MARCIA FERREIRA QUEIROZ

GENOMIC ANALYSES OF *Calonectria pteridis* AND STANDARD AREA DIAGRAM SETS FOR SEVERITY ASSESSMENT OF CALONECTRIA LEAF BLIGHT

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

VIÇOSA - MINAS GERAIS
2022

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

T
Q3g
2022

Queiroz, Marcia Ferreira, 1989-
Genomic analyses of *Calonectria pteridis* and standard area diagram sets for severity assessment of calonectria leaf blight / Marcia Ferreira Queiroz. – Viçosa, MG, 2022.
1 tese eletrônica (85 f.): il. (algumas color.).

Texto em inglês.
Orientador: Acelino Couto Alfenas.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Fitopatologia, 2022.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2023.343
Modo de acesso: World Wide Web.

1. Fungos fitopatogênicos. 2. *Cylindrocladium pteridis*. 3. Mancha-de-calonectria. 4. Eucalipto - Doenças e pragas. 5. Genoma. I. Alfenas, Acelino Couto, 1950-. II. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. III. Título.

CDD 22. ed. 632.46

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

MARCIA FERREIRA QUEIROZ

GENOMIC ANALYSES OF *Calonectria pteridis* AND STANDARD AREA DIAGRAM SETS FOR SEVERITY ASSESSMENT OF CALONECTRIA LEAF BLIGHT

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: December 02, 2022.

Assent:

Documento assinado digitalmente
gov.br
MARCIA FERREIRA QUEIROZ
Data: 14/06/2023 16:33:17-0300
Verifique em https://validar.iti.gov.br

Marcia Ferreira Queiroz
Author

Documento assinado digitalmente
gov.br
ACELINO COUTO ALFENAS
Data: 19/06/2023 08:41:57-0300
Verifique em https://validar.iti.gov.br

Acelino Couto Alfenas
Adviser

To the loves of my life, my parents,
Ana Paula and José Carlos, and
my son, Samir Lorenzo

I DEDICATE.

ACKNOWLEDGEMENT

First, to God who is the basis of everything in my life, always guiding and lightening my journey.

I would like to express my immense gratitude to my parents, Ana Paula and José Carlos, to whom I owe everything I am, for their constant support, encouragement and unconditional love, thank you for giving me wings and allowing me to fly.

To my beloved son, Samir Lorenzo, for allowing me to feel unconditional love, making my days happier. He is my greatest gift and reason to persevere.

To my beloved brothers and sisters, Ana Carla, Ana Caroline, Wagner, and Cleriston, for teaching me in so many ways about life as I strive to be the best example I can be.

I would like to express my sincere gratitude to my Advisor, Prof. Acelino Couto Alfenas for his mentoring, friendship, enthusiasm, confidence, constant teaching transmitted and opportunity of getting a Ph.D. in a renowned postgraduate program, a unique experience, full of learning and enrichment in both scientific and personal terms. I am also very grateful to Prof. Rafael Ferreira Alfenas (Ph.D co-advisor) for his mentoring, friendship, confidence, patience and teachings.

To Prof. Emerson Del Ponte, Ph.D Pedro Vidigal, Ph.D Samuel Alves, Ph.D Túlio Morgan, Ph.D Fernando Fernandes and Ph.D Kaique Alves for their unconditional support and direct collaboration in the development of this work.

To Camila Paixão, who I was lucky to have as an intern on this project, I am very grateful for all her competence, dedication and help.

I also wish to express thanks to Brazilian Forest Companies, Suzano S/A and Clonar Resistência a Doenças Florestais for financing and providing the plant material and infrastructure for inoculation.

I also wish to thanks to the friends of the Forest Pathology Laboratory (Patomol) for friendship, technical support and for always helping me.

To Márcia Brandão for her friendship, excellent work and full support with bureaucratic issues.

Special thanks to all amazing friends I made in Viçosa-MG for supporting me and for the many good times we shared during this journey.

To the Federal University of Viçosa, especially to the Plant Pathology Graduate Program for the opportunity of getting a Ph.D.

To the professors of the Plant Pathology Graduate Program for the constant teachings transmitted.

I also wish to express my gratitude to the Grupo de Estudos em Fitopatologia (GEAFIP), in which had the honor to participate as scientific coordinator in 2019, a valuable professional experience.

I also thank the secretariat of the Plant Pathology Graduate Program, especially to Sara for always being so solicitous and ready to help.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for granting the scholarship.

And finally, to those who, directly or indirectly, contributed to the development of this work, many thanks.

ABSTRACT

QUEIROZ, Marcia Ferreira, D.Sc., Universidade Federal de Viçosa, December, 2022. **Genomic analyses of *Calonectria pteridis* and standard area diagram sets for severity assessment of Calonectria leaf blight.** Adviser: Acelino Couto Alfenas.

Calonectria leaf blight (CLB) caused by several *Calonectria* species represents a serious threat to the sustainability of commercial plantations and nurseries of *Eucalyptus* in worldwide. In Brazil, *C. pteridis* is the main causal agent of CLB in eucalyptus plantations. The selection and planting of resistant eucalyptus genotypes are the main strategies of control of this disease. A comprehensive understanding of biology, genetics and pathogenicity of *Calonectria* species, as well as assessing of disease severity accurately and reliably for screening of resistant eucalyptus clones are critical knowledge to an efficient management of CLB. Therefore, in chapter 1of this Thesis, we present *de novo* assembly of the nuclear genome of *C. pteridis* LPF059, which represents a baseline for further investigations. In chapter 2, we performed a comprehensive genomic analysis study that allowed us to maximize our understanding of genetics, lifestyle, reproduction, and pathogenicity of *C. pteridis* compared to other *Calonectria* species. The assembly of *C. pteridis* LPF059 genome shows a high completeness with a large set of coding sequences related to CAZymes and secondary metabolites. The size and gene content of the mitochondrial genome of Calonectria of *C. pteridis* LPF059 is comparable to the mitochondrial genome of other *Calonectria* species. Based on mating type genes *C. pteridis* is a heterothallic species. *Calonectria henricotiae* and *C. pseudonaviculata* are classified as mesotrophs and the other species in this study as vasculartrophs due to their differences in CAZymes content and therefore nutrient acquisition. In addition, *Calonectria* species considered as saprophytic - "apathogenic"- or broad host-range showed a higher content of pathogenicity/virulence-related genes than species that are narrow host-range, which indicates that the loss of certain genes may affected their ability to infect multiple plant host species. In chapter 3, we evaluated the effect of rater instruction and two standard area diagram sets (SADs), that varied on scale structure, on the accuracy and reliability of visual estimates of CLB severity. It was found that the instruction is critical for inherently less accurate raters and given the importance of visual assessment of disease symptoms, we recommend the use of amended-linear set for CLB severity assessment. This study maximizes our knowledge on the biology, genetics, lifestyle, evolution, and pathogenicity of *C. pteridis* compared to other *Calonectria* species, providing valuable genomics data for future investigations. Moreover, it improves our understanding of

fundamental issues on development and usage of SADs, besides valuables SADs to be used on CLB severity assessment for screening of resistant eucalyptus clones.

Keywords: *Cylindrocladium pteridis*. Genome sequencing. CAZymes. Lifestyle. Fungal pathogenicity. Diagrammatic scale rating. Disease resistance.

RESUMO

QUEIROZ, Marcia Ferreira, D.Sc., Universidade Federal de Viçosa, dezembro de 2022. **Análises genômicas de *Calonectria pteridis* e conjuntos de diagramas de área padrão para avaliação da severidade da mancha-de-calonectria**. Orientador: Acelino Couto Alfenas.

A mancha-de-calonectria, causada por várias espécies de *Calonectria,* representa uma séria ameaça à sustentabilidade de plantações comerciais e viveiros de *Eucalyptus* em todo o mundo. No Brasil, *C. pteridis* é o principal agente causal da mancha-de-calonectria em plantações de eucalipto. A seleção e o plantio de genótipos de eucalipto resistentes são as principais estratégias de controle dessa doença. Uma compreensão abrangente da biologia, genética e patogenicidade das espécies de *Calonectria*, bem como a avaliação acurada e confiável da severidade da doença para a seleção de clones de eucalipto resistentes, são conhecimentos essenciais para um manejo eficiente da mancha-de-calonectria. Portanto, no Capítulo 1 desta tese, apresentamos a montagem *de novo* do genoma nuclear de *C. pteridis* LPF059, o qual representa uma referência para investigações adicionais. No Capítulo 2, realizamos um estudo abrangente de análise genômica que nos permitiu maximizar a nossa compreensão da genética, estilo de vida, reprodução e patogenicidade de *C. pteridis* em comparação com outras espécies de *Calonectria*. A montagem do genoma de *C. pteridis* LPF059 apresenta alta completude, com um grande conjunto de sequências codificadoras relacionadas a CAZymes e metabólitos secundários. O tamanho e o conteúdo gênico do genoma mitocondrial de *C. pteridis* LPF059 são comparáveis aos de outras espécies de *Calonectria*. Com base nos genes de *mating type*, *C. pteridis* é uma espécie heterotálica. *Calonectria henricotiae* e *C. pseudonaviculata* são classificadas como mesotróficas, enquanto as outras espécies deste estudo são classificadas como vasculotróficas devido às diferenças em seu conteúdo de CAZymes e, consequentemente, na aquisição de nutrientes. Além disso, as espécies de *Calonectria* consideradas saprofíticas - "apatogênicas" - ou de ampla gama de hospedeiros apresentaram um conteúdo maior de genes relacionados à patogenicidade/virulência do que as espécies de restrita gama de hospedeiros, o que indica que a perda de certos genes pode ter afetado sua capacidade de infectar várias espécies de plantas hospedeiras. No capítulo 3, avaliamos o efeito de instruções fornecidas aos avaliadores e de dois conjuntos de diagramas de área padrão (DAPs), as quais variam em estrutura, na acurácia e reprodutibilidade das estimativas visuais da severidade da mancha-de-calonectria. Descobrimos que as instruções são essenciais para avaliadores naturalmente menos precisos e, dada a importância da avaliação visual dos sintomas da doença, recomendamos o

uso do conjunto de diagramas linear ajustado para a avaliação da severidade da mancha-de-calonectria. Este estudo amplia nosso conhecimento sobre a biologia, genética, estilo de vida, evolução e patogenicidade de *C. pteridis* em comparação com outras espécies de *Calonectria*, fornecendo dados genômicos valiosos para investigações futuras. Além disso, melhora nossa compreensão sobre questões fundamentais relacionadas ao desenvolvimento e uso de DAPs, além de fornecer valiosos DAPs para serem utilizados na avaliação da severidade da mancha-de-calonectria para seleção de clones de eucalipto resistentes.

Palavras-chave: *Cylindrocladium pteridis*. Sequenciamento do genoma. CAZymes. Estilo de vida. Patogenicidade fúngica. Escala diagramática. Resistência a doenças.

SUMMARY

GENERAL INTRODUCTION

Fungal plant pathogens can cause diseases worldwide that devasting economic, social, and ecological impacts. Among nectriaceous fungi, *Calonectria* species cause important diseases in numerous plant hosts and are broadly distributed in different regions of the world (LI et al., 2022; LIU et al., 2020). To date, there are at least 132 described species of *Calonectria*, including species that occupy a wide range of ecological niches and lifestyles (LIU et al., 2020; SALCEDO-SARMIENTO et al., 2021; SANCHEZ-GONZALEZ et al., 2022). These species reside in two main phylogenetic groups known as the Prolate Group and the Shaero-Naviculate Group, based on the shape of the vesicles in their conidiogenous apparatuses (LIU et al., 2020; PHAM et al., 2019). At the time, 11 species complexes are defined in *Calonectria*, distributed in Prolate Group (*C. brassicae, C. candelabrum, C. colhounii, C. cylindrospora, C. gracilipes, C. mexicana, C. pteridis, C. reteaudii* and *C. spathiphylli* species complexes), and Sphaero-Naviculate Group (*C. kyotensis* and *C. naviculata* species complexes) (LIU et al., 2020).

Calonectria spp. are soil-borne fungi (CROUS, 2002) and plant pathogenic *Calonectria* species cause important diseases, including cutting rot, damping-off, red crown rot, root rot, seedling rot, shoot blight, stem canker and leaf blight in numerous agricultural, horticultural and forestry crops (CROUS, 2002; LOMBARD et al., 2010). Eucalyptus, an economically important forestry species is severely threatened by several *Calonectria* species, which cause Calonectria leaf blight (CLB) disease in tropical and subtropical region in worldwide (BOSE et al., 2023). In Brazil, CLB is as one of the main constraints to *Eucalyptus* cultivation (ALFENAS et al., 2009, 2015; BOSE et al., 2023), and *C. pteridis* is the most common species reported in the commercial plantations, mainly in warmer and humid regions (ALFENAS et al., 2015; FERREIRA et al., 1995).

Calonectria pteridis Crous, M.J. Wingf. & Alfenas, 1993 (= *Cylindrocladium pteridis*) belongs to the *C. pteridis* complex, which currently comprises seven described species predominantly present in North and South America (LIU et al., 2020), and it was first described in 1926 causing brown spot on *Rumohra adiantiformis* (leather fern) in Florida (WOLF, 1926). Since then, it has been reported worldwide in several plant hosts besides *Eucalyptus* (FARR; ROSSMAN, 2022). In Brazil, *C. pteridis* was first reported in 1995 causing leaf spot and severe defoliation in *Eucalyptus grandis* plantations from Southeastern Bahia (FERREIRA et al., 1995). Since then, *C. pteridis* is the most common species found in the Brazilian eucalyptus plantations causing CLB (ALFENAS et al., 2009, 2013, 2015).

CLB is characterized by light gray small rounded or elongated leaf spots with a reddish-purple border, often surrounded by a callus, usually containing conidial sporulation on the abaxial leaf surface, easily observed in the early stages of infection. With the disease progress, the lesions coalesce, turn light brown and cover a large proportion of the leaf blade, leading to severe defoliation, which reduces the photosynthetic area and consequently the tree volumetric growth in highly susceptible genotypes (ALFENAS et al., 2013; FERREIRA et al., 1995).

The main control method for CLB is the use of resistant genotypes due to the existence of inter and intraspecific variability for resistance in *Eucalyptus* (ALFENAS et al., 2016). Considering that fungal pathogens have developed sophisticated mechanisms of pathogenicity and virulence to suppress host defense responses and cause disease (RAUWANE et al., 2020), understanding the biology and pathogenesis of plant pathogens is critical to the efficiency and durability of control methods. Thus, the identification of pathogenicity related genes is important to understand the mechanisms involved in pathogenicity and virulence of a plant pathogen (LELWALA et al., 2019).

Advances in genome sequencing have improved investigations in several areas of plant pathology, such as population genetics, evolution, reproductive biology and lifestyle and

pathogenicity (AYLWARD et al., 2017; LI et al., 2020), however, there is no genome sequence data available for *C. pteridis*. The genome sequencing of this important pathogen will facilitate the identification of putative genes related to mating behavior, pathogenicity and virulence, for example. Draft genome sequence of several fungal pathogens has been used to infer genes involved in pathogenicity (KUMAR et al., 2017; RIVERA et al., 2018; VERMA et al., 2016). Insights into the pathogenicity repertoire of *C. pteridis* and comparative genomics analysis will significantly expand the knowledge base of this genus, contributing to future investigations and better understanding of the mechanisms involved in pathogenicity and virulence.

Additionally, screening and selection of sources of resistance to CLB are required for planting of resistant eucalyptus genotypes. Usually, the selection has been based on defoliation (ALFENAS et al., 2016; REZENDE et al., 2019)**,** but use of visual assessment of disease severity has been recommended as a more practical method for assessing severity before the occurrence of drastic plant defoliation, reducing the effects of the natural senescence on assessments (FREITAS et al., 2019). Visual estimation of severity must provide accurate and reliable estimates. In this regard, the accuracy of the estimates can be improved when aided by standard area diagram sets (SADs) (DEL PONTE et al., 2017).

This study aimed: i) To obtain a *de novo* assembly of the nuclear genome of *C. pteridis* LPF059; ii) To characterize structurally and functionally the nuclear and mitochondrial genome of *C. pteridis* LPF059 and identify mating type idiomorphs in *C. pteridis*; iii) To classify the trophic phenotype of *Calonectria* species based on the CaZymes content and identify and compare putative pathogenicity- related gene profile in *Calonectria* species ; and iv) To evaluate the effect of the rater training on the quality of visual estimates of CLB and compare the accuracy and reliability of visual disease estimates using different SAD structures to quantify CLB severity.

REFERENCES

ALFENAS, A. C. et al. **Clonagem e doenças do eucalipto**. 2. ed. Viçosa-MG: Editora UFV, 2009.

ALFENAS, R. F. et al. Mass spore production and inoculation of Calonectria pteridis on Eucalyptus spp. under different environmental conditions. **Tropical Plant Pathology**, v. 38, n. 5, p. 406–413, set. 2013.

ALFENAS, R. F. et al. Diversity and potential impact of Calonectria species in Eucalyptus plantations in Brazil. **Studies in Mycology**, v. 80, p. 89–130, 2015.

ALFENAS, R. F. et al. Screening of Corymbia and Eucalyptus species for resistance to Calonectria pteridis leaf blight. **Forest Pathology**, v. 46, n. 1, p. 76–81, 2016.

AYLWARD, J. et al. A plant pathology perspective of fungal genome sequencing. **IMA Fungus**, v. 8, n. 1, p. 1–15, 1 jun. 2017.

BOSE, R. et al. Calonectria leaf blight of Eucalyptus: A global review. **Annals of Applied Biology**, v. 182, n. 1, p. 6–28, 1 jan. 2023.

CROUS, P. W. **Taxonomy and pathology of Cylindrocladium (Calonectria) and allied genera**. Saint Paul, Minnesota: The American Phytopathological Society, 2002.

DEL PONTE, E. M. et al. Standard area diagrams for aiding severity estimation: Scientometrics, pathosystems, and methodological trends in the last 25 years. **Phytopathology**, v. 107, n. 10, p. 1161–1174, 2017.

FARR, D.; ROSSMAN, A. **Fungal Databases, U.S. National Fungus Collections, ARS, USDA**, 2022. Available from: https://nt.ars-grin.gov/fungaldatabases/. Retrieved November 11 2022.

FERREIRA, F. et al. Mancha-de-pteridis: doença foliar de eucalipto em áreas tropicais brasileiras. **Fitopatologia Brasileira**, v. 20, p. 107–110, 1995.

FREITAS, R. G. et al. Genetic diversity and aggressiveness of Calonectria pteridis in Eucalyptus spp. **Plant Pathology**, v. 68, n. 5, p. 869–877, 1 jun. 2019.

KUMAR, A. et al. Draft genome sequence of Karnal bunt pathogen (Tilletia indica) of wheat provides insights into the pathogenic mechanisms of quarantined fungus. **PLOS ONE**, v. 12, n. 2, p. e0171323, 1 fev. 2017.

LELWALA, R. V. et al. Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in Colletotrichum tanaceti. **PLOS ONE**, v. 14, n. 5, p. e0212248, 1 maio 2019.

LI, J. Q. et al. Mating genes in calonectria and evidence for a heterothallic ancestral state. **Persoonia: Molecular Phylogeny and Evolution of Fungi**, v. 45, p. 163–176, 1 dez. 2020.

LI, J. Q. et al. Calonectria in the age of genes and genomes: Towards understanding an important but relatively unknown group of pathogens. **Molecular Plant Pathology**, v. 23, n. 7,

p. 1060–1072, 1 jul. 2022.

LIU, Q. L. et al. Reconsideration of species boundaries and proposed DNA barcodes for Calonectria. **Studies in Mycology**, v. 97, 1 set. 2020.

LOMBARD, L. et al. Species concepts in Calonectria (Cylindrocladium). **Studies in Mycology**, v. 66, p. 1–13, 2010.

PHAM, N. Q. et al. Ten new species of Calonectria from Indonesia and Vietnam. **Mycologia**, v. 111, n. 1, p. 78–102, 2 jan. 2019.

RAFIEI, V.; VÉLËZ, H.; TZELEPIS, G. The Role of Glycoside Hydrolases in Phytopathogenic Fungi and Oomycetes Virulence. **International Journal of Molecular Sciences**, v. 22, n. 17, 1 set. 2021.

RAUWANE, M. E. et al. Pathogenicity and Virulence Factors of Fusarium graminearum Including Factors Discovered Using Next Generation Sequencing Technologies and Proteomics. **Microorganisms**, v. 8, n. 2, 1 fev. 2020.

REZENDE, E. H. et al. Evaluation of resistance in eucalyptus grandis progenies against leaf spot of cylindrocladium and kirramyces. **Summa Phytopathologica**, v. 45, n. 3, p. 295–301, 1 jul. 2019.

RIVERA, Y. et al. Genome analysis of the ubiquitous boxwood pathogen Pseudonectria foliicola. **PeerJ**, v. 2018, n. 8, p. e5401, 24 ago. 2018.

SALCEDO-SARMIENTO, S. et al. Elucidating the interactions between the rust Hemileia vastatrix and a Calonectria mycoparasite and the coffee plant. **iScience**, v. 24, n. 4, p. 102352, 23 abr. 2021.

SANCHEZ-GONZALEZ, E. I. et al. Two new species of Calonectria (Hypocreales, Nectriaceae) causing Eucalyptus leaf blight in Brazil. **MycoKeys** 91: 169-197, v. 91, p. 169–197, 29 jul. 2022.

VERMA, S. et al. Draft genome sequencing and secretome analysis of fungal phytopathogen Ascochyta rabiei provides insight into the necrotrophic effector repertoire. **Scientific Reports** 2016 6:1, v. 6, n. 1, p. 1–14, 19 abr. 2016.

WOLF, F. A. BROWN LEAFSPOT OF LEATHER LEAF FERN. **Journal of the Elisha Mitchell Scientific Society**, v. 42, p. 55–62, 1926.

