DANYLLO AMARAL DE OLIVEIRA

SNPS MARKERS FOR IDENTIFICATION OF EUCALYPTUS SPECIES AND HYBRIDS AND RNA-SEQ TO IDENTIFY GENES ASSOCIATED WITH THE EUCALYPTUS PHYSIOLOGICAL DISORDER

LAVRAS - MG
2024

DANYLLO AMARAL DE OLIVEIRA

SNPS MARKERS FOR IDENTIFICATION OF EUCALYPTUS SPECIES AND HYBRIDS AND RNA-SEQ TO IDENTIFY GENES ASSOCIATED WITH THE EUCALYPTUS PHYSIOLOGICAL DISORDER

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Doutor

Prof. Dr. Evandro Novaes
Orientador

LAVRAS - MG
2024

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Oliveira, Danyllo Amaral de.

SNPs markers for identification of eucalyptus species and hybrids and RNA-seq to identify genes associated with the eucalyptus physiological disorder / Danyllo Amaral de Oliveira. - 2023.

67 p.

Orientador(a): Evandro Novaes.

Tese (doutorado) - Universidade Federal de Lavras, 2023.

Bibliografia.

1. Genotipagem. 2. Taxonomia. 3. Expressão gênica. I. Novaes, Evandro. II. Título.

Fonte: Universidade Federal de Lavras (2024).

DANYLLO AMARAL DE OLIVEIRA

MARCADORES SNPS PARA IDENTIFICAÇÃO DE ESPÉCIES E HÍBRIDOS DE EUCALIPTO E RNA-SEQ PARA IDENTIFICAÇÃO DE GENES ASSOCIADOS AO TRANSTORNO FISIOLÓGICO DO EUCALIPTO

SNPS MARKERS FOR IDENTIFICATION OF EUCALYPTUS SPECIES AND HYBRIDS AND RNA-SEQ TO IDENTIFY GENES ASSOCIATED WITH THE EUCALYPTUS PHYSIOLOGICAL DISORDER

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Doutor

APROVADA em 15 de dezembro de 2023

Dr. Welison Andrade Pereira UFLA
Dr. Paulo Eduardo Ribeiro Marchiori UFLA
Dr. Matias Kirst UNIVERSITY OF FLORIDA
Dr. Edival Angelo Valverde Zauza SUZANO S/A

Prof. Dr. Evandro Novaes
Orientador

LAVRAS - MG
2024

Aos meus pais Irenaldo Oliveira e Rejane Oliveira
por me mostrar o caminho da educação como
construção de pessoas do bem
Dedico

AGRADECIMENTOS

Agradeço a Deus por ter me dado força para a conclusão de uma etapa de extrema importância na minha vida.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro e concessão da bolsa de estudos. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Aos meus professores e amigos do Programa de Pós-graduação em Genética e Melhoramento de Plantas - PPGM da Universidade Federal de Lavras por todos os ensinamentos e momentos de aprendizado durante o período de doutoramento.

Ao Instituto de Pesquisas Florestais – IPEF, na pessoa do Dr. Paulo Henrique Müller da Silva e a Embrapa Recursos Genéticos e Biotecnologia na pessoa do Dr. Dario Grattapaglia pela parceria no primeiro artigo dessa tese.

À empresa Suzano papel e celulose, unidade de biotecnologia FuturaGene, e todos os seus colaboradores, em especial a Rodrigo Graça, Antonio Porto, Magnus Kerber, Bruna Brumer, José Mateus e Dror Avisar, pela parceria e colaboração na excussão do segundo artigo desta tese. A troca de experiências na execução deste projeto foi fundamental para o meu crescimento profissional.

Ao professor Matias Kirst e todos os membros do *Forest Genomics Lab* na *University of Florida* por me receber durante o meu período de doutorado sanduíche. Foi um período de muito aprendizado científico e cultural.

Agradecimento especial ao meu professor, orientador e amigo Evandro Novaes, o qual me ajudou durante toda a caminhada do doutoramento com muita paciência, humildade e inteligência.

A minha família, aos meus pais Irenaldo Oliveira e Rejane Oliveira, obrigado pelo incentivo na minha educação, e aos meus irmãos Jamilly Oliveira e Daniel Oliveira obrigado pelo apoio nessa jornada.

Agradeço a minha companheira Lana Carvalho pela ajuda durante todas as fases dessa caminhada, desde a inscrição no processo seletivo para o doutorado até a defesa desta tese.

Muito obrigado!

RESUMO

Eucalyptus spp. é um gênero arbóreo economicamente importante com cerca de 700 espécies. Características favoráveis como o rápido crescimento, adaptabilidade a diferentes ecossistemas, grande diversidade genética dentro e entre espécies, boa densidade da madeira e alto rendimento na fabricação de papel e celulose tornam o eucalipto importante na silvicultura. Empresas de produtos florestais e centros de pesquisa buscam por novas tecnologias para auxiliar no melhoramento genético do eucalipto frente aos novos desafios. A aplicação da genômica na eucaliptocultura possui como principal objetivo auxiliar o melhoramento genético, diminuindo o tempo dos ciclos de seleção. Todavia, a genômica pode também ser utilizada como ferramenta no estudo de desafios ainda não solucionados. Neste documento são pontuados dois desafios da eucaliptocultura e a utilização de técnicas genômicas como ferramenta para ajudar o melhoramento do gênero. O primeiro artigo desta tese traz a identificação de espécies e híbridos interespecíficos de eucalipto com base na utilização de marcadores SNP (polimorfismo de um único nucleotídeo) com ampla cobertura do genoma. O segundo artigo traz uma abordagem de RNA-seq para identificar genes e rotas metabólicas associadas ao distúrbio fisiológico do eucalipto, na tentativa de iluminar o possível agente etiológico da doença que ainda é desconhecido. No primeiro estudo os marcadores SNPs foram capazes de identificar a maior parte das espécies e seus híbridos interespecíficos de acordo com as principais classificações filogenéticas do gênero, além de comparar a identificação genômica com o pedigree anotado por melhoristas. Como o gênero *Eucalyptus* possui várias espécies (>600) e classificação taxonômica complexa, este estudo oferece uma alternativa molecular para os melhoristas identificarem de forma objetiva e precisa a composição genômica de seus híbridos. Nosso estudo mostra que a composição genômica dos genitores pode ser diferente daquela esperada pelas anotações do seu pedigree. No segundo estudo, a comparação da expressão gênica entre clones suscetíveis e resistentes mostrou que o distúrbio fisiológico afeta diversas rotas metabólicas, sendo portanto, uma doença complexa a nível molecular. Rotas de sinalização de estresses e genes relacionados a produção de energia e fermentação foram diferencialmente expressos entre clones suscetíveis e resistentes. Terpenóides e outras vias também foram identificadas como importantes para a resistência ou suscetibilidade de clones à PDE. Desta forma, ambos os estudos apresentam a utilização de diferentes técnicas genômicas e geração de amplos bancos de dados para ajudar em dois problemas importantes para o melhoramento da eucaliptocultura: identificação de espécies e híbridos, e de genes envolvidos na resposta de clones resistentes ao distúrbio fisiológico do *Eucalyptus*.

Palavras-chave: *Eucalyptus*; Híbridos; Taxonomia; Genotipagem; Expressão gênica.

ABSTRACT

Eucalyptus spp. is an economically important tree genus with about 700 species. Favorable characteristics such as rapid growth, adaptability to different ecosystems, great genetic diversity within and between species, good wood density, and high yield in paper and pulp manufacturing make *Eucalyptus* important in forestry. Forest product companies and research centers seek new technologies to assist in the genetic improvement of *Eucalyptus* in the face of new challenges. The application of genomics in *Eucalyptus* plantation has the main objective of assisting genetic improvement, reducing the time of selection cycles. However, genomics can also be used as a tool in the study of challenges that have not yet been solved. This document points out two challenges of *Eucalyptus* cultivation and the use of genomic techniques as a tool to help improve the genus. The first article in this thesis brings the identification of species and interspecific hybrids of *Eucalyptus* based on the use of SNP markers (single nucleotide polymorphism) with wide genome coverage. The second article brings an RNA-seq approach to identify genes and metabolic pathways associated with the physiological disorder of *Eucalyptus* to illuminate the possible etiological agent of the disease that is still unknown. In the first study, the SNP markers were able to identify most of the species and their interspecific hybrids according to the main phylogenetic classifications of the genus, in addition to comparing the genomic identification with the pedigree annotated by breeders. As the *Eucalyptus* genus has several species (>600) and complex taxonomic classification, this study offers a molecular alternative for breeders to objectively and accurately identify the genomic composition of their hybrids. Our study shows that the genomic composition of the parents can be different from that expected by the annotations of their pedigree. In the second study, the comparison of gene expression between susceptible and resistant clones showed that the physiological disorder affects several metabolic pathways, therefore, it is a complex disease at the molecular level. Stress signaling pathways, genes related to energy production, and fermentation pathway were differentially expressed between susceptible and resistant clones. Terpenoids and other pathways were also identified as important for the resistance or susceptibility of clones to PDE. Thus, both studies present the use of different genomic techniques and the generation of large databases to help in two important problems for the improvement of *Eucalyptus* plantations: identification of species and hybrids, and of genes involved in the response of resistant clones to the physiological disorder of *Eucalyptus*.

Keywords: *Eucalyptus*; Hybrids; Taxonomy; Genotyping; Gene Expression.

SUMÁRIO

PRIMEIRA PARTE
INTRODUÇÃO GERAL

INTRODUÇÃO

Eucalyptus spp. é um gênero arbóreo economicamente importante. Com cerca de 700 espécies, este gênero é atualmente plantado em 95 países ao redor do mundo (ZHANG; WANG, 2021). No Brasil, a eucaliptocultura ocupa cerca de 8 milhões de hectares plantados (SERVIÇO FLORESTAL BRASILEIRO, 2019). O país se destaca em plantios florestais em âmbito mundial, pois detêm grande parte da tecnologia de silvicultura do eucalipto (ASSIS; ABAD; AGUIAR, 2015). A eucaliptocultura no Brasil atinge as mais altas produtividades florestais do mundo, e atende demandas por produtos madeireiros e não madeireiros nos mercados interno e externo.

Características favoráveis como o rápido crescimento, adaptabilidade a diferentes ecossistemas, resistência a pragas e doenças, boa densidade da madeira e alto rendimento na fabricação de papel e celulose fazem com que haja uma grande aceitação de plantios florestais com o gênero *Eucalyptus* (ASSIS; ABAD; AGUIAR, 2015). A possibilidade de hibridação interespecífica e clonagem de espécies de eucalipto são outras importantes características para o avanço dos programas de melhoramento da cultura (GRATTAPAGLIA; KIRST, 2008). Programas de melhoramento de eucalipto exploram as características complementares de pelo menos 10 espécies do gênero, os quais desenvolvem novos híbridos utilizando as características complementares das espécies (MPHAHLELE et al., 2020; JONES et al., 2016; (GRATTAPAGLIA; KIRST, 2008). Um exemplo de sucesso da técnica de hibridação interespecífica é o híbrido entre *E. grandis* e *E. urophylla,* atualmente, o mais plantado pelas empresas produtoras de papel e celulose no Brasil.

Devido a importância da eucaliptocultura, empresas de produtos florestais e centros de pesquisa buscam tecnologias para o aumento de produtividade e redução do tempo do ciclo de melhoramento. Estudos de indução de florescimento (DE OLIVEIRA CASTRO et al., 2021), seleção precoce (FRIAS et al., 2020), hibridação (ASSIS; BAUER; TAFAREL, 1993), investigação de problemas fitossanitários (FERNANDES SANTOS et al., 2022), redução da interação genótipos x ambientes (DE OLIVEIRA et al., 2023) entre outros trabalhos auxiliam o melhoramento do eucalipto. Novas abordagens como a utilização de *Big Data* genômicos são exemplos de novas estratégias utilizadas o melhoramento da espécie (XU et al., 2023). Estudos de associação genética ampla (GWAS), seleção genômica (GS) (GRATTAPAGLIA, 2022; MPHAHLELE et al., 2020), e multiômicas são exemplos de abordagens genômicas consideradas promissoras no melhoramento do eucalipto (XU et al., 2023; WANG et al., 2018).

A genômica florestal pode ser associada a estudos com diferentes objetivos (BORTHAKUR et al., 2022). Na eucaliptocultura a genômica possui o objetivo de melhorar genótipos de forma mais rápida que a utilização de métodos convencionais (DE OLIVEIRA CASTRO et al., 2021). A utilização de marcadores moleculares, como por exemplo os marcadores SNPs, são amplamente utilizados nos estudos de melhoramento molecular. Contudo, outras abordagens, como a biologia de sistemas e genômica computacional, permitem que a identificação de genes que regulam características complexas, além do entendimento de mecanismos de adaptação a estresses bióticos e abióticos. Estes estudos podem ser pontos de partida para a aplicação de novas técnicas da edição gênica para a solução de desafios (BORTHAKUR et al., 2022).

Neste documento são apresentadas duas abordagens genômicas. Na primeira foram genotipados milhares de SNPs, de forma ampla no genoma de espécies de eucalipto e híbridos interespecíficos com o objetivo de avaliar se esses marcadores conseguem separar e identificar as espécies, bem como a composição genômica dos híbridos interespecíficos. Com isso, foi possível comparar a composição esperada, conforme o pedigree anotado pelos melhoristas, com a aquela identificada de forma objetiva pelas análises genômicas com 27K SNPs. O estudo indica que os SNPs podem orientar os melhoristas de eucalipto na escolha de genitores e potenciais clones já que esses marcadores conseguem estimar a composição genômica das árvores. Informação relevante na condução dos programas de melhoramento de eucalipto.

O segundo estudo utilizou o RNA-seq para identificar genes e vias metabólicas associados ao Distúrbio Fisiológico do eucalipto. O objetivo foi iluminar as possíveis causas dessa doença que ainda possui agente etiológico desconhecido. Alguns pesquisadores citam como causas dessas desordens as mudanças climáticas, efeitos nutricionais e alguns fatores genéticos envolvidos (CÂMARA et al., 2018). Neste sentido, acredita-se que análises genômicas de grande eficiência possam ser úteis na elucidação dessa desordem. A técnica de RNA-Seq permite avaliar a expressão de todos os genes do eucalipto em qualquer parte da planta e em qualquer condição ambiental. Com isso, é possível comparar a expressão dos genes em clones de *Eucalyptus* historicamente resistentes e suscetíveis a doença (WANG; GERSTEIN; SNYDER, 2009; HAN et al., 2015). A obtenção de genes diferencialmente expressos em condições e épocas que favoreçam ou não o desenvolvimento da enfermidade, especialmente no início do distúrbio, pode contribuir para a identificação das vias metabólicas e dos processos fisiológicos afetados pela doença. Neste estudo, foi possível observar na comparação do metabolismo de clones suscetíveis e resistentes, alterações nas vias de

sinalização, incremento da fermentação, produção de metabólitos secundários, além de genes possivelmente relacionados com a resistência e/ou suscetibilidade dos clones.

Portanto, neste estudo duas técnicas genômicas foram utilizadas em diferentes contextos na eucaliptocultura. Estas foram importantes para ajudar nos desafios e gerar conhecimento para a eucalipto cultura. Os marcadores SNPs apresentaram um alto potencial de identificação de espécies e híbridos da espécie sendo um estudo base para o desenvolvimento de novas tecnologias. No mesmo caminho, a técnica de RNA-seq foi capaz de mostrar as diferenças metabólicas importantes entre clones suscetíveis e resistentes de eucalipto ao distúrbio fisiológico. A identificação destas rotas metabólicas é importante foi importante para aumentar o entendimento desta doença complexa.

REFERÊNCIAS

ASSIS, T. F.; ABAD, J. I. M.; AGUIAR, A. M. Melhoramento Genético do Eucalipto. Em: SCHUMACHER, M. V.; VIEIRA, M. (Ed.). **Silvicultura do eucalipto no Brasil**. Santa Maria-RS: UFSM, 2015. p. 225–247. 2015.

ASSIS, T. F.; BAUER, J. F. S.; TAFAREL, G. SINTETIZAÇÃO DE HÍBRIDOS DE *Eucalyptus* POR CRUZAMENTOS CONTROLADOS. **Ciência florestal**, v. 3, n. 1, p. 161–170, 1993.

BORTHAKUR, D.; BUSOV, V.; CAO, X. H.; DU, Q.; GAILING, O.; ISIK, F.; KO, J.-H.; LI, C.; LI, Q.; NIU, S.; QU, G.; VU, T. H. G.; WANG, X.-R.; WEI, Z.; ZHANG, L.; WEI, H. Current status and trends in forest genomics. **Forestry Research**, v. 2, n. 1, p. 0–0, 2022. doi: 10.48130/fr-2022-0011.

CÂMARA, A. P.; OLIVEIRA, J. T. S.; BOBADILHA, G. S.; VIDAURRE, G. B.; TOMAZELLO FILHO, M. SOLIMAN, E. P. Physiological disorders affecting dendrometric parameters and eucalyptus wood quality for pulping wood. **CERNE**, v. 24, n. 1, p. 27–34, 2018. doi: 10.1590/01047760201824012480.

DE OLIVEIRA, A. F. C. F.; LIMA, J. L.; NOVAES, E.; CARNEIRO, V. Q.; RAMALHO, M. A. P. Clonal composites as a strategy for mitigating the clones × environments interaction in *Eucalyptus*. **Cerne**, v. 29, n. 1, 2023. doi: 10.1590/01047760202329013122.

DE OLIVEIRA CASTRO, C. A.; DOS SANTOS, G. A.; TAKAHASHI, E. K.; PIRES NUNES, A. C.; SOUZA, G. A.; DE RESENDE, M. D. V. Accelerating *Eucalyptus* breeding strategies through top grafting applied to young seedlings. **Industrial Crops and Products**, v. 171, 1 nov. 2021. doi: 10.1016/j.indcrop.2021.113906.

FERNANDES SANTOS, C. A.; DA COSTA, S. R.; BOITEUX, L. S.; GRATTAPAGLIA, D.; SILVA-JUNIOR, O. B. Genetic associations with resistance to *Meloidogyne enterolobii* in guava (Psidium sp.) using cross-genera SNPs and comparative genomics to *Eucalyptus* highlight evolutionary conservation across the Myrtaceae. **PLoS ONE**, v. 17, n. 11 nov. 2022. doi: 10.1371/journal.pone.0273959.

FRIAS, Y. A.; FERREIRA, E. A.; CRUZ, V. H.; ALVES, D. P.; PRADO, E. P.; LIMA, R. C.; CRUZ, C.

D.; TOMAZ, R. S. Early selection efficiency for recommendation of *Eucalyptus* sp. **International Journal for Innovation Education and Research**, v. 8, n. 3, p. 304–315, 1 mar. 2020. doi: 10.31686/ijier.vol8.iss3.2231.

GRATTAPAGLIA, D. Twelve Years into Genomic Selection in Forest Trees: Climbing the Slope of Enlightenment of Marker Assisted Tree Breeding. **Forests**, 1 out. 2022. doi: 10.3390/f13101554.

GRATTAPAGLIA, D.; KIRST, M. *Eucalyptus* applied genomics: From gene sequences to breeding tools. **New Phytologist**, v. 179, n. 4, p. 911–929, 2008. doi: 10.1111/j.1469-8137.2008.02503.x.

HAN, Y.; GAO, S.; MUEGGE, K.; ZHANG, W.; ZHOU, B. Advanced applications of RNA sequencing and challenges. **Bioinformatics and Biology Insights**, v. 9, p. 29–46, 2015. doi: 10.4137/BBI.S28991.

JONES, R. C.; NICOLLE, D.; STEANE, D. A.; VAILLANCOURT, R. E.; POTTS, B. M. High density, genome-wide markers and intra-specific replication yield an unprecedented phylogenetic reconstruction of a globally significant, speciose lineage of Eucalyptus. **Molecular Phylogenetics and Evolution**, v. 105, p. 63–85, 1 dez. 2016. doi: 10.1016/j.ympev.2016.08.009.

MPHAHLELE, M. M.; ISIK, F.; MOSTERT-O'NEILL, M. M.; REYNOLDS, S. M.; HODGE, G. R.; MYBURG, A. A. Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*. **Tree Genetics and Genomes**, v. 16, n. 4, 1 ago. 2020. doi: 10.1007/s11295-020-01443-1.

SERVIÇO FLORESTAL BRASILEIRO, 2019. Disponível em <http://www.florestal.gov.br/>

WANG, J. P.; MATTHEWS, M. L.; WILLIAMS, C. M.; SHI, R.; YANG, C.; TUNLAYA-ANUKIT, S.; CHEN, H. C.; LI, Q.; LIU, J.; LIN, C. Y.; NAIK, P.; SUN, Y. H.; LOZIUK, P. L.; YEH, T. F.; KIM, H.; GJERSING, E.; SHOLLENBERGER, T.; SHUFORD, C. M.; SONG, J.; MILLER, Z.; HUANG, Y. Y.; EDMUNDS, C. W.; LIU, B.; SUN, Y.; LIN, Y. C. J.; LI, W.; CHEN, H.; PESZLEN, I.; DUCOSTE, J. J.; RALPH, J.; CHANG, H. M.; MUDDIMAN, D. C.; DAVIS, M. F.; SMITH, C.; ISIK, F.; SEDEROFF, R.; CHIANG, V. L. Improving wood properties for wood utilization through multi-omics integration in lignin biosynthesis. **Nature Communications**, v. 9, n. 1, 1 dez. 2018. doi: 10.1038/s41467-018-03863-z.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq : a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, n. 1, p. 57–63, 2009. doi: 10.1038/nrg2484.RNA-Seq.

XU, T.; LIU, Z.; ZHAN, D.; PANG, Z.; ZHANG, S.; LI, C.; KANG, X.; YANG, J. Integrated transcriptomic and metabolomic analysis reveals the effects of polyploidization on the lignin content and metabolic pathway in *Eucalyptus*. **Biotechnology for Biofuels and Bioproducts**, v. 16, n. 1, 1 dez. 2023. doi: 10.1186/s13068-023-02366-4.

ZHANG, Y. X.; WANG, X. J. Geographical spatial distribution and productivity dynamic change of *Eucalyptus* plantations in China. **Scientific Reports**, v. 11, n. 1, 1 dez. 2021. doi: 10.1038/s41598-021-97089-7.

