2006) version 4.5, using BLASTp tool with an E-value of 1e-10. We identified and grouped the pathogenicity-related genes in six groups: loss of pathogenicity, reduced virulence, unaffected virulence, increased virulence, lethal, and others. Additional, other searches were also conducted against the Database of Fungal Virulence Factors (DFVF) (Lu et al., 2012) with an E-value of 1e-10.

Comparative genomic analysis

To evaluate the pathogenicity mechanisms of *Ceratocystis fimbriata* LPF1912 during its infection in eucalyptus trees, the sequences of the genome of *Ce. eucalypticola* (Accession

Number: LJOA00000000.1) (Wingfield et al., 2015) and *Calonectria pseudoreteaudii* (Accession Number: MOCD00000000.1) (Ye et al., 2017) were downloaded from the Genome database (https://www.ncbi.nlm.nih.gov/genome).

Gene prediction and annotation, and secretome analysis of the downloaded sequences were done following the same methods as described above for *Ce. fimbriata* LPF1912. Additionally, the Markov clustering program OrthoMCL version 2.0 (Li et al., 2003) was used to identify orthologous genes among the three eucalyptus fungal pathogens species. Thus, predicted protein sequences were submitted to reciprocal similarity searches using BLASTp tool, considering an E-value of 1e-10 as threshold for the significant alignments. Putative orthologous were then clustered with OrthoMCL using an inflation value of 1.5 and similarity of 50%.

RESULTS

Features of *Ceratocystis fimbriata* LPF1912 genome

After trimming and filtering, a final data-set containing 11,821,390 reads was selected. The genome of *Ce. fimbriata* LPF1912 was assembled with four different softwares (A5-Miseq, SPAdes, CLC Genomics Workbench, and SOAPdenovo2) to identify the best assembly. The results of each assembly were compared in regards to genome size, N50, number of assembled scaffolds, and the size of the largest scaffold, with completeness assessment done using Benchmarking Universal Single-Copy Orthologs (BUSCO). The assembly using A5-miseq showed the best results (Supplementary Table S1). The assembly contains 788 scaffolds with a combined length of about 31.6 Mb, which is similar to the size estimated by k-mer analysis (30.2 - 30.6 Mb) (Supplementary Fig. S1A). The N50 of the assembled genome is 182 Kb and the scaffolds were sequenced with an average coverage of about 80-fold (Supplementary Fig. S1B).

Ceratocystis fimbriata LPF1912 genome was found to contain 8,042 putative coding genes of which 7,390 were coding DNA sequences (CDS) (Table 1). To assess the completeness of the genome, we compared the results of BUSCO analysis of *Ce. fimbriata* LPF1912 genome with other assemblies of *Ceratocystis* species (available at the GenBank database). *Ce. fimbriata* LPF1912 genome analysis using 1,315 BUSCO groups expected for Ascomycota, resulted in 97.1% complete groups (96.9% single, 0.2% duplicated), 2.0% fragmented, and 0.9% missed (Supplementary Fig. S1C), indicating that the assembled genome covers the majority of the coding regions.

Repetitive sequences represent 5.84% of the *Ce. fimbriata* LPF1912 genome, including retroelements [LINEs (0.11%) and LTR (1.19%)], DNA transposons (2.25%), small RNA (0.14%), and simple and low complexity repeat regions (2.14%) (Table 1; Supplementary Table S2). A total of 120 ribosomal RNAs genes was predicted, including 112 rRNA 5S, six rRNA 5.8S, one rRNA 18S, and one rRNA 28S (Table 1). In addition, 532 transfer RNAs were found with many associated with the transport of glycine, arginine, leucine, methionine, valine amino acids.

Functional annotation and secretome prediction for *Ceratocystis fimbriata* LPF1912 genome

The proteins encoded by the 7,390 predicted CDS were functionally annotated and classified through similarity searches using Blast2GO. Among the predicted proteins, 7,135 sequences (96.6%) significantly aligned with sequences of NCBI nr database and 5,896 sequences (79.8%) were annotated with GO terms (Supplementary Table S3).

Of the proteins with assigned GOs, 2,053 proteins were annotated with terms of all the three primary categories: cellular component (CC), molecular function (MF), and biological process (BP). Additionally, 666, 872, and 128 of annotated proteins had only GO terms of CC, MF and BP categories, respectively (Supplementary Table S3). The most abundant GO terms were related to “integral component of membrane” and “membrane” (GO:0016021, GO:0016020; 1,303, 400), “nucleus” (GO:0005634; 473) and “cytoplasm” (GO:0005737; 247) for CC category; "ATP binding" (GO:0005524; 545), " nucleic acid binding" (GO:0003676; 406) and "zinc ion binding" (GO:0008270; 352) for MF category; “oxidation-reduction process” (GO:0055114; 402), “transmembrane transport” (GO:0055085; 188) and “protein phosphorylation” (GO:0006468; 160) for BP category. In addition, a total of 626 secreted proteins in the *Ce. fimbriata* LPF1912 genome, which represents 8.5% of all predicted proteins (Fig. 1A).

Comparative genomic analysis

To better understand the molecular pathogenicity mechanisms of *Ce. fimbriata* LPF1912 (LPF1912) in *Eucalyptus*, we performed a comparative analysis with the predicted proteome and secretome of two fungal pathogens known to also infect *Eucalyptus* spp., *Ceratocystis eucalypticola* (*Ceu*) and *Calonectria pseudoreteaudii* (*Cal*). A total of 7,234 proteins (614

secreted) were predicted in the genome of *Ceu,* while 16,861 proteins (1,429 secreted) were predicted for *Cal* (Fig. 1A).

Analysis of orthologous proteins revealed that LPF1912, *Ceu,* and *Cal* share 4,919 clusters (60.4%) with 16,164 proteins (5,198 of LPF1912, 5,186 of *Ceu*, and 5,780 of *Cal*), which corresponds to the core proteome conserved among these three species (Fig. 1B). In addition, LPF1912 and *Ceu* share 1,774 clusters (21.8%) with 3,777 proteins (1,922 of LPF1912 and 1,885 of *Ceu*), suggesting a close relationship between these two pathogens (Fig. 1B). *Cal* was found to contain the largest group of singletons with 1,398 clusters (17.2%) with 5,933 proteins (Fig. 1B). Analysis of secretome showed that LPF1912 and *Ceu* share more clusters of orthologous (n = 375; 52.0%) than all three species together (n = 165; 22.9%), which reinforce the similarity between these two pathogens (Fig. 1B). As for the proteome, *Cal* also contained the largest group of singletons (n = 171; 23.7%) of secreted proteins.

Comparative analysis of KOG functional profiles of proteome revealed a very similar profile for LPF1912 and *Ceu* different from *Cal* (Fig. 1C). Although *Cal* had the largest total number of predicted proteins, LPF1912 and *Ceu* had a larger number of proteins that belongs to group A (RNA processing and modification) (Fig. 1C). However, *Cal* had a high number of proteins classified in KOG groups M (Cell wall/membrane/envelope biogenesis) and Q (biosynthesis, transport, and catabolism of secondary metabolites) (Fig. 1C).

Analysis of the KOG profiles of secretome revealed differences among the three pathogens (Fig. 1D) with the KOG groups K (Transcription) and Z (Cytoskeleton) were exclusive to LPF1912 and *Ceu* (Fig. 1D). Two groups (A: RNA processing and modification; L: Replication, recombination, and repair) were specific to secreted proteins of *Cal* (Fig. 1D). KOG groups T (Signal transduction mechanisms), U (Intracellular trafficking, secretion, and vesicular transport), H (Coenzyme transport and metabolism), and P (Inorganic ion transport and metabolism) were more frequent in LPF1912 and *Ceu* (Fig. 1D). In *Cal*, a large majority of secreted proteins was classified in KOG groups M (Cell wall/membrane/envelope biogenesis), V (Defense mechanisms), C (Energy production and conversion), I (Lipid transport and metabolism) and Q (Energy production and conversion; and Secondary metabolites biosynthesis, transport, and catabolism) (Fig. 1D).

To identify potential pathogenicity-associated genes in LPF1912, *Ceu*, and *Cal* genomes we performed similarity searches of all their predicted genes and secreted proteins against the Pathogen-Host Interaction (PHI) database. Similar results were obtained in the analysis of all predicted genes (Fig. 2A). However, LPF1912 and *Ceu* had a high proportion of secreted proteins related to virulence than *Cal* (Fig. 2A).

Investigations of the fungal-specific virulence genes shared among the three species, using the DFVF, showed clustering of the proteome and secretome of the pathogens. Among all predicted genes, 300 clusters of virulence genes were common to these three eucalyptus pathogens (Fig. 2B). However exclusive clusters were detected, with *Cal* having 65 exclusive clusters of virulence genes (Fig. 2B). Although LPF1912 and *Ceu* did not have exclusive virulence genes, they shared 28 clusters of virulence-related proteins (Fig. 2B). In terms of virulence-related proteins, LPF1912 and *Ceu* shared five clusters (Fig. 2B; Table 2), while *Cal* had 44 exclusive clusters of virulence-related proteins (Fig. 2B; Table 2).