CHAPTER 1-Draft genome sequence of *Calonectria pteridis*, the causal agent of Calonectria leaf blight on eucalyptus

Marcia F. Queiroz,[a] Samuel A. Santos,[a]* Pedro M. P. Vidigal,[b] Túlio Morgan,[c] Fernando M. Fernandes,[a] Rafael F. Alfenas,[a] Talyta G. Zarpelon,[d] and Acelino C. Alfenas[a#]

[a]Laboratory of Forest Pathology, Department of Plant Pathology/Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, MG, 36570-900, MG, Brazil.

[b]Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Viçosa, MG, 36570-900, MG, Brazil.

[c]Departament of Microbiology/Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, MG, 36570-900, Brazil.

[d]Suzano Papel e Celulose, Aracruz, ES, Brazil.

Address correspondence to Acelino C. Alfenas, aalfenas@ufv.br.

*Present address: Samuel A. Santos, Plant Health Program, Fiber R&D Asia Pacific Resources International Ltd (APRIL), Pangkalan Kerinci 28654, Riau, Sumatra, Indonesia.

This chapter has been published as an original article at on Microbiology Resource Announcements: DOI: https://doi.org/10.1128/mra.00284-22.

ABSTRACT

Here, we report the draft genome sequence of *Calonectria pteridis*, the causal agent of Calonectria leaf blight in eucalyptus plantations in Brazil. The 58,373,473-bp genome assembly consists of 1,167 scaffolds, with a GC content of 50.21%. These genomic data can contribute to future studies involving the biology, adaptability, and pathogenicity of *C. pteridis*.

ANNOUNCEMENT

Calonectria pteridis (*Cylindrocladium pteridis*; Ascomycota, Sordariomycetes) is the causal agent of Calonectria leaf blight (CLB), a devastating foliar disease affecting eucalyptus plantations in Brazil (1, 2). Here, we present the characterization of the genome of *C. pteridis* LPF059, which was originally isolated from eucalyptus tree leaves with typical symptoms of CLB in Pará State, Brazil, and was deposited in the Culture Collection of the Laboratory of

Forest Pathology (LPF), Universidade Federal de Viçosa (Viçosa, Brazil) (3). Mycelium grown on potato dextrose broth medium for 48 h at 26°C and 180 rpm was used for DNA extraction with the DNeasy Plant minikit (Qiagen) according to the manufacturer's instructions. DNA quality and quantity were assessed through 1% agarose gel electrophoresis and use of a Qubit 2.0 fluorometer. A DNA library (insert size, ~350 bp) was prepared using the NEBNext Ultra II DNA library preparation kit (New England Biolabs, USA) following a standard protocol and was sequenced using the Illumina NovaSeq 6000 platform. Raw data (26,416,738 reads [2 × 150-bp paired-end reads]) were processed with FastQC v. 0.11.5 (4). Adapters and low-quality reads were trimmed with Trimmomatic v. 0.39 (Phred quality scores of ≥30) (5).

Trimmed reads (25,499,888 paired-end reads) were used to estimate the genome size by k-mer frequency and distribution using Jellifish v. 2.2.6 (6) to produce histograms of 17-, 21-, and 25-mers, which were processed by GenomeScope (7). The *ab initio* assembly of LPF059 genome was performed using SPAdes v. 3.12.0 (8), by testing different odd k-mer values (from 21 to 125). Scaffolding and gap-closing steps for the assembled contigs were performed by Redundans v. 0.14a with default parameters (9). To remove potential contaminants and mitochondrial sequences, all assembled scaffolds were compared with a bacterium-virus database and the mitochondrial genome of *Calonectria ilicicola* (GenBank accession number NC_046826.1) using BLASTn v. 2.6.0 (E values of $1e^{-25}$) (10). The completeness of the genome assembly was evaluated using BUSCO v. 5.0.0 (11) with the ascomycota_odb10 database.

The estimated genome size of *C. pteridis* LPF059 ranged from 56,565,321 to 56,655,434 bp. The genome assembly contains 1,167 scaffolds, with a total size of 58,373,473 bp, an N_{50} value of 322,939 bp, and a GC content of 50.21%. The longest scaffold was 1,316,024 bp. BUSCO completeness analysis showed 97.7% completeness (96.8% as single-copy genes and 0.9% as duplicated genes). This work provided a basis for future studies

on genetic population structure, comparative genomics, and transcriptomics, which may lead to a better understanding of the biology, adaptability, and pathogenesis of *C. pteridis*.

Data availability

The assembled genome sequence of *C. pteridis* isolate LPF059 was deposited in NCBI GenBank under accession number JAKZGU000000000. Raw data are available at the SRA under accession numbers SRR20046038 and SRR20046039. The BioProject and BioSample accession numbers are PRJNA812993 and SAMN26443383, respectively.

ACKNOWLEDGMENTS

This work was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Suzano S/A. The authors also thank the Núcleo de Análise de Biomoléculas (NuBioMol) for suportting in data analysis. NuBioMol is financially supported by the following Brazilian agencies: Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), CNPq, Financiadora de Estudos e Projetos (Finep) and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).

REFERENCES

1. Ferreira FA, Alfenas AC, Moreira AM, Demuner NL. 1995. Mancha-de-pteridis: doença foliar de eucalipto em áreas tropicais brasileiras. Fitopatol Bras 20:107–110.

2. Alfenas RF, Lombard L, Pereira OL, Alfenas AC, Crous PW. 2015. Diversity and potential impact of *Calonectria* species in Eucalyptus plantations in Brazil. Stud Mycol 80:89–130.

3. Alfenas RF, Pereira OL, Freitas RG, Freitas CS, Dita MAD, Alfenas AC. 2013. Mass spore production and inoculation of *Calonectria pteridis* on *Eucalyptus* spp. under different environmental conditions. Trop Plant Pathol 38:406–413.

4. Andrews S. 2016. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl 30:2114–2120.

6. Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764–770.

7. 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.

8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, 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.

9. Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 44:e113.

10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

11. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

CHAPTER 2- Genomic analyses of *Calonectria pteridis,* the causal agent of Calonectria leaf blight on eucalyptus

Marcia F. Queiroz,[a] Túlio Morgan,[b] Samuel A. Santos,[a]* Pedro M. P. Vidigal,[c] Fernando M. Fernandes,[a] Camila F. Paixão,[a] Rafael F. Alfenas,[a] and Acelino C. Alfenas[a#]

[a]Laboratory of Forest Pathology, Department of Plant Pathology/Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, MG, 36570-900, MG, Brazil.

[b]Departament of Microbiology/Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, Viçosa, MG, 36570-900, Brazil.

[c]Núcleo de Análise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Viçosa, MG, 36570-900, MG, Brazil.

Correspondence author: Acelino C. Alfenas, aalfenas@ufv.br.

*Present address: Samuel A. Santos, Plant Health Program, Fiber R&D Asia Pacific Resources International Ltd (APRIL), Pangkalan Kerinci 28654, Riau, Sumatra, Indonesia.

Abstract

Calonectria pteridis is the main causal agent of Calonectria leaf blight in commercial plantations of eucalyptus in Brazil. In this study, we characterized the 54.8 Mb genome of *C. pteridis* LPF059, which contains 17,388 coding sequences comprising 44% of the genome – 16,964 coding proteins genes, 78 ribosomal RNA, and 346 transfer RNA. A total of 1,775 sequences were predicted as secreted proteins and about 6% of the genome consists of repetitive sequences. GO terms were assigned to approximately 89% of the coding sequences with the most representative functional categories related to biosynthesis of secondary metabolites and carbohydrate metabolism. The mitochondrial genome was found to be 60.8 kb with introns belonging to the LAGLIDADG and GIY-YIG family classes of homing endonuclease. Analysis of mating type locus and PCR amplification confirmed that *C. pteridis* LPF059 is heterothallic. Comparative genomic analysis of *C. pteridis* with other *Calonectria* species showed high content of CAZymes and secondary metabolites among them. The majority of *Calonectria* species were considered as vasculartrophs, except *C. henticotiae* and *C. pseudonaviculata* (mesotrophs). Annotation of unique orthologous clusters of *C. pteridis* and exclusive orthologous clusters shared by *C. pteridis* related to secondary metabolism, transport, and chitin synthase may play an important role in its pathogenicity or contribute to stress tolerance during the saprophytic phases. In addition, phylogenomic analysis of a concatenated alignment of

2,308 single-copy orthologs genes revealed that grouping of *Calonectria* species is in line with previous phylogenetic analyses based on DNA barcodes. Our results provide valuable genomic data for further investigation and expand our knowledge on the biology, lifestyle and pathogenicity of *C. pteridis*.

Keywords: *Cylindrocladium pteridis*. CAZymes. Mitochondrial genome. Mating type. Pathogenicity.

INTRODUCTION

Calonectria species (*Hypocreales*, *Nectriaceae*) are broadly distributed in different regions of the world. At the time, there are at least 132 described species of *Calonectria*, including species that occupy a wide range of ecological niches and lifestyles (Liu et al. 2020; Salcedo-Sarmiento et al. 2021; Sanchez-Gonzalez et al. 2022).

Plant pathogenic *Calonectria* species cause important diseases on many agricultural, horticultural and forestry plants, comprehending more than 335 plant species, distributed among nearly 100 plant families (Crous 2002; Lombard et al. 2010a). For instance, *C. pauciramosa* has been reported on numerous woody, herbaceous, and ornamental plants worldwide, where it causes disease, such as cutting rot, damping-off, root and leaf blight (Crous 2002; Lombard et al. 2010b; Li et al. 2021). *C. ilicicola* has been reported mainly as economically important pathogen on soybeans and peanuts, causing Cylindrocladium black rot (CBR) and Red crow rot (RCR), respectively (Gai et al. 2017; Liu et al. 2021). Other disease, boxwood blight caused by *C. henricotiae* and *C. pseudonaviculata* is considered a major threat to boxwood (*Buxus* sp.) production (Yang et al. 2021). In forestry crops, most reports of *Calonectria* species occur in *Acacia* spp, *Pinus* spp. and *Eucalyptus* spp., where several species are responsible for a variety of disease symptoms, such as damping-off, cutting rot, stem cankers, root and collar rot, and leaf blight (Crous 2002; Lombard et al. 2010a, c; Soares et al.

2019). To date, 37 species of *Calonectria* distributed in eight species complexes cause Calonectria leaf blight (CLB) of *Eucalyptus*(Sanchez-Gonzalez et al. 2022; Bose et al. 2023)., a serious disease in eucalyptus nurseries and plantations in tropical and subtropical regions, such as India, Australia, Indonesia, Vietnam, China, and Brazil (Bose et al. 2023). In Brazil, C. *pteridis* is the most common species reported in the eucalyptus plantations, causing CBL, mainly in warmer and humid regions (Ferreira et al. 1995; Alfenas et al. 2015).

Calonectria pteridis belongs to the *C. pteridis* complex, which currently comprises seven described species predominantly present in North and South America. In Brazil, *C. pteridis* was first reported in 1995 causing leaf spot and severe defoliation in *Eucalyptus grandis* plantations from Southeastern Bahia (Ferreira et al. 1995).

CLB is characterized by light gray small rounded or elongated leaf spots with a reddish-purple border. As the disease progress, the lesions become light brown and may cover a large proportion of the leaf blade, leading to serious defoliation, which is a limiting factor of eucalyptus trees growth (Alfenas et al. 2013).

Despite being one of the most devastating diseases in *Eucalyptus* plantations, little is known about genetic, biology, adaptability, and mechanisms of pathogenicity of *Calonectria* species causal agents of CLB, including *C. pteridis*. So far, however, a small number of studies have focused on analyzing the mechanisms of pathogenicity in *C. pseudoreteaudii* on Eucalyptus (Chen et al. 2015; Ye et al. 2017, 2018). *Calonectria pseudoreteaudii* is an aggressive species in eucalyptus plantations in China and Southeast Asia (Ye et al. 2017) and, these previous studies suggested that the production of toxin and cell-wall-degrading enzymes are among the mechanisms utilized to establishment of CLB.

The availability public genomic sequence data have benefited the research community in the study of fungal genetic, fungal biology and provides a better understanding of genome evolution, pathogenicity and lifestyle of pathogens (Aylward et al. 2017; Lelwala et al. 2019).

In this regard, it should also be noted that genome data are not used but provide background for observational and experimental data. More recently, a greater number of *Calonectria* genomic sequences have been publicly available, including the genome sequence of *C. pteridis* (Queiroz et al. 2022). The genome sequence of this important pathogen is a good source for identifying putative genes related to mating type, pathogenicity and virulence. Insights into the pathogenicity repertoire of *C. pteridis* and comparative genomics will significantly expand the knowledge base of this genus.

Our aim in this study was to characterize the nuclear and mitochondrial genome sequence of *Calonectria pteridis* LPF059, investigate its mating behavior and compare it with other *Calonectria* species to reveal features that may provide a better understanding of the pathogenicity and lifestyle of these important plant pathogens.

MATERIAL AND METHODS

Nuclear genome annotation

The 1,167 scaffolds sequences of the genome of *C. pteridis* LPF059 (GenBank accession number JAKZGU000000000) were submitted the *ab initio* gene prediction using Braker v. 2.1.6 pipeline C (https://github.com/Gaius-Augustus/BRAKER) (Hoff et al. 2019) using default parameters, selecting "--softmasking" option and with protein evidence derived from hypocreales_odb10 lineage dataset (total of 48,588 genes) and proteins of *C. ilicicola* F018 (total of 17, 308 genes) (Liu et al. 2021) as a closely related species. The completeness of the gene prediction was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 5.0.0 (https://busco.ezlab.org/) (Simão et al. 2015) with the ascomycota_odb10 lineage dataset. The rRNA genes were annotated by Barrnap v. 0.9 (https://github.com/tseemann/barrnap) using Eukaryota database, while tRNA genes were identified with tRNAscan-SE v. 2.0.6 (http://lowelab.ucsc.edu/tRNAscan-SE/) using

Eukaryotic database and default parameters (Chan et al. 2021). Repeat sequences were predicted and classified using RepeatModeler v.2.0.1 (http://www.repeatmasker.org/RepeatModeler/) and RepeatMasker v. 4.1.1 (http://www.repeatmasker.org). For that, RepeatModeler v. 2.0.1 was used to generate a library containing *ab initio* models of repetitive elements. The resultant repeat library was then combined with the Fungi database of RepBase library v.20181026 (Bao et al. 2015) and Dfam v.3.4 database (Storer et al. 2021). Combined library was then used in RepeatMasker v. 4.1.1 to predict repetitive sequences.

Functional annotations for predicted proteins, including Gene Ontology terms and COG functional categories were obtained using EggNOG-mapper v. 2.1.9 (http://eggnog-mapper.embl.de/) and PANNZER2 (http://ekhidna2.biocenter.helsinki.fi/sanspanz/) web servers.

Mitochondrial genome

The mitochondrial genome (mtDNA) was assembled together with the sequences of the nuclear genome of *C. pteridis* isolate LPF059 obtained by Queiroz et al. (2022) using the Illumina NovaSeq 6000 platform with 2x150-bp paired-end reads. The *ab initio* assembly was done using SPADES v. 3.12.0 (Bankevich et al. 2012) and mitochondrial genome was identified by performing a BLAST search on nuclear genome assembly using of *C. ilicicola* (NC_046826.1). The resulting mitochondrial sequences were annotated using MITOS web server (http://mitos.bioinf.uni-leipzig.de/index.py) (Bernt et al. 2013) and MFannot tool (https://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) with the genetic code set to four (mold). Complete ribosomal RNA genes were checked by alignment with *Calonectria* mitochondrial genome sequences of the NCBI non-redundant (nr) database using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) searches with default parameters. All

annotations for the *C. pteridis* LPF059 mitochondrial genome were illustrated in circular a plot using the GenomeVx tool (http://wolfe.ucd.ie/GenomeVx/).

Identification of mating type idiomorphs in *C. pteridis* LPF059 and PCR analysis

MAT1 idiomorphs were identified using the amino acid sequence of MAT1 idiomorphs reported in *Calonectria* species (MAT1-1-1, MAT1-1-2, MAT1-1-3, MAT1-2-1, and MAT1-2-12) (LI et al., 2020) as query in a local BLASTp search (e-value cutoff 1e-10) against a database created from amino acid sequence of *C. pteridis* LPF059. MAT analysis locus was performed in the CLC main workbench v. 21.0.5. The functional domains of the MAT genes were determined using the Conserved Domain search on NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al. 2011).

To provide PCR-based identification of the MAT1 idiomorph, DNA was extracted using the protocol described by (Specht et al. 1982). DNA quality and quantity were assessed through 1% agarose gel electrophoresis and NanoDrop 2000c spectrophotometer (Thermo Scientific), respectively. MAT1-1-1 (primers Cal_MAT111_F and Cal_MAT111_R) and MAT1-2-1 (primers Cal_MAT121_F and Cal_MAT121_R) genes were amplified by PCR from nine isolates of *C. pteridis* following protocol described by (Li et al. 2020). The amplified fragments were separated on 1% agarose via electrophoresis and visualized under Uv light.

Orthogroups search and phylogenenomic analysis

To construct the phylogenomic tree of *C. pteridis* and other *Calonectria* species, the publicly available genome sequences of 15 *Calonectria* species were downloaded from NCBI Genbank. Then, coding DNA sequences (CDS) prediction was performed using Braker v. 2.1.6 pipeline C (Hoff et al. 2019) as described above for *C. pteridis* LPF059. Additionally, the predicted proteomes of 8 other Ascomycota fungi were downloaded from NCBI Genbank. The completeness of the genes predicted for all species was evaluated using Benchmarking

Universal Single-Copy Orthologs (BUSCO) v. 5.0.0 (https://busco.ezlab.org/) (Simão et al. 2015). Genomes with < 5% of missing genes against the ascomycota_odb10 lineage dataset were used in the analysis (Table 1).

Orthogroups were identified using OrthoFinder v.2.5.4 (https://github.com/davidemms/OrthoFinder) (Emms and Kelly 2019), with multiple sequence alignments generated using MAFFT v.7.305 (Katoh and Standley 2013) and maximum likelihood trees using FastTree v.2.1.9 (Price et al. 2010). OrthoFinder infers orthogroups at each node in the species tree by analyzing the rooted gene trees.

Concatenated and trimmed sequence alignments from the single-copy orthologous genes shared across all 24 fungal species were extracted from the OrthoFinder results and used for maximum likelihood phylogenetic analysis with RaxML-NG v.1.0.3 (Kozlov et al. 2019) Phylogenomic tree was constructed using the best fitting model [general matrix (JTT)+I+G4] obtained using modeltest-NG (Darriba et al. 2020) and 1000 bootstrap replicates. *Sclerotinia sclerotiorum* and *Botrytis cinerea* was used as outgroup to root the phylogenomic tree.

Comparative genomic analysis

Orthogroups present among the *Calonectria* species (Tabela 1) were retrieved from Orthofinder results. The core proteome was defined by orthologous clusters conserved among all *Calonectria* species. Species-unique orthologous clusters were defined as orthogroups present in a species and absent in every other species.

To identify potential secreted proteins from the predicted proteomes of 16 *C*alonectria species, the web analysis tool SECRETOOL was used (http://genomics.cicbiogune.es/SECRETOOL/Secretool_Full.php) with default parameters, except for SignalP (cut-off probability 0.5) (Cortázar et al. 2014). Furthermore, Deeploc-1.0 (http://www.cbs.dtu.dk/services/DeepLoc/) (Almagro et al. 2017) was used for prediction of

the subcellular localization by creating a recurrent neural network relying on protein sequence information per se. Proteins determined as being secreted by SECRETOOL and/or located in the extracellular space by Deeploc were selected as secreted proteins.

Putative virulence, pathogenicity and effector-associated proteins were identified by performing BLASTp searches (E value of $1e^{-10}$) of predicted proteomes against Pathogen-Host Interaction database (PHI-base) (Urban et al. 2017) version v.4.12 and Database of Fungal Virulence Factors (DFVF). The identification of proteins potentially involved in carbohydrate metabolism (CAZymes) was done using dbCAN2 meta server (https://bcb.unl.edu/dbCAN2) (Zhang et al. 2018). In addition, the trophic lifestyle prediction was carried out using the CATAStrophy pipeline with dbcan_version 10 (Hane et al. 2020). To predict potential secondary metabolite biosynthetic (SMB) gene clusters was used the Fungal AntiSMASH server (https://antismash.secondarymetabolites.org/) (Blin et al. 2019).

Functional annotation of predicted proteomes for other *Calonectria* species was done following the same methods as described above for *C. pteridis* LPF059.

RESULTS

Gene annotation and repetitive sequences content of *C. pteridis* LPF059 nuclear genome

A total of 17,812 genes was predicted in the nuclear genome, including 17,388 CDS, 78 rRNA (5S =75, 5.8S =1, 18S=1, 28S=1), and 346 tRNA. BUSCO completeness analysis showed 99.8% (96.5% as single-copy and 3.3% as duplicated genes).

The repetitive content represents 5.93% (3,462,544 bp) of the genome, with the majority of sequences assigned as unknown (2.94%), followed by transposable elements (TEs) types LTR (1.39%), simple repeats (0.71%), DNA transposons (0.37%), TEs type LINE (0.24%), low complexity (0.12%), rolling circles (0.07%), small RNA (0.05%), TEs type SINEs (0.03%) and satellites (0.02%).

Functional annotation for *C. pteridis* LPF059 genome

About 89% (15,383 sequences) of CDS was assigned with GO terms (Figure 1A). The most representative categories for cellular component were related to "cell and cell part" (GO:0044464, GO:0005623; 4894), "organelle" (GO:0043226; 4168) and "membrane" (GO:0016020; 2662). For molecular function category, it was found "organic cyclic compound binding" and "heterocyclic compound binding"(GO:0097159, GO:1901363; 3234, 3223), "ion binding" (GO:0043167, 3135), "hydrolase activity" (GO:0016787, 1988) and "oxidoreductase activity" (GO:0016491, 1900); while for biological process, the most common were "organic substance metabolic process" (GO:0071704, 5032), "primary metabolic process" (GO:0044238, 4713), "cellular metabolic process" (GO:0044237, 4683), "nitrogen compound metabolic process" (GO:0006807, 4244). In addition, a total of 12, 929 (~ 75%) predicted CDS were classified on clusters of orthologous groups functional categories related to biosynthesis of secondary metabolites, transport, and catabolism (Q; 1142), and carbohydrate transport and metabolism (G; 1058) (Figure1B).

