



**Universidade Federal do Pampa**

**UNIVERSIDADE FEDERAL DO PAMPA**

**CAMPUS SÃO GABRIEL**

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS**

**JORDANA CAROLINE NAGEL**

**CARACTERIZAÇÃO DA DIVERSIDADE MOLECULAR DE NOGUEIRA-PECÃ  
(*Carya illinoensis*) BASEADA EM MARCADORES MOLECULARES**

**TESE DE DOUTORADO**

**SÃO GABRIEL  
2022**

**JORDANA CAROLINE NAGEL**

**CARACTERIZAÇÃO DA DIVERSIDADE MOLECULAR DE NOGUEIRA-PECÃ  
(*Carya illinoensis*) BASEADA EM MARCADORES MOLECULARES**

Tese de Doutorado apresentada ao Programa de Pós-Graduação *Stricto sensu* em Ciências Biológicas da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Doutor em Ciências Biológicas.

Orientador: Prof. Dr. Valdir Marcos Stefenon

Coorientador: Prof. Dr. Igor Poletto

**SÃO GABRIEL  
2022**

Ficha catalográfica elaborada automaticamente com os dados fornecidos  
pelo(a) autor(a) através do Módulo de Biblioteca do  
Sistema GURI (Gestão Unificada de Recursos Institucionais) .

N147c Nagel, Jordana Caroline

CARACTERIZAÇÃO DA DIVERSIDADE MOLECULAR DE NOGUEIRA-PECÃ  
(*Carya illinoensis*) BASEADA EM MARCADORES MOLECULARES /  
Jordana Caroline Nagel.

119 p.

Tese(Doutorado)-- Universidade Federal do Pampa, DOUTORADO  
EM CIÊNCIAS BIOLÓGICAS, 2022.

"Orientação: Valdir Marcos Stefenon".

1. cpDNA. 2. nDNA. 3. marcadores SSR. 4. plastoma. 5.  
sequenciamento de DNA. I. Título.

**JORDANA CAROLINE NAGEL**

**CARACTERIZAÇÃO DA DIVERSIDADE MOLECULAR DE NOGUEIRA-PECÃ (*Carya illinoensis*) BASEADA EM MARCADORES MOLECULARES**

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Doutor em Ciências Biológicas.

Dissertação defendida e aprovada em: 07 de junho de 2022.

Banca examinadora:

---

Prof. Dr. Valdir Marcos Stefenon

Orientador

UFSC

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Adriana Cibebe de Mesquita Dantas

UERGS

---

Prof. Dr. Adelar Mantovani

UDESC

---

Prof. Dr. Filipe de Carvalho Victória

UNIPAMPA



Assinado eletronicamente por **Valdir Marcos Stefenon, Usuário Externo**, em 08/06/2022, às 16:26, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

---



Assinado eletronicamente por **FILIPE DE CARVALHO VICTORIA, PROFESSOR DO MAGISTERIO SUPERIOR**, em 09/06/2022, às 13:12, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

---



Assinado eletronicamente por **Adriana Cibele de Mesquita Dantas, Usuário Externo**, em 27/06/2022, às 16:39, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

---



Assinado eletronicamente por **Adelar Mantovani, Usuário Externo**, em 27/06/2022, às 16:52, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

---



A autenticidade deste documento pode ser conferida no site [https://sei.unipampa.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0807111** e o código CRC **28EEADFC**.

---

## AGRADECIMENTOS

Resumir em palavras a gratidão que sinto é uma tarefa que por mais que eu tente, será sempre incompleta. Tornar-se doutora é uma conquista individual, mas só pude trilhar esse caminho, porque foi compartilhado com pessoas e forças que me inspiraram, apoiaram, acolheram e principalmente me ajudaram. Chegar a um ponto da vida e perceber que eu pude e posso contar com grandes pessoas na minha vida é uma honra. E, eu devo isso a todos vocês.

Dedico a você Isabella, luz da vida, é por você e sempre será filha, por todo carinho, amor, compreensão e paciência. Mamãe te ama!

A toda minha família que sempre me incentivou nos estudos e buscou sempre me ajudar nos momentos de dificuldade. Principalmente aos meus pais, Beloni e Romeu, pelo apoio e a confiança que sempre depositaram em mim. O sonho do meu pai era ter uma filha doutora, finalmente, não é pai! Imensa gratidão por tudo, suporte emocional, financeiro e com o meu bem mais precioso, a Isabella. Sem vocês nada disso seria possível!

Agradeço a Universidade Regional Integrada do Alto Uruguai e das Missões - Santo Ângelo, Universidade Federal de Santa Catarina e a Universidade Federal do Pampa – São Gabriel, por oferecerem a estrutura e os recursos necessários para o desenvolvimento do meu doutorado.

Ao CNPq, Capes, IBPecan pelo apoio financeiro para a pesquisa.

Aos professores pelo ensinamento, dedicação e profissionalismo. Aos professores e revisores que avaliaram, julgaram e contribuíram com esse trabalho.

Aos meus colegas Lenon, Carol, Daniel, Júlia, Rafael, Dalvan e Suelen pela ajuda e as risadas durante os períodos de laboratório, com certeza isso tornou a pesquisa mais leve! A Dra. Deise Sarzi pela ajuda com os softwares e por disponibilizar um tempo para uma aula, mesmo a distância. Ao Dr. Tales Poletto pelos auxílios no laboratório, discussões do trabalho e pelas imagens e materiais cedidos para a Tese. Ao prof. Dr. Igor Poletto pela parceria de trabalho. A Dra. Laura Sommer por disponibilizar um tempo para a leitura e contribuição na escrita. Obrigada por fazerem parte dessa história, agradeço grandemente pela amizade, pelos ensinamentos e por fazer a minha vida mais feliz, principalmente nos períodos em Florianópolis. Vocês significam muito pra mim!

As minhas amigas Flávia e Raquel que disponibilizaram um lugar para estadia durante a realização das pesquisas, obrigada de coração pelo apoio! Aos amigos/as Thiago, Karen, Natália, Ranieri, Laura e Daiane que compreenderam minhas ausências nesse período e nunca

deixaram de me apoiar! Obrigada pela amizade e carinho de sempre.

Ao meu orientador, professor Dr. Valdir Marcos Stefenon que vem me acompanhando e instruindo desde 2011, do mestrado ao doutorado. Obrigada pela orientação, paciência, transmissão do conhecimento, experiência, e por sempre me incentivar a ser cada vez melhor. Você é um “espelho profissional” para mim! Sua colaboração e apoio fez com que esse trabalho tenha sido realizado com grande sucesso. Posso afirmar, com toda a certeza, que tenho muita sorte em ter te escolhido para essa caminhada. Espero que continuemos com nossa parceria por muitos anos!

E aos demais que contribuíram, direta ou indiretamente, para a realização deste trabalho.

A todos MUITO OBRIGADA!

*“Que todos os nossos esforços estejam sempre focados no desafio à impossibilidade. Todas as grandes conquistas humanas vieram daquilo que parecia impossível.”*

**Charles Chaplin**

## RESUMO

A noqueira-pecã (*Carya illinoensis* [Wangenh.] K. Koch) pertence à família Juglandaceae, e tem distribuição predominante nas regiões temperadas do Hemisfério Norte. O interesse pelos seus frutos tem apresentado um crescimento exponencial no Brasil e outros países, especialmente por seu alto valor nutricional. Além disso, é uma opção de renda para agricultores familiares, devido ao seu potencial de uso em áreas de reserva legal, tornando-se uma espécie de interesse econômico e social, principalmente no Sul do Brasil. Devido à carência de informações genéticas, falta de controle de polinização, e de registro de coleta de propágulos nos pomares, os materiais genéticos em cultivo, geralmente, não têm procedência completamente conhecida. Nesse sentido, a caracterização molecular a partir de marcadores moleculares pode solucionar esse problema, identificando e distinguindo cultivares, e propiciando, simultaneamente, as bases para o planejamento de um eventual programa de melhoramento genético da espécie. Além disso, contribuiria, ainda, para a melhoria do sistema de manejo dos pomares, com repercussão direta na produtividade. O presente trabalho objetivou sequenciar o DNA plastidial da cultivar Imperial, utilizando este, para desenvolver marcadores microssatélite e posteriormente avaliar concomitantemente com marcadores microssatélites plastidiais universais e marcadores microssatélites nucleares espécie-específicos a diversidade genética existente entre as cultivares plantadas no Rio Grande do Sul. No sequenciamento completo do DNA (cpDNA) de *C. illinoensis* cv Imperial, constatou-se que o genoma possui 160.818 pares de base (pb) de comprimento, estrutura quadripartida com LSC de 90.041 pb, SSC de 18.791 pb e dois IRs de 25.993 pb. Um total de 78 regiões codificadoras de proteínas, 37 codificadoras de tRNA e oito regiões codificadoras de rRNA foram previstas. Para a prospecção, caracterização e validação dos primeiros marcadores microssatélites plastidiais para a espécie foram utilizados dados obtidos no sequenciamento do plastoma *C. Illinoensis* cv Imperial. Para a validação dos marcadores, foi realizado o isolamento do DNA genômico de folhas coletadas de 13 cultivares e os 10 pares de *primers* plastidiais prospectados foram amplificados via reação de PCR e os alelos separados via eletroforese em gel de agarose. O teste de transferibilidade *in silico* dos marcadores resultou em amplificação de 10 *loci* nas cultivares de *C. illinoensis*, nove em espécies do gênero *Carya* e sete no gênero *Juglans*, sugerindo seu potencial para aplicação em estudos genéticos dessas espécies. Na validação dos marcadores em cultivares de *C. illinoensis*, apenas o *locus cpCill* não amplificou para a cultivar Choctaw, e não houve

amplificação para os *loci cpCil8* e *cpCil9* na cultivar Chickasaw. Para a análise da divergência genética das cultivares foram testados 10 marcadores microssatélites nucleares (SSR) e 13 plastidiais (cpSSR). Através da análise de coordenadas principais (PCoA) dos nSSR, os dois componentes explicaram, no total, 21,39% da variabilidade total observada, enquanto para os cpSSR o total de 84,65%. Apesar do valor de correlação cofenética alto (cpSSR= 0,9931 e nSSR= 0,9791), os valores de *bootstrap* nos agrupamentos maiores foram baixos. Os resultados obtidos neste estudo evidenciaram alta diversidade genética tanto para marcadores nucleares ( $I= 2,365$ ) quanto plastidiais ( $I= 2,183$ ). Esta estimativa de diversidade genética é bem significativa, e este resultado pode ser utilizado para determinar o padrão genético das cultivares analisadas. Este é o primeiro estudo de desenvolvimento e caracterização de marcadores microssatélites plastidiais espécie-específico para *C. illinoensis* com potencial aplicabilidade para gerar informações a respeito da diversidade genética entre as cultivares da espécie. Os dados obtidos até o momento nos permitem auxiliar no aprofundamento de análises da diversidade genética, pois por meio de estudos recentes a polinização aberta promove alta segregação genética, e a propagação através de sementes pode ter introduzido centenas de variedades de noqueira-pecã pelo mundo. Essa grande diversidade na forma dos frutos, das árvores, a qualidade das nozes e suas diferentes formas de se reproduzir, ainda causa uma grande confusão quanto à nomenclatura e a identificação.

**Palavras-chave:** cpDNA, nDNA, marcadores SSR, cloroplastos, sequenciamento de DNA, plastoma.

## ABSTRACT

The pecan tree (*Carya illinoensis* [Wangenh.] K. Koch) belongs to the Juglandaceae family, and has a predominant distribution in the temperate regions of the Northern Hemisphere. Interest in its fruits has shown exponential growth in Brazil and other countries, especially for its high nutritional value. In addition, it is an income option for family farmers, due to its potential for use in legal reserve areas, becoming a kind of economic and social interest, especially in southern Brazil. Due to the lack of genetic information, lack of pollination control, and lack of records of propagule collection in orchards, genetic materials in cultivation generally do not have a completely known origin. In this sense, molecular characterization from molecular markers can solve this problem, identifying and distinguishing cultivars, and simultaneously providing the basis for planning an eventual program for the genetic improvement of the species. In addition, it would also contribute to the improvement of the orchard management system, with direct repercussions on productivity. The present work aimed to sequence the plastid DNA of the Imperial cultivar, using it to develop microsatellite markers and later to evaluate, concomitantly with universal plastid microsatellite markers and species-specific nuclear microsatellite markers, the genetic diversity existing among the cultivars planted in Rio Grande do Sul. In the complete DNA sequencing (cpDNA) of *C. illinoensis* cv Imperial, it was found that the genome is 160,818 base pairs (bp) long, with a quadripartite structure with LSC of 90,041 bp, SSC of 18,791 bp and two IRs of 25,993 bp. A total of 78 protein-coding regions, 37 tRNA-coding regions and eight rRNA-coding regions were predicted. For the prospection, characterization and validation of the first plastid microsatellite markers for the species, data obtained from the sequencing of the plastome *C. illinoensis* cv Imperial were used. To validate the markers, genomic DNA was isolated from leaves collected from 14 cultivars and the 10 pairs of plastid primers prospected were amplified via PCR reaction and the alleles separated via agarose gel electrophoresis. The *in silico* transferability test of the markers resulted in amplification of 10 *loci* in *C. illinoensis* cultivars, nine in species of the genus *Carya* and seven in the genus *Juglans*, suggesting its potential for application in genetic studies of these species. In the validation of markers in *C. illinoensis* cultivars, only the *cpCil1* locus did not amplify for the Choctaw cultivar, and there was no amplification for the *cpCil8* and *cpCil9* loci in the Chickasaw cultivar. For the analysis of the genetic divergence of the cultivars, 10 nuclear (SSR) and 13 plastid (cpSSR) microsatellite markers were tested. Through the analysis of principal coordinates (PCoA) of

the nSSR, the two components explained, in total, 21.39% of the total variability observed, while for the cpSSR the total of 84.65%. Despite the high cophenetic correlation value (cpSSR= 0.9931 and nSSR= 0.9791), *bootstrap* values in the larger clusters were low. The results obtained in this study showed high genetic diversity for both nuclear ( $I= 2.365$ ) and plastid ( $I= 2.183$ ) markers. This estimate of genetic diversity is very significant, and this result can be used to determine the genetic pattern of the analyzed cultivars. This is the first study on the development and characterization of species-specific plastid microsatellite markers for *C. illinoensis* with potential applicability to generate information about genetic diversity among cultivars of the species. The data obtained so far allow us to help in deepening the analysis of genetic diversity, because through recent studies, open pollination promotes high genetic segregation, and propagation through seeds may have introduced hundreds of pecan varieties around the world. This great diversity in the form of the fruits, the trees, the quality of the nuts and their different ways of reproducing, still causes a great confusion regarding the nomenclature and the identification.

Keywords: cpDNA, nDNA, SSR markers, chloroplasts, DNA sequencing, plastoma.

## LISTA DE FIGURAS

### INTRODUÇÃO GERAL

**Figura 1.** (A) noqueira-pecã; (B) fruto maduro; (C) frutos verdes; (D) flor pistilada; (E) flor estaminada; (2E) detalhe da flor estaminada; (F) amêndoa ..... 22

### CAPÍTULO 2

**Figura 1.** Amplificação e eletroforese *in silico*. O SimGel corresponde ao marcador ou *ladder* em pares de bases, sendo a região avaliada entre 90 e 500 pb. São *primers* que demonstraram estar de acordo com os critérios propostos durante a validação..... 73

**Figura 2.** Número de *loci* amplificados em cada um dos gêneros analisados. À esquerda pode-se observar a árvore filogenética que representa as relações de distância entre os seis gêneros que pertencem à tribo Juglandoideae, adaptado de Xiang et al. (2014)..... 76

**Figura 3.** Géis de eletroforese (agarose 2,5%) com os produtos da amplificação dos 10 *locus* SSR plastidiais selecionados a partir do sequenciamento do plastoma de *C. illinoensis* cv Imperial, para as 17 amostras de noqueira-pecã, para cada marcador, e a esquerda e direita com *ladder* de 100pb..... 80

### CAPÍTULO 3

**Figura 1.** . Número de alelos por *locus* para cada um dos 10 nSSR analisados nas cultivares de *C. illinoensis*. ..... 99

**Figura 2.** Número de alelos por *locus* para cada um dos 13 cpSSR analisados nas cultivares de *C. illinoensis*. ..... 99

**Figura 3.** Análise de coordenadas principais das cultivares de Noqueira-pecã, com base nos dados genéticos moleculares nucleares..... 103

**Figura 4.** Dendrograma de UPGMA baseada na distância euclidiana de 15 amostras de noqueira-pecã com marcadores nucleares..... 104

**Figura 5.** Análise dos componentes principais das cultivares de noqueira-pecã, com base nos dados genéticos moleculares plastidiais..... 109

**Figura 6.** Dendrograma de UPGMA baseada na distância euclidiana de 17 amostras de noqueira-pecã com marcadores plastidiais..... 111

## LISTA DE TABELAS

### INTRODUÇÃO GERAL

<b>Tabela 1.</b> Protocolo CTAB 2% modificado utilizado para extrações de DNA .....	34
---	----

### CAPÍTULO 2

<b>Tabela 1.</b> Regiões microssatélite localizadas no genoma plastidial de <i>C. illinoensis</i> cv Imperial .....	70
---	----

<b>Tabela 2.</b> <i>Loci</i> microssatélites prospectados considerados ótimos para teste nas variedades de <i>Carya illinoensis</i> .....	72
---	----

<b>Tabela 3.</b> Transferibilidade de <i>primers</i> microssatélite em espécies da família Juglandaceae.....	75
--	----

<b>Tabela 4.</b> Número da amostra e as respectivas cultivares de <i>C. illinoensis</i> .....	79
---	----

<b>Tabela 5.</b> Tamanho do alelo observado (pb) obtidas para as 17 amostras de <i>C. illinoensis</i> . ...	81
---	----

<b>Tabela 5.</b> Tamanho do alelo observado (pb) obtidas para as 17 amostras de <i>C. illinoensis</i> (Continuada).....	82
---	----

### CAPÍTULO 3

<b>Tabela 1.</b> Número da amostra e as respectivas cultivares de <i>C. illinoensis</i> para cada numeração. ....	94
---	----

<b>Tabela 2.</b> <i>Loci</i> microssatélite prospectados e universais testados nas variedades de <i>Carya illinoensis</i> .....	96
---	----

<b>Tabela 3.</b> <i>Loci</i> microssatélite nucleares espécie-específicos para <i>Carya illinoensis</i> .....	97
---	----

<b>Tabela 4.</b> Número de alelos por locos ( $A$ ), Número efetivo de alelos por locos ( $A_e$ ), Índice de Shannon ( $I$ ), heterozigosidade observada ( $H_O$ ) e esperada ( $H_E$ ), Índice de fixação ( $F$ ), para os 10 <i>locus</i> microssatélite nucleares em 11 cultivares de <i>C. illinoensis</i> .....	101
--	-----

<b>Tabela 5.</b> Índices de diversidade para os 13 <i>locus</i> microssatélite plastidiais em 13 cultivares de <i>C. illinoensis</i> .....	106
--	-----

## LISTA DE ABREVIATURAS E SIGLAS

- °C – Graus Celsius
- AFLP** – Amplified fragment length polymorphism - Polimorfismo de comprimento de fragmento amplificado
- ATP** – Adenosine TriPhosphate – Adenosina trifosfato
- BSA** – Bovine Serum Albumin - Albumina de soro bovino
- ccmp2*** – Nomenclatura de *primer*
- CTAB** – Cationic hexadecyl trimethyl ammonium bromide - Catiônico de hexadecil trimetil amônio
- cDNA** – Complementary DNA – DNA complementar
- cm** – Centímetros
- cpCil*** – Nomenclatura de *primer*
- cpDNA** - Chloroplast DNA – DNA plastidial
- CDS** – Região codificante
- cpSSR** – Marcador molecular microsatélite localizado no genoma plastidial
- DNA** - Ácido desoxirribonucleico
- dNTP** – Trifosfato desoxinucleotídeos
- EST** – Expressed sequence tags
- IBGE** - Instituto Brasileiro de Geografia e Estatística
- IBPecan** – Instituto Brasileiro de Pecanicultura
- h*** – Diversidade de haplótipos
- ha** – Hectares
- INC** - International Nut and Dried Fruit Council
- INTA** - Instituto Nacional de Tecnología Agropecuária
- IR** – Regiões de repetição invertida
- ISSR** - Inter Simple Sequence Repeats - Repetições de sequência inter simples
- Kg** – Quilograma
- LFGDV** – Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal
- LSC** – Large single-copy - Região de cópia única longa
- m** – Metros
- MAPA** – Ministério da Agricultura, Pecuária e Abastecimento
- MgCl<sub>2</sub>** – Cloreto de magnésio
- mRNA** – RNA mensageiro
- Na*** – Número de alelos
- Ne*** – Número efetivo de alelos
- NCBI** – National Center for Biotechnology Information
- NGS** – Next generation Sequencing - Sequenciamento de nova geração
- ng** – Nanograma
- nm** – Nanômetro
- nSSR** – Marcador molecular microsatélite localizado no genoma nuclear
- ntcp9*** – Nomenclatura de *primer*
- ntcp40*** – Nomenclatura de *primer*

**PAGE** – Polyacrylamide gel electrophoresis - Eletroforese em Gel de Poliacrilamida  
**pb** – Pares de base  
**PCR** – Polymerase Chain Reaction - Reação em cadeia da polimerase  
***pm-cin*** – Nomenclatura de *primer*  
***pm-ca*** – Nomenclatura de *primer*  
***pm-ga*** – Nomenclatura de *primer*  
**PVP** – Polivinilpirrolidona  
**RAPD** – Random amplified polymorphic DNA - DNA polimórfico amplificado ao acaso  
**RFLP** – Restriction fragment Length Polymorphism - Polimorfismo de comprimento de fragmento de restrição  
**RNA** – Ácido ribonucleico  
**RNase** – Nucleases que catalisam a degradação do ácido ribonucleico  
**Rs** – Riqueza alélica  
**s** – Segundos  
**SNP** – Single Nucleotide Polymorphism - Polimorfismo de nucleotídeo simples  
**SPCR** – Software para análise de PCR *in silico*  
**SRA** – Sequence read archive - Arquivo de leitura de sequência  
**SSC** – Short single-copy - Região de cópia única curta  
**SSR** – Simple sequence repeats - Sequências Simples Repetidas  
***uh*** – Diversidade não-viesada  
**UFMS** – Universidade Federal de Santa Maria  
**UFSC** – Universidade Federal de Santa Catarina  
**UNIPAMPA** – Universidade Federal do Pampa  
**UV** – Ultravioleta  
**μL** – Microlitro  
**μM** – Micromolar  
***1-A05*** – Nomenclatura de *primer*

## SUMÁRIO

<b>1 INTRODUÇÃO GERAL .....</b>	<b>19</b>
<b>2 REVISÃO BIBLIOGRÁFICA.....</b>	<b>21</b>
2.1 <i>Carya illinoensis</i> .....	21
2.2 Diversidade genética .....	24
2.3 Marcadores moleculares .....	25
2.4 Marcadores microssatélite .....	26
2.5 Marcadores microssatélite plastidiais e nucleares .....	27
2.6 Prospecção de marcadores microssatélite .....	28
2.7 Transferibilidade de marcadores microssatélite .....	30
2.8 Divergência genética .....	31
<b>3 JUSTIFICATIVA.....</b>	<b>32</b>
<b>4 OBJETIVOS .....</b>	<b>33</b>
4.1 Objetivo geral .....	33
4.2 Objetivos específicos.....	33
<b>5 MATERIAL E MÉTODOS .....</b>	<b>33</b>
5.1 Material vegetal .....	34
5.2 Extração e quantificação do DNA .....	34
5.3 Sequenciamento do genoma plastidial de <i>C. illinoensis</i> cv. Imperial.....	35
5.4 Prospecção e desenho dos <i>primers</i> SSR.....	35
5.5 Validação dos marcadores SSR plastidiais de <i>C. illinoensis</i> .....	35
5.6 Transferibilidade dos marcadores SSR plastidiais de <i>C. illinoensis in silico</i> .....	35
5.7 Caracterização genética das cultivares de <i>C. illinoensis</i> .....	36
<b>6 REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>36</b>
<b>Capítulo 01</b>	
<b>Características estruturais, evolutivas e filogenômicas do genoma plastidial de <i>Carya illinoensis</i> cv. Imperial.....</b>	<b>45</b>
<b>Capítulo 02</b>	
<b>Prospecção, caracterização e validação de marcadores SSR a partir do cloroplasto da noqueira-pecã “Imperial” .....</b>	<b>63</b>

### **Capítulo 03**

<b>Análise da divergência genética entre cultivares de <i>Carya illinoensis</i> (Juglandaceae) através de marcadores microssatélite plastidiais e nucleares .....</b>	<b>90</b>
<b>CONSIDERAÇÕES FINAIS.....</b>	<b>119</b>

## 1 INTRODUÇÃO GERAL

A *Carya illinoensis* [Wangenh.] K. Koch (nogueira-pecã) é uma espécie frutífera florestal, nativa dos Estados Unidos (REIGER, 2004), e tem distribuição predominante nas regiões temperadas do Hemisfério Norte. No contexto sul-americano, o cultivo brasileiro de noqueira-pecã se destaca como um dos principais países produtores do fruto. A estimativa de área plantada no Brasil supera os oito mil hectares, ocorrendo em propriedades inferiores a 15 hectares, geralmente de base familiar, sendo que mais de 70% da área está no Rio Grande do Sul, seguido por Paraná e Santa Catarina (MARTINS et al., 2018).