Differentially expressed genes (DEGs) of *Ceratocystis fimbriata* LPF1912 during its interaction with eucalyptus plants

A total of 70,973,156 reads were obtained from the six RNA-seq libraries that mapped to the LPF1912 genome. After filtering and trimming, 70,622,485 clean reads were used to identify the DEGs of *Ce. fimbriata* LPF1912. PCA analysis showed that the two biological replicates from each group had no confounding effects (Supplementary Fig. S2) and replicates showed a high correlation (Supplementary Fig. S3). The RNA-Seq reads mapping to the genome of *Ce. fimbriata* LPF1912 increased according to the infection time, especially for the susceptible eucalyptus *pool* (Supplementary Table S4).

The total number of DEGs at the three incubation periods were 37 and 34 in the susceptible and resistant *pool*, respectively (Fig. 3A-B). In the eucalyptus susceptible *pool*, six DEGs were continuously observed at all infection periods, among them five were down-regulated and one up-regulated (Fig. 3A). However, in the eucalyptus resistant *pool*, nine DEGs were common at all infection phases, and all were down-regulated (Fig. 3B). The highest number of DEGs were detected in the 12 vs 24 hours after inoculation (hai) (Fig. 3C). Comparing susceptible vs resistant *pool*, 42 down-regulated genes were detected, and among of them 18 were common, 14 were specific from the susceptible *pool*, and 10 were exclusive of the resistant *pool*. There were no exclusively up-regulated genes observed in the resistant *pool* (Fig. 3C), though in the susceptible *pool* there were three common and two specific up-regulated genes (Fig. 3C).

Functional annotation and pathogenicity-related DEGs

Functional analysis of KOGs showed that on the susceptible *pool* a majority (75.7%) of DEGs had no-hit to known proteins (Fig. 4A). However, in the resistant *pool*, 10 (29.4%) down-regulated genes were related to information storage, processing, cellular processes, and signaling (Fig. 4B), which indicate that at 24 hai cells fungal were negatively impacted by host resistance mechanisms.

Among all DEGs, 11 DEGs were associated with virulence and pathogenicity according to DFVF (Table 3). The majority were down-regulated in the R-12h vs R-24h (Table 3). GO enrichment analysis revealed that these genes were related to cytoskeleton organization, protein ubiquitination, RNA polymerase, and translation regulation (Table 3). Furthermore, three pathogenicity-related DEGs in the S-12h vs S-24h were associated with polygalacturonase and kinase activity (Table 3).

DISCUSSION

In this study, we present a high-quality whole-genome assembly (31.8 Mb) for *Ceratocystis fimbriata* LPF1912 (LPF1912), a serious and widespread fungal pathogen in Brazilian eucalyptus plantations that causes CW. LPF1912 has a compact genome, with almost a half of the genome comprised of coding sequence (13.8 Mb; 43%) and only 1.85 Mb (~6%) of repetitive sequence. The total number of predicted genes is similar to other *Ceratocystis* species, such as *Ce. fimbriata* type isolate (Wilken et al., 2013), *Ce. eucalypticola* (Wingfield et al., 2015), *Ce. cacaofunesta* (Molano et al., 2018), and *Ce. manginecans* (van der Nest et al., 2014).

To identify key pathogenicity-related genes of *Ceratocystis fimbriata* in *Eucalyptus* spp., we compared the LPF1912 genome with *Ceratocystis eucalypticola* (*Ceu*) and *Calonectria pseudoreteaudii* (*Cal*), which also are known to infect *Eucalyptus* spp. causing CW in South Africa (van Wyk et al., 2012) and CLB in China (Chen et al., 2013, 2011; Ye et al., 2017), respectively. During our investigation, notable similarities in proteome and secretome profiles between LPF1912 and *Ceu* were observed. It was expected since LPF1912 and *Ceu* cause the same disease in *Eucalyptus* spp. and have a close taxonomic relationship (Fourie et al., 2015; Oliveira et al., 2016, 2015). Further comparisons of proteome or secretome using OrthoMCL, identified clusters that are highly conserved between the two *Ceratocystis* pathogens, as well as which genes that are unique to each eucalyptus fungal pathogen. In addition, KOG enrichment analysis of secreted proteins revealed that *Cal* displayed unique proteome and secretome profiles in comparison to the *Ceratocystis* species.

Since secreted proteins play crucial roles during host infection of fungal pathogenic (de Sain and Rep, 2015; Molano et al., 2018), we performed a BLASTp analysis of the three eucalyptus fungal pathogens secretome against the DFVF, which is a specific database for fungal pathogen-specific (Lu et al., 2012). LPF1912 and *Ceu* shared five exclusive clusters with DFVF, while *Cal* had 44 exclusive clusters. GO enrichment analysis revealed an interesting difference among the three eucalyptus fungal pathogens (Table 2). Three exclusive gene clusters shared by LPF1912 and *Ceu* were related to the trehalose metabolic process, proteolysis, and cell wall metabolic process. Trehalose is a disaccharide with protein and membrane-stabilizing capability that occurs naturally in plants and plays an important role in stress tolerance (Grennan, 2007). Furthermore, cell wall-degrading enzymes (CWDEs) have been considered as a pathogenicity factor for many plant pathogenic fungi (Annis and Goodwin, 1997; Brito et al., 2006; Tonukari, 2003), especially fungal pathogens that cause wilt diseases (Chang et al., 2016; Kubicek et al., 2014). On the other hand, among the 44 clusters exclusive to *Cal*, five were related to lipid and cutin metabolism. Since *Cal* preferably infects the leaf (Chen et al., 2011), these genes have a crucial role in infection as the eucalyptus leaf surface contains large amounts of lipid compounds such as cutin and waxes (Guzmán et al., 2014) that provides a physical barrier against pathogens (Serrano et al., 2014; Ziv et al., 2018).

Additionally, to study the pathogenicity mechanisms of LPF1912 we evaluated the DEGs during its infection in eucalyptus plants resistant and susceptible to CW at three times after inoculation. Among the DEGs, 11 had homologs in the DFVF, therefore, they were associated with virulence and pathogenicity (Lu et al., 2012). Of them, the majority were down-regulated in the resistant *pool* comparisons, especially in the R-12h vs R-24h (Table 3). Such genes were involved in essential biological processes to fungal growth such as cytoskeleton organization, protein ubiquitination, RNA polymerase, and translation regulation (Hershko et al., 2000; Riquelme et al., 2018; Xiang and Plamann, 2003). Response time is considered as a crucial key for plant defense, thus there is a correlation between time response and effective resistance (Shen et al., 2017). Often, effective response in several plants resistant to fungal pathogens occur up to 24h after inoculation (Dobon et al., 2016; Li et al., 2016; H. Liu et al., 2017; Lu et al., 2018; Nirmala et al., 2011; Romeis et al., 2000; Yadav et al., 2016). In this context, our results indicate that at the beginning of infection the fungus tried to infect and colonize, however, due to the effective resistance response of the eucalyptus plants resistant to CW, the fungal growth was negatively affected at 24h after inoculation.

In the susceptible *pool*, the total number of expressed genes of the fungus was slightly higher than the resistant *pool*, likely due to the absence of plant efficient defense mechanisms

(Lo Presti et al., 2015). Interestingly, three pathogenicity-related DEGs in the S-12h vs S-24h were associated with CWDEs (endopolygalacturonase) and kinase activity (Table 3). Polygalacturonase is one of the most important enzyme family responsible for the depolymerization of cell walls (Quoc and Chau, 2017), and have been considered as a virulence factor to many fungal pathogens (C. Q. Liu et al., 2017; Mori et al., 2008; Oeser et al., 2002; Shieh et al., 1997; Ye et al., 2017). Furthermore, Protein kinases (PK) are a major class of signaling molecules that catalyze reversible phosphorylation of a large proportion of cellular proteins, modulating protein activity and gene expression (Cohen, 2000). The coordinated action of multiple PK pathways integrates a variety of external and internal cues to orchestrate key processes during fungal infection, such as appressorium formation (Xu and Hamer, 1996), which represents a crucial step to host infection by many phytopathogenic fungi. The crucial role of protein kinases in pathogenicity has been demonstrated for several phytopathogenic fungi (Turrà et al., 2014).

As presented in this study, the three eucalyptus fungal pathogens showed different pathogenicity mechanisms. Since LPF1912 and *Ceu* shared the same pathogenicity-related genes, which were much different from *Cal*, our findings suggest that the pathogenicity is mainly related to the nature of infection of CW and CLB. In addition, using RNA-Seq we identified and confirmed that CWDEs and PK play a crucial role during *Ceratocystis fimbriata* infection on eucalyptus plants.