Mitochondrial genome size and gene annotation

The mtDNA of *C. pteridis* LPF059 was found in a single scaffold with 60,801 bp (Figure 2), comprising 12 core protein-coding genes (*cox1*, *cox2*, *cox3*, *cob*, *nad1*, *nad2*, *nad4, nad4L, nad5, atp6, atp8, atp9*), 26 tRNAs, two rRNAs, one ribosomal protein S3 (*rps3*), one ribosomal protein S5 (*rps5*), and the RNase P (*rnpB*). Seven (*cox1*, *cox2*, *cox3*, *cob*, *nad1*, *nad2*, *nad5*) out of 12 protein-coding genes are interrupted by introns belong to group I, which contain open read frames (ORFs) coding for homing endonucleases (LAGLIDADG and GIY-YIG families). In addition, 7 ORFs were detected in intergenic regions – three LAGLIDADG, two GIY-YIG, and two encoding hypothetical proteins (Figure 2).

Mating type idiomorph in *C. pteridis* LPF059 genome

Local BLASTp searches identified three proteins of *C. pteridis* genome with high level of cover and similarity to MAT1-1 idiomorph of *C. pseudopteridis*, *C. pauciramosa,* and *C. amazonica* (Figure 3). MAT1-1 contained a predicted protein with an MATalpha_HMG box domain (MAT 1-1-1; cdd:pfam04769), a predicted protein containing a PPF domain (MAT1-1-2; cdd:pfam17043) and a predicted protein containing an HMG box domain (MAT1-1-3; cdd:cd01389), located within an approximately 4.3 kb region. The predicted MAT1-1-1, MAT1-1-2 and MAT1-1-3 genes have a CDS about 1.2 kb (2 introns), 1.4 kb (4 introns) and 366 bp (1 intron), respectively (Figure 3). In addition, validation by PCR amplification confirmed the presence of *MAT1-1* in 5 isolates of *C. pteridis* and *MAT1-2* in other 4 different isolates (Figure 4).

Phylogenomic analysis

A total of 4,712 orthogroups was shared among all species, of these 2,308 was assigned as single-copy genes and used for phylogenomic analysis of 16 *Calonectria* species plus 8 other ascomycetes fungi (Figure 5). Phylogenetic trees generated by Orthofinder v.2.5.4 pipeline using FastTree v.2.1.9 or by RaxML-NG v.1.0.3 presented similar topology, both with all branches maximally supported. ML analysis identified two major clades representing the *Calonectria* species (Figure 5), which harbor species belong to 8 species complexes, including *C. colhounii*, *C. reteaudii*, *C. pteridis*, *C. candelabra*, *C. cylindrospora*, *C. mexicana* species complexes in clade 1, and *C. naviculata* and *C. kyotensis* species complexes in clade 2.

Calonectria pteridis was found closely related to *C. colhounii* and *C. reteaudi* species complex. Among all ascomycete fungi evaluated in this study, *Neonectria ditissima* was more related with *Calonectria* species (Figure 5).

Comparative genomic analysis of *C. pteridis* with other *Calonectria* species

The total predicted putative proteins ranged from 13,665 to 23,467 among 16 *Calonectria* species (Table 2). Of these, about 1,300 to 2,300 sequences were annotated as secreted proteins (Table 2). For *C. pteridis*, around 1,775 sequences were found as secreted proteins, representing approximately 10% of its proteome. In addition, following their genome size, *C. naviculata* and *C. ilicicola* presented large proteome than other *Calonectria* species (Table 2). In contrast, *C. pseudonaviculata* and *C. henricotiae* showed the smallest genomes and consequently the smallest proteomes as well. The same trend was also observed for genome size vs. secretome, with few exceptions (Table 2). Overall, the secretome ranged from 9.6 to 11.2 % of the predicted proteome for all *Calonectria* species.

Consistent with its larger genome and proteome size, *C. naviculata* had the highest number of homologs in PHI-base (n= 5,954) and DFVF (n= 1,514), followed by C. *multiphialidica* (n= 5,034; 1401) and *C. ilicicola* (n= 4,949; 1,359) (Figure 6A; 6B). A total of 4,325 and 1219 homologs were recorded in *C. pteridis* from the PHI-base and DFVF, respectively, with a profile similar to *C. crousiana* (n= 4305; 1215). Futher, genes correlated with unaffected pathogenicity (UP) and reduced virulence (RV) were identified in higher frequency for all species (Figure 6A). Similar results were obtained in the analyses of CAZymes, where *C. naviculata* had the higher number of sequences (n= 926) assigned to different CAZymes classes, followed by C. *multiphialidica* (n=886) and *C. honkhonghensis* (n= 856). The CAZymes repertoire of *C. pteridis* were relatively smaller showing the same number of sequences as *C. crousiana* (n =752). The most abundant classes of CAZymes were glycosyl hydrolases (GH) for all species (Figure 6C). The antiSMASH analyses identified that *C. naviculata* contained the highest number of SMB clusters (n= 132), followed by *C. pseudoretaudii* (n = 123) and *C. multiphialidica* (n = 122). The composition of the SMB clusters of *C. pteridis* was most similar to *C. fujianensis*. SMB clusters associated with

polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) are predominant in the predicted proteomes of all species (Figure 6D). Only *C. naviculata* contained one cluster of Arylpolyene and one of Hserlactone. On the other hand, only one cluster of Fungal-RiPP was detected on *C. pteridis* and *C. hongkongensis* (Figure 6D). A small number of homologs in PHI-base and DFVF, as well as CAZymes repertoire and clusters of SMB were identified for *C. henricotiae* and *C. pseudonaviculata* (Figure 6D).

The trophic phenotypes of *Calonectria* species analyzed was predicted using the complete CAZymes repertoire. Relative centroid distance (RCD) scores were calculated using CATAStrophy. Species were predicted as members of a major trophic class where RCD = 1 and assigned additional affinities for other classes or sub-classes where RCD $\geq$ 0.95. All *Calonectria* species were classified according to common trophic terms as necrotrophs. However, considering the alternative classification, most species were defined as vasculartroph in a major trophic class and sub-class affinity, except *C. henricotiae* and *C. pseudonaviculata*, which were classified in a major trophic class as mesotroph and vasculartroth in sub-class affinity (Table 3).

In the orthology analysis, the predicted proteome of these *Calonectria* species formed 19,267 orthologous clusters of which 9,893 (51.34%) were shared among all species, representing the core proteome conserved among these species (Figure 7A).

The species-unique clusters were most abundant in *C. naviculata* with 329 unique clusters, consisting of 1,222 proteins. A total of 50 clusters was found exclusively in *C. ilicicola* with 81 proteins, while *C.pteridis*, *C.multiphialidica,* C. *pseudoreteaudii* and *C.leucothoes* had 23, 16, 8 and 6 unique clusters with 57, 48, 21 and 12 proteins, respectively. A smaller number of unique clusters were found in *C. hawksworthii, C. pauciramosa*, *C. crousiana*, *C. hongkongensis*, which had 3, 3, 2 and 2 unique clusters with 8, 6, 4, and 6 proteins, respectively. Only one cluster with 2 proteins was unique for *C. honghensis*, *C. kyotensis*, *C. henricotiae* and

C. pseudonaviculata. No unique cluster was found for *C. fujianensis* and *C.aciculata* (Figure 7A.

Analysis of secretome revelated 2,524 orthologous clusters shared among all species, representing approximately 25% of core proteome. As for proteome, *C. naviculata* contained the most abundant species-unique clusters of secreted proteins. Nonetheless, *C. pteridis* showed more species-unique clusters (n=8) of secreted proteins than *C.ilicicola* (n = 2). No unique cluster of secreted proteins was found *C.pauciramosa*, *C. crousiana*, *C. hongkongensis*, *C.honghensis*, *C. kyotensis*, *C. henricotiae*, *C.pseudonaviculata*, *C. fujianensis* and *C.aciculata* (Figure 7B).

Investigations of unique orthologous clusters showed that only *C. ilicicola*, *C. leucothoes*, *C. multiphialidica* and *C. naviculata* had clusters related to pathogenicity and virulence pathways (Table 4). The pathogenicity-associated proteins in unique clusters of *C. ilicicola* were related secondary metabolite biosynthetic process (GO:0044550) with homologs in the DFVF identified as a reducing polyketide synthase (PKS) involved in mycotoxin biosynthesis. Similarly, *C. leucothoes* showed one unique cluster related to the secondary metabolite biosynthetic process (GO:0044550) with homologs in the DFVF described as a non-ribosomal peptide synthetase (NPKS) involved in siderophore and mycotoxin biosynthesis. Furthermore, the pathogenicity-associated proteins in unique cluster of *C. multiphialidica* were related oxidoreductase activity (GO:0016491) with homologs in DFVF, which are described as NADP-dependent oxidoreductase RED1 involved in mycotoxin biosynthesis. Finally, in *C. naviculata*, pathogenicity-associated proteins in four unique clusters were related to kinase activity (GO:0016301), oxidoreductase activity, acting on CH-OH group of donors (GO:0016614), transmembrane transport (GO:0055085), and acyltransferase activity (GO:0016747), which had homologs in DVDF defined as transcription factor SKN7, alcohol oxidase, MDR efflux pump ABC3, acetyl-CoA acetyltransferase erg10B, respectively (Table

4). Moreover, some of the unique clusters included secreted proteins (OG0017222; OG0016333) (Table 4).

Additionally, results of orthologous clusters search revealed that *C. pteridis* had the highest number of clusters shared exclusively with *C. naviculata* (n = 130), followed by *C. hawksworthii* (n= 22), *C.ilicicola* (n=15), *C. pauciramosa* (n=12), *C. leucothoes* (n=12), *C. pseudoreteaudii* (n=8) and *C. multiphialidica*. On the other hand, *C. pseudonaviculata*, *C. crousiana*, *C. hongkongensis*, *C. kyotensis*, *C. henricotiae*, *C. fujianensis*, *C. aciculata* and *C. hongnensis* shared a small number of exclusive clusters with *C. pteridis* (Figure 8). Analysis of exclusive orthologous clusters shared with *C. pteridis* showed that *C. naviculata* shared two exclusive clusters associated to pathogenicity and virulence pathways (Table 5), which were assigned to GO terms related to oxidoreductase activity, acting on CH-OH group of donors (GO:0016614) and to the secondary metabolite biosynthetic process (GO:0044550) with similarity to sequences in DFVF related to 3-ketoacyl-CoA reductase activity and polyketide synthase/peptide synthetase, respectively. Further, mutants for these sequences in PHI database were annotated with phenotypes of "reduced virulence/loss of pathogenicity" and "effector" (Table 5).

C. crousiana, *C. pseudonaviculata* and *C. hawksworthii* shared one exclusive cluster with *C. pteridis* each related to transferase activity (GO:0016740), chitin synthase activity (GO:0004100) and nucleotide binding (GO:0000166), which had homologs in DFVF described as non-reducing polyketide synthase CTB1, chitin synthase and ABC transporter FGM5/MRP-like transporter 1, respectively. The proteins included in these clusters were similar to sequences that resulted in phenotypes of reduced virulence and/or loss of pathogenicity after mutation according PHI database (Table 5).

DISCUSSION

Calonectria pteridis, a member of the *C. pteridis* species complex, is the most common species reported in the commercial eucalyptus plantations in Brazil that causes CLB, mainly in warmer and humid regions (Alfenas et al. 2015). *Calonectria pteridis* LPF059 has approximately 44% (25,583,181 bp) of the genome comprised of coding sequence and only 5.93% (3,462,544 bp) of repetitive sequences. Similar to *C. pseudoreteaudii* (Ye et al. 2018), the most of classified repetitive sequences are TEs (~34%). The number of coding sequences in *C. pteridis* LPF059 was more similar to *C. pseudoreteaudi* and *C. crousiana* predicted in this study. The most represented functional classes in the *C. pteridis* LPF059 proteome were related to unknown function, secondary metabolites biosynthesis and carbohydrate metabolism. These same functional categories have been reported for rapidly evolving gene families of *C. naviculata* and *C. multiphialidica* (Rogers et al. 2022).

The mitochondrial sequences available in *Calonectria* species are scarce (Gai et al. 2020; Salgado-Salazar et al. 2022). Among fungal species, the mitochondrial genome size and genes content can vary among species and even among strains within of the same species (Sandor et al. 2018; Gai et al. 2020). The size of the *C. pteridis* LPF059 mitochondrial genome (60,801 bp) were comparable to that of *C.ilicicola* GDBL60, which had 60,843 bp (Gai et al. 2020). Typically, fungal mitochondrial genome contains 14 protein-coding genes involved in the ATP synthesis and in the electron transport chain (Sandor et al. 2018). In contrast, *C. pteridis* LPF059 had 12 core protein-coding genes consistent with a previous study that, when evaluating 685 mitochondrial genomes from different fungal species found that approximately 36,6% of mitochondrial genomes analyzed have absent at least one of the 14 core coding genes (Fonseca et al. 2021). All intra and intergenic introns belong to the LAGLIDADG and GIY-YIG family classes of homing endonuclease.

Homothalic and heterothallic mating systems have been reported in *Calonectria* species (Li et al. 2020). In a previous study, PCR amplification of the mating genes was used to investigate mating strategies of 70 *Calonectria* species, but *C. pteridis* was not included. Our *in silico* analysis identified a single mating type idiomorph in *C. pteridis* LPF059 suggesting that this species is heterothallic, corroborating previous statement for species in *C. pteridis* complex (Liu et al. 2020). The MAT1-1 idiomorph identified in *C. pteridis* LPF059 harbor the MAT1-1-1, MAT1-1-2 and MAT1-1-3 genes according to a previous study of heterothallic *Calonectria* species (Li et al. 2020). *Calonectria pteridis* was confirmed as heterothallic fungus, requiring a partner of the alternate mating type to initiate the sexual life cycle, by amplification of either MAT1-1-1 or MAT1-2-1 in different isolates. PCR amplification of the mating genes using primers designed by Li et al. (2020) can be employed for generating information in population genetics and epidemiology studies of *C. pteridis.*

The genus *Calonectria* is a diverse group that accommodates at least 132 species presently (Liu et al. 2020; Sanchez-Gonzalez et al. 2022; Bose et al. 2023), including species with diverse lifestyles, such as saprobe, mycoparasite and several aggressive plant pathogens (Lombard et al. 2010a; Liu et al. 2020; Salcedo-Sarmiento et al. 2021). At the time of initiating this study, few *Calonectria* genomes were publicly available for analysis and comparison.

The maximum likelihood tree constructed using single copy orthologous genes showed 100% confidence in tree topology for *Calonectria* spp. and other ascomycete fungi with all branches supported by bootstrap values of 100.

Consistent with previous results using DNA barcodes (Pham et al. 2019; Liu et al. 2020), there was a clear separation of the species complexes in accordance with two main phylogenetic groups, clade 1 and clade 2, corresponding respectively to Prolate Group and the Sphaero-Naviculate Group, which are differentiated based on the shape of the vesicles emerging from their conidiogenous apparatuses (Pham et al. 2019). Further, as verified by Liu et al. (2020), *C.*

honghensis is closely related to *C. aciculata*. These species have been distinguished by the dimensions of its macroconidia (Liu et al. 2020). Similarly, C. *pseudonaviculata* is closely related to *C. henricotieae* and, they have been distinguished genetically based on fungicide sensitivity, thermotolerance and genome wide SSRsequence profiles (Gehesquière et al. 2016; LeBlanc et al. 2019). *Calonectria pseudoturangicola* CMW 47496 available in GenBank was reduced to synonymy with *C. kyotensis* by Liu et al. (2020). Thus, *C. pseudoturangicola* was treated as *C. kyotensis* in this study and, it is closely related to C. *hongkongensis*. These species have been distinguished by the dimensions of its macroconidia and vesicles (Liu et al. 2020).

To better understand the molecular underpinnings of pathogenicity in *Calonectria*, predicted proteome of 16 species were screened to identify putative genes involved in PHI process against PHI-base, fungal-specific virulence genes against DFVF, CAZymes repertoire and SMB gene clusters. Thus, considering these multiple annotation analyses, *C. naviculata* exhibited the highest frequency of sequences annotated in all analyses, followed by *C. multiphialidica*, except for SMB clusters. In this case, *C. pseudoreteaudii* and *C. multiphialidica* had a similar number of SMB clusters. On the order hand, *C. henricotiae* and *C. pseudonaviculata* had smaller number of sequences annotated in all analyses. Interestingly, these results indicated a possible correlation between proteome size and the content of pathogenicity/virulence genes of *Calonectria* species examined in this study, so that *C. naviculata* had the largest proteome and C. *henricotiae* and *C. pseudonaviculata* the smallest proteomes. These species belong to the *C. naviculata* species complex, and the time of differentiation was estimated to be approximately 26 million years ago (Malapi-Wight et al. 2019; Chen et al. 2022) with number of gene families expanded by 1794 and contracted by 979 in *C. naviculata* (Chen et al. 2022).

The rapid expansion of the gene families involved in pathogenesis has been reported for *C. naviculata* and *C. multiphialidica*, while rapid contraction has been verified for *C.*

henricotiae and *C. pseudonaviculata*, explained by the different lifestyle between these species (Rogers et al. 2022). Indeed, the loss and gain of pathogenicity gene families play an important role in host specificity and pathogenic adaptation (Baroncelli et al. 2016). *Calonectria naviculata* and *C. multiphialidica* have not been reported as plant pathogens but have been identified in leaf litter and soil (Liu et al. 2020), while *C. pseudonaviculata* and *C. henricotiae* cause disease in *Buxus* sp. (Malapi-Wight et al. 2019). *Calonectria pseudonaviculata* also causes disease in *Sarcoccocca* and *Pachysandra*, other plants in the Buxaceae family (Malapi-Wight et al. 2016; LeBlanc et al. 2018).

The most abundant CAZyme class among *Calonectria* species were glycoside hydrolases (GH) and auxiliary activities (AA), except for *C. henricotiae* and *C. pseudonaviculata*, which showed a greater number of GH and glycosyltransferase (GT) classes. GH plays a role in the catabolism of carbohydrates based on the hydrolysis of glycosidic bonds of glycosides and are often known as cell wall degrading enzymes (CWDEs) (Rafiei et al. 2021). On the other hand, GT carries out the transfer of sugar moieties resulting in the formation of glycosidic bonds during the biosynthesis of disaccharides, oligosaccharides and polysaccharides (Lairson et al. 2008).

Considering that degrade plant cell wall carbohydrates can be essential during the infection and decomposition of host plant tissue, CAZymes profiles can be used as indicators of trophic phenotypes. Biotrophic fungi tend a have reduced set genes enconding CAZymes than necrotrophic and hemibiotrophic fungi (Zhao et al. 2013).

The trophic behaviors and lifestyles for *Calonectria* species remains less explored or unclear (Salgado-Salazar et al. 2022; Rogers et al. 2022). The traditional classification of trophic behaviors is mainly based on observational data. CATAStrophy method provides a non-biased way to predict the trophic phenotypes of filamentous plant pathogens based on their

CAZyme gene content, as a complementary method to the traditional tripartite trophic classification in biotroph, hemibiotroph and necrotroph (Hane et al. 2020)

Our analysis of trophic phenotypes using CATAStrophy pipeline based on CAZymes content revealed an interesting difference among *Calonectria* species. In a common classification, all *Calonectria* species in this study are classified as necrotrophic, but based on new classification, most species are vasculartrophic, while *C. henricotiae* and *C. pseudonaviculata* was predicted primarily mesotrophic, showing affinity with vasculartrophic. Vasculartrophs are most similar in CAZymes content to the PB (Broad host-range polymertrophs) subclass, which exihibit wide host-ranges and complex and elongated life cycles that may indicate prolonged saprotrophic or biotrophic phases prior to necrotrophy. This group also includes species commonly classified as wilts, anthracnoses, and rots. Mesotrophs corresponding best to hemibiotrophs, contains facultative biotrophic species that have longer latent periods than necrotrophs and do not use toxins as a primary virulence determinant (Hane et al. 2020).

The lack of knowledge of the trophic behavior and lifestyle of *C. henricotiae* e *C. pseudonaviculata* was considered by (Rogers et al. 2022) a challenge to interpret how the mechanism of nutrient acquisition affect evolution of gene families involved in pathogenesis. Thus, although this study does not include investigations about evolution of gene families, the predicted lifestyle for *Calonectria* species appears provide a better interpretation of our results related to investigations of pathogenicity and virulence genes. Therefore, *Calonectria* species classified as vasculartrophic, represented by species saprophytic -“apathogenic”- or broad host-range, have a higher content of putative genes involved in pathogenicity and virulence than those classified as mesotrophic, represented by species that exhibit limited host-range, *i.e. C. henricotiae* e *C. pseudonaviculata*.

Among vasculartrophic species, only *C. naviculata* and *C. multiphiliadica* have been treated as saprobic (apathogenic) species (Rogers et al. 2022). However, the similar lifestyle based in CAZymes content suggests similar capabilities for cell degradation and nutrient processing. This consideration can be supported by the recent report of *C. hongkonghensis* isolated from diseased *E. gunnii* roots, which previously had only been isolated from *Eucalyptus* plantation soil (Zhang et al. 2022). *Calonectria hongkonghensis* was classified as vasculartrophic in this study.