SEGUNDA PARTE – ARTIGOS

ARTIGO 1 - GENOME-WIDE ANALYSIS HIGHLIGHTS GENETIC ADMIXTURE IN EXOTIC GERMPLASM RESOURCES OF *EUCALYPTUS* AND UNEXPECTED ANCESTRAL GENOMIC COMPOSITION OF INTERSPECIFIC HYBRIDS

Artigo publicado na revista Plos One

PLOS ONE

RESEARCH ARTICLE

Genome-wide analysis highlights genetic admixture in exotic germplasm resources of *Eucalyptus* and unexpected ancestral genomic composition of interspecific hybrids

Danyllo Amaral de Oliveira[1], Paulo Henrique Muller da Silva[2], Evandro Novaes[1], Dario Grattapaglia[3]*

1 Departamento de Biologia, Universidade Federal de Lavras, Lavras, MG, Brazil, **2** Instituto de Pesquisas e Estudos Florestais, Piracicaba, SP, Brazil, **3** Plant Genetics Laboratory, EMBRAPA Genetic Resources and Biotechnology, Brasilia, DF, Brazil

* dario.grattapaglia@embrapa.br

Check for updates

OPEN ACCESS

Citation: de Oliveira DA, da Silva PHM, Novaes E, Grattapaglia D (2023) Genome-wide analysis highlights genetic admixture in exotic germplasm resources of *Eucalyptus* and unexpected ancestral genomic composition of interspecific hybrids. PLoS ONE 18(8): e0289536. https://doi.org/10.1371/journal.pone.0289536

Editor: Chunxian Chen, USDA/ARS, UNITED STATES

Received: May 5, 2023

Accepted: July 20, 2023

Published: August 8, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0289536

Copyright: © 2023 de Oliveira et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Abstract

Eucalyptus is an economically important genus comprising more than 890 species in different subgenera and sections. Approximately twenty species of subgenus *Symphyomyrtus* account for 95% of the world's planted eucalypts. Discrimination of closely related eucalypt taxa is challenging, consistent with their recent phylogenetic divergence and occasional hybridization in nature. Admixture, misclassification or mislabeling of *Eucalyptus* germplasm resources maintained as exotics have been suggested, although no reports are available. Moreover, hybrids with increased productivity and traits complementarity are planted worldwide, but little is known about their actual genomic ancestry. In this study we examined a set of 440 trees of 16 different *Eucalyptus* species and 44 interspecific hybrids of multi-species origin conserved in germplasm banks in Brazil. We used genome-wide SNP data to evaluate the agreement between the alleged phylogenetic classification of species and provenances as registered in their historical records, and their observed genetic clustering derived from SNP data. Genetic structure analyses correctly assigned each of the 16 species to a different cluster although the PCA positioning of *E. longirostrata* was inconsistent with its current taxonomy. Admixture was present for closely related species' materials derived from local germplasm banks, indicating unintended hybridization following germplasm introduction. Provenances could be discriminated for some species, indicating that SNP-based discrimination was directly proportional to geographical distance, consistent with an isolation-by-distance model. SNP-based genomic ancestry analysis showed that the majority of the hybrids displayed realized genomic composition deviating from the expected ones based on their pedigree records, consistent with admixture in their parents and pervasive genome-wide directional selection toward the fast-growing *E. grandis* genome. SNP data in support of tree breeding provide precise germplasm identity verification, and allow breeders to objectively recognize the actual ancestral origin of superior hybrids to more realistically guide the program toward the development of the desired genetic combinations.

Funding: This work was supported by FAPESP (Foundation for Scientific Research of São Paulo State) competitive grant number 2017/24609-5 to PHMS, FAP-DF (Foundation for Scientific Research of the Federal District) competitive grant RECGENOMICS 00193-00000924/2021-92 to DG and CNPq (Brazilian National Council for Scientific and Technological Development) research productivity fellowship to PHMS, EN and DG. There was no additional external funding received for this study and the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Eucalyptus is a highly diverse genus of tree species from Australia and neighboring islands ruling the forest landscape across large part of Australia and neighboring islands. Besides its keystone ecological role in native forests, the genus includes the most widely commercially planted hardwood tree species in tropical and subtropical regions of the world [1, 2]. Fast growth, wide adaptability to a globally broad diversity of tropical and subtropical environments, combined with multipurpose wood properties for energy, solid wood products, pulp and paper, have secured the superior position of the eucalypts in current world forestry [3].

As particularly speciose, the genus *Eucalyptus* has received significant attention to its phylogenetic organization. The classic taxonomy of eucalypts with more than 700 species [4], has now been expanded to include over 890 species [5], but taxonomic issues still remain as species delimitations are still being actively investigated. The genus was originally subdivided in 13 subgenera but two of them, *Angophora* and *Corymbia*, were recently recognized as separate genera based on molecular evidences [6, 7]. The largest subgenus *Symphyomyrtus* includes 470 species organized in 15 sections which are mostly well separated using molecular marker data [8, 9]. Nevertheless, some discrepancies exist between DNA marker and morphology-based classifications at the lower taxonomic level, consistent with recent divergence of taxa, characters convergence and occasional hybridization in natural populations, generating hybrid swarms and zones of intergradation between species of the same section in the wild [10–13].

Most of the economically important eucalypts species belong to subgenus *Symphyomyrtus*, and, within it, to three specific sections, *Exsertaria*, *Latoangulatae* and *Maidenaria* [14]. These sections include approximately twenty species that account for more than 95% of the world's planted eucalypts. Among those, *Eucalyptus grandis*, *E. urophylla* (sect. *Latoangulatae*), *E. camaldulensis* (sect. *Exsertaria*) and *E. globulus* (sect. *Maidenaria*) make up more than 80% of the planted areas [15]. The sexual compatibility among species within and across these three main sections have been important drivers of breeding programs especially in tropical and subtropical countries, where large extensions of forests are planted with hybrid material [10, 12, 16, 17]. Hybrid breeding coupled to clonal propagation has allowed the aggregation and exploitation of important characteristics from different species and provenances in highly productive hybrid clones [18, 19]. In Brazil, an estimated 80% of eucalypts plantations are established with first- or second-generation hybrids involving mainly *E. urophylla* and *E. grandis* [17, 20]. The remaining 20% are mostly pure species material of the two former species, or hybrids with *E. camaldulensis* and *E. pellita* for greater drought tolerance, and to a much lesser extent with *E. globulus* to improve wood quality for cellulose. Additionally, *E. dunnii* and *E. benthamii* are also used as pure species or in hybrid combinations in areas subject to frost [21].

A number of studies in the last fifteen years have approached and largely established the challenge of resolving lower-level, within section taxonomy in *Eucalyptus* using different genome-wide DNA marker data [8, 9, 22–24]. However, issues remain for species in section *Latoangulatae*, for example, due to their intermediate nature, when compared to more densely clustered taxa in other sections [8]. Admixture of *Eucalyptus* species in their native range has been reported [11, 25], reflecting the phylogenetic fluidity that still exist in some taxa. However, misclassification, mislabeling and hybridization of eucalypts germplasm resources maintained as exotics in different countries has been suggested, but reports are only anecdotal. Little or no data exist on such incidences, or on the overall current status of gene banks in countries where eucalypt make up a large proportion of planted forests.

Genome-wide studies looking at large numbers of eucalypt species have used DArT markers, genotyped originally with probe arrays [26] and, more recently, by genome complexity reduction with restriction enzyme digestion followed by high-throughput sequencing [27].

This DNA assay has been valuable as it provides simultaneous discovery and genotyping of Single Nucleotide Polymorphisms (SNP) within and across species, facilitating genus-wide phylogenetic studies. However, some challenges remain for this SNP genotyping method due to variable sequencing coverage and irregular sampling of loci causing variable genotype reproducibility and ultimately limited data portability across studies in highly heterozygous genomes such as those of the eucalypts [28–30]. The development of *Eucalyptus* multispecies SNP arrays based on industry-level "gold standard" technology has provided a worldwide usable platform allowing seamless and precise data exchange across studies [31].

Although species discrimination using DNA data is largely settled, less attention has been devoted to looking at provenance variation within species. This is particularly important for breeding programs that take advantage of matching distinctive provenance characteristics to specific sites in exotic environments, or aim at deliberately exploiting provenance and species complementarity by building specific genomic compositions by interspecific hybridization [2, 12, 18]. Likewise, few studies have examined the possibility of using DNA data to describe the actual genomic ancestral composition of hybrids, including those derived from more than two parental species. Knowledge of the actual genomic composition of complex hybrids of distinctive performance would allow directing more deliberate selection strategies in hybrid breeding programs. Earlier studies using microsatellite markers indicated that provenances of *E*. *grandis* could be distinguished but not for *E*. *urophylla* and *E*. *camaldulensis*, and some hybrid clones could be assigned to their most likely ancestral species, although with incomplete resolution [32]. Using SNP data, preliminary analyses have shown that provenances within species could be distinguished for *E*. *grandis* and *E*. *urophylla* [31, 33] but not for *E*. *camaldulensis*, consistent with the latter being more prone to hybridization or a remnant of an ancient widespread taxon [8].

The current eucalypt SNP arrays have been used to estimate recombination rates and carry out dense linkage mapping [34], build relationship matrices for genomic selection in several species, reviewed in [35], and understand the consequences of artificial selection [36]. No studies to date, however, have evaluated their ability to characterize germplasm material in gene banks. Questions frequently arise regarding the verification of the alleged species classification, the possibility of discriminating provenances and determining the genomic composition of hybrid clones of unknown or uncertain origin derived from successive generations of interspecific recombination. In this study we examined a large set of germplasm accessions including 440 *Eucalyptus* trees of 16 species and 44 interspecific hybrids currently conserved or used in Brazil. We used genome-wide SNP data to evaluate the agreement between the alleged phylogenetic classification of species and provenances as registered in their historical records, and their observed genetic clustering obtained from genomic data, agnostic to any prior phylogenetic information. We focused on the main planted species of *Symphyomyrtus* given their outstanding relevance in terms of germplasm use and conservation. Additionally, we used SNP data to examine the actual genomic makeups of hybrids derived from interspecific crosses involving two of more species, and compare them with their expected composition based on the recorded ancestral species.

Material and methods

Plant material

The study involved a germplasm set of 440 trees belonging to 16 *Eucalyptus* species of five sections of subgenus *Symphyomyrtus* and 44 interspecific hybrid clones (Table 1). These trees are conserved in species/provenance/progeny trials and clonal banks at the Anhembi Experimental Research Station of the Institute for Forestry Research (IPEF) in Brazil (22.7897˚ S,

Table 1. Section, provenances, source and number of individual trees sampled for each of the 16 *Eucalyptus* species included in the study.

Species	Section	Provenance	Source*	Number of individuals
E. argophloia	Adnataria	Narromine	CSIRO 12716	16
E. deglupta	Equatoria	Philipines	CSIRO 21163	18
E. brassiana	Exsertaria	Cape York Peninsula	CSIRO 1188	18
E. camaldulensis	Exsertaria	Nott's Crossing Katherine River	IPEF—Brazil	26
E. tereticornis	Exsertaria	Unknown	Suzano—Brazil	7
E. tereticornis	Exsertaria	MT Garnet	CSIRO 20768	10
E. tereticornis	Exsertaria	N of Mareeba	CSIRO 15370	12
E. tereticornis	Exsertaria	Mitchell R Oaky CK	CSIRO 16645	12
E. grandis	Latoangulata	Atherton	Suzano—Brazil	16
E. grandis	Latoangulata	Coffs Harbour	Suzano—Brazil	20
E. longirostrata	Latoangulata	Starkvale Creek	CSIRO 20007	12
E. longirostrata	Latoangulata	Coominglah SF	CSIRO 19312	12
E. longirostrata	Latoangulata	Goodger	CSIRO 20943	12
E. pellita	Latoangulata	West of Hopevale	CSIRO 21018	16
E. pellita	Latoangulata	NW of Kuranda	CSIRO 17859	15
E. robusta	Latoangulata	David Low Wy Brisbane	CSIRO 17004	16
E. robusta	Latoangulata	Byfield SF	CSIRO 15945	14
E. saligna	Latoangulata	20 km N Helidon	CSIRO 16942	12
E. saligna	Latoangulata	Kroombit Tops	CSIRO 20483	12
E. saligna	Latoangulata	Richmond Range	CSIRO 20483	12
E. urophylla	Latoangulata	Multiple provenances	IPEF—Brazil	42
E. benthamii	Maidenaria	Kedumba Valley	Klabin- Brazil	24
E. dunnii	Maidenaria	Unknown	CMPC—Brazil	28
E. globulus	Maidenaria	Unknown	CMPC—Brazil	26
E. nitens	Maidenaria	Unknown	Klabin—Brazil	20
E. viminalis	Maidenaria	Unknown	Klabin—Brazil	12
Total				440

* The CSIRO seed lots were sampled in native populations from Australia. The Brazilian germplasm, maintained as exotic resources, are indicated by the institution or forest company where the germplasm is conserved or utilized for breeding. These Brazilian germplasm resources are at least one generation removed from the original seedlots introductions.

https://doi.org/10.1371/journal.pone.0289536.t001

48.1280˚ W), or in the gene banks of some associated forest-based companies. For six species (*E. grandis*, *E. longirostrata*, *E. pellita*, *E. robusta*, *E. saligna* and *E. tereticornis*), samples were analyzed for more than one provenance. The original locations of the species and provenances were plotted on top of a base map of world country boundaries shapefile of Australia, publicly available under a Creative Commons Attribution 4.0 International Public License (https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Boundaries) using the R package tmap (Fig 1). Plant material included: (1) individual trees sampled in species/provenance trials established with original seeds collected in Australia for which the CSIRO (Commonwealth Scientific and Industrial Research Organisation) seedlot number is known, and (2) individual trees collected in Brazilian germplasm banks at least one generation removed from the original introductions, maintained by IPEF or by three associated forestry companies (Suzano, Klabin, Vallourec and CMPC Celulose Riograndense), sometimes with unknown provenance origin (Table 1). In addition, 44 interspecific hybrid clones obtained by controlled interspecific crosses of two or more species were studied to compare their SNP-realized versus pedigree-expected genomic composition. These hybrids were grouped into five

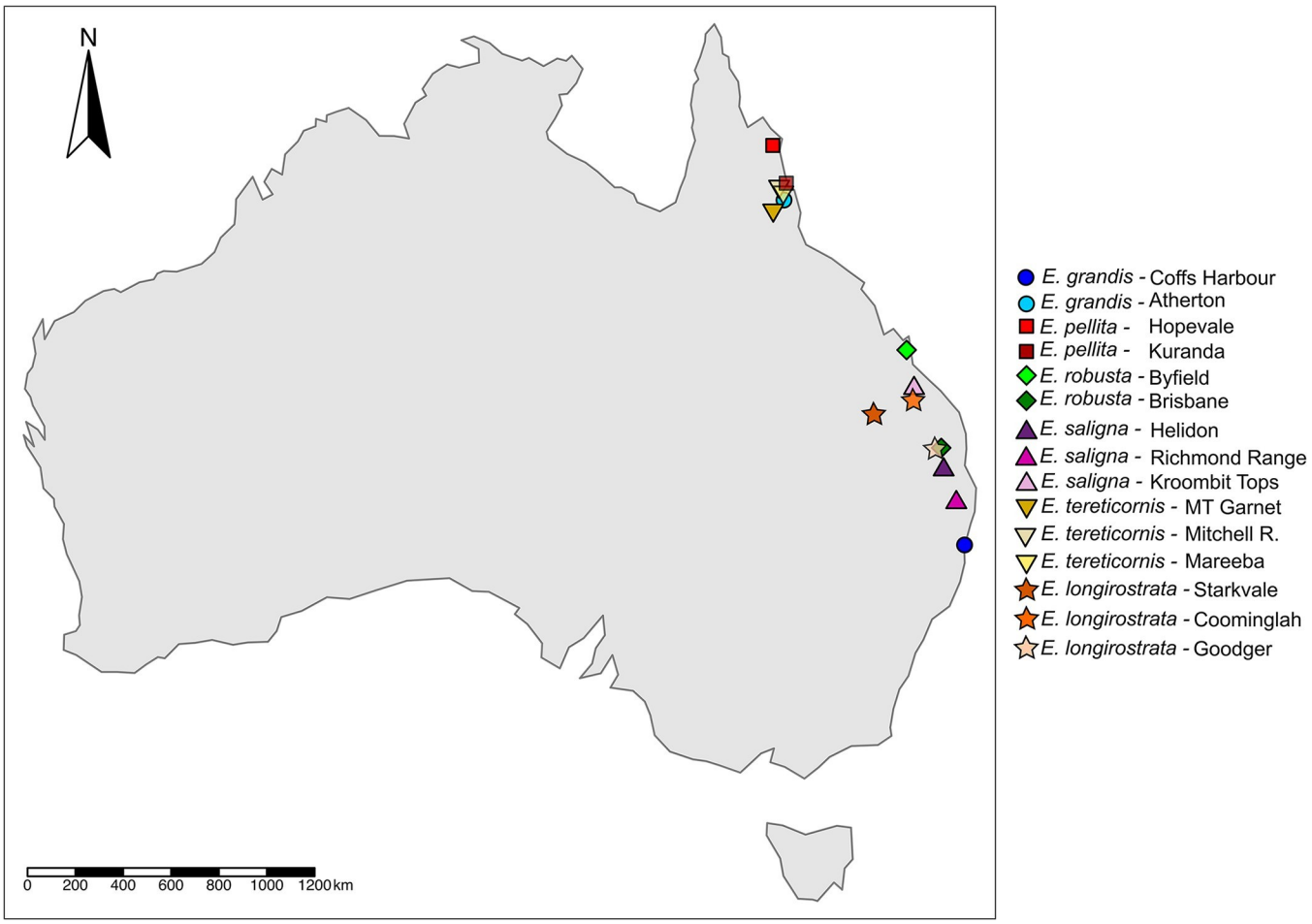

Fig 1. *Eucalyptus* species for which provenances were studied, plotted on their respective geographic locations on a publicly available basemap reprinted under a CC BY 4.0 license with permission from The World Bank (https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Boundaries).

https://doi.org/10.1371/journal.pone.0289536.g001

classes (Hybrids 1 to 5) according to the *Symphyomyrtus* sections of the component *Eucalyptus* species registered in their pedigrees (Table 2).

SNP genotyping and filtering

Total genomic DNA was extracted with an optimized Sorbitol/CTAB protocol [37]. DNA samples were sent to ThermoFisher (Santa Clara, CA) for SNP genotyping with the 72K Eucalyptus Axiom Array developed for *Eucalyptus* and *Corymbia* species (https://www.thermofisher.com/order/catalog/product/551134; Grattapaglia D. and Silva-Junior O.B., unpublished). This Axiom Array is a second-generation Eucalyptus SNP platform with 68,055 SNPs specific to the *Eucalyptus* genome, 28,177 of them shared with the previously developed Infinium EUChip60k [31], and 4,147 specific to the genome of its sister genus *Corymbia*, these latter ones not used in this study.

SNPs with more than 10% missing data and with minor allele frequency (MAF) below 5% were removed using PLINK v1.9 [38] using parameters–maf 0.05 –geno 0.1. A total of 48,645 SNPs passed these filtering thresholds. The dataset was further pruned of 21,347 SNPs that were in linkage disequilibrium (LD) with other markers to remove redundant information

Table 2. List of the 44 *Eucalyptus* hybrids studied, obtained by controlled interspecific crosses of two or more species, classified into five groups according to the sections of *Symphyomyrtus* involved in the cross (Hybrids 1 –Hybrids 5).

Hybrid	Sections involved	Individual ID	Parental species crossed
Hybrids 1	*Latoangulatae x Exsertaria*	*Hyb-1*	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-2	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-3	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-4	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-5	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-6	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-7	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-8	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-9	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-10	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-11	*E. urophylla x E. camaldulensis*
Hybrids 1		Hyb-12	*E. camaldulensis x E. grandis*
Hybrids 2	*Latoangulatae x Maidenaria*	Hyb-13	*E. dunnii x E. urophylla*
Hybrids 2		Hyb-14	*(E. dunnii x E. grandis) x E. dunnii*
Hybrids 2		Hyb-15	*(E. grandis x E. saligna) x (E. urophylla x E. globulus)*
Hybrids 2		Hyb-16	*((E. grandis x E. urophylla) x (E. grandis x E.globulus)) x ((E. dunnii x E. grandis) x (E. urophylla x E. globulus))*
Hybrids 2		Hyb-17	*E. dunnii x E. urophylla*
Hybrids 2		Hyb-18	*E. dunnii x E. urophylla*
Hybrids 2		Hyb-19	*E. dunnii x E. urophylla*
Hybrids 2		Hyb-20	*(E. grandis x E. urophylla) x (E. urophylla x E. globulus)*
Hybrids 2		Hyb-21	*E. grandis x E. globulus*
Hybrids 2		Hyb-22	*(E. urophylla x E. globulus) x (E. dunnii x E. grandis)*
Hybrids 2		Hyb-23	*E. saligna x (E. grandis x E. globulus)*
Hybrids 2		Hyb-24	*E. urophylla x (E. grandis x E. globulus)*
Hybrids 2		Hyb-25	*(E. dunnii x E. grandis) x E.globulus*
Hybrids 2		Hyb-26	*E. saligna x E. globulus*
Hybrids 2		Hyb-27	*((E. grandis x E. urophylla) x (E. grandis x E. globulus)) x E. benthamii*
Hybrids 2		Hyb-28	*(E. urophylla x E. grandis) x ((E. dunnii x E. grandis) x (E. urophylla x E. globulus))*
Hybrids 2		Hyb-29	*(E. urophylla x E. grandis) x E. benthamii*
Hybrids 2		Hyb-30	*E. grandis x E. benthamii*
Hybrids 2		Hyb-31	*(E. urophylla x E. grandis) x E. globulus*
Hybrids 2		Hyb-32	*E. urophylla x E. benthamii*
Hybrids 2		Hyb-33	*E. urophylla x E. dunnii*
Hybrids 2		Hyb-34	*E. urophylla x E. dunnii*
Hybrids 2		Hyb-35	*E. urophylla x E. globulus*
Hybrids 2		Hyb-36	*E. urophylla x E. globulus*
Hybrids 3	*Latoangulatae*	Hyb-37	*E. grandis x E. urophylla*
Hybrids 3		Hyb-38	*E. grandis x E. pellita*
Hybrids 3		Hyb-39	*E. urophylla x E. grandis*
Hybrids 3		Hyb-40	*E. urophylla x E. grandis*
Hybrids 3		Hyb-41	*E. urophylla x E. saligna*
Hybrids 4	[a]	Hyb-42	*E. dunnii x E. globulus*
Hybrids 5	[b]	Hyb-43	*(E. dunnii x E. grandis) x E. camaldulensis*
Hybrids 5		Hyb-44	*E. camaldulensis x (E. urophylla x E. globulus)*

[a] *Maidenaria x Maidenaria;*

[b] *Exsertaria x Latoangulatae x Maidenaria*

https://doi.org/10.1371/journal.pone.0289536.t002

and avoid regions of the genome with a disproportionate influence on the results, that could potentially distort the representation of genome-wide structure [39]. LD pruning was performed using PLINK parameter–indep-pairwise 50 5 0.3. With the retained 27,298 SNPs, the rate of per individual missing data was below 10% for all samples, except for one sample of *E. grandis* from Coffs Harbor. This sample had 64.9% missing genotypes and was removed from further analyses. Ultimately, genetic analyses were performed with a dataset of 27,298 SNPs genotyped in 484 individual trees.