CONCLUSIONS

In this study, we presented a characterization of the genome of *Ceratocystis fimbriata* LPF1912, the causal agent of Ceratocystis Wilt in eucalyptus plantations in Brazil. The assembly exhibited a higher level of genome quality and completeness compared to other *Ceratocystis* genomes. Additionally, we compared the LPF1912 genome with the other two fungal pathogens that are also found to infect eucalyptus. We performed a general characterization of secretome to three pathogens and observed that they had a minimal quorum of conserved proteins, as well as specific clusters of pathogenicity-related genes. Finally, we evaluated the transcriptome of LPF1912 during its infection in eucalyptus plants, mapping DEGs that had a molecular function related to pathogenicity. Moreover, we provide a high-quality dataset that may be useful in future studies of genomic and transcriptomic to target and develop strategies to manage *Ceratocystis fimbriata* sensu stricto group.

Acknowledgments The authors thank to ‘Conselho Nacional de Desenvolvimento Cientifico e Tecnológico’ (CNPq) and ‘Coordenação de Aperfeicoamento de Pessoal de Nivel Superior’ (CAPES) for the fellowship granted to first the author. To Clonar Resistência a Doenças Florestais for providing the plant material and the inoculation facilities. To Núcleo de Análise de Biomoléculas (NuBioMol) for supporting data analysis. NuBioMol is supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), CAPES, CNPq, Financiadora de Estudos e Projetos (Finep) and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI). To Claudia Lang by text reading and review.

Data availability The sequencing raw data from from two MiSeq libraries and scaffolds sequences of the genome assembled were deposited in the Sequence Read Archive and WGS portal from NCBI under the following accession numbers: SRR9831094, SRR9831095, VNIJ00000000.

Table 1 Genome and transcriptome statistics of assembly and annotation of *Ceratocystis fimbriata* LPF1912 genome

Genome assembly	
Estimated genome size[a] (Mb)	30.2 - 30.6
Total size of assembly[b] (Mb)	31.6
Total number of assembled scaffolds	788
Largest scaffold (bp)	802,649
Smallest scaffold (bp)	641
N50[c] scaffold length (bp)	181,961
GC content of the whole genome (%)	48.09
Heterozygosity (%)	0.12
Repetitive sequences[d] (Mb; %)	1.85; 5.84
Coding sequence (Mb; %)	13.8; 43
Number of genes	8,042
Coding DNA Sequences (CDS) (Mb; %)	7,390 (12.2; 38.5)
GC content in coding regions (%)	52.03
Number of ribosomal RNAs (rRNAs)	120
Number of transfer RNAs (tRNAs)	532
Transcriptome assembly	
Total size of assembly (bp)	4,034,972
N50 contig length (bp)	611
Total number of assembled contigs	7,936
Largest contig (bp)	3,768
Smallest contig (bp)	201
BUSCO assessment (%)	
Complete	22.1
Complete and single-copy	20.0
Complete and duplicate	2.1
Fragmented	45.5
Missing	32.4

[a] Estimation based on the k-mers frequency distributions of 17-, 21-, and 25-mers.

[b] Combined length of all assembled scaffolds.

[c] N50 is defined as the sequence length of the shortest scaffold at 50% of the total genome length.

[d] Includes retroelements, DNA transposons, small RNA, satellites, and simple and low complexity repeat regions. Further details in Supplementary Table S4.

Table 2 Gene Ontology (GO) of the five most representative orthologous genes clusters from secreted proteins of *Ceratocystis fimbriata* LPF1912 (LPF1912), *Ce. Eucalypticola* (*Ceu*), and *Calonectria pseudoreteaudii* (*Cal*) associated with pathogenicity according to the Database of Fungal Virulence Factors (DFVF)

Cluster	gene IDs	GO terms[a]	DFVF uniprot IDs
		Shared between LPF1912 and *Ceu*	
1	g926 (LPF1912); g762 (*Ceu*)	CC:GO:0016021 (integral component of membrane); MF:GO:0042124 (1,3-beta-glucanosyltransferase activity); BP:GO:0099132 (ATP hydrolysis coupled cation transmembrane transport)	O13318; P43076; P0C955; Q2KN79; P0C7S9
2	g495 (LPF1912); g4150 (*Ceu*)	MF:GO:0004555 (alpha-trehalase activity); BP:GO:0005991 (trehalose metabolic process)	O42622; Q156F4; Q8NJR9
3	g4863 (LPF1912); g5211 (*Ceu*)	CC:GO:0009277 (fungal-type cell wall); MF:GO:0016757 (chitin binding); BP:GO:0006037 (cell wall chitin metabolic process)	Q5AJC0
4	g1412 (LPF1912); g7013 (*Ceu*)	MF:GO:0004190 (aspartic-type endopeptidase activity); BP:GO:0006508 (proteolysis)	Q4WFS2
5	g7226 (LPF1912); g6059 (*Ceu*)	MF:GO:0003824 (catalytic activity); BP:GO:0009116 (nucleoside metabolic process)	P0CY34
		Exclusive to *Cal*	
1	g14066; g5035	CC:GO:0005576 (extracellular region); MF:GO:0050525 (cutinase activity)	Q00298; Q8TGB8; Q9Y7G8
2	g16528	CC:GO:0005576 (extracellular region); BP:GO:0009405 (pathogenesis); MF:GO:0050525 (cutinase activity)	P10951; P11373
3	g11471	BP:GO:0006629 (lipid metabolic process)	Q6WER3
4	g16799	MF:GO:0004620 (phospholipase activity); BP:GO:0016042 (lipid catabolic process)	Q9UWF6
5	g9755	MF:GO:0004806 (triglyceride lipase activity); BP:GO:0016042 (lipid catabolic process)	Q5APG1

[a] CC: cellular component. MF: molecular function. BP: biological process.

Table 3 Differentially expressed genes of *Ceratocystis fimbriata* LPF1912 associated with virulence and pathogenicity according to the Database of Fungal Virulence Factors (DFVF)

Gene ID	Comparison[a]	Log_2FC[b]	Gene ontology[C]	DFVF UniProt ID	Protein name
g588	R-12h vs R-24h	-1.82	CC:GO:0005737 (cytoplasm); MF:GO:0016616 (oxidoreductase activity); BP:GO:0016567 (protein ubiquitination)	Q5ADS0	Ubiquitin
g821	R-12h vs R-24h	-1.33	CC:GO:0005737 (cytoplasm); MF:GO:0005525 (GTP binding); BP:GO:0007010 (cytoskeleton organization)	C1GM22	Tubulin alpha chain
*g1560	S-12h vs S-24h	-2.99	CC:GO:0005576 (extracellular region); MF:GO:0004650 (polygalacturonase activity); BP:GO:0071555 (cell wall organization)	O94100	Endopolygalacturonase 1
g2595	R-12h vs R-18h; R-12h vs R-24h	-1.97; -2.11	CC:GO:0016021 (integral component of membrane); BP:GO:0055085 (transmembrane transport)	Q9P8L8	DHA14-like major facilitator
g2733	S-12h vs S-24h	-2.52	CC:GO:0005829 (cytosol); MF:GO:0016301 (kinase activity); BP:GO:0016310 (phosphorylation)	F5HAA8	Uncharacterized protein
g3813	R-12h vs R-18h; R-12h vs R-24h	-2.34; -2.58	CC:GO:0005737 (cytoplasm)	Q59LF3	Regulator of cytoskeleton and endocytosis RVS167
g5458	R-12h vs R-18h; R-12h vs R-24h; S-12h vs S-24h	-1.53; -2.87; -2.37	CC:GO:0005634 (nucleus); MF:GO:0000981 (RNA polymerase II); BP:GO:0006357 (regulation of transcription by RNA polymerase II)	A8QJI7	XlnR
g5845	R-12h vs R-24h	-2.43	MF:GO:0003729 (mRNA binding)	Q1DXH0	Polyadenylate-binding protein
g6125	S-12h vs S-24h; S-18h vs S-24h	-2.02; -1.78	MF:GO:0004674 (kinase activity); BP:GO:0006468 (protein phosphorylation)	Q5K2R7	Putative serine/threonine kinase
g6308	R-12h vs R-18h; R-12h vs R-24h	-1.49; -2.31	MF:GO:0004535 (ribonuclease activity); BP:GO:0090503 (exonucleolytic)	Q5A8L5	CCR4-NOT core DEDD family RNase subunit
g6635	R-12h vs R-24h	-1.68	CC:GO:0005840 (ribosome); MF:GO:0003735 (structural constituent of ribosome); BP:GO:0006412 (translation)	Q5ADS0	Ubiquitin

[a] R-12h: resistant *pool* at 12 h after inoculation. R-18h: resistant *pool* at 18 h after inoculation. R-24h: resistant *pool* at 24 h after inoculation. S-12h: susceptible *pool* at 12 h after inoculation. S-18h: susceptible *pool* at 18 h after inoculation. S-24h: susceptible *pool* at 24 h after inoculation.

[b] Log_2 of fold-change (FC) value.

[C] CC: cellular component. MF: molecular function. BP: biological process.

*Secreted protein.

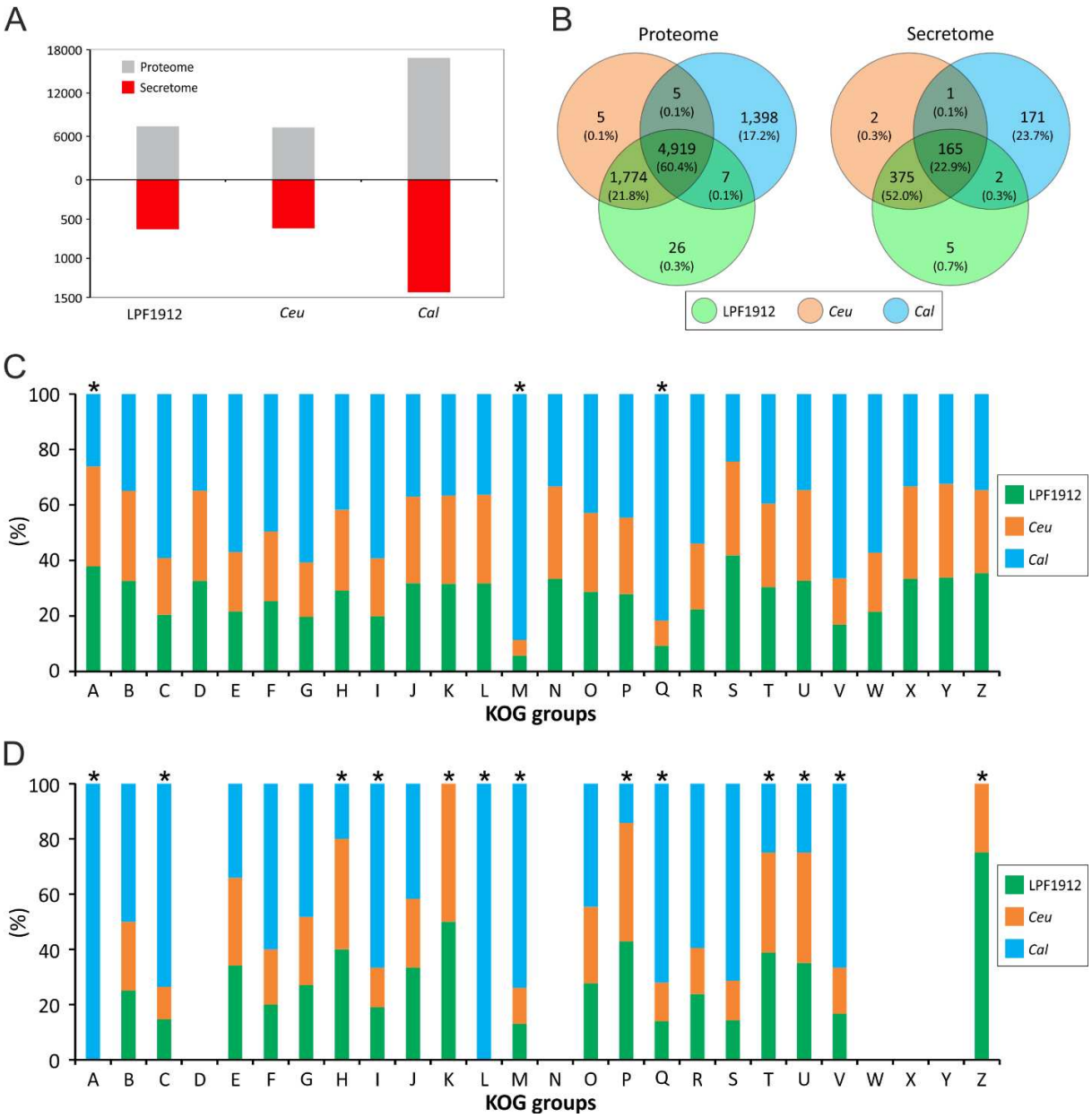

Fig. 1. Comparative genomics analysis of *Ceratocystis fimbriata* LPF1912, *Ceratocystis eucalypticola* (*Ceu*), and *Calonectria pseudoreteaudii* (*Cal*). (A) Number of all predicted genes and secreted proteins. (B) Venn diagrams showing unique and shared orthologous genes among three species. (C e D) EuKaryotic Orthologous Groups (KOG) enrichment analysis for all predicted genes and secreted proteins, respectively. Asterisks state the major different groups among the three species. **KOG groups:** [A] RNA processing and modification; [B] Chromatin structure and dynamics; [C] Energy production and conversion; [D] Cell cycle control, cell division, chromosome partitioning; [E] Amino acid transport and metabolism; [F] Nucleotide transport and metabolism; [G] Carbohydrate transport and metabolism; [H] Coenzyme transport and metabolism; [I] Lipid transport and metabolism; [J] Translation, ribosomal structure and biogenesis; [K] Transcription; [L] Replication, recombination and repair; [M] Cell wall/membrane/envelope biogenesis; [N] Cell motility; [O] Posttranslational modification, protein turnover, chaperones; [P] Inorganic ion transport and metabolism; [Q] Secondary metabolites biosynthesis, transport and catabolism; [R] General function prediction only; [S] Function unknown; [T] Signal transduction mechanisms; [U] Intracellular trafficking, secretion, and vesicular transport; [V] Defense mechanisms; [W] Extracellular structures; [X] Unnamed protein; [Y] Nuclear structure; and [Z] Cytoskeleton. The color key is described in the figure.

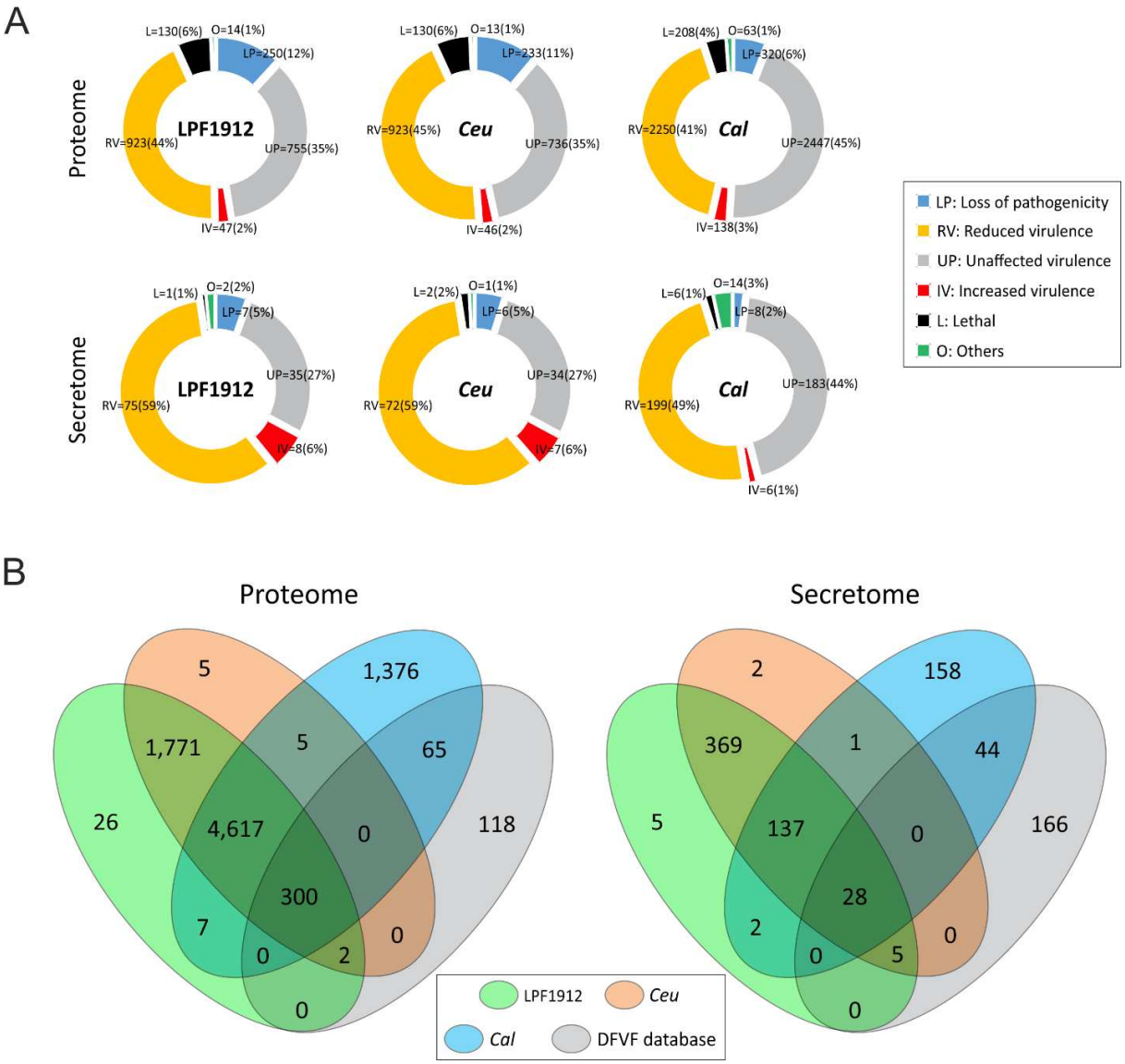

Fig. 2. Comparative genomics analysis of *Ceratocystis fimbriata* LPF1912 (LPF1912), *Ceratocystis eucalypticola* (*Ceu*), and *Calonectria pseudoreteaudii* (*Cal*). (A) Classification of all predicted genes and secreted proteins from each species on the PHI database. (B) Venn diagrams showing unique and shared orthologous genes among all predicted genes or secreted proteins of the three species and DFVF database proteins. Color key is described in the figure.