Several secondary metabolites produced by plant pathogens are often phytotoxic and can be essential for pathogenicity and contribute to virulence (Howlett 2014; Xu et al. 2021). A phytotoxin PF1070A was demonstrated has a role in the virulence of *C. ilicicola* (Ochi et al. 2011). An additional report of secondary metabolites revealed a high production capacity of secondary metabolite genes in *C. pseudoreteaudii*, including NRPS and PKS genes up-regulated in mycelia grown *Eucalyptus* tissue medium culture (Ye et al. 2018). This result supports our result for SMB clusters analysis, which showing a high number of predicted SMB clusters among *Calonectria* species.

Comparative genomic analysis revealed high similarity in the proteome (51.34%) and secretome (25%) of 16 *Calonectria* species. Investigations of unique clusters revealed that *C. ilicicola* has a one unique cluster related to the secondary metabolite biosynthetic process with homologs in the DFVF identified as a reducing polyketide synthase (PKS) involved in mycotoxin biosynthesis. Similarly, *C. leucothoes* showed one unique cluster identified as secreted and related to the secondary metabolite biosynthetic process with hits in the DFVF described as a nonribosomal peptide synthetase (NPKS) involved in siderophore and octapeptide fusaoctaxin A biosynthesis. These PKS and NPKS are enzymes involved in the secondary metabolite production that has a role important in the pathogenicity or virulence found in *Cochliobolus heterostrophus* (Baker et al. 2006), *Alternaria alternata* (Harimoto et al.

2007) and *Fusarium graminearum* (Sieber et al. 2014). Additionally, one unique cluster of *C. multiphialidica* was related to oxidoreductase activity with homologs in DFVF involved in the mycotoxin biosynthesis.

Consistent with its larger individual repertoire of genes homologs in DFVF, C. *naviculata* has the highest number of unique clusters (n=4), which were associated to oxidoreductase activity, kinase activity, transmembrane transport, and acyltransferase activity with homologs in DVDF defined as alcohol oxidase, transcription factor SKN7, MDR efflux pump ABC3, acetyl-CoA acetyltransferase erg10B, respectively. Transcription factor SKN7 regulates heat shock genes in response to oxidative stress and genes involved in cell wall integrity in response to osmotic changes (Basso et al. 2017). Furthermore, transmembrane proteins such as MDR efflux pump ABC can expulse a wide range of quite different substrates, important for the defense of fungal cells to toxic compounds. As *C. naviculata* has only been isolated from soil, such osmoregulation and detoxification mechanisms seem to be important for stress tolerance and permanence in this environment.

Finally, investigations of clusters exclusive shared by *C. pteridis* and other *Calonectria* species identified two exclusive clusters associated with pathogenicity and virulence shared by *C.pteridis* and *C. naviculata*, which were related to oxidoreductase activity and to the secondary metabolite biosynthetic process. In addition, three exclusive clusters with homologs in the DFVF, individually related to transferase activity, chitin synthase activity and nucleotide binding were shared by *C. pteridis* and *C. crousiana*, *C. pseudonaviculata* and *C. hawksworthii*, respectively. In this case, sequences showing homologs in DFVF described as non-reducing polyketide synthase CTB1, chitin synthase and ABC transporter FGM5/MRP-like transporter 1, respectively. Frequently oxidoreductases are required for the biosynthesis of secondary metabolites (Rose et al. 2002). Chitin synthases genes were related to a potential role in pathogenicity of *C. ilicicola* during infection of peanut (Chen et al. 2022). Furthermore, it has

been documented that rather than conferring an advantage to the microorganism during plant colonization, secondary metabolite toxins may allow a producer organism to escape predators during the saprophytic phases of its life in its ecological niche (Howlett 2014).

CONCLUSIONS

This study presents the characterization of the nuclear and mitochondrial genome of *Calonectria pteridis* LPF059, the main causal agent of Calonectria leaf blight in commercial eucalyptus plantations in Brazil. The nuclear genome shows a high completeness of gene repertoire and with the most of coding sequence related to secondary metabolites biosynthesis and carbohydrate metabolism. Mitochondrial genome size and gene content is comparable to the mitochondrial genome of other *Calonectria* species *C. pteridis* is heterotallic and it has a high potential to produce CAZymes and secondary metabolites, as occurs in other species of *Calonectria*, mainly those related to cell wall degradation and toxin production. Despite this, *Calonectria* species differ in their trophic behavior. Thus, *C. henricotiae* and *C. pseudonaviculata* are classified as mesotrophs and the other species in this study as vasculartrophs due to their differences in CAZymes content and therefore nutrient acquisition. Furthermore, species that are so far saprophytic -“apathogenic”- or broad host-range have a higher content of pathogenicity/virulence-related genes, including CAZymes and secondary metabolites, than species that are narrow host-range. Potentially, these genes are not only related to pathogenicity/virulence, but also to adaptation and stress tolerance in different environments, which indicates that the loss of certain genes may contribute to the reduction of the host range of *Calonectria* species.

ACKNOWLEDGMENTS

This work was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The authors also thank the Núcleo de Análise de Biomoléculas

(NuBioMol) for suportting in data analysis. NuBioMol is financially supported by the following Brazilian agencies: Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), CNPq, Financiadora de Estudos e Projetos (Finep) and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).

REFERENCES

Alfenas RF, Lombard L, Pereira OL, et al (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

Alfenas RF, Pereira OL, Freitas RG, et al (2013) Mass spore production and inoculation of Calonectria pteridis on Eucalyptus spp. under different environmental conditions. Trop Plant Pathol 38:406–413. https://doi.org/10.1590/S1982-56762013000500005

Almagro AJJ, Sønderby CK, Sønderby SK, et al (2017) DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387–3395. https://doi.org/10.1093/BIOINFORMATICS/BTX431

Aylward J, Steenkamp ET, Dreyer LL, et al (2017) A plant pathology perspective of fungal genome sequencing. IMA Fungus 8:1–15. https://doi.org/10.5598/IMAFUNGUS.2017.08.01.01/FIGURES/5

Baker SE, Kroken S, Inderbitzin P, et al (2006) Two polyketide synthase-encoding genes are required for biosynthesis of the polyketide virulence factor, T-toxin, by Cochliobolus heterostrophus. Mol Plant Microbe Interact 19:139–149. https://doi.org/10.1094/MPMI-19-0139

Bankevich A, Nurk S, Antipov D, et al (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

Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:1–6. https://doi.org/10.1186/S13100-015-0041-9/TABLES/2

Baroncelli R, Amby DB, Zapparata A, et al (2016) Gene family expansions and contractions are associated with host range in plant pathogens of the genus Colletotrichum. BMC Genomics 17:1–17. https://doi.org/10.1186/S12864-016-2917-6/FIGURES/2

Basso V, Znaidi S, Lagage V, et al (2017) The two-component response regulator Skn7 belongs to a network of transcription factors regulating morphogenesis in Candida albicans and independently limits morphogenesis-induced ROS accumulation. Mol Microbiol

106:157–182. https://doi.org/10.1111/MMI.13758

Bernt M, Donath A, Jühling F, et al (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313–319. https://doi.org/10.1016/J.YMPEV.2012.08.023

Blin K, Shaw S, Steinke K, et al (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/NAR/GKZ310

Bose R, Banerjee S, Pandey A, et al (2023) Calonectria leaf blight of Eucalyptus: A global review. Ann Appl Biol 182:6–28. https://doi.org/10.1111/AAB.12800

Chan PP, Lin BY, Mak AJ, Lowe TM (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res 49:9077–9096. https://doi.org/10.1093/NAR/GKAB688

Chen Q, Guo W, Feng L, et al (2015) Transcriptome and proteome analysis of Eucalyptus infected with Calonectria pseudoreteaudii. J Proteomics 115:117–131. https://doi.org/10.1016/J.JPROT.2014.12.008

Chen X, Luo M, Wu W, et al (2022) Virulence-Associated Genes of Calonectria ilicola, Responsible for Cylindrocladium Black Rot. J fungi (Basel, Switzerland) 8:. https://doi.org/10.3390/JOF8080869

Cortázar AR, Aransay AM, Alfaro M, et al (2014) SECRETOOL: integrated secretome analysis tool for fungi. Amino Acids 46:471–473. https://doi.org/10.1007/S00726-013-1649-Z

Crous PW (2002) Taxonomy and pathology of Cylindrocladium (Calonectria) and allied genera. The American Phytopathological Society, Saint Paul, Minnesota

Darriba Di, Posada D, Kozlov AM, et al (2020) ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol 37:291–294. https://doi.org/10.1093/MOLBEV/MSZ189

Emms DM, Kelly S (2019) OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol 20:1–14. https://doi.org/10.1186/S13059-019-1832-Y/FIGURES/5

Ferreira F., Alfenas AC, Moreira A., Demuner N. (1995) Mancha-de-pteridis: doença foliar de eucalipto em áreas tropicais brasileiras. Fitopatol Bras 20:107–110

Fonseca PLC, De-Paula RB, Araújo DS, et al (2021) Global Characterization of Fungal Mitogenomes: New Insights on Genomic Diversity and Dynamism of Coding Genes and Accessory Elements. Front Microbiol 12:. https://doi.org/10.3389/fmicb.2021.787283

Gai Y, Deng Q, Chen X, et al (2017) Phylogenetic diversity of Calonectria ilicicola causing Cylindrocladium black rot of peanut and red crown rot of soybean in southern China. J Gen Plant Pathol 83:273–282. https://doi.org/10.1007/s10327-017-0724-9

Gai Y, Pan R, Peng X (2020) A phylogenomic tree of fungi: evolutionary relationships among Calonectria ilicicola and 586 fungal mitochondrial genomes. http://www.tandfonline.com/action/authorSubmission?journalCode=tmdn20&page=instructions 5:1709–1711. https://doi.org/10.1080/23802359.2020.1749163

Gehesquière B, Crouch JA, Marra RE, et al (2016) Characterization and taxonomic reassessment of the box blight pathogen Calonectria pseudonaviculata, introducing Calonectria henricotiae sp. nov. Plant Pathol 65:37–52. https://doi.org/10.1111/ppa.12401

Hane JK, Paxman J, Jones DAB, et al (2020) "CATAStrophy," a Genome-Informed Trophic Classification of Filamentous Plant Pathogens – How Many Different Types of Filamentous Plant Pathogens Are There? Front Microbiol 10:. https://doi.org/10.3389/FMICB.2019.03088/FULL

Harimoto Y, Hatta R, Kodama M, et al (2007) Expression profiles of genes encoded by the supernumerary chromosome controlling AM-toxin biosynthesis and pathogenicity in the apple pathotype of Alternaria alternata. Mol Plant Microbe Interact 20:1463–1476. https://doi.org/10.1094/MPMI-20-12-1463

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-Genome Annotation with BRAKER. Methods Mol Biol 1962:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5

Howlett BJ (2014) Secondary Metabolites Produced by Plant Pathogens. Nat Prod Discourse, Divers Des 159–169. https://doi.org/10.1002/9781118794623.CH9

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/MOLBEV/MST010

Kozlov AM, Darriba D, Flouri T, et al (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455. https://doi.org/10.1093/BIOINFORMATICS/BTZ305

Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem 77:521–555. https://doi.org/10.1146/ANNUREV.BIOCHEM.76.061005.092322

LeBlanc N, Gehesquière B, Salgado-Salazar C, et al (2019) Limited genetic diversity across pathogen populations responsible for the global emergence of boxwood blight identified using SSRs. Plant Pathol 68:861–868. https://doi.org/10.1111/ppa.13003

LeBlanc N, Salgado-Salazar C, Crouch JA (2018) Boxwood blight: an ongoing threat to ornamental and native boxwood. Appl Microbiol Biotechnol 102:4371–4380. https://doi.org/10.1007/s00253-018-8936-2

Lelwala R V., Korhonen PK, Young ND, et al (2019) Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in Colletotrichum tanaceti. PLoS One 14:e0212248. https://doi.org/10.1371/JOURNAL.PONE.0212248

Li JQ, Barnes I, Liu FF, et al (2021) Global genetic diversity and mating type distribution of

calonectria pauciramosa: An important wide-host-range plant pathogen. Plant Dis 105:. https://doi.org/10.1094/PDIS-05-20-1050-RE

Li JQ, Wingfield BD, Wingfield MJ, et al (2020) Mating genes in calonectria and evidence for a heterothallic ancestral state. Persoonia Mol Phylogeny Evol Fungi 45:163–176. https://doi.org/10.3767/persoonia.2020.45.06

Liu HH, Wang J, Wu PH, et al (2021) Whole-Genome Sequence Resource of *Calonectria ilicicola*, the Casual Pathogen of Soybean Red Crown Rot. Mol Plant Microbe Interact 34:848–851. https://doi.org/10.1094/MPMI-11-20-0315-A

Liu QL, Li JQ, Wingfield MJ, et al (2020) Reconsideration of species boundaries and proposed DNA barcodes for Calonectria. Stud Mycol 97:. https://doi.org/10.1016/j.simyco.2020.08.001

Lombard L, Crous PW, Wingfield BD, Wingfield MJ (2010a) Species concepts in Calonectria (Cylindrocladium). Stud Mycol 66:1–13. https://doi.org/10.3114/sim.2010.66.01

Lombard L, Crous PW, Wingfield BD, Wingfield MJ (2010b) Multigene phylogeny and mating tests reveal three cryptic species related to Calonectria pauciramosa. Stud Mycol 66:15–30. https://doi.org/10.3114/sim.2010.66.02

Lombard L, Zhou XD, Crous PW, et al (2010c) Calonectria species associated with cutting rot of Eucalyptus. Persoonia Mol Phylogeny Evol Fungi 24:1–11. https://doi.org/10.3767/003158510X486568

Malapi-Wight M, Salgado-Salazar C, Demers JE, et al (2016) Sarcococca blight: Use of whole-genome sequencing for fungal plant disease diagnosis. Plant Dis 100:1093–1100. https://doi.org/10.1094/PDIS-10-15-1159-RE

Malapi-Wight M, Veltri D, Gehesquière B, et al (2019) Global distribution of mating types shows limited opportunities for mating across populations of fungi causing boxwood blight disease. Fungal Genet Biol 131:. https://doi.org/10.1016/j.fgb.2019.103246

Marchler-Bauer A, Lu S, Anderson JB, et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:. https://doi.org/10.1093/NAR/GKQ1189

Ochi S, Yoshida M, Nakagawa A, Natsume M (2011) Identification and activity of a phytotoxin produced by Calonectria ilicicola, the causal agent of soybean red crown rot. Can J Plant Pathol 33:347–354. https://doi.org/10.1080/07060661.2011.593558

Pham NQ, Barnes I, Chen SF, et al (2019) Ten new species of Calonectria from Indonesia and Vietnam. Mycologia 111:1–25. https://doi.org/10.1080/00275514.2018.1522179

Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5:. https://doi.org/10.1371/JOURNAL.PONE.0009490

Queiroz MF, Santos SA, Vidigal PMP, et al (2022) Draft Genome Sequence of Calonectria pteridis, the Causal Agent of Calonectria Leaf Blight on Eucalyptus. Microbiol Resour

Announc 11:. https://doi.org/10.1128/MRA.00284-22

Rafiei V, Vélëz H, Tzelepis G (2021) The Role of Glycoside Hydrolases in Phytopathogenic Fungi and Oomycetes Virulence. Int J Mol Sci 22:. https://doi.org/10.3390/IJMS22179359

Rogers LW, Koehler AM, Crouch JA, et al (2022) Comparative genomic analysis reveals contraction of gene families with putative roles in pathogenesis in the fungal boxwood pathogens Calonectria henricotiae and C. pseudonaviculata. BMC Ecol Evol 22:1–16. https://doi.org/10.1186/S12862-022-02035-4/TABLES/2

Rose MS, Yun SH, Asvarak T, et al (2002) A decarboxylase encoded at the Cochliobolus heterostrophus translocation-associated Tox1B locus is required for polyketide (T-toxin) biosynthesis and high virulence on T-cytoplasm maize. Mol Plant Microbe Interact 15:883–893. https://doi.org/10.1094/MPMI.2002.15.9.883

Salcedo-Sarmiento S, Aucique-Pérez CE, Silveira PR, et al (2021) Elucidating the interactions between the rust Hemileia vastatrix and a Calonectria mycoparasite and the coffee plant. iScience 24:102352. https://doi.org/10.1016/J.ISCI.2021.102352

Salgado-Salazar C, Romberg MK, Blomquist C, et al (2022) Lifestyle, mating type and mitochondrial genome features of the plant pathogen Calonectria hawksworthii (Hypocreales, Nectriaceae) as revealed by genome analyses. https://doi.org/101080/0706066120222065534 44:723–736.

Sanchez-Gonzalez EI, De T, Farias Soares P, et al (2022) Two new species of Calonectria (Hypocreales, Nectriaceae) causing Eucalyptus leaf blight in Brazil. MycoKeys 91 169-197 91:169–197. https://doi.org/10.3897/MYCOKEYS.91.84896

Sandor S, Zhang Y, Xu J (2018) Fungal mitochondrial genomes and genetic polymorphisms. Appl Microbiol Biotechnol 102:9433–9448. https://doi.org/10.1007/S00253-018-9350-5

Sieber CMK, Lee W, Wong P, et al (2014) The Fusarium graminearum Genome Reveals More Secondary Metabolite Gene Clusters and Hints of Horizontal Gene Transfer. PLoS One 9:e110311. https://doi.org/10.1371/JOURNAL.PONE.0110311

Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/BIOINFORMATICS/BTV351

Soares ID, Auer CG, Dos Santos AF, et al (2019) First report of Calonectria Leaf Blight caused by Calonectria metrosideri on Eucalyptus benthamii in Brazil. Plant Dis. 103

Specht CA, DiRusso CC, Novotny CP, Ullrich RC (1982) A method for extracting high-molecular-weight deoxyribonucleic acid from fungi. Anal Biochem 119:158–163. https://doi.org/10.1016/0003-2697(82)90680-7

Storer J, Hubley R, Rosen J, et al (2021) The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA 12:1–14. https://doi.org/10.1186/S13100-020-00230-Y/FIGURES/8

Urban M, Cuzick A, Rutherford K, et al (2017) PHI-base: a new interface and further additions for the multi-species pathogen-host interactions database. Nucleic Acids Res 45:D604–D610. https://doi.org/10.1093/NAR/GKW1089

Xu D, Xue M, Shen Z, et al (2021) Phytotoxic Secondary Metabolites from Fungi. Toxins (Basel) 13:. https://doi.org/10.3390/TOXINS13040261

Yang X, McMahon MB, Ramachandran SR, et al (2021) Comparative analysis of extracellular proteomes reveals putative effectors of the boxwood blight pathogens, Calonectria henricotiae and C. pseudonaviculata. Biosci Rep 41:. https://doi.org/10.1042/BSR20203544

Ye X, Liu H, Jin Y, et al (2017) Transcriptomic analysis of Calonectria pseudoreteaudii during various stages of Eucalyptus infection. PLoS One 12:. https://doi.org/10.1371/journal.pone.0169598

Ye X, Zhong Z, Liu H, et al (2018) Whole genome and transcriptome analysis reveal adaptive strategies and pathogenesis of Calonectria pseudoreteaudii to Eucalyptus. BMC Genomics 19:. https://doi.org/10.1186/s12864-018-4739-1

Zhang H, Yohe T, Huang L, et al (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/NAR/GKY418

Zhang Y, Chen C, Chen C, et al (2022) Identification and Characterization of Calonectria Species Associated with Plant Diseases in Southern China. J Fungi 8:719. https://doi.org/10.3390/JOF8070719/S1

Zhao Z, Liu H, Wang C, Xu JR (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics 14:. https://doi.org/10.1186/1471-2164-14-274

Figures and Tables

Table 1. Details and Genbank accessions of genomic data included in the analyses in this study.

Species name	Strain	Substrate	Geographic origin	Genome size (Mb)	Genbank acession	Reference
Calonectria aciculata	CMW 47645	Leaves (*E.urophylla x E. grandis*)	China	61.59	GCA_013406995.1	Liu et al. (2019)
C. crousiana	CMW 27249	Leaves (*E. grandis*)	China	58.07	GCA_013406985.1	Liu et al. (2019)
C. fujianensis	CMW 27257	Leaves (*E. grandis*)	China	61.45	GCA_013406965.1	Liu et al. (2019)
C. hawksworthii	S6964	Persea americana grafting seedling	Californnia, USA	64.77	GCA_020975415.1	Salgado-Salazar et al. (2022)
C. henricotiae	JAC13-131	*Buxus* sp.	Germany	53.03	GCA_020623695.1	LeBlanc et al. (2021)
C. henricotiae	CB077	*Buxus* sp.	Belgium	47.54	GCA_004380935.1	LeBlanc et al. (2021)
C. honghensis	CMW 47669	Soil (*Eucalyptus* plantation)	China	61.74	GCA_013403855.1	Liu et al. (2019)
C. hongkongensis	CMW 47271	Soil (*Eucalyptus* plantation)	China	61.88	GCA_017140755.1	Li et al. (2020)
C. ilicicola	JXLN31	Stem (*Arachis hypogaea*)	China	69.86	GCA_014621995.1	Gai et al.(2017)
C. ilicicola	ZJHZ01	Stem (*Glycine max*)	China	70.36	GCA_014622005.1	Gai et al. (2017)
C. leucothoes	CBS 109166	leaf (*Leucothoe axillaris*)	Florida, USA	63.06	GCA_002179835.1	Malapi-Wight et al. (2019)
C. multiphialidica	CBS 112678	Soil	Cameroon	62.9	GCA_020623665.1	LeBlanc et al. (2021)
C. naviculata	CBS 101121	leaf litter	Brazil	65.72	GCA_003031705.1	Malapi-Wight et al. (2019)
C. pauciramosa	CBS 138824	*E. grandis*	South Africa	62.55	GCA_017140785.1	Li et al. (2020)
C. pauciramosa	CMW 7592	*E. grandis*	Uruguay	62.32	GCA_017140655.1	Li et al. (2020)
C. pseudonaviculata	CBS 139394	*Sarcococca hookeriana* var. *humilis*	Maryland, USA	51.44	GCA_001696505.1	Malapi-Wight et al. (2016)
C. pseudonaviculata	CBS 139395	*Buxus sempervirens*	Connecticut, USA	54.98	GCA_004380915.1	Malapi-Wight et al. (2016)
C. pseudoreteaudii	YA51	*Eucalyptus* sp.	China	63.69	GCA_001879505.1	Ye et al. (2018)
C. pseudoturangicola[a]	CMW 47496	Soil (*Eucalyptus* plantation)	China	62.1	GCA_013403825.1	Liu et al. (2019)
C. pteridis	LPF059	Leaves (*Eucalyptus* sp.)	Brazil	58.4	GCA_022837005.1	Queiroz et al. (2021)
Fusarium graminearum	PH-14	NA	NA	36.45	GCF_000240135.3	Cuomo et al. (2007)
Neonectria ditissima	R09/05	*Malus domestica*	United Kingdom	45.71	GCA_001306435.1	Gomez-Cortecero et al. (2015)
Trichoderma reesei	QM6	NA	NA	33.4	GCF_000167675.1	Martinez et al. (2008)
Verticillium dahliae	VdLs.17	NA	NA	33.8	GCF_000150675.1	Broad Institute of MIT and Harvard
Neurospora crassa	OR74A	NA	NA	41	GCF_000182925.2	Broad Institute of MIT and Harvard
Pyricularia oryzae	MG8	Rice	NA	41	GCF_000002495.2	Dean et al. (2005)
Botrytis cinerea	B05.10	NA	NA	42.6	GCF_000143535.2	Wageningen University, Biosystematics Group
Sclerotinia sclerotiorum	1980 UF-70	NA	NA	38.45	GCF_000146945.2	Broad Institute of MIT and Harvard

[a] *C. pseudoturangicola* was treated as a synonym of *C. kyotensis* by Liu et al (2020).

NA: Information not available.

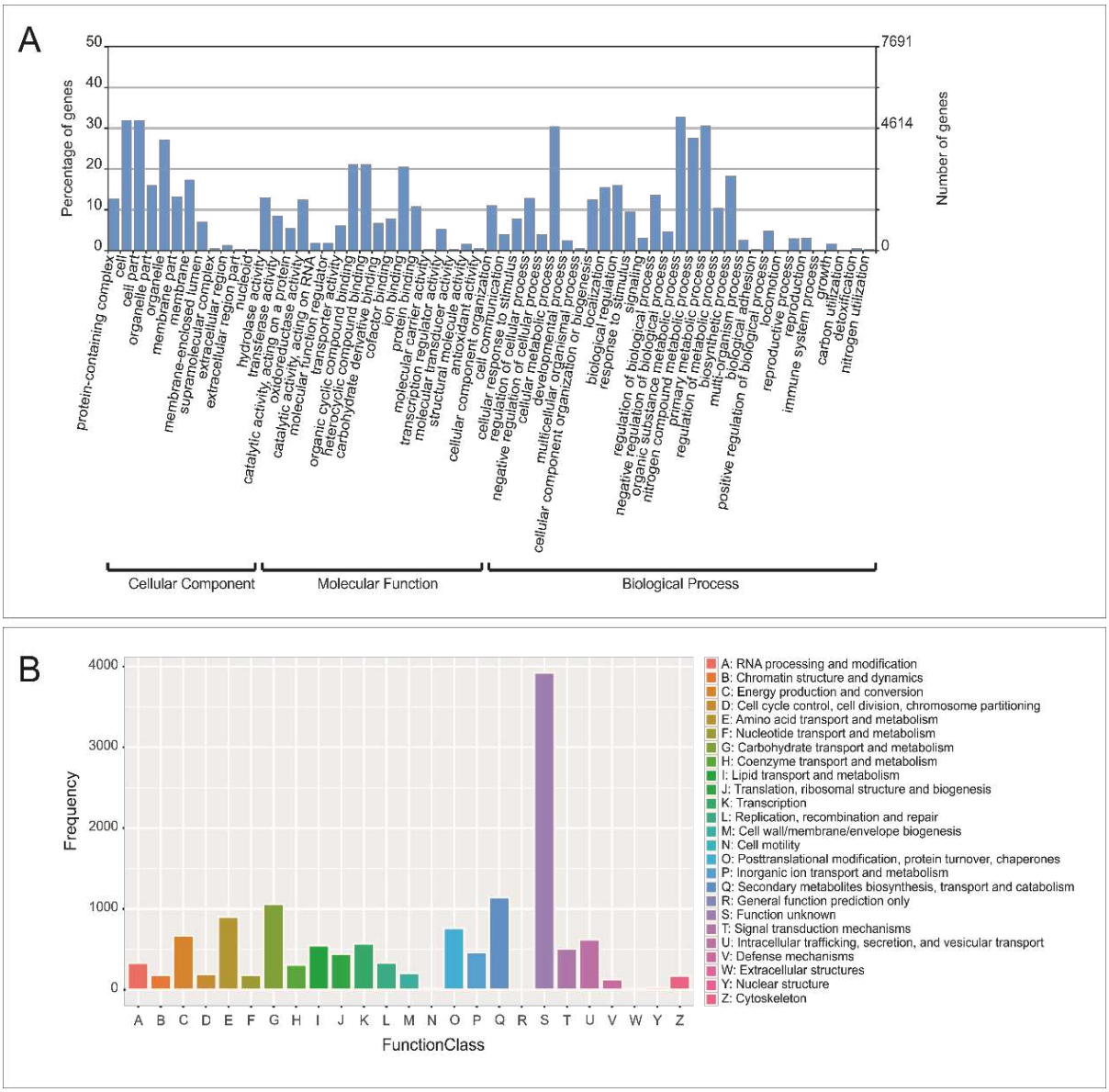

Fig. 1 Functional annotation of the predicted proteins of *Calonectria pteridis* LPF059 genome based on gene ontology (GO), showing more representative terms involved in cellular component, molecular function, and biological process (A), and frequency of eggNOG functional categories (COG) (B).

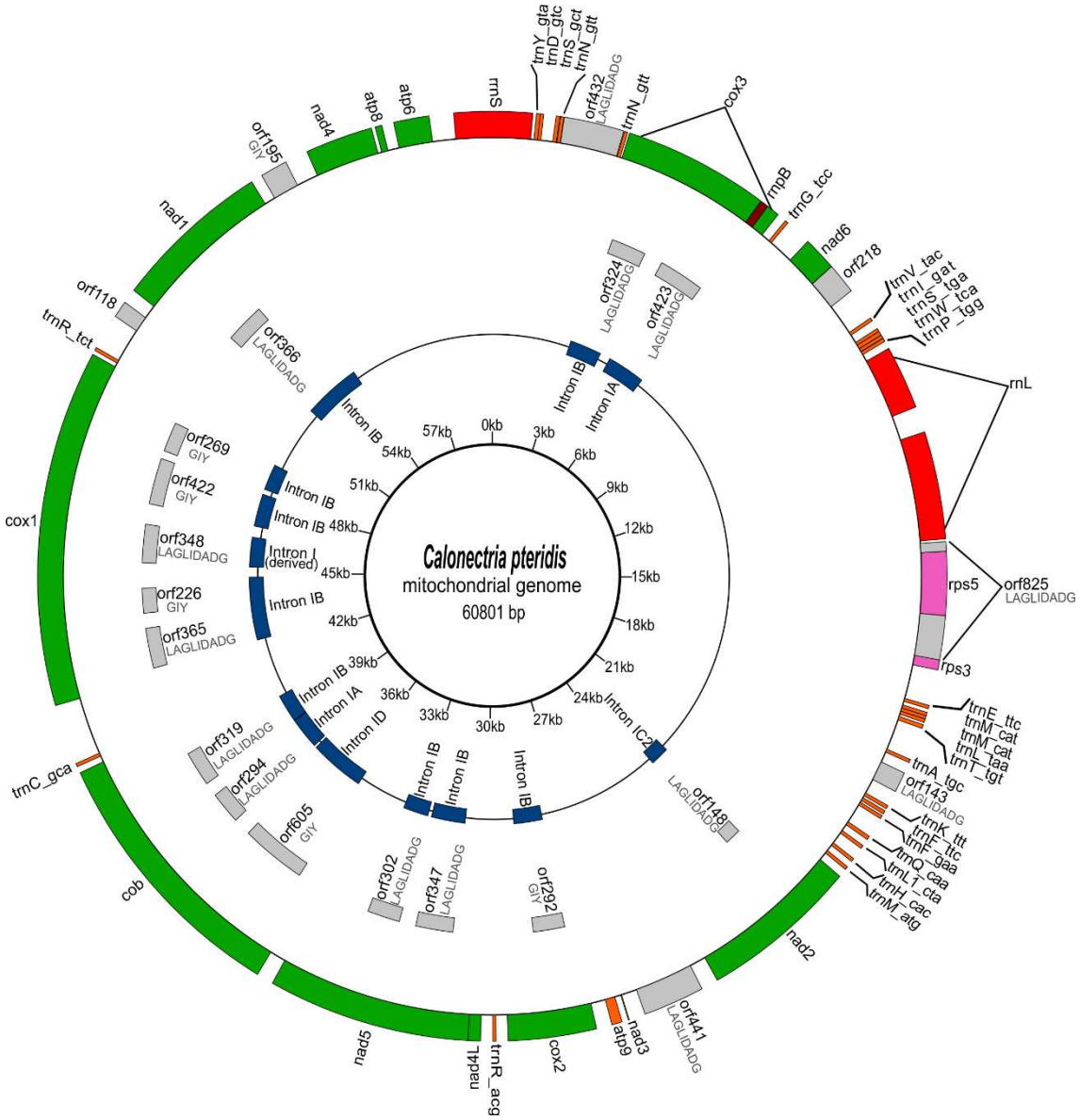

Fig. 2 Circular map of the mitochondrial genome of *Calonectria pteridis* LPF059. The circular map was generated using GenomeVx tool. Genes are presented in a clockwise direction (forward). Green: core protein-coding genes; orange: genes of transfer RNAs; red: genes of large and small ribosomal subunits; pink: genes encoding ribosomal proteins S3 and S5; brown: gene of RNase P (*rnpB*); dark blue: introns; gray: open reading frame (ORFs) encoding homing endonucleases and hypothetical proteins.

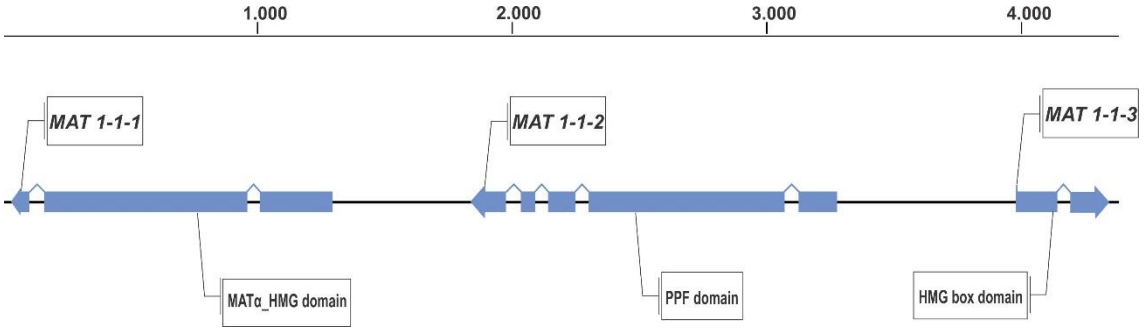

Fig. 3 Schematic organization of *MAT1-1* locus of *Calonectria pteridis* LPF059 genome.

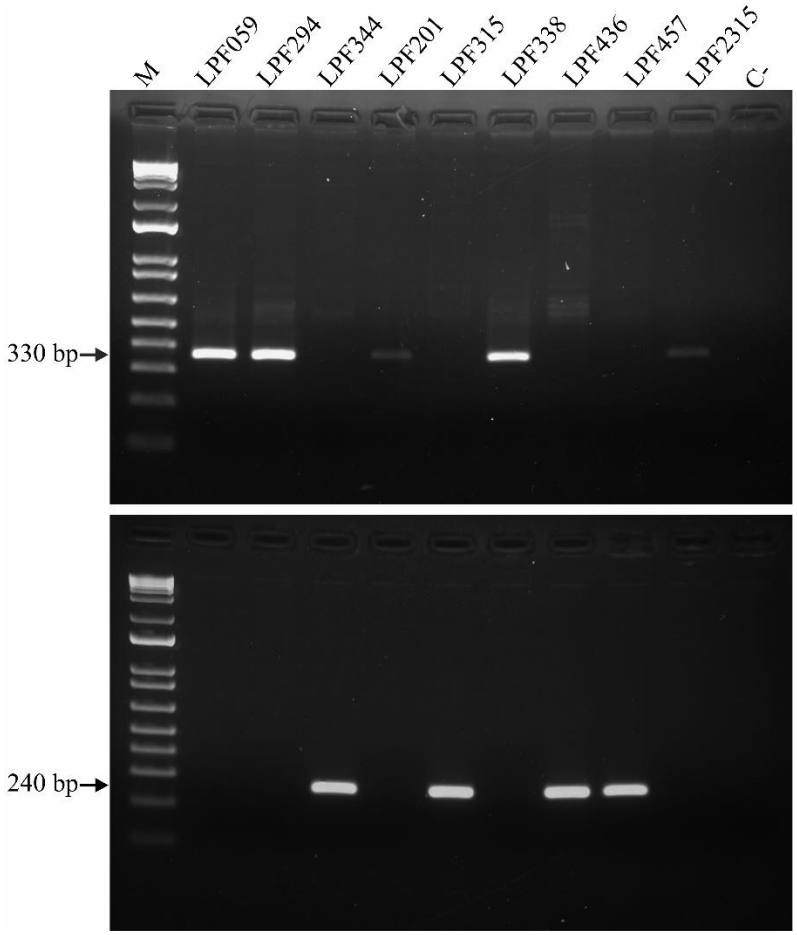

Fig. 4 PCR amplification of mating type sequences from nine *Calonectria pteridis* isolates by primers MAT1-1-1 and MAT1-2-1 (Li et al., 2020). Lane M, 1 kb DNA ladder; lane C-, negative control (distilled water only). Fragment size is indicated on the left side, showing amplicon of *MAT1-1* as 330 bp, while amplicon of *MAT2-1* is 240 bp.

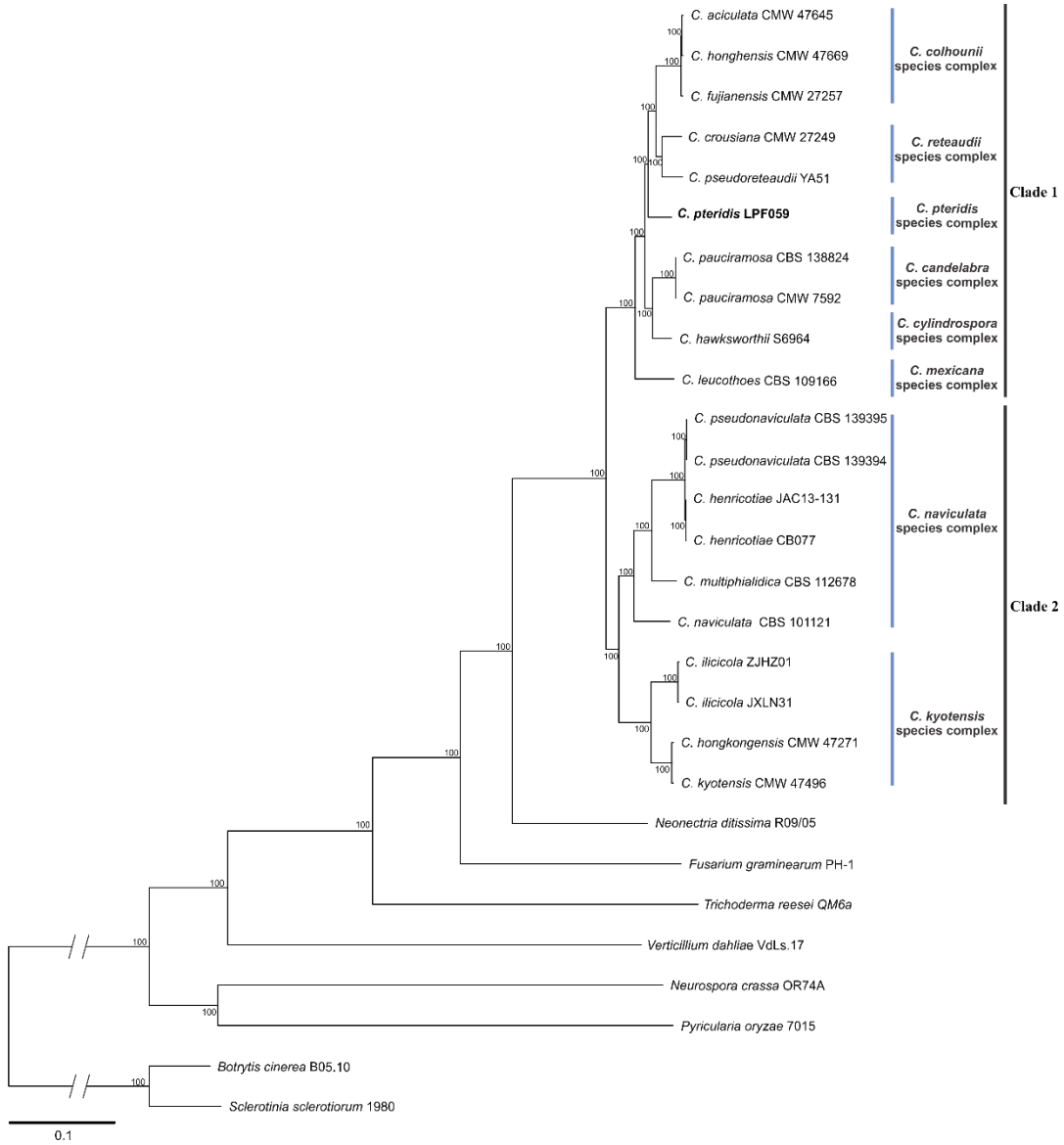

Fig 5. The best maximum likelihood (ML) tree constructed based on 2,308 single-copy orthologous genes shared among 16 *Calonectria* species and 8 other ascomycetes fungi. Node numbers show bootstrap support values (1000 replicates).

Table 2. Number of predicted genes and secreted proteins of 16 *Calonectria* species.

Species name	Genbank acession	Genome size (Mb)	Proteome	Secretome
Calonectria naviculata	GCA_003031705.1	65.72	23,467	2,325
C. ilicicola	GCA_014621995.1	69.86	20,565	2,020
C. ilicicola	GCA_014622005.1	70.36	20,533	2,006
C. pseudoturangicola	GCA_013403825.1	62.1	19,700	2,107
C. multiphialidica	GCA_020623665.1	62.9	19,582	2,008
C. hongkongensis	GCA_017140755.1	61.88	19,581	2,080
C. hawksworthii	GCA_020975415.1	64.77	19,349	2,091
C. leucothoes	GCA_002179835.1	63.06	18,705	2,081
C. pauciramosa	GCA_017140785.1	62.55	18,565	2,038
	GCA_017140655.1	62.32	18,452	2,032
C. aciculata	GCA_013406995.1	61.59	18,373	1,993
C. honghensis	GCA_013403855.1	61.74	18,132	1,946
C. fujianensis	GCA_013406965.1	61.45	18,07	1,933
C. pseudoreteaudii	GCA_001879505.1	63.69	18,021	1,917
C. pteridis	GCA_022837005.1	58.4	17,388	1,775
C. crousiana	GCA_013406985.1	58.07	17,033	1,826
C. henricotiae	GCA_004380935.1	47.54	14,253	1,389
	GCA_020623695.1	53.03	13,633	1,323
C. pseudonaviculata	GCA_001696505.1	51.44	13,665	1,322
	GCA_004380915.1	54.98	13,659	1,326

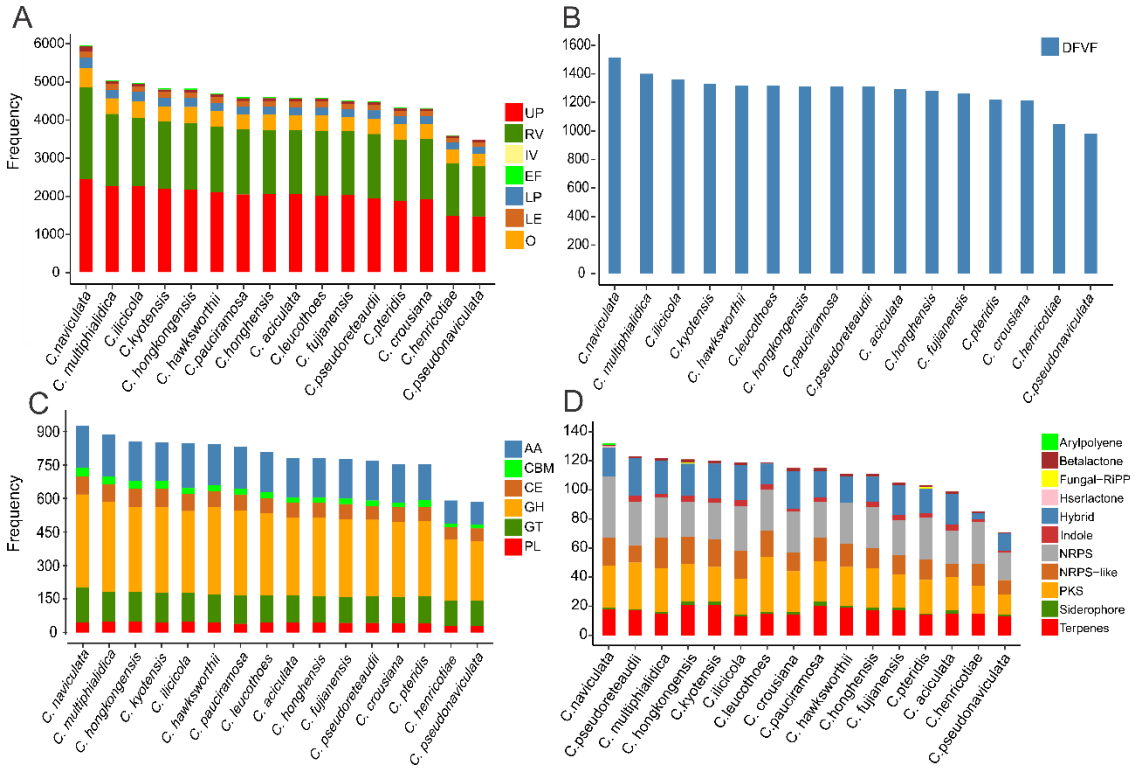

Fig 6. Composition of different pathogenicity-related genes predicted for *Calonectria pteridis* and other *Calonectria* species. The number of genes in each gene category (y axis) plotted for each species (x axis). (A) Homologs in the Pathogen-host interaction (PHI) database: UP, unaffected pathogenicity"; RV, reduced virulence; IV, increased virulence; EF, effector; LP, loss of pathogenicity; LE, lethal; (B) Homologs in Database of Fungal Virulence Factors (DVFV); (C) CAZyme classes: AA, Auxiliary activity; CBM, Carbohydrate-binding modules; CE, Carbohydrate esterase; GH, Glycoside hydrolase; GT, glycosyltransferase; PL, Polysaccharide lyase; (D) Secondary metabolite biosynthetic (SMB) gene clusters.

Table 3. Trophic CATAStrophy classifications for *Calonectria* species.

Species	Lifestyle class		
	Common class[a]	Novel major class[b]	Novel sub-class affinities[c]
Calonectria aciculata	necrotroph	vasculartroph	vasculartroph
C. crousiana	necrotroph	vasculartroph	vasculartroph
C. fujianensis	necrotroph	vasculartroph	vasculartroph
C. hawksworthii	necrotroph	vasculartroph	vasculartroph
C. henricotiae	necrotroph	mesotroph	vasculartroph
C. honghensis	necrotroph	vasculartroph	vasculartroph
C. hongkongensis	necrotroph	vasculartroph	vasculartroph
C. ilicicola	necrotroph	vasculartroph	vasculartroph
C. leucothoes	necrotroph	vasculartroph	vasculartroph
C. multiphialidica	necrotroph	vasculartroph	vasculartroph
C. naviculata	necrotroph	vasculartroph	vasculartroph
C. pauciramosa	necrotroph	vasculartroph	vasculartroph
C. pseudoreteaudii	necrotroph	vasculartroph	vasculartroph
C. kyotensis	necrotroph	vasculartroph	vasculartroph
C. pteridis	necrotroph	vasculartroph	vasculartroph
C. pseudonaviculata	necrotroph	mesotroph	vasculartroph

[a] Common trophic class literature derived.
[b] Major trophic class based in relative centroid distance (RCD) score of 1.
[c] Affinity for trophic sub-class based in RCD ≥0.95.

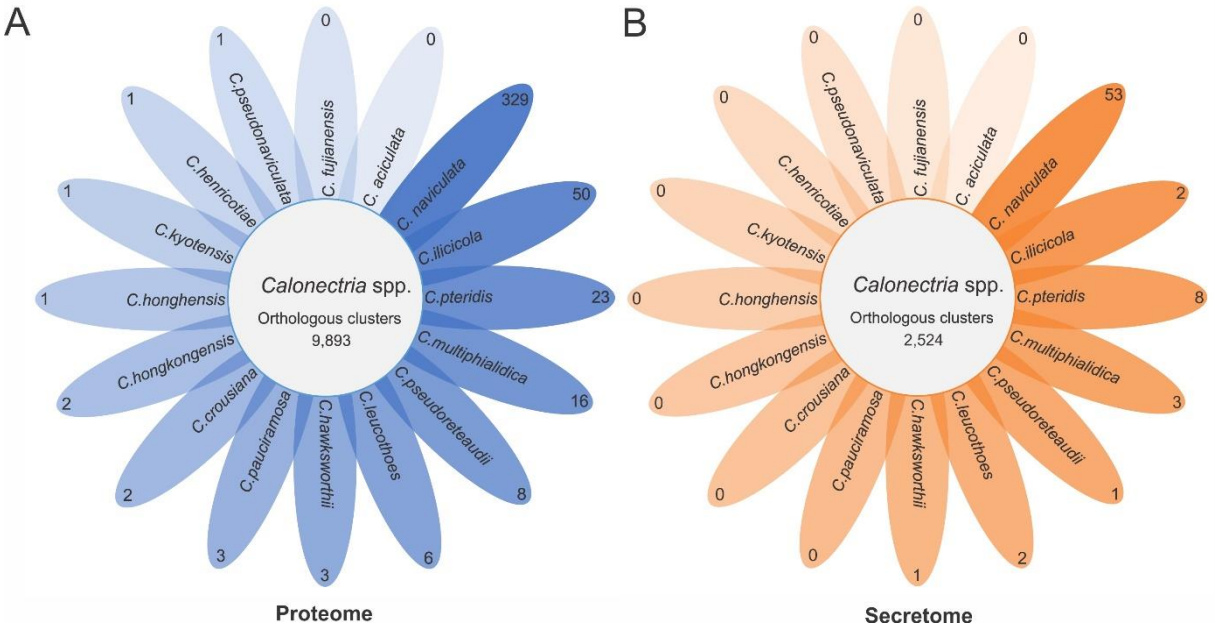

Fig 7. Orthologous clusters unique and shared among all predicted proteome and secreted proteins of 16 *Calonectria* species.

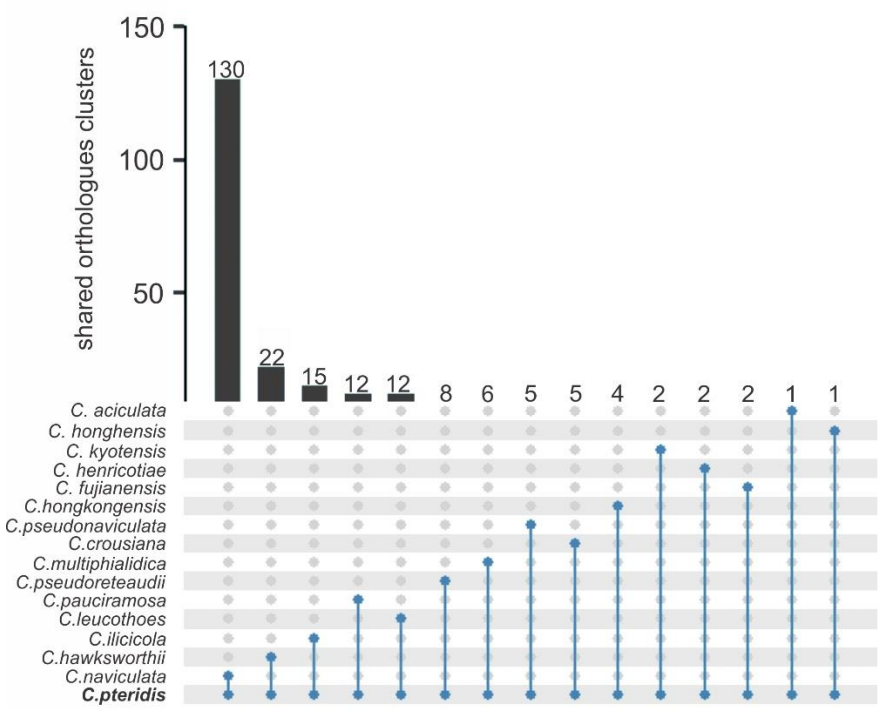

Fig 8. Venn analysis of exclusive orthologs clusters shared with *C. pteridis* LPF059.

Table 4. Gene Ontology (GO) of the unique orthologous clusters with pathogenicity-associated proteins according to the Database of Fungal Virulence Factors (DFVF) and Pathogen-Host Interaction (PHI) database.

Species name	Orthologues cluster	Gene IDs	GO terms[a]	DFVF uniprot ID	PHI-base phenotype
C. ilicicola	OG0014247	g4449, g4450	MF:GO0016746 (acyltransferase activity), MF: GO:0016491 (oxidoreductase activity), MF: GO0031177 (phosphopantetheine binding), BP: GO:0044550 (secondary metabolite biosynthetic process)	N4WHE3	reduced virulence
C. leucothoes	OG0017222	g7615*, g13084*	MF:GO0031177(phosphopantetheine binding), MF:OG:0008168 (methyltransferase activity), MF:OG0016874 (ligase activity), BP: GO:0044550 (secondary metabolite biosynthetic process), BP: OG:0017000 (antibiotic biosynthetic process)	C9K7B5 I1SAJ7	loss of pathogenicity; reduced virulence
C. multiphialidica	OG0015817	g10776, g13732, g17822	MF:GO:0016491(oxidoreductase activity), MF:GO:0016628 (oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor)	Q8NJQ2	reduced virulence
C. naviculata	OG0007974	g13856, g22334, g5358	CC:GO:0005737 (cytoplasm), CC:GO:0016020 (membrane), MF:GO:0000155 (phosphorelay sensor kinase activity), MF:GO:0016301 (kinase activity), MF:GO:0016772 (transferase activity, transferring phosphorus-containing groups), BP:GO:0000160 (phosphorelay signal transduction system), BP:GO:0006355 (regulation of DNA-templated transcription), BP:GO:0016310 (phosphorylation), BP:GO:0018106 (peptidyl-histidine phosphorylation)	Q5A4X5, Q9C1Q7, Q8X1E7	reduced virulence
	OG0016333	g7480*, g19344	MF:GO:0050660(flavin adenine dinucleotide binding), MF:GO:0016614 (oxidoreductase activity, acting on CH-OH group of donors)	Q9P304	reduced virulence
	OG0001941	g2025, g1506, g4583, g15530, g21068, g10651	CC:membrane (GO:0016020), CC:GO:0055052 (ATP-binding cassette (ABC) transporter complex, substrate-binding subunit-containing), MF:GO:0005524 (ATP binding); MF:GO:0000166 (nucleotide binding), MF:GO:0016787 (hydrolase activity), BP:GO:0055085 (transmembrane transport)	Q3Y5V5	loss of pathogenicity
	OG0019093	g11634, g22861	MF:GO:0016747 (acyltransferase activity, transferring groups other than amino-acyl groups)	Q4WCL5	lethal

*Secreted proteins; [a]CC: Component cellular, MF: Molecular function; BB: Biological process.

Table 5. Gene Ontology (GO) of the exclusive orthologous clusters shared with *C.pteridis* LPF059 and associated with pathogenicity according to the Database of Fungal Virulence Factors (DFVF) and Pathogen-Host Interaction (PHI) database.

Orthologues cluster	Gene IDs	GO terms	DFVF uniprot IDs	PHI phenotype
OG0017234	g7048(C.*pteridis*), g10232(*C.naviculata*)	CC:GO:0110165 (cellular anatomical entity), MF:GO:0016614 (oxidoreductase activity, acting on CH-OH group of donors), BP:GO:0006629 (lipid metabolic process), BP:GO:0009058 (biosynthetic process); BP:GO:0044237 (cellular metabolic process)	A4QTE3	reduced virulence; loss of pathogenicity
OG0016293	g3240(*Ca.pteridis*), g21234(*Ca.naviculata*)	MF:GO:0004315(3-oxoacyl-[acyl-carrier-protein] synthase activity), MF:GO:0008483(transaminase activity), MF:GO:0030170 (pyridoxal phosphate binding),MF:GO:0031177 (phosphopantetheine binding), MF:GO:0016746(acyltransferase activity), MF:GO:0008168 (methyltransferase activity), BP:GO:0044550(secondary metabolite biosynthetic process)	Q6ZX14	effector
OG0019562	g5601(*C.pteridis*), g1189(*C.crousiana*)	MF:GO:0016740 (transferase activity), BP:GO:0008152 (metabolic process)	Q6DQW3	reduced virulence
OG0017057	g16410(*C.pteridis*), g2817 (*C.pseudonaviculata*)	CC:GO:0016020 (membrane), CC:GO:0005886 (plasma membrane), MF:GO:0004100 (chitin synthase activity), MF:GO:0016757 (glycosyltransferase activity)	Q8TFN4, G4XVC0	reduced virulence; loss of pathogenicity
OG0020322	g8837(*C.pteridis*), g14588(*C.hawksworthii*)	CC:GO:0016020 (membrane), MF:GO:0000166 (nucleotide binding)	I1S2J9, Q5A762	reduced virulence

CHAPTER 3-Effect of rater instruction and standard area diagram structure on the accuracy of estimates of severity in Calonectria leaf blight

Marcia F. Queiroz[1]†, Kaique S. Alves[1]†, Camila F. Paixão[1], Rafael F. Alfenas[1], Emerson M. Del Ponte[1], Acelino C. Alfenas[1]*

[1]*Departamento de Fitopatologia, Universidade Federal de Viçosa, 36570-900, Viçosa, MG, Brazil*

*Correspondence author: Acelino C. Alfenas
Email: aalfenas@ufv.br

† These authors contributed equally to this study

This chapter has been submitted as an original article and still under revision at Tropical Plant Pathology.

ABSTRACT

The objectives of this study were to evaluate the effect of rater instruction and two standard area diagrams, that varied on scale structure, on the accuracy and reliability of visual estimates of calonectria leaf blight (CLB) severity. The minimum and the maximum severities of the newly developed SADs were 0.8% and 90%, respectively, with eight intermediate severity values for the linear SAD (10, 20, 30, 40, 50, 60, 70, 80%) and amended-linear SAD (3, 5, 10, 15, 30, 45, 60, 75%). For validation, one group (21 raters) used the amended-linear SAD and another group (22 raters) used the linear SAD. For each group, three sequential assessments were made: 1) without any aid; 2) following specific instructions; and 3) using the respective SAD. Two raters with the highest overall accuracy and two with significant gains in accuracy after training were chosen, from each SAD group, for further evaluating the accuracy of unaided estimates after SAD use. The linear and amended-linear SADs similarly improved the overall accuracy of the estimates (CCC from 0.93 to 0.96). The overestimation of severity was reduced at the lower end of the scale (< 25% severity) when using the amended-linear SAD. A trend of increasing reliability was observed as raters progressed through the rounds of assessments, but there was no statistical difference. The less accurate raters during unaided estimates from both SAD groups improved their accuracy after the instruction and maintained the same level of accuracy in the two rounds after SAD use. Instruction was critical for inherently less accurate raters. We recommend the amended-linear SAD for assessments of CLB severity, which minimizes the overestimation at the lower severity levels.

Keywords: Diagrammatic scale, Phytopathometry, *Calonectria pteridis*

INTRODUCTION

Calonectria leaf blight (CLB) caused by *Calonectria* spp. is one of the most devastating foliar diseases of eucalyptus plantations (Alfenas et al. 2015; Pham et al. 2019). In Brazil, *Calonectria pteridis* is the most common species associated with CLB on eucalyptus trees in warmer and humid regions (Alfenas et al. 2015). Symptoms of CLB include leaf and shoot blight, and can result in intense defoliation of trees that lead to significant negative effects on the growth of eucalyptus trees (Alfenas et al. 2013). In general, foliar symptoms are characterized by numerous small rounded or elongated spots randomly distributed in the leaf blade. In the early stages of the symptoms, the lesions are light gray with a peripheral narrow, dark gray humid halo, containing conidial sporulation often associated with young lesions on the abaxial leaf surface. As the disease progresses, the lesions become light or dark brown, often surrounded by a callus and sometimes delimited by a reddish-purple border. The lesions may coalesce and cover a large proportion of the leaf blade, leading to severe defoliation, which reduces the photosynthetic area and consequently the tree volumetric growth in highly susceptible genotypes (Ferreira et al. 1995; Alfenas et al. 2013).

The main control method for CLB is the use of resistant genotypes due to the existence of inter and intraspecific variability for resistance in eucalyptus (Alfenas et al. 2016). Thus, screening and selection of sources of resistance to CLB are required and are usually based on defoliation (Alfenas et al. 2016; Rezende et al. 2019). Nonetheless, the use of visual assessment of disease severity has been recommended as a more practical method for assessing severity before the occurrence of drastic plant defoliation, reducing the effects of the natural senescence on assessments (Freitas et al. 2019).

Measurements of plant disease severity are needed for many applications of plant pathology such as quantifying yield losses, monitoring and forecasting epidemics, comparing different methods of disease control, characterization of pathogen aggressiveness, and

screening plant genotypes for host resistance (Bock et al. 2010; Araújo et al. 2018; Freitas et al. 2019; Vikram et al. 2021). Plant disease severity is most commonly expressed as the percentage of plant tissue area affected by the presence of disease symptoms or signs of the pathogen (Nutter et al. 1991; Bock et al. 2021b). Visual estimation of severity, *i.e.* for assigning a value or score for the severity perceived by the human eye, must provide accurate and reliable estimates. Accuracy determines the closeness of the visual estimate to the actual value and reliability can be defined as the tendency of replicate estimates of the same specimen(s) obtained under different conditions to be close to one another (Nutter et al. 1991; Madden et al. 2007). However, disease severity is relatively more difficult to assess accurately and inaccurate estimates can lead to erroneous conclusions and inappropriate decisions in disease management (Bock et al. 2020). Thus, the accuracy of the estimates can potentially be improved when aided by standard area diagram sets (SADs) (Del Ponte et al. 2017, 2022).

SADs are illustrations (pictorial or graphic representation) of diseased plants or plant parts selected with distinct severity levels, which are employed to improve the accuracy and reliability of visual estimates of disease severity on a specimen (Bock et al. 2016, 2021b). Although SADs are well established, most commonly, they have been designed and validated for diseases that lack this tool, evaluating only the impact on accuracy and reliability, but not addressing specific aspects related to SAD-specific factors and aspects of raters that may affect improvements on the accuracy of the estimates, such as SAD structure and rater training (Del Ponte et al. 2017). For instance, what characteristics of the incremental severity interval can maximize the accuracy of severity estimates? Is there a significant gain in accuracy and/or reliability that justify having more illustrations at low severity levels? Can the rater training affect gains in the accuracy and reliability of severity estimates? Is there a tendency for raters after training to maintain or maximize gains in accuracy over time even without the aid of

SADs? These and other questions are of fundamental importance for a better understanding of SAD development and usage.

To evaluate CLB on eucalyptus, two SADs were elaborated with severity values incrementing logarithmically by (Damasceno et al. 2014). However, there is evidence that the relationship between visually estimated and actual disease severity is linear, not logarithmic (Bock et al. 2010). Furthermore, the knowledge about sources of errors for visual estimates of disease severity has been improved over the years (Bock et al. 2010, 2021a; Del Ponte et al. 2017). Thus, several recommended practices have been provided to improve accuracy and reliability and minimize the risk of type II errors (Del Ponte et al. 2017). These include, for example, the use of diagrams distributed approximately linearly and spaced no more than 15% apart; and in the severity range of 0 to10%, the inclusion of additional diagrams (±2) considering this range can be particularly prone to overestimation. In some studies where quantitative ordinal scales were designed based on a 10% linear scale with additional classes at <10%, it was demonstrated that an amended linear scale maximizes the accuracy of estimated disease severity (Chiang et al. 2014; Liu et al. 2019). Moreover, the error of the estimates also can be reduced by training based on specialized computer software or instruction (Bardsley and Ngugi 2013). The instruction refers to written or oral descriptions of symptoms, how to delineate them, and a description of how the rating scales used for assessment should be implemented (Bock et al. 2021a).

Considering that the effects of SAD-specific factors and other aspects that may affect the rater's performance remain less explored or with unclear impacts on accuracy and reliability of severity assessments, the objectives of this study were to: (i) assess and compare the accuracy and reliability of visual disease estimates using linear and amended-linear SAD structures different SAD structures to quantify CLB severity, and (ii) evaluate the effect of rater instruction on the quality of visual estimates of CLB.

MATERIAL AND METHODS

Plant inoculation procedures

Conidial suspensions of *Calonectria pteridis* isolate LPF059 (from the culture collection of the Forest Pathology Laboratory, Universidade Federal de Viçosa, Brazil) were produced as previously described (Alfenas et al. 2013). Leaves of 90-day-old plants of *Eucalyptus grandis* x *Eucalyptus urophylla* hybrid clones (CLR158, CLR222, CLR221, and CLR236) were inoculated on both sides by spraying 200 mL/plant of inoculum suspension at 2.5 x 10^4 conidia/mL. After inoculation, the plants were kept in the greenhouse at 26 °C for 48 h with intermittent water nebulization for 10 s at 10-min intervals. Subsequently, they were irrigated for 2 min at 1-h intervals until the collection of symptomatic leaves ten days after inoculation.

Image acquisition

A total of 326 leaves exhibiting various levels of severity (from 0.87% to 95%) were individually digitized using a flatbed scanner (HP LaserJet Pro 200 color MFP M276nw). A blue background was used as a standard for all images. Images were stored in JPG format with a resolution of 300 dpi. The percentage area diseased was manually determined on each leaf using the ImageJ software (Schneider et al. 2012), based on the color threshold method to discriminate between healthy and diseased tissue.

SAD development

Two SADs with different structures were developed that both 10 images of eucalyptus leaves, each depicting a range of severities. The first, the one with amended-linear increments, had three diagrams added between the minimum (0.8%) and 15% severity, and five additional diagrams were included at a 15-percentage point interval until reaching a maximum of 90%

severity (Fig. 1A). The second structure, the linear SAD, had severity levels ranging from 0.8% to 90% distributed equally following a fixed 10-percentage point interval (Fig. 1B).

SAD validation

A subset of 46 images of symptomatic leaves with severity values distributed as evenly as possible in the range from 2.57 to 95.30 was used by two groups of raters to check whether SAD use, structure and rater instruction influenced the accuracy, precision, and reliability of the visual estimates without and with the aids. A total of 43 raters participated in the validation, and all had no experience in evaluating CLB severity. Two groups were randomly defined. One group (21 raters) was assigned to the amended-linear SAD and another group (22 raters) to the linear SAD.

Evaluations were performed remotely via online sessions, in which raters had to perform the assessment on each leaf and record their scores using an online form. For severity assessment, each image was projected onto a Google slides presentation and the assessment of each leaf had to be performed during a fixed assessment time of 30 seconds.

The assessments were performed in three rounds. First, the raters assigned severity values without the aid of the SAD and without instruction. In this step, the raters did not have access to any information about the pathosystem prior to performing the evaluation. Subsequently, the same images were projected in a different order and the raters assessed CLB severity yet without the aid of the SAD but with instruction. The instruction included how to recognize CLB symptoms, distinguish the symptomatic tissues from the non-symptomatic and how the severity can be visually underestimated when leaf lesions have a reddish-purple border. A clear definition of disease severity (percentage area diseased) was provided to raters. Furthermore, in this round, we showed the raters six different images displaying incremental actual severity values. Finally, in the third round, the same images, with the order randomized,

were projected and the raters assessed CLB severity with the aid of the respective SAD assigned for that group. In this step, instructions about how the SAD should be used as a guide for interpolation of their best estimate were provided. The raters were not informed about the existence of a second SAD structure.

Rater-level agreement metrics

Firstly, a general summary of rater-level absolute measurement errors (difference between estimated and actual severity) was performed and analyzed graphically, and the variability of these errors was further evaluated by calculating their respective standard deviations.

The level of agreement between the rater's severity estimates and actual severity was measured using Lin's concordance correlation coefficient (CCC) analysis (Lin 1989). The CCC, also denoted as ρ_c, measures the overall accuracy between estimated and actual severity values, *i.e*. the level of agreement between these two variables. ρ_c ranges from -1 to 1 ($-1 \leq \rho_c \leq 1$) with $\rho_c = 1$ meaning that estimated and actual values are at a maximum agreement, *i.e*. they are equal. Furthermore, ρ_c is given by the product of two parameters: The generalized bias coefficient (C_b) and Precision (Pearson's r). Estimates of ρ_c, r, and C_b for each individual rater were obtained in R (R Core Team 2022) using the 'CCC()' function from the *DescTools* package (Signorell 2022).

Interrater reliability

The interrater reliability, *i.e*, the overall concordance between raters, was evaluated using the intraclass correlation coefficient (ICC). The ICC was calculated for each SAD structure (linear or amended linear) and assessment round (Unaided no-instruction, Unaided

after-instruction, SAD-aided) combination. To estimate the mean ICC and respective 95% confidence intervals, we used the function 'icc()' from the R package *irr* (Gamer et al. 2019).

Effect of SAD structure and instruction

Accuracy, bias, and precision were modeled using generalized linear mixed-effects models, in which the interaction between aid (Unaided no-instruction, Unaided after-instruction, SAD-aided) and SAD structure (linear and amended-linear) were added as fixed effects and the raters as random intercepts. Models for accuracy and precision parameters (ρ_c and Pearson's r) used the Gaussian family, while the Beta family with the *logit* link was used for the bias parameter (C_b). Post-hoc comparisons were performed using the Tukey test ($p < 0.05$). For that, the functions 'emmeans()' and 'cld()' from the emmeans (Lenth 2022) and multcomp (Hothorn et al. 2008) R packages were used.

Window-based accuracy

This analysis was made to assess whether the levels of accuracy are affected by the level of severity in the testing set of images. Therefore, calculations of rater-level agreement metrics (ρ_c, C_b, and Pearson's r) were performed at fixed 15 p.p. intervals or windows, ranging from 0 to 100, with a 5 p.p. step, as depicted in Fig. 2.

Accuracy of unaided estimates over time following SAD use

This experiment aimed to evaluate the raters' accuracy during subsequent rounds of estimates after instruction and performing a round of assessments using the aid of the SADs. A new subset of 50 images (50% used in the previous phases of assessments) was used in two subsequent rounds of unaided severity assessments after 7 and 14 days of the third evaluation round (SAD-aided). Four raters were chosen from the linear SAD assessment group and four

raters were chosen from the amended-linear SAD assessment group. From each group, two raters who obtained the best overall accuracy and two raters with the most significant gains in accuracy after instruction were selected. Raters' unaided assessments were performed as in the first phase of evaluations (unaided-no instruction; described above). Rater-level agreement metrics were calculated as described for each rater in each subsequent assessment.

RESULTS

Estimation errors

A general trend of severity overestimation was observed at lower levels of severity with a sharp shift to underestimation when scoring leaves with severity over 25% (Fig. 3). This pattern was more evident in the unaided rounds of assessments, either with or without training. The SAD-aided estimates using the amended-linear SAD were slightly more effective in reducing deviations at lower levels of disease, compared with the linear SAD. Overall, the standard deviation of the absolute error decreased across the consecutive round of assessments from 9.78 (unaided no-instruction) to 9.00 (unaided after-instruction) to 7.20 (SAD-aided) for the amended-linear SAD group and 9.36 (unaided no-instruction) to 8.46 (unaided after-instruction) to 7.77 (SAD-aided) for the linear SAD group.

Accuracy, precision, and Interrater reliability

In general, unaided estimates had reasonably good accuracy, with a mean CCC higher than 0.9. Nevertheless, there was an overall trend of improvement in assessment accuracy as the raters moved through the assessment rounds (Fig. 4). However, no statistically significant improvement in accuracy was observed after the raters' instruction. Nonetheless, a significant increment in accuracy was observed when assessments were SAD-aided (compared to unaided

no- instruction round; see Fig. 4). Accuracy values increased from 0.93 to 0.96 using both amended-linear and linear SAD.

Compared to the first round of assessments (unaided no-instruction), the majority of raters had their accuracy increased when assessing severity using any of the SADs (Fig. 5A). Raters with lower baseline accuracy (unaided no-instruction) had higher gains than inherently accurate raters. Two raters (one from each SAD group) rater increased accuracy by around 0.10. However, lesser gains in accuracy were observed as degrees of aid and instruction increased from unaided after-instruction to SAD-aided and from unaided no-instruction to unaided after-instruction (Fig 5B and Fig 5C).

Higher values (>0.97) of generalized bias coefficient (C_b) were obtained among the SAD groups and assessment rounds (Table 1). Generally, values were very similar among the groups, besides a significantly lower value for the unaided (no-instruction) estimates from the linear SAD group of raters, compared to the amended-linear SAD group. Moreover, a slight, but significant, increase in C_b was obtained from SAD-aided assessments using the linear SAD, compared to the previous rounds (Table 1). Precision estimates (Pearson's r) increased across the rounds of assessment for both raters' groups, however, significant increments (compared to unaided no-instruction) were only obtained on the SAD-aided rounds (Table 1). No differences were observed between raters' groups within each round of assessment.

The interrater reliability estimates were very similar between SAD groups within each round of assessments. There was a trend of increase in the values of ICC as raters progress through the rounds of assessments (for both SAD groups), but confidence intervals overlapped, indicating no significant difference (Table 1).

Window-based accuracy

Window-based (intervals of 15 percent point) overall accuracy, precision, and generalized bias varied greatly across different levels of disease severity (Fig. 6A). Regarding overall accuracy, an oscillation pattern was observed for both SAD structures, in which lower values were observed around the intervals of 20% to 35%, 40% to 55%, and 60% to 75% severity. The amended-linear SAD favored more accurate estimates at lower levels of severity. However, window-based precision was not improved at any interval of severity (Fig. 6B). On the other hand, the most evident effect of SAD structures was on the generalized bias. The amended-linear SAD favored increases in the bias coefficient at lower severity (<30%), while not providing increments at intervals over 35%. Furthermore, the linear SAD favored high gains in bias coefficient at intervals with higher levels of severity (>35%), while not providing improvements at lower intervals (Fig. 6C).

Persistent accuracy of unaided estimates after SAD use

The raters that were less accurate in the first round of assessment (unaided no-instruction) from both SAD groups improved their CCC after the instruction. The same level of accuracy was maintained in the following round, using the SAD. After 7 and 14 days, when making unaided assessments, these raters preserved the level of accuracy observed in the SAD-aided assessments (Fig. 7). The raters that were more accurate in the first round kept a consistent level of accuracy on all the subsequent rounds of assessment, including the last two unaided rounds.

DISCUSSION

SADs have been developed and validated for different pathosystems. They were aimed at standardizing the assessments and obtaining more accurate and reliable disease severity

estimates. However, issues related to SAD-specific factors or aspects of raters that may affect improvements in the accuracy and reliability remained less explored (Bardsley and Ngugi 2013; Schwanck and Del Ponte 2014; Franceschi et al. 2020; Belan et al. 2020). Thus, based on the best operating practices for developing and using SADs (Del Ponte et al. 2017; Bock et al. 2021a), we developed two SADs, each with ten severity values, one with severity increasing following the amended-linear (15% increment) and another with linear increments (10% increment). For validation purposes, we provided training instructions to the raters before the second round of unaided assessments. Furthermore, the persistence of accuracy of unaided estimates in two rounds after SADs use was also assessed. Corroborating previous statements (Bock et al. 2010, 2021a; Del Ponte et al. 2017), we found an overall overestimation at low severity levels (<25%). This overestimation was more evident when not using any aid, regardless of instruction level. In contrast, using the amended-linear SAD the overestimation was reduced at the lower end. These results are consistent with the suggestion that the inclusion of additional reference diagrams on a linear scale at the lower end of the severity range on a linear scale would reduce the errors of estimates within that interval (Schwanck and Del Ponte 2014; Del Ponte et al. 2017). Also the results of Chiang et al. (2014) and Liu et al. (2019) indicated that an amended-linear ordinal scale with additional classes at low severity improved the accuracy of estimates.

For determining the overall accuracy, we used Lin's concordance correlation coefficient (Lin 1989). It combines the measures of accuracy and precision, providing a more appropriate method to test agreement or accuracy (Madden et al. 2007; Bock et al. 2010). Raters with a CCC of around 0.90 or higher can be considered accurate. However, a visual rater rarely will show a consistent CCC greater than 0.95. In most cases, raters with training instruction and/or SADs will have a CCC ≥0.85 to 0.95 (Bock et al. 2021a). Thus, the mean CCC of unaided estimates in our study indicated accurate estimates (CCC ≥0.93 to 0.95). Nevertheless, when

disease severity was estimated with any of the two SADs, significant improvement in accuracy was observed compared with unaided no-instruction estimates. Indeed, the more biased and imprecise the unaided estimate, the greater the benefits to using SADs (Del Ponte et al. 2022). In our study, analysis of the individual components of CCC, the generalized bias coefficient (C_b), and the precision (Pearson`s r), indicated a significant reduction of bias and increase of precision on linear SAD estimates, as well as a significant increase in precision on amended-linear SAD estimates, both compared with estimates unaided no-instruction. Despite that, based on concordance analysis, there was no significant advantage for any specific type of SAD structure. This may be explained because additional reference diagrams were placed at the lower end of the severity range for amended-linear SADs and so reduction of overestimation was found in that range. In contrast, linear SADs were more effective in reducing errors at higher levels of disease (>25%), also where there are more reference diagrams. In fact, the most obvious effect of SAD structures was on C_b verified through window-based accuracy analysis. In this case, the amended-linear SAD reduced the bias at lower severity (<30%), while not reducing at intervals over 35%. On the other hand, linear SAD favored reduction in the bias at intervals with higher levels of severity (>35%), while not providing improvements at lower intervals.

The lower severity range is important for many studies in plant pathology , as it is a low severity disease that needs to be estimated with greater accuracy because these assessments serve as the basis for epidemiological studies, or comparing treatments of most interest, such as fungicide effectiveness and host resistance (Chiang et al. 2016; Chiang and Bock 2022). As noted, there is a tendency to overestimate disease severity at lower levels, which is especially greater when foliar lesions are small and numerous (Bock et al. 2010), such as occur at the earliest stage of CLB, thus leading to the false impression that the lesions occupy more area

than they actually do. Therefore, it is particularly important to consider any approach that can reduce errors, especially in the lower disease severity range.

Interestingly, there was no significant difference after instruction, although there was a tendency for improvement in the accuracy and reliability of the estimates. Raters vary in their intrinsic ability to estimate disease severity. In this regard, some raters can be inherently more accurate, yet others can be inaccurate (Bock et al. 2020). Then, such results, in general, can be explained because the rater's baseline accuracy was high from the start, with few exceptions. Consequently, unaided estimates had low biases and were reasonably precise, accurate, and reliable. Indeed, exploring the results of gains in accuracy, the more inaccurate the unaided estimates, the greater the benefits of using instruction. The raters who were less accurate in unaided no-instruction estimates from both SAD groups improved their accuracy after the instruction and maintained the same level of accuracy in the two rounds (7 and 14 days) after the use of SAD. These results reinforced the observation that wherever possible raters should be experienced, what may perhaps be achieved through training or instruction (Bock et al. 2015, 2021a).

The instruction in identification and disease assessment provided an opportunity for the raters to recognize the disease symptoms and, this experience of symptom recognition has an impact on the ability to estimate severity accurately (Bardsley and Ngugi 2013; Bock et al. 2020). Thus, although raters varied in their ability to estimate disease severity, the instruction appears to be a way to bring the inherently less accurate raters closer to the raters inherently more accurate in terms of their ability in this study. However, it remains undefined for how long the level of accuracy will be maintained in the estimates unaided after instruction and how often the instruction should be repeated to maintain the accuracy of the estimates. To verify this issue, it would be necessary to carry out more unaided assessment rounds after a instruction round and SADs use until it was possible to detect a reduction in the level of accuracy.

In conclusion, the results from the present study show in general SADs developed increased the accuracy of estimates of CLB severity. Instruction appears to be critical to obtaining more accurate and reliable estimates, especially for inherently less accurate raters, and therefore, whenever possible, it should be provided to raters. Given the importance of visual assessment of severity for the selection of sources of resistance to CLB and, considering the symptoms characteristic of the disease, we recommend the employed the amended-linear set for assessments of CLB severity, which minimizes the overestimation at low levels of severity.

ACKNOWLEDGMENTS

This work was funded by 'Conselho Nacional de Desenvolvimento Científico Tecnológico' (CNPq) and FAPEMIG. K.S.A. is thankful to 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil' (CAPES) – Finance code 001 for a scholarship. The authors also acknowledge Clonar Resistência a Doenças Florestais for providing the plant material and infrastructure for inoculation, to all the raters for their contribution to the SADs validation and to Dr. Marston Franceschini (Forest Pathology Laboratory/Bioagro/UFV) for revising the manuscript.

Author Contribution Statement

M.F.Q., K.S.A., E.M.D., and A.C.A. conceived and designed the research; K.S.A. performed the data analysis; M.F.Q., K.S.A., and C.F.P. conducted experimental procedures; E.M.D., R.F.A. and A.C.A. oversaw the research; M.F.Q., K.S.A., and E.M.D. wrote the manuscript.

Data Availability Statement

The data that support the findings of this study are openly available in the Open Science Framework at http://doi.org/10.17605/OSF.IO/WZ3JE.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Alfenas RF, Freitas RG, Pereira OL, Coutinho MM, Zarpelon TG, Cândido TS, Alfenas AC (2016) Screening of *Corymbia* and *Eucalyptus* species for resistance to *Calonectria pteridis* leaf blight. Forest Pathology 46:76–81.

Alfenas RF, Lombard L, Pereira OL, Alfenas AC, Crous PW (2015) Diversity and potential impact of *Calonectria* species in *Eucalyptus* plantations in Brazil. Studies in Mycology 80:89–130.

Alfenas RF, Pereira OL, Freitas RG, Freitas CS, Dita MAD, Alfenas AC (2013) Mass spore production and inoculation of *Calonectria pteridis* on *Eucalyptus* spp. under different environmental conditions. Tropical Plant Pathology 38:406–413.

Araújo ER, Alves DP, Higashikawa FS (2018) Cultivar resistance and chemical, biological and fertilizer treatments for controling *Botrytis* leaf blight of onion. Tropical Plant Pathology 43:160–164.

Bardsley SJ, Ngugi HK (2013) Reliability and accuracy of visual methods to quantify severity of foliar bacterial spot symptoms on peach and nectarine. Plant Pathology 62:460–474.

Belan LL, Belan LL, da Matta Rafael A, Gomes CAG, Alves FR, de Jesus Junior, WC, Moraes WB (2020) Standard area diagram with color photographs to estimate the severity of coffee leaf rust in *Coffea canephora*. Crop Protection 130:105077.

Bock CH, Barbedo JGA, Del Ponte EM, Bohnenkamp D, Mahlein A-K (2020) From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. Phytopathology Research 2:1-30.

Bock CH, Chiang K-S, Del Ponte EM (2021a) Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. Tropical Plant Pathology 47:25–42.

Bock CH, El Jarroudi M, Kouadio LA, Mackels C, Chiang K-S, Delfosse P (2015) Disease Severity Estimates—Effects of Rater Accuracy and Assessment Methods for Comparing Treatments. Plant Disease 99:1104–1112.

Bock CH, Hotchkiss MW, Wood BW (2016) Assessing disease severity: accuracy and reliability of rater estimates in relation to number of diagrams in a standard area diagram set. Plant Pathology 65:261–272.

Bock CH, Pethybridge SJ, Barbedo JGA, Esker PD, Mahlein A-K, Del Ponte EM (2021b) A phytopathometry glossary for the twenty-first century: towards consistency and precision in intra- and inter-disciplinary dialogues. Trop plant pathology 47:14–24.

Bock CH, Poole GH, Parker PE, Gottwald TR (2010) Plant Disease Severity Estimated

Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging. Critical Reviews in Plant Sciences 29:59–107.

Chiang K-S, Bock CH (2022) Understanding the ramifications of quantitative ordinal scales on accuracy of estimates of disease severity and data analysis in plant pathology. Tropical Plant Pathology 47:58–73.

Chiang K-S, Bock CH, Lee I-H, El Jarroudi M, Delfosse P (2016) Plant Disease Severity Assessment-How Rater Bias, Assessment Method, and Experimental Design Affect Hypothesis Testing and Resource Use Efficiency. Phytopathology 106:1451–1464.

Chiang K-S, Liu S-C, Bock CH, Gottwald TR (2014) What Interval Characteristics Make a Good Categorical Disease Assessment Scale? Phytopathology 104:575–585.

Damasceno VFF, Furtado EL, Ferreira Filho PJ (2014) Comparação de dois métodos de elaboração e validação de escala diagramática para a quantificação da severidade da mancha de *Cylindrocladium* em eucalipto. Summa phytopathologica 40:248–255.

Del Ponte EM, Cazón LI, Alves KS, Pethybridge SJ, Bock CH (2022) How much do standard area diagrams improve accuracy of visual estimates of the percentage area diseased? A systematic review and meta-analysis. Tropical Plant Pathology 47:43–57.

Del Ponte EM, Pethybridge SJ, Bock CH, Michereff SJ, Machado FJ, Spolti P (2017) Standard area diagrams for aiding severity estimation: Scientometrics, pathosystems, and methodological trends in the last 25 years. Phytopathology 107:1161–1174.

Ferreira FA, Alfenas AC, Moreira AM, Demuner NL (1995) Mancha-de-pteridis: doença foliar de eucalipto em áreas tropicais brasileiras. Fitopatologia Brasilieira 20:107–110.

Franceschi VT, Alves KS, Mazaro SM, Godoy CV, Duarte HSS, Del Ponte EM (2020) A new standard area diagram set for assessment of severity of soybean rust improves accuracy of estimates and optimizes resource use. Plant Pathology 69:495–505.

Freitas RG, Alfenas RF, Guimarães LMS, Badel JL, Alfenas AC (2019) Genetic diversity and aggressiveness of *Calonectria pteridis* in Eucalyptus spp. Plant Pathology 68:869–877.

Gamer M, Lemon J, Singh <puspendra.pusp22@gmail.com> IFP (2019) irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1.

Hothorn T, Bretz F, Westfall P (2008) Simultaneous Inference in General Parametric Models. Biometrical Journal 50:346–363.

Lenth RV (2022) emmeans: Estimated marginal means, aka least-squares means. R package version 1.8.6.

Lin LI-K (1989) A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45:255–268.

Liu HI, Tsai JR, Chung WH, Bock CH, Chiang KS (2019) Effects of Quantitative Ordinal Scale Design on the Accuracy of Estimates of Mean Disease Severity. Agronomy 9:565.

Madden LV, Hughes G, van den Bosch F (2007) The study of plant disease epidemics. APS Press, Saint Paul.

Nutter FW Jr, Teng PS, Shokes FM (1991) Disease assessment terms and concepts. Plant Disease 75:1187–1188.

Pham NQ, Barnes I, Chen S, Liu F, Dang QN, Pham TQ, Lombard L, Crous PW, Wingfield MJ (2019) Ten new species of *Calonectria* from Indonesia and Vietnam. Mycologia 111:78–102.

R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rezende EH, Duin IM, Coelho TAV, Soares ID, Higa AR, Santos AF, Silva LD, Auer CG (2019) Avaliação da resistência de progênies de *Eucalyptus grandis* para mancha foliar de *Cylindrocladium* e *Kirramyces*. Summa Phytopathologica 45:295–301.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675.

Schwanck AA, Del Ponte EM (2014) Accuracy and Reliability of Severity Estimates Using Linear or Logarithmic Disease Diagram Sets in True Colour or Black and White: a Study Case for Rice Brown Spot. Journal of Phytopathology 162:670–682.

Signorell A (2022) DescTools: Tools for descriptive statistics. R Package Version 0.99.38.

Vikram P, Sehgal D, Sharma A, Bhavani S, Gupta P, Randhawa M, Pardo N, Basandra D, Srivastava P, Singh S, Sood T, Sansaloni CP, Rahman H, Singh S (2021) Genome-wide association analysis of Mexican bread wheat landraces for resistance to yellow and stem rust. PLoS One 16:e0246015.

Figures and Tables

Table 1. Marginal means and 95% confidence intervals (•) for the two components of Lin's concordance coefficient, the generalized bias coefficient (C_b) and the precision (Pearson`s $r$), for rater`s estimates of calonectria leaf blight severity in eucalyptus' leaves using the amended-linear and linear SADs for each phase of assessment. The level of agreement between raters (interrater reliability) is given by the intraclass correlation coefficient (ICC).

	Lin's concordance analysis components[1]				Interrater reliability	
	Generalized bias coefficient (C_b)		Precision (r)		ICC	
	Am-Linear	Linear	Am-Linear	Linear	Am-Linear	Linear
Unaided (No instruction)	0.98 **a A** (0.97, 0.98)	0.97 **b B** (0.97, 0.98)	0.95 **b A** (0.94, 0.96)	0.95 **b A** (0.94, 0.96)	0.82 (0.76, 0.88)	0.84 (0.78, 0.89)
Unaided (After instruction)	0.98 **a A** (0.98, 0.99)	0.98 **b A** (0.97, 0.98)	0.96 **ab A** (0.95, 0.97)	0.96 **ab A** (0.95, 0.97)	0.84 (0.78, 0.89)	0.86 (0.81, 0.91)
SAD-Aided	0.99 **a A** (0.98, 0.99)	0.99 **a A** (0.98, 0.99)	0.97 **a A** (0.96, 0.98)	0.97 **a A** (0.96, 0.97)	0.87 (0.81, 0.91)	0.86 (0.80, 0.90)

[1] Means followed by a common letter are not significantly different by the Tukey-test at the 5% level of significance. Uppercase letters should be compared within the rows, while lowercase should be compared within the columns for each statistic.

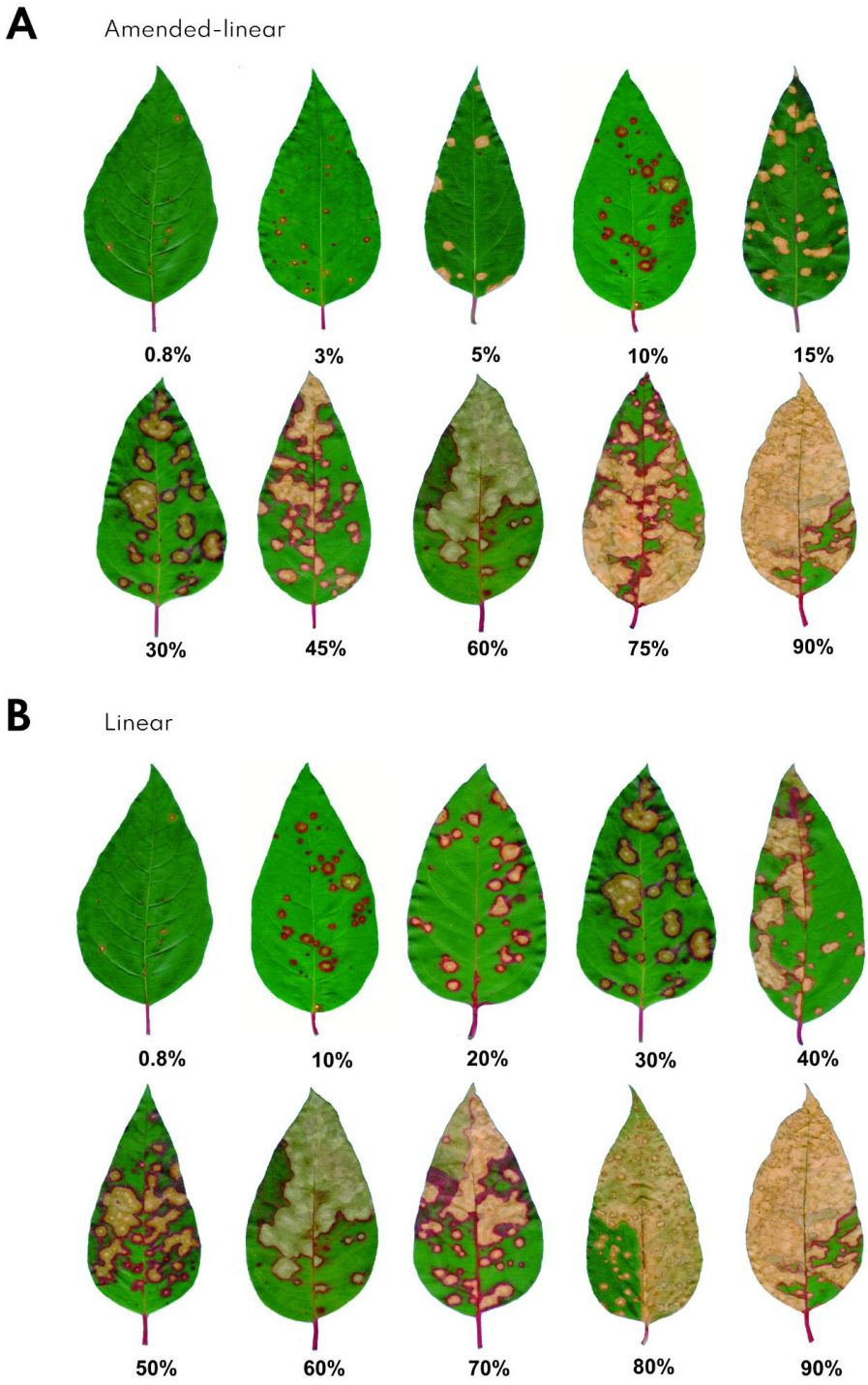

Fig. 1. Standard area diagrams (SADs) for aiding visual estimates of the severity of calonectria leaf blight (*Calonectria* spp.) in eucalyptus leaves using two different scale structures: (A) amended-linear and (B) linear

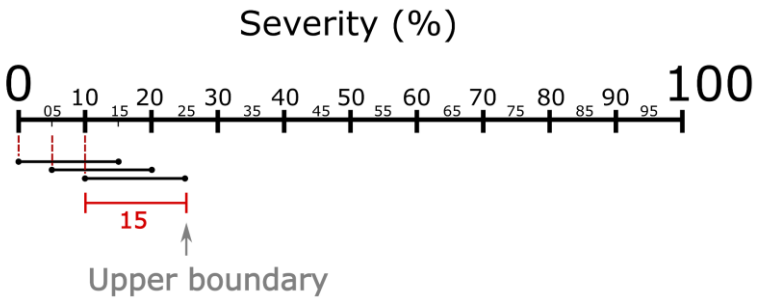

Fig. 2. Diagram of the windows used to calculate window-based accuracy.

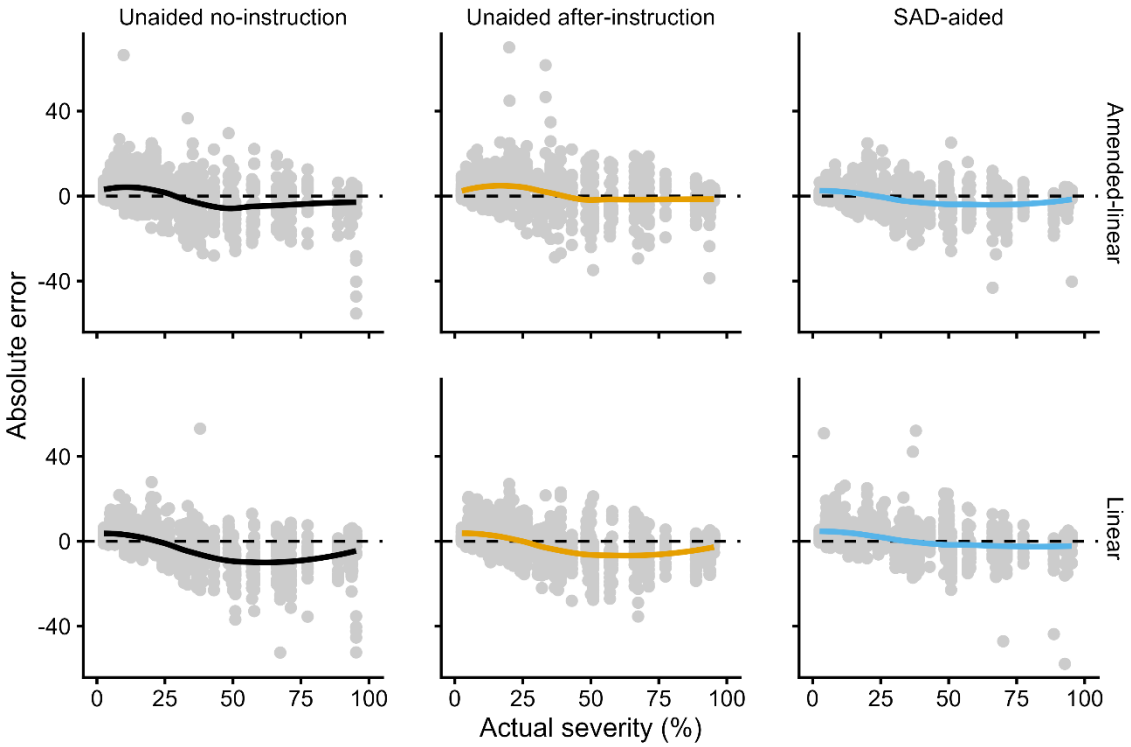

Fig. 3. Scatter plot for the relationship between raters` absolute measurement errors and leaf actual severity for each combination of SAD structure (Amended-linear and Linear) and evaluation step (Unaided no-instruction, Unaided after instruction, and SAD-aided) for assessing Calonectria leaf blight severity in eucalyptus leaves. The dots in each plot represent the individual error values and the lines represent the smooth function estimated using the loess method.

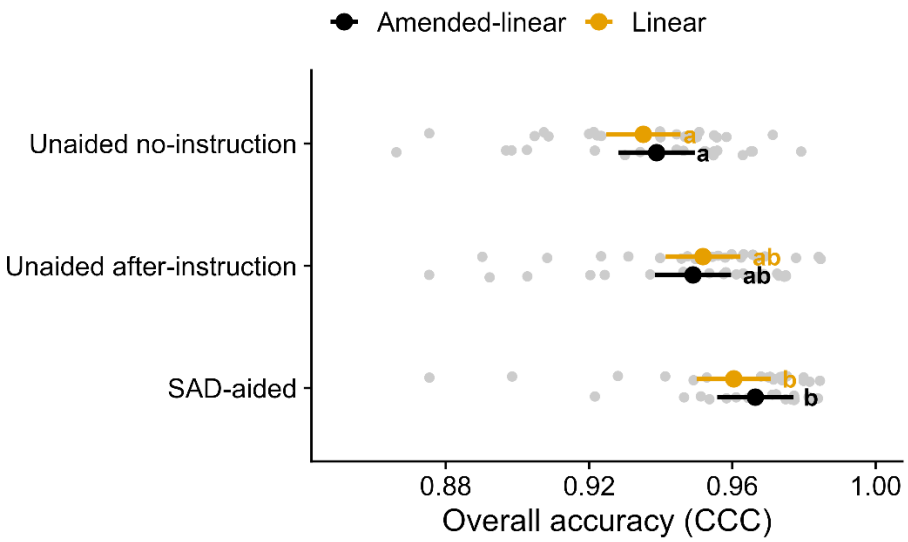

Fig. 4. Overall accuracy of rater`s estimates of Calonectria leaf blight severity in eucalyptus leaves using the amended-linear and linear SADs (colors) for each assessment phase (y-axis). Grey dots represent raters' individual Lin's concordance correlation coefficient (CCC or ρ_c). Error bars represent the 95% confidence interval for the mean CCC. Letters represent the Tukey mean comparison test groupings (5% probability), in which groups with the same color should be compared. Groups accompanied by a common letter are not significantly different.

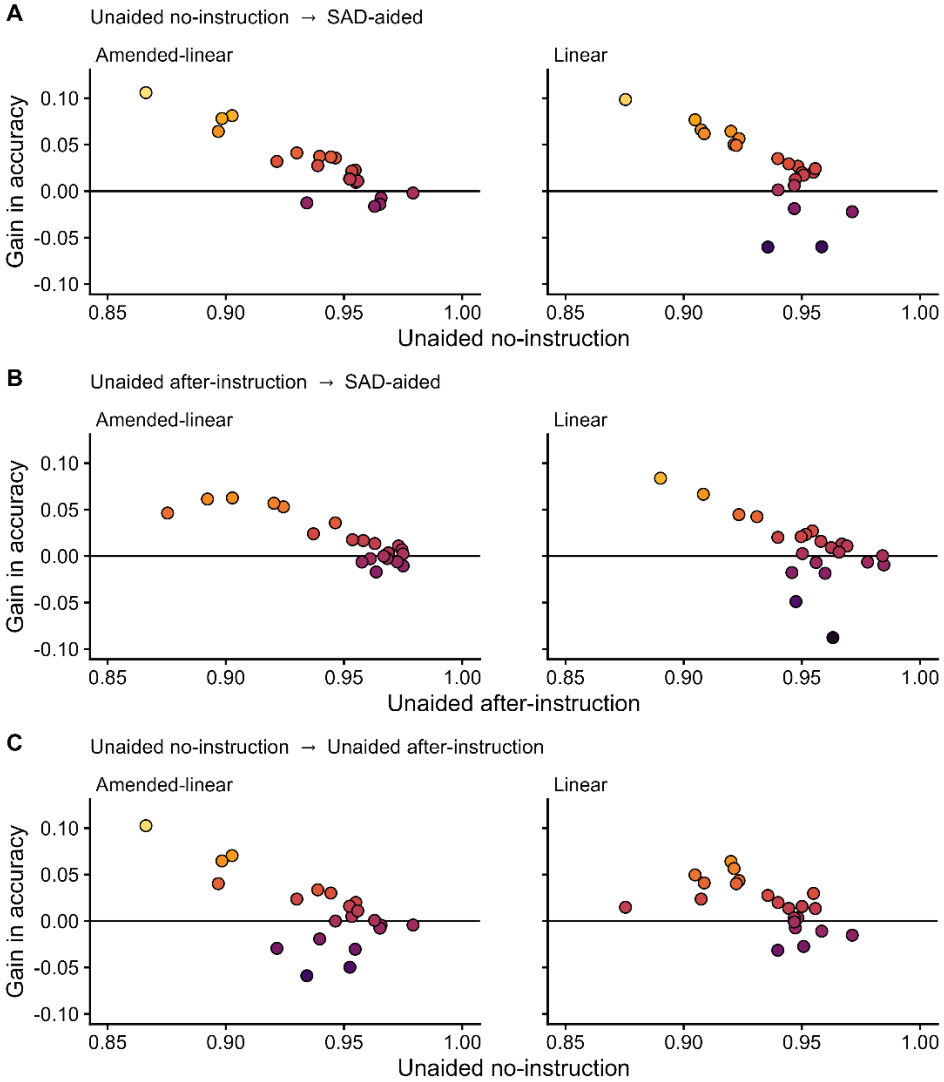

Fig. 5. Relationship between unaided (no-training or after-instruction) overall accuracy (Lin's concordance correlation coefficient; ρ_c) and gain in accuracy for the amended-linear (left-hand column) and linear (right-hand column) SAD groups. (A) Gain in accuracy from unaided no-instruction to SAD-aided rounds of assessment; (B) Gain in accuracy from unaided after-instruction to SAD-aided rounds of assessment; (C) Gain in accuracy from unaided no-instruction to unaided after-instruction round of assessment.

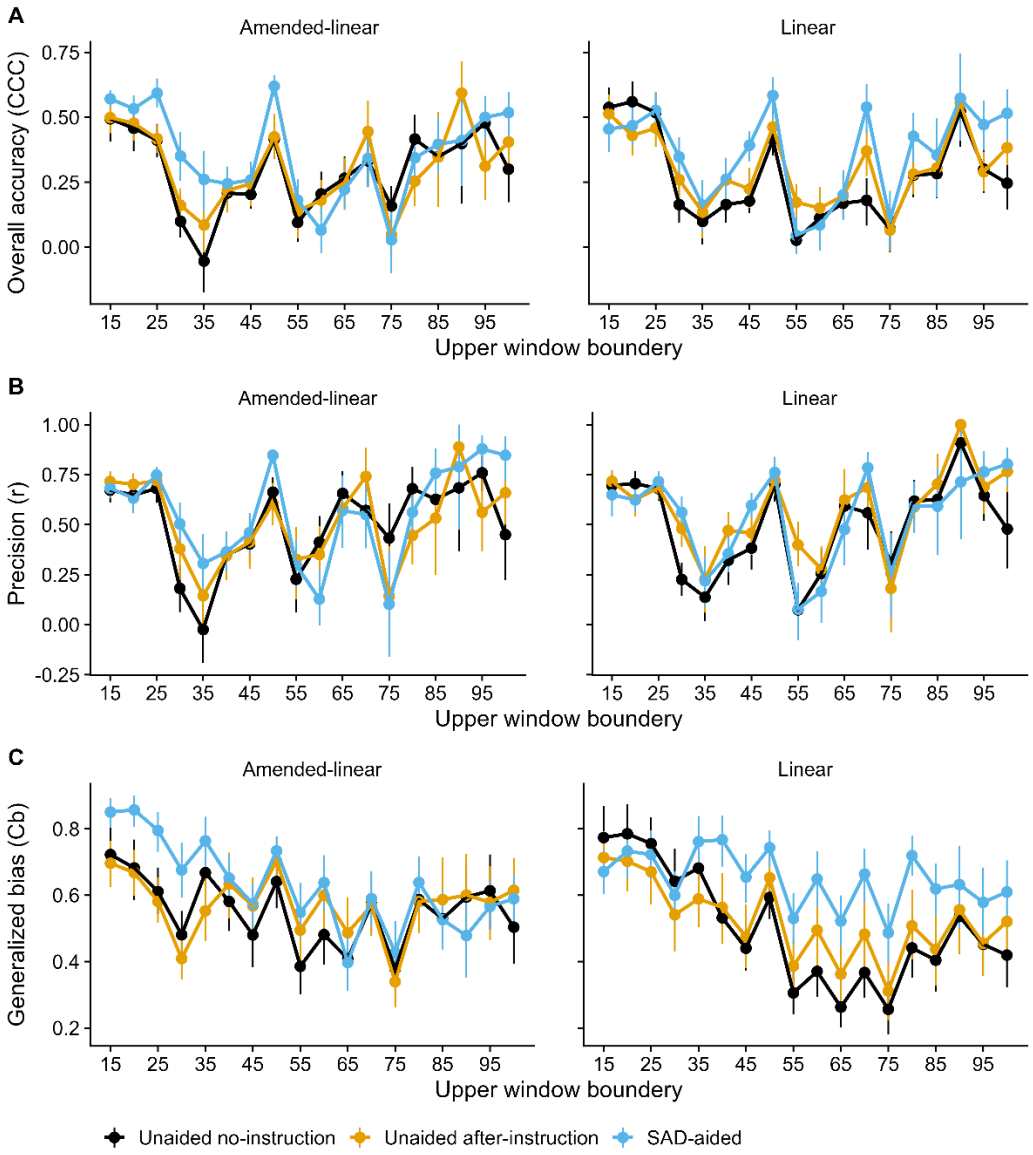

Fig. 6. Window-based (intervals of 15 percent points; see Fig. 2) estimates of (A) overall accuracy (Lin's concordance correlation coefficient; ρ_c), (B) precision (Pearson's correlation coefficient), and (C) bias coefficient for each round of severity assessment (unaided no-instruction, unaided after instruction, and SAD-aided) and for each standard area diagram structure (amended linear and linear). Solid points and error bars indicate bootstrapped means and 95% confidence intervals.

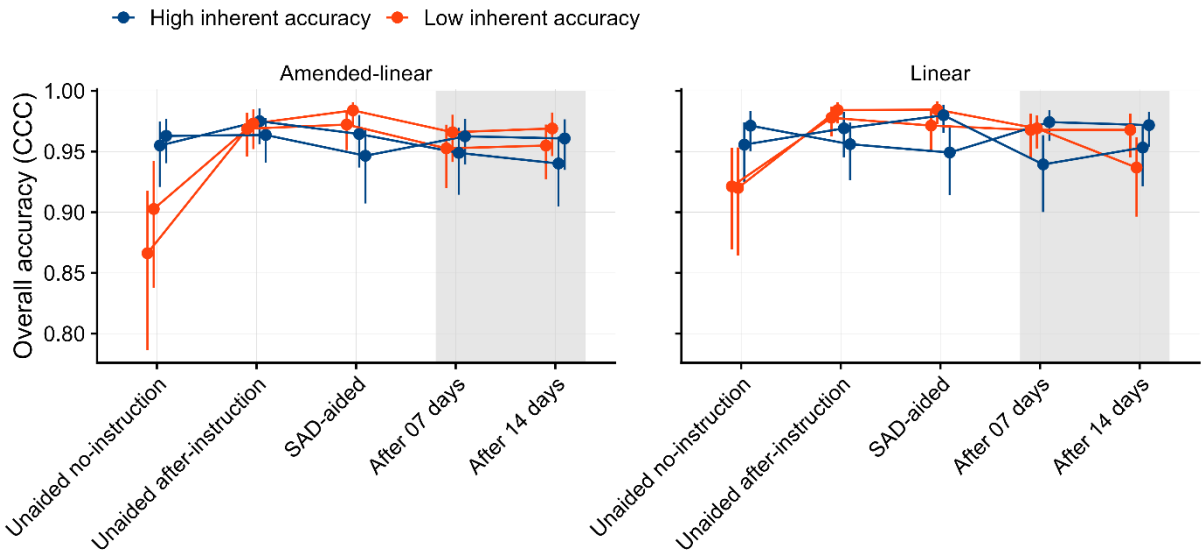

Fig. 7. Accuracy of severity estimates over time. Estimated Lin's concordance correlation coefficient (CCC or ρ_c) across the five rounds of severity assessment (x-axis) for eight selected raters. Four from the amended-linear SAD group and four from the linear SAD group. Two of each group were highly inherently accurate and the other two were less inherently accurate in the unaided no-instruction round of assessment. The gray shaded area highlights the two last rounds of unaided assessments, done 7 and 14 days after the SAD-aided round.