A cultura de noqueira-pecã foi introduzida por intermédio de norte-americanos, em meados dos anos de 1900, nos municípios de Santa Bárbara e Americana, no Estado de São Paulo (BACKES e IRGANG, 2004). No Estado do Rio Grande do Sul, a cidade de Anta Gorda foi à pioneira no cultivo de noqueira-pecã, que teve início em 1943, quando houve o ingresso de quatro mudas oriundas de Kentucky (EUA), que existem no município até hoje e são consideradas as ancestrais da maior parte das noqueiras-pecã encontradas no município e na região (FRONZA et al., 2013).

A cultura da noqueira-pecã começou a ser explorada de maneira econômica apenas na década de 1970 (RASEIRA, 1990). Conforme Duarte e Ortiz (2001), essa expansão deveu-se a um programa federal de incentivos para florestamento e reflorestamento por meio de leis de incentivos fiscais. Nessa época grandes pomares foram implantados, alguns superando 100.000 mudas.

A área de cultivo com esta espécie tem se expandindo por se tratar de um fruto nobre com elevado valor de mercado e simples manejo, além da utilização na confeitaria, a madeira da noqueira-pecã é muito apreciada em pisos e mobiliários sendo possível o consórcio com outras culturas e animais. Deste modo, aperfeiçoa-se a área de produção, principalmente em propriedades familiares (MARCHIORI, 1997).

Existem mais de mil cultivares da espécie, que exibem variação na forma dos frutos, qualidade da noz, arquitetura da árvore e características reprodutivas. Esse fato deve-se ao alto índice de segregação gênica, em função da reprodução sexual através de sementes, que se constitui no principal método de propagação (ALMEIDA et al., 2002, DE OLIVEIRA et al., 2021). As mais importantes cultivares comerciais de noqueira-pecã existentes no Brasil são provenientes dos Estados Unidos, sendo que, das 41 cultivares registradas, somente duas foram desenvolvidas no Brasil (MAPA, 2021). As cultivares mais plantadas no Brasil são norte-americanas, dentre elas, destacam-se Barton, Shawnee, Cape Fear, Caddo, Choctaw e

MoneyMaker. Além destas, alguns produtores optam por produção através de mudas derivadas de pés-francos, as quais não são enxertadas e demandam um período mais longo para entrar em fase de plena produção.

A *C. illinoensis*, ainda que seja bastante comercializada por seus frutos exibirem elevados valores nutricionais, é desprovida de informações genéticas. Existem muitos materiais, mas não se sabe se eles são realmente diferentes, em decorrência da insuficiência de controle de polinização e de coleta de propágulos nos plantios, uma vez que sua propagação é efetuada principalmente por enxertia de borbulhia, no verão, ou de garfagem, no inverno, sobre porta enxertos oriundos de sementes (DONADIO, 1998).

Os marcadores microssatélite, também conhecidos como SSR (*Simple Sequence Repeats*), são marcadores moleculares eficientes para gerar informações que auxiliam em diversos estudos genéticos. Além de úteis para construção de mapas genéticos (GUPTA e VARSHNEY, 2000), os SSR têm sido utilizados para estudar diversidade de populações naturais (BARROSO et al., 2010) e de espécies cultivadas (ALMEIDA et al., 2009), permitindo analisar desde indivíduos até espécies proximamente relacionadas, devido ao fato de que as sequências de DNA que flanqueiam os microssatélites, são geralmente conservadas dentro de uma mesma espécie ou entre espécies de gêneros correlatos, o que permite o desenho de *primers* para ampliações específicas desses *locus* (OLIVEIRA et al., 2006).

Principalmente devido à sua natureza específica, codominância e alto polimorfismo, os marcadores microssatélite têm sido empregados de forma eficaz para avaliar a diversidade genética, diferenciação, estrutura genética intrapopulacional e fluxo gênico interpopulacional (NAGEL et al., 2015; STEFENON et al., 2016). O genoma plastidial pode exibir herança uniparental, com algumas exceções, ausência de recombinação e é conservado, sendo os microssatélites plastidiais indicados para utilização em estudos evolutivos, filogeográficos, filogenéticos e de fluxo gênico. Nesse sentido, a caracterização molecular a partir de marcadores de DNA pode solucionar este problema, distinguindo cultivares e acessos, e propiciando, simultaneamente, as bases para o planejamento de um eventual programa de melhoramento genético. Além disso, contribuiria, também para o progresso do sistema de manejo dos pomares, com repercussão direta na produtividade, pela necessidade do emprego de materiais genéticos distintos em decorrência das cultivares apresentarem maturação das estruturas reprodutivas em épocas diferentes. Assim, a liberação do pólen ocorre em tempos diferentes e, quanto maior for o período de polinização, maior será o sucesso de frutificação e de fecundação (STELLA e LUCCHESI, 2015).

Até o momento, não há registro na literatura de marcadores microssatélite plastidiais

desenvolvidos em *C. illinoinesis*. Assim, com a caracterização e validação dos marcadores SSR espécie-específicos, estudos genéticos poderão ser realizados empregando essas marcas e assim avançar na caracterização genética da espécie. Além disso, os marcadores desenvolvidos para *C. illinoinesis* poderão apresentar-se como uma ferramenta promissora à transferibilidade e utilização em outras espécies da família Juglandaceae, por exemplo. A presente tese está organizada em três capítulos, precedidos pela revisão bibliográfica. O primeiro capítulo aborda a caracterização estrutural, evolutiva e filogenômica do genoma plastidial de *C. illinoinesis* cv. Imperial. O segundo capítulo apresenta a prospecção, caracterização e validação dos primeiros marcadores microssatélite plastidiais desenvolvidos para a espécie, assim como a transferebilidade *in silico* para espécies pertencentes à família Juglandaceae. O terceiro capítulo descreve a aplicação de marcadores SSR plastidiais e nucleares na caracterização e análise da divergência genética entre cultivares de *C. illinoinesis*.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 *Carya illinoensis*

*Carya illinoensis* (Wangenh) C. Koch é uma espécie florestal frutífera de clima temperado, popularmente conhecida como noqueira-pecã, pertencente à família Juglandaceae, originária dos Estados Unidos, nativa das matas ciliares da América do Norte, desde o Nebraska, Iowa, Illinois, Texas, até Oaxaca, no sul do México (FRONZA et al., 2013; MARCHIORI, 1997; ORO, 2007).

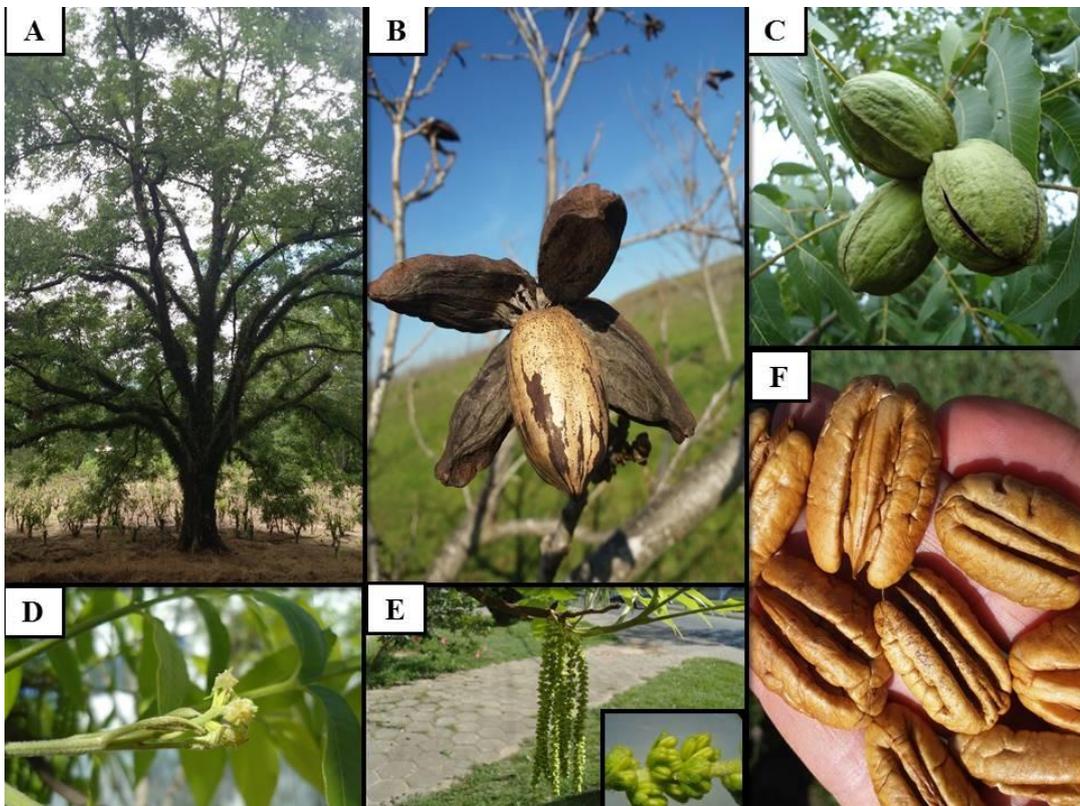
Atualmente, os maiores produtores de noz-pecã são os Estados Unidos e o México, mas também vem sendo cultivada na Argentina, Uruguai, Chile, Brasil e Austrália. No Brasil, o cultivo da *Carya illinoensis* se estende da região Sul ao Estado de Minas Gerais, sendo que, na região Sul, em particular, as maiores extensões de área plantada estão nas regiões do Vale do Taquari, Rio Pardo e Central do Estado do Rio Grande do Sul (POLETTTO et al., 2015).

No Brasil, a noqueira-pecã foi introduzida em meados de 1900 por norte-americanos, nos municípios de Santa Bárbara e Americana, no estado de São Paulo. A madeira é de boa qualidade, com 720 kg.m<sup>-3</sup> de peso específico, podendo ser utilizada em pisos e móveis. A planta é caducifólia e atinge até 50 m de altura e 150 cm de diâmetro. Sua propagação é feita

por sementes, e as cultivares são enxertadas, sendo que a produção de nozes se inicia entre seis e 10 anos de idade (BACKES; IRGANG, 2004).

Segundo definição de Rieger (2006), a espécie é monóica com inflorescências pistiladas pendentes em grupos laterais de 2-3 inflorescências, possuindo apenas anteras protegidas por brácteas esverdeadas. É decídua, com folhas compostas e imparipinadas, com folíolos oblongo-lanceolados de margem serrilhada, casca de coloração acinzentada e tronco ereto relativamente curto. Segundo o mesmo autor, devido à extensa área de ocorrência da espécie, há também uma ampla diferença entre a maturação das nozes nas distintas regiões dos Estados Unidos.

Marchiori (1997) classifica as inflorescências pistiladas como amentilhos pendentes. As inflorescências estaminadas são bem pouco evidentes e também aparecem em ramos laterais terminais. A parede do ovário é fundida à bráctea ou involúcro de tecido e a parede externa do ovário, juntamente com o involúcro, torna-se a parte carnosa do fruto. As flores são incompletas, pois não possuem pétalas e sépalas (Figura 1). Ocorre dicogamia na espécie, isto é, a maturação das flores estaminadas e pistiladas ocorrem em períodos diferentes na mesma planta, por isso há necessidade de plantio de linhas de outra cultivar polinizadora para que ocorra a fecundação em pomares.



**Figura 1.** (A) noqueira-pecã; (B) fruto maduro; (C) frutos verdes; (D) flor pistilada; (E) flor estaminada; (2E) detalhe da flor estaminada; (F) amêndoa. **Fonte:** Autor, 2022.

Os padrões de herança para o tipo de dicogamia (protoginia ou protandria) foram determinados a partir de estudos de cruzamento realizados pelo programa de melhoramento de pecã do Departamento de Agricultura dos Estados Unidos, em Brownwood, Texas. Um clone (a cultivar Mahan) produziu progênie totalmente protogínica ao passo que cruzamentos entre clones protândricos geraram apenas progênies protândricas. Um único par de alelos gera o tipo de dicogamia, com a protoginia sendo dominante (PP ou Pp) em relação à protandria, que é recessiva (pp) (THOMPSON e ROMBERG, 1985). A polinização é do tipo anemófila, sendo que seu pólen pode ser carregado pelo vento até 50 metros de distância (FRONZA et al., 2013; POLLETO et al., 2015; RASEIRA, 1990; STELLA e LUCCHESE, 2015).

A propagação da *C. illinoensis* é normalmente conseguida por enxertia de cultivares melhoradas sobre mudas de porta-enxertos. Contudo, esse método de propagação, apesar de produzir mudas clonais dotadas de uniformidade, tem como desvantagem o tempo que demanda, o custo e a baixa sobrevivência das mudas (RENUKDAS et al., 2010).

É uma espécie de grande importância em sistemas agroflorestais, pois, além de produzir frutos saborosos, possui madeira de boa qualidade, considerada nobre em seu país de origem, e muito empregada na fabricação de móveis vergados (BELTRAME et al., 2012; GATTO et al. 2012). Seus frutos possuem alto valor nutricional, sendo uma fonte de potássio, tiamina, zinco, cobre, magnésio, fósforo, niacina, ácido fólico, ferro e vitamina B6, e, também, são uma ótima fonte de fibras, além de ser ricos em ácido oleico, que é uma gordura monoinsaturada que pode auxiliar na prevenção de doenças cardíacas (MARTINS et al., 2018).

Os frutos são comercializados para o consumo *in natura* e para emprego na indústria farmacêutica e, além das nozes, outras partes da planta têm sido empregadas, principalmente as folhas. Também são reportados outros usos da *C. illinoensis*, como em reposições de áreas de preservação permanente, auxiliando a rentabilidade das propriedades familiares, as quais, por conta da manutenção de áreas de preservação, têm que abandonar áreas de cultivo agrícola, e podem fazer do cultivo da espécie, uma alternativa produtiva (CRUZ et al., 2014; STELLA e LUCCHESE, 2015; TERABE et al., 2008).

Dentre os fatores a serem levados em consideração na produção de nozes, está à escolha das cultivares a serem implantadas, sendo importante que o plantio seja realizado com mais de uma cultivar, pois cada uma exibe características individualizadas em relação à precocidade de produção e polinização, resistência a doenças, tamanho do fruto, rendimento e qualidade (POLETO et al., 2012; STELLA e LUCCHESE, 2015). Nesse sentido, as principais cultivares de *C. illinoensis* plantadas no Brasil são provenientes dos Estados

Unidos, sendo as mais importantes: Mahan, Frotscher, Schley, Success e Moneymaker Barton, Shawnee, Cape Fear, Chickasaw e Choktaw, Desirable, Melhorada, Imperial, Importada, Burkett, Chpecear, Shoshon (POLETTTO et al., 2012; ROVANI et al., 2015). Muitas dessas cultivares foram desenvolvidas no final do século XIX e anos iniciais do século XX por viveiristas que não realizavam registros e que não empregavam práticas de proteção contra polinizações cruzadas indesejadas em seus pomares. Como consequência, somente o parental de semente de muitas cultivares é conhecido, ou nenhum deles (CONNER e WOOD, 2001). No Brasil, não há muitos dados na literatura técnica ou científica, de informações relacionadas à caracterização morfoagronômica nem tampouco molecular das cultivares utilizadas.

Na Universidade da Geórgia, nos Estados Unidos, 100 marcadores RAPD foram utilizados para estimar a relação genética entre 43 cultivares comerciais de *C. illinoensis*, nas quais estão incluídas cinco (Desirable, Barton, Mahan, Moneymaker e Success) das sete cultivares conhecidas avaliadas no presente estudo, sendo obtido um *fingerprint* distinto para cada um dos genótipos estudados (CONNER e WOOD, 2001). Igualmente, também foi realizada uma análise da variabilidade genética em embriões somáticos de *C. illinoensis* via marcadores AFLP, em que, embriões derivados da mesma linha de cultura apresentaram alta similaridade. No entanto, dentro de uma mesma linha houve algumas diferenças entre embriões, que exibiram polimorfismo elevado (VENDRAME et al., 1999).

Além disso, um estudo analisou três conjuntos diferentes de características, a análise morfológica, química e genética (utilizando marcadores AFLPs) de acessos de noz-pecã (*C. illinoensis*) para avaliar alguns acessos de pecã crescendo no Sul do Brasil. Dados morfométricos, químicos e genéticos podem ser usados isolados ou combinados para determinar padrões de diversidade, caracterizar cultivares, e identificar plantas com produtos comerciais, farmacêuticos ou nutricionais potenciais (POLETTTO et al, 2020).

## **2.2 Diversidade genética**

As variações morfológicas, fisiológicas e comportamentais entre plantas e populações são consideradas relevantes diante das pressões de seleção do ambiente, sendo essas variações resultantes da variabilidade de alelos presentes na população (YOUNG et al., 1996). Assim, em termos de definição, a diversidade genética quantifica essa variabilidade genética existente (HUGHES et al., 2008).

Atualmente, com o desenvolvimento de ferramentas moleculares é possível quantificar

e caracterizar a diversidade genética das populações a partir de índices genéticos frequentemente utilizados em estudos de genética de populações, tais como: as frequências de genótipos e alelos, a proporção de *loci* polimórficos, a heterozigosidade observada e esperada ou ainda por meio do parâmetro de diversidade alélica (TORO e CABALLERO, 2005).

O conhecimento da diversidade genética entre genótipos de uma população de interesse é relevante para programas de melhoramento, pois permite a organização do germoplasma e uma amostragem mais eficiente de genótipos (NIENHUIS et al. 1995). Esta estimativa pode ser obtida a partir de marcadores moleculares e/ou caracteres agrônômicos (morfológicos), sendo que a primeira se sobressai por não ter interferências ambientais.

O estudo da diversidade genética entre acessos de uma cultura, além de permitir a identificação de materiais genéticos muito próximos ou duplicados, sugere aqueles genótipos mais distantes geneticamente, os quais poderão ser recomendados para futuros programas de policruzamentos no desenvolvimento de cultivares melhoradas (SCAPIM et al., 2002).

A maioria das espécies exploradas comercialmente teve sua diversidade genética diminuída drasticamente em consequência da domesticação e das técnicas de seleção e melhoramento de plantas (SAAVEDRA e SPOOR, 2002). Além disso, muitos genótipos se perdem pela substituição por novas cultivares gerando, muitas vezes, o desaparecimento de variedades locais. Novas cultivares, em geral, apresentam base genética estreita, isto é, muito aparentadas entre si, e a predominância de um restrito número de genótipos ocupando grandes áreas de plantio tem sido considerada um risco para a agricultura, podendo ocasionar uma erosão genética (BORÉM e MIRANDA, 2005).

### **2.3 Marcadores moleculares**

Os marcadores moleculares representam características de DNA que são responsáveis por distinguir dois ou mais organismos, sendo estas herdadas geneticamente (MILACH, 1998). Atualmente, existem diversos tipos de marcadores moleculares baseados em DNA e estes se diferenciam pela tecnologia empregada para detectar variabilidade e diferenças entre organismos, assim como pelos custos, facilidades de utilização, consistência e repetibilidade (MILACH, 1998; WEISING et al., 2005).

Os marcadores de DNA são pouco influenciados pelo ambiente (BERED et al., 1997), o que constitui grande vantagem em análises genéticas, pois elimina a necessidade de homogeneização das condições em que os indivíduos a serem estudados estão se desenvolvendo. Os marcadores geram uma grande quantidade de informações sobre a

identidade genética, diversidade, frequência gênica de um determinado germoplasma, bem como permitem avaliar relações filogenéticas e, também, a criação de estratégias de conservação dos recursos genéticos.

Os marcadores podem ser aplicados na análise da diversidade genética e seleção de genitores; proteção de cultivares e *fingerprinting* do DNA (ou seja, padrão molecular do DNA); análise da pureza genética de sementes; mapeamentos de genes e características complexas; melhoramento assistido por marcadores moleculares, como, retrocruzamentos assistidos e seleção assistida, entre outros (FALEIRO, 2007; GUIMARÃES et al., 2009; LOPES et al., 2002; MILACH, 1999; TOPPA e JADOSKI, 2013).

A disponibilidade de marcadores genéticos é fundamental na biologia e no melhoramento de plantas para a caracterização e seleção de plantas, para o mapeamento de genes e para a avaliação e análises de diversidade genética (RUDD et al., 2005). A escolha do marcador a ser empregado depende muito de qual o resultado se espera, e da espécie em estudo, por isso, é importante identificar as características de interesse e estabelecer objetivos claros para o programa de melhoramento ou conservação (MILACH, 1998). Para exemplificar, os marcadores RAPD, AFLP e microssatélites são mais apropriados para testes de paternidade, estudos de variabilidade dentro da mesma espécie, identidade genética (AZOFEIFA-DELGADO, 2006; CONNER e WOOD, 2001; FALEIRO, 2007; MILACH, 1999).

## **2.4 Marcadores microssatélite**

Os marcadores microssatélite também conhecidos como marcadores SSR (Sequências Simples Repetidas), são unidades curtas de 2 a 5 pb repetidas em tandem (FALEIRO, 2007). Estes marcadores estão localizados em grandes quantidades e uniformemente distribuídos pelo genoma das plantas, além de serem codominantes, multialélicos e com alta heterozigosidade (ZUCCHI, 2003), sendo utilizados para medir e monitorar similaridade, variabilidade e relação genética entre genótipos, construir mapas genéticos e encontrar marcadores para características de interesse (MEHLENBACHER, 1995; PIGATO e LOPES, 2001; SALLA et al., 2002; UPADHYAY et al., 2004). Existem diversas vias para desenvolver marcadores microssatélites que são posteriormente usados em estudos genéticos e programas de melhoramento de espécies florestais. Exemplos destas vias são as bibliotecas genômicas, bibliotecas de cpDNA e sequenciamento, sendo a eficácia destas inquestionável. Entretanto, o custo de desenvolvimento destas vias é muito alto, embora o sequenciamento já

possua um custo mais reduzido (KANTETY et al., 2002).

As vantagens dos marcadores SSR vão desde sua codominância, o alto nível de polimorfismo que pode ser detectado até sua alta reprodutibilidade. Mas como dito anteriormente, existe um alto custo requerido no desenvolvimento de *primers* específicos, o que gera desvantagens (FALEIRO, 2007).

Para testes em laboratório é necessário amplificar microssatélites via PCR utilizando *primers* específicos para flanquear o DNA de interesse (FALEIRO, 2007). Esta metodologia gera muitos custos e a partir disto é importante realizar a validação *in silico* dos *primers* antes da síntese dos mesmos (SARZI et al. 2019) e testar a transferibilidade de marcadores entre espécies (PETRY et al. 2019). De acordo com a literatura, obteve-se sucesso ao realizar a transferibilidade de *primers* entre algumas espécies da mesma família (SCHIAVON et al., 2009; LEITE et al., 2007; OLIVEIRA et al., 2010).

## 2.5 Marcadores microssatélite plastidiais e nucleares

Sequências que flanqueiam os microssatélites são bastante conservadas e utilizadas para o desenho de *primers* e a amplificação dos microssatélites. A diferença de tamanho entre os fragmentos amplificados permite detectar o polimorfismo. Essa diferença se deve ao número de repetições dentro dos microssatélites. Além disso, os microssatélites parecem ter uma distribuição frequente e aleatória, permitindo uma cobertura completa do genoma (RALLO et al., 2000).

As vantagens do uso de microssatélites sobre outros marcadores moleculares nos estudos genéticos são a reprodutibilidade, simplicidade e rapidez, pequena quantidade de DNA requerida, baixo custo de utilização, grande poder de resolução e altos níveis de polimorfismo (CHEN et al., 1997; LITT e LUTY, 1989). Além disso, os marcadores microssatélite podem ser usados para análises de transferibilidade, *pedigrees* e populações (BRONDANI et al., 1998). As principais desvantagens são o alto custo requerido no desenvolvimento de *primers* específicos, quando eles não estão disponíveis para a espécie a ser estudada (LITT e LUTY, 1989; MORGANTE e OLIVIERI, 1993; QUELLER et al., 1993).