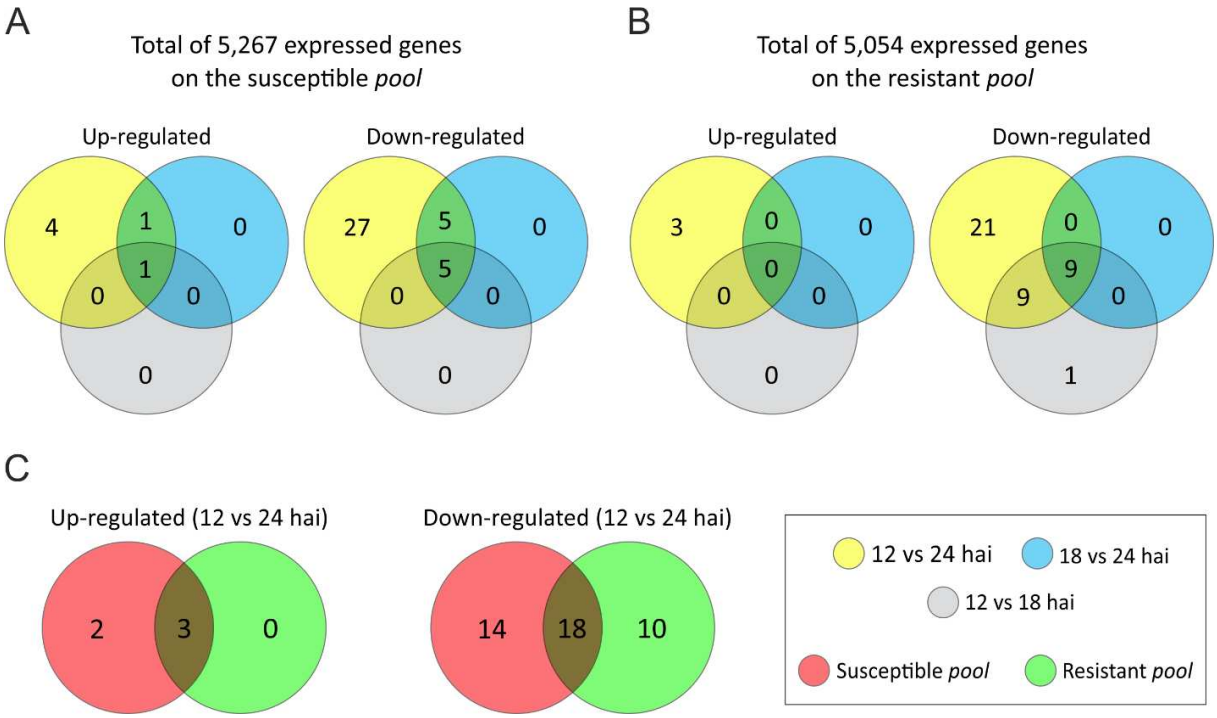

Fig. 3. Venn diagrams of differentially expressed genes (DEGs) of *Ceratocystis fimbriata* LPF1912 during three incubation periods of its infection in eucalyptus susceptible and resistant clones. (A and B) Venn diagrams represent DEGs from three incubation periods. (C) Venn diagrams represent DEGs from 12 vs 24 hours after inoculation (hai). Color key is described in the figure.

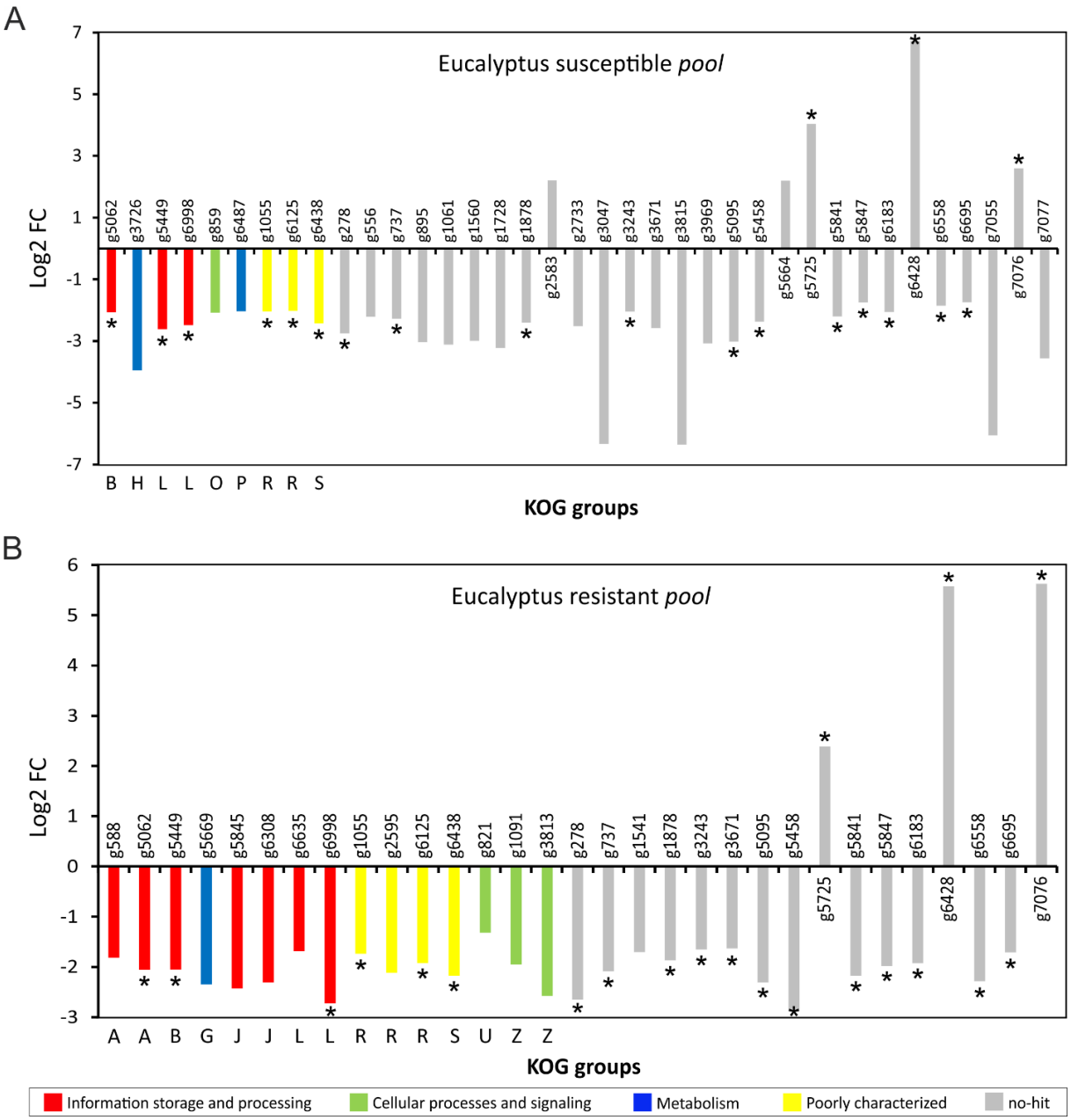

Fig. 4. EuKaryotic Orthologous Groups (KOG) enrichment analysis of differentially expressed genes (DEGs) of *Ceratocystis fimbriata* LPF1912 from 12 12 vs 24 hai. (A) Eucalyptus susceptible *pool*. (B) Eucalyptus resistant *pool*. **KOG groups:** [A] RNA processing and modification; [B] Chromatin structure and dynamics; [C] Energy production and conversion; [D] Cell cycle control, cell division, chromosome partitioning; [E] Amino acid transport and metabolism; [F] Nucleotide transport and metabolism; [G] Carbohydrate transport and metabolism; [H] Coenzyme transport and metabolism; [I] Lipid transport and metabolism; [J] Translation, ribosomal structure and biogenesis; [K] Transcription; [L] Replication, recombination and repair; [M] Cell wall/membrane/envelope biogenesis; [N] Cell motility; [O] Posttranslational modification, protein turnover, chaperones; [P] Inorganic ion transport and metabolism; [Q] Secondary metabolites biosynthesis, transport and catabolism; [R] General function prediction only; [S] Function unknown; [T] Signal transduction mechanisms; [U] Intracellular trafficking, secretion, and vesicular transport; [V] Defense mechanisms; [W] Extracellular structures; [X] Unnamed protein; [Y] Nuclear structure; and [Z] Cytoskeleton. *DEGs common in both eucalyptus *pools*.