Statistical and population genetics analyses

Basic population genetics parameters were estimated, such as the average minor allele frequency (MAF), observed (H_o) and expected heterozygosity (H_e). Analyses were performed in R (R Core Team 2020) using packages adegenet v2.1.3 [40] and hierfstat v 0.5–7 [41]. The data was input into R in FSTAT format after transformation with PGDSpider v2.1.1.5 [42].

fastSTRUCTURE v. 1.0 [43] was run with the 27,298 SNPs to infer population structure for the 484 individual trees. Analyses were performed with the number of clusters K varying from 2 to 30 and option—seed = 100. The input was the binary version (BED) of the PED file from PLINK. The most likely model was selected using the supervised estimators of [44] implemented in the StructureSelector [45] web server (https://lmme.ac.cn/StructureSelector/). Cluster assignment for each of the samples was visualized with barplots in R, using packages pophelper v2.3.0 [46] and gridextra v2.3 [47]. Additional fastSTRUCTURE analyses were carried out separately for individual eucalypt sections and species to assess resolution at within taxa levels for provenances differentiation.

Genomic composition of the hybrids was initially obtained from the unsupervised inference provided by fastSTRUCTURE, and compared with the recorded pedigree information. Specifically, for fastSTRUCTURE annotation, we used the meanQ file, which provided the probabilities of each sample belonging to each of the clusters found. Subsequently, a supervised analysis was carried out using ADMIXTURE, a software for model-based estimation of ancestry in unrelated individuals [48]. For this analysis, samples defined as being from pure species with ~99% probability in the initial fastSTRUCTURE analysis, were used as reference populations to infer the genomic composition of the hybrids' genomes. A simple matching genetic distance among individual trees was also estimated and groups visualized with a principal component analysis (PCA) on the genetic distance matrix, where distances among trees were represented in a cartesian graph with PC1 and PC2. These analyses were performed in R using packages adegenet [40] ape v5.4 [49] and pegas v0.14 [50]. PCA biplots were visualized using ggplot2 v.3.3.2 package [51].

Results

SNP diversity across species

After filtering and LD pruning, the final SNP dataset of 27,298 SNPs (S1 File) had a very low percentage of missing data (<3%) for all germplasm sets (species, provenances and hybrids), corroborating the good performance of the multi-species SNP array for population genomics and molecular breeding across eucalypt taxa. The percentage of polymorphic loci per population ranged from 29.7% for *E. deglupta* to over 93% for *E. urophylla* and the hybrids involving crosses between distant sections *Latoangulatae* and *Maidenaria* (Table 3). Overall, there was no significant difference in the proportion of polymorphic SNPs among the different eucalypt sections (ANOVA F-value = 1.14, p-value = 0.37). The average MAF across taxa was similar, within 0.1 and 0.15 for most taxa but *E. urophylla*, *E. grandis*, *E. camaldulensis* and the hybrids had a slightly higher average MAF.

Table 3. Summary of the proportion of polymorphic SNPs (Minor Allele Frequency; MAF > 0.05) and their average MAF for the 27,298 filtered and LD pruned SNPs, and genetic diversity parameters (observed (H_o) e expected (H_e) heterozygosity) for each species and hybrid (see Table 2) germplasm source.

Species (Provenance)	N	% SNPs MAF > 0.05	Average MAF	Average H_e	Average H_o
E. argophloia	16	42.23%	0.104	0.136	0.188
E. benthamii	24	42.63%	0.112	0.151	0.175
E. brassiana	18	55.34%	0.124	0.167	0.167
E. camaldulensis	26	68.88%	0.151	0.205	0.200
E. deglupta	18	29.81%	0.091	0.114	0.179
E. dunnii	28	55.54%	0.119	0.161	0.166
E. globulus	26	58.34%	0.128	0.172	0.176
E. grandis (Atherton)	16	68.02%	0.162	0.219	0.214
E. grandis (Coffs Harbour)	20	79.03%	0.180	0.245	0.240
E. longirostrata (Coominglah)	12	42.90%	0.098	0.132	0.156
E. longirostrata (Goodger)	12	41.38%	0.095	0.128	0.155
E. longirostrata (Starkvale)	12	42.19%	0.096	0.130	0.150
E. nitens	20	38.94%	0.083	0.336	0.158
E. pellita (Hopevale)	16	61.56%	0.132	0.180	0.182
E. pellita (Kuranda)	15	61.28%	0.137	0.186	0.168
E. robusta (Brisbane)	16	56.76%	0.125	0.170	0.183
E. robusta (Byfield)	14	51.61%	0.120	0.161	0.174
E. saligna (Helidon)	12	59.54%	0.142	0.191	0.187
E. saligna (Kroombit Tops)	12	60.92%	0.142	0.192	0.209
E. saligna (Richmond Range)	12	60.56%	0.143	0.193	0.202
E. tereticornis	7	44.83%	0.120	0.385	0.198
E. tereticornis (Mount Garnet)	10	58.40%	0.143	0.192	0.209
E. tereticornis (Mareeba)	12	61.09%	0.146	0.196	0.209
E. tereticornis (Mitchell R.Oaky C)	12	60.10%	0.144	0.193	0.209
E. urophylla	42	93.67%	0.219	0.297	0.272
E. viminalis	12	60.16%	0.122	0.166	0.178
Hybrids 1	12	90.29%	0.225	0.304	0.307
Hybrids 2	24	93.94%	0.229	0.309	0.318
Hybrids 3	5	74.70%	0.205	0.272	0.316
Hybrids 4	1	32.43%	0.163	0.188	0.334
Hybrids 5	2	54.22%	0.186	0.231	0.333

https://doi.org/10.1371/journal.pone.0289536.t003

Population structure analysis

StructureSelector analysis of the fastSTRUCTURE results indicated the most likely model with K = 18 taxonomic clusters (S2 File). This model correctly assigned each of the 16 species to a different cluster (Fig 2; S3 File). In the case of *E. saligna* some individuals were additionally separated according to provenance and the hybrids were assembled in a separate highly admixed cluster (Fig 2). Admixture at the individual level was seen in allegedly pure species trees. Some *E. camaldulensis* individuals were classified as being admixed with *E. tereticornis*, some *E. urophylla* individuals admixed with *E. grandis*, and a few additional admixed individuals were seen that were expected pure (Fig 2). At the higher taxonomic level of sections within subgenus *Symphyomyrtus*, models with smaller numbers of clusters easily separated eucalypt sections. For example, at K = 2, section *Maidenaria* detached from the rest. With K = 3, *Latoangulatae* and *Maidenaria* split, with occasional admixture seen in individuals of some species. With K = 4, species of *Exsertaria* separated from the other sections. Surprisingly,

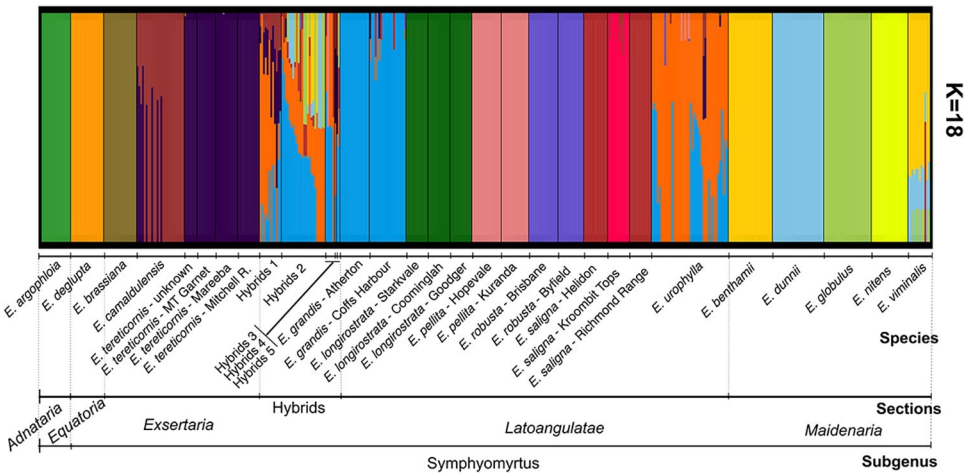

Fig 2. Population structure analysis of the 440 trees of 16 *Eucalyptus* species and 44 hybrids classified according to the section of their component species involved (Hyb 1 to 5).

https://doi.org/10.1371/journal.pone.0289536.g002

however, *E. longirostrata* that belongs to section *Exsertaria*, was clustered together with *E. deglupta* and *E. argophloia* that belong to two other different sections (S4 File).

The SNPs dataset could not differentiate most provenances within species when all individuals were analyzed together, except for *E. saligna* from Kroombit Tops (Fig 2). This provenance was assigned to a separate group from Helidon and Richmond Range provenances, which in turn were clustered together. All the provenances for the other species (*E. tereticornis*, *E. grandis*, *E. longirostrata*, *E. pellita*, *E. pilularis* and *E. robusta*) could not be discriminated even at higher K's (S5 File). Only when species were analyzed individually, fastSTRUCTURE modeling resolved some of the provenances. This was the case of the two *E. grandis* provenances from Atherton and Coffs Harbor and *E. robusta* from Brisbane and Byfield. Somewhat separate clustering was also seen for *E. longirostrata* from Starkvale, and *E. tereticornis* from Mount Garnet, although some individuals either displayed admixture or were not clustered accordingly (Fig 3). Lastly, provenances of some species clearly could not be distinguished. This occurred with *E. pellita*, *E. longirostrata* from Coominglah and Goodger and with *E. tereticornis* from Mitchell Road (Oaky Creek) and Mareeba.

Determination of ancestral species composition of hybrids

The ancestral genomic composition of hybrids estimated with both fastSTRUCTURE and ADMIXTURE were compared to their respective pedigree expected composition (Fig 4; S6 File). The supervised analysis carried out using ADMIXTURE resulted, in general, in similar genomic composition as those obtained with fastSTRUCTURE, although some differences were seen for example in Hybrids 1, where the fastSTRUCTURE model indicated the unexpected presence of *E. tereticornis* genome. Overall, there were only nine out of the 41 hybrids for which the SNP-based composition closely matched the pedigree expected one. This happened for hybrids Hyb-31, Hyb-32, Hyb-33, Hyb-34, Hy-35, Hyb-36, Hyb-38, Hyb-39, Hyb-40 e Hyb-41, almost all of them simple F_1 hybrids. For all other hybrids, small to large deviations were observed.

For a considerable number of hybrids, additional unanticipated species from those recorded in the pedigree, were observed in their composition (Fig 4). For example, hybrids Hyb-1 through Hyb-11 in the Hybrids 1 group were expected to be F_1's of *E. urophylla* and *E.*

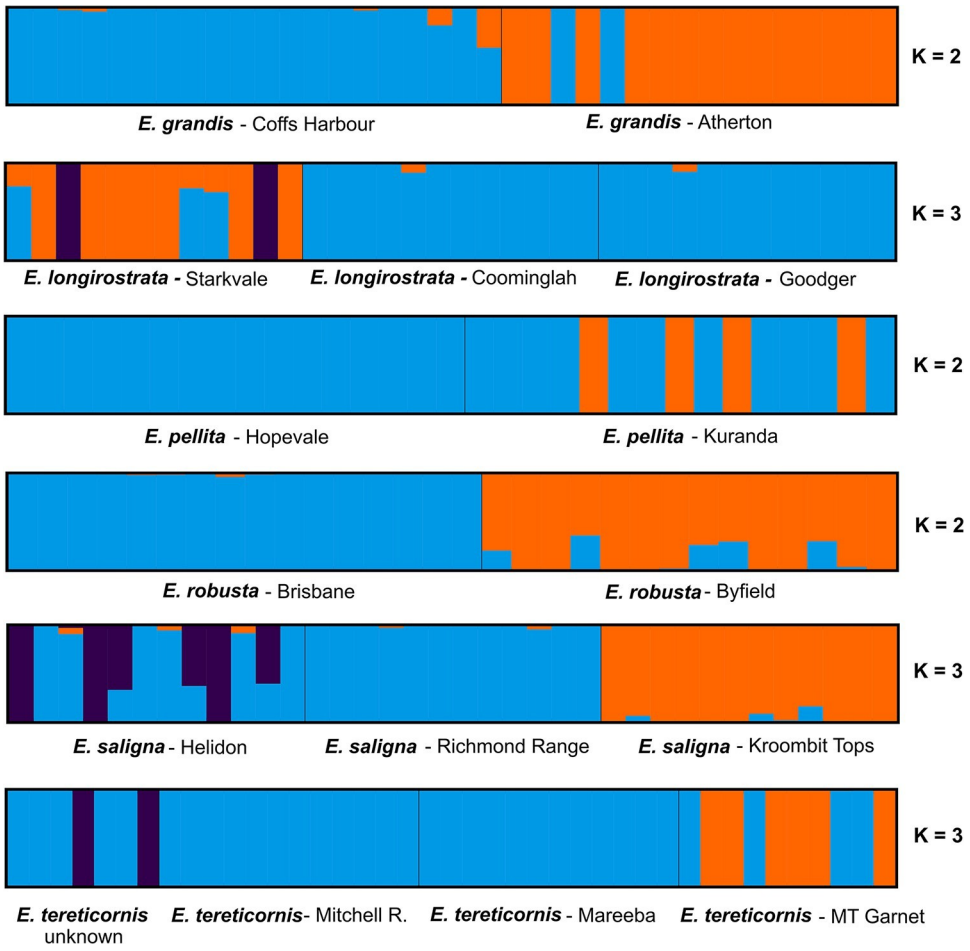

Fig 3. Population structure analysis of each *Eucalyptus* species separately for which more than one provenance was studied.

https://doi.org/10.1371/journal.pone.0289536.g003

camaldulensis. However, eight of them showed variable amounts of *E. grandis* genome in the ADMIXTURE analysis while the fastSTRUCTURE model suggested the presence of *E. tereticornis* genome more frequently than that of *E. camaldulensis*. The unexpected presence of *E. grandis* genome was again seen in several other hybrids in the Hybrids 2 group (ex. Hyb-17, Hyb-18, Hyb-19, Hyb-26). Furthermore, in this group of hybrids none or a considerably less than expected proportion of the genome was detected coming from the recorded species of *Maidenaria* involved in the crosses, namely *E. dunnii* and *E. globulus*. *E. dunnii* genome was not detected in six of the 14 hybrids and *E. globulus* in seven of 12 where it should have been observed (Fig 4). For example, in hybrids Hyb-13, Hyb-17, Hyb-18, Hyb-19, Hyb-21 and Hyb-26 expected to be F_1 hybrids of *Latoangulatae* species (*E. urophylla*, *E. grandis* or *E. saligna)* with *Maidenaria* species *(E. dunnii* or *E. globulus*), the SNP data showed little or no sign of the two temperate species genomes and an unexpected or larger than expected proportions of the genome of *E. grandis*. Finally, there were cases where the presumed genomic composition was completely different from the SNP-estimated one. For example, hybrid Hyb-42 was expected to be a *E. dunnii x E. globulus* hybrid, when in fact it involved mainly species of *Latoangulatae* with *E. camaldulensis*, suggesting mislabeling.

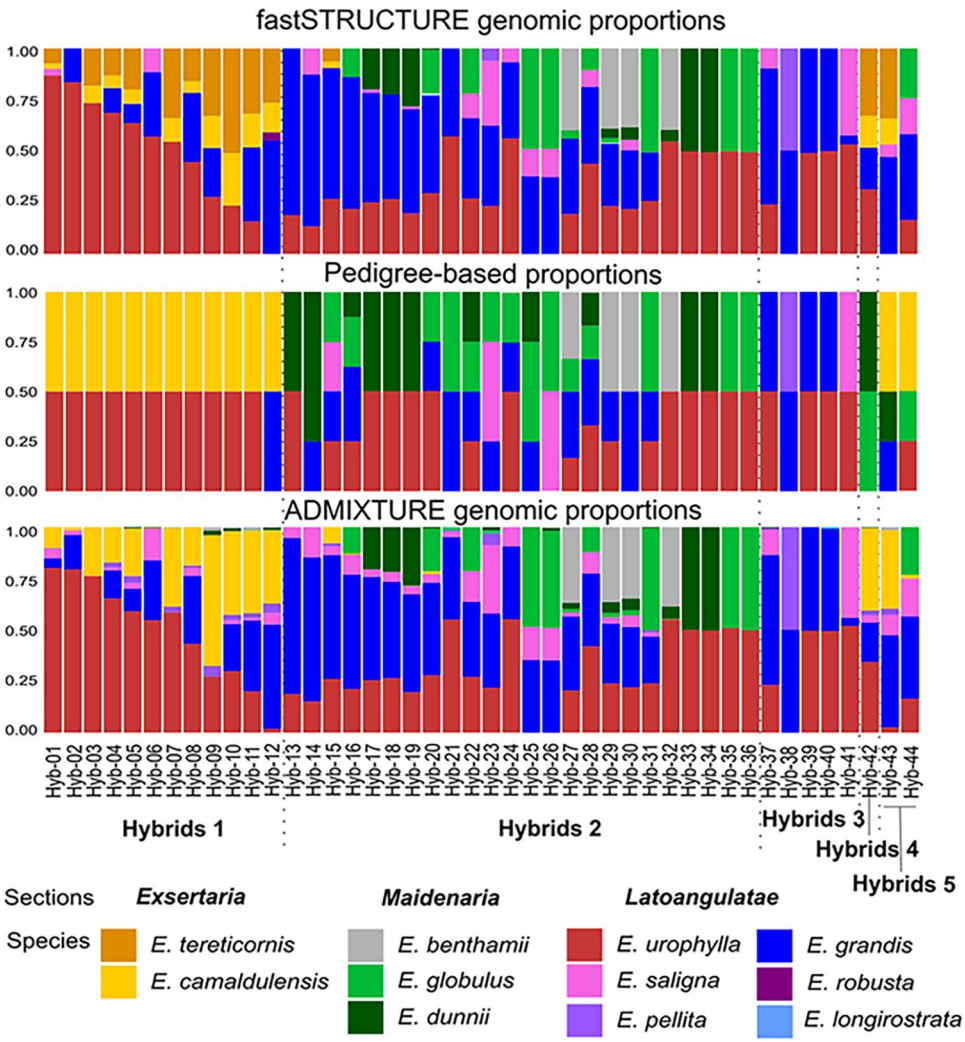

Fig 4. Analysis of the ancestral species' genomic composition of the 44 interspecific hybrids studied. The genomic proportions estimated by unsupervised inference with fastSTRUCTURE (top panel) and by a supervised model with species data as reference with ADMIXTURE (bottom panel), were compared with the expected composition from pedigree information (middle panel). The species were categorized into sections according to Brooker's (2000) classification.

https://doi.org/10.1371/journal.pone.0289536.g004

Genetic distances among species, provenances and hybrids

Overall, the PCA plot based on the genetic distance matrix positioned the different species and sections as expected, clustering phylogenetically closer species of the same section (Fig 5). A clear exception, however, was seen for *E. longirostrata*, taxonomically classified in section *Latoangulatae*. The PCA placed it away from *Latoangulatae* and closer to *E. argophloia* and *E. deglupta*. These two species belong to two different sections but they clustered together, considerably separated from all other species. In most cases, the PCA analysis had no resolution to discriminate provenances within species. In line with the fastSTRUCTURE results, exceptions were the Kroombit Tops provenance of *E. saligna*, and the two provenances each of *E. robusta* and *E. grandis* that were separated in the PCA.

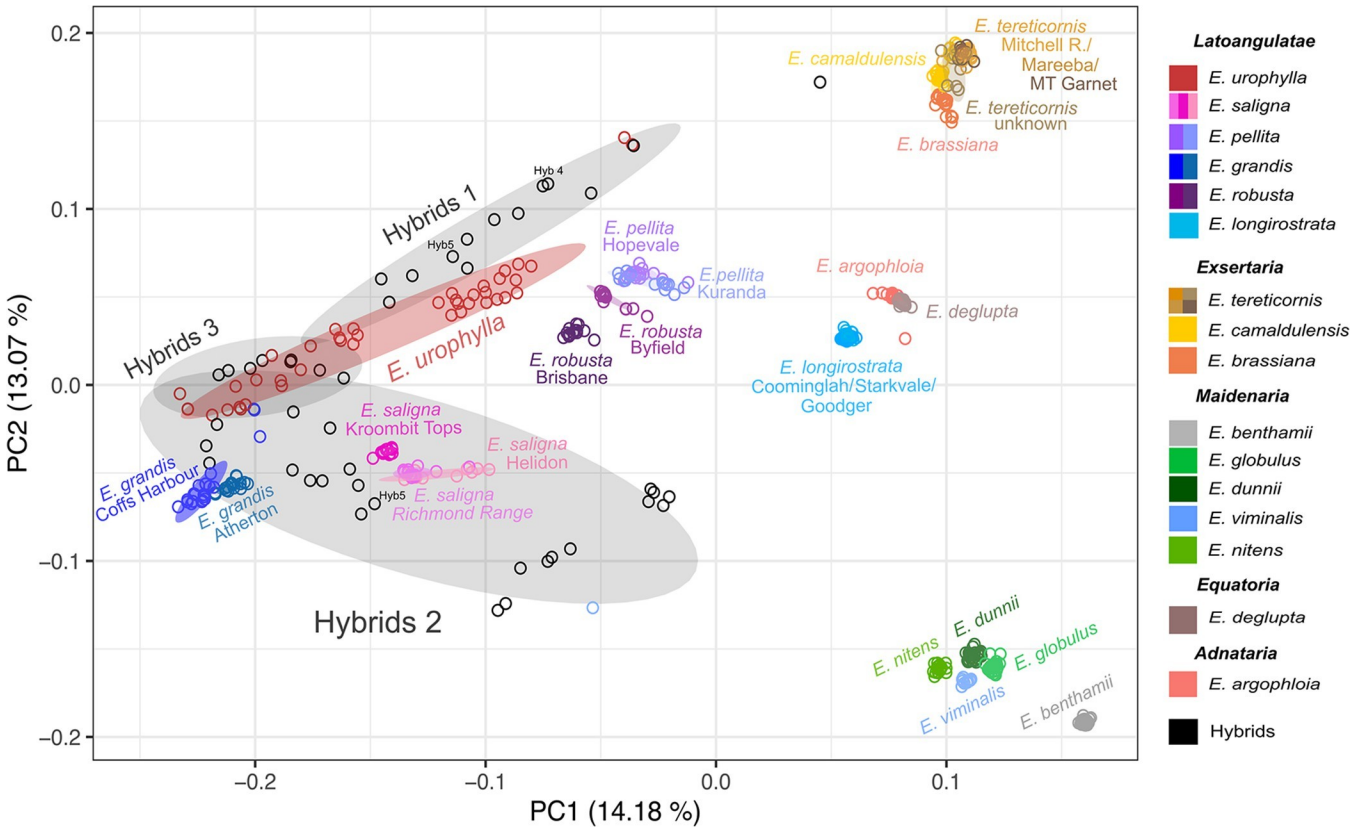

Fig 5. PCA scatter plot of the 484 *Eucalyptus* individuals in the first two principal components. Samples colored by species, provenances and hybrids were grouped in their respective classes according to the taxonomic sections involved in the cross. The ellipses depict the 95% confidence interval for the distribution of each species or hybrid group.

https://doi.org/10.1371/journal.pone.0289536.g005

Discussion

Genome-wide eucalypt species SNPs diversity

Consistent with the initial validation data provided alongside the EuCHIP60k development [31], our results corroborate that the current SNPs arrays platforms offer effective power to carry out genetic diversity analysis of the main eucalypt species planted worldwide. Within species, the proportion of polymorphic SNPs showed some variation, although for the vast majority, over 40% of the SNPs were informative and the average MAF was generally above 0.13, despite the relatively limited sample sizes analyzed (Table 3). Higher proportions of polymorphic SNPs, above 68% up to 93%, and higher average MAF were observed for *E. grandis*, *E. camaldulensis* and *E. urophylla*. These results may be explained in part from the somewhat larger sample sizes analyzed. In the case of *E. urophylla* the alleged mixture of provenances might have contributed to the higher diversity. A second explanation for the higher SNP diversity in these three species involves potential admixture due to unintended interspecific hybridization. These three species are widely used to generate interspecific hybrids in Brazil and the structure analysis results indicated admixture in the *E. urophylla* and *E. camaldulensis* trees (see below).

A third possible explanation for the higher SNP diversity observed in *E. grandis*, *E. camaldulensis* and *E. urophylla* is some ascertainment bias derived from the discovery panels used in the initial SNP discovery for the development of the EuCHIP60K. Although SNP discovery was carried out on sequence data for 240 trees of 12 species, a proportionally larger amount of

sequence data was obtained for these three species when compared to the others [31]. Large proportions of informative SNPs (58–60%) were also seen for species of *Maidenaria*, consistent with the fact *that E*. *globulus* was also an important target of sequence production during SNP discovery. Larger proportions of polymorphic SNPs and higher average MAF were also observed in the different hybrids. This was evidently expected, given the transmission to the hybrid of alternative SNP alleles fixed in each parental species.

Except for *E*. *urophylla*, *E*. *grandis*, *E*. *camaldulensis* and the hybrids, the results suggest that the rate of SNP polymorphism might depend more on the level of genetic diversity captured in the specific sample of individuals than on the particular species analyzed. This in turn indicates that the SNP set used delivers largely equivalent numbers of polymorphic SNPs between any pairwise taxa within the main sections of subgenus *Symphomyrtus*. This indicates good potential for the selection of ancestry informative SNPs sets [52] that appear in substantially different frequencies between species, provenances or populations in this phylogenetic group. The expansion of the number of species and provenances and the specific selection of ancestry informative SNPs at the species and provenances levels would constitute an obvious follow-up of this study.

SNPs recover the expected species structure but admixture is present

Genome-wide SNP data provided the necessary resolution to check and validate the phylogenetic classification of germplasm sources of the eucalypt species sampled in this study. The most likely model for the SNP dataset found k = 18 clusters, allowing clearcut discrimination of the five sections and the 16 species sampled of subgenus *Symphyomyrtus*, while reliably indicating the admixed composition of hybrids (Fig 2). This result substantiates what a number of previous phylogenetic studies have shown using different types of DNA marker data such as ribosomal ITS, chloroplast DNA, microsatellites and DArT (reviewed in [2]), and more recent studies that further expanded the sampling of taxa and individuals within taxa [8]. The evolutionary history 'written' in the genome of these *Symphyomyrtus* species is generally consistent with their current phylogenetic organization within this subgenus.

Differently from several previous reports that examined germplasm sampled exclusively in their center of origin, our study included material conserved in exotic conditions from variable sources (Table 1). In general, for the species' germplasm that came directly from original sources in Australia, the genetic structure splits were clearcut. For species that included material from unknown origins or collected in germplasm banks established in Brazil, occasional admixture was seen. The *E*. *camaldulensis* trees showed admixture with *E*. *tereticornis*, *E*. *viminalis* individuals showed admixture with *E*. *dunnii*, and *E*. *urophylla* sampled from multiple provenances established in Brazil displayed significant admixture with *E*. *grandis* (Fig 2). For some of these germplasm sources our data indicate that accidental hybridization might have taken place once the germplasm was introduced in Brazil. In the exotic habitat under different ecoclimatic conditions, reproductive barriers between eucalypt species such as geographic distance and flowering phenology that maintain species apart in their natural range, are relaxed or even broken, facilitating hybridization [53]. The paradigmatic example is the famous eucalypt hybrid swarm of the Rio Claro Arboretum established upon the introduction of *Eucalyptus* species in Brazil in 1904 [53, 54]. Several species were planted side by side, and seeds collected from that germplasm generated very heterogenous plantation forests, where some hybrids of unknown origin and outstanding performance were selected and are still planted or used in breeding programs today [17, 18]. The results of our study point to the development of ancestry informative SNPs that should allow reconstructing and understanding the recombination history of these hybrids.

The *E*. *viminalis* germplasm sample also showed evidences of admixture with *E*. *dunni* and *E*. *globulus* at k = 18. This sample of trees was from an advanced generation germplasm source established in Brazil but with unknown origin in Australia. Hybridization between these temperate species of section *Maidenaria* once introduced in Brazil cannot be ruled out, although less likely than for the previously mentioned species of *Exsertaria* and *Latoangulatae*, since *Maidenaria* species flower and produce seed less conspicuously in the tropics [53]. When a model with k = 20 was tested, *E*. *viminalis* individuals clearly split (S5 File) with no evidence of hybrid constitution. This result highlights the long-standing challenge with admixture modeling, whereby the most likely selection of K clusters is a difficult problem to automate in a way that is effectively robust [39].

The graphical projection of the different species and hybrids in the PCA was generally consistent with the phylogenetic expectations (Fig 5). Complementing the structure analysis, the PCA provided additional information regarding the genetic distance among the different taxa. *E*. *deglupta* and *E*. *argophloia* were placed at a considerable distance from the main section of *Symphyomyrtus*. The fact that they clustered together was however unexpected, since they are classified in distinct sections. These two species are currently part of *Symphyomyrtus* [55] and while no contention exists regarding the classification of *E*. *argophloia*, *E*. *deglupta* has originally been classified in subgenus *Minutifructis* [4]. The three main sections of interest in the subgenus were clearly separated and contained the expected species, exception made for *E*. *longirostrata* that clustered away from its section *Latoangulatae* and distant from *Exsertaria* as well.

Samples of *E*. *longirostrata* have been examined in the most extended molecular phylogenetic study of terminal taxa of sections *Maidenaria*, *Exsertaria* and *Latoangulatae* to date [8]. That study produced a phylogeny that largely matched the morphological treatment of sections, although sections *Exsertaria* and *Latoangulatae* were shown to be polyphyletic. Several inconsistencies between the morphological classification and the molecular phylogeny were described, and a number of taxa in *Latoangulatae* were deemed polyphyletic at the species level. A polyphyletic group is one that shows mixed evolutionary origin, descended from more than one ancestor, with taxa sharing homoplasies, typically explained as a result of convergent evolution, complicating the correct taxonomical classification [56]. *E*. *longirostrata* was itself deemed polyphyletic, classified within series *Lepidotae-Fimbriatae* and clustered into *Latoangulatae* IV, a clade considerably distant from *Latoangulatae* II where *E*. *grandis*, *E*. *pellita*, *E*. *robusta* and the section type species *E*. *saligna* belong. Furthermore, those authors suggest that all *Latoangulatae* species other than those in *Latoangulatae* II would be better placed in other taxonomic sections to reflect the phylogeny revealed in their study. The most recent classification of the eucalypts [14, 55] however, classified *E*. *longirostrata* into a different section, *Pumilio*. In our study, the sharp split of *E*. *longirostrata* from *Latoangulatae* and *Exsertaria* (Fig 5), provides further molecular evidence for this most recent taxonomic classification placing the species in a separate section.

Provenance discrimination is strongly dependent on geographical distance

With the exception of one provenance of *E*. *saligna*, all other Eucalyptus provenances could not be discriminated when all 484 samples were analyzed together (Fig 2). When species were analyzed separately, provenances could be discriminated for some species but not for others (Figs 3, 5). Looking at the geographical position of the sampled provenances (Fig 1), a pattern emerged suggesting that SNP-based discrimination was strongly dependent of geographical distance. The two provenances of *E*. *grandis* (Atherton and Coffs Harbor), separated in the structure and PCA analyses, are located at more than 2,000 km apart. The same happened with

provenances Byfield and Brisbane of *E*. *robusta* at ~700 km from each other, and *E*. *saligna* Kroombit Tops provenance located at >700 km from the other two *E*. *saligna* provenances. All other provenances that were loosely or otherwise not discriminated are located at less than 200–300 km apart. These results indicate an isolation-by-distance (IBD) model of population structure for the provenances sampled for these species. The genetic similarity between populations will decrease exponentially as the geographic distance between them increases, because of the limiting effect of geographic distance on rates of gene flow [57].

A number of studies in *Eucalyptus* have looked at the prevalence of genetic structure between populations located at various geographic distances. These studies have generally shown that an IBD model fits well the observed data, with genetic distances between provenances strongly positively correlated with geographic distances [24, 58, 59]. A recent landscape study based on very dense DNA data obtained by whole genome sequencing in *E*. *albens and E*. *sideroxylon*, also found strong support for IBD in both species [60]. Taken together, ours and others' results indicate that clearcut distinction of *Eucalyptus* germplasm sources in what regards provenance variation, might not be straightforward even with a dense panel of SNPs, unless provenances are geographically distant or provenance-informative SNP markers are specifically identified and used. As a result, what breeders may call as different provenances could in effect be members of the same continuous population despite several kilometers of physical distance, if gene flow is ubiquitous. It must be mentioned, however, that our study suffered from limited and somewhat uneven sampling of provenances that might have contributed to a greater difficulty in distinguishing some of them. It has been shown that subpopulations with reduced sampling tend to be merged together in genetic structure analyses, and uneven sampling may lead to downward-biased estimates of the true number of subpopulations [44]. Larger sample sizes for the provenances studied should allow better estimation of allele frequencies and possibly selection of ancestry informative, provenance-specific SNPs for greater discrimination power.

Genomic composition of hybrids indicates directional selection toward tropical genomes

Our genome-wide data showed that the majority of the hybrids studied (35 out of 44) displayed genomic composition deviating from the expected one based on pedigree information (Fig 4). This result is important in view of the long standing and widespread adoption of deliberate breeding strategies toward the selection of elite hybrid clones with specific anticipated genomic composition, especially in tropical countries (reviewed in [12]). This in turn highlights one more important application of using dense, high-quality array-based SNP data in support of breeding programs. SNP data not only provide precise germplasm identity verification, but more importantly allow the breeder to objectively recognize the actual ancestral origin of superior hybrids in order to discard unwanted hybrid combinations or to more realistically guide the breeding program toward the development of the desired genetic material.

For the sample of hybrids studied in this work, the lack of adherence between the expected genomic composition and the actual one suggests at least two hypotheses. Notwithstanding the possibility of mislabeling errors during controlled crosses, as likely the case for hybrids Hyb-13, Hyb-14 and Hyb-42, the second and most probable hypothesis is pervasive genetic admixture of the parents involved in the original interspecific cross. Given the frequently unknown introduction history, followed by local intermating in Brazil in the last 120 years, as discussed previously, there is a considerable possibility that the presumed parents were themselves misclassified. Moreover, because hybrids tend to be produced by crossing good performing parents in the breeding program, it is quite possible that actually some of the parents

used were themselves hybrids, distorting the expected composition of the resulting hybrid offspring. Species within the same sections of *Symphyomyrtus* that display overlapping morphological features and easily hybridize would be more prone to such occurrences. Clearcut examples were six supposedly F_1 hybrids that in principle did not involve *E*. *grandis*, but where the SNP data revealed its presence (Hyb-13, Hyb-17, Hyb-18, Hyb-19, Hyb-26, Hyb-42). Likewise, several F_1 hybrids of *E*. *urophylla* with *E*. *camaldulensis* (Hybrids 1 group) showed variable amounts of *E*. *grandis* genome in their composition, and the presence of *E*. *tereticornis* genome more frequently than that of *E*. *camaldulensis* (Fig 4). Admixture of *E*. *grandis* genome into the *E*. *urophylla* parents and difficulties in morphologically discriminating *E*. *camaldulensis* germplasm from *E*. *tereticornis* could readily explain these results.

Besides the presence of *E*. *grandis* as an unexpected species in the genomically realized pedigree, the observation of larger than expected proportions of *E*. *grandis* genome was also seen for all hybrids where this species was involved. Fourteen hybrids derived from advanced generation recombinant intercrosses involved one or both hybrid parents with three or more species represented, *E*. *grandis* being one of them (ex. Hyb-14, Hyb-15, Hyb-16, Hyb-20, Hyb-22 through Hyb-25, Hyb-27 through Hyb-29, Hyb-32, Hyb-43 and Hyb-44) (Table 2). The pedigree-expected proportions were estimated based on the final presumed participation of each single species in the pedigree, assuming balanced Mendelian inheritance and recombination rates in the previous hybrid generations with no selection. For all these 20 hybrids, the SNP data showed, however, a consistently higher proportion of *E*. *grandis* genome in the hybrid composition. Aside from unintended admixture in the original parents, the ubiquitous unexpected presence or higher than anticipated proportion of *E*. *grandis* genome in the vast majority of hybrids, strongly suggests genome-wide directional selection for this species' genome throughout the breeding history of these complex hybrid clones. This should not be surprising given that volume growth is the main breeding target, and that *E*. *grandis* is well known for its fast growth [53]. Our data therefore not only corroborates the pivotal role of *E*. *grandis* in hybrid breeding, but also shows that its actual participation is considerably larger than expected and frequently unintended. Moreover, our data also demonstrate that in hybrids between species of *Latoangulatae* and *Exsertaria* with species of *Maidenaria* (Hybrids 2 group), the actual participation of the latter, such as *E*. *globulus*. *E*. *dunnii* and *E*. *benthamii* in the final hybrid's genome composition is less than expected, consistent with strong selection against the less adapted temperate genomes in tropical environments.

Concluding remarks

In conclusion, we have shown that the current *Eucalyptus* multi-species SNP array platform, provides a valuable tool to look at within taxa variation in *Symphyomyrtus*, to investigate population structure and track the genomic ancestry of individual clones. As the current "gold standard" in the high-throughput SNP genotyping industry, SNP arrays provide full data portability across studies carried out at different times. This represents a crucial advantage for the construction of legacy SNP databases for multiple *Eucalyptus* species and populations when compared to reduced representation genotyping by sequencing methods. SNP array data portability across studies allows effortless data consolidation across time for comparative studies and meta-analyses, that should be valuable for resolving taxonomic issues that still persist in the eucalypts. We are aware, however, that for eucalypt species phylogenetically distant from subgenus *Symphyomyrtus*, the current SNP array will not provide equivalent numbers of informative SNPs due to a higher genomic divergence [31].

We have also shown that while species classification is well resolved at the genome-wide level, provenance discrimination is not always so. It depends essentially on geographical

distance, consistent with an isolation by distance model, and likely to be impacted by sample size. Further studies with larger samples sizes and the identification of provenance specific SNPs are warranted. Finally, our results are novel in that they objectively show, based on SNP data, that unplanned genetic admixture should not be a surprise in exotic germplasm sources not only in Brazil but likely in other countries, especially among phylogenetically closer species that easily hybridize in exotic environments. Moreover, the genomic ancestral composition of control-crossed hybrids in Brazil indicated that strong selection takes place in favor of tropical genomes and more specifically that of *E*. *grandis*. SNP-based auditing of hybrids' genomic composition could be introduced as a standard practice in hybrid breeding programs to more truthfully guide the program toward the development of the desired genetic material.

Supporting information

S1 File. SNP genotype data. Complete dataset for the filtered 27,298 SNPs obtained with the 72k Eucalyptus Axiom Array for the 484 individuals studied.
(CSV)

S2 File. Supervised estimators of k clusters. Results of the four supervised estimators of Puechmaille (2016) to detect the number of clusters implemented in the web server Structure-Selector (Li and Liu 2018) indicating that the germplasm set is most likely structured in 18 clusters after modelling with a variable number of k from 2 to 30 using FastStructure.
(DOCX)

S3 File. Output fastSTRUCTURE at k = 18. Output of the meanQ values of the fastSTRUCTURE analysis of the 484 individuals studied with K = 18.
(XLSX)

S4 File. Structure analysis plot at k = 2 to 4. Population structure analyses of the *Eucalyptus* species and hybrids clustered with variable numbers of clusters (*K*) from 2 to 4 separating the *Eucalyptus* sections (*Maidenaria*, *Latoangulatae* e *Exsertaria*), while displaying admixture in species of section *Latoangulatae*.
(DOCX)

S5 File. Structure analysis plot at k = 20 to 30. Population structure analyses of the *Eucalyptus* species and hybrids clustered with variable numbers of clusters (K) from 20 to 30, beyond the most likely model with K = 18.
(DOCX)

S6 File. Output fastSTRUCTURE & ADMIXTURE of hybrids' composition. Output of the meanQ values of the unsupervised fastSTRUCTURE analysis (sheet A) and supervised ADMIXTURE analysis (sheet B) of the ancestral genomic composition of the 44 hybrids.
(XLSX)

Acknowledgments

We would like to thank the IPEF cooperative tree breeding program (PCMF) affiliated companies, highlighting CMPC, Klabin, Suzano and Vallourec for providing materials for the study. Special thanks to the Experimental Stations of Forestry Sciences at ESALQ/USP for maintaining a large genetic collection of eucalypts in partnership with IPEF (current agreement: 1013868) and to prof. Alexandre S. Coelho for his assistance with computational facilities.

Author Contributions

Conceptualization: Paulo Henrique Muller da Silva, Evandro Novaes, Dario Grattapaglia.

Data curation: Danyllo Amaral de Oliveira, Paulo Henrique Muller da Silva, Evandro Novaes, Dario Grattapaglia.

Formal analysis: Danyllo Amaral de Oliveira, Paulo Henrique Muller da Silva, Evandro Novaes.

Funding acquisition: Paulo Henrique Muller da Silva, Dario Grattapaglia.

Investigation: Danyllo Amaral de Oliveira, Evandro Novaes.

Methodology: Paulo Henrique Muller da Silva, Evandro Novaes.

Project administration: Paulo Henrique Muller da Silva, Evandro Novaes.

Resources: Paulo Henrique Muller da Silva, Evandro Novaes, Dario Grattapaglia.

Software: Evandro Novaes.

Supervision: Evandro Novaes.

Validation: Dario Grattapaglia.

Writing – original draft: Danyllo Amaral de Oliveira, Dario Grattapaglia.

Writing – review & editing: Danyllo Amaral de Oliveira, Evandro Novaes, Dario Grattapaglia.

References

1. Myburg AA, Potts BM, Marques CM, Kirst M, Gion JM, Grattapaglia D, et al. Eucalyptus. In: C K, editor. Genome mapping and molecular breeding in plants Vol. 7: Forest trees. New York, NY, USA: Springer; 2007. p. 115–60.
2. Grattapaglia D, Vaillancourt R, Shepherd M, Thumma B, Foley W, Külheim C, et al. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. Tree Genet Genomes. 2012; 3:463–508. https://doi.org/10.1007/s11295-012-0491-x
3. Iglesias-Trabado G, Wilstermann D. Eucalyptus universalis. Global cultivated eucalypt forests map June 2009. Version 1.0.2: www.gitforestry.; 2009 [cited accessed 21 April 2023].
4. Brooker MIH. A new classification of the genus Eucalyptus L'Her. (Myrtaceae). Australian Systematic Botany. 2000; 13(1):79–148.
5. Slee A, Brooker M, Duffy S, West J. EUCLID: Eucalypts of Australia. Collingwood, Australia: CSIRO Publishing; 2006.
6. Steane DA, Nicolle D, Vaillancourt RE, Potts BM. Higher-level relationships among the eucalypts are resolved by ITS-sequence data. Australian Systematic Botany. 2002; 15:49–62.
7. Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, et al. Chloroplast genome analysis of Australian eucalypts–Eucalyptus, Corymbia, Angophora, Allosyncarpia and Stockwellia (Myrtaceae). Molecular Phylogenetics and Evolution. 2013; 69(3):704–16. https://doi.org/10.1016/j.ympev.2013.07.006 PMID: 23876290
8. Jones RC, Nicolle D, Steane DA, Vaillancourt RE, Potts BM. High density, genome-wide markers and intra-specific replication yield an unprecedented phylogenetic reconstruction of a globally significant, speciose lineage of Eucalyptus. Molecular Phylogenetics and Evolution. 2016; 105:63–85. https://doi.org/10.1016/j.ympev.2016.08.009 PMID: 27530705
9. Steane DA, Nicolle D, Sansaloni CP, Petroli CD, Carling J, Kilian A, et al. Population genetic analysis and phylogeny reconstruction in Eucalyptus (Myrtaceae) using high-throughput, genome-wide genotyping. Mol Phylogenet Evol. 2011; 59(1):206–24. Epub 2011/02/12. https://doi.org/10.1016/j.ympev.2011.02.003 PMID: 21310251.
10. Griffin AR, Burgess IP, Wolf L. Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit.—a review. Australian Journal of Botany. 1988; 36:41–66.