O DNA circular plastidial (cpDNA) é principalmente conservado em espécies de plantas, contendo aproximadamente 130 genes, e compõem um genoma bastante confiável e informativo. Usado principalmente para taxonomia molecular, código de barras de DNA e como um recurso de informação de filogenia. As regiões de cópia única curta e longa (SSC e

LSC respectivamente) fornecem uma variedade de informações de genes codificadores que variam de complexo fotossistêmico, sintases de ATP a maturases, e RNAs ribossômicos (SUGIURA, 1992; JENSEN, 2013; MACHADO et al., 2017).

As sequências de cpDNA são úteis para resolver a filogenia de plantas em níveis altos de evolução por causa de suas taxas mais baixas de substituição silenciosa de nucleotídeos (RAUBESON e JANSEN, 2005). As sequências de muitos genes podem superar o problema de múltiplas substituições que causam a perda de informações filogenéticas entre as linhagens (LOCKHART et al., 1999). Além disso, caracteres estruturais em cpDNAs, como inversões de ordem/segmento de gene, expansão/contração das regiões de repetição invertida (IR) e perda/ganho de genes, podem servir como marcadores poderosos para inferência filogenética (RAUBESON e JANSEN 2005).

Dentre as regiões mais informativas do genoma de cloroplasto destacam-se os DNA microssatélites (cpSSR), que são repetições relativamente curtas e seu tamanho varia de 1 a 6 pares de bases. A natureza desses marcadores provavelmente está relacionada a erros ocorridos na replicação do DNA causados pela DNA polimerase (NAHUM, 2004).

O DNA plastidial tem menor porcentagem de componente não codificante em comparação com o DNA nuclear, embora os SSR sejam abundantes em genomas de cloroplasto. Em contraste com os marcadores de DNA nuclear que são herdados tanto da semente quanto do pólen, o cpDNA é herdado apenas por via materna nas angiospermas, com algumas exceções. É considerado um marcador altamente polimórfico que pode ser usado para rastrear divergências por meio de isolamento geográfico (PROVAN et al., 2001).

Enquanto *loci* SSR plastidiais (cpSSR) têm expressão haploide, os *loci* nucleares (nSSR) são marcadores codominantes, isto é, possibilitam identificar se o indivíduo estudado é homocigoto ou heterocigoto. Essa natureza codominante e o alto polimorfismo torna os nSSRs marcadores preferenciais para estudos de fluxo gênico, testes de paternidade e identificação de cultivares. Por sua vez, a combinação de cpSSRs e nSSRs permite análises complementares muito mais informativas.

## **2.6 Prospecção de marcadores microssatélite**

O desenvolvimento de marcadores microssatélite inicia com a prospecção das regiões SSR, posteriormente com o desenho dos iniciadores (também chamados de oligonucleotídeos, ou *primers*) complementares às regiões que flanqueiam a região SSR e é finalizada com a validação dos marcadores por meio da técnica de amplificação via PCR e visualização do

polimorfismo entre indivíduos (MASON, 2015; VIEIRA et al., 2016).

Com relação à prospecção dos SSR, existem alguns métodos que podem ser empregados. O primeiro deles é por meio da pesquisa em banco de dados, como por exemplo, no GenBank, para conferir se há dados de sequência genômica disponíveis da espécie de interesse para continuar com a identificação de regiões SSR na sequência disponível (MASON, 2015).

O outro método é a própria construção de bibliotecas genômicas clássicas enriquecidas para regiões microssatélites, onde basicamente é preciso inicialmente fragmentar o DNA e ligá-lo a adaptadores que serão inseridos em vetores para a transformação em *Escherichia coli* e após o DNA amplificado é sequenciado (VIEIRA et al., 2016).

E por fim, o outro método de prospecção dos SSR é via sequenciamento de DNA de nova geração (NGS – *Next Generation Sequencing*), sendo as plataformas Illumina as mais utilizadas para a detecção de SSR (VIEIRA et al., 2016). Este procedimento tem a vantagem de identificar grande quantidade de sequências contendo as regiões SSR para o desenvolvimento dos marcadores SSR, sendo que atualmente muitos estudos utilizaram NGS que determinaram muitas sequências de genoma ou transcriptoma para a descoberta de *locus* SSR em plantas (ZALAPA et al., 2012).

Os *primers* são um conjunto de oligonucleotídeos específicos que flanqueiam a sequência-alvo de SSR, sendo um iniciador direto (*forward*) para a direção 5' – 3' e outro iniciador reverso (*reverse*) para a direção 3' – 5' da fita complementar (MASON, 2015). Portanto, após desenhar e sintetizar os oligonucleotídeos, é necessário realizar a otimização das reações de amplificação via técnica de PCR, onde nesta etapa são realizados os ajustes de concentração dos reagentes, dos marcadores e de temperatura ideal para cada par de *primers*. A visualização do produto da PCR pode ser feita por meio de eletroforese em gel de agarose para verificar o tamanho dos fragmentos amplificados, porém para a visualização de polimorfismos outras duas técnicas podem ser utilizadas pelo fato de apresentarem maior sensibilidade: eletroforese em gel de poliacrilamida (PAGE - *Polyacrylamide Gel Electrophoresis*) e eletroforese em gel capilar (MASON, 2015).

Outra estratégia relevante é a possibilidade de transferibilidade dos marcadores microssatélites entre espécies relacionadas (VARSHNEY et al., 2005; WANG; BARKLEY e JENKINS, 2009), viabilizando os estudos com espécies em que ainda não validaram marcadores espécie-específicos. Fagundes et al (2016) enfatiza a dificuldade no desenvolvimento de marcadores SSR e a importância de testar a transferibilidade entre as espécies

## 2.7 Transferibilidade de marcadores microssatélite

Uma grande limitação dos marcadores microssatélite é o desenvolvimento de *primers*, que necessitam ser desenhados para cada *locus*. Apenas de 10 a 20 % dos *primers* isolados são informativos. A construção de biblioteca genômica envolve um processo trabalhoso, de alto custo e especialização nas técnicas para a caracterização do marcador. Porém, após os testes e obtenção dos *primers* para uma espécie, o custo e trabalho são reduzidos (GUPTA et al., 2002 apud VASCONCELOS, 2016).

Existem, basicamente, dois tipos de bibliotecas gênicas: a biblioteca genômica e a biblioteca de cDNA. Uma biblioteca genômica representa um conjunto de clones individuais, totalizando o genoma inteiro de um organismo. São utilizadas para a identificação de genes ou fragmentos de DNA específicos (PINTO, 2019). Esse método baseia-se no DNA genômico de alta qualidade e fragmentado usando enzimas de restrição. A enzima depende do comprimento médio desejado de fragmentos de DNA, da repetição de microssatélites a ser encontrada e do tipo de extremidades (coesivas ou contundentes) dos fragmentos de restrição. O DNA fragmentado é, então, selecionado por tamanho, preferencialmente, obtendo pequenos fragmentos (300-700 pb). Dependendo do método de fragmentação, os fragmentos de DNA são ligados em um vetor plasmídeo comum, diretamente ou após a ligação a adaptadores específicos, para a identificação das regiões microssatélites (ZANE et al., 2002).

Enquanto a biblioteca de cDNA contém apenas os genes que codificam produtos proteicos, se diferenciando da biblioteca genômica por usar mRNA, que a partir da transcriptase reversa é convertida em cDNA. Este tipo de biblioteca é utilizado com a finalidade de obter EST (*Expressed Sequence Tags*) (ZANE et al., 2002).

Atualmente, existe um grande número de marcadores microssatélite para genomas de plantas completamente sequenciados. Usando alguns programas de computador, os dados de sequências para os genes e os clones de cDNA podem ser baixados do GenBank e digitalizados para identificação de SSR, que são tipicamente referidos como EST-SSR ou microssatélites gênicos (VARSHNEY et al., 2005; ZANE et al., 2002).

Os microssatélites genômicos e gênicos podem ser transferidos para outras espécies baseados na homologia entre estas espécies (VARSHNEY et al., 2005), utilizando *primers* específicos heterólogos. Assim, facilitando a obtenção de dados em relação a espécies que não tem *primers* desenvolvidos. Os já desenvolvidos podem ser usados na identificação de espécies filogeneticamente próximas (VARSHNEY et al., 2005).

Tanto marcadores microssatélite gênicos como os genômicos podem ser transferidos

entre espécies. Porém, devido a uma maior conservação evolutiva e por estarem localizados em regiões expressas, estudos revelam que marcadores gênicos tendem a apresentar uma maior taxa de transferibilidade (KALIA et al., 2011).

## **2.8 Divergência genética**

Na realização de cruzamentos em programas de melhoramento, o conhecimento e o uso da diversidade genética disponível são de grande importância. Além de proporcionar informações sobre a variação genética existente entre genótipos, os estudos sobre diversidade genética têm possibilitado, também, a identificação de parentais divergentes para serem utilizados nos cruzamentos, buscando-se recombinações gênicas que possibilitem a obtenção de genótipos superiores. A utilização de cruzamentos entre parentais divergentes é importante, também, por evitar a depressão endogâmica (GOUVÊA, 2009).

Cruz e Carneiro (2003) relatam que melhoristas têm recomendado para a formação de população-base, o intercruzamento entre cultivares superiores e divergentes. Essa divergência pode ser avaliada a partir de caracteres agrônomicos, morfológicos, moleculares, entre outros. Dias (1994) ressalta que a hibridação envolvendo tipos parentais divergentes é passível de produzir segregantes transgressivos favoráveis em gerações avançadas. Logo, estudos sobre divergência genética entre parentais podem ter grande aplicabilidade no melhoramento. Segundo Dias et al. (1997), o conhecimento da divergência pode ser utilizado, também, na avaliação, descrição e classificação de recursos genéticos que visem a sua preservação.

A divergência genética tem sido avaliada por meio de técnicas biométricas, baseadas na quantificação da heterose, ou por processos preditivos. Dentre os métodos fundamentados em modelos biométricos, citam-se as análises dialélicas, que exigem a avaliação dos parentais e de todas suas combinações híbridas. Quanto aos métodos preditivos, dispensam a obtenção prévia das combinações híbridas (CRUZ et al., 2004; CRUZ e CARNEIRO, 2003). Em relação aos métodos preditivos, Dias (1994) ressalta que a seleção de combinações promissoras baseada no desempenho dos genótipos per se tem natureza preditiva, significando economia de recursos financeiros, de tempo e mão de obra, sendo que, a seleção de parentais pelo próprio desempenho evita que grande número de cruzamentos seja realizado, em contraste aos métodos de quantificação da heterose que tem natureza realizada. Já Vieira et al. (2005) relatam que é de fundamental importância que o melhorista conheça profundamente o germoplasma disponível, em termos de desempenho agrônomico por si só, capacidade de combinação e dissimilaridade genética (divergência). Sendo que a estimativa da divergência

genética cresce em importância, pois quando combinada com o conhecimento do comportamento por si só dos parentais, pode ser uma alternativa à realização de cruzamentos dialélicos.

Segundo Cruz (2001), há uma grande variedade de procedimentos multivariados, sendo mais utilizados aqueles que avaliam a similaridade ou dissimilaridade entre acessos, com o objetivo de recomendar intercruzamentos entre parentais divergentes em programas de hibridação ou para administrar informações disponíveis em bancos de germoplasma. Na predição da divergência genética, vários métodos podem ser aplicados, sendo a escolha do método mais adequado determinada pela precisão desejada pelo pesquisador, pela facilidade da análise e pela forma como os dados foram obtidos (CRUZ et al., 2004).

Vários são os procedimentos multivariados e entre eles estão a Distância Euclidiana, Análise por Componentes Principais. Segundo Cruz, (2006) são finalidades de cada análise: a Distância Euclidiana é uma medida de dissimilaridade de grande importância em estudos de divergência genética em que se procura identificar parentais a serem utilizados em programas de hibridação; a Análise por Componentes Principais transforma um conjunto de variáveis em um novo conjunto, cujas variáveis são independentes entre si. Essa análise avalia a importância relativa das variáveis em estudos da diversidade genética

### **3 JUSTIFICATIVA**

O setor da fruticultura apresenta um crescimento ascendente na utilização dos frutos secos, sendo que a nível global se constituem, principalmente, por nozes, amêndoas, avelã, pistache, castanhas e pecãs (INTA, 2021). Os frutos secos produzidos no Brasil estão englobados, principalmente, pelos seguintes tipos de nozes: Castanha-do-Brasil; Castanha-de-caju; Macadâmia e Pecã (ou Pecan) (ORO, 2007). O Rio Grande do Sul nas décadas de 1960-1970 passou a ser explorado comercialmente, especialmente na Região Sul do Brasil. Segundo dados da Embrapa e do International Nut and Dried Fruit (INC), a produção em 2019 projetou o Brasil mundialmente, com safra recorde de 3,5 mil toneladas, tornando o país o quarto maior produtor do fruto.

Entretanto, atualmente, há expansão do cultivo, devido, principalmente, ao alto preço pago pela noz e pela composição nutricional benéfica a saúde. O cultivo de nogueira-pecã também denominado de pecanicultura e ou cultivo da pecaneira propicia além do fruto, a

madeira, que é utilizada pela indústria na fabricação de móveis (GATTO, 2006). A casca do fruto da noqueira-pecã é utilizada para a infusão de chá, que é rica em compostos fenólicos e taninos (PRADO, 2008).

Estudos relacionados à caracterização genética de cultivares de *C. illinoensis* utilizadas no Estado do Rio Grande do Sul podem auxiliar tanto na conservação da variabilidade genética como na melhoria das técnicas de manejo da produção dessa espécie. Além disso, podem constituir um subsídio importante para técnicos e produtores, fornecendo informações precisas acerca das cultivares e, como consequência, aumentando a qualidade do produto a ser comercializado e, desta forma, sua rentabilidade.

## **4 OBJETIVOS**

### **4.1 Objetivo geral**

Avaliar e caracterizar a diversidade molecular de cultivares *Carya illinoensis* plantadas no Brasil utilizando marcadores moleculares plastidiais e nucleares.

### **4.2 Objetivos específicos**

- a) Sequenciar e caracterizar o plastoma da cultivar Imperial;
- b) Desenvolver marcadores moleculares a partir do plastoma da cultivar Imperial;
- c) Validar os marcadores microssatélite plastidiais em cultivares de *C. illinoensis* cultivadas no Brasil;
- d) Caracterizar e analisar a divergência genética entre cultivares de *C. illinoensis* plantadas no Brasil, de maneira comparativa utilizando marcadores microssatélite plastidiais e nucleares.

## **5 MATERIAL E MÉTODOS**

As atividades referentes ao presente estudo foram desenvolvidas no Laboratório de Biologia Molecular da Universidade Federal do Pampa (UNIPAMPA) e no Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal da Universidade Federal de Santa Catarina (UFSC).

## 5.1 Material vegetal

Foram utilizadas 13 cultivares da espécie como material vegetal: Imperial (3), Moneymaker (1), Barton (2), Desirable (2), Jackson (1), Farley (1), Sucess (1), Chickasaw (1), Shawnee (1), Stuart (1), Importada (1), Choctaw (1) e Maham (1). Algumas cultivares são produzidas e comercializadas por diferentes viveiristas e, por isso, foram amostradas em duplicata ou triplicata, totalizando 17 plantas. O material foi gentilmente cedido pelo engenheiro florestal Tales Poletto, no âmbito do projeto “Caracterização molecular de cultivares da noqueira-pecã”, desenvolvido através da parceria entre UNIPAMPA, UFSC, UFSM E IBPecan. Após a coleta das folhas, estas foram imediatamente mantidas em PVP-40 2% (polivinilpirrolidona), e subsequentemente utilizadas para extração de DNA.

## 5.2 Extração e quantificação do DNA

O DNA genômico das 17 amostras (13 cultivares) foi extraído a partir das folhas usando o método CTAB (DOYLE e DOYLE, 1987), com modificações (Tabela 1). A quantificação do DNA foi realizada em gel de agarose 2%. A quantidade e a qualidade do DNA isolado foram verificadas através do espectrofotômetro NanoDrop 1000® (Thermo Fisher Scientific, Wilmington, EUA) em relação a concentração e razões de absorbância 260/280 e 260/230.

**Tabela 1.** Protocolo CTAB 2% modificado utilizado para extrações de DNA.

<b>Etapas</b>	<b>Aproximadamente 50 miligramas de amostra em microtubos de 2 ml</b>
<b>1</b>	Maceração: 7 Minutos em nitrogênio líquido mais dois ciclos de 30 segundos no equipamento TissueLyser®
<b>2</b>	Acrescentar 1 mL de CTAB 2%, 3µL de Proteinase K, 2µL de RNase A
<b>3</b>	30 minutos a 65°C
<b>4</b>	Acrescentar 800µL de Clorofórmio/Álcool Isoamílico (CIA) e homogeneizar por inversão por 3 minutos
<b>5</b>	Centrifugar a 14000 RPM por 5 minutos
<b>6</b>	Transferir 600µL do sobrenadante para um novo tubo
<b>7</b>	Acrescentar 400µL de Álcool Isopropílico a -20°C e manter por pelo menos 30 minutos a -20°C
<b>8</b>	Centrifugar a 10000 RPM por 5 minutos e descartar o líquido
<b>9</b>	Acrescentar 1mL de Álcool Etílico 70% a -20°C e manter por 5 minutos
<b>10</b>	Centrifugar a 14000 RPM por 5 minutos e descartar o álcool
<b>11</b>	Acrescentar 1mL de Álcool Etílico Absoluto (99%) a -20°C e manter por 3 minutos descartando em seguida
<b>12</b>	Deixar o pellet secar e diluir em 100µL de TE

### 5.3 Sequenciamento do genoma plastidial de *C. illinoensis* cv. Imperial

Para o sequenciamento do genoma de *C. illinoensis* cv. Imperial o cpDNA foi isolado e usado para a preparação da biblioteca genômica com o sistema Ion OneTouch2™ usando o kit Ion PGM™ Template OT2 400. O sequenciamento foi realizado usando o kit Ion PGM™ Sequencing 400 no Ion PGM™ System com um Ion 318™ Chip v2. Os dados de sequência bruta foram depositados no banco de dados NCBI Sequence Read Archive (SRA) sob o número SSR10382885, Bioproject PRJNA587009, Biosample SAMN13174479. A montagem e anotação do genoma foram realizadas através de ferramentas específicas de bioinformática (descritos no CAPÍTULO 1).

### 5.4 Prospecção e desenho dos *primers* SSR

A prospecção de marcadores microssatélite foi realizado através do software SSR locator (MAIA et al., 2008) para encontrar SSR-di, tri e tetra. Para validar os marcadores foi realizada a simulação de eletroforese em gel com ajuda do software SIMGEL, incluído no pacote SPCR (CAO et al., 2005) (descritos no CAPÍTULO 2).

### 5.5 Validação dos marcadores SSR plastidiais de *C. illinoensis*

Para validar os marcadores foi realizada a simulação de eletroforese em gel com ajuda do software SIMGEL, incluído no pacote SPCR (CAO et al., 2005). A validação seguiu alguns critérios de avaliação como não apresentar *loci* sobrepostos, os alelos devem estar entre 90 e 500 pb, quando dois *loci* estiverem no tamanho desejado e não sobrepostos, eles devem ser testados em laboratório, *primers* ótimos devem ter apenas um *locus* na região desejada e mais que dois *loci* na região avaliada, foi desconsiderado para análises em laboratório. Após a validação *in silico*, foram sintetizados *primers* para 10 cpSSR da cultivar Imperial, os quais foram validados através da amplificação via PCR e separação por eletroforese em gel de agarose 2,5% para as 17 amostras (13 cultivares) coletadas (descritos no CAPÍTULO 2).

### 5.6 Transferibilidade dos marcadores SSR plastidiais de *C. illinoensis in silico*

Os *primers* desenvolvidos a partir de *C. illinoensis* cv Imperial foram testados

quanto à transferibilidade *in silico* através de buscas das sequências plastidiais obtidas no banco de dados do NCBI (National Center for Biotechnology Information). Os testes foram realizados em espécies da família Juglandaceae, a qual a espécie *C. illinoensis* pertence (descritos no CAPÍTULO 2).

### 5.7 Caracterização genética das cultivares de *C. illinoensis*

A caracterização genética foi feita com 23 *locus* SSR, sendo 13 marcadores microssatélite plastidiais, 10 prospectados do genoma plastidial de *C. illinoensis* cv. Imperial, além dos marcadores desenvolvidos para o genoma plastidial de tabaco *ccmp2*, *ntcp9* e *ntcp40*. Além desses, foram utilizados 10 marcadores SSR nucleares desenvolvidos para *C. illinoensis*. A caracterização dos genótipos foi realizada por meio de eletroforese em gel de agarose 2,5%. Os géis foram corados com GelRed e visualizados em transiluminador de luz UV. O *software* TotalLab TL120 1D v2009 foi usado para análise das bandas no gel de agarose 2,5%. A partir da análise das bandas, foram estimadas as frequências alélicas e os índices de diversidade genética, o número de alelos (*A*), o número efetivo de alelos (*A<sub>e</sub>*), diversidade de haplótipos (*h*), diversidade não-viesada (*uh*), a similaridade genética detectada pelos marcadores microssatélite foi explorada com Análises de Coordenadas Principais (PCoA) e a mensuração da diversidade genética molecular foi usado o Índice de Diversidade de Shannon (SHANNON, 1948) por meio do programa GenAlEx versão 6.5 (PEAKALL e SMOUSE, 2012).

O *software* Past 4.03 (HAMMER et al., 2001) foi utilizado para elaborar um dendrograma baseado no algoritmo euclidiano, baseados em uma correlação cofenética entre as cultivares (descritos no CAPÍTULO 3).

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, W.A.B. de. et al. Anatomia e ultra estrutura de embriões somáticos de pecan (*Carya illinoensis*). **Magistra**, Cruz das Almas, v. 14, n. 1, jan./jun., 2002.

ALMEIDA, V.C. et al. In situ and genetic characterization of *Gossypium barbadense* L. from the States of Para and Amapa, Brazil. **Pesquisa Agropecuária Brasileira**, v.44, n.7, p. 719-25, 2009.

AZOFEIFA-DELGADO, Á. Uso de marcadores moleculares en plantas; Aplicaciones en frutales del trópico. **Agronomía Mesoamericana**, 17(2): 221-242. 2006.

BACKES, P.; IRGANG, B. Árvores do Sul. **Guia de identificação & interesse Ecológico**. As principais espécies nativas Sul-Brasileiras. Santa Cruz do Sul: Instituto Souza Cruz, 204 p. 2004.

BARROSO, P.A.V. et al. In situ conservation and genetic diversity of three populations of *Gossypium mustelinum* Miers (ex Watt). **Genetic Resources and Crop Evolution**, v. 57, n.3, p.343- 9, 2010.

BELTRAME, R. et al. Resistência ao impacto da madeira de açoita-cavalo em diferentes condições de umidade. *Cerne*, Lavras v.16, n.4, 2010. Disponível em: <[http://www.dcf.ufla.br/cerne/artigos/03-02-20115854v16\\_n4\\_artigo%2009.pdf](http://www.dcf.ufla.br/cerne/artigos/03-02-20115854v16_n4_artigo%2009.pdf)>. Acesso em: 19 jun. 2021.

BERED, F.; BARBOSA NETO, J. F.; CARVALHO, F. I. F. Marcadores moleculares e sua aplicação no melhoramento genético de plantas. **Ciência Rural**, 27: 513- 520. 1997.

BORÉM A; MIRANDA GV. **Melhoramento de plantas**. Viçosa: UFV. 552p. 2005.

BRONDANI, R. P. V. BRONDANI, C.; TARCHINI, R.; GRATTAPAGLIA, D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetics**, v. 97, n. 5–6, p. 816– 827, 1998.

BRUNEL, D. A microsatellite marker in *Helianthus annuus* L. **Plant Molecular Biology**, v. 24, p. 397–400, 1994.

CAO, Y.; WANG, L.; XU, K.; KOU, C.; ZHANG, Y.; WEI, G.; HE, J.; WANG, Y.; ZHAO, L. **Information theory-based algorithm for in silico prediction of PCR products with whole genomic sequences as templates**. *BMC Bioinformatics*, v.6, n.190, 2005.

CHEN, K.; TANKSLEY, S. D. High-Resolution Mapping and Functional Analysis of se2.1 : A Major Stigma Exsertion Quantitative Trait Locus Associated With the Evolution From Allogamy to Autogamy in the Genus *Lycopersicon*. **Genetics**, v. 168, p. 1563–1573, 2004.

CONNER, P. J.; WOOD, B. W. Identification of Pecan Cultivars and Their Genetic Relatedness as determined by Randomly Amplified Polymorphic DNA Analysis. **J. AMER. SOC. HORT. SCI.** 126(4):474–480. 2001.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 585p. Viçosa: UFV, 2003.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 480p. Viçosa: UFV, 2004.

CRUZ, C.D. **Programa Genes: Análise multivariada e simulação**. Viçosa: UFV. 2006. 175p.

CRUZ, F. G. et al. Efectividad biológica de extractos de *Carya illinoensis*, para el control de *Meloidogyne incógnita*. **Revista Mexicana de Ciencias Agrícolas**, Vol.5 Núm.7, p. 1317-1323, 2014.

CRUZ, C.D. A informática no melhoramento genético. In: Nass, L.L.; Valois, A.C.C.; Melo, I.S.; Valadares-Ilgis, M.C. **Recursos genéticos e melhoramento: plantas**. Rondonópolis: Fundação MT, p.1085-1118, 2001.

DE OLIVEIRA L.O.; BEISE D.C.; DOS SANTOS D.D.; NAGEL, J.C.; POLETTO, T.; POLETTO, I.; STEFENON, V.M. Molecular markers in *Carya illinoensis* (Juglandaceae): from genetic characterization to molecular breeding, **The Journal of Horticultural Science and Biotechnology**, DOI: 10.1080 / 14620316.2021.1892534. 2021.

DIAS, L.A. dos S.; KAGEYAMA, P.Y.; CASTRO, G.C. Divergência fenética multivariada na preservação de germoplasma de cacau (*Theobroma cacao* L.). **Agrotrópica**, v.9, p.29-40, 1997.

DIAS, L.A.S. **Divergência genética e fenética multivariada na predição de híbridos e preservação de germoplasma de cacau (*Theobroma cacao* L.)**. 94p. Tese (Doutorado) - Escola Superior de Agricultura “Luiz de Queiroz” , Universidade de São Paulo, Piracicaba. 1994.

DONADIO, L.C. **Frutas Exóticas**. Jaboticabal, FUNEP, 1998. 279p.

DOYLE, J.J.; DOYLE, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, v.19, p.11-15, 1987.

DUARTE, V.; ORTIZ, E. R. N. Podridão de *Phytophthora* da amêndoa e da casca da noqueira pecan. In: LUZ, E. D. M. N. et al. (Ed.). **Doenças causadas por *Phytophthora* no Brasil**. Campinas: Rural, 2001. p. 493-508. FILIPPIN, I. L. **Viabilidade econômica do cultivo de noqueira pecã em áreas de reserva legal e de preservação permanente**. 2011. 74 f. Dissertação (Mestrado em Ciência e Tecnologia de Sementes) - Universidade Federal de Pelotas, Pelotas. 2011.

FAGUNDES, B. S., L. F. DA SILVA, R. M. GIACOMIN, D. SECCO, J. A. DÍAZ-CRUZ; P. R. DA-SILVA. Transferability of microsatellite markers among Myrtaceae species and their use to obtain population genetics data to help the conservation of the Brazilian Atlantic Forest. **Tropical Conservation Science** 9(1): 408-422. 2016.

FALEIRO, F. G. **Marcadores genético-moleculares aplicados a programas de conservação e uso de recursos genéticos**. Planaltina – DF, Embrapa Cerrados, 102p., 2007.

FIELD, D.; WILLS, C. Long, polymorphic microsatellites in simple organisms. **Proceedings. Biological sciences / The Royal Society**, v. 263, n. 1367, p. 209–15, 1996.

FRONZA, D.; POLETTO, T.; HAMANN, J.J. **O cultivo da noqueira-pecã**. Santa Maria: Universidade Federal de Santa Maria, Colégio Politécnico, Núcleo de Fruticultura Irrigada, 2013. KREUZ, Carlos Leomar; SOUZA, Alceu; CLEMENTE, Ademir. Custos de produção, expectativas de retorno e de riscos do agronegócio mel no planalto norte de Santa Catarina. **Revista Custos e agronegócio on line**. v. 4, n. 1, p. 46-61, Jan/abr. 2008. Disponível em: . Acesso em: 10 jun. de 2021.

GATTO, D. A. **Características tecnológicas do vergamento das madeiras de *Luehea divarticata*, *Carya illinoensis* e *Platanus x acerifolia* como subsidio para o manejo florestal**. 2006. 115f. Tese (Doutorado em Engenharia Florestal) – Curso de Pós-Graduação

em Engenharia Florestal, Universidade Federal de Santa Maria, Santa Maria, 2006.

GOUVÊA, L. R. L. **Divergência genética em seringueira estimada através de técnicas multivariadas e marcadores moleculares microssatélites**. Dissertação (Mestrado) – Instituto Agronômico (IAC), São Paulo. 2009.

GUIMARÃES, C. T. et al. Marcadores moleculares e suas aplicações no melhoramento genético. **Informe Agropecuário**, Belo Horizonte, v.30, n.253, nov./dez., 2009.

GUPTA, P.K.; VARSHNEY, R.K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. **Euphytica**, v.113, n.3, p.163-85, 2000.

HAMMER, Oyvind et al. PAST: Paleontological statistics software package for education and data analysis. **Palaeontologia electronica**, v. 4, n. 1, p. 9, 2001

HUGHES, A.R.; INOUE, B.D.; JOHNSON, M.T.; UNDERWOOD, N.; VELLEND, M. Ecological consequences of genetic diversity. **Ecological Letters**, v. 11, p. 609-623. 2008.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Lavoura permanente. 2014. Disponível em: <<http://www.ibge.gov.br/estadosat/temas.php?sigla=rs&tema=lavourapermanente2014>>. Acesso em: 12 jul. de 2021.

INTA. Instituto Nacional de Tecnología Agropecuaria. Frutos secos: evolución del mercado internacional y situación argentina. General Roca, Rio Negro, Junho, 2015. Disponível em: <<http://inta.gov.ar/documentos/frutos-secos-evolucion-delmercado-internacional-y-situacion-argentina>>. Acesso em: 8 jun. 2021.

JENSEN R.G. **Biochemistry of the chloroplast**. *Biochemistry Plants*. 1:273-313. 2013.

KALIA, R. J.; RAI, M. K.; KALIA, S.; SINGH, R.; DHAWAN, A.K. Microsatellite markers: an overview of the recent progress in plants. **Euphytica**, v. 177, p.309–334, 2010.

LI Y.C.; KOROL A.B.; FAHIMA T.; BEILES A.; NEVO E. **Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review**. *Mol Ecol*. 2002 Dec; 11(12):2453-65.

KANTETY, R.V.; ROTA, M.L.; MATHEWS, D.E.; SORRELLS, M.E. **Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat**. *Plant. Mol. Biol. Rep.* v. 48, p.501-510, 2002.

KUHN R.M et al. O banco de dados do navegador do genoma UCSC: atualização 2007 10.1093/nar/gkl928. **Nucl Acids Res.** 35: D668–673. doi: 10.1093/nar/gkl928. 2007.

LEITE, T. L.; MARCO, A. F.; TARCHETTI, B. D.; FERREIRA, M. A. J. F.; AMARAL, Z. P. S.; BUSO, G. S. C. **Análise de transferibilidade de primers microssatélites de *Cucumis melo* para *Curcubita moschata* e *Luffa cylindrica***. Brasília, 2007.

LITT, M.; LUTY, J. A. A Hypervariable Microsatellite Revealed by In Vitro Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. **American Journal of Human Genetics**, v. 44, p. 397–401, 1989.

LOCKHART P.J.; HOWE C.J.; BARBROOK A.C.; LARKUM A.W.D.; PENNY D. **Spectral analysis, systematic bias, and the evolution of chloroplasts**, Mol Biol Evol, vol. 16 (pg. 573-576). 1999.

LOPES, R. et al. Marcadores moleculares dominantes (RAPD e AFLP): Aspectos técnicos e interpretação genética. **Biociência**, n 29, 2002.

MACHADO L.O.; VIEIRA L.N.; STEFENON V.M.; PEDROSA F.O.; SOUZA E.M.; GUERRA M.P.; NODARI R.O. **Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences**. Genetica (2017) 145:163–174. 2017.

MAIA, L.C.; PALMIERI, D.A.; DE SOUZA, V.Q.; KOPP, M.M.; DE CARVALHO, F.I.; COSTA DE OLIVEIRA, A. SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. International Journal of Plant Genomics, v. 2008, n. 41 p. 26-96. 2008.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. Registro Nacional de Cultivares. Disponível em: <[http://extranet.agricultura.gov.br/php/proton/cultivarweb/cultivares\\_registradas.php](http://extranet.agricultura.gov.br/php/proton/cultivarweb/cultivares_registradas.php)> Acesso em: 12 jul. 2021.

MARCHIORI, J.N.C. **Dendrologia das angiospermas: das magnoliáceas às flacourtiáceas**. Santa Maria: Ed. da UFSM. 271 p. 1997.

MARTINS, C. R., CONTE, A.; FRONZA, D.; FILIPPINI ALBA, J. M.; HAMANN, J. J.; BILHARVA, M. G.; MALGARIM, M. B.; FARIAS, R. M.; DE MARCO, R.; REIS, T. S. **Situação e perspectiva da noqueira-pecã no Brasil**. Pelotas: Embrapa Clima Temperado, 2018. 31 p. (Embrapa Clima Temperado. Documentos, 462).

MASON, A.S. SSR Genotyping. In: BATLEY, J. **Plant Genotyping**. Springer, New York, NY, p. 77-89. 2015.

MILACH, S. C. K. Marcadores de DNA: aplicações no melhoramento de plantas. **Biociência**, n. 5. 1998.

MILACH, S. C. K. Marcadores moleculares nos recursos genéticos e no melhoramento de plantas. In: **Recursos Genéticos e Melhoramento de Plantas para o Nordeste Brasileiro**. QUEIRÓZ, M. A. de.; GOEDERT, C. O.; RAMOS, S. R. R. Petrolina – PE: Embrapa Semi-Árido/ Brasília – DF: Embrapa Recursos genéticos e Biotecnologia, nov. 1999.

MEHLENBACHER, S.A. **Classical and molecular approaches to breeding fruit and nut crops for disease resistance**. Hortscience, Alexandria, v.30, n.3,p.466- 477. 1995.

MOKOCHINSKI, F. M. **Estimativa de produção, caracterização física e perfil químico de amêndoas de noqueira-pecã**. Universidade Estadual do Centro-Oeste. Guarapuava- PR, 2015.

MORGANTE, M.; OLIVIERI, A. M. PCR-amplified microsatellites as markers in plant genetics. **The Plant Journal**, v. 3, n. 1, p. 175–82, 1993.

NAGEL J.C.; CECONI D.E.; POLETTO I.; STEFENON V.M. Historical gene flow within and among populations of *Luehea divaricata* in the Brazilian Pampa. **Genetica** 143: 317-329. <http://dx.doi.org/10.1007/s10709-015-9830-9>. 2015.

NAHUM, L.A. **Evolução dos genomas**. In: Mاتيoli S.R. (Ed). *Biologia Molecular e Evolução*. Holos editora, Ribeirão Preto, São Paulo. p. 82-96. 2004.

NIENHUIS, J. et al. Genetic relationships among cultivars and landraces of lima bean (*Phaseolus lanatus* L.) as measured by RAPD markers. **Journal of the America Society for Horticultural Science**, Alexandria, v.120, n.2, p.300-306, 1995.

OLIVEIRA, E.J. et al. Origin, evolution and genome distribution of microsatellites. **Genetics and Molecular Biology**, v.29, n.2, p.294-307, 2006.

OLIVEIRA, M. S. P.; SANTOS, J. B.; AMORIM, E. P.; FERREIRA, D.F. **Variabilidade genética entre acessos de açaizeiro utilizando marcadores microssatélites**. *Ciência & Agrotecnologia*, Lavras, v. 34, n. 5, 1253-1260. 2010.

ORO, T. **Composição nutricional, compostos bioativos e vida de prateleira de noz e óleo prensado a frio de noz-pecã [*Carya illinoensis* (Wangenh.) C. Koch]**, 2007. 105f. Dissertação (Mestrado em Ciência de Alimentos) – Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, Florianópolis, 2007.

ORTIZ, ORTIZ, E. R. N.; CAMARGO, L. E. A. Doenças da Nogueira Pecan. In: KIMATI, H.; AMORIM, L.; BERGAMIN FILHO, A.; CAMARGO, L. E. A.; REZENDE, J. A. M. (Eds.). **Manual de fitopatologia: doenças das plantas cultivadas**. 4. ed. São Paulo: Agronômica Ceres, 1997. cap. 54, p. 530-536.

ORTIZ, E. R. N.; CAMARGO, L. E. A. Doenças da Nogueira Pecan. In: KIMATI, H. et al. (Eds.). **Manual de fitopatologia: doenças das plantas cultivadas**. 4. ed. São Paulo: Agronômica Ceres. p. 501–505. 2005.

PIGATO, S.M.P.C.; LOPES, C.R. Avaliação da variabilidade genética em quatro gerações de *Eucalyptus urophylla* S.T. Blake por meio de marcador molecular RAPD. **Scientia Forestalis**, Piracicaba, n.60, p.119-133, 2001.

PINTO, N. C.. **Tecnologia do DNA Recombinante**. [S. l.], 2019. Disponível em: [http://www2.iq.usp.br/docente/nadja/QBQ3401\\_aula9.pdf](http://www2.iq.usp.br/docente/nadja/QBQ3401_aula9.pdf). Acesso em: 4 jan. 2022.

POLETTO, T. et al. Análise de características dos frutos de cultivares de noqueira-pecã cultivadas no Rio Grande do Sul. In: SIMPÓSIO DE ENSINO, PESQUISA E EXTENSÃO, XVI, **Anais...** Santa Maria: UNIFRA, 2012.

POLETTO, T. et al. Dormancy overcome methods in seeds of pecan *Carya illinoensis* (Wangenh.) K. Koch]. **Revista Árvore**, v.39, n.6, p.1111-1118. Available from: doi: 10.1590/0100-67622015000600014. 2015.

POLETTO, T., POLETTO, I., SILVA, L.M.M., MUNIZ, M.F.B., REINIGER, L.R., RICHARDS, N., STEFENON, V.M. Morphological, chemical and genetic analysis of southern Brazilian pecan (*Carya illinoensis*) accessions. **Scientia Horticulturae**, 261, 108863. doi:10.1016/j.scienta.2019.108863. 2020.

- PRADO, A. C. P. do. **Avaliação da atividade antioxidante da casca e torta de noz-pecã [*Carya illinoensis* (Wangenh) C. Koch]**, 2008. 131f. Dissertação (Mestrado em Ciência de Alimentos) – Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, Florianópolis, 2008.
- PROVAN J.; POWELL W.; HOLLINGSWORTH P.M. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. **Trends Ecol Evol**. 2001 Mar 1; 16(3):142-147.
- QUELLER, D. C.; STRASSMANN, J. E.; HUGHES, C. R. Microsatellites and Kinship. **Trends in Ecology and Evolution**, v. 8, n. 8, p. 285–288, 1993.
- RALLO, P.; DORADO, G.; MARTIN, A. Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). **Theoretical and Applied Genetics**, v. 101, p. 984–989, 2000.
- RASEIRA, A. **A cultura da noqueira pecã (*Carya illinoensis*)**. Embrapa Clima Temperado, Pelotas. Comunicado Técnico nº 63. 3p. 1990.
- RAUBESON L.A.; JANSEN R.K. A rare chloroplast-DNA structural mutation is shared by all conifers, **Biochem Syst Ecol**, vol. 20 (pg. 17-24).1992.
- RAUBESON, L. A.; JANSEN R. K. **Chloroplast genomes of plants**. Pp. 45–68 in R. Henry, ed. Diversity and evolution of plants—genotypic and phenotypic variation in higher plants. CABI Publishing, Oxfordshire, United Kingdom. 2005.
- REIGER, M., 2004. **Pecan – *Carya illinoensis* (Wangenh.) K. Koch**. Georgia: University of Georgia. Disponível em: <<http://www.uga.edu/fruit%20/peca.htm>>. Acesso em: 18 jun. 2021.
- RENUKDas, N. N.; MANOHARAN, M.; JR, J. O. G. In vitro propagation of pecan (*Carya illinoensis* (Wangenh) K. Koch). **Plant Biotechnology** 27, 211-215. 2010.
- RIEGER, M. **Introduction to Fruit Crops**. New York: food products press. 2006.
- ROVANI, F. F. M.; WOLLMANN, C. A.; MONTEIRO, A. Exigências bioclimáticas e riscos climáticos ao cultivo da noqueira pecã (*Carya illinoensis*) no Rio Grande do Sul, Brasil. In: **Atas do X Congresso Da Geografia Portuguesa: Os Valores da Geografia**, ROXO, M. J. et al., Ed. Associação Portuguesa de Geógrafos, Lisboa, set. 2015.
- RUDD, S.; SCHOOF, H.; MAYER, K. PlantMarkers--a database of predicted molecular markers from plants. **Nucleic Acids Research**, v. 33, p. 628–632, 2005.
- SAAVEDRA, G.; SPOOR, W. Genetic base broadening in autogamous crops: *Lycopersicon esculentum* Mill. as a model. **Managing Plant Genetic Diversity**, v.443, p.291-299. 2002.
- SALLA, M.F.S.; RUAS, C.F.; RUAS, P.M.; CARPENTIERI-PÍPOLO, V. **Uso de marcadores moleculares na análise da variabilidade genética em acerola (*Malpighia emarginata* D.C.)**. Revista Brasileira de Fruticultura, Jaboticabal, v.24, n.1, p.15- 22. 2002.
- SARZI, D. S. et al. Discovery and characterization of SSR markers in *Eugenia uniflora* L. (Myrtaceae) using low coverage genome sequencing. **Anais da Academia Brasileira de**

**Ciências**, v.91, n.1, p.1-8. Available from: <Available from: <http://dx.doi.org/10.1590/0001-3765201920180420>>. doi: 10.1590/0001-3765201920180420. 2019.

SCAPIM, C. A. et al. Análise dialéctica e heterose de populações de milho-pipoca. **Bragantia**, v. 61, n. 03, p. 219-230. 2002

SCHIAVON, A. L.; SANTOS, M. A.; SOUZA, R. C.; HUNGRIA, M. **Transferibilidade de marcadores microssatélites de soja para feijão comum (*Phaseolus vulgaris*)** In: IV Jornada acadêmica da Embrapa soja, Londrina, n. 312, p. 180-185. 2009.

STEFENON V.M., NAGEL J.C., CECONI D.E. e POLETTO I. Evidences of genetic bottleneck and fitness decline in *Luehea divaricata* populations from southern Brazil. **Silva Fennica** 50: 1566. 2016.

STELLA, A. L. S.; LUCCHESI, O. A. Avaliação da bibliografia livre como subsídio aos sistemas de cultivo de noqueira-pecã (*Carya illinoensis* (Wangenh) K. Koch). In: SEMINÁRIO DE INOVAÇÃO E TECNOLOGIA, 5., 2015, Ijuí. **Anais...** Ijuí: UNIJUÍ, p. 1-7. 2015.

SUGIURA M. The chloroplast genome. **Plant Molecular Biology**. 19:149-168. Thompson, T.E. & Romberg, L.D. 1985 Inheritance of heterodichogamy in pecan *J. Hered.* 76 456 458. 1992.

TERABE, N. I.; MARTINS, C. M.; HOMECHIN, M. Microrganismos associados a frutos de diferentes cultivares de noz pecan. **Ciência e Agrotecnologia**, Lavras, v. 32, n. 2, p. 659-662, mar./abr. 2008.

TOPPA, E. V. B.; JADOSKI, C. J. O uso dos marcadores moleculares no melhoramento genético de plantas. **Scientia Agraria Paranaensis**, v. 12, n. 1, p. 1- 5, 2013.

TORO, M. A.; CABALLERO, A. Characterization and conservation of genetic diversity in subdivided populations. **Philosophical Transactions of the Royal Society of London B: Biological Sciences**, v. 360, n. 1459, p. 1367-1378. 2005.

UPADHYAY, A.; JAYADEV, K.; MANIMEKALAI,R.; PARTHASARATHY, V.A. **Genetic relationship and diversity in Indian coconut accessions based on RAPD markers**. *Scientia Horticulturae* , Amsterdam, v.99, n.3-4, p.353-362. 2004.

VARSHNEY, R.K.; GRANER, A.; SORRELLS, M.E. Genic microsatellite markers in plants: features and applications. **TRENDS in Biotechnology**, v. 23, n. 1, p. 48-55. 2005.

VASCONCELOS, A. **Análise de diversidade genética em populações de *Campomanesia Adamantium* (Cambess.) O.Berg. Myrtaceae por meio de marcadores microssatélites transferíveis**. Monografia (Biotecnologia) – Universidade federal da grande Dourados, Dourados, 2016.

VENDRAME, W. A.; KOCHERT, G.; WETZSTEIN, H. Y. AFLP analysis of variation in pecan somatic embryos. **Plant Cell Reports**, 18: 853–857. 1999.

VIEIRA, E.A.; CARVALHO, F.I.F; OLIVEIRA, A.C.; BENIN, G.; ZIMMER, P. D.; SILVA, J.A.G.; MARTINS, A.F.; BERTAN, I., SILVA, G.O.; SCHIMIT, D.A. M. Comparação entre

medidas de distância genealógica, morfológica e molecular em aveia em experimentos com e sem a aplicação de fungicida. **Bragantia**, v.64, p.51-60, 2005.

VIEIRA, M.L.C.; SANTINI, L.; DINIZ, A.L.; MUNHOZ, C.D.F. Microsatellite markers: what they mean and why they are so useful. **Genetics and molecular biology**, v. 39, n. 3, p. 312-328. 2016.

WANG M.L.; BARKLEY N.A.; JENKINS T.M. **Microsatellite markers in plants and insects**. Part I: Applications and biotechnology. *Genes Genomes Genomics*, 3: 54–67. 2009.

WEISING, K.; NYBOM.; H, WOLFF, K. e KAHL G. **DNA fingerprinting in plants: Principles, Methods, and Applications**. 2nd edition. CRC Press, Boca Raton, p.444, 2005.

WEISSENBACH, J.; GYAPAY, G; DIB, C.; VIGNAL, A.; MORISSETTE, J.; MILLASSEU, P.; VAYSSEIX, G.; LATHROP, M. A second-generation linkage map of the human genome. **Nature**, v. 359, p. 794–801, 1992.

YOUNG, A.; BOYLE, T.; BROWN, T. The Population Genetic Consequences of Habitat Fragmentation for Plants. **Trends in Ecology and Evolution**, 11, 413-418. [http://dx.doi.org/10.1016/0169-5347\(96\)10045-8](http://dx.doi.org/10.1016/0169-5347(96)10045-8). 1996.

ZALAPA, J. E.; CUEVAS, H.; ZHU, H., STEFFAN, S.; SENALIK, D.; ZELDIN, E; MCCOWN, B.; HARBUT, R.; SIMON, P. Using nextgeneration sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. **American journal of botany**, v. 99, n. 2, p. 193-208, 2012.

ZANE, L. BARGELLONI, L., PATARNELLO, T., Strategies for microsatellite isolation: a review. **Molecular Ecology**, v. 11, n.1, p. 1–16, 2002.

ZUCCHI, M. I.; BRONDANI, R. P. V.; PINHEIRO, J. B.; CHAVES, L. J.; COELHO, A. S. G.; VENCOVSKY, R. Genetic structures and gene flow in *Eugenia dysenterica* DC in the Brazilian Cerrado utilizing SSR markers. **Genetics and Molecular Biology**, v. 26, p. 449-

**CAPÍTULO 1. Structural, evolutionary and phylogenomic features of the plastid genome of *Carya illinoensis* cv. Imperial**

Artigo publicado na Revista **Annals of Forest Research (Ann. For. Res. 63(1): 3-18, 2020 DOI: 10.15287/afr.2019.1413)** sob autoria de Jordana Caroline Nagel, Lilian de Oliveira Machado, Rafael Plá Matielo Lemos, Cristiane Barbosa D'Oliveira Matielo, Tales Poletto, Igor Poletto, Valdir Marcos Stefenon.

# Structural, evolutionary and phylogenomic features of the plastid genome of *Carya illinoensis* cv. Imperial

Jordana Caroline Nagel<sup>1,2</sup>, Lilian de Oliveira Machado<sup>3</sup>, Rafael Plá Matielo Lemos<sup>1</sup>, Cristiane Barbosa D'Oliveira Matielo<sup>1</sup>, Tales Poletto<sup>4</sup>, Igor Poletto<sup>1</sup>, Valdir Marcos Stefenon<sup>3,1</sup>

**Nagel J.C., de Oliveira Machado, L., Lemos R.P.M., Barbosa D'Oliveira Matielo C., Poletto T., Poletto I., Stefenon V.M.,** 2020. Structural, evolutionary and phylogenomic features of the plastid genome of *Carya illinoensis* cv. Imperial. Ann. For. Res. 63(1): 3-18.

**Abstract.** The economically most important nut tree species in the world belong to family Juglandaceae, tribe Jungladeae. Evolutionary investigations concerning species from this tribe are important for understanding the molecular basis driving the evolution and systematics of these species. In this study, we release the complete plastid genome of *C. illinoensis* cv. Imperial. Using an IonTorrent NGS platform we generated 8.5 x 10<sup>8</sup> bp of raw sequences, enabling the assemblage of the complete plastid genome of this species. The plastid genome is 160,818 bp long, having a quadripartite structure with an LSC of 90,041bp, an SSC of 18,791 bp and twoIRs of 25,993 bp. A total of 78 protein-coding, 37 tRNA-coding, and 8 rRNA-coding regions were predicted. Bias in synonymous codon usage was detected in cultivar Imperial and three tRNA-coding regions were identified as hotspots of nucleotide divergence, with high estimations of dN/dS ratio. The high fraction of SSR loci prospected in non-coding regions may provide informative genetic markers, useful to a wide range of genetic researches. Despite the significant structural differences among plastid genomes, the phylogenetic relationships among species is supported by the whole plastid genome analysis, supporting the monophyly of subtribes Caryinae and Juglandinae within family Juglandaceae.

**Keywords:** Pecan, chloroplast genome, phylogenomics, Juglandaceae

**Authors.** <sup>1</sup>Federal University of the Pampa, Graduate Program in Biological Sciences, Campus São Gabriel, São Gabriel, RS, Brazil | <sup>2</sup>Universidade Regional Integrada do Alto Uruguai e das Missões, Campus Santo Ângelo, Santo Ângelo, RS, Brazil | <sup>3</sup>Federal University of Santa Catarina, Graduate Program in Plant Genetic Resources, Florianópolis, SC, Brazil | <sup>4</sup>Federal University of Santa Maria, Graduate Program in Forest Engineering, Santa Maria, RS, Brazil.

<sup>5</sup> Corresponding author: Valdir Marcos Stefenon (valdirstefenon@gmail.com)

**Manuscript** received November 3, 2019; revised February 10, 2020; accepted February 18, 2020; online first March 2<sup>nd</sup>, 2020.

## Introduction

In most plant species, the plastid DNA shows maternal inheritance, low recombination frequency, and a relaxed evolution rate. These characteristics make plastid genomes substantially distinctive from the nuclear genomes (Wolfe et al. 1987) and very useful in a wide range of studies about evolutionary relationships in plants. The principal function of the plastid is to carry out photosynthesis, but other major cellular functions also occur in this organelle, including synthesis of starch, fatty acids, pigments and amino acids (Wicke et al. 2011).

The known plastid genomes of green plants typically contain from 100 to 120 genes, of which approximately 70-88 are protein-coding genes, 33-35 are tRNAs and four are rRNAs (Wicke et al. 2011). With the progress of the next generation sequencing (NGS) technologies, complete plastid genomes have been sequenced for several plant species, generating a wide set of genomic resources, which enable the development of molecular markers and more refined taxonomic and evolutionary studies. The development of genetic markers from plastid genome sequences has significantly contributed to studies about gene flow (plastid SSR markers; Perdereau et al. 2014), phylogeography (SNPs and Indels in plastid genes; Stefenon et al. 2019a), and hybridization/introgression (plastid SSR markers; Curtu et al. 2007) in forest tree species. Moreover, identification of genome rearrangements becomes increasingly important, especially as high levels of rearrangement have been observed among both eukaryotes and prokaryotes (Wicke et al. 2011). Such rearrangements may also be very useful to understand phylogenetic and evolutionary trends within and among plant groups. Knowledge about genome rearrangements, gene content, recombination events, loss of genes, and gene transfer to the nucleus are of great importance for understanding evolutionary events in plants (Vieira et al. 2016, Bock

2017, Lopes et al. 2017).

*Carya illinoensis* (Wangenh) K. Koch, popularly known as pecan (Figure 1a-c, Supporting Information), is a deciduous tree species of the tribe Juglandae (Juglandaceae family), native to the temperate zones of North America. Several species of tribe Juglandae - as *Juglans regia* (walnut), *Cyclocarya paliurus* (wheel wingnut), *Carya sinensis* (Chinese hickory) and *Carya illinoensis* (hickory or pecan) - are important crop tree species cultivated in several regions of the world aiming at the commercial production of nuts. Pecan is one of the most important nut crop species in the world, cultivated in several countries in North America (USA and Mexico), South America (Peru, Argentina, and Brazil), Africa (South Africa and Egypt), Asia (Israel and China) and Australia (Poletto et al. 2018). The species was introduced in Brazil with commercial interest during the 1870s and farmers needed to select cultivars adapted to the climatic and biological conditions where the orchards were established (Poletto et al. 2015). Currently, more than 40 cultivars are commercially planted in large areas of Southern Brazil. Despite the economic significance of *Carya* as a crop tree species and the importance of plastid genomes for comparative evolutionary analysis and molecular taxonomy, only the plastid genomes of *Carya sinensis* (Hu et al. 2016), *Carya kweichowensis* (Yeh et al. 2018), and *Carya cathayensis* (Zhai et al. 2019) were published to date, while one unpublished sequence of the plastid genome of *Carya illinoensis* is deposited in the NCBI database (Genbank ID MH909599.1). Besides, few molecular studies have been performed concerning *C. illinoensis* cultivars planted in southern Brazil (e.g Poletto et al. 2019). Considering the scarcity of genomic studies available for *Carya* species and aiming to generate novel genomic resources for *Carya illinoensis*, we sequenced, assembled and characterized the complete plastid genome of *C. illinoensis* cv. Imperial using next-generation sequencing

(NGS) technology. Here, we report the main finds obtained from this initiative concerning the genes present and the codon usage bias in the plastid genome of this species. In addition, we identified potentially polymorphic plastid SSR markers and also performed a comparative evolutionary analysis of plastid genomes between *C. illinoensis* cv. Imperial and other tree species of the tribe Juglandae using publicly available genomic datasets, thus revealing some features of pecan evolution.

## Materials and methods

### Plastid genome sequencing and assembling

Healthy leaves of an adult individual of *C. illinoensis* cv. Imperial were sampled in the municipality of Anta Gorda, Rio Grande do Sul State, southern Brazil. A voucher of this sample was deposited under the number HBEI1624 in the Bruno Edgar Irgang Herbarium of the Federal University of the Pampa, Brazil. Intact chloroplasts were isolated from the leaves as described by Matielo et al. (2019) and the plastid DNA (cpDNA) was isolated using the CTAB method (Doyle and Doyle 1987). The quality of the isolated DNA was checked using a NanoVue™ spectrophotometer (GE Healthcare). The isolated cpDNA was used for library preparation with Ion OneTouch2™ System using the Ion PGM™ Template OT2 400 Kit. The sequencing was performed using Ion PGM™ Sequencing 400 kit on the Ion PGM™ System with an Ion 318™ Chip v2. Raw sequence data were deposited in the NCBI Sequence Read Archive (SRA) database under number SSR10382885, Bioproject PRJ-NA587009, Biosample SAMN13174479.

The plastid genome of *C. illinoensis* cv. Imperial was assembled using a reference-guided approach with the plastid genome of *C. illinoensis* (NC041449.1) as a reference, in the CLC Genomics Workbench software. The mean coverage of the sequencing was deter-

mined as the total size of the sequenced reads divided by the size of the assembled plastid genome. Annotation of the plastid genome was conducted using the GeSeq (Tillich et al 2017) and the cpGAVAS (Liu et al. 2012) platforms. For GeSeq, annotation started from four references chloroplast annotations (*Arabidopsis thaliana*, *Castanea mollissima*, *Juglans nigra* and *J. regia*). The software tRNAscan (Chan and Lowe 2019) and Aragorn (Laslett and Camback 2004) were used for searching the tRNAs, while the physical circular map of the plastid genome was built using Organellar Genome DRAW software (Lohse et al. 2013).

### Characterization of plastid genome features in *C. illinoensis* cv. Imperial

Relative synonymous codon usage (RSCU) of all protein-coding genes were determined using MEGA 6.0 software (Tamura et al. 2013). RSCU corresponds to the proportion of the observed occurrence of a codon to its expected occurrence if all the synonymous codons of a particular amino acid are used evenly. Prospection of simple sequence repeats (SSRs) loci in the plastid genome of *C. illinoensis* cv. Imperial was performed using the Perl script MISA (Beier et al. 2017) setting minimum thresholds for search at ten for mononucleotide repeats, six for dinucleotide repeats and five for tri-, tetra-, penta- and hexanucleotide repeats. The location of the SSR loci within the plastid genome was determined using the cpGAVAS (Liu et al. 2012) platform.

### Evolutionary and phylogenetic relationships of *C. illinoensis* cv. Imperial within Juglandae

Different comparisons were performed among *Carya illinoensis* cv. Imperial and 11 species of Juglandae tribe (*Carya illinoensis* (Wangenh) K.Koch, *Carya kweichowensis* Kuang & A.M. Lu, *Carya sinensis* Dode, *Cyclocarya paliurus* (Batal.) Iljinsk., *Juglans cinerea* L., *Juglans regia* L., *Juglans major* (Torr.)

A. Heller, *Juglans sigillata* Dode, *Juglans hopeiensis* Hu, *Juglans mandshurica* Maxim., and *Juglans cathayensis* Dode), whose plastid genome sequences were downloaded from the Genbank database. Species nomenclature in this study follows the International Plant Names Index (www.ipni.org). An initial pairwise comparison of the gene order of plastid genomes was performed through a dot plot analysis using the software Mafft online service (Kato et al. 2017). Translocations, inversions or indels occurring in a set of genes are visualized in the pairwise comparison as displacements of the positive and/or negative slopes representing the LSC, SSC (positive slope) and IR (negative slope) regions of the plastid genomes compared. Boundaries of the IRa, IRb, SSC and LSC regions (IR/SSC and IR/LSC boundaries) and sizes of each region were determined using the online platform IRscope (Amiryousefi et al. 2018). Aiming to compare the number and type of SSR loci in all species, these ubiquitous regions were prospected in the plastid genome sequences of all other 11 species, as described above for *C. illinoensis* cv. Imperial.

A phylogenomic analysis was performed using the complete plastid genome sequences of *C. illinoensis* cv. Imperial and more 11 species of tribe Juglandae. The plastid genome sequence of *Castanea mollissima* Blume (Fagaceae) was used as outgroup. The Genbank IDs of all downloaded sequences are given in Figure 3c. The 13 sequences were aligned using the software Mafft online service (Kato et al. 2017) and the phylogenetic tree was constructed with the Neighbor-Joining algorithm. Support of the analysis was determined through 500 bootstrap replicates using the same software.

Rearrangements and inversions among Juglandae plastid genomes were visualized using the default parameters of the Multiple Genome Alignment software MAUVE 2.4.0 (Darling 2004). Hotspots of sequence divergence were determined using the sliding windows analysis with the complete plastid genomes of 11 spe-

cies of tribe Juglandae with a window length of 400 bp and step size of 100 bp, using the DnaSP v.5 software (Librado and Rozas 2009). *Carya kweichowensis* was excluded from this analysis giving the significant difference in length of the plastid genome in comparison to all other species.

The ratio of the nonsynonymous ( $dN$ ) to synonymous ( $dS$ ) substitutions -  $dN/dS$  ratio - of the genomic regions identified as hotspots of divergence was performed to investigate the patterns of selection occurring within Juglandae. The complete regions containing CDSs, introns and intergenic spacers corresponding to each hotspot were extracted from all species analyzed in this study and aligned individually using the MUSCLE algorithm (Edgar 2004) as implemented in MEGA 6.0 (Tamura et al. 2013), with pairwise deletion set to gaps/missing data treatment. To each alignment, the  $dS$  and  $dN$  values were calculated using MEGA under the Kimura 2-parameters model. The pairwise and the overall mean distances were estimated for each hotspot region using the same software.

## Results

### General features of the plastid genome of *C. illinoensis* cv. Imperial

The sequencing effort generated a total of 5,639,849 raw reads, representing 852,882,019 assembled nucleotides, which corresponds to a mean plastid genome coverage of 5,303.4×

The plastid genome of *Carya illinoensis* cv. Imperial (GenBank ID MN221384) presents 160,818 bp in length and the traditional quadripartite structure of plastid genomes (Figure 1d, Supporting Information). The large single-copy (LSC) region is composed of 90,041 bp with a GC content of 33.73%, and the small single-copy (SSC) region contains 18,791 bp with a GC content of 29.89%. The LSC and the SSC regions are separated by two inverted repeat regions (IRs) of 25,993 bp each, with

42.58% of GC content. The overall GC content of this plastid genome was 36.14%. A total of 123 genes, including 78 protein-coding genes, 37 tRNA-coding genes, and eight rRNA-coding genes were predicted.

A total of seven protein-coding genes, seven tRNA-coding genes, and four rRNA-coding genes are located within each IR region of the chloroplast genome of *C. illinoensis* cv. Imperial (Table 1). The SSC region presents 12 protein-coding genes and one tRNA-coding gene, while the LSC region presents 61 protein-coding genes and 22 tRNA-coding regions (Table 1). Eighteen genes presented introns, while the *rps12* gene is trans-spliced, with the 5'-end located in the LSC and the 3'-end duplicated in the IR regions. The tRNA-coding gene *trnE-UUC* has two copies and the *trnM-CAU* has three copies.

### Comparative analysis among chloroplast genomes of Juglandaceae species

Cultivar Imperial presents a plastid genome-

with one base deletion and 21 single nucleotide polymorphisms (SNPs) in comparison to the plastid genome of *Carya illinoensis* MH909599.1 deposited in the Genbank database. Concerning the other species of the family Juglandaceae, the size of the plastid genome of cultivar imperial is also very similar (Figure 1a-b), except to *Carya kweichowensis*. This species has a plastid genome 14,495 bases longer than the plastid genome of *C. illinoensis* cv. Imperial (Table 2). The IR regions of *C. illinoensis* cv. Imperial were only 65 to 34 bp shorter than the IR regions of *Carya sinensis* and *Cyclocarya* respectively.

Concerning the plastid genomes of *Juglans* species included in this study, *C. illinoensis* cv. Imperial revealed a difference in the length of the IR regions ranging from 31 bp to 227 bp (Table 2, Figure 1b). Excluding *C. kweichowensis*, the coefficient of variation of the plastid genome size (total plastid genome, LSC, SSC, and IRs) ranged from 0.25% to 0.88% (Table 2). On the other hand, each IR region of *C. kweichowensis* is 40,943 bp in

**Table 1** Distribution of the coding regions within the plastid genome of *C. illinoensis* cv. Imperial, according to the plastid region

Region	Category	Genes
IRs <sup>1</sup>	Protein-coding	<i>ndhB*</i> , <i>rpl2*</i> , <i>rpl23</i> , <i>rps12</i> <sup>3*</sup> , <i>rps7</i> , <i>ycf1</i> <sup>2</sup> , and <i>ycf2</i>
	tRNA-coding	<i>trnA-UGC*</i> , <i>trnM-CAU</i> , <i>trnI-GAU*</i> , <i>trnL-CAA</i> , <i>trnE-UUC*</i> , <i>trnR-ACG</i> and <i>trnV-GAC</i>
	rRNA-coding	<i>rrn16</i> , <i>rrn23</i> , <i>rrn4.5</i> and <i>rrn5</i>
SSC	Protein-coding	<i>ndhF</i> , <i>rps32</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhG</i> , <i>ndhI</i> , <i>psaC</i> , <i>ccsA</i> , <i>ndhA*</i> , <i>ndhH</i> , <i>rps15</i> and <i>ycf1</i> <sup>2</sup>
	tRNA-coding	<i>trnL-UAG</i>
LSC	Protein-coding	<i>rps19</i> , <i>rpl22</i> , <i>rps3</i> , <i>rpl16</i> , <i>rpl14</i> , <i>rpsB</i> , <i>infA</i> , <i>rpl36</i> , <i>rps11</i> , <i>rpoA</i> , <i>petA</i> , <i>petD</i> , <i>petB</i> , <i>petG</i> , <i>petL</i> , <i>pbfl</i> , <i>psbH</i> , <i>psbT</i> , <i>psbB</i> , <i>psbE</i> , <i>psbJ</i> , <i>psbL</i> , <i>clpP*</i> , <i>rps12</i> <sup>3*</sup> , <i>rpl20</i> , <i>rps18</i> , <i>rpl33</i> , <i>psaJ</i> , <i>psaI</i> , <i>cemA</i> , <i>ycf4</i> , <i>accD</i> , <i>rbcl</i> , <i>atpB</i> , <i>atpE</i> , <i>ndhC</i> , <i>ndhK</i> , <i>ndhJ</i> , <i>rps4</i> , <i>ycf3</i> <sup>3*</sup> , <i>psaA</i> , <i>psaB</i> , <i>rps14</i> , <i>psbZ</i> , <i>psbC</i> , <i>psbD</i> , <i>psbM</i> , <i>petN</i> , <i>rpoB</i> , <i>rpoC1</i> <sup>*</sup> , <i>rpoC2</i> , <i>rps2</i> , <i>atpl</i> , <i>atpH</i> , <i>atpF*</i> , <i>atpA</i> , <i>psbI</i> , <i>psbK</i> , <i>rps16*</i> , <i>matK</i> , <i>psbA</i>
	tRNA-coding	<i>trnP-UGG</i> , <i>trnW-CCA</i> , <i>trnM-CAU</i> <sup>4</sup> , <i>trnV-UAC*</i> , <i>trnF-GAA</i> , <i>trnL-UAA*</i> , <i>trnS-GGA</i> , <i>trnT-UGU</i> , <i>trnG-GCC</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnE-UUC*</i> , <i>trnY-GUA</i> , <i>trnD-GUC</i> , <i>trnC-GCA*</i> , <i>trnR-UCU</i> , <i>trnS-CGA</i> , <i>trnS-GCU</i> , <i>trnQ-UUG</i> , <i>trnK-UUU*</i> , <i>trnH-GUG</i>

Note. <sup>1</sup>All coding regions within the IR regions are duplicated, <sup>2</sup>A large part of the *ycf1* gene is located within the SSC, <sup>3</sup>Trans-spliced gene, <sup>4</sup>This gene has two copies within the LSC region, \*Coding regions containing introns.

**Table 2** Size of plastid genomes of Juglandaceae species in base pairs (bp)

Species	Total length	LSC	SSC	IR
<i>Carya illinoensis</i> cv. Imperial	160,818	90,041	18,791	25,993
<i>Carya illinoensis</i> MH909599.1	160,819	90,042	18,791	25,993
<i>Carya kweichowensis</i>	175,313	89,858	3,569	40,943
<i>Carya sinensis</i>	160,195	89,541	18,538	26,058
<i>Cyclocarya paliurus</i>	160,562	90,007	18,477	26,039
<i>Juglans cathayensis</i>	159,730	89,333	18,351	26,023
<i>Juglans mandshurica</i>	159,729	89,845	18,352	25,766
<i>Juglans hopeiensis</i>	159,714	89,316	18,352	26,023
<i>Juglans sigillata</i>	160,351	89,871	18,412	26,034
<i>Juglans regia</i>	160,367	89,872	18,423	26,036
<i>Juglans major</i>	160,276	89,829	18,397	26,025
<i>Juglans cinerea</i>	160,288	89,803	18,417	26,034
Coefficient of variation (%)*	0.25	0.29	0.88	0.31

Note. \*The coefficient of variation was computed excluding *Carya kweichowensis*.

length, that is, 1.6-fold larger than each IR region of *C. illinoensis* (Figure 2b), while the SSC region of *C. kweichowensis* is only 3,569 bp in length, i.e., about 6-fold shorter than the SSC of *C. illinoensis*.

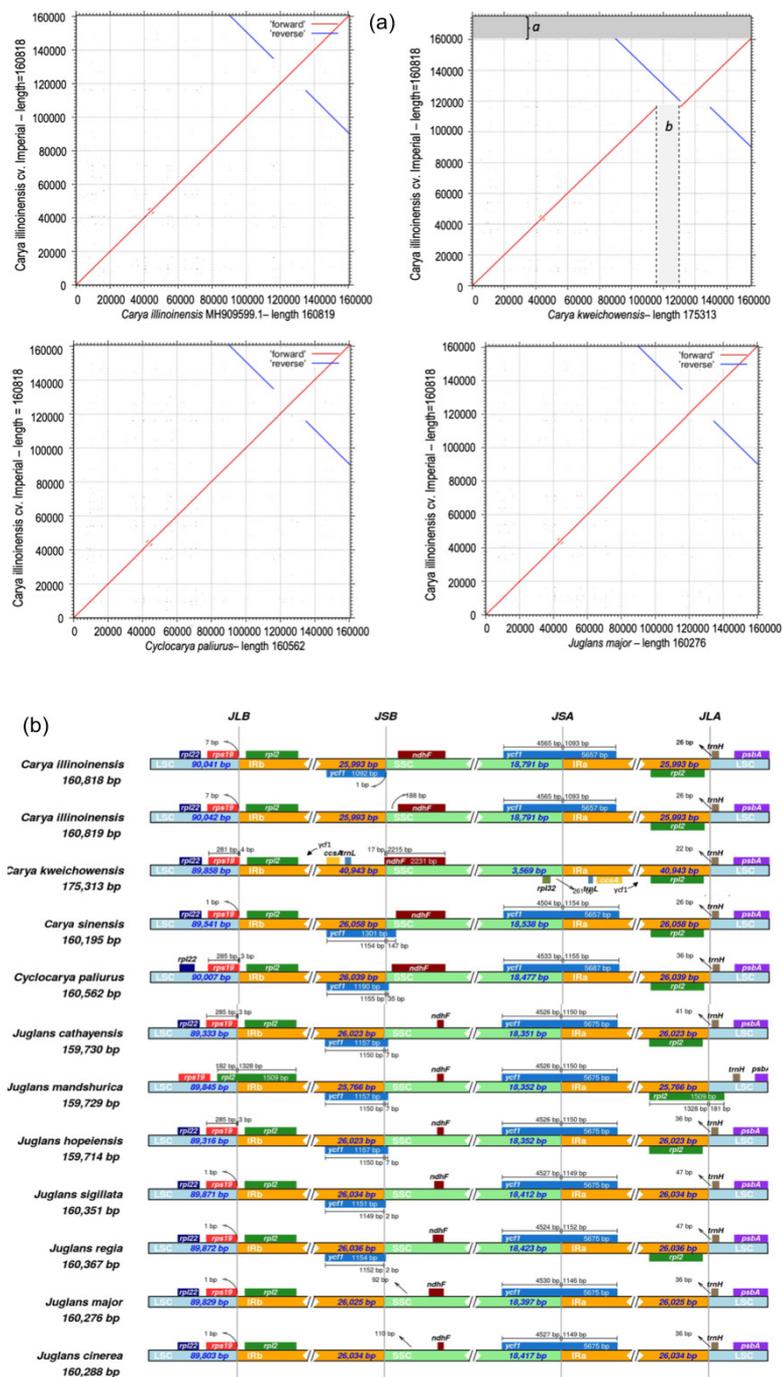
The IRa/LSC border is located between genes *rpl2* (in the IRa region) and *trnH* (in the LSC region) in all species (Figure 2b). In *J. mandshurica*, 181 bp of the *rpl2* gene are positioned within the LSC region. The IRa/SSC border is filled by the *ycf1* gene in all species except *Carya kweichowensis*, in which this gene is located internally in the IR regions, distant to the borders with the SSC and LSC regions. The genes *rps19* in the LSC region and *rpl2* in the IRb region define the IRb/LSC border in all species. In *J. mandshurica*, 82 bp of the *rpl2* gene are inside the LSC region. A comparatively long *ndhF* gene is found in the border IRb/SSC in *Carya* and *Cyclocarya*. This gene is much shorter in *Juglans*. A segment of the *ycf1* gene is found in the IRb/SSC border in *C. illinoensis*, *C. sinensis*, *Cyclocarya*, *J. cathayensis*, *J. mandshurica*, *J. sinensis*, *J. paliurus*, *J. hopeiensis*, *J. sigillata* and *J. regia* (Figure 2b).

Detailed comparisons of the IR/SSC and IR/LSC junction sites among species of Juglandaceae (Figure 2b) show little variation. Variations in IR boundaries were mainly observed in *C. kweichowensis*, and some substantial differ-

ences in the size of the *ndhF* gene in species of *Juglans* in comparison to *C. illinoensis*, *C. sinensis*, and *Cyclocarya*.

The synonymous codons in angiosperms genomes retain different usage frequencies, i.e., codon usage biases. In *C. illinoensis* cv. Imperial, codon usage bias was revealed, with a high proportion of synonymous codons presenting the nucleotides A or U in the third position (Figure 2a). This pattern is consistent with most of the plastid genomes and may be linked firstly to the high A/T content of the plastid genome (A/T = 63.86%), although selective pressure cannot be discarded (see results of the sliding window analysis below). Codons CUG(L), CAC(H), GAC(D), CGC(R) and GGC(G) can be considered under-represented (RSCU < 0.6), while codons UUA(L) and AGA(R) are over-represented (RSCU > 1.6, Barbhuiya et al. 2019).

Twenty-nine codons presented RSCU values higher than 1.0, meaning they are being used more often than the expected. On the other hand, 33 codons are being used less frequently than expected, presenting a RSCU < 1.0. As expected, codons AUG(M) and UGG(W), the unique codons for methionine and tryptophan respectively, presented an RSCU = 1.0 (Figure 2a). The relative synonymous codon usage (RSCU) of the plastid genome of *C. illinoensis* cv. Imperial revealed a higher proportion



**Figure 1** Relationship among plastid genomes of *Carya illinoensis* cv. Imperial and species of Juglandae. (a) Dot plot analyses of the plastid genome of *C. illinoensis* cv. Imperial against three Juglandaceae species. The positive red slope means that the pair of sequences compared is in the same orientation. The negative blue slope denotes that the pair of sequences compared can be aligned, but their orientation is opposite. In the dot plot against *C. kweichowensis*, the gray region “a” highlights the 14,495 bp without correspondent sequence in *C. illinoensis* cv. Imperial (as well as in the other species of Juglandaceae). In the same dot plot, region “b” highlights a gap, which means that no similarity was identified between this region, present in the plastid genome of *C. kweichowensis*, and the plastid of *C. illinoensis* cv. Imperial. (b) Boundaries at the junctions of the LSC, IRa, IRb, and SSC regions of the plastid genomes of *Carya*, *Cyclocarya*, and *Juglans* species.



This pattern is coherent with the large size of the *Carya kweichowensis* plastid genome because besides to expansion and contraction of the IR and SSC boundary regions, repetitive genomic elements are also believed to enlarge genome sizes (Stefenon et al. 2019b). With the prospection parameters employed, no C/G SSR was found in the plastid genomes of *C. illinoensis* and *Cyclocarya paliurus*.

### Phylogenetic relationships and evolution of juglandeae based on the whole plastid genome analysis

The phylogenomic analysis based on the complete plastid sequences of 12 species of the tribe Juglandeae using *Castanea mollissima* (Fagaceae) as outgroup (Figure 3c) supports the phylogenetic studies based on morphological, nuclear and plastid genes. High bootstrap calculations (BP = 100%) support all clades.

The phylogenomic tree (Figure 3c) clustered *Carya illinoensis* cv. Imperial with *C. illinoensis* MH909599.1 with 100% bootstrap support. *Carya kweichowensis* clustered with *Carya sinensis* (BP = 100%) as a sister group of *C. illinoensis* clade (BP = 100%). These species comprise the subtribe Caryinae within tribe Juglandeae. All seven species of *Juglans* formed a monophyletic clade sister to *Cyclocarya paliurus* (BP = 100%). *Juglans* and *Cyclocarya* comprise subtribe Juglandinae.

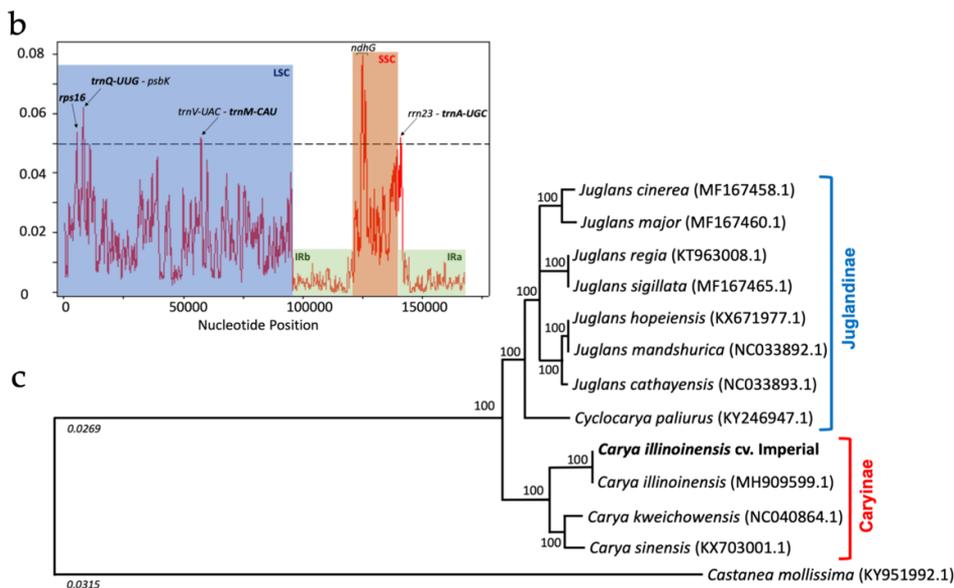
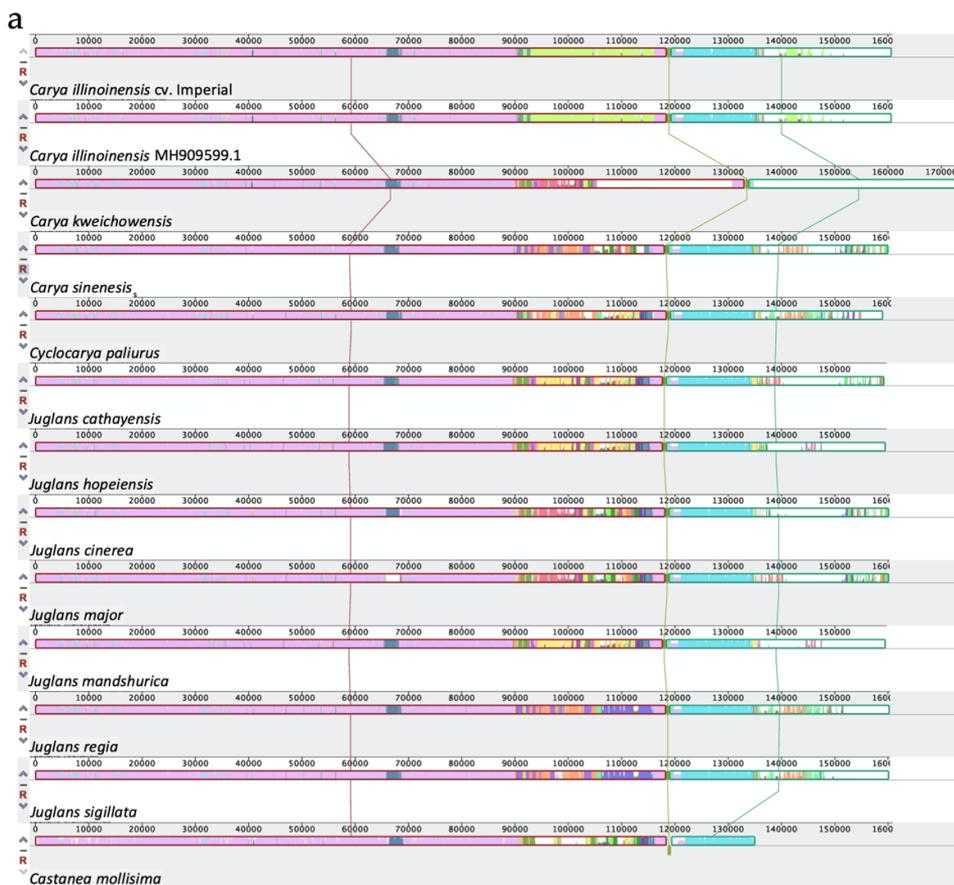
Genes rearrangements were observed among the evaluated species of Juglandeae

and *Castanea mollissima* (Figure 3a). Forty-two homologous DNA regions are not collinear between *C. illinoensis* cv. Imperial and *C. illinoensis* MH909599.1 (data not shown). *Carya sinensis* presents 195 and *Carya kweichowensis* presents 236 non-collinear homologous DNA regions in comparison to *C. illinoensis*, respectively. *Cyclocarya* presented relatively larger non-collinear homologous DNA regions in comparison to the species of *Carya* (Figure 3a).

The non-collinearity of homologous DNA among *Juglans* species is also restricted to minor regions, with a high similarity between *J. regia* and *J. sigillata* and between *J. cinerea* and *J. major* (data not shown), supporting the phylogenetic clustering of these species (Figure 3c). As expected, the larger difference concerning the Local Collinear Blocks (LCBs) is observed with species of Juglandaceae in comparison to *Castanea mollissima* (Figure 3a).

Based on the sliding window analysis, five hotspots of plastid nucleotide divergence were identified in this study (Figure 3b). Three of these hotspots are located within the LSC region (the *rps16* gene, the *trnQ-UUG* - *psbK* region, and the *trnV-UAC* - *trnM-CAU* region), one in the SSC region (the *ndhG* gene) and one in the IRb region (the *rrn23* - *trnA-UGC* region). In each of the three hotspot sites composed of two coding-regions, the largest portion of the hotspot sequence was related to one tRNA gene (*trnQ-UUG*, *trnM-CAU*, and *trnA-UGC*, Figure 3b). According to the codon

**Figure 3** Evolutionary relationships of species from tribe Juglandeae. (a) Multiple plastid genome alignment with species from tribe Juglandeae and *Castanea mollissima* (Fagaceae). Different colors represent different Local Collinear Blocks (LCBs), indicating sequence rearrangements and/or translocation. (b) Sliding window analysis of aligned whole plastid genomes of the 12 species of tribe Juglandeae. The regions with high nucleotide variability ( $\pi > 0.050$ ) and approximated limits of the SSC, LSC, IRa, and IRb regions are indicated. (c) Phylogenomic relationship among species of Juglandeae, based on the NJ algorithm and using *Castanea mollissima* (Fagaceae) as outgroup. Number at nodes are the bootstrap support after 500 replicates. Numbers below branches are their respective length. Plastid genome sequences of all species were downloaded from Genbank (ID number given after each species name), except for *C. illinoensis* cv. Imperial, which plastid genome sequence was generated in this study.



usage analysis (Figure 2a), codons UUG(L) and CAU(H) are used more often than expected, while the UGC(C) is used less often than expected without bias in codon usage. This suggests that the bias in codon usage can be an effect of evolutive selection, reflected in the presence of tRNA genes in the hotspots of plastid nucleotide divergence.

The  $dN/dS$  ratio of the five hotspots of diversity revealed values compatible with a pattern of positive selection ( $dN/dS > 1$ ). The mean values of  $dN/dS$  ratio overall codon were (i)  $rps16 = 3.9$ , (ii)  $trnQ-UUG - psbK = 4.25$ , (iii)  $trnV-UAC - trnM-CAU = 4.25$ , (iv)  $ndhG = 4.25$ , and (v)  $rrn23 - trnA-UGC = 4.25$ . The species pairwise distance and the mean overall distance ( $d$ ) based on the mean number of substitutions per site (Tables S1-S5 – Supporting information) and the maximum parsimony tree based on the matrix of species pairwise distances (Figure S2– Supporting information) revealed different patterns for each hotspot. The highest overall distance among populations ( $d = 0.514$ ) was retrieved in the  $trnQ-UUG - psbK$  region, followed by the  $rrn23 - trnA-UGC$  region ( $d = 0.418$ ). The mean overall distance for the other hotspot regions of nucleotide diversity were lower than 0.009. The clustering of the species in the maximum parsimony trees was different for each region and presented low bootstrap support for most of the branches, except for the  $trnQ-UUG - psbK$  region which presented bootstrap values higher than 62%.

## Discussion

*Carya* species are adapted to a broad range of climate types in North America, resulting in the existence of numerous cultivars and hybrid lines of *C. illinoensis* and worldwide commercial cultivation (Huang et al. 2019). Thus, understanding the genomic structure and evolutive history of this species is crucial for the conservation and exploitation of the species' genetic resources. In reporting the complete

plastid genome sequence of *Carya illinoensis* cv. Imperial and comparing evolutive and structural parameters of the plastid genomes, we add to the scientific literature, novel genomic resources and evolutionary insights about this species.

The general structure, the number, and the category of genes in the plastid genome of *Carya illinoensis* cv. Imperial (GenBank ID MN221384) is comparable to most angiosperms (Hu et al. 2016, Dong et al. 2017, Hu et al. 2017, Zhai et al. 2019) and very similar to most species of Juglandaceae, with small size differences ranging from 1 to 1089 bp (Table 2). An expressive exception is the plastid genome of *C. kweichowensis*, which underwent a very large expansion of the IRs resulting in a meaningful increase in its genome size and a severe reduction in the SSC region (Table 2). It seems that this Chinese endemic species experienced a particular event of IR expansion/SSC contraction, not shared by other species of the tribe Juglandaceae, resulting in quite evident differences in the size of the IRa, IRb, and SSC regions. Expansion and contraction of the IR regions and the LSC/IR/SSC boundaries have been recognized as the main cause of the size variation of angiosperm plastid genomes (Dugas et al. 2015). Different from the IR expansion observed in *C. kweichowensis*, the IR/SSC and IR/LSC boundaries are highly conserved among all species of Juglandaceae included in this study, with only some subtle expansion/contraction events of the IRs (Figure 1b). Sequence variabilities, especially in the *yef1* gene at IR-SSC junction and in the *rps19* and *rpl22* genes at IR-LSC junction, are frequently observed as a result of expansion and contractions events by gene conversion (Zhu et al. 2016, Lopes et al. 2018).

Although numerous factors such as mutational pressure, translational selection, compositional constraints, and gene length may cause bias in codon usage patterns, we suggest that the biased usage towards codons with bases A and U in the third position (26 out of 28 preferentially used codons) observed

in *C. illinoensis* cv. Imperial is an effect of the AT-rich composition of the species' plastid genome. A similar pattern of codon usage bias and amino acid frequency has been reported in plastid genomes of other species as *Angelica polymorpha* (Apiaceae) (Park et al. 2019). However, the existence of three tRNA-coding regions with biased usage and that were also identified as hotspots of nucleotide divergence suggest the occurrence of evolutive selection over these codons. High estimations of  $dN/dS$  ratio suggest these hotspots of divergence are experiencing positive selection in Juglandaceae.

SSR loci are repetitive elements found in nuclear and organellar genomes and are widely employed in genetic studies (Lemos et al. 2018). The plastid SSR loci identified in this study have potential usefulness as molecular markers for studies of species and even cultivars differentiation. The difference in number of SSR loci among species (Figure 2c) means that some of them are species-specific and can be characterized as markers for species differentiation. Moreover, some SSR loci revealed polymorphism when *C. illinoensis* cv. Imperial and *C. illinoensis* MH909599.1 are compared (Table 3, Supporting Information) and can be used for the development of haplotypic genetic barcodes for cultivars identification. The high fraction of SSR loci prospected in non-coding regions (intergenic regions and introns) of *C. illinoensis* cv. Imperial may provide informative genetic markers and their deep characterization *in silico* and the laboratory is an ongoing project of our research group. Plastid markers are uniparentally inherited and lack allele recombination being, therefore, convenient to a wide range of genetic researches including cultivars identification, population genetics, gene flow, hybridization/introgression, and characterization of species evolutionary history.

Approximately 83% of the identified SSR loci are located within introns and intergenic regions. SSR loci located within intergenic regions and introns evolve faster than CDSs

(Rogalski et al. 2015) and are, therefore, more useful as molecular markers. Thus, these SSR loci prospected in the plastid genome of *C. illinoensis* cv. Imperial have the potential for the development of markers for genetic studies at population, species and genus levels. Species-specific SSR markers have generated an elevated number of polymorphic loci than those employing universal primers (Wheeler et al. 2014). The difference in the number of SSR loci observed among species in this study highlights the possibility of selecting species-specific markers for each investigated species.

Moreover, we show that despite the significant structural differences among plastid genomes, as LCBs rearrangements and differences in the IR/SSC and IR/LSC boundaries, the phylogenetic relationship among species is supported when complete plastid genome sequences are analyzed as a single molecule. Overall, the species structure retrieved in our phylogenomic tree resembles the phylogenetic relationships of Juglandaceae species based on plastid genes (Stanford et al. 2000; Manos et al. 2007), nuclear ITS (Stanford et al. 2000), and morphological data (Manos et al. 2007). Moreover, our phylogenomic analysis supports the monophyly of subtribes Caryinae and Juglandinae within family Juglandaceae and the monophyly of the genus *Juglans* is corroborated.

Single nucleotide variants, indels, and large structural variants were reported for plastid genomes of *Juglans* species (Hu et al. 2017). In our study, the distribution of the Local Colinear Blocks (Figure 3a) reveals low difference among species (with exception of the quite larger plastid genome of *C. kweichowensis*) and represents the phylogenetic pattern of the species, with phylogenetic closer species sharing the position of the LCBs, without significant gene rearrangements.

Moreover, we investigated whether any plastid gene of the studied species underwent positive selection. The sliding window analysis identified a total of five hotspots of nu-

cleotide divergence, including protein-coding genes (*rps6* and *ndhG*) and intergenic regions (*trnQ*-UUG – *psbK*, *trnV*-UAC – *trnM*-CAU, and *rrn23* – *trnA*-UGC). Six protein-coding genes and three intergenic regions were identified as a hotspot of nucleotide diversity among species of *Juglans* (Hu et al. 2017). However, the hotspots observed among *Juglans* species are all different from the hotspots we identified in our study. Among species of tribe Brasicaceae, the *rps16* gene was also found to be a hotspot of nucleotide divergence, presenting a signature of positive selection (Lopes et al. 2018). In our study, the *dN/dS* ratio of the hotspots of nucleotide diversity also revealed a signature of positive selection for all regions. Although the effect of the positive selection on the protein function and the adaptive capacity are still poorly understood, several studies have reported signatures of positive selection on plastid genes (Lopes et al. 2018).

## Conclusion

In conclusion, the findings of this study have important implications in the areas of genetics, evolution, conservation, breeding, and biotechnology of *Carya illinoensis*. Here we reported the complete plastid genome of *C. illinoensis* cv. Imperial, a pecan variety cultivated in southern Brazil and revealed the existence of minor differences among plastid genomes of species of tribe Juglandaeae. On the other hand, the presence of signatures of positive selection in some genes and intergenic regions was also observed. Moreover, this plastid genome rises as an important genomic resource, enabling the improvement of phylogenomic analyses based on whole plastid genome sequences and also the development of plastid SSR markers with wide applications in population genetics, cultivars identification, and genetic improvement.

## Acknowledgments

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil, Grant n. 302501/2017-7) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brazil, Finance code 001) for the financial support, scholarships and grants awarded. We thank Prof. Dr. Luiz Fernando Würdig Roesch for supporting the NGS facilities.

## Data archiving statement

The whole plastid genome sequence of *Carya illinoensis* cv. Imperial is deposited in the Genbank database under number MN221384. Raw sequence data were deposited in the NCBI Sequence Read Archive (SRA) database under number SSR10382885, Bioproject PRJ-NA587009, Biosample SAMN13174479.

## References

- Amiryousefi A., Hyvönen J., Poczai P., 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030-3031. DOI: 10.1093/bioinformatics/bty220
- Beier S., Thiel T., Münch T., Scholz U., Mascher M., 2017. MISA-web: a web server for SSR prediction. *Bioinformatics* 33:2583-2585. DOI: 10.1093/bioinformatics/btx198
- Bock R., 2017. Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annu Rev Genet.* 51:1-22. DOI: 10.1146/annurev-genet-120215-035329
- Chan P.P., Lowe T.M., 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol Biol.* 1962:1-14. DOI: 10.1007/978-1-4939-9173-0\_1
- Curtu A.L., Gailing O., Finkeldey R., 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology* 7:218. DOI: 10.1186/1471-2148-7-218
- Darling, A.C.E., 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394-1403. DOI: 10.1101/gr.2289704
- Dong W., Xu C., Li W., Xie X., Lu Y., Liu Y., Jin X., Suo Z., 2017. Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA

- sequences. *Front. Plant Sci.* 8:1148. DOI: 10.3389/fpls.2017.01148
- Doyle J.J., Doyle J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull Bot Soc Am.* 19:11-15.
- Dugas D.V., Hernandez D., Koenen, et al., 2015. Mimosoid legume plastid genome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. *Sci. Rep.* 5:1-13. DOI: 10.1038/srep16958
- Edgar R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797. DOI: 10.1093/nar/gkh340
- Hu Y., Chen X., Feng X., Woeste K.E., Zhao P., 2016. Characterization of the complete chloroplast genome of the endangered species *Carya sinensis* (Juglandaceae). *Conservation Genet Resour* 8:467-470. DOI: 10.1007/s12686-016-0601-4
- Hu Y., Woeste K.E., Zhao P., 2017. Completion of the chloroplast genomes of five chinese Juglans and their contribution to chloroplast phylogeny. *Front. Plant Sci.* 7:1955. DOI: 10.3389/fpls.2016.01955
- Huang Y., Xiao L., Zhang Z. et al., 2019. The genomes of pecan and Chinese hickory provide insights into *Carya* evolution and nut nutrition. *GigaScience* 8: 1-17. DOI: 10.1093/gigascience/giz036
- Katoh K., Rozewicki J., Yamada K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* bbx108. DOI: 10.1093/bib/bbx108
- Laslett D., Canback B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11-16. DOI: 10.1093/nar/gkh152
- Lemos R.P.M., Matielo C.B.D'O., Beise D.C., Rosa V.G., Sarzi D.S., Roesch L.F.W., Stefenon V.M., 2018. Characterization of plastidial and nuclear SSR Markers for understanding invasion histories and genetic diversity of *Schinus molle* L. *Biology* 7:43. DOI: 10.3390/biology7030043
- Librado P., Rozas J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452. DOI: 10.1093/bioinformatics/btp187
- Liu C., Shi L., Zhu Y., Chen H., Zhang J., Lin X., Guan X., 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13: 715. DOI: 10.1186/1471-2164-13-715
- Lohse M., Drechsel O., Kahlau S., Bock R., 2013. Organellar Genome - DRAW - a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41: W575-W58. DOI: 10.1093/nar/gkt289
- Lopes A.S., Pacheco T.G., Santos K.G., Vieira L.N., Guerra M.P., Nodari R.O., Souza E.M., Pedrosa F.O., Rogalski M., 2017. The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep.* 37: 307-328. DOI: 10.1007/s00299-017-2231-z
- Lopes A.S., Pacheco T.G., Nimz T., Vieira L.N., Guerra M.P., Nodari R.O., Souza E.M., Pedrosa F.O., Rogalski M., 2018. The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* 247: 1011-1030. DOI: 10.1007/s00425-018-2841-x
- Manos P.S., Soltis P.S., Soltis D.E., et al., 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. *Syst. Biol.* 56: 412-430. DOI: 10.1080/10635150701408523
- Matielo C.B.D'O., Lemos R.P.M., Sarzi D.S., Machado L.O., Beise D.C., Dobbler P.C.T., Castro R.M., Fett M.S., Roesch L.F.W., Camargo F.O., Stefenon V.M., 2019. Whole plastid genome sequences of two drug-type Cannabis: insights into the use of plastid in forensic analyses. *Journal of Forensic Sciences* DOI: 10.1111/1556-4029.14155
- Park I., Yang S., Kim W.J., et al., 2019. Sequencing and comparative analysis of the chloroplast genome of *Angelica polymorpha* and the development of a novel indel marker for species identification. *Molecules* 24: 138. DOI: 10.3390/molecules24061038
- Perdereau P.C., Kelleher C.T., Douglas G.C., Hodkinson T.R., 2014. High levels of gene flow and genetic diversity in Irish populations of *Salix caprea* L. inferred from chloroplast and nuclear SSR markers. *BMC Plant Biology* 14:202. DOI: 10.1186/s12870-014-0202-x
- Poletto I., Muniz M.F.B., Poletto T., Stefenon V.M., Baggio C., Ceconi D.E., 2015. Germination and development of pecan cultivar seedlings by seed stratification. *Pesq. Agropec. Bras.* 50:1232-1235. DOI: 10.1590/S0100-204X2015001200014
- Poletto T., Stefenon V.M., Poletto I., Muniz M.F.B., 2018. Pecan propagation: Seed mass as a reliable tool for seed selection. *Horticulturae* 4: 26. DOI: 10.3390/horticulturae4030026
- Poletto T., Poletto I., Silva L.M.M., Muniz M.F.B., Reinger L.R.S., Richards N., Stefenon V.M., 2019. Morphological, chemical and genetic analysis of southern Brazilian pecan (*Carya illinoensis*) accessions. *Scientia Horticulturae*. DOI: 10.1016/j.scienta.2019.108863
- Rogalski M., Vieira L.N., Fraga H.P., Guerra M.P., 2015. Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* 6: 586. DOI: 10.3389/fpls.2015.00586
- Stanford A.M., Harden R., Parks C.R., 2000. Phylogeny and biogeography of Juglans (Juglandaceae) based on matK and its sequence data. *Am. J. Bot.* 87:872-882. DOI: 10.2307/2656895
- Stefenon V.M., Kablunde G., Lemos R.P.M., Rogalski M., Nodari R.O., 2019a. Phylogeography of plastid DNA sequences suggests post-glacial southward demograph-

- ic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. Scientific Reports 9:2752. DOI: 10.1038/s41598-019-39308-w
- Stefenon V.M., Sarzi D.S., Roesch L.F.W., 2019b. High throughput sequencing analysis of *Eugenia uniflora*: insights into repetitive DNA, gene content and potential biotechnological applications. 3 Biotech 9:200. DOI: 10.1007/s13205-019-1729-1
- Tamura K., Stecher G., Peterson D., Filipiński A., Kumar S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725-2729. DOI: 10.1093/molbev/mst197
- Tillich M., Lehwark P., Pellizzer T., 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45: W6-W11. DOI: 10.1093/nar/gkx391
- Vieira L.N., Rogalski M., Faoro H., Fraga H.P., Anjos K.G., Picchi G.F.A., Nodari R.O., Pedrosa F.O., Souza E.M., Guerra M.P., 2016. The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. Tree Genet Genomes 12: 10. DOI: 10.1007/s11295-016-0968-0
- Wheeler G.L., Dorman H.E., Buchanan A., Challagundla L., Wallace L.E., 2014. A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. Appl Plant Sci. DOI: 10.3732/apps.1400059
- Wicke S., Schneeweiss G.M., dePamphilis C.W., Müller K.F., Quandt D., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76: 273-297. DOI: 10.1007/s11103-011-9762-4
- Wolfe K.H., Li W.H., Sharp P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84:9054-9058. DOI: 10.1073/pnas.84.24.9054
- Ye L., Fu C., Wang Y., Liu J., Gao L., 2018. Characterization of the complete plastid genome of a Chinese endemic species *Carya kweichowensis*. Mitochondrial DNA Part B: Resources 3:492-493. DOI: 10.1080/23802359.2018.1464414
- Zhai D-C., Yao Q., Cao X-F., Hao Q-Q., Ma M-T., Pan J., Bai X-H., 2019. Complete chloroplast genome of the wild-type Hickory *Carya cathayensis*. Mitochondrial DNA Part B: Resources 4:1457-1458. DOI: 10.1080/23802359.2019.1598815
- Zhu A., Guo W., Gupta S., Fan W., Mower J.P., 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol 209:1747-1756. DOI: 10.1111/nph.13743

## Supporting information

**Figure S1.** *Carya illinoensis* cv. Imperial. (a) Fruits in intermediate stage of ripening. (b) Ripe nuts and (c) the kernel of the nut. (d) Physical map of the plastid genome of *C. illinoensis* cv. Imperial. The darker gray area in the inner circle represents the GC content, while light gray is the AT content. Coding regions containing introns are marked with one asterisk.

**Tables S1-S5.** Mean number of base substitutions per site and the corresponding standard errors of the hotspot sites of nucleotide divergence among 11 species of the tribe Juglandaceae.

**Figure S2.** Maximum Parsimony analysis of 11 species of tribe Juglandaceae, using the hotspot regions of nucleotide divergence, determined through a sliding window analysis.

**Table S6.** Examples of homologous SSR loci in *Carya illinoensis* cv. Imperial and *Carya illinoensis* MH909599.1. Loci 1 and 2 have the same motif, size and position in the plastid genome of both individuals. Loci 3 and 4 differ in the start and end position due to the one base deletion in *C. illinoensis* cv. Imperial. Loci 5 to 7 have differences concerning the number of repeats of the motif, representing putatively polymorphic markers.



**CAPÍTULO 2. Prospecção, caracterização e validação de marcadores SSR a partir do cloroplasto da noqueira-pecã “Imperial”**

## RESUMO

A noqueira-pecã (*Carya illinoensis*) é uma árvore frutífera com centro de origem em regiões temperadas do Hemisfério Norte, do Sul dos Estados Unidos até o México. Seu principal produto econômico é a noz-pecã, muito apreciada para consumo *in natura* ou processada. O estudo de diversidade genética e caracterização de cultivares são importantes no processo de melhoramento da espécie. Uma das técnicas utilizadas para estudo de diversidade e caracterização de cultivares é o desenvolvimento de marcadores moleculares altamente polimórficos, como as regiões microssatélite. O presente estudo teve como objetivo prospectar, caracterizar e validar marcadores microssatélite plastidiais em cultivares de *C. illinoensis* e testar a transferibilidade *in silico*, destes marcadores, em espécies da família Juglandaceae. Para tal, foram utilizados dados obtidos no sequenciamento e montagem do genoma plastidial da cultivar Imperial, para posterior prospecção de microssatélite e o desenho de *primers*. Foram identificadas 147 regiões microssatélite no genoma plastidial montado de *C. illinoensis* cv Imperial, sendo a frequência de trinucleotídeos igual a 47,6%, seguidos de 44,9% para mononucleotídeos e 5% de dinucleotídeos. A validação dos marcadores foi realizada *in silico* com o genoma plastidial da cultivar Imperial e com os genomas plastidiais de quatro cultivares (Lakota, Pawnee, Whichita e 87MX3-2.1) obtidos no banco de dados do GenBank/NCBI. Da validação *in silico* dos *primers*, de um total de 12 *primers* testados, 10 amplificaram apenas em um *locus* na região avaliada, demonstrando serem viáveis. Os 10 pares de *primers* plastidiais prospectados foram então validados através da amplificação via reação de PCR e eletroforese em gel de agarose, do DNA de 13 cultivares de *C. illinoensis*. O teste de transferibilidade *in silico* dos marcadores resultou em amplificação de 10 *loci* nas cultivares de *C. illinoensis*, nove em espécies do gênero *Carya* e sete no gênero *Juglans*, sugerindo seu potencial para aplicação em estudos genéticos dessas espécies. A transferibilidade de marcadores foi relativamente alta quando comparada com dados presentes na literatura. Porém, o fato de algumas espécies terem uma transferibilidade baixa, pode se dar também pelo baixo número de sequências disponíveis no banco de dados. Na validação dos marcadores em cultivares de *C. illinoensis*, apenas o *locus cpCil1* não amplificou para a cultivar Choctaw, e não houve amplificação para os *loci cpCil8* e *cpCil9* na cultivar Chickasaw. Marcadores SSR apresentam um alto custo para seu desenvolvimento e, por isso, uma alternativa de barateamento é de extremo interesse. Com a prospecção, caracterização e validação destes marcadores SSRs espécie-específicos, outros estudos genéticos poderão ser realizados utilizando essas marcas e assim avançar na caracterização genética da espécie e como subsídio a programas de melhoramento.

**Palavras-chave:** cpDNA, cpSSR, *Carya illinoensis*, *in silico*.

## 1 INTRODUÇÃO

Dentre as espécies pertencentes ao gênero *Carya*, destaca-se *Carya illinoensis* (Wangenh.) K. Koch, conhecida popularmente como noqueira-pecã (MARCHIORI, 1997). As florestas naturais de *C. illinoensis* abrangem uma faixa contínua do México até Illinois, nos Estados Unidos, sendo que os mais densos bosques de noqueira-pecã se configuram nos cumes aluviais (apenas além da faixa normal de inundação) ao longo dos rios afluentes de Oklahoma, Texas e Louisiana (MCWILLIAMS, 2013). No Brasil, o Rio Grande do Sul apresenta a maior produção de noqueira-pecã da América Latina, sendo *C. illinoensis* a principal espécie cultivada, destinando-se, unicamente, à produção de nozes comestíveis (GATTO et al. 2008). Além disso, a espécie é utilizada para a produção de lenha, quando o pomar é renovado sendo também empregada na arborização de parques e praças (BACKES e IRGANG, 2004).

A *C. illinoensis* é uma planta monóica, com flores pistiladas e estaminadas separadas, que ocorrem entre os meses de outubro e novembro. Os frutos são do tipo drupa, normalmente agrupados em cachos de três a sete unidades, com epicarpo que se separa do fruto na maturação, a parte aproveitável varia de 40 a 60%. No Brasil, o período de maturação do fruto acontece de março a maio, determinado pelas condições edafoclimáticas, fitossanitárias, nutricionais e, especialmente genéticas (DUARTE e ORTIZ, 2001). Normalmente a maturação ocorre entre 180 e 200 dias após a fecundação das flores (RASEIRA 1990). Entretanto, para que as plantas atinjam a produtividade esperada, é imprescindível que o pomar seja constituído de plantas pertencentes ao grupo de cultivares protândricas, ou seja, àquelas em que as flores estaminadas chegam primeiro à maturidade (Desirable, Western, Cheyenne, Caddo, Cape Fear, Success, Barton, Cherokee), e das cultivares protogínicas, em que as flores pistiladas chegam primeiro à maturidade (Choctaw, Wichita, Kiowa, Shawnee, Apache, Mohawk, Shoshoni, Chickasaw, Tejas, Elliott, Stuart, Comanche e Schley) (RASEIRA, 1990). As cultivares são propagadas por enxertia, e a produção de nozes inicia-se a partir quatro até oito anos após o plantio (BACKES e IRGANG, 2004).

Estudos genéticos relacionados à caracterização das cultivares de *C. illinoensis* plantadas no Brasil são escassos (DE OLIVEIRA et al. 2021). Portanto, é importante a utilização de marcadores moleculares com o objetivo de gerar informação sobre a diversidade genética, para auxiliar na elaboração de estratégias de programas de melhoramento da

espécie. Marcadores moleculares têm sido utilizados como ferramentas auxiliares nas diferentes etapas do melhoramento genético, desde a caracterização de germoplasma até as etapas finais de desenvolvimento e seleção de plantas melhoradas (FERREIRA e FALEIRO, 2008; FERREIRA e GRATAPAGLIA, 1998). Entre os marcadores moleculares disponíveis, os marcadores microssatélite apresentam grandes vantagens como a alta reprodutibilidade, codominância, alto conteúdo de polimorfismo e alelismo múltiplo (CHEN et al., 1997; LITT e LUTY, 1989). Conhecidos também como SSR (*Simple Sequence Repeats*), são marcadores moleculares eficientes para gerar informações que auxiliam em diversos estudos genéticos. As regiões microssatélite estão amplamente distribuídas no genoma nuclear dos eucariotos, além de serem localizadas também nos genomas mitocondrial e plastidial (POWELL et al., 1995; WANG et al., 2009).

O genoma plastidial (cpDNA) difere do genoma nuclear, na maioria dos casos, por ser haplóide e não recombinante (NOCK et al., 2011). A análise do genoma plastidial é utilizada para estudos evolutivos e ecológicos, com o objetivo de avaliar os padrões uniparentais de fluxo gênico dentro e entre populações, com algumas exceções, além disso, para classificar táxons, estudar filogenia, filogeografia (STEFENON et al., 2019) e gerar *barcoding* (PROVAN et al., 2001; MCPHERSON et al., 2013; WHEELER et al., 2014). Em plantas, é complexo detectar polimorfismos em populações com genomas de cloroplasto devido ao baixo grau de substituição e baixo percentual de mutação (MCPHERSON et al., 2013). Os marcadores microssatélite plastidiais (cpSSR) apresentam-se como potencial ferramenta para contornar esta dificuldade devido às modificações intraespecíficas (PROVAN et al., 2001).

Com os progressos na tecnologia de sequenciamento, o número e o tamanho das sequências aumentaram e hoje muitos genomas nucleares e de cloroplasto foram totalmente sequenciados. Os cpSSR são adequados para detectar gargalos em populações naturais, para medir a diversidade citoplasmática e a variação genética em material de melhoramento de plantas (FJELLHEIM et al., 2006). Em estudos de parentesco, os cpSSR funcionam como marcadores parentais paternos ou maternos, atuando como medidas diretas de fluxo de pólen ou semente. Os *loci*, por não serem recombinantes, fornecem aspectos que correspondem ao parental não sendo necessário inferir a fase gamética dos *loci* heterozigotos como nos microssatélites nucleares.

Até o momento, não há registro na literatura de marcadores microssatélite plastidiais desenvolvidos em *C. illinoensis*. Sendo assim, com a prospecção e validação dos marcadores SSR plastidiais (cpSSR) espécie-específicos, estudos genéticos poderão ser realizados utilizando essas marcas para promover o avanço nos estudos de caracterização

genética da espécie. Além disso, os marcadores plastidiais desenvolvidos para *C. illinoensis* podem ser uma ferramenta promissora através da transferibilidade para outras espécies da família viabilizando os estudos com espécies em que ainda não validaram marcadores espécie-específicos. Com isso, o objetivo deste estudo foi prospectar, caracterizar e validar marcadores cpSSR para estudos da diversidade genética de cultivares de *C. illinoensis*.

## 2 MATERIAL E MÉTODOS

O material vegetal foi gentilmente cedido pelo Engenheiro Florestal Tales Poletto, no âmbito do projeto “Caracterização molecular de cultivares da noqueira-pecã”, desenvolvido através da parceria entre UNIPAMPA, UFSC, UFSM e IBPecan.

### 2.1 Prospecção e caracterização dos marcadores cpSSR

O desenvolvimento de marcadores microssatélite e o desenho dos *primers* SSR foram realizados a partir do genoma plastidial da cultivar Imperial, sendo utilizado o sequenciamento e montagem reportados por Nagel et al. (2020). A prospecção de marcadores microssatélite plastidiais e o desenho de *primers* foram realizados através do software SSR locator (MAIA et al., 2008). Inicialmente, foram definidos limites mínimos para pesquisa: dez para repetições de mononucleotídeos, três para repetições de dinucleotídeos e cinco para repetições de tri-, tetra-, penta- e hexanucleotídeos. Os microssatélites adjacentes foram classificados como compostos, quando localizados a uma distância menor do que 25 (pb). Ou seja, o *software* considera duas regiões SSR com distâncias de até 25 (pb) como sendo uma mesma região composta.

### 2.2 Validação *in silico* dos marcadores cpSSR

A simulação da PCR *in silico* foi realizada para o genoma plastidial de todas as cultivares de *C. illinoensis* depositados no banco de dados de cpDNA do Taxonomy browser no site do NCBI (*National Center for Biotechnology Information*, disponível em: <http://www.ncbi.nlm.nih.gov/>). Foram obtidos dados de sequências de cpDNA para as cultivares Lakota, Pawnee, Whichita, Imperial e 87MX3-2.1 (disponíveis no GenBank sob os

números MH909600.1, MH909599.1, MN977124.1 e MN221384.1, respectivamente).

Por fim, foi empregado o programa SPCR, o qual fez a simulação da PCR *in silico*, necessitando de especificações exatas de temperatura de anelamento (45°C) e porcentagem de Guanina/Citosina (30%), para a primeira PCR *in silico* de *C. illinoensis*, seguida de cada uma das cultivares.

Para validar os marcadores *in silico* foi realizada a simulação de eletroforese em gel com a ajuda do software SIMGEL, incluído no pacote SPCR (CAO et al., 2005). A validação seguiu alguns critérios de avaliação: (i) não apresentar *loci* sobrepostos e (ii) os alelos apresentarem um tamanho próximo ao estimado durante a prospecção. Quando dois *loci* apresentaram tamanho desejado e sem sobreposição, foram testados em laboratório. Marcadores ótimos devem possuir apenas um *locus* na região desejada. Marcadores que não atingiram os dois critérios acima foram desconsiderados para análises em laboratório.

### **2.3 Transferibilidade dos marcadores *in silico***

Os marcadores desenvolvidos a partir do genoma plastidial de *C. illinoensis* cv Imperial foram testados quanto à transferibilidade *in silico*. Foram realizados os mesmos procedimentos de PCR e eletroforese virtual, utilizando sequências de genomas plastídias obtidas no banco de dados do NCBI para espécies da família Juglandaceae, a qual a espécie *C. illinoensis* pertence, dentre elas, quatro cultivares de *C. illinoensis*, seis espécies do gênero *Carya*, uma espécie do gênero *Cyclocarya* e 11 espécies do gênero *Juglans*.

### **2.4 Validação dos marcadores em cultivares de *C. illinoensis***

Para validação dos marcadores nas cultivares em laboratório, amostras foliares de 17 plantas (13 cultivares) foram coletadas em três municípios no Estado do Rio Grande do Sul (Anta Gorda, Estrela e Don Feliciano). Estas foram embaladas e imediatamente mantidas em PVP-40 2% (polivinilpirrolidona) para conservação da integridade do material genético durante o transporte ao Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal (LFDGV) da Universidade Federal de Santa Catarina (UFSC). Após conservação, as amostras foram mantidas em freezer para a posterior realização dos procedimentos e análises genéticas. Os tecidos foliares foram macerados e a extração de DNA total foi realizada por meio do método CTAB (DOYLE e DOYLE, 1987), com modificações.

Posteriormente, para verificar a qualidade do DNA extraído, realizou-se uma

eletroforese horizontal em gel de agarose 2%, corado com GelRed® e tampão para eletroforese TBE (Tris Borato-EDTA) na concentração de 1x. A quantidade e a qualidade do DNA isolado foram verificadas através do espectrofotômetro NanoDrop 1000® (Thermo Fisher Scientific, Wilmington, EUA) em relação à concentração e razões de absorbância 260/280 e 260/230.

Na etapa de validação *in silico* dos marcadores foram selecionados 10 *loci* que atenderam os critérios de avaliação determinados. O protocolo de amplificação dos 10 marcadores microssatélite plastidiais foi por meio de reações de PCR em Termociclador Veriti®. Para as reações de amplificação via PCR foram realizadas para um volume final de 15 µL, contendo 2,0 µl de DNA, 0,9 µl de dNTP (Sinapse Inc, Hollywood, Florida, USA), 0,2 µl de Taq Polimerase DNA (Ludwig), 1,5 µl tampão 10X (Ludwig), 1,20 µL de BSA, 0,75 µl de MgCl<sub>2</sub> (Ludwig), 1,0 µM de iniciador forward, 1,20 µM de iniciador reverse. A programação no Termociclador Veriti® foi realizada da seguinte maneira: 95°C por 3 min; seguido de 30 ciclos de 95°C por 15 s, temperatura de 60° de anelamento por 30 s, 72°C por 30 s; seguido de oito ciclos de 94°C por 30s, temperatura de 53°C por 45 s, 72°C por 45s; e a etapa de extensão final a 72°C por 10 min, sendo mantidos a 4°C.

A caracterização dos genótipos foi realizada por meio de eletroforese em gel de agarose 2,5%. Os géis foram corados com GelRed e visualizados em um transiluminador de luz UV.

## 2.6 Análise dos dados

O *software* TotalLab TL120 1D v2009 (Nonlinear Dynamics Ltd., Newcastle, Inglaterra) foi usado para análise das bandas no gel de agarose 2,5%. Primeiramente, fez-se a seleção das linhas/colunas, definindo-se o local de cada poço em que foi pipetada a amostra. Após, fez-se a detecção das bandas manualmente e, em seguida, a calibração do marcador de peso molecular 100 pb. Posteriormente, os valores dos pares de bases foram fornecidos por coluna. A partir de então, o programa calcula uma série de algoritmos para se verificar os picos de cada banda/coluna no gel, assim como os limites das bandas, esses limites podem ser ajustados manualmente. Ao lado da imagem do gel, o programa fornece um gráfico com a posição da intensidade (em pixels) da banda. Em seguida, foi selecionada cada coluna, sendo que o programa disponibiliza uma tabela informando os valores dos pares de bases.

### 3 RESULTADOS E DISCUSSÃO

#### 3.1 Prospecção e caracterização dos marcadores

Foram identificadas 147 regiões microssatélite no genoma plastidial de *C. illinoensis* cv Imperial (Tabela 1). A frequência de trinucleotídeos no genoma plastidial foi igual a 47,6%, seguidos de 44,9% para mononucleotídeos e 5% de dinucleotídeos (Tabela 1). De acordo com Azevedo (2007), os motivos dinucleotídeo, trinucleotídeo e tetranucleotídeos, encontrados em regiões microssatélite são os mais frequentes e polimórficos, embora haja vasta frequência de hexanucleotídeos, estes apresentam baixo polimorfismo.

Os *loci* selecionados para a validação em cultivares de *C. illinoensis* (ver a seguir), encontram-se todos em regiões intergênicas, não codificadoras do genoma plastidial da espécie.

**Tabela 1.** Regiões microssatélite localizadas no genoma plastidial de *C. illinoensis* cv Imperial.

Motivos	Quantidade de sequências	Porcentagem
Mononucleotídeos	66	44,9%
Dinucleotídeos	7	5%
Trinucleotídeos	70	47,6%
Tetranucleotídeos	2	1%
Pentanucleotídeos	1	<0,5%
Hexanucleotídeos	1	<0,5%
<b>Total</b>	<b>147</b>	<b>100%</b>

**Fonte:** Autor, 2022.

Os resultados obtidos neste estudo exibem diferenças, principalmente pelo quantitativo de amostras estudadas, assim como as regiões onde estão presentes os SSR. Em regiões intergênicas a incidência de SSR tende a ser maior (VIEIRA et al., 2016). A baixa frequência de motivos do tipo di e tetranucleotídeos foram observados para os marcadores desenvolvidos. Ebrahimi et al. (2018), analisando microsequências no gênero *Juglans* (Juglandaceae), detectaram a presença de motivos di-, tri- e tetranucleotídeos no genoma plastidial, mas as frequências variaram consideravelmente entre as espécies. O número total de motivos identificados no genoma plastidial foi pequeno em comparação com nSSR, mas as

diferenças no número de motivos entre as espécies, e na presença ou ausência de motivos entre as espécies foram grandes o suficiente para que pudessem ser usados para compreender a evolução dos plastídios. Foram encontrados em maior número os motivos AT, sendo os dinucleotídeos, as repetições AT/TA mais comuns (46%), os trinucleotídeos (ATT/ATA) também estavam presentes, porém raros (um a três motivos por genoma). Nesta mesma análise, os autores verificaram que os motivos de trinucleotídeos eram dependentes da espécie, com um a dois motivos identificados por espécie.

Em relação às regiões cpSSR identificadas em estudos com algumas espécies arbóreas cultivadas, a maioria foi classificada como mononucleotídeos (GEORGE et al. 2015, ANDRADE et al., 2018). Em 31 espécies de *Eucalyptus* o número total de cpSSR variou entre 71 para *Eucalyptus melliodora* e 135 para *Eucalyptus aromaphloia*, superior ao observado para *Pinus* spp., mas também com predominância de mononucleotídeos, e apenas 0,48% de di- e 3,57% de repetições de trinucleotídeos para espécies do gênero (ANDRADE et al., 2018). Além disso, 270 cpSSR foram observados para *P. koraiensis*, com predominância de tipo mononucleotídeo (GEORGE et al. 2015). Para o gênero *Pinus*, 1542 cpSSR foram desenvolvidos para 20 espécies, 56,15% delas em regiões não codificante, sendo a maior parte dos cpSSR classificada como mononucleotídeos (1333), mas também foram localizados 53 dinucleotídeos, 61 tri-, 62 tetra-, 17 penta- e 16 hexanucleotídeos repetidos (MORO et al. 2019). Para as espécies de *Pinus* as repetições de dinucleotídeos são muito mais abundantes do que tri- ou tetranucleotídeos (SMITH et al., 1994). Em espécies de cereais, os trinucleotídeos cpSSR são mais frequentes (54-78%), seguidos pelos dinucleotídeos (17,1-40,4%) e repetições de tetranucleotídeos (3-6%) (VARSHNEY et al., 2002).

Em estudo para família Juglandaceae, dentre eles para a espécie *Cyclocarya paliurus* (espécie monotípica da China Subtropical), foram encontradas o total de 153 sequências que continham repetições de microsatélite e 137 com regiões flangeadoras relativamente longas foram usadas para projetar *primers* (FAN et al., 2013). Na aplicação de 12 marcadores polimórficos EST-SSR nucleares em uma noqueira chinesa endêmica (*Juglans cathayensis*), usando tecnologia de sequenciamento de última geração, observou-se que as repetições de dinucleotídeos foram as mais frequentes (8.969; 39,9%), seguidas por repetições de trinucleotídeos (2.993; ou 13,3%) e repetições de tetranucleotídeos (310; ou 1,4%) (DANG et al., 2015). Para *Juglans hopeiensis* as repetições de mononucleotídeos foram as mais frequentes (7.208; ou 43,2%), seguidas por repetições de dinucleotídeos (6.180; ou 37,0%), então por repetições de trinucleotídeos (2.288; ou 13,7%) (HU et al., 2015). Esses trabalhos realizados foram a partir de DNA nuclear e, mesmo sendo espécies da mesma família,

demonstram uma variação na percentagem de motivos até mesmo dentro do gênero.

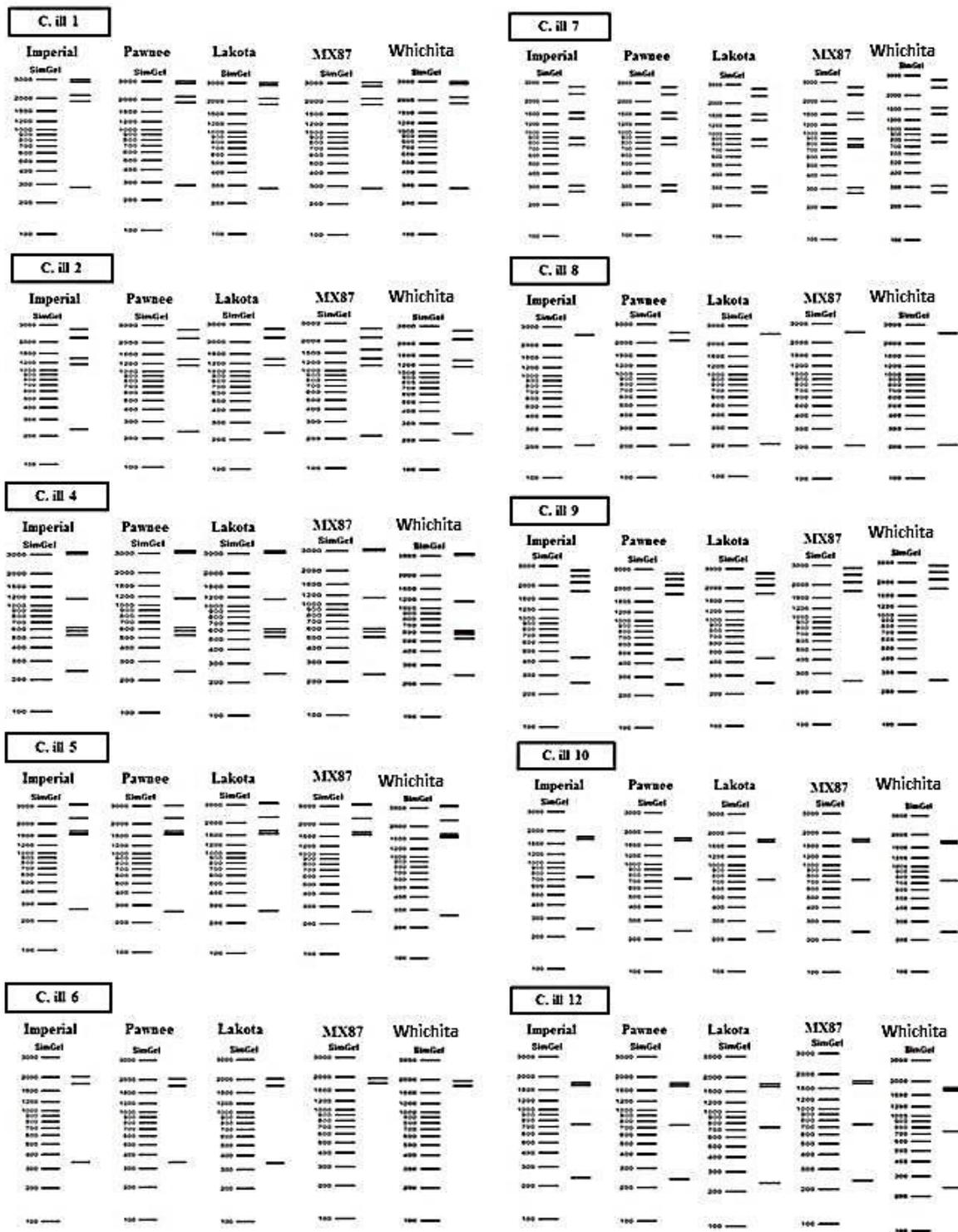
### 3.1 Validação *in silico* dos marcadores cpSSR de *Carya illinoensis*

Foi realizada uma amplificação *in silico* confrontando inicialmente as sequências de *C. illinoensis* através do software SPCR (CAO et al, 2005) utilizando uma faixa de 90 a 500 pb. Dos 12 *loci* prospectados, 10 foram considerados ótimos e aptos a serem sintetizados para teste (Tabela 2). A figura 1 demonstra o resultado da validação *in silico*.

**Tabela 2.** *Loci* microssatélites prospectados considerados ótimos para teste nas cultivares de *Carya illinoensis*.

<i>Locus</i>	<i>Primer sequence (5'-&gt;3')</i>	<b>RepeatMotif</b>	<b>Product Size</b>
<i>cpCil1</i>	GATCGGATTATATTTGTTTG TAAATGAAGTAGTTGGGTCT	(AAT) <sub>4</sub>	280
<i>cpCil2</i>	AGGAGATTCCCATATCTATC AATAAACTACGCAGAAAATG	(TA) <sub>10</sub>	237
<i>cpCil4</i>	AAATAAAGTAAGAAGGGCTC GACGAAATCAAAGAGGTACT	(AT) <sub>6</sub> -(AT) <sub>7</sub>	244
<i>cpCil5</i>	TGTACCCTCTTCAAATTCTA AATAACCGATTAGTCCTTCT	(ATT) <sub>6</sub>	267
<i>cpCil6</i>	CCTGAATATGATTCTACCAA GTGAAAAGGAGAACACATAG	(AATA) <sub>3</sub>	190
<i>cpCil7</i>	TCTTATCAATAAAATTTCCG GGAATAGTTCCTTTCATTCT	(ATT) <sub>5</sub>	270
<i>cpCil8</i>	AACGTTTTTATTGTTGTGTT TTTCGTAAAATCTCGTCTAC	(TTA) <sub>4</sub>	210
<i>cpCil9</i>	ATTACTCAAAGAATCGACA ATTTGTTTATTCCTTATCCC	(AT) <sub>6</sub>	257
<i>cpCil10</i>	CTTTCTGGATGTAGATGATG CGACGTAATTTTCATAGAGTC	(TA) <sub>6</sub>	239
<i>cpCil12</i>	CGACGTAATTTTCATAGAGTC CTTTCTGGATGTAGATGATG	(AT) <sub>6</sub>	239

Fonte: Autor, 2022.



**Figura 1.** Amplificação e eletroforese *in silico*. O SimGel corresponde ao marcador ou *ladder* em pares de bases, sendo a região avaliada entre 90 e 500 pb. São *primers* que demonstraram estar de acordo com os critérios propostos durante a validação. (Fonte: Autor, 2022).

Não existe critério para a definição do número mínimo de unidades repetitivas que caracterize um motivo microssatélite (YU et al. 1999). Porém, existe uma relação direta entre o número de repetições e o número de alelos em um microssatélite (WEBER et al. 1990). No entanto, microssatélites curtos apresentaram maior polimorfismo do que os longos (YU et al. 1999).

### **3.2 Transferibilidade *in silico* dos marcadores cpSSR**

Os resultados da transferibilidade *in silico* dos *primers* de *C. illinoensis* para outras espécies de Juglandaceae demonstraram que a grande maioria dos *loci* amplificaram com sucesso nas cultivares e em espécies do gênero *Carya* (Tabela 3), a porcentagem de amplificação entre os *loci* variou entre 30,4% (*cpCil1*) e 95,7% (*cpCil2*, 5, 9, 10 e 12). O número de *loci* amplificados por espécie apresentou ampla variação, os 10 marcadores obtiveram amplificação bem-sucedida nas cultivares de *C. illinoensis*, nove em espécies do gênero *Carya* e sete no gênero *Juglans*, sugerindo seu potencial para aplicação em estudos genéticos dessas espécies. Portanto, constatada a baixa transferibilidade para a espécie *Juglans ailanthifolia*, e a não amplificação dos *primers* *cpCil1* e *cpCil7* para todas as espécies do mesmo gênero.

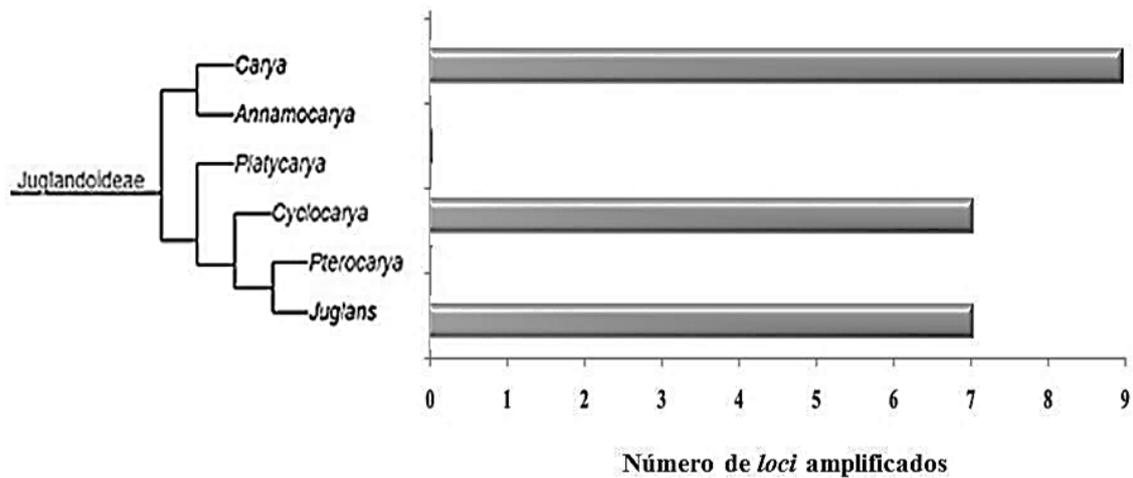
As regiões flanqueadoras dos microssatélites são bastante conservadas, o que possibilita a utilização dos *primers* desenvolvidos para uma espécie em outras espécies pertencentes ao mesmo gênero ou até mesmo gêneros distantes da mesma família (FERREIRA e GRATTAPAGLIA, 1998, BARBARÁ et al., 2007). O sucesso de transferência entre espécies é maior para as alógamas ou com sistema de reprodução misto e quando o tamanho do genoma da espécie alvo é menor do que o da espécie da qual foram originados os microssatélites (BARBARÁ et al., 2007).

**Tabela 3.** Transferibilidade de *primers* microsatélite em espécies da família Juglandaceae.

Gênero/Espécies	<i>cpCil1</i>	<i>cpCil2</i>	<i>cpCil4</i>	<i>cpCil5</i>	<i>cpCil6</i>	<i>cpCil7</i>	<i>cpCil8</i>	<i>cpCil9</i>	<i>cpCil10</i>	<i>cpCil12</i>	No. de loci
<b><i>Carya</i></b>											
<i>C. illinoensis</i> cv Imperial	X	X	X	X	X	X	X	X	X	X	10
<i>C. illinoensis</i> cv Whichita	X	X	X	X		X	X	X	X	X	9
<i>C. illinoensis</i> cv Pawnee	X	X	X	X	X	X	X	X	X	X	10
<i>C. illinoensis</i> cv Lakota	X	X	X	X	X	X	X	X	X	X	10
<i>C. illinoensis</i> cv 87MX3-2.11	X	X	X	X		X	X	X	X	X	9
<i>C. sinensis</i>			X	X	X	X	X	X	X	X	9
<i>C. ovata</i>	X	X	X	X		X	X	X	X	X	9
<i>C. laciniosa</i>	X	X	X	X		X	X	X	X	X	9
<i>C. kweichowensis</i>		X	X	X	X	X	X	X	X	X	9
<i>C. hunanensis</i>		X		X	X	X	X	X	X	X	8
<i>C. cathayensis</i>		X	X	X	X	X	X	X	X	X	9
<b><i>Cyclocarya</i></b>											
<i>C. paliurus</i>		X	X	X	X			X	X	X	7
<b><i>Juglans</i></b>											
<i>J. sigillata</i>		X	X	X	X			X	X	X	7
<i>J. regia</i>		X	X	X	X			X	X	X	7
<i>J. nigra</i>		X	X	X	X		X	X	X	X	8
<i>J. microcarpa</i>		X	X	X	X		X	X	X	X	8
<i>J. mandshurica</i>		X	X	X	X			X	X	X	7
<i>J. hopeiensis</i>		X	X	X	X			X	X	X	7
<i>J. hindsii</i>		X	X	X	X		X	X	X	X	8
<i>J. grande</i>		X	X	X	X			X	X	X	7
<i>J. cinerea</i>		X	X	X	X		X	X	X	X	8
<i>J. cathayensis</i>		X	X	X	X			X	X	X	7
<i>J. ailanthifolia</i>		X									1
<b>% de transferibilidade</b>	30,4	95,7	91,3	95,7	78,3	47,8	65,2	95,7	95,7	95,7	

Fonte: Autor, 2022.

Com relação ao número de *loci* amplificados por gênero, obteve-se uma média de nove que amplificaram em espécies do gênero *Carya* à qual pertence *C. illinoensis*, sete na espécie do gênero *Cyclocarya* e sete em *Juglans* (Figura 2).



**Figura 2.** Número de *loci* amplificados em cada um dos gêneros analisados. À esquerda pode-se observar a árvore filogenética que representa as relações de distância entre os seis gêneros que pertencem à tribo Juglandoideae, adaptado de Xiang et al. (2014).

Os microssatélites plastidiais desenvolvidos a partir de *C. illinoensis* cv Imperial demonstraram um elevado grau de transferibilidade entre diversas espécies da tribo Juglandoideae. Isto sugere que as regiões que flanqueiam as sequências SSR, identificadas nesta biblioteca, são bem conservadas entre as espécies estudadas, mesmo pertencendo a gêneros mais distantes filogeneticamente de *Carya*, à qual *C. illinoensis* pertence. Todos os pares de *loci* descritos neste trabalho amplificaram *in silico* em pelo menos para uma espécie dentro dos gêneros analisados da família Juglandaceae e, embora de um modo geral as taxas de sucesso tenham diminuído com o aumento da distância evolutiva, isto não foi constante para todos os marcadores e espécies analisados (Tabela 3).

A percentagem de *loci* amplificados *in silico* em *Juglans* (70%), gênero mais distante de *Carya*, foi igual à do gênero *Cyclocarya* (70%), que é mais próxima filogeneticamente. No entanto, é importante ressaltar que o número de espécies testadas para cada gênero não foi o mesmo e que essa diferença pode ter contribuído para a maior ou menor percentagem de transferibilidade em determinado gênero. Esse é o caso, por exemplo, do gênero *Cyclocarya*, que apresentou equivalência em percentagem de transferibilidade, mas teve uma única espécie testada (*C. paliurus*). Outro aspecto a ser levado em consideração, é que, a transferência não depende somente da relação filogenética entre as espécies, mas também de outros fatores como tamanho e complexidade do genoma e região na qual os microssatélites estão localizados (codificadoras ou não codificadoras) (OLIVEIRA et al. 2006, BARBARÁ et al. 2007).

O caminho mais comum empregado em estudos com cpSSR é a utilização de marcadores desenvolvidos em outras espécies. Ebert e Peakall (2009) realizaram uma revisão

bibliográfica sobre as aplicações de cpSSR para uma ampla variedade de espécies de plantas e 64% dos estudos retornados aplicaram ensaios de marcadores cpSSR que foram desenvolvidos para espécies não relacionadas. Constituindo a maioria desses estudos, observou-se um subconjunto dos 10 marcadores *ccmp* desenvolvidos por Weising e Gardner (1999) ou os 36 marcadores *ntcp* de Bryan et al. (1999), ambos derivados do tabaco (*Nicotiana tabacum*). Novos estudos vêm surgindo, os quais têm se baseado em análises de polimorfismos de cpSSR universais ou sequenciamento de regiões intergênicas (COLLEVATTI et al., 2003; LIAN et al., 2003; LIRA et al., 2003; MARTINS et al., 2006). Devido ao fato de ser uma molécula altamente conservada entre as espécies de angiospermas, o DNA plastidial permite essa utilização em nível universal, não sendo necessário, como no caso de DNA nuclear, isolar sequências específicas de SSR (AZEVEDO et al., 2008). O uso de marcadores universais tem benefícios em termos de custos de desenvolvimento, no entanto, como estratégia para acessar a variação do cpSSR em espécie nova, é inerentemente limitado (AZEVEDO et al., 2008).

Com a escassez de estudos relacionados à cpSSR na família Juglandaceae, observam-se estudos com transferibilidade de *locus* microssatélite nucleares não codificantes e expressos entre espécies taxonomicamente relacionadas, como observado para o gênero *Juglans* (Juglandaceae) (ROSS-DAVIS e WOESTE, 2008) e para o gênero *Carya* (GRAUKE et al. 2010). Estudos prévios relataram altas taxas de transferência dentro de um táxon, com o declínio de sucesso na transferibilidade com o aumento da distância genética na análise com nSSR (SAHA et al. 2006; BARBARÁ et al. 2007; GRAUKE et al., 2010; ROSS-DAVIS e WOESTE, 2008). Ross-Davis e Woeste (2008) testaram a transferibilidade de 10 *loci* nucleares desenvolvidos para *Juglans nigra* em 14 espécies do mesmo gênero e na espécie *C. illinoensis*, indicando o alto nível de homologia entre o genoma destas espécies. Os *loci* amplificados na maioria das espécies examinadas eram altamente polimórficos, mostrando 18–32 alelos por *locus* entre as espécies, entretanto, o polimorfismo dentro dos táxons foi geralmente baixo, variando de 1,3 para *J. xquadrangulata* a 13 para *J. cinerea*. Isso provavelmente ocorreu devido ao tamanho limitado da amostra, quatro dos dez *loci* foram amplificados em todas as espécies. A transferência de três marcadores para *C. illinoensis* observou-se incomum o que apontou para uma forte conservação desses *loci* ou uma divergência relativamente recente dos gêneros (HALE et al. 2005). Fan et al. (2013) desenvolveram 28 marcadores microssatélite nucleares para *Ciclocarya paliurus*. Neste estudo, a transferibilidade dos marcadores foi testada em *Juglans regia*, *Pterocarya stenoptera* e *Platycarya strobilacea*. Vinte e um (75,0%), 22 (78,6%) e 15 (53,6%) *loci* foram

amplificados com sucesso em *J. regia*, *P. stenoptera* e *P. strobilacea*, respectivamente. Mais da metade foi amplificado com sucesso em três espécies intimamente relacionadas, sendo assim, esses marcadores representaram um conjunto de ferramentas moleculares úteis para estudos de genética populacional em outras espécies de Juglandaceae.

Em um estudo em *Pterocarya stenoptera*, 2452 microssatélite nucleares foram descobertos no transcriptoma. Cem *loci* foram selecionados aleatoriamente para verificar a eficiência da amplificação. Destes *loci* testados, 15 apresentaram polimorfismo. Este foi o primeiro conjunto de marcadores microssatélite desenvolvidos especificamente para *P. stenoptera* o que se tornou útil para estudos de genética populacional e genética da paisagem da espécie. Todos os marcadores foram amplificados com sucesso na espécie relacionada *P. hupehensis*, sugerindo que eles também poderiam ser usados para estudar outras espécies relacionadas à *Pterocarya* (WANG et al., 2018).

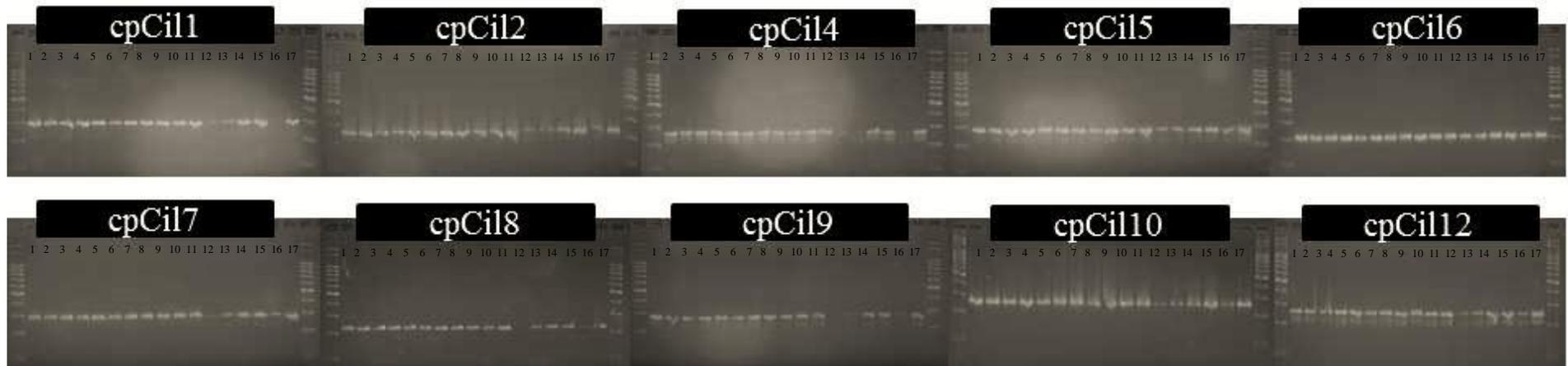
Os resultados deste trabalho também indicaram ótimo potencial de transferência dos marcadores microssatélite plastidiais no mesmo gênero (91%) e entre gêneros mais distantes (de 65,5 a 70%) da família Juglandaceae. Este foi um resultado importante, pois os marcadores cpSSR prospectados neste trabalho podem contribuir para o estudo de espécies de Juglandaceae, diminuindo o custo e tempo necessários para o desenvolvimento de marcadores específicos para cada espécie. Os microssatélites encontrados no genoma plastidial, são indicados para utilização em estudos evolutivos, filogeográficos, filogenéticos e de fluxo gênico (MACHADO et al., 2022). Eventos de deriva genética, efeito fundador e gargalos genéticos (*bottlenecks*) são analisados de maneira mais detalhada utilizando cpSSR e são considerados marcadores complementares ao SSR nucleares (POWELL et al., 1995; PROVAN; POWELL; HOLLINGSWORTH, 2001). Esta contribuição se torna ainda mais relevante se considerarmos que não há estudos de transferibilidade em espécies da família, relacionados à cpDNA, e que Juglandaceae possui importância econômica relevante nomeadamente na produção de nozes-comuns (*Juglans*), pecãs (*Carya illinoensis*) e hícória (*Carya*). A noz produzida pela espécie *Juglans regia* é uma das principais produzidas comercialmente em nível mundial (CHRISTENHUSZ e BYNG, 2016).

### **3.3 Validação dos marcadores em cultivares de *C. illinoensis***

Foram realizadas as reações de PCR conforme descritas na metodologia para 17 amostras (13 cultivares; Tabela 4) com os 10 marcadores prospectados e os produtos de PCR foram separados através de eletroforese em gel de agarose 2,5% (Figura 3).

**Tabela 4.** Número da amostra e as respectivas cultivares de *C. illinoensis*.

<b>Amostra</b>	<b>Cultivar</b>
1	Moneymaker
2	Barton
3	Desirable
4	Desirable
5	Jackson
6	Farley
7	Imperial
8	Imperial
9	Imperial
10	Sucess
11	Barton
12	Chickasaw
13	Shawnee
14	Stuart
15	Importada
16	Choctaw
17	Mahan



**Figura 3.** Géis de eletroforese (agarose 2,5%) com os produtos da amplificação dos 10 *locus* SSR plastidiais selecionados a partir do sequenciamento do plastoma de *C. illinoensis* cv Imperial, para as 17 amostras de noqueira-pecã, para cada marcador, e a esquerda e direita com *ladder* de 100pb. (Fonte: Autor, 2022).

### 3.4 Genotipagem

Com o auxílio do *software* TotalLab TL120, foram verificadas as bandas nos géis para conferir o posicionamento dos *loci* amplificados (Tabelas 5).

**Tabela 5.** Tamanho do alelo observado (pb) obtidas para as 17 amostras de *C. illinoensis*.

<b>Cultivar</b>	<b><i>cpCil1</i></b>	<b><i>cpCil2</i></b>	<b><i>cpCil4</i></b>	<b><i>cpCil5</i></b>	<b><i>cpCil6</i></b>
MoneyMaker	351	286	285	329	217
Barton	356	277	294	320	214
Desirable	351	285	292	315	212
Desirable	350	288	295	317	208
Jackson	353	281	288	329	206
Farley	355	272	283	320	205
Imperial	353	267	292	318	211
Imperial	351	263	293	317	216
Imperial	345	276	290	311	208
Sucess	347	276	296	315	214
Barton	340	260	300	318	213
Chickasaw	351	290	289	327	212
Shawnee	357	291	295	326	211
Stuart	349	285	300	321	214
Importada	350	272	291	321	210
Choctaw	--	297	283	319	214
Mahan	353	271	296	315	212

**Fonte:** Autor, 2022.

**Tabela 5.** Tamanho do alelo observado (pb) para as 17 amostras de *C. illinoensis* (continuada).

Cultivar	<i>cpCil7</i>	<i>cpCil8</i>	<i>cpCil9</i>	<i>cpCil10</i>	<i>cpCil12</i>
Moneymaker	351	304	252	264	304
Barton	349	302	248	263	304
Desirable	343	300	246	259	304
Desirable	344	310	243	253	303
Jackson	340	310	245	259	303
Farley	345	306	245	258	301
Imperial	344	304	245	257	301
Imperial	348	306	246	255	300
Imperial	344	304	248	249	300
Sucess	349	306	253	250	299
Barton	353	310	253	250	299
Chickasaw	342	--	--	246	300
Shawnee	350	314	257	248	300
Stuart	347	314	257	242	300
Importada	346	316	251	240	298
Choctaw	350	314	254	245	300
Mahan	327	314	253	244	299

Fonte: Autor, 2022.

Os 10 marcadores plastidiais prospