Research, Society and Development, v. 11, n. 6, e45311629227, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i6.29227

Marker-assisted selection and protein identification in *Eucalyptus*

Seleção assistida por marcadores e identificação de proteínas em *Eucalyptus*

Selección asistida por marcadores e identificación de proteínas en *Eucalyptus*

Received: 04/14/2022 | Reviewed: 04/24/2022 | Accept: 05/01/2022 | Published: 03/05/2022

Crislaine Costa Calazans
ORCID: https://orcid.org/0000-0002-6100-0608
Federal University of Sergipe, Brazil
E-mail: cris.calazans@yahoo.com.br
Valdinete Vieira Nunes
ORCID: https://orcid.org/0000-0002-0272-0971
Federal University of Sergipe, Brazil
E-mail: val.ufs@gmail.com
Juliana Lopes Souza
ORCID: https://orcid.org/0000-0002-7514-0361
Federal University of Sergipe, Brazil
E-mail: juliana_lopes_souza@live.com
Natali Aparecida Santana
ORCID: https://orcid.org/0000-0003-3638-9348
Federal University of Sergipe, Brazil
E-mail: nataliknd@gmail.com
Fernanda Evangelista de Almeida
ORCID: https://orcid.org/0000-0002-6297-9731
Federal University of Sergipe, Brazil
E-mail: fernanda.evangelista14@gmail.com
Renata Silva-Mann
ORCID: https://orcid.org/0000-0001-5993-3161
Federal University of Sergipe, Brazil
E-mail: renatamann@academico.ufs.br

Abstract

Marker-assisted selection aims to improve quantitative traits in *Eucalyptus*, and molecular markers allow the identification of (the expression) of challenging alleles. Aiming to overcome this challenge our work presents a systematic review of the use of marker-assisted selection and candidate proteins to identify markers associated with traits *of Eucalyptus*. The research carried out a prospection of content obtained by keywords in the scientific databases Scopus and Web of Science. The bibliographic data generated was imported using the Bibliometrix package in R software. The nucleic acid sequences were prospected in the Genetic Sequence Database – GenBank. The proteins were prospected in UNIPROT and the 3D structures were obtained at the Swiss Bioinformatics Institute - SIB. The prospection resulted in 63 scientific articles after duplicates were removed, from 41 journals, with an average publication of 12.4 articles per year and an average of 25.2 citations per document. China and Brazil stand out in the number of publications. RAPD technique was the most used in studies to obtain alleles with traits of interest in *Eucalyptus*. The ISSR and SSR markers were highlighted for studies on genotype resistance. The sequences associated with resistance to galling organisms in *Eucalyptus* are related to *TAC1* transcription regulatory proteins, mediator 2; zinc finger protein 11, associated with the expressed genes *TAC1*, *At5g64680*, and *ZFP11*.
Keywords: Bibliometrics; Forest improvement; MAS; Teaching.

Resumo

A seleção assistida por marcadores visa melhorar características quantitativas em *Eucalyptus*, e marcadores moleculares permitem a localização de alelos com expressão de difícil identificação. Assim, apresenta-se uma revisão sistemática de artigos científicos relacionados ao uso de seleção assistida por marcadores e realiza-se a prospecção de proteínas candidatas a marcadores associados a características de interesse econômico em espécies do gênero *Eucalyptus*. A pesquisa foi realizada a partir da prospecção de conteúdos obtidos por palavras-chaves nas bases científicas Scopus e Web of Science. Os dados bibliográficos gerados foram importados utilizando o pacote Bibliometrix no software R. As sequências de ácidos nucleicos foram prospectadas no Genetic Sequence Database – GenBank, e as proteínas foram prospectadas no UNIPROT, e no Instituto Suíço de Bioinformática - SIB foram obtidas as estruturas 3D das proteínas. Foram prospectados 74 artigos científicos, dos quais foram retiradas as duplicatas, totalizando 63 documentos de 41 revistas, com publicação média de 12,4 artigos por ano e média de 25,2 citações por documento. A China e o Brasil se destacam em número de publicações. A técnica de RAPD foi a mais utilizada em estudos para obtenção de alelos com características de interesse em *Eucalyptus*. Os marcadores ISSR e

SSR foram destaques para genótipos resistentes a pragas e doenças. As sequências associadas a resistência a organismos galhadores em *Eucalyptus* estão relacionadas a proteínas reguladoras de transcrição *TAC1*, proteínas associadas ao mediador 2, proteína de dedo de zinco 11, as quais estão associadas aos genes expressos *TAC1*, *At5g64680* e *ZFP11*.

Palavras-chave: Bibliometria; Melhoramento florestal; SAM; Ensino.

Resumen

La selección asistida por marcadores tiene como objetivo mejorar los rasgos cuantitativos en *Eucalyptus*, y los marcadores moleculares permiten la identificación (la expresión) de alelos desafiantes. Con el objetivo de superar este desafío, nuestro trabajo presenta una revisión sistemática en el uso de selección asistida por marcadores y proteínas candidatas para identificar marcadores asociados con rasgos de *Eucalyptus*. La investigación se llevó a cabo una prospección de contenido obtenido por palabras clave en las bases de datos científicas Scopus y Web of Science. Los datos bibliográficos generados fueron importados utilizando el paquete Bibliometrix en el software R. Las secuencias de ácidos nucleicos fueron prospectadas en la Base de Datos de Secuencias Genéticas – GenBank. Las proteínas fueron prospectadas en UNIPROT y las estructuras 3D fueron obtenidas en el Swiss Bioinformatics Institute - SIB. La prospección resultó en 63 artículos científicos después de la eliminación de duplicados, de 41 revistas, con una publicación promedio de 12,4 artículos por año y un promedio de 25,2 citas por documento. China y Brasil se destacan en el número de publicaciones. La técnica RAPD fue la más utilizada en los estudios para obtener alelos con caracteres de interés en *Eucalyptus*. Se destacaron los marcadores ISSR y SSR para estudios de resistencia de genotipos. Las secuencias asociadas con la resistencia a los organismos irritantes en *Eucalyptus* están relacionadas con las proteínas reguladoras de la transcripción TAC1, mediador 2; proteína con dedos de zinc 11, asociada con los genes expresados TAC1, At5g64680 y ZFP11.

Palabras clave: Bibliometría; Mejora forestal; SAM; Enseñanza.

1. Introduction

Eucalyptus, Myrtaceae, is one of Australia's most common plant genera, with over 700 species, most of which fall into two of eight subgenera: *Symphyomyrtus* (563 species) and *Eucalyptus* (124 species) (Nicolle, 2019). The species of this genus stand out for providing a greater return in biomass volume when compared to most native species in Brazil, as they have a fast growth rate and, therefore, are species widely used in the productive sector (Samal et al., 2021).

Due to growing environmental and economic pressure, the increase in demand for renewable fuels and products has given prominence to eucalyptus forests, as they represent the most abundant form of biomass in the world, consisting of lignocellulosic material. Also, the residues from this activity have the potential to be used for high yield of cellulose biomass and bioethanol (Jaleta et al., 2016; Trevorah et al., 2018).

In Brazil, there are about four million hectares of planted eucalyptus forest, and the success of these plantations is the result of detailed site selection, the use of intensive cultural practices, the selection of the best species and genotypes, and genetic breeding (Food and Agriculture Organization of the United Nations [FAO], 2021).

Different genotypes must be obtained and selected for genetic tree improvement, whether for increased productivity, tolerance to water restriction, or resistance to pests or diseases. Employing molecular markers is possible to identify loci associated with quantitative traits in different plant species. Molecular markers are divided into biochemical (protein markers) and DNA markers (Viana et al., 2012).

Molecular markers allow the identification of the genome segment related to a phenotype, contributing to genetic variance. These markers have the advantage of discriminating genotypes and assessing the genetic variability existing within and among individuals of the same species and different species (Souza, 2015; Souza et al., 2008). Some examples of DNA-based molecular markers are Restriction Fragment Length Polymorphism (RFLP) (Young & Tanksley, 1989), Amplified Fragment Length Polymorphism (AFLP) (Savelkoul et al., 1999; Silva-Mann et al., 2005), Random Amplified Polymorphic DNA (RAPD) (Williams et al., 1993), Inter Simple Sequence Repeat (ISSR) (Godwin et al., 1997; Ribeiro et al., 2017), Simple Sequence Repeat (SSR) (Pradeep Reddy et al., 2002), Diversity Arrays Technology (DArT) (Jaccoud et al., 2001), and Single Nucleotide Polymorphism (SNP) (Chagné et al., 2007).

The RAPD marker emerged from modifications of the Polymerase Chain Reaction (PCR), a technique that comprises the synthesis of millions of copies of a specific segment of DNA in the presence of the DNA polymerase enzyme. RAPD markers eliminate the need to previously know the sequence to be amplified, and it is mediated at the beginning of amplification by a 10-base nitrogen primer, a randomly amplified DNA polymorphism (Agarwal et al., 2008).

ISSR markers have the advantages of a high degree of polymorphism and reproducibility. It consists of amplifying DNA segments located between two identical microsatellite regions. It uses repeated DNA sequences (SSR) as a single primer in PCR reactions (Kumar et al., 2009) and is a cost-effective technique. DArT is a marker-based on microarray hybridization to genotype hundreds of loci in a single assay, applying to any species with no need for genome knowledge (Sansaloni et al., 2011).

Among the various applications of DNA-based molecular markers, Marker Assisted Selection (SAM) stands out as a method of selecting traits of economic interest based on information from molecular DNA markers. SAM considers the molecular pattern of individuals, that is, the genotype rather than just the phenotype, and is based on mapping and associating markers with genes that control a particular trait of interest (Goetze et al., 2017; Maldonado et al., 2017; 2018). The diverse molecular markers developed for many plant species allow the genetic identification of phenotypic traits using DNA-based approaches, including loci for quantitative traits (Quantitative Trait Toci - QTL), associative mapping (Mapping Association - AM) or Linkage Imbalance (Linkage Disequilibrium - LD) and genomic selection (Zhang et al., 2018).

Therefore, our study aims to investigate the main techniques used in marker-assisted selection for species of the genus *Eucalyptus* and to prospect proteins related to economically relevant traits from molecular data.

2. Methodology

A review of scientific articles on marker-assisted selection in Eucalyptus was carried out, evaluating publications on the scientific bases Scopus (http://www.scopus.com) and Web of Science (http://www.webofknowledge.com), using the terms: “assisted selection”, “markers”, “MAS”, “Eucalyptus” interspersed by the operators "AND" and "OR" and prospected in the title, abstract and/or keywords of scientific articles published and indexed in the bases until the search date.

Data referring to scientific publications were exported in BibTex format, and the information was extracted and analyzed using the Bibliometrix package of the R software (R Core Team, 2020). The variables were statistically analyzed, and the graphics were generated using the Biblioshiny platform from citation data, scientific collaboration, countries of origin, and keywords. From October 2020 to January 2022, research protocols based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). The PRISMA Statement and its extensions are a minimal evidence-based set of recommendations designed primarily to encourage transparent and complete communication. This set of guidelines was developed to assist authors and to ensure that all aspects of research are reported accurately and transparently (Sarkis-Onofre et al., 2021). The development of a protocol is an essential component of the systematic review process; it ensures that a systematic review is carefully planned and that what is planned is explicitly documented before the review begins, thus promoting consistent conduct of the review team, accountability, research integrity, and transparency of the eventually completed review (Moher et al., 2015). Therefore, this methodology was applied due to its application of eligibility criteria for the selection of articles. Articles classified in of the following three reasons were excluded:

Article not found

The article does not refer to the object of the study

Article outside the scope of the search, as detailed in Table 1

All articles generated after the search was refined were analyzed in total.

Table 1. Eligibility criteria were applied in a bibliometric study involving Marker Assisted Selection.

Scientific base	Search protocol	Chronology	Reasons for Exclusion
Scopus	SEARCH WITHIN (Article title, Abstract, Keywords): “assisted selection” AND (“markers” OR “MAS”) AND “Eucalyptus” Refined by: PUBLICATION DATE: (All years)	1945 – 2022	(1) article not found; (2) article does not refer to the study organism; (3) article outside the scope of the research.
Web of Science	TOPIC: “assisted selection” AND (“markers” OR “MAS”) AND “Eucalyptus” Refined by: TYPES OF DOCUMENTS: (ARTICLE) AND PUBLICATION DATE: (All years)		

Source: Authors (2022).

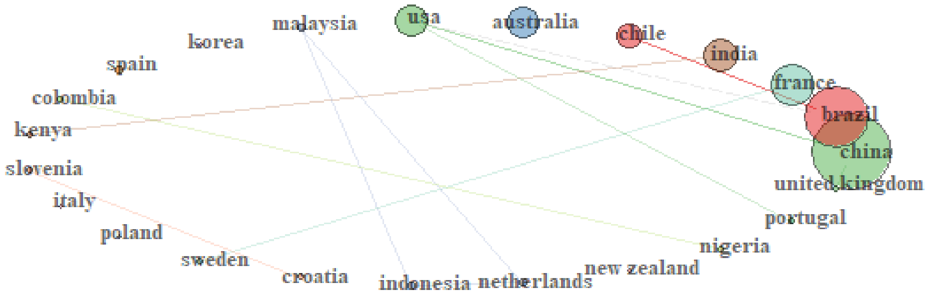
The sequences described in the publications were inserted as entries to identify sequence-related proteins and the relationship with expressed genes, and these nucleotide sequences were obtained in FASTA format. In the protein bank, UNIPROT (https://www.uniprot.org/), sequences in FASTA format were entered as input, and a BLAST was performed to obtain the corresponding proteins. From these proteins in FASTA format, it was possible to associate genes expressed from the model organism *Arabidopsis thaliana* using tblastn. The 3D structures of the proteins were accessed at the Swiss Institute of Bioinformatics—SIB (https://www.sib.swiss). The analysis of the biological profile of gene expression was performed on the e-Plant platform based on the model species *Arabidopsis thaliana* (L.) Heynh (Waese et al., 2017).

3. Results and Discussion

3.1 Scientific prospection

Seventy-four scientific articles from the Scopus (27) and Web of Science (47) databases were prospected. After duplicates were removed there was a total of 63 published documents distributed in 41 journals. As for the exclusion criteria, one article was excluded since it was impossible to find it in its entirety for analysis (reason 1). Twenty-six articles were excluded from the research because they did not deal with the species under study (reason 2); seven documents were excluded because they did not align with the theme (reason 3). There was an average of 12.4 publications per year for the topic in question and 25.2 citations per document. Of the total number of authors, 227 authors, 305 were co-authors, four wrote single-authored documents, and 223 were multi-author works. An average of 2,787 articles were referenced per year. China and Brazil stand out in publications with 19% and 18% of publications, respectively, followed by Australia (14%), France (13%), and the United States with 8% of the documents surveyed. Figure 1 shows the leading countries in collaboration between researchers, highlighting publications in partnership between researchers from the United States of America (USA) with China, Brazil, Portugal, and Chile.

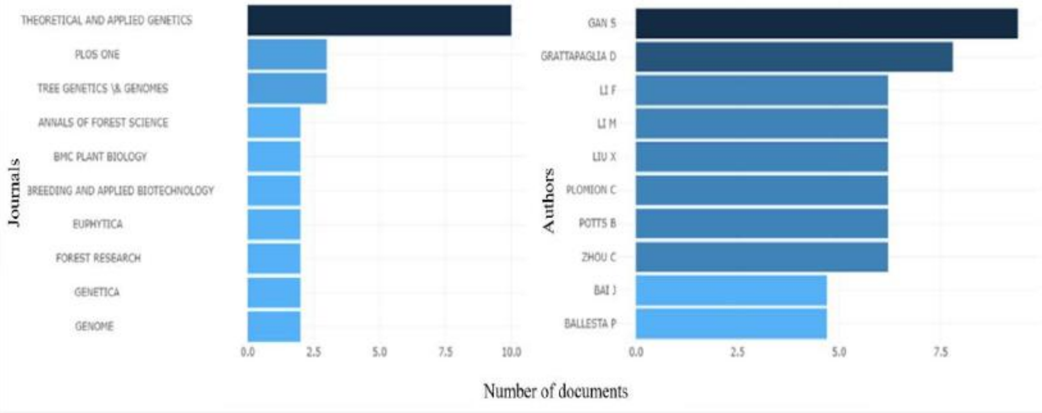
Figure 1. Collaboration between countries in many publications from 1945 to 2022.



Source: Authors (2022).

The Theoretical and Applied Genetics journal was the most cited, followed by Plos One and Tree Genetics and Genomes (Figure 2). Theoretical and Applied Genetics publishes original research and review articles in all critical areas of modern genetics, genomics, and plant biotechnology (Theoretical and Applied Genetics, 2022). The other journals highlighted in the research have a lower contribution in publications and present a publication scope of works in the areas of biotechnology, plant breeding of forest species, or genetics of forest species. The author Siming Gan stood out by developing works to build moderate density molecular maps for the genomes of *Eucalyptus urophylla* S. T. Blake and *Eucalyptus tereticornis* Smith using RAPD markers (Gan et al., 2003). Studies aimed at constructing these maps are essential for detecting quantitative trait loci (QTLs), marker-assisted selection (SAM), and comparative mapping based on the complete genome in *Eucalyptus* breeding programs. Dario Grattapaglia presents himself as one of the forerunners in the combined use of the "pseudo-testcross" mapping strategy and RAPD markers to map quantitative trait loci (QTLs) controlling characteristics related to vegetative propagation in *Eucalyptus* (Grattapaglia et al., 1996; Grattapaglia et al., 2004).

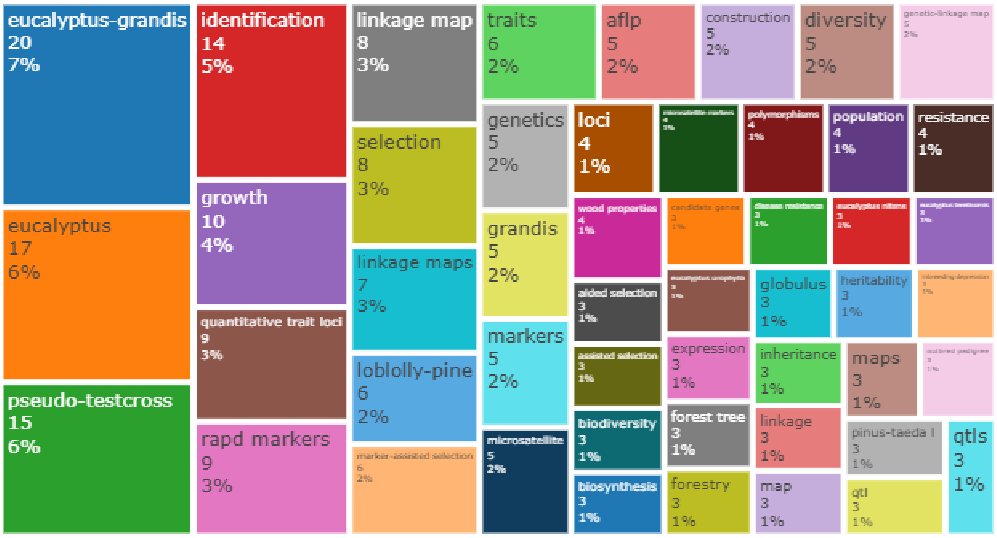
Figure 2. Main journals and authors of articles published in the Web of Science and Scopus databases on assisted selection and markers in the *Eucalyptus* genus from 1945 to 2022.



Source: Authors (2022).

In Figure 3, it is possible to observe the main thematic areas of publications related to the use of molecular markers in selecting desired characteristics in *Eucalyptus*. The terms in evidence were: “*Eucalyptus grandis*” (7%), “*Eucalyptus*” (6%), “pseudo-testcross” (6%), “identification” (5%), “growth” (4%), and “RAPD markers” and “quantitative trait loci” (3%).

Figure 3. Main thematic areas of articles published in the Web of Science and Scopus databases on assisted selection and markers in the *Eucalyptus* genus from 1945 to 2022.



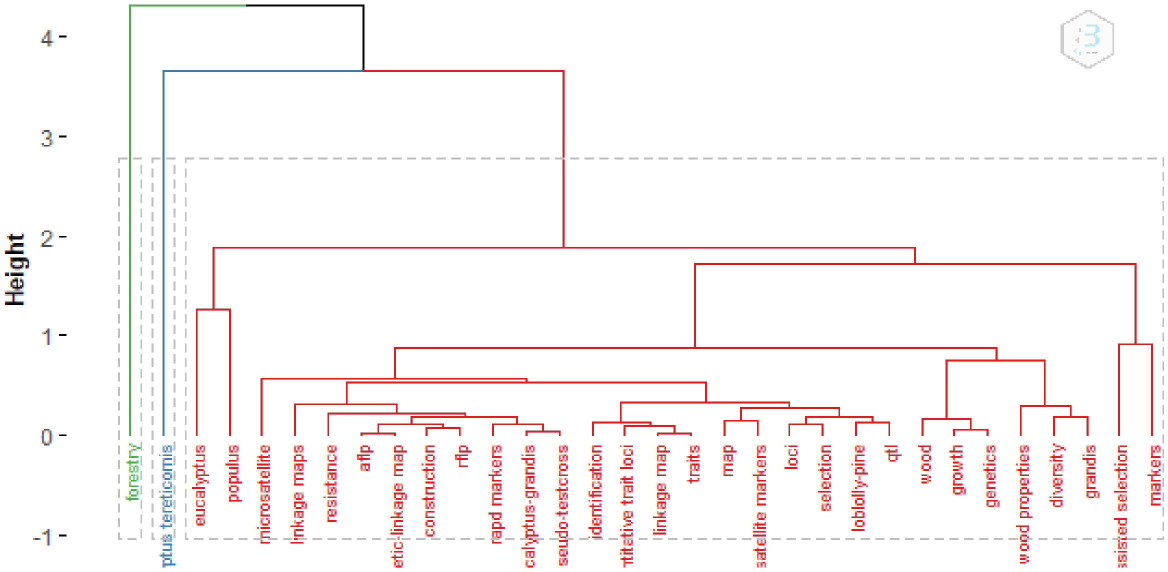
Source: Authors (2022).

The more significant number of publications containing the term "*Eucalyptus grandis*" may be related to the importance of the work entitled "The genome of *Eucalyptus grandis*" with the participation of the author Grattapaglia, which describes the genome of *E. grandis* and is the first reference to the order Myrtales (Naidoo et al., 2014), published in Nature in 2014 and cited in 539 articles.

AFLP markers are cited in works that combine the pseudo-testcross strategy used to generate genetic maps in clones of different *Eucalyptus* species (Marques et al., 1999). The use of the AFLP marker was also reported in evaluating wood quality components in the selection of genotypes (Arcade et al., 2002).

Quantitative trait loci at different ages are of particular interest in tree breeding programs. Essential practices for marker-assisted selection have been reported to identify genetic associations for growth in *Eucalyptus* using SSR markers (Maldonado et al., 2018). From the association of terms in a dendrogram, it was possible to identify the formation of three main groups: green, blue, and red. The red group is divided into subgroups, composing study themes (Figure 4).

Figure 4. Dendrogram formed from the most frequent terms in articles searching for marker-assisted selection in *Eucalyptus* from 1945 to 2022.



Source: Authors (2022).

The subgroups characterize the application of technologies aimed at the use of markers for (1) identification of resistance to diseases - "linkage maps", "resistance", "AFLP", "etic-linkage map", "construction", " RFLP", "RAPD markers", "*Eucalyptus grandis*", "pseudo-testcross"; (2) Map construction and QTL detection – "identification", "Qualitative trait loci", "linkage map", "traits", "map", "loci", "selection", "loblolly-pine", "QTL "; (3) Vegetative growth, wood properties and genetic diversity – "wood", "growth", "genetics", "wood properties", "diversity", "grandis".

The ISSR and SSR markers allowed the identification of information about the genetic control of resistance in plants and have great potential in detecting resistance alleles to pests and diseases in *Eucalyptus*. Cited markers differentiated plants that showed higher and lower tolerance to pests and diseases, showing that selection assisted by molecular markers can be a viable alternative for breeders (Gomes et al., 2019; Zhang et al., 2018).

With SAM, important advances were made in forest selection and tree improvement, especially in *Eucalyptus* breeding programs, according to the results observed in the literature. This technique allows for efficiency in the genotype selection process, eliminating the influence of the environment on the expression of the characteristics of interest (Gomes et al., 2019). This is because predictions are more accurate for traits controlled by a few quantitative trait loci (QTLs), which control a relatively large proportion of the phenotypic variation (Lee et al., 2017).

Early flowering and abundance of flowers have long been considered desirable traits in *Eucalyptus* breeding programs. Eucalyptus *cladocalyx* flowers provide a source of nectar for honey production in arid ecosystems (Contreras-Soto, Ballesta, Ruiz, & Mora, 2016). Single sequence repeat (SSR) markers have been associated with early flowering in *Eucalyptus* (Thamarus et al., 2002; Farro et al., 2013). Three *loci* associated with early flowering were identified associated with ISSR markers linked to early flowering and flower abundance of *E. cladocalyx*, accounting for 10-16% of the phenotypic variation, while two *loci* were responsible for 11-13% of the phenotypic variation in flowering intensity (Contreras-Soto et al., 2016).

Identifying genomic regions associated with complex traits can contribute to increasing the efficiency of small-scale breeding programs. Six significant genomic regions can be used as selection criteria for a genetic improvement of more than one trait (growth, flowering, and survival traits) (Ballesta et al., 2015). In identifying genes associated with disease resistance, an effective strategy is the mapping of many codominant multi-allelic microsatellite markers (SSR). The Embra125 marker has potential use in assisted selection, with a consequent increase in efficiency during the selection of rust-resistant genotypes in *Eucalyptus* (Rosado et al., 2010).

Genomic *loci* associated with resistance to gall-inducing insects were identified for the first time in plants and provide a valuable contribution to the understanding of the genetic basis underlying plant resistance to gall wasps. Seven SSR markers were associated with resistance to *Leptocybe invasa* in *E. grandis*, of which four associations were verified in *E. tereticornis*. Five SSR markers are homologous to known genes or predicted proteins of the seven significant *loci* detected in *E. grandis* (Zhang et al., 2018). C2H2 zinc proteins are one of the most prominent families of transcription factors in plants and have been found to participate in several signal transduction pathways, including pathogen defense and stress responses. Another protein is protein tyrosine phosphatase which functions as an inducer of dose-dependent mitotic progression (UNIPROT, 2022).

Linkage disequilibrium (LD) mapping can be used to identify alleles associated with characteristics of wood quality in forest populations. Depending on the extent of the LD, it is possible to identify alleles in candidate genes associated with a trait. In *Eucalyptus nitens* the *CCR* gene, an essential gene of the lignin pathway, affects the physical properties of the secondary cell wall, such as rigidity and strength (Thumma et al., 2005). Maps for *E. globulus* demonstrated specific genes and ESTs responsible for lignin production, cell wall formation, flowering, and cambium development used for QTL mapping of wood, fiber, and floral traits. Maps based on the segregation of codominant *loci* are transferable to other eucalyptus species and provide a basis for investigating genome organization and evolution in the genus (Thamarus et al., 2002).

E. cladocalyx at different ages were studied to identify stable marker-trait associations for growth in 47 half-sib families. Using the SSR, *EMBRA32*, *EMBRA191*, and *EMBRA208* markers, a significant association was found between height and Diameter at Breast Height (DBH) (Maldonado et al., 2018). RAPD markers are used to detect traits related to the wood quality and growth in *E. grandis* x *E. urophylla* hybrids (Rocha et al., 2007) and anomalies related to morphological characteristics such as branching, reduced height, smaller and asymmetrical leaves, and deleterious recessive leaf shape of seedlings in *E. grandis* (Fuchs et al., 2015). Although RAPD markers have been used in many studies involving gene mapping to genetically characterize *Eucalyptus* species with technological advances, the use of this marker is no longer recommended for this type of analysis. The DArT marker combined with next-generation sequencing technologies (NGS) generates linkage maps of *Eucalyptus* as a reference genome showing promising results (Sansaloni et al., 2011). This marker can be used in comparative genomic studies, characterize targeted genes, and help with selection in *Eucalyptus* and related taxa (Li et al., 2015).

3.2 Identification of proteins and associations from molecular data

The genomic resources developed for the cultivation of *Eucalyptus* accelerated its use in the breeding programs by selection assisted by markers, leading to the availability of improved hybrids worldwide. Table 2 shows compiled information on the main markers used to identify characteristics of economic importance in *Eucalyptus*.

Table 2. Molecular markers in *Eucalyptus*. Characteristics, markers, genes, species, and references.

Traits	Markers	Genes	Species	References
Flowering and fruiting	SSR ISSR	*ISO1-970, ISO1-500, RCA7-890, RCA7-110*	*Eucalyptus cladocalyx* F.Muell *E. globulus* Labill *Eucalyptus* spp.	Thamarus *et al*. (2002); Farro *et al*. (2013); Ballesta *et al*. (2015) Contreras-Soto *et al*. (2016);
Disease resistance	ISSR SSR	*UBC807, UBC808, UB823, UBC834, UBC855, UBC866, UBC873, UBC890 EMBRA125, EMBRA181, EMBRA227, EMBRA333, EMBRA321, EUCESSR0930, EUCeSSR479, EUCeSSR683*	*Eucalyptus* spp.	Rosado *et al*. (2010); Zhang *et al*. (2018); Gomes *et al*. (2019)
Construction of molecular maps	RAPD		*E. urophylla* S.T.Blake e *E. tereticornis* Sm.	Gan *et al*. (2003)
Genetic diversity	SSR	*EMBRA158, EMBRA91, EMBRA70, EMBRA208, EMBRA155, EMBRA127, EMBRA56, EMBRA11, EMBRA4*	*E. cladocalyx* F.Muell	Mora, Arriagada, Ballesta and Ruiz (2017)
Vegetative propagation/ wood properties	RAPD SSR DArT AFLP SRR RFLP	*PIL18-1, PIL26-1, PIL26-6, PIL38-6, HDR18, HDR26, HDR38, HDR18-8, HDR26-8, HDR26-1, HDR38-1, VIG26-1, VIG38-1, VIG26-5, EMBRA 40, EMBRA195, EMBRA 207, CRC 47, EMBRA125, EMBRA1071, EMBRA1656, EMBRA208, EMBRA191*	*E. grandis* W.Hill, *E. urophylla* S.T.Blake, *E. tereticornis* Sm., *E. camaldulensis* Dehnh., *E. cladocalyx* F.Muell, *E. globulus* Labill, *Eucalyptus spp., E. globulus* Labill	Verhaegen *et al*. (1997); Marques *et al*, (1999); Thamarus *et al*. (2002); Thamarus *et al*. (2004); Thumma *et al*. (2005); Freeman, Whittock, Potts and Vaillancourt (2009); Freeman *et al*. (2013); Shanmugapriya, Bachpai, Ganesan and Yasodha (2015); Li (2015); Maldonado *et al*. (2018)
Identification of anomalies	RAPD	*PR10, BETV1*	*E. grandis* W.Hill	Fuchs *et al*. (2015)

Source: Authors (2022).

The RAPD technique was the most used in studies to obtain alleles in *Eucalyptus*. The ISSR and SSR markers were highlighted in studies to detect pest and disease-resistant genotypes. A preference for ISSR markers was observed, limiting itself to the study of disease-response-related primers. ISSR markers present promising results for molecular analysis as they align in specific regions of the genome (Inter *loci* SSR), and DNA segments are amplified by simple reactions, producing a reproductive pattern of amplified fragments (Domingues et al., 2017). In Table 3 are presented ISSR primers used in disease resistance studies in *Eucalyptus*.

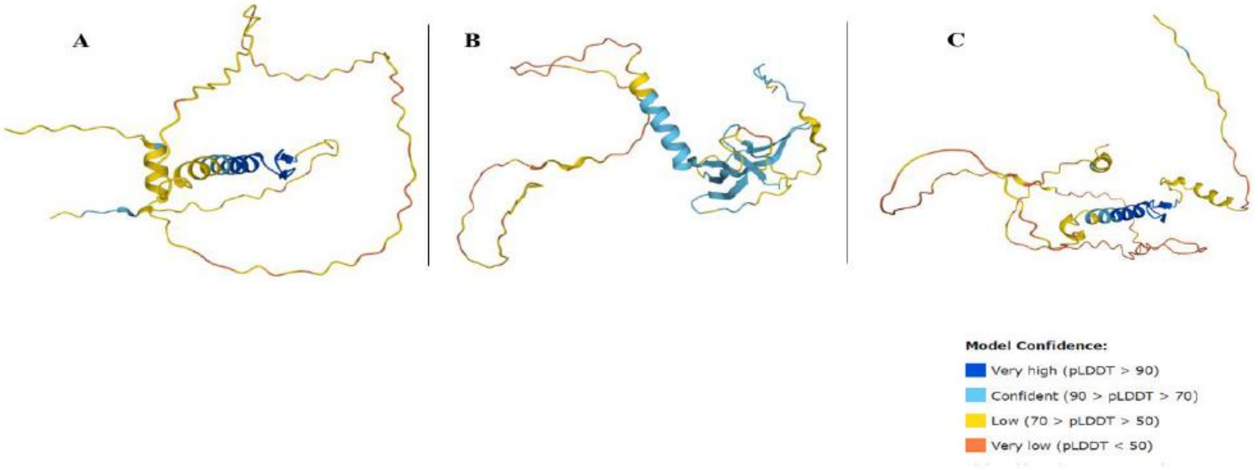
Table 3. ISSR primers used in plant disease resistance studies in *Eucalyptus*. Primer, sequences, annealing temperature (TA).

Primer	Sequences	TA (ºC)
UBC807	$(AG)_8T$	43
UBC 808	$(AG)_8C$	47
UBC823	$(TC)_8C$	52
UB 834	$(AG)_8YT$	47
UBC855	$(AC)\ _8YT$	50
UBC866	$C(TCC)_5TC$	55
UBC873	$(GACA)_4$	52
UBC890	$VHV(GT)_7$	56

Source: Authors (2022).

However, SSR markers were used to identify genomic *loci* in *Eucalyptus*, the genes Embra333, EUCeSSR0930, Embra321, which are related to the response *Leptocybe invasa* attack (Zhang et al., 2018) are homologous to the known genes *AT1G14410*, *AT2G42410* and *AT5G64680*. The original sequences of these genes were obtained and searched in BlastX against the NCBI non-redundant protein database to obtain predicted proteins. the three-dimensional structures of proteins with functions related to gall wasp resistance in *Eucalyptus* were verified at the Swiss Institute of Bioinformatics (SIB) (Figure 5). The sequences associated with resistance to galling organisms in *Eucalyptus* were: TAC1, At5g64680, and ZFP11. Moreover, have genes expressed in the model species *Arabidopsis thaliana* (L.) Heynh, respectively: *AT5G64680*, *AT1G14410*, *AT2G42410*.

Figure 5. Three-dimensional structure of proteins related to resistance to galling organisms obtained by homology with *Eucalyptus grandis*. A. TAC1; B. At5g64680; 2. C. ZFP11.

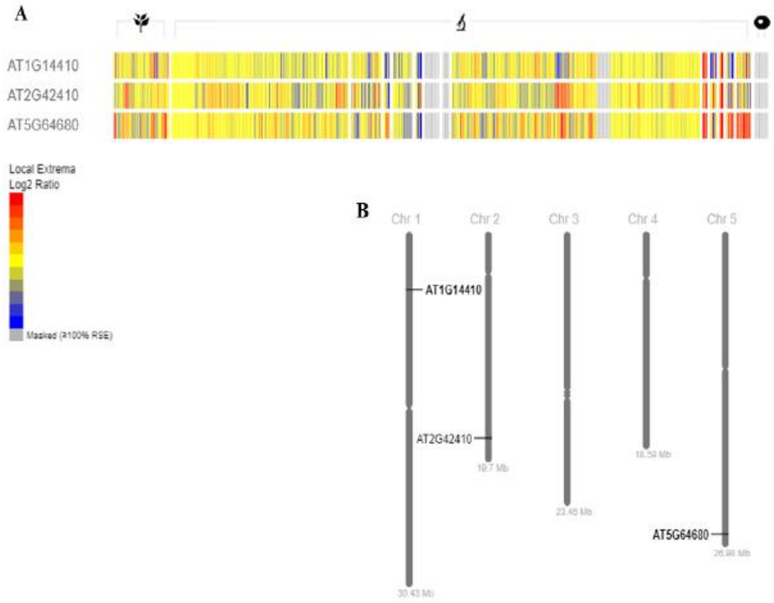


Source: Swiss Institute of Bioinformatics [SIB] (2022).

In plants with overexpression of *ZFP11*, the likely transcription factor can regulate cell division and growth, have abnormal leaf morphology, early flowering, and most plants are sterile. Proteins that bind to RNA and that can regulate gene expression were preliminarily characterized. The zinc finger domain presents a highly conserved arrangement of cysteines and histidine, C-x(8)-Cx(5)-C-x(3)-H, belonging to the C3H1 type of zinc fingers that is conserved and interacts with mRNAs, preferentially in the 3' untranslated region (HALL, 2005).

The expression levels of the *AT1G14410*, *AT2G42410*, and *AT5G64680* genes and the chromosomal location were obtained in e-Plant and are presented in Figure 6. e-Plant allows us to visualize the natural connections between DNA sequences and, natural variation (polymorphisms), molecular structures, protein-protein interactions, and gene expression patterns by combining multiple data visualization tools with a user interface (WAESE et al., 2017).

Figure 6. A - Heat map displays expression levels and corresponding cell locations; B – Chromosomal location; of the *AT1G14410*, *AT2G42410*, and *AT5G64680* genes.



Source: e-Plant (2022).

Based on the expression levels of the prospected genes, there is a pattern of similarity both in the expression levels at the leaf level and at the cell and subcellular levels. The *AT1G14410* gene is located on chromosome 1, *AT2G42410* on chromosome 2, and *AT5G64680* on chromosome 5. The zinc finger is one of the most abundant structural motifs involved in the interaction between nucleic acids and proteins. Zinc finger proteins are involved in several aspects of gene regulation (Laity et al., 2001). Zinc finger domains (ZnFs) are common, relatively small protein motifs that fold around one or more zinc ions. In addition to their role as a DNA binding module, ZnFs have been shown to mediate protein: protein and protein: lipid interactions. This small zinc-binding domain, often found in clusters containing different binding specificities, can facilitate multiple intermolecular interactions between nucleic acids and proteins (Matthews & Sunde, 2002).

4. Conclusion

The RAPD technique was the most used in studies to obtain alleles with characteristics of interest in *Eucalyptus*. The vital role of these markers is related to their random amplification that could flank regions of genes to express proteins potentially useful for SAM.

The ISSR and SSR markers were highlighted in studies to identify pest and disease-resistant genotypes. As codominant markers, it is possible to detect very informative alleles and detect hybrids in crossings.

The candidate genes associated with resistance to galling organisms in *Eucalyptus* can be evaluated by transcriptional regulatory proteins TAC1, proteins associated with mediator 2, zinc finger protein 11, and the expressed genes *TAC1*, *At5g64680*, and *ZFP11*.

The evidence observed in data mining in this article open frameworks. As perspectives for future studies, it can be elected research to support the comparative framework to study new hybrid combinations.

We emphasize the sine qua non-importance to carry out studies like this in the frame to define strategies to access techniques to establish very robust studies, mainly in species like *Eucalyptus* aiming for improvements in breeding programs.

Also, the development of new research focusing on bioinformatics tools is suggested to expand and improve studies with different marker-assisted selection techniques.

References

Agarwal, M., Shrivastava, N., & Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports, 27*(4), 617-631. https://doi.org/10.1007/s00299-008-0507-z

Arcade, A., Faivre-Rampant, P., Pâques, L. E., & Prat, D. (2002). Localisation of genomic regions controlling microdensitometric parameters of wood characteristics in hybrid larches. *Annals of Forest Science*, *59*(5-6), 607-615. https://doi.org/10.1051/forest:2002046

Ballesta, P., Mora, F., Ruiz, E., & Contreras-Soto, R. (2015). Marker-trait associations for survival, growth, and flowering components in *Eucalyptus cladocalyx* under arid conditions. *Biologia plantarum*, *59*(2), 389-393.

Bandaranayake, C. K., & Kearsey, M. J. (2005). Genome mapping, QTL analysis and MAS: Importance, principle, constraints and application in coconut. *Plant Genetic Resources Newsletter*, (142), 47.

Chagné, D., Batley, J., Edwards, D., & Forster, J. W. (2007). *Single Nucleotide Polymorphism Genotyping in Plants*. In: Oraguzie, N. C., Rikkerink, E. H. A., Gardiner, S. E., Silva, H. N. (eds) Association Mapping in Plants. Springer, New York, NY. https://doi.org/10.1007/978-0-387-36011-9_5

Contreras-Soto, R., Ballesta, P., Ruiz, E., & Mora, F. (2016). Identification of ISSR markers linked to flowering traits in a representative sample of *Eucalyptus cladocalyx*. *Journal of forestry research*, *27*(2), 239-245.

Domingues, S. D., Neves, A. F., Mangolin, C. A., & Machado, M. F. P. S. (2017). Seleção de primers para análise de Inter Simple Sequence Repeats (ISSR) em Cereus sp. (Cactaceae) Selection of primers for Inter Simple Sequence Repeats (ISSR) in *Cereus* sp.(Cactaceae). *Revista Biotecnologia & Ciência* v, 6(2), 46-54.

e-Plant. (2022). The bio-analytic resource for plant biology. http://bar.utoronto.ca/

Farro, A. P., Bortoloto, T. M., Oda, S., Mello, E., & Marino, C. L. (2013). Identification of molecular marker linked to early flowering in *Eucalyptus grandis*. Rev Ins Fla, (24), 149-157.

Food and Agriculture Organization of the United Nations. (2021). *Eucalyptus Species*. http://www.fao.org/3/AC121E/ac121e04.htm

Freeman, J. S., Potts, B. M., Downes, G. M., Pilbeam, D., Thavamanikumar, S., & Vaillancourt, R. E. (2013). Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytologist*, *198*(4), 1121-1134.

Freeman, J. S., Whittock, S. P., Potts, B. M., & Vaillancourt, R. E. (2009). QTL influencing growth and wood properties in *Eucalyptus globulus*. *Tree Genetics & Genomes*, *5*(4), 713-722.

Fuchs, M. C. P., Tambarussi, E. V., Lourenção, J. C., Nogueira, L. M., Bortoloto, T. M., González, E. R., Oda, S., & Marino, C. L. (2015). Molecular marker associated with a deleterious recessive anomaly in *Eucalyptus grandis* seedlings. *Annals of Forest Science*, *72*(8), 1043–1052. https://doi.org/10.1007/s13595-015-0502-9

Gan, S., Shi, J., Li, M., Wu, K., Wu, J., & Bai, J. (2003). Moderate-density molecular maps of *Eucalyptus urophylla* S. T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. *Genetica*, *118*(1), 59-67. https://doi.org/10.1023/A:1022966018079

Goetze, M., Passaia, G., & Sperb-Ludwig, F. (2017). *Marcadores moleculares baseados em restrição: AFLP e suas variações. In Marcadores Moleculares na Era genômica: Metodologias e Aplicações* (p.181). https://www.lume.ufrgs.br/bitstream/handle/10183/206114/001056131.pdf?sequence=1%0Ahttps://www.sbg.org.b r/sites/default/fil es/e_book_marcadores_moleculares_sbg_2017_final.pdf

Gomes, C. A. F. C., Pereira, F. B., De Oliveira Garcia, F. A., Garret, A. T. D. A., De Siqueira, L., & Tambarussi, E. V. (2019). Inoculation of *Ceratocystis fimbriata* ellis & halsted in *Eucalyptus* spp. And evaluation pf genetic diversity by ISSR markers. *Scientia Forestalis/Forest Sciences*, *47*(123), 579-587. https://doi.org/10.18671/scifor.v47n123.19

Grattapaglia, D., Bertolucci, F. L. G., Penchel, R., & Sederoff, R. R. (1996). Genetic mapping of quantitative trait loci controlling growth and wood quality traits in Eucalyptus grandis using a maternal half-sib family and RAPD markers. *Genetics*, *144*(3), 1205-1214.

Grattapaglia, D., Ribeiro, V. J., & Rezende, G. D. S. P. (2004). Retrospective selection of elite parent trees using paternity testing with microsatellite markers: An alternative short term breeding tactic for Eucalyptus. *Theoretical and Applied Genetics*, *109*(1), 192-199. https://doi.org/10.1007/s00122-004-1617-9

Hall, T. M. Multiple modes of RNA recognition by zinc finger proteins. *Curr Opin Struct Biol*. 15(3):367-373. 2005.

Jaccoud, D., Peng, K., Feinstein, D., & Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, *29*(4), 25. https://doi.org/10.1093/nar/29.4.e25

Jaleta, D., Mbilinyi, B., Mahoo, H., & Lemenih, M. (2016). *Eucalyptus* Expansion as Relieving and Provocative Tree in Ethiopia. *Journal of Agriculture and Ecology Research International*, *6*(3), 1-12.

Kumar, P., Gupta, V., Misra, A., Modi, D., & Pandey, B. (2009). Potential of Molecular Markers in Plant Biotechnology. *Plant Omics*, 2(4), 141–162.

Laity, J. H., Lee, B. M., & Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Current opinion in structural biology*, 11(1), 39-46.

Lee, S. H., Clark, S., & Werf, J. H. J. V. D. (2017). Estimation of genomic prediction accuracy from reference populations with varying degrees of relationship. *PLoS ONE*, *12*(12), 1-22. https://doi.org/10.1371/journal.pone.0189775

Li, F., Zhou, C., Weng, Q., Li, M., Yu, X., Guo, Y., Wang, Y., Zhang, X., & Gan, S. (2015). Comparative genomics analyses reveal extensive chromosome colinearity and novel quantitative trait loci in eucalyptus. *PLoS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0145144

Maldonado, C., Contreras-Soto, R. I., Gerhardt, I. F. S., Amaral, A. T. Jr., & Mora, F. (2018). Stable marker-trait associations for growth across different ages in *Eucalyptus cladocalyx* through the use of microsatellites. *Scientia Forestalis*, *46*(119). https://doi.org/10.18671/scifor.v46n119.04

Marques, C. M., Vasquez-Kool, J., Carocha, V. J., Ferreira, J. G., O'Malley, D. M., Liu, B. H., & Sederoff, R. (1999). Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. *Theoretical and Applied Genetics*, *99*(6), 936-946. https://doi.org/10.1007/s001220051400

Matthews, J. M., & Sunde, M. (2002). Zinc fingers--folds for many occasions. *IUBMB life*, 54(6), 351-355.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews, 4(1), 1-9. https://doi.org/10.1186/2046-4053-4-1

Mora, F., Arriagada, O., Ballesta, P., & Ruiz, E. (2017). Genetic diversity and population structure of a drought-tolerant species of *Eucalyptus*, using microsatellite markers. *Journal of Plant Biochemistry and Biotechnology*, *26*(3), 274-281.

Naidoo, S., Külheim, C., Zwart, L., Mangwanda, R., Oates, C. N., Visser, E. A., Wilken, F. E., Mamni, T. B., & Myburg, A. A. (2014). Uncovering the defence responses of eucalyptus to pests and pathogens in the genomics age. *Tree Physiology*, *34*(9), 931-943. https://doi.org/10.1093/treephys/tpu075

Nicolle, D. (2019). *Classification of the Eucalypts (Angophora, Corymbia and Eucalyptus)*. Version 4.

R Core Team. (2020). *R: A Language and environment for statistical computing*. [Software]. Vienna: R Foundation on Statistical Computing (1.3.1073).

Savelkoul, P. H. M., Aarts, H. J. M., Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., & Lenstra, J. A. (1999). Amplified-fragment length polymorphism analysis: the state of an art. *Journal of clinical microbiology*, *37*(10), 3083-3091.

Shanmugapriya, A., Bachpai, V. K. W., Ganesan, M., & Yasodha, R. (2015). Association analysis for vegetative propagation traits in *Eucalyptus tereticornis* and *Eucalyptus camaldulensis* using simple sequence repeat markers. *Proceedings of the National Academy of Sciences*, India Section B: Biological Sciences, *85*(2), 653-658.

Swiss Institute of Bioinformatics. (2022). The 3D structures of the proteins. https://www.sib.swiss/

Ribeiro, D. O., Silva-Mann, R., Carvalho, S. V. Á., Souza, E. M. S. D., Vasconcelos, M. C., & Blank, A. F. (2017). Genetic variability in *Jatropha curcas* L. from diallel crossing. *Genetics and Molecular Research*, *16*(2): gmr16029651.

Rocha, R. B., Barros, E. G., Cruz, C. D., Rosado, A. M., & Araújo, E. F. (2007). Mapping of QTLs related with wood quality and developmental characteristics in hybrids (*Eucalyptus grandis* x *Eucalyptus urophylla*). *Revista Árvore*, *31*(1), 13-24. https://doi.org/10.1590/S0100-67622007000100003

Rosado, T. B., Tomaz, R. S., Ribeiro Junior, M. F., Rosado, A. M., Guimarães, L. M. D. S., Araújo, E. F. D., & Cruz, C. D. (2010). Detection of QTL associated with rust resistance using IBD-based methodologies in exogamic *Eucalyptus* spp. populations. *Crop Breeding and Applied Biotechnology*, 10, 321-328.

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings*, *5*(7), 1-2. https://doi.org/10.1186/1753-6561-5-s7-p54

Samal, B., Vanapalli, K. R., Dubey, B. K., Bhattacharya, J., Chandra, S., & Medha, I. (2021). Influence of process parameters on thermal characteristics of char from co-pyrolysis of eucalyptus biomass and polystyrene: Its prospects as a solid fuel. *Energy*, 232, 121050.

Sarkis-Onofre, R., Catalá-López, F., Aromataris, E., & Lockwood, C. (2021). How to properly use the PRISMA Statement. Systematic Reviews, 10(1), 1-3. https://doi.org/10.1186/s13643-021-01671-z

Silva-Mann, R., Vieira, M. G., Machado, J. C., Bernardino Filho, J. R., Salgado, K. C., & Stevens, M. R. (2005). AFLP markers differentiate isolates of *Colletotrichum gossypii* from *C. gossypii* var. cephalosporioides. *Fitopatologia Brasileira*, 30, 169-172.

Souza, D. C. L. (2015). Técnicas moleculares para caracterização e conservação de plantas medicinais e aromáticas: uma revisão. *Revista Brasileira de Plantas Medicinais*, *17*(3), 495-503. https://doi.org/10.1590/1983-084X/13_071

Souza, G. A., Carvalho, M. R.O., Martins, E. R., Guedes, R. N. C., & Oliveira, L. O. (2008). Diversidade genética estimada com marcadores ISSR em populações brasileiras de *Zabrotes subfasciatus*. *Pesquisa Agropecuária Brasileira*, *43*(7), 843-849. https://doi.org/10.1590/S0100-204X2008000700008

Thamarus, K., Groom, K., Bradley, A., Raymond, C. A., Schimleck, L. R., Williams, E. R., & Moran, G. F. (2004). Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. *Theoretical and Applied Genetics*, *109*(4), 856-864.

Thamarus, K., Groom, K., Murrell, J., Byrne, M., & Moran, G. (2002). A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fiber, and floral traits. *Theoretical and Applied Genetics*, *104*, 379-387.

Theoretical and Applied Genetics. *International Journal of Plant Breeding Research*. https://www.springer.com/journal/122/aims-and-scope

Thumma, B.R., Nolan, M.F., Evans, R., & Moran, G. (2005). Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in Eucalyptus spp. *Genetics*, 171, 1257-1265.

Trevorah, R. M., Huynh, T., Vancov, T., & Othman, M. Z. (2018). Bioethanol potential of Eucalyptus obliqua sawdust using gamma-valerolactone fractionation. *Bioresource Technology*, 250, 673-682. https://doi.org/10.1016/j.biortech.2017.11.084

Verhaegen, D., Plomion, C., Gion, J. M., Poitel, M., Costa, P., & Kremer, A. (1997*)*. Quantitative trait dissection analysis in Eucalyptus using RAPD markers: Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theoretical and Applied Genetics*, *95*(4), 597-608.

Viana, J. M. S., Cruz, C. D., & Barros, E. G. (2012). *Genética - Fundamentos*. Universidade Federal de Viçosa (2a). Editora UFV.

Waese, J., & Provart, N. J. (2017). “The bio-analytic resource for plant biology,” in Plant Genomics Databases. *Humana Press*, 119–148.

Williams, J. G., Hanafey, M. K., Rafalski, J. A., & Tingey, S. V. (1993). Genetic analysis using random amplified polymorphic DNA markers. In Methods in enzymology. *Academic Press*, 218, 704-740.

Young, N. D., Tanksley, S. D. (1989). Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theoret. Appl. Gene *Theoretical and Applied Genetics*, 77, 95-101. https://doi.org/10.1007/BF00292322

Zhang, M., Zhou, C., Song, Z., Weng, Q., Li, M., Ji, H., Mo, X., Huang, H., Lu, W., Luo, J., Li, F., & Gan, S. (2018). The first identification of genomic loci in plants associated with resistance to galling insects: A case study in *Eucalyptus* L’Hér. (Myrtaceae). *Scientific Reports*, *8*(1), 1-10. https://doi.org/10.1038/s41598-018-20780-9

Zhang, M., Zhou, C., Song, Z., Weng, Q., Li, M., Ji, H., & Gan, S. (2018). The first identification of genomic loci in plants associated with resistance to galling insects: a case study in *Eucalyptus* L'Hér (Myrtaceae). *Scientific reports*, 8(1), 1-10.