REFERENCES

Alfenas, A.C., Zauza, E.A.V., Mafia, R.G., Assis, T.F., 2009. Clonagem e doenças do eucalipo, 1st ed. Editora UFV, Viçosa, MG state, Brazil.

Alfenas, R.F., Lombard, L., Pereira, O.L., Alfenas, A.C., Crous, P.W., 2015. Diversity and potential impact of Calonectria species in Eucalyptus plantations in Brazil. Stud. Mycol. 80, 89–130. https://doi.org/10.1016/J.SIMYCO.2014.11.002

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data [WWW Document]. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 7.18.19).

Annis, S.L., Goodwin, P.H., 1997. Recent advances in the molecular genetics of plant cell wall-degrading enzymes produced by plant pathogenic fungi. Eur. J. Plant Pathol. 103, 1–14. https://doi.org/10.1023/A:1008656013255

Baker, C.J., Harrington, T.C., Krauss, U., Alfenas, A.C., 2003. Genetic Variability and Host Specialization in the Latin American Clade of Ceratocystis fimbriata. Phytopathology 93, 1274–1284. https://doi.org/10.1094/PHYTO.2003.93.10.1274

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V, Sirotkin, A. V, Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Bartholomé, J., Mandrou, E., Mabiala, A., Jenkins, J., Nabihoudine, I., Klopp, C., Schmutz, J., Plomion, C., Gion, J.-M., 2015. High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. New Phytol. 206, 1283–1296. https://doi.org/10.1111/nph.13150

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. https://doi.org/10.1038/nbt.3519

Brito, N., Espino, J.J., González, C., 2006. The Endo-β-1,4-Xylanase Xyn11A Is Required for Virulence in Botrytis cinerea. Mol. Plant Microbe Interact. 19, 25–32. https://doi.org/10.1094/MPMI-19-0025

Bruce, M., Neugebauer, K.A., Joly, D.L., Migeon, P., Cuomo, C.A., Wang, S., Akhunov, E., Bakkeren, G., Kolmer, J.A., Fellers, J.P., 2014. Using transcription of six Puccinia triticina races to identify the effective secretome during infection of wheat. Front. Plant Sci. 4, 520. https://doi.org/10.3389/fpls.2013.00520

Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K.K., Jurka, J., Michelmore, R.W., Dubcovsky, J., 2011. Next Generation Sequencing Provides Rapid Access to the Genome of Puccinia striiformis f. sp. tritici, the Causal Agent of Wheat Stripe Rust. PLoS One 6, e24230. https://doi.org/10.1371/journal.pone.0024230

Chan, P.P., Lowe, T.M., 2019. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol. Biol. 1962, 1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

Chang, H.-X., Yendrek, C.R., Caetano-Anolles, G., Hartman, G.L., 2016. Genomic characterization of plant cell wall degrading enzymes and in silico analysis of xylanses and polygalacturonases of Fusarium virguliforme. BMC Microbiol. 16, 147. https://doi.org/10.1186/s12866-016-0761-0

Chen, Q.Z., Guo, W.S., Ye, X.Z., Huang, X.P., Wu, Y.Z., 2013. Identification of Calonectria associated with Eucalyptus leaf blight in Fujian Province. J. Fujian For. Coll. 32, 176–182.

Chen, S., Huang, T., Zhou, Y., Han, Y., Xu, M., Gu, J., 2017. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. BMC Bioinformatics 18, 80. https://doi.org/10.1186/s12859-017-1469-3

Chen, S.F., Lombard, L., Roux, J., Xie, Y.J., Wingfield, M.J., Zhou, X.D., 2011. Novel species of Calonectria associated with Eucalyptus leaf blight in Southeast China. Persoonia 26, 1–12. https://doi.org/10.3767/003158511X555236

Cohen, P., 2000. The regulation of protein function by multisite phosphorylation--a 25 year update. Trends Biochem. Sci. 25, 596–601.

Coil, D., Jospin, G., Darling, A.E., 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31, 587–589. https://doi.org/10.1093/bioinformatics/btu661

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

de Sain, M., Rep, M., 2015. The Role of Pathogen-Secreted Proteins in Fungal Vascular Wilt Diseases. Int. J. Mol. Sci. 16, 23970–93. https://doi.org/10.3390/ijms161023970

Dobon, A., Bunting, D.C.E., Cabrera-Quio, L.E., Uauy, C., Saunders, D.G.O., 2016. The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. BMC Genomics 17, 380. https://doi.org/10.1186/s12864-016-2684-4

Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. J. Mol. Biol. 300, 1005–1016. https://doi.org/10.1006/JMBI.2000.3903

Ferreira, E.M., Harrington, T.C., Thorpe, D.J., Alfenas, A.C., 2010. Genetic diversity and interfertility among highly differentiated populations of Ceratocystis fimbriata in Brazil. Plant Pathol. 59, 721–735. https://doi.org/10.1111/j.1365-3059.2010.02275.x

Ferreira, F.A., Demuner, A.M.M., Demuner, N.L., Pigato, S., 1999. Murcha-de-Ceratocystis em eucalipto no Brasil. Fitopatol. Bras. 24, p.284.

Ferreira, F.A., Maffia, L.A., Barreto, R.W., Demuner, N.L., Pigatto, S., 2006. Symptomatology of Ceratocystis wilt in eucalyptus. Rev. Árvore 30, 155–162.

Ferreira, M.A., Harrington, T.C., Gongora-Canul, C.C., Mafia, R.G., Zauza, E.A. V., Alfenas, A.C., 2013. Spatial-temporal patterns of Ceratocystis wilt in *Eucalyptus* plantations in Brazil. For. Pathol. 43, 153–164. https://doi.org/10.1111/efp.12013

Fourie, A., Wingfield, M.J., Wingfield, B.D., Barnes, I., 2015. Molecular markers delimit cryptic species in Ceratocystis sensu stricto. Mycol. Prog. 14, 1020. https://doi.org/10.1007/s11557-014-1020-0

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883

Grattapaglia, D., Kirst, M., 2008. *Eucalyptus* applied genomics: from gene sequences to breeding tools.

New Phytol. 179, 911–929. https://doi.org/10.1111/j.1469-8137.2008.02503.x

Grennan, A.K., 2007. The role of trehalose biosynthesis in plants. Plant Physiol. 144, 3–5. https://doi.org/10.1104/pp.104.900223

Guimarães, L.M. da S., Titon, M., Lau, D., Rosse, L.N., Oliveira, L.S.S., Rosado, C.C.G., Christo, G.G.O., Alfenas, A.C., 2010. Eucalyptus pellita as a source of resistance to rust, ceratocystis wilt and leaf blight. Crop Breed. Appl. Biotechnol. 10, 124–131.

Guzmán, P., Fernández, V., Graça, J., Cabral, V., Kayali, N., Khayet, M., Gil, L., 2014. Chemical and structural analysis of Eucalyptus globulus and E. camaldulensis leaf cuticles: a lipidized cell wall region. Front. Plant Sci. 5, 481. https://doi.org/10.3389/fpls.2014.00481

Hershko, A., Ciechanover, A., Varshavsky, A., 2000. The ubiquitin system. Nat. Med. 6, 1073–1081. https://doi.org/10.1038/80384

Huynh, M.D., Page, J.T., Richardson, B.A., Udall, J.A., 2015. Insights into Transcriptomes of Big and Low Sagebrush. PLoS One 10, e0127593. https://doi.org/10.1371/journal.pone.0127593

IBÁ, 2019. Indústria Brasileira de Árvores - Report 2019. Sao Paulo, Brazil. https://iba.org/datafiles/publicacoes/relatorios/iba-relatorioanual2019.pdf. Accessed 01 Nov 2019

Joly, D.L., Feau, N., Tanguay, P., Hamelin, R.C., 2010. Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (Melampsora spp.). BMC Genomics 11, 422. https://doi.org/10.1186/1471-2164-11-422

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462–467. https://doi.org/10.1159/000084979

Kim, H.A., Lim, C.J., Kim, S., Choe, J.K., Jo, S.-H., Baek, N., Kwon, S.-Y., 2014. High-Throughput Sequencing and De Novo Assembly of Brassica oleracea var. Capitata L. for Transcriptome Analysis. PLoS One 9, e92087. https://doi.org/10.1371/journal.pone.0092087

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. 305, 567–580. https://doi.org/10.1006/JMBI.2000.4315

Kubicek, C.P., Starr, T.L., Glass, N.L., 2014. Plant Cell Wall–Degrading Enzymes and Their Secretion in Plant-Pathogenic Fungi. Annu. Rev. Phytopathol. 52, 427–451. https://doi.org/10.1146/annurev-phyto-102313-045831

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Li, L., Stoeckert, C.J., Roos, D.S., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189. https://doi.org/10.1101/gr.1224503

Li, W., Liu, Y., Wang, Jing, He, M., Zhou, X., Yang, C., Yuan, C., Wang, Jichun, Chern, M., Yin, J., Chen, W., Ma, B., Wang, Y., Qin, P., Li, S., Ronald, P., Chen, X., 2016. The durably resistant rice cultivar Digu activates defence gene expression before the full maturation of *M agnaporthe oryzae* appressorium. Mol. Plant Pathol. 17, 354–368. https://doi.org/10.1111/mpp.12286

Liu, C.Q., Hu, K. Di, Li, T.T., Yang, Y., Yang, F., Li, Y.H., Liu, H.P., Chen, X.Y., Zhang, H., 2017. Polygalacturonase gene pgxB in Aspergillus niger is a virulence factor in apple fruit. PLoS One 12, e0173277. https://doi.org/10.1371/journal.pone.0173277

Liu, H., Guo, Z., Gu, F., Ke, S., Sun, D., Dong, S., Liu, W., Huang, M., Xiao, W., Yang, G., Liu, Y.,

Guo, T., Wang, H., Wang, J., Chen, Z., 2017. 4-Coumarate-CoA Ligase-Like Gene OsAAE3 Negatively Mediates the Rice Blast Resistance, Floret Development and Lignin Biosynthesis. Front. Plant Sci. 7, 2041. https://doi.org/10.3389/fpls.2016.02041

Liu, J.J., Sturrock, R.N., Sniezko, R.A., Williams, H., Benton, R., Zamany, A., 2015. Transcriptome analysis of the white pine blister rust pathogen Cronartium ribicola: de novo assembly, expression profiling, and identification of candidate effectors. BMC Genomics 16, 678. https://doi.org/10.1186/s12864-015-1861-1

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., Kahmann, R., 2015. Fungal Effectors and Plant Susceptibility. Annu. Rev. Plant Biol. 66, 513–545. https://doi.org/10.1146/annurev-arplant-043014-114623

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, L., Rong, W., Massart, S., Zhang, Z., 2018. Genome-Wide Identification and Expression Analysis of Cutinase Gene Family in Rhizoctonia cerealis and Functional Study of an Active Cutinase RcCUT1 in the Fungal–Wheat Interaction. Front. Microbiol. 9, 1813. https://doi.org/10.3389/fmicb.2018.01813

Lu, T., Yao, B., Zhang, C., 2012. DFVF: database of fungal virulence factors. Database 2012, bas032. https://doi.org/10.1093/database/bas032

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Yunjie, Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Yong, Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, Jian, Lam, T.-W., Wang, Jun, 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1, 18. https://doi.org/10.1186/2047-217X-1-18

Mafia, R.G., Ferreira, M.A., Zauza, E.A. V., Silva, J.F., Colodette, J.L., Alfenas, A.C., 2013. Impact of Ceratocystis wilt on eucalyptus tree growth and cellulose pulp yield. For. Pathol. 43, 379–385. https://doi.org/10.1111/efp.12041

Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–70. https://doi.org/10.1093/bioinformatics/btr011

Meinhardt, L.W., Costa, G.G., Thomazella, D.P., Teixeira, P.J.P., Carazzolle, M., Schuster, S.C., Carlson, J.E., Guiltinan, M.J., Mieczkowski, P., Farmer, A., Ramaraj, T., Crozier, J., Davis, R.E., Shao, J., Melnick, R.L., Pereira, G.A., Bailey, B.A., 2014. Genome and secretome analysis of the hemibiotrophic fungal pathogen, Moniliophthora roreri, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. BMC Genomics 15, 164. https://doi.org/10.1186/1471-2164-15-164

Molano, E.P.L., Cabrera, O.G., Jose, J., do Nascimento, L.C., Carazzolle, M.F., Teixeira, P.J.P.L., Alvarez, J.C., Tiburcio, R.A., Tokimatu Filho, P.M., de Lima, G.M.A., Guido, R.V.C., Corrêa, T.L.R., Leme, A.F.P., Mieczkowski, P., Pereira, G.A.G., 2018. Ceratocystis cacaofunesta genome analysis reveals a large expansion of extracellular phosphatidylinositol-specific phospholipase-C genes (PI-PLC). BMC Genomics 19, 58. https://doi.org/10.1186/s12864-018-4440-4

Mori, T., Jung, H.-Y., Maejima, K., Hirata, H., Himeno, M., Hamamoto, H., Namba, S., 2008. Magnaporthe oryzae endopolygalacturonase homolog correlates with density-dependent conidial germination. FEMS Microbiol. Lett. 280, 182–188. https://doi.org/10.1111/j.1574-6968.2008.01062.x

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., Goodstein, D.M., Dubchak, I., Poliakov, A., Mizrachi, E.,

Kullan, A.R.K., Hussey, S.G., Pinard, D., van der Merwe, K., Singh, P., van Jaarsveld, I., Silva-Junior, O.B., Togawa, R.C., Pappas, M.R., Faria, D.A., Sansaloni, C.P., Petroli, C.D., Yang, X., Ranjan, P., Tschaplinski, T.J., Ye, C.-Y., Li, T., Sterck, L., Vanneste, K., Murat, F., Soler, M., Clemente, H.S., Saidi, N., Cassan-Wang, H., Dunand, C., Hefer, C.A., Bornberg-Bauer, E., Kersting, A.R., Vining, K., Amarasinghe, V., Ranik, M., Naithani, S., Elser, J., Boyd, A.E., Liston, A., Spatafora, J.W., Dharmwardhana, P., Raja, R., Sullivan, C., Romanel, E., Alves-Ferreira, M., Külheim, C., Foley, W., Carocha, V., Paiva, J., Kudrna, D., Brommonschenkel, S.H., Pasquali, G., Byrne, M., Rigault, P., Tibbits, J., Spokevicius, A., Jones, R.C., Steane, D.A., Vaillancourt, R.E., Potts, B.M., Joubert, F., Barry, K., Pappas, G.J., Strauss, S.H., Jaiswal, P., Grima-Pettenati, J., Salse, J., Van de Peer, Y., Rokhsar, D.S., Schmutz, J., 2014. The genome of Eucalyptus grandis. Nature 510, 356–362. https://doi.org/10.1038/nature13308

Nirmala, J., Drader, T., Lawrence, P.K., Yin, C., Hulbert, S., Steber, C.M., Steffenson, B.J., Szabo, L.J., von Wettstein, D., Kleinhofs, A., 2011. Concerted action of two avirulent spore effectors activates Reaction to Puccinia graminis 1 (Rpg1)-mediated cereal stem rust resistance. Proc. Natl. Acad. Sci. 108, 14676–14681. https://doi.org/10.1073/pnas.1111771108

Oeser, B., Heidrich, P.M., Müller, U., Tudzynski, P., Tenberge, K.B., 2002. Polygalacturonase is a pathogenicity factor in the Claviceps purpurea/rye interaction. Fungal Genet. Biol. 36, 176–186. https://doi.org/10.1016/s1087-1845(02)00020-8

Oliveira, L.S.S., Damacena, M.B., Guimarães, L.M.S., Siqueira, D.L., Alfenas, A.C., 2016. Ceratocystis fimbriata isolates on Mangifera indica have different levels of aggressiveness. Eur. J. Plant Pathol. 145, 847–856. https://doi.org/10.1007/s10658-016-0873-2

Oliveira, L.S.S., Guimarães, L.M.S., Ferreira, M.A., Nunes, A.S., Pimenta, L.V.A., Alfenas, A.C., 2015. Aggressiveness, cultural characteristics and genetic variation of Ceratocystis fimbriata on Eucalyptus spp. For. Pathol. 45, 505–514. https://doi.org/10.1111/efp.12200

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786. https://doi.org/10.1038/nmeth.1701

Pierleoni, A., Martelli, P., Casadio, R., 2008. PredGPI: a GPI-anchor predictor. BMC Bioinformatics 9, 392. https://doi.org/10.1186/1471-2105-9-392

Quoc, N.B., Chau, N.N.B., 2017. The Role of Cell Wall Degrading Enzymes in Pathogenesis of Magnaporthe oryzae. Curr. Protein Pept. Sci. 18, 1019–1034. https://doi.org/10.2174/1389203717666160813164955

Riquelme, M., Aguirre, J., Bartnicki-García, S., Braus, G.H., Feldbrügge, M., Fleig, U., Hansberg, W., Herrera-Estrella, A., Kämper, J., Kück, U., Mouriño-Pérez, R.R., Takeshita, N., Fischer, R., 2018. Fungal Morphogenesis, from the Polarized Growth of Hyphae to Complex Reproduction and Infection Structures. Microbiol. Mol. Biol. Rev. 82, e00068-17. https://doi.org/10.1128/MMBR.00068-17

Romeis, T., Piedras, P., Jones, J.D., 2000. Resistance gene-dependent activation of a calcium-dependent protein kinase in the plant defense response. Plant Cell 12, 803–816. https://doi.org/10.1105/tpc.12.5.803

Rosado, C.C.G., Guimarães, L.M.D.S., Titon, M., Lau, D., Rosse, L., Resende, M.D.V. De, Alfenas, A.C., 2010. Resistance to Ceratocystis Wilt (Ceratocystis fimbriata) in Parents and Progenies of Eucalyptus grandis x E. urophylla. Silvae Genet. 59, 99–106. https://doi.org/10.1515/sg-2010-0012

Ross-Davis, A.L., Stewart, J.E., Hanna, J.W., Kim, M.-S., Knaus, B.J., Cronn, R., Rai, H., Richardson, B.A., McDonald, G.I., Klopfenstein, N.B., 2013. Transcriptome of an Armillaria root disease pathogen reveals candidate genes involved in host substrate utilization at the host-pathogen

interface. For. Pathol. 43, 468–477. https://doi.org/10.1111/efp.12056

Roux, J., Wingfield, M., 2009. *Ceratocystis* species: emerging pathogens of non-native plantation *Eucalyptus* and *Acacia* species. South. For. a J. For. Sci. 71, 115–120. https://doi.org/10.2989/SF.2009.71.2.5.820

Serrano, M., Coluccia, F., Torres, M., L'Haridon, F., Métraux, J.-P., 2014. The cuticle and plant defense to pathogens. Front. Plant Sci. 5, 274. https://doi.org/10.3389/fpls.2014.00274

Shen, Y., Liu, N., Li, C., Wang, X., Xu, X., Chen, W., Xing, G., Zheng, W., 2017. The early response during the interaction of fungal phytopathogen and host plant. Open Biol. 7, 170057. https://doi.org/10.1098/rsob.170057

Shieh, M.T., Brown, R.L., Whitehead, M.P., Cary, J.W., Cotty, P.J., Cleveland, T.E., Dean, R.A., 1997. Molecular genetic evidence for the involvement of a specific polygalacturonase, P2c, in the invasion and spread of Aspergillus flavus in cotton bolls. Appl. Environ. Microbiol. 63, 3548–3552.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467. https://doi.org/10.1093/nar/gki458

Tan, M.H., Austin, C.M., Hammer, M.P., Lee, Y.P., Croft, L.J., Gan, H.M., 2018. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (Amphiprion ocellaris) genome assembly. Gigascience 7, 1–6. https://doi.org/10.1093/gigascience/gix137

Tonukari, N.J., 2003. Enzymes and fungal virulence. J. Appl. Sci. Environ. Manag. 7, 5–8. https://doi.org/10.4314/jasem.v7i1.17158

Turrà, D., Segorbe, D., Di Pietro, A., 2014. Protein Kinases in Plant-Pathogenic Fungi: Conserved Regulators of Infection. Annu. Rev. Phytopathol. 52, 267–288. https://doi.org/10.1146/annurev-phyto-102313-050143

van der Nest, M.A., Bihon, W., De Vos, L., Naidoo, K., Roodt, D., Rubagotti, E., Slippers, B., Steenkamp, E.T., Wilken, P.M., Wilson, A., Wingfield, M.J., Wingfield, B.D., 2014. Draft genome sequences of Diplodia sapinea, Ceratocystis manginecans, and Ceratocystis moniliformis. IMA Fungus 5, 135–140. https://doi.org/10.5598/imafungus.2014.05.01.13

van Wyk, M., Roux, J., Nkuekam, G.K., Wingfield, B.D., Wingfield, M.J., 2012. Ceratocystis eucalypticola sp. nov. from Eucalyptus in South Africa and comparison to global isolates from this tree. IMA Fungus 3, 45–58. https://doi.org/10.5598/imafungus.2012.03.01.06

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Wilken, P.M., Steenkamp, E.T., Wingfield, M.J., de Beer, Z.W., Wingfield, B.D., 2013. Draft nuclear genome sequence for the plant pathogen, Ceratocystis fimbriata. IMA Fungus 4, 357–358. https://doi.org/10.5598/imafungus.2013.04.02.14

Wingfield, B.D., Barnes, I., Wilhelm de Beer, Z., De Vos, L., Duong, T.A., Kanzi, A.M., Naidoo, K., Nguyen, H.D.T., Santana, Q.C., Sayari, M., Seifert, K.A., Steenkamp, E.T., Trollip, C., van der Merwe, N.A., van der Nest, M.A., Markus Wilken, P., Wingfield, M.J., 2015. IMA Genome-F 5: Draft genome sequences of Ceratocystis eucalypticola, Chrysoporthe cubensis, C.

deuterocubensis, Davidsoniella virescens, Fusarium temperatum,Graphilbum fragrans, Penicillium nordicum, and Thielaviopsis musarum. IMA Fungus 6, 493–506. https://doi.org/10.5598/imafungus.2015.06.02.13

Winnenburg, R., Baldwin, T.K., Urban, M., Rawlings, C., Köhler, J., Hammond-Kosack, K.E., 2006. PHI-base: a new database for pathogen host interactions. Nucleic Acids Res. 34, D459–D464. https://doi.org/10.1093/nar/gkj047

Xiang, X., Plamann, M., 2003. Cytoskeleton and motor proteins in filamentous fungi. Curr. Opin. Microbiol. 6, 628–633. https://doi.org/10.1016/J.MIB.2003.10.009

Xu, J.R., Hamer, J.E., 1996. MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus Magnaporthe grisea. Genes Dev. 10, 2696–706.

Yadav, I.S., Sharma, A., Kaur, S., Nahar, N., Bhardwaj, S.C., Sharma, T.R., Chhuneja, P., 2016. Comparative Temporal Transcriptome Profiling of Wheat near Isogenic Line Carrying Lr57 under Compatible and Incompatible Interactions. Front. Plant Sci. 7, 1943. https://doi.org/10.3389/fpls.2016.01943

Ye, X., Liu, H., Jin, Y., Guo, M., Huang, A., Chen, Q., Guo, W., Zhang, F., Feng, L., 2017. Transcriptomic Analysis of Calonectria pseudoreteaudii during Various Stages of Eucalyptus Infection. PLoS One 12, e0169598. https://doi.org/10.1371/journal.pone.0169598

Zauza, E.A. V., Alfenas, A.C., Harrington, T.C., Mizubuti, E.S., Silvai, J.F., 2004. Resistance of *Eucalyptus* Clones to *Ceratocystis fimbriata*. Plant Dis. 88, 758–760. https://doi.org/10.1094/PDIS.2004.88.7.758

Ziv, C., Zhao, Z., Gao, Y.G., Xia, Y., 2018. Multifunctional Roles of Plant Cuticle During Plant-Pathogen Interactions. Front. Plant Sci. 9, 1088. https://doi.org/10.3389/fpls.2018.01088

FINAL REMARKS

- The *de novo* assembled haploid nuclear genome of *Austropuccinia psidii* epitype guava (*Psidium guajava*) isolate contains 787 scaffolds with a combined length of 672 Mbp and an N50 scaffold length of 1.4 Mbp. A total of 20,694 genes were predicated, including 20,184 protein-coding, 29 rRNAs, and 480 tRNAs genes. Additionally, about 546 Mbp (81.37%) of *A. psidii* genome consists of repetitive sequences. Of these, the majority (60.39%) were classified as transposable elements.

- The *de novo* transcriptome assembly of urediniospores of *A. psidii* contains 46,742 contigs with a combined length of 71.39 Mbp, which had a high BUSCO completeness (92.2%) with the Basidiomycota database.

- Gene expression profile analysis by RNA-Seq of two contrasting *Eucalyptus grandis* genotypes in resistance to rust (*A. psidii*) revealed a constitutive overexpression of several resistance-related genes in the resistant genotype compared to the susceptible one.

- The *de novo* assembled nuclear genome of 21 fungal pathogens of the *Ceratocystis fimbriata* complex had an average size of 30.4 Mbp, with 7,551 protein-coding, 111 rRNAs, and 366 tRNAs genes. All genomes showed a high BUSCO completeness (>90%) with the Sordariomyceta database. Furthermore, independently on the host species and geographic location, comparative genomic analysis showed a high level of similarity among all *Ceratocystis* genomes, with a shared set of 6,141 conserved orthologous clusters genes. This represents 81.33% of the average number of protein-coding genes.

- Transcriptome analysis of *C. fimbriata* LPF1912 isolate during its infection in 16 eucalyptus genotypes, as well as comparative genomic analysis with *C. eucalypticola* and *Calonectria pseudoreteaudii*, which also infect eucalyptus, revealed different pathogenicity-related genes among the three eucalyptus fungal pathogens.