11. Potts BM, Barbour RC, Hingston AB, Vaillancourt RE. Turner Review No. 6: Genetic pollution of native eucalypt gene pools—identifying the risks. Australian Journal of Botany. 2003; 51(1):1–25. ISI:000181019700001.

12. Potts BM, Dungey HS. Hybridisation of Eucalyptus: Key issues for breeders and geneticists. New Forest. 2004; 27:115–38.

13. Butcher PA, McDonald MW, Bell JC. Congruence between environmental parameters, morphology and genetic structure in Australia's most widely distributed eucalypt, Eucalyptus camaldulensis. Tree Genet Genomes. 2009; 5(1):189–210. https://doi.org/10.1007/s11295-008-0169-6

14. Nicolle D, Jones RC. A revised classification for the predominantly eastern Australian Eucalyptus subgenus Symphyomyrtus sections Maidenaria, Exsertaria, Latoangulatae and related smaller sections (Myrtaceae). Telopea. 2018; 21:129–45. https://doi.org/10.7751/telopea12571.

15. Harwood C, editor New introductions—doing it right. Proceedings of the Conference "Developing a Eucalypt Resource for New Zealand"; 2011; Blenheim, New Zealand.

16. dos Santos GA, Nunes ACP, de Resende MDV, Silva LD, Higa A, de Assis TF. An index combining volume and Pilodyn penetration to study stability and adaptability of Eucalyptus multi-species hybrids in Rio Grande do Sul, Brazil. Australian Forestry. 2016; 79(4):248–55. https://doi.org/10.1080/00049158.2016.1237253

17. Rezende GDSP, Resende MDV, Assis TF. Eucalyptus Breeding for Clonal Forestry. In: Fenning T, editor. Challenges and Opportunities for the World's Forests in the 21st Century. Dordrecht: Springer Netherlands; 2014. p. 393–424.

18. Grattapaglia D, Kirst M. *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytologist. 2008; 179(4):911–29. https://doi.org/10.1111/j.1469-8137.2008.02503.x ISI:000258266200005. PMID: 18537893

19. Assis T. Hybrids and mini-cutting: a powerful combination that has revolutionized the Eucalyptus clonal forestry. BMC Proceedings. 2011; 5(Suppl 7):I18. https://doi.org/10.1186/1753-6561-5-S7-I18

20. Lima BM, Cappa EP, Silva-Junior OB, Garcia C, Mansfield SD, Grattapaglia D. Quantitative genetic parameters for growth and wood properties in Eucalyptus "urograndis" hybrid using near-infrared phenotyping and genome-wide SNP-based relationships. PloS one. 2019; 14(6):e0218747. https://doi.org/10.1371/journal.pone.0218747 PMID: 31233563

21. Paludeto JGZ, Grattapaglia D, Estopa RA, Tambarussi EV. Genomic relationship-based genetic parameters and prospects of genomic selection for growth and wood quality traits in Eucalyptus benthamii. Tree Genet Genomes. 2021; 17(4):20. https://doi.org/10.1007/s11295-021-01516-9 WOS:000679453800001.

22. McKinnon GE, Vaillancourt RE, Steane DA, Potts BM. An AFLP marker approach to lower-level systematics in Eucalyptus (Myrtaceae). American Journal of Botany. 2008; 95(3):368–80. https://doi.org/10.3732/ajb.95.3.368 PMID: 21632361

23. Hudson CJ, Freeman JS, Myburg AA, Potts BM, Vaillancourt RE. Genomic patterns of species diversity and divergence in Eucalyptus. New Phytologist. 2015; 206(4):1378–90. https://doi.org/10.1111/nph.13316 PMID: 25678438

24. Rutherford S, Rossetto M, Bragg JG, McPherson H, Benson D, Bonser SP, et al. Speciation in the presence of gene flow: population genomics of closely related and diverging Eucalyptus species. Heredity. 2018; 121(2):126–41. https://doi.org/10.1038/s41437-018-0073-2 PMID: 29632325

25. von Takach Dukai B, Jack C, Borevitz J, Lindenmayer DB, Banks SC. Pervasive admixture between eucalypt species has consequences for conservation and assisted migration. Evolutionary Applications. 2019; 12(4):845–60. https://doi.org/10.1111/eva.12761 PMID: 30976314

26. Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, Myburg AA, et al. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. Plant Methods. 2010; 6:16. Epub 2010/07/01. https://doi.org/10.1186/1746-4811-6-16 [pii] PMID: 20587069.

27. Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, et al. Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proceedings. 2011; 5(Suppl 7):P54. https://doi.org/10.1186/1753-6561-5-S7-P54

28. Myles S. Improving fruit and wine: what does genomics have to offer? Trends in Genetics. 2013; 29(4):190–6. https://doi.org/10.1016/j.tig.2013.01.006 PMID: 23428114

29. Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, et al. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular Ecology. 2013; 22(11):3165–78. https://doi.org/10.1111/mec.12089 PMID: 23110526

30. Lowry DB, Hoban S, Kelley J, L., Lotterhos K, E., Reed L, K., Antolin M, F., et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation.

Molecular Ecology Resources. 2016; 17(2):142–52. https://doi.org/10.1111/1755-0998.12635 PMID: 27860289

31. Silva-Junior OB, Faria DA, Grattapaglia D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing 240 *Eucalyptus* tree genomes across 12 species. New Phytologist. 2015; 206(4):1527–40. https://doi.org/10.1111/nph.13322 PMID: 25684350

32. Faria DA, Mamani EMC, Pappas GJ, Grattapaglia D. Genotyping systems for Eucalyptus based on tetra-, penta-, and hexanucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. Tree Genet Genomes. 2011; 7(1):63–77. https://doi.org/10.1007/s11295-010-0315-9 ISI:000286462800006.

33. Correia L, Faria D, Grattapaglia D. Comparative assessment of SNPs and microsatellites for fingerprinting, parentage and assignment testing in species of Eucalyptus. BMC Proceedings. 2011; 5(Suppl 7): P41. https://doi.org/10.1186/1753-6561-5-S7-P41

34. Silva-Junior OB, Grattapaglia D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of Eucalyptus grandis. New Phytologist. 2015; 208:830–45. https://doi.org/10.1111/nph.13505 PMID: 26079595.

35. Grattapaglia D. Twelve Years into Genomic Selection in Forest Trees: Climbing the Slope of Enlightenment of Marker Assisted Tree Breeding. Forests [Internet]. 2022; 13(10).

36. Mostert-O'Neill MM, Reynolds SM, Acosta JJ, Lee DJ, Borevitz JO, Myburg AA. Genomic evidence of introgression and adaptation in a model subtropical tree species, Eucalyptus grandis. Molecular Ecology. 2021; 30(3):625–38. https://doi.org/10.1111/mec.15615 PMID: 32881106

37. Inglis PW, Pappas MdCR, Resende LV, Grattapaglia D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. PLosOne. 2018; 13(10):e0206085. https://doi.org/10.1371/journal.pone.0206085 PMID: 30335843

38. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81 (3):559–75. Epub 2007/07/25. https://doi.org/10.1086/519795 PMID: 17701901.

39. Liu C-C, Shringarpure S, Lange K, Novembre J. Exploring Population Structure with Admixture Models and Principal Component Analysis. In: Dutheil JY, editor. Statistical Population Genomics. New York, NY: Springer US; 2020. p. 67–86.

40. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008; 24(11):1403–5. https://doi.org/10.1093/bioinformatics/btn129 PMID: 18397895

41. Goudet J. hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes. 2005; 5(1):184–6. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

42. Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012; 28(2):298–9. https://doi.org/10.1093/bioinformatics/btr642 PMID: 22110245

43. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics. 2014; 197(2):573–89. https://doi.org/10.1534/genetics.114.164350 PMID: 24700103

44. Puechmaille SJ. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Molecular Ecology Resources. 2016; 16(3):608–27. https://doi.org/10.1111/1755-0998.12512 PMID: 26856252

45. Li Y-L, Liu J-X. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources. 2018; 18(1):176–7. https://doi.org/10.1111/1755-0998.12719 PMID: 28921901

46. Francis RM. pophelper: an R package and web app to analyse and visualize population structure. Molecular Ecology Resources. 2017; 17(1):27–32. https://doi.org/10.1111/1755-0998.12509.

47. Auguie B. gridExtra. Miscellaneous Functions for "Grid" Graphics.2017.

48. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 2009; 19(9):1655–64. https://doi.org/10.1101/gr.094052.109 PMID: 19648217

49. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019; 35(3):526–8. https://doi.org/10.1093/bioinformatics/bty633 PMID: 30016406

50. Paradis E. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. 2010; 26(3):419–20. https://doi.org/10.1093/bioinformatics/btp696 PMID: 20080509

51. Wickham H. ggplot2. Elegant Graphics for Data Analysis. Cham, Switzerland: Springer 2016. 260 p.

52. Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, et al. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Science International: Genetics. 2007; 1(3):273–80. https://doi.org/10.1016/j.fsigen.2007.06.008 PMID: 19083773

53. Eldridge K, Davidson J, Harwood C, van Wyk G. Eucalypt domestication and breeding. Oxford: Clarendon Press; 1993. 288 p.

54. Brune A, Zobel BJ. Genetic base populations, gene pools and breeding populations for *Eucalyptus* in Brazil. Silvae Genet. 1981; 30(4–5):146–9.

55. Nicolle D. Classification of the eucalypts (Angophora, Corymbia and Eucalyptus) Version 6. 2022. Available from: http://www.dn.com.au/Classification-Of-The-Eucalypts.pdf.

56. Beentje H, Williamson J. The Kew Plant Glossary: An Illustrated Dictionary of Plant Terms: Kew; 2010.

57. Kimura M, Weiss GH. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics. 1964; 49(4):561–76. https://doi.org/10.1093/genetics/49.4.561 PMID: 17248204

58. Jones TH, Vaillancourt RE, Potts BM. Detection and visualization of spatial genetic structure in continuous Eucalyptus globulus forest. Molecular Ecology. 2007; 16(4):697–707. https://doi.org/10.1111/j.1365-294X.2006.03180.x PMID: 17284205

59. Steane DA, Potts BM, McLean E, Collins L, Prober SM, Stock WD, et al. Genome-wide scans reveal cryptic population structure in a dry-adapted eucalypt. Tree Genetics & Genomes. 2015; 11(3):33. https://doi.org/10.1007/s11295-015-0864-z

60. Murray KD, Janes JK, Jones A, Bothwell HM, Andrew RL, Borevitz JO. Landscape drivers of genomic diversity and divergence in woodland Eucalyptus. Molecular Ecology. 2019; 28(24):5232–47. https://doi.org/10.1111/mec.15287 PMID: 31647597

ARTIGO 2 - AN RNA-SEQ APPROACH REVEALS THE METABOLIC CHANGES IN RESPONSE TO THE PHYSIOLOGICAL DISORDER OF *EUCALYPTUS* AND ITS RECOVERY

Artigo redigido de acordo com as normas da revista *Tree physiology*

AN RNA-SEQ APPROACH REVEALS THE METABOLIC CHANGES IN RESPONSE TO THE PHYSIOLOGICAL DISORDER OF *EUCALYPTUS* AND ITS RECOVERY

Abstract

Eucalyptus plantations are expanding around the world, which poses new challenges to the culture. The Physiological Disorder of *Eucalyptus* (PDE), despite not being a recent problem, has been threatening the *Eucalyptus* plantations in some regions of Brazil. The disorder causes apical dieback, over budding, bark shredding and significantly decreases the forest yield. Several hypotheses have already been tested as the cause of the disorder, however, the causal agent remains unknown. This study aimed to identify the metabolic changes due to the PDE by comparing susceptible and resistant clones under field conditions with and without PDE. An RNA-seq approach was used to evaluate differentially expressed genes (DEGs) in three different locations: two with high PDE incidence (BA1 and BA2) and another with no PDE history (SP), in two seasons (during and after the disorder, when plants had symptoms and were recovering). Differential expression analyzes were performed using the bioconductor package DESeq2, contrasting susceptible vs. resistant clones within each location. Lists with DEGs were used in Gene Ontology (GO) enrichment analyses using AgriGO web tool. Pathways activated and repressed under PDE were visualized using MapMan. Many GO terms and pathways related to stress signaling, signaling hormone synthesis, and signal transduction pathways were activated. Genes related to processes related to energy production and fermentation were differentially regulated between susceptible and resistant clones. Terpenoids and other pathways were identified as important to the resistance or susceptibility of clones to PDE. In conclusion, our study indicates that the physiological disorder causes widespread metabolic changes in several pathways of *Eucalyptus*. This study serves as a basis for illuminating the disorder and a source of new hypotheses to the cause of this important disorder.

Keywords: Plant stress, tree metabolism, differential expressed genes, abiotic factors

Introduction

Eucalyptus is a genus that contain fast-growing and high-yielding tree species planted worldwide. These woody plants have many uses, such as pulp and paper production, charcoal, lumber, and essential oil industries (Assis et al. 2015). Although eucalypts are characterized by high adaptation to many soil and climate conditions (e.g. high tolerance to infertile soil and drought), new challenges are appearing as plantations increase to new areas with different climatic conditions (Florêncio et al. 2022). Since the 2000's, a new disorder was found in a coastal region of Brazil, called the Physiological Disorder of *Eucalyptus* (PDE). More specifically, the disorder was found in *Eucalyptus* plantations in the South of Bahia State and North of Espírito Santo State in Brazil.

The PDE is a disorder that causes apical dieback, over budding, bark shredding and

significantly decreases the forest yield. Even though many hypotheses have been raised about the PDE cause, its etiology is still unknow. This disorder often affects plants with six to eight months after the plantation, causing high financial loses to the companies (Bueno et al. 2020). After PDE incidence, a recovery season of susceptible clones occur but the trees lose their industry viability. Research performed by the Brazilian forestry companies have identified susceptible clones, as well as the hotspots where the PDE happen. These findings indicate that the PDE expression is influenced by both genetic and environmental components (Câmara et al. 2018, Almeida et al. 2013).

Many hypotheses were developed about the causal agent of PDE. Evaluations about biotic agents of diseases like bacteria and fungi were test, but no isolated microorganism could induce the PDE or similar symptoms (Ferreira et al. 1989). Nowadays, most breeders and researchers involved with PDE studies do not believe that the causal agent is biotic. Evidences point to PDE as an abiotic disorder, as it occurs in some specific sites often after the rainy season (Almeida et al. 2013, Leite et al. 2014). Metal toxicity like manganese also were studied as a potential cause in flooded sites (Leite et al. 2014). The manganese intoxicated plants show symptoms close to the eucalypts affected by PDE. However, it is known that manganese increases the symptoms but is not the principal agent of the PDE (Harguindeguy et al. 2018). Another idea is the alternation of a drought season followed by flooding that can cause root hypoxia and death (Jardim et al. 2018, Ferreira et al. 1989). Under this hypothesis, affected plants should be mainly on lower altitudes, but PDE has also been found on higher lands. New studies using different methodologies can raise new hypotheses about the etiology of the PDE.

Knowledge about the differential expression of genes between susceptible and resistant clones under PDE conditions can shed light into problem. The RNA-seq is known to be an efficient technique for the quantification and evaluation of gene expression (Metzker 2010, Van Dijk et al. 2014). Many RNA-seq studies of plant stress, such as heat, drought, flood, salt, heavy metal stress point to key genes and pathways involved in the plants' response to these conditions. Metabolic Pathways and Gene Ontology terms enrichment using the differential expressed genes (DEGs) between susceptible and resistant genotypes can aid the identification of the main molecular responses to the PDE.

Comparing the activated pathways and important DEGs between susceptible and resistant clones to PDE may highlight the causes of this disorder. This study aimed to identify the metabolic

changes due to the Physiological Disorder of *Eucalyptus* comparing susceptible and resistant clones under contrasting field conditions using a RNA-seq approach. The clones were grown in three experiments: two in sites with high incidence of PDE and another in a site without history of the disorder.

Material and methods

Field experiment

Three *Eucalyptus* field experiments were planted on April 2021 using a randomized complete blocks design with three replicates, square plots of 4x4 plants and two border lanes. Trees were planted using a 3x2 m spacing. The treatments compared in these three experiments were five *Eucalyptus* clones: two known as highly resistant to the PDE (clone 1 - C1 and clone 2 - C2), an intermediate clone (clone 3 – C3) which shows less PDE symptoms, and two clones highly susceptible (clone 4 – C4 and clone 5 – C5) to the PDE (Figure 1).

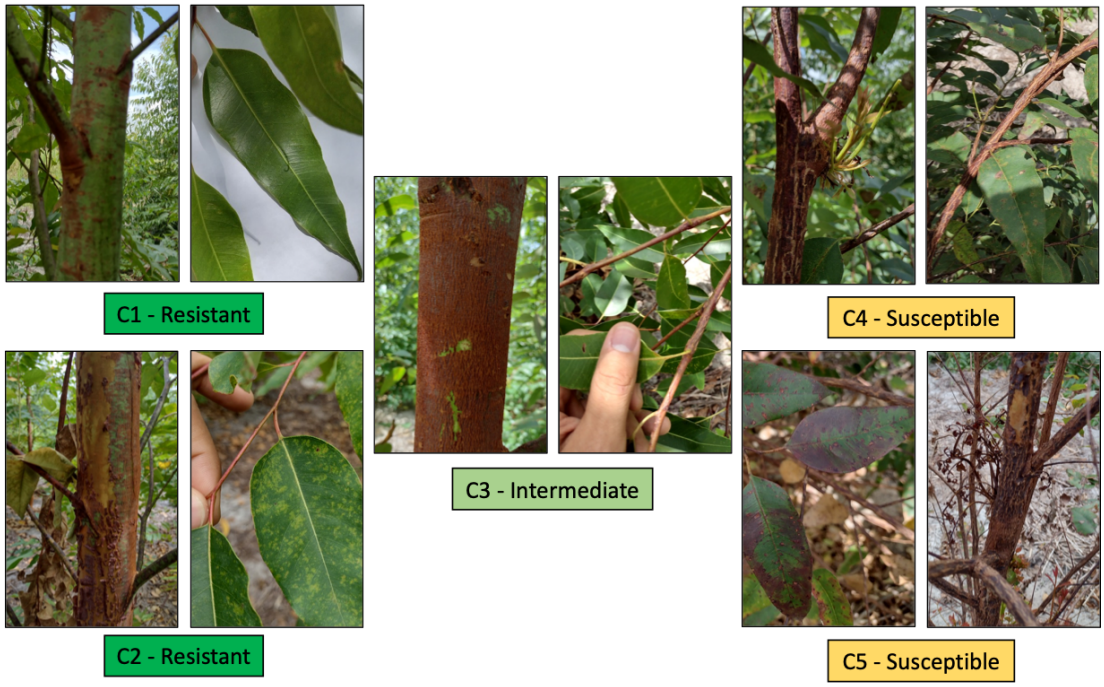

Figure 1. Phenotypic response of the five clones studied here, during the PDE conditions (Season 1) at site BA1.

The three sites with different climatic conditions (Figure S1) were chosen according to

the history of occurrence of the PDE in Brazil. Two experiments were in the south of Bahia State, Brazil, with high incidence of PDE. The other site was in the State of São Paulo State, Brazil, with no historical occurrence of the PDE (Figure 2). Planting and topdressing fertilization were performed on the experiments based on the physical and chemical analyses of the three soils. Weeds and ants were controlled with the standard chemical control practices employed by Suzano SA company.

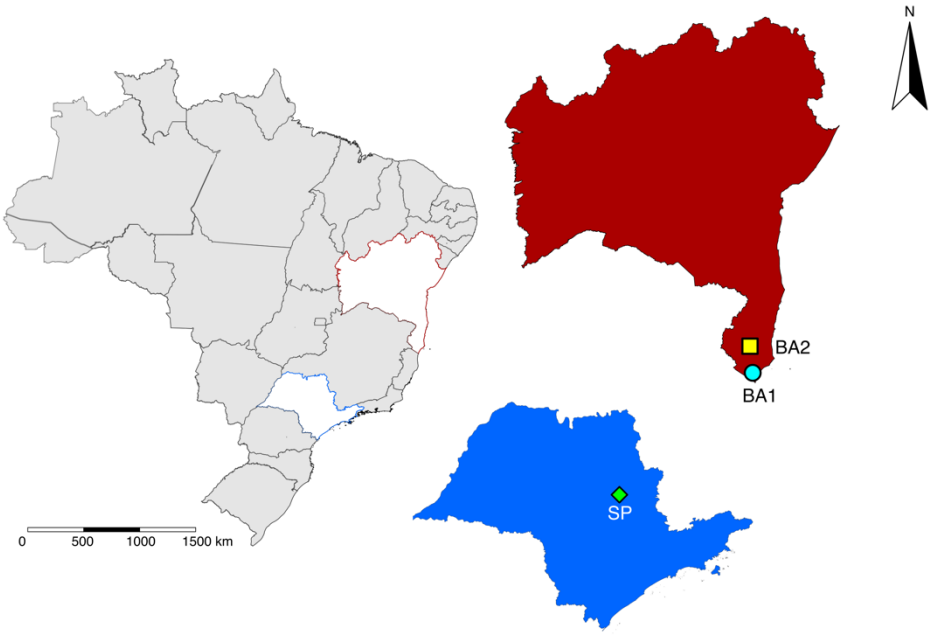

Figure 2. Map locating the three experimental sites in São Paulo State (SP, in blue), with no history of PDE, and two sites in Bahia State (BA, in red) with high incidence of PDE.

RNA extraction and sequencing

Young shoots, containing the apical meristem and a pair of leaves, were sampled from each experiment in two seasons at the same daytime (morning) for RNA-extraction. These tissues were chosen, as they are generally the first to start to show PDE symptoms. The first collection (Season 1 - S1) happened at ten months after planting (February 2022), when the susceptible clones had high incidence of PDE symptoms. The other sampling time (Season 2 – S2) was performed at 16 months after planting (August 2022) when the susceptible clones started to recover from the PDE symptoms. Samples from three biological replicates were collected from all genotypes in each season from each site, totaling 90 samples:

$$5\ clones \times 3\ sites \times 3\ replicates \times 2\ seasons = 90\ RNA\ samples$$

The RNA extraction was carried out using the Invitrogen PureLink RNA Mini kit. The RNA quality was evaluated using Nanodrop, Qubit and agarose gel electrophoresis. RNA samples were sent to the BGI company for the RNA sequencing using RNA stabilization columns (GTR5025-S Gentegra).

Reads mapping and gene expression analyses

The RNA-seq data were downloaded into a high-performance Linux server. First, the quality of the DNA sequences was evaluated using the FastQC software (http://www.bioinformatics.babraham.ac.uk/ projects/fastqc/). After attesting their quality, sequencing reads were mapped against the *Eucalyptus grandis* genome v.2.0, available on Phytozome 13 (https://phytozome-next.jgi.doe.gov/info/Egrandis_v2_0), using the STAR aligner v.2.7.10 (Dobin et al. 2013). Count files provided by STAR were merged into a matrix using R v. 4.3.1 (R Core Team 2022). This count matrix had the number of reads from each sample (in the matrix columns) mapped in each gene (rows) and was used for differential gene expression analyses. A Principal Component Analysis (PCA) were done using the rlog normalized count matrix using the gridExtra package.

The differentially expressed genes (DEG) were identified using the bioconductor package DESeq2 (Love et al. 2014). We used the interaction design model ~ *genotype* + *condition* + *genotype:condition* for the analyses following the DESeq2 manual (available at: https://bioconductor.org/packages/release/bioc/html/DESeq2.html). The main model design used was ~ *pheno_season* + *site* + *pheno_season:site*, where the *pheno_season* effect is the comparison between susceptible vs. resistant clones in each season; *site* is the effect of the three sites of experiments; *pheno_season:site* is the interaction between phenotypes within season with sites. The second design used was ~*clone_season* + *site* + *clone_season:site*, where *clone_season* means the comparison between individual clones in each season; *site* means the three sites of experiments; *clone_season:site* is the interaction between clones within season with sites (Table 1). For this second design we compared two known full-siblings clones with contrasting phenotypic response to PDE: resistant clone C2 × susceptible clone C5.

Using results from DESeq2, many contrasts were analyzed (Table 1). For example, clones susceptible vs. resistant within each site, contrasts between full-siblings clones with contrasting phenotypic response in each site, resistant clones' differences between Season 1 (disorder season)

and Season 2 (recovery season) or susceptible clones' differences between Season 1 and Season 2 as well as individual clones' comparison between Season 1 and Season 2. A filter of FDR > 0.05 and logfold2chage > |2| was applied to access the highly significant DEGs. Venn diagram was constructed using the online tool Venny (https://bioinfogp.cnb.csic.es/tools/venny/index.html) to evaluate the number of shared DEG among contrasts. The *Eucalyptus* genome annotation file was merged to the DEG lists to identify the functional categories of activated and repressed genes under PDE.

Table 1. Models used in DESeq2 and contrasts that result in DEGs lists.

Models	Comparison	Contrasts
~phenotype_season + site + phenotype_season:site	Phenotype model Interaction term	interaction susceptible_S1 vs. resistant_S1 : ambient BA1 vs. SP
		interaction susceptible_S1 vs. resistant_S1 : ambient BA2 vs. SP
		interaction susceptible_S2 vs. resistant_S2 : ambient BA1 vs. SP
		interaction susceptible_S2 vs. resistant_S2 : ambient BA2 vs. SP
	Phenotype comparison inside each individual ambient	susceptible_S1 vs. resistant_S1 within SP
		susceptible_S1 vs. resistant_S1 within BA1
		susceptible_S1 vs. resistant_S1 within BA2
		susceptible_S2 vs. resistant_S2 within SP
		susceptible_S2 vs. resistant_S2 within BA1
		susceptible_S2 vs. resistant_S2 within BA2
	Phenotype comparison among seasons inside each individual ambient	resistant_S1 vs. resistant_S2 within SP
		resistant_S1 vs. resistant_S2 within BA1
		resistant_S1 vs. resistant_S2 within BA2
		susceptible_S1 vs. susceptible_S2 within SP
		susceptible_S1 vs. susceptible_S2 within BA1
		susceptible_S1 vs. susceptible_S2 within BA2
~clone_season + site + clone_season:site	Clone model Interaction term	interaction C5_S1 vs. C2_S1 : ambient BA1 vs. SP
		interaction C5_S1 vs. C2_S1 : ambient BA2 vs. SP
		interaction C5_S2 vs. C2_S2 : ambient BA1 vs. SP
		interaction C5_S2 vs. C2_S2 : ambient BA2 vs. SP
	Full-sibling clones comparison inside each individual ambient	C5_S1 vs. C2_S1 within SP
		C5_S1 vs. C2_S1 within BA1
		C5_S1 vs. C2_S1 within BA2
		C5_S2 vs. C2_S2 within SP
		C5_S2 vs. C2_S2 within BA1
		C5_S2 vs. C2_S2 within BA2

GO enrichment and pathway activation/repression analyses

Functional analyses were performed with the list of DEGs for each individual contrast, using the Egrandis_297_v2.0.annotation_info file from Phytozome v.13. The GO terms enrichment analyses were carried out using the Fisher's Test, terms with FDR < 0.05 significance level were

highlighted on the web software AgriGO.v2.0 (Tian et al. 2017). *Eucalyptus grandis* genome gene identifiers (locus ID) were used for each DEG with default parameters on AgriGO to identify significantly enriched GO terms. Significant terms were analyzed with REVIGO (Supek et al. 2011) to reduce the redundant terms. Pathways activation and repression were visualized using the MapMan software (Usadel et al. 2009). The input files were DEG lists generated by DESeq2 for each contrast and expression (logfold2change) level for each gene.

Results

Physiological disorder changes the expression of a large number of genes

On average 17.73 M reads per sample were mapped onto the *E. grandis* genome. A Principal Component Analysis (PCA) with the normalized count matrix data, clustered the samples according to the genotypes. Clone C3, with intermediate resistance to PDE, had more dissimilar gene expression pattern as it clustered far from the other clones (Figure S2). The models used for the differentially expressed genes (DEGs) analyses obtained good convergence (Figure S3).

The highest number of DEGs was found at the season and sites with occurrence of the disorder (Figure 3). Sites BA2 and BA1, with high incidence of PDE, had a large number of DEGs in Season 1 (2,2722 and 2,446 DEGs), when susceptible clones had high incidence of PDE symptoms. On the other hand, in Season 2, when clones were recovering from the disorder, the number of DEG was lower (1,772 and 1,377 DEGs). The number of DEGs in SP, where there is no incidence of the disorder, was similar in the two collection seasons (Figure 3A). In general, during the stress period, the clones showed higher numbers of upregulated DEGs, while in the recovery period the number of downregulated DEGs was higher in BA.

Shared DEGs among sites were found for each season (Figure 3B). It was verified the existence of exclusive genes for each location, as well as the existence of genes that were exclusive for BA1 and BA2, places where the PDE occurs. The intersection between BA1 and BA2 points to a decrease of 107 DEGS from the disorder incidence period to the recovery period.

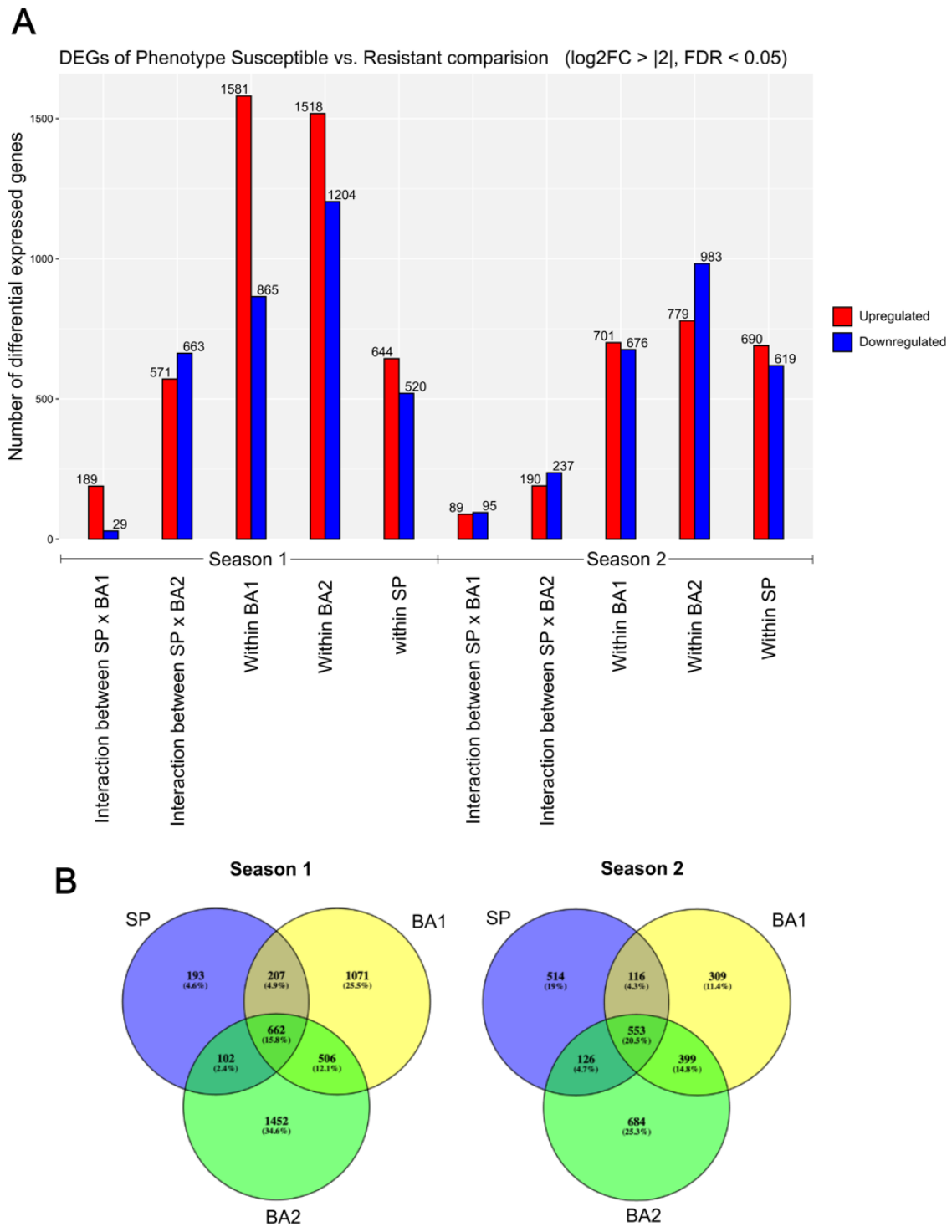

Figure 3. Number of DEGs in each contrast and shared among them. **A.** Number of down and upregulated genes in susceptible clones in each contrast. **B.** Venn diagram showing the number of shared genes among sites in each season.

Gene ontology enrichment among differentially expressed genes

Only the Gene Ontology (GO) terms of Molecular Function and Biological Process were enriched among the DEGs from the different contrasts (Figure 4). In terms of Molecular Function, a significant catalytic activity was observed in the two sites where *Eucalyptus* is affected by PDE. Enrichment of catalytic activity term (GO:0003824) was present in genes activated in both susceptible (upregulated) and resistant (downregulated) clones. Exopeptidase-related term (GO:0008238) was also enriched among genes upregulated in susceptible genotypes in BA1-S1, as well as in both sites with PDE during the recovery season (S2). Terms related to terpene synthase activity (GO:0010333 and its "child" terms) were upregulated in resistant clones during both seasons of sites with PDE.

Terms related to the Biological Process indicate that resistant clones and susceptible clones have different signaling activity regardless of the location where PDE occurs. For example, terms such as signaling (GO:0023052), phosphorylation (GO:0016310), protein phosphorylation (GO:0006468), signal transduction (GO:0007165) were enriched among both up and downregulated genes (i.e. activated in susceptible or resistant clones). GO terms also show that resistant clones seem to activate their reproductive potential. Reproductive process (GO:0022414), pollination (GO:0009856), recognition of pollen (GO:0048544), reproduction (GO:0000003) were some of the terms enriched among genes upregulated in resistant clones in almost all environments, except in BA2-S1. During the disorder, susceptible clones presented GO terms related to increased cellulose and cell wall production (GO:0044036 and GO:0016998). Enrichment was also observed for two terms related to chromosome condensation, chromosome condensation (GO:0030261) and mitotic chromosome condensation (GO:0007076), in susceptible clones during the disorder period.

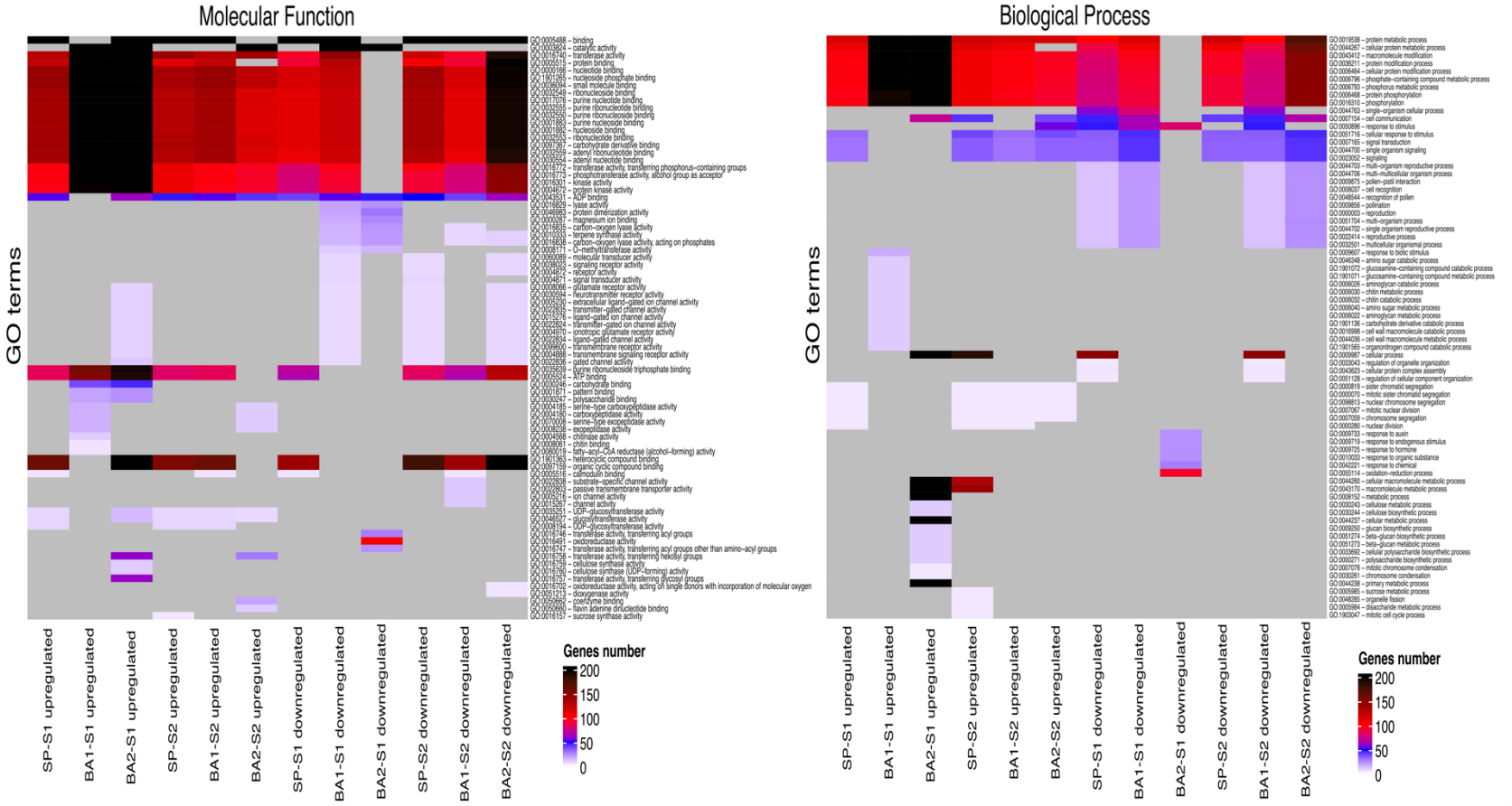

Figure 4. GO terms related to molecular function and biological process significantly enriched (FDR < 0.05) among up and downregulated genes between susceptible vs. resistant *Eucalyptus* clones in different sites and seasons.

Metabolic changes in susceptible vs. resistant clones

Many metabolic pathways were differentially regulated in susceptible vs. resistant clones in the different sites and seasons, with and without PDE. Some of these metabolic pathway changes seen to be due to constitutive genetic differences in resistant vs. susceptible clones. For example, sucrose degradation pathway, fermentation, terpenoid production, phenylpropanoid production, receptor kinases were activated in all environments and periods, regardless of the occurrence or not of the disorder. However, the expression of the genes in these pathways varied in susceptible vs. resistant clones with season and site of sampling.

Genes related to signaling proteins (kinases), protein degradation and modification, as well as transcription factors were more differentially regulated during the period and in sites where the plants were under stress (Figure 5). Plant hormones genes which act as stress signaling genes were also upregulated during the disorder. Jasmonate, IAA, ABA and salicylic acid were examples of hormones with altered gene expression between susceptible and resistant clones in different environments and collections. Many glutaredoxin-related genes were highly activated at sites where PDE occurs, mainly in disorder-susceptible genotypes (Figure 5).

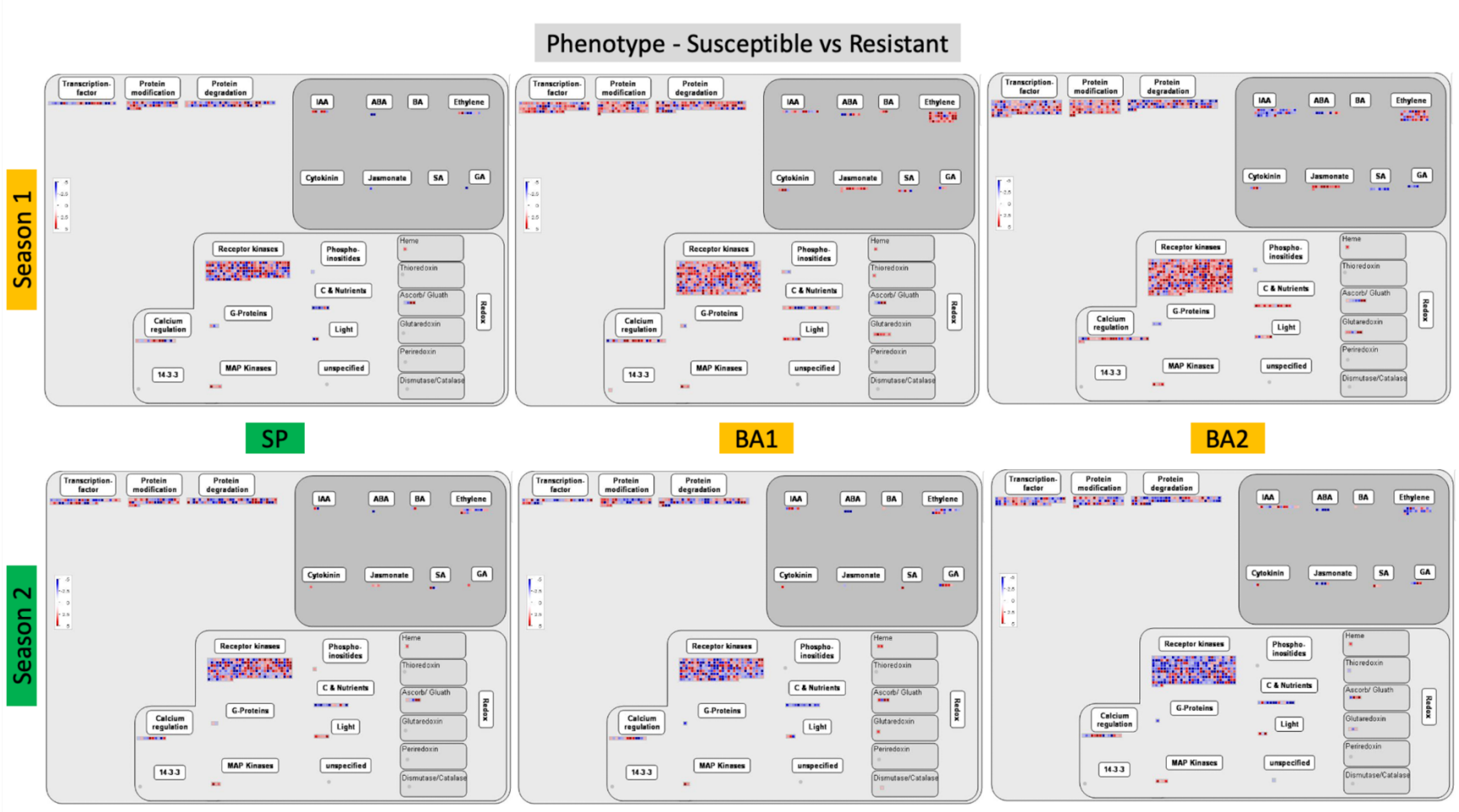

Figure 5. MapMan overview of regulatory pathways depicting differentially expressed genes between susceptible vs. resistant *Eucalyptus* clones within each ambientes (SP, BA1, BA2) during the seasons 1 and 2. Analyses and figure were generated in the MapMan software. Red squares depict upregulated genes and blue square downregulated genes in susceptible clones.

Susceptible versus resistant clones had many metabolic differences as observed by the large number of DEG in many pathways (Figure 6). For example, genes related to photorespiration tetrapyrrole synthesis showed higher expression in season S1. A chlorophylase II gene was upregulated in BA2 but downregulated in BA1. Another regulated pathway was the respiration tricarboxylic acid cycle (TCA), with susceptible clones having alpha carbonic anhydrase 7 gene expression activated during the disorder season. Aconitate hydratase genes, also from the TCA pathway, were upregulated in susceptible clones in sites and seasons BA1-S1, BA1-S2 and BA2-S2.

Some genes related to macromolecule synthesis were activated only in clones present in sites with a history of PDE (Figure 6). Genes related to phospholipid synthesis stand out, being differentially expressed between susceptible and resistant clones only in BA1 and BA2. Other important activated pathways were the cell wall-related pathways. Synthesis, signaling and cell wall modification pathways had DEG. Cell wall signaling genes were found to be downregulated at BA2-S1, such as some LRR (leucin rich repeat) and FLA7 (fascilin-like arabinogalactan 7). A cell wall modification pathway was activated in the first sampling season in both sites with PDE and in the second season only in BA2. Expansin and xyloglucan transglycosylase were genes differentially expressed in these pathways. However, in BA1-S1, genes of cell wall modification pathway were upregulated, while in BA2-S1 there were up and downregulated genes, and in BA2-S2 there were only downregulated genes. Cell wall pectin esterase family had genes with differential expression only in the two sites with PDE, and it was only expressed in the period of occurrence of the PDE.

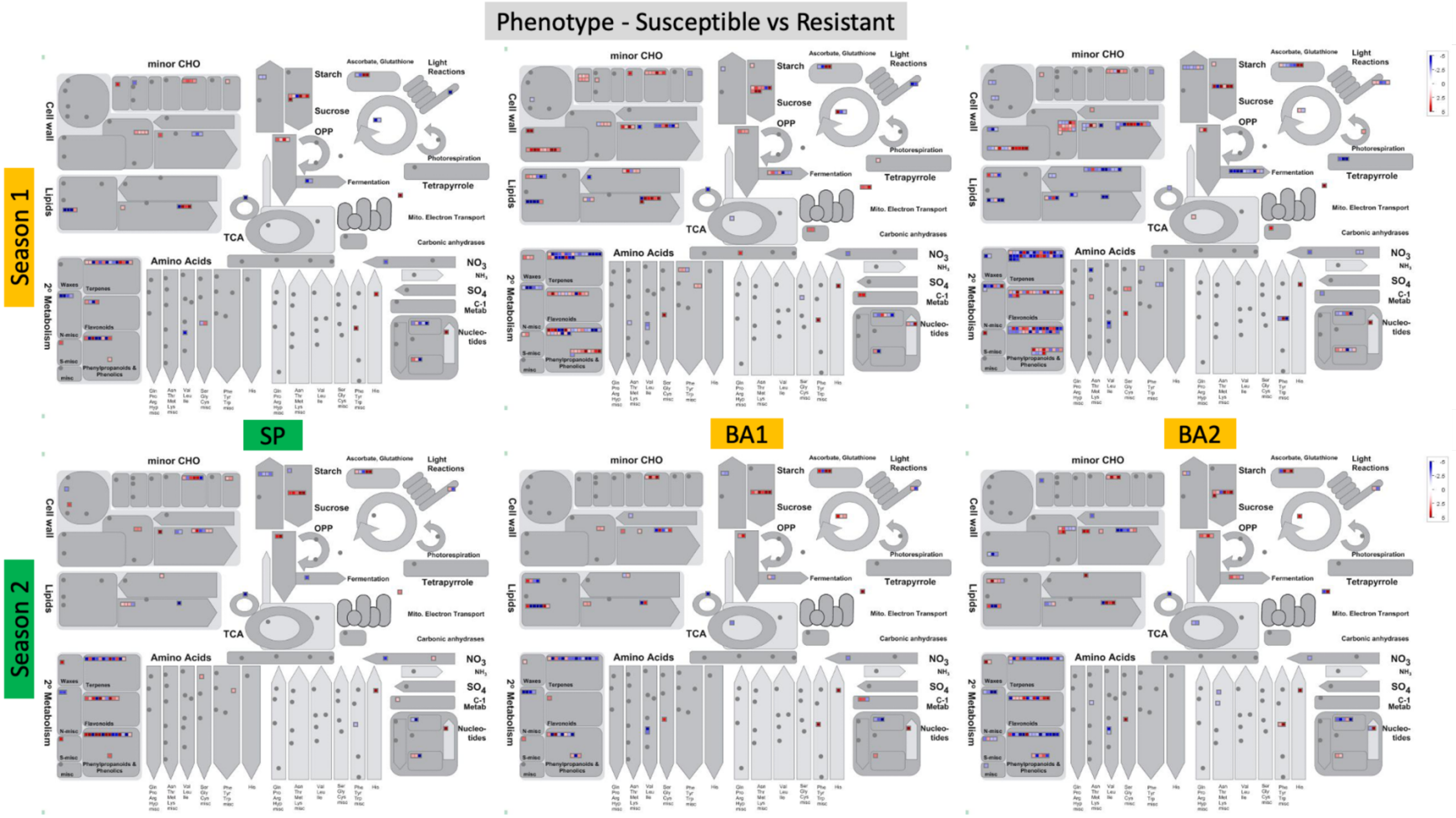

Figure 6. Metabolism overview showing differentially expressed genes between susceptible vs. resistant clones within each site (SP, BA1, BA2) during the seasons 1 and 2 (during and after PDE symptoms). Analyses and figure were generated in the MapMan software. Red squares depict upregulated genes and blue square downregulated genes in susceptible clones.

A metabolic pathway related to fermentation was altered in different sites. Acetaldehyde dehydrogenase and pyruvate descarboxylase genes, of the fermentation pathway, were more differentially expressed in sites BA1 and BA2 during season 1 (Figure 6). In the glycolysis pathway, pyruvate descarboxylase genes were more expressed in resistant clones in BA2-S1 (Figure 7). However, when we evaluated these genes in BA1-S1, a site that started to recover in season 1, the pyruvate descarboxylase genes were more expressed in susceptible clones. Moreover, when we evaluated the glycolysis pathway in the second (recovery) season of site BA2, susceptible genotypes induced pyruvate descarboxylase genes. Comparing the seasons in susceptible genotypes, there was a higher expression of L-lactate dehydrogenase genes related to lactate production during the disorder (Season 1) (Figure S4). On the other hand, resistant clones seem to activate the ethanol production, through a pyruvate descarboxylase gene, during the season S1 with PDE compared to the recovery season S2 (Figure S4).

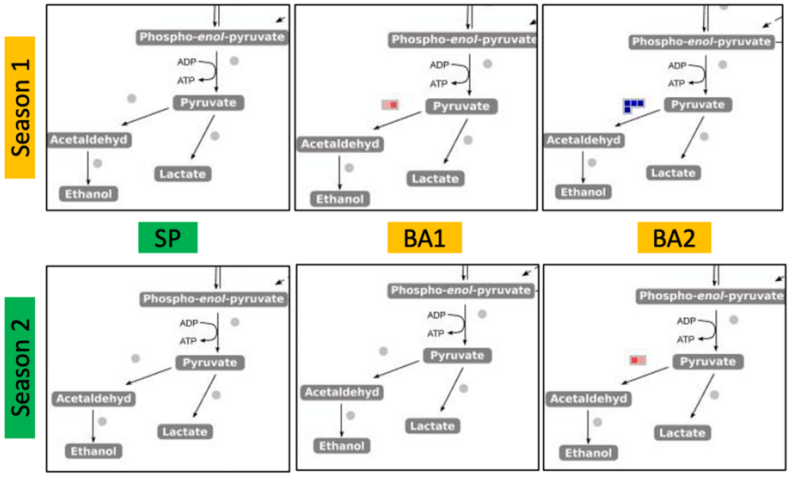

Figure 7. Glycolysis or fermentation pathway with differentially expressed genes in susceptible vs. resistant clones within each site (SP, BA1, BA2) and seasons 1 and 2 (during and after PDE symptoms). Analyses and figure were generated in the MapMan software. Red squares depict upregulated genes and blue square downregulated genes in susceptible clones.

An important result that agreed with the GO enriched categories was the differential expression of genes in secondary metabolism (Figure 8). In general, some pathways had increased expression in both susceptible and resistant clones, such as phenylpropanoids, lignin and lignans. Genes related to the production of terpenoids, and alkaloids were more expressed in the resistant clones during

the first season in sites with PDE. In contrast, simple phenols genes were activated in susceptible genotypes during the first season. Several heat shock proteins (HSP) were activated in both susceptible and resistant clones. Comparison between susceptible and resistant clones in different locations showed the influence of the disorder in HSP.

Differential gene expression between full sibling clones and seasons

Contrasts between full sibling clones C5 vs. C2 showed the activation of pathways like those observed in the comparison of susceptible vs. resistant clones. However, more DEGs were observed in these pathways when using the contrast C5 vs. C2 at the individual sites (Figure S5). For example, cell walls related pathways in BA2-S1 have large amount of DEGs when we compare only the comparison between phenotypes. TCA and carbonic anhydrases also were highligthed with more genes expressed in clones' comparison.

Gene expression and resistance level

When we analyzed specific genes, we were able to find candidates associated with the level of resistance of the clones. A gene of tocopherol pathway, PDS-1 (Eucgr.A02023), related to the production of p-hydroxyphenylpyruvate dioxygenase (HPPDase), was significantly more expressed in resistant genotypes in all seasons. In addition, its expression level was correlated with the degree of tolerance of the clones to PDE (Figure 9). Similar pattern was observed in other genes, such as ascorbate peroxidase 3 (Eucgr.E00522) and glutathione-S-transferase 6 (Eucgr.H01166), which were more expressed in resistant clones. Many genes related to gamma-irradiation and mitomycin c induced (GMI1) were more expressed in susceptible clones (Figure S6). These genes are candidates of molecular markers of the resistance/susceptibility to PDE.

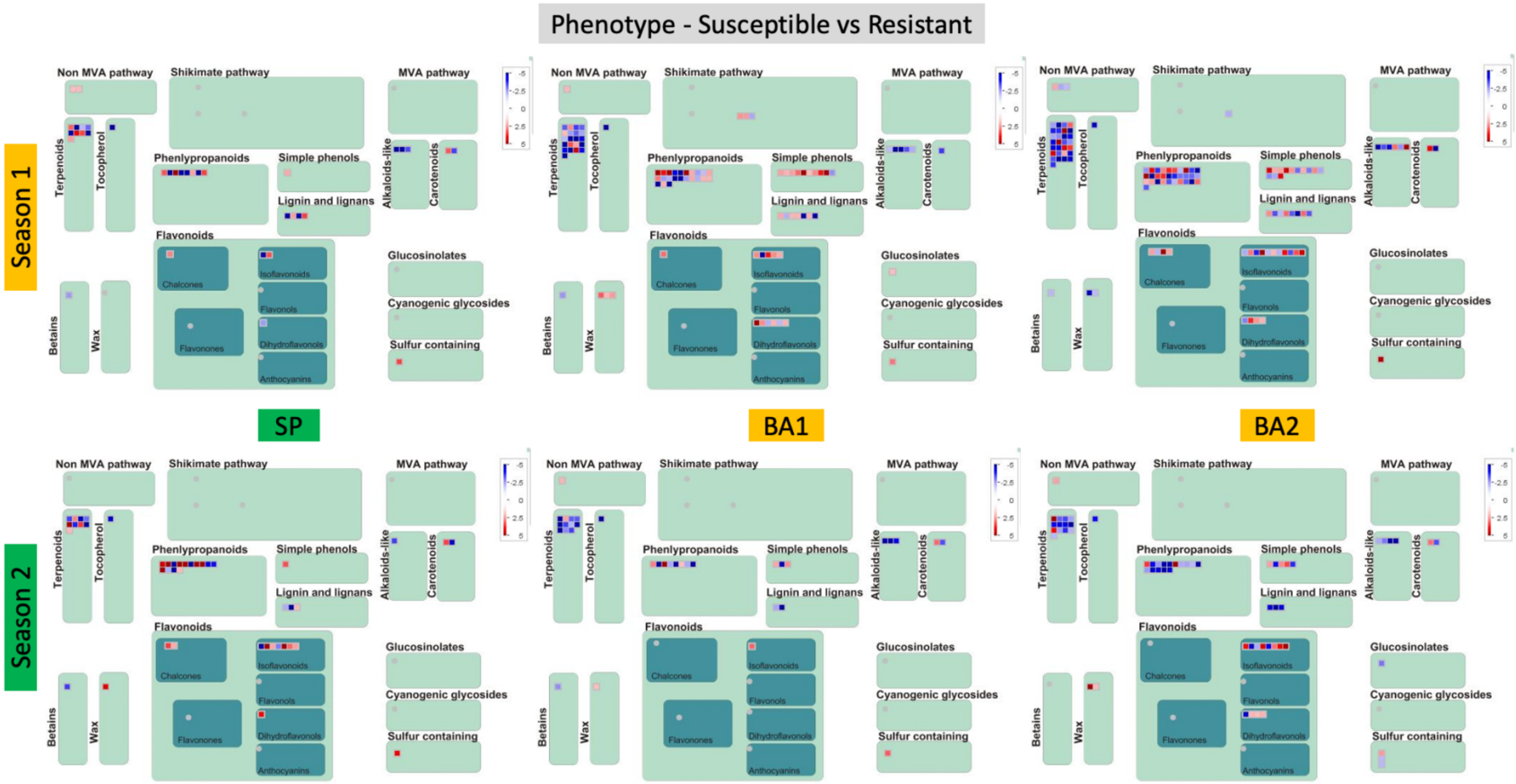

Figure 8. Secondary metabolism pathway showing differentially expressed genes between susceptible vs. resistant clones within each ambientes (SP, BA1, BA2) during the seasons 1 and 2 (during and after PDE symptoms). Analyses and figure were generated in the MapMan software. Red squares depict upregulated genes and blue squares downregulated genes in susceptible clones.

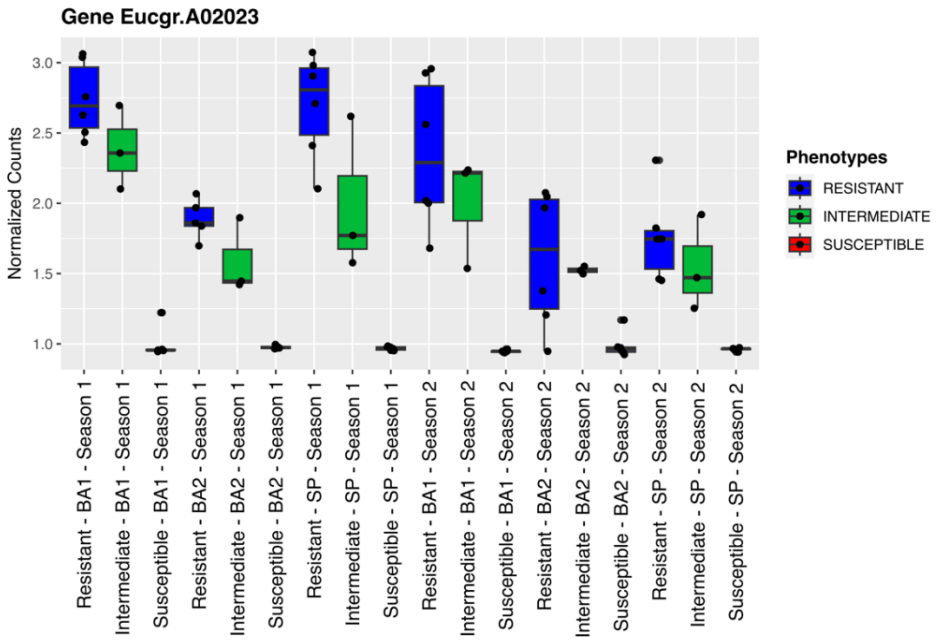

Figure 9. PDS-1 gene (Eucgr.A02023) expression difference among Susceptible and Resistant clones. Counts were normalized using the rlog normalization method.

Discussion

Physiological Disorder of *Eucalyptus* (PDE) is a complex stress. The possibility of the causal agent being a singular environmental variable, interaction of several abiotic factors or even biotic factors cannot be ruled out when we analyze the results obtained in this study. From the comparison of the metabolism of susceptible and resistant clones to PDE, many genes and metabolic pathways were activated, making it difficult to pinpoint any cause. Understanding the metabolic regulation of plants to each type of stress is a challenge (Da Ros et al. 2023). It is known that many signaling pathways and hormones are commonly regulated in plants under various types of stress (Da Ros et al. 2023, Medina et al. 2019). For example, hormone molecules that signal stress were altered in *Eucalyptus* clones subjected to water stress (Teshome et al. 2023). Under stress caused by PDE hormone as ABA, ethylene, and jasmonates were expressed. The higher number of DEGs in BA1-S1 and BA2-S1 (Figure 3) and the increase of DEG involved in signal transduction pathway may be related to a plant defense response to the PDE causal agent.

In this study, genes involved in plant adaptation to stressful conditions were differentially regulated. The activation of reproduction pathways that occurred in resistant clones, as well as the expression of thioredoxin gene family responsible for maintaining cellular redox homeostasis are

examples of a possible adaptive response (Meyer et al. 2008, Nuruzzaman et al. 2012). Thioredoxin is known to be involved in the response to different types of stress, adaptability and development in plants. Exopeptidase activity during plant recovery may also be related to an adaptation of clones to PDE conditions. Exopeptidases have proteolytic action, hydrolyzing proteins into amino acid residues (Miazek et al. 2008). Control of proteolysis is studied as one of the post-transcriptional factors in plant gene regulation under stress and functions in plant immunity (Godson et al. 2021, Miazek et al. 2008). Proteins degraded or inactivated by stress can be fragmented, recycled for the synthesis of new secondary metabolites or as energy sources for stress adaptation.

In this study, secondary metabolites were more activated in resistant clones under stress conditions, when compared to susceptible clones. Phenolic compounds are examples of secondary metabolites considered in many plants as protective agents against stress caused by pathogens and ultraviolet radiation (Dai et al. 2010, Loreto et al. 2010). Furthermore, the accumulation of toxic molecules like ROS, due to stress, can be toxic for plants. Proteases help eliminate these molecules, preventing damage to plant tissues (Godson et al. 2021). In this study, serine carboxypeptidase-like protease (SCLP) genes were expressed at sites with PDE. It is known that SCLP genes are involved in oxidative stress tolerance, indicating that the gene may be involved in the regulation of defense responses against oxidative stress and pathogen infection (Xu et al. 2021).

Genes related to energy machinery, such as those involved in the citric acid cycle, are important regulators identified in plants under stress. The aconitate gene were differentially expressed in sites under stress. The products of these genes are found in the mitochondria and can act in several metabolic pathways that generate ATP and developmental processes, such as programmed cell death, oxidative stress, and hypersensitive response (Kesawat et al. 2022). Also related to energy sources, chlorophyllases were regulated in the study. These enzymes catalyze the degradation of chlorophyll into various products and can be stimulated by ethylene under flooding conditions (Panda et al. 2021). Chlorophyllases play an important role during leaf senescence, which may explain why this gene expression is higher in periods when PDE occurs. It is known that the reduction in chlorophyll production can be an indicator of stress due to excess light. Excess light can lead to the production of ROS (Panda et al. 2021). Thus, in our results, the resistant clones showed higher expressions of the tetrapyrole pathway during incidence of the disorder, which indicates a possible regulation of the amount of chlorophyll to avoid damage due to high light incidence.

ROS can be considered molecules that lead to programmed cell death (PCD) in plants (Demidchik 2015). One of the characteristics of plants that induce programmed cell death is mitochondrial condensation, disruption of organelles and chromatin condensation (Latrasse et al. 2016, Rybaczek et al. 2015, Díaz-Tielas et al. 2012). GO terms associated with chromosomal condensation at A3-S1 may be related to these events. Many gamma-irradiation and mitomycin c induced genes with double strand break repair protein function (Böhmdorfer et al. 2011) were upregulated in the susceptible clones, indicating DNA strands breakage and activation of the DNA repair mechanisms, which are characteristic of PCD (Latrasse et al. 2016).

Activation of the fermentation pathway in sites with PDE might be a result of the heavy rains present in these locations. However, the fermentation pathway can be activated in aerobically conditions activating the production of known anaerobic proteins. Enzymes like alcohol dehydrogenase and pyruvate decarboxylase act as a pyruvate regulator and associated with plant growth in aerobic condition (Panda et al. 2021, Ventura et al. 2020, Zabalza et al. 2009). Under oxygen limitation, free radicals are transformed into ROS when it becomes available again. It is known that the PDE occurs in places with large variations in rainy and dry periods, which cannot be ruled out as an aggravating factor for the disorder. Another point to highlight is the degree of toxicity between alcohol and lactate. At cellular pH, alcohol is less aggressive compared to lactate, since it does not alter cellular pH (Miro et al. 2013). In this study, when comparing the metabolism of *Eucalyptus* between seasons, it was observed that susceptible genotypes in BA2 (Figure S3) express more L-lactate dehydrogenase in the first season (during the disorder). On the other hand, resistant clones activated the pyruvate decarboxylase. Some studies show that activation of the lactate pathway may be the beginning of alcoholic fermentation. A switch from lactate production to lower toxicity ethanol is performed to tolerate anoxic conditions in more flooding resistant plants (Bailey-Serres et al. 2008).

Incidence of PDE is higher in regions with high temperature and high air humidity (e.g. BA1 and BA2). Both factors are known to be involved in reduction of evapotranspiration in plants (Chia et al. 2022). Also These two factors are known to induce production of ROS and heat shock proteins (Georgii et al. 2017). Acetaldehyde dehydrogenase (ALDH), in addition to being the key enzyme in the anaerobiosis process, is transformed into acetate as part of the TCA cycle when reoxygenation occurs (Bailey-Serres et al. 2008). ALDH, in addition to abiotic stress tolerance, acts in male sterility restoration, embryo development and seed viability and maturation (Kotchoni

et al. 2010). Additionally, ALDHs act as “aldehyde scavenger” during lipid peroxidation to mitigate oxidative/electrophilic stress caused by ROS (Singh et al. 2013).

Phenylpropanoids can be related to lignin production (Naoumkina et al. 2010), and cell wall modifying genes were highly expressed in clones under PDE conditions. It is known that production of aerenchyma is an important characteristic of hypoxia-resistant plants (Medina et al. 2019). The assembly of aerenchyma requires cell wall modification. However, alterations in the expression of genes related to the cell wall under conditions of high and low temperatures and CO_2 availability have already been reported (Feltrim et al. 2022, Le Gall et al. 2015). Therefore, the gene expression changes in cell wall observed in our study may be a response to climate changes. Genes related to heat stress (heat shock proteins) were also activated in plant metabolism under PDE conditions. It is known that the regions where the disorder occurs have high average annual temperatures. Therefore, heat stress may be an aggravating factor for the occurrence of the disorder.

Observation of the metabolic pathways and genes differentially expressed in *Eucalyptus* plants during and after the incidence of PDE is important for understanding the disorder. Knowledge of the metabolic changes of plants under different stresses and comparison with the pathways activated in *Eucalyptus* clones that suffer from the physiological disorder can help shed light on the causes of this stress. Our study indicates that the physiological disorder causes widespread metabolic changes in several pathways of *Eucalyptus*. Clones affected by the PDE showed changes in glycolysis, cell wall, secondary metabolites and stress signaling pathways when compared to the resistant clones during and after the incidence of the disorder in different sites. Genes related to antioxidants were found to be more expressed in resistant clones. This study provides information that can illuminate the molecular mechanisms involved in PDE resistance and create new hypotheses for the PDE cause.

Supplementary data

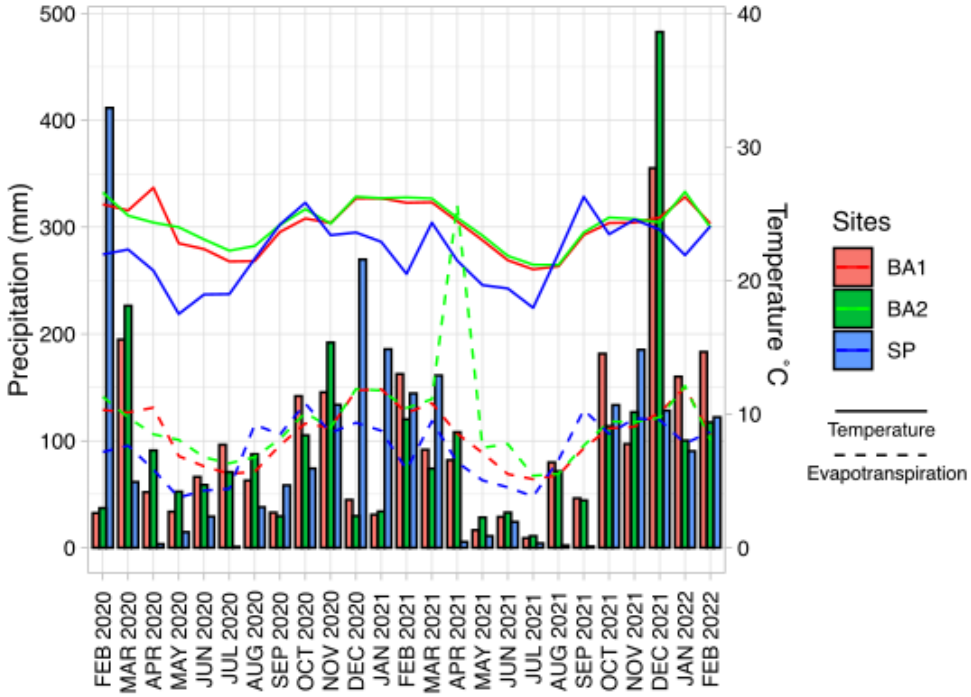

Figure S1. Climate data in sites with (BA1 e BA2) and without DFE (SP).

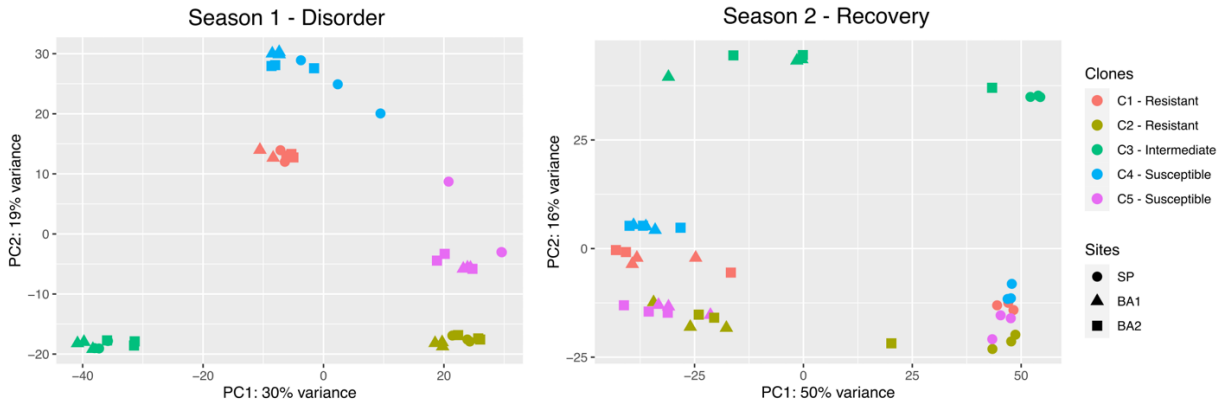

Figure S2. Principal components analysis with the genotype samples in different sites and seasons

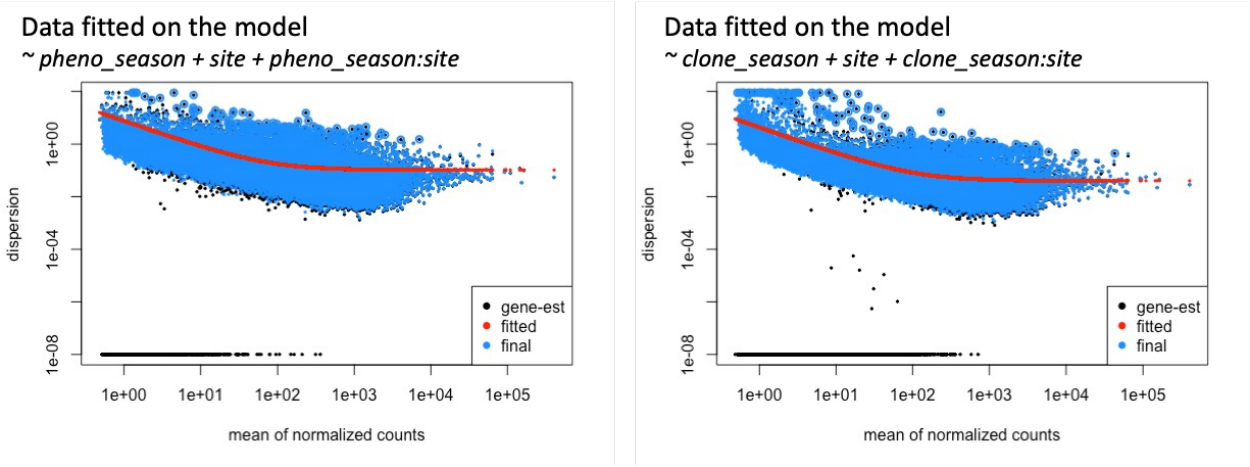

Figure S3. Models fitting analyses using the designs *~phenotype_season + site + phenotype_season:site* and *~clone_season + site + clone_season:site.*

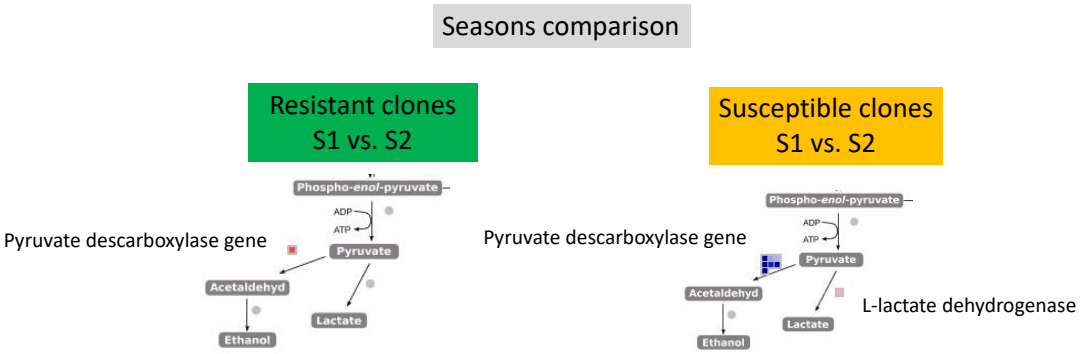

Figure S4. Comparison between seasons for resistant and susceptible clones. Analyses and figure generated in the MapMan software. Red squares are related to upregulated genes and blue square are related to downregulated genes.

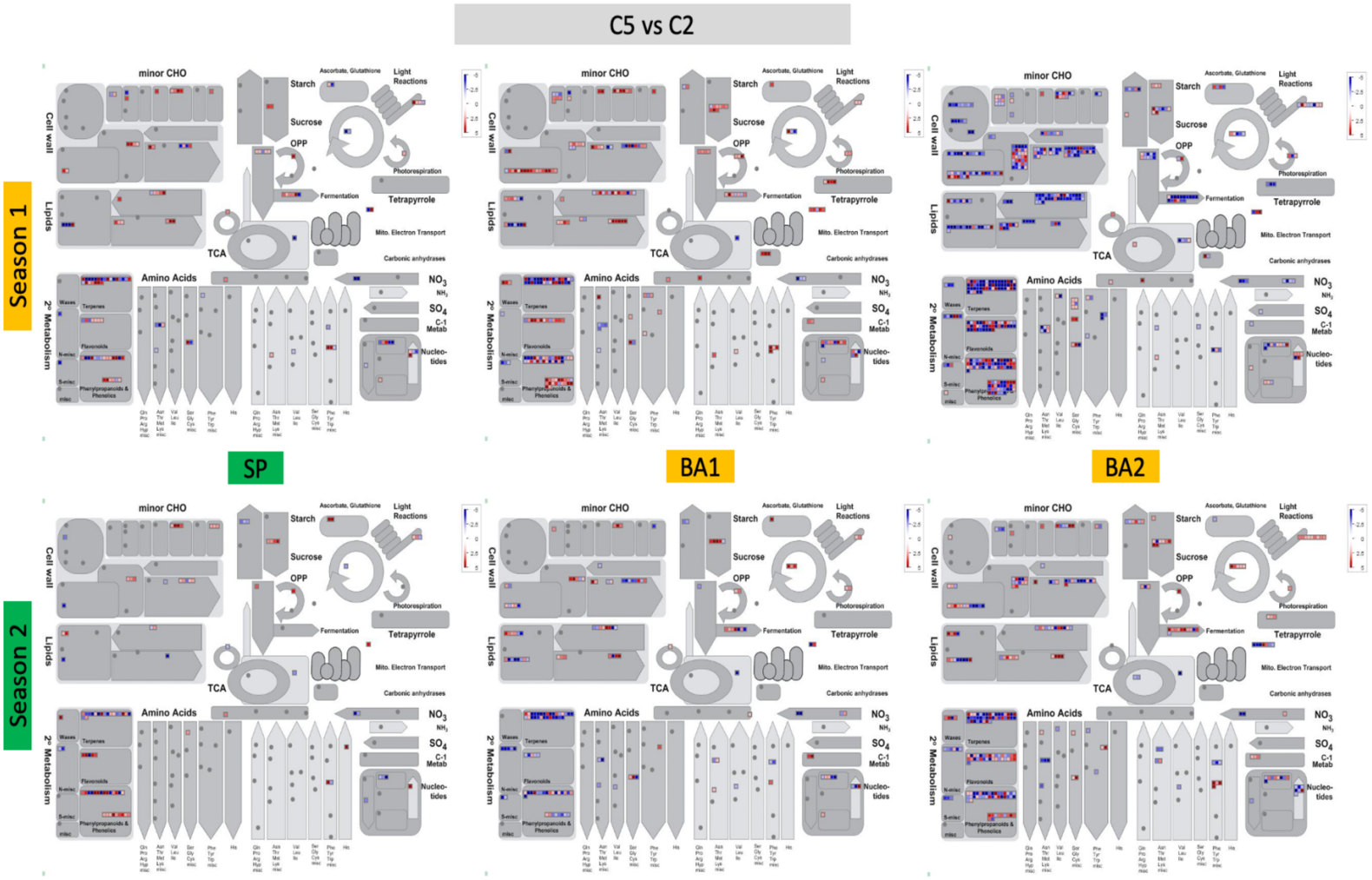

Figure S5. Metabolism overview showing differentially expressed genes between clone C5 vs. clone C2 within each site (SP, BA1, BA2) during the seasons 1 and 2. Analyses and figure generated in the MapMan software. Red squares are related to upregulated genes and blue square are related to downregulated genes.

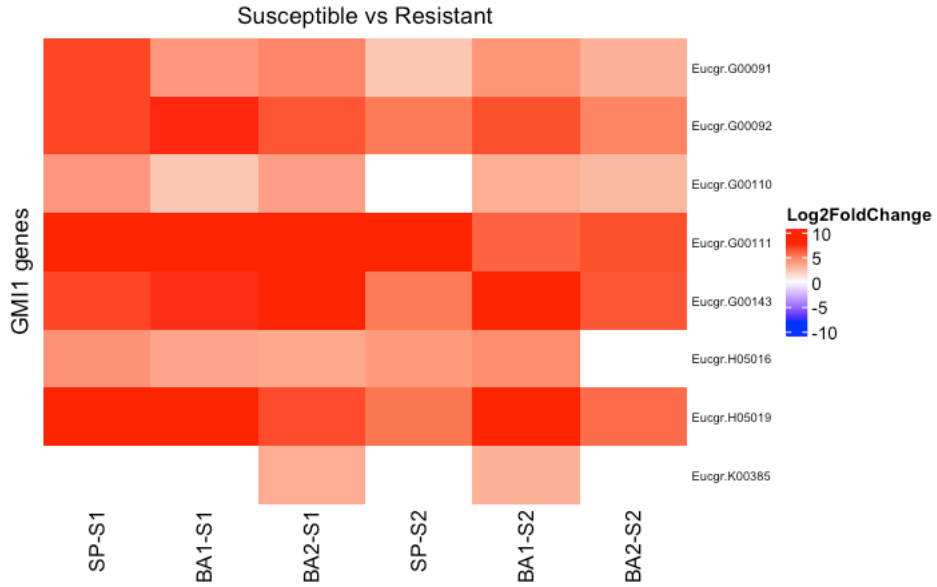

Figure S6. Heatmap showing differentially expressed genes GMI1 genes between Susceptible vs. Resistant clones within each site (SP, BA1, BA2) during the seasons 1 and 2.

References

Almeida A, Ribeiro A, Leite F (2013) Relação entre a seca dos ponteiros do eucalipto e o clima no vale da bacia hidrográfica do rio doce. Engenharia Ambiental 10:5–13.

Assis TF, Abad JIM, Aguiar AM (2015) Melhoramento Genético do Eucalipto. In: Schumacher MV, Vieira M (eds) Silvicultura do eucalipto no Brasil. UFSM, Santa Maria-RS, pp 225–247.

Bailey-Serres J, Voesenek LACJ (2008) Flooding stress: Acclimations and genetic diversity. Annu Rev Plant Biol 59:313–339.

Böhmdorfer G, Schleiffer A, Brunmeir R, Ferscha S, Nizhynska V, Kozák J, Angelis KJ, Kreil DP, Schweizer D (2011) GMI1, a structural-maintenance-of-chromosomes-hinge domain-containing protein, is involved in somatic homologous recombination in Arabidopsis. Plant Journal 67:420–433.

Bueno IGA, Picoli EA de T, Isaias RM dos S, Lopes-Mattos KLB, Cruz CD, Kuki KN, Zauza EAV (2020) Wood anatomy of field grown eucalypt genotypes exhibiting differential dieback and water deficit tolerance. Curr Plant Biol 22:2214–6628.

Câmara AP, Oliveira JTS, Bobadilha GS, Vidaurre GB, Tomazello Filho, M. Soliman EP (2018) Physiological disorders affecting dendrometric parameters and *Eucalyptus* wood quality for pulping wood. Cerne 24:27–34.

Chia SY, Lim MW (2022) A critical review on the influence of humidity for plant growth forecasting. IOP Conf. Ser.: Mater. Sci. Eng. 1257

Da Ros L, Bollina V, Soolanayakanahally R, Pahari S, Elferjani R, Kulkarni M, Vaid N, Risseuw E, Cram D, Pasha A, Esteban E, Konkin D, Provart N, Nambara E, Kagale S (2023) Multi-omics atlas of combinatorial abiotic stress responses in wheat. The Plant Journal 1-18.

Dai J, Mumper RJ (2010) Plant phenolics: Extraction, analysis and their antioxidant and anticancer properties. Molecules 15:7313–7352.

Demidchik V (2015) Mechanisms of oxidative stress in plants: From classical chemistry to cell biology. Environ Exp Bot 109:212–228.

Dianese JG, Haridasan M, Moraes TSA (1984) Tolerance to 'Mal do Rio Doce', a major disease of *Eucalyptus* in Brazil. Tropical Pest Management 30 3: 247–252.

Díaz-Tielas C, Graña E, Sotelo T, Reigosa MJ, Sánchez-Moreiras AM (2012) The natural compound trans-chalcone induces programmed cell death in *Arabidopsis thaliana* roots. Plant Cell Environ 35:1500–1517.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Feltrim D, Gupta B, Gundimeda S, Kiyota E, Júnior APD, Cintra LC, Mazzafera P (2022) Exposure of *Eucalyptus* to varied temperature and CO_2 has a profound effect on the physiology and expression of genes related to cell wall formation and remodeling. Tree Genet Genomes 18

Ferreira FA (1989) Doenças abióticas do eucalipto. In F. A. Ferreira (Ed.), Patologia Florestal 1 ed., p. 570.

Florêncio GWL, Martins FB, Fagundes FFA (2022) Climate change on *Eucalyptus* plantations and adaptive measures for sustainable forestry development across Brazil. Ind Crops Prod 188

Georgii E, Jin M, Zhao J, Kanawati B, Schmitt-Kopplin P, Albert A, Winkler JB, Schäffner AR (2017) Relationships between drought, heat and air humidity responses revealed by transcriptome-metabolome co-analysis. BMC Plant Biol 17

Godson A, Van Der Hoorn RAL (2021) The front line of defence: A meta-analysis of apoplastic proteases in plant immunity. J Exp Bot 72:3381–3394.

Harguindeguy I, Castro GF De, Novais SV, Vergutz L, Araujo W, Novais R (2018) Physiological Responses to Hypoxia and Manganese in Eucalyptus Clones with Differential Tolerance to Vale do Rio Doce Shoot Dieback. Rev Bras Cienc Solo 42:1–16.

Jardim JM, Jardim CM, Colodette JL (2018) Understanding the pulping and bleaching perfomances of *Eucalyptus* woods affected by physiological disturbance. Tappi Journal 17 2: 633-642.

Kesawat MS, Kherawat BS, Ram C, Singh A, Dey P, Gora JS, Misra N, Chung SM, Kumar M (2022) Genome-Wide Identification and Expression Profiling of Aconitase Gene Family Members Reveals Their Roles in Plant Development and Adaptation to Diverse Stress in *Triticum aestivum*

L. Plants 11

Kotchoni SO, Jimenez-Lopez JC, Gao D, Edwards V, Gachomo EW, Margam VM, Seufferheld MJ (2010) Modeling-dependent protein characterization of the rice aldehyde dehydrogenase (ALDH) superfamily reveals distinct functional and structural features. PLoS One 5

Latrasse D, Benhamed M, Bergounioux C, Raynaud C, Delarue M (2016) Plant programmed cell death from a chromatin point of view. J Exp Bot 67:5887–5900.

Le Gall H, Philippe F, Domon JM, Gillet F, Pelloux J, Rayon C (2015) Cell wall metabolism in response to abiotic stress. Plants 4:112–166.

Leite FP, Novais RF, Silva IR, Félix NB, Neves J, Medeiros A, MC V, Villani E (2014) Manganese accumulation and its relation to '*Eucalyptus* Shoot Blight in the Vale do Rio Doce'. Rev Bras Cienc Solo 38:193–204.

Loreto F, Schnitzler JP (2010) Abiotic stresses and induced BVOCs. Trends Plant Sci 15:154–166.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:1–21.

Medina EF, Mayrink GCV, Dias CR, Vital CE, Ribeiro DM, Silva IR, Merchant A (2019) Physiological and biochemical responses of *Eucalyptus* seedlings to hypoxia. Ann For Sci 76

Metzker ML (2010) Sequencing technologies the next generation. Nat Rev Genet 11:31–46.

Meyer Y, Siala W, Bashandy T, Riondet C, Vignols F, Reichheld JP (2008) Glutaredoxins and thioredoxins in plants. Biochim Biophys Acta Mol Cell Res 1783:589–600.

Miazek A, Zagdańska B (2008) Involvement of exopeptidases in dehydration tolerance of spring wheat seedlings. http://merops.

Miro B, Ismail AM (2013) Tolerance of anaerobic conditions caused by flooding during germination and early growth in rice (*Oryza sativa* L.). Front Plant Sci 4

Naoumkina MA, Zhao Q, Gallego-Giraldo L, Dai X, Zhao PX, Dixon RA (2010) Genome-wide analysis of phenylpropanoid defence pathways. Mol Plant Pathol 11:829–846.

Nuruzzaman M, Sharoni AM, Satoh K, Al-Shammari T, Shimizu T, Sasaya T, Omura T, Kikuchi S (2012) The thioredoxin gene family in rice: Genome-wide identification and expression profiling under different biotic and abiotic treatments. Biochem Biophys Res Commun 423:417–423.

Panda D, Barik J (2021) Flooding Tolerance in Rice: Focus on Mechanisms and Approaches. Rice Sci 28:43–57.

Rybaczek D, Musialek MW, Balcerczyk A (2015) Caffeine-induced premature chromosome condensation results in the apoptosis-like programmed cell death in root meristems of *Vicia faba*. PLoS One 10

Singh S, Brocker C, Koppaka C, Chen Y, Jackson BC, Matsumoto A, Thompson DC, Vasiliou V (2013) Aldehyde dehydrogenases in cellular responses to oxidative/electrophilic stress. Free Radical Biology and Medicine, 56: 89-10. http://dx.doi.org/10.1016/j.freeradbiomed.2012.11.010

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One 6

Teshome DT, Zharare GE, Ployet R, Naidoo S (2023) Transcriptional reprogramming during recovery from drought stress in *Eucalyptus grandis*. Tree Physiol 43:979–994.

Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res 45:W122–W129.

Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M (2009) A guide to using MapMan to visualize and compare Omics data in plants: A case study in the crop species, Maize. Plant Cell Environ 32:1211–1229.

Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. Trends Genet 30:418–426.

Ventura I, Brunello L, Iacopino S, Valeri MC, Novi G, Dornbusch T, Perata P, Loreti E (2020) Arabidopsis phenotyping reveals the importance of alcohol dehydrogenase and pyruvate decarboxylase for aerobic plant growth. Scientific Reports 10:1-14

Xu X, Zhang L, Zhao W, Fu L, Han Y, Wang K, Yan L, Li Y, Zhang XH, Min DH (2021) Genome-wide analysis of the serine carboxypeptidase-like protein family in *Triticum aestivum* reveals TaSCPL184-6D is involved in abiotic stress response. BMC Genomics 22

Zabalza A, Dongen JT, Forehlich A, Oliver SN, Faix B, Gupta KJ, Schmalzin E, Orcaray L, Rovuela M, Geigenberger P (2009) Regulation of Respiration and Fermentation to Control the Plant Internal Oxygen Concentration. Plant physiology 149:1087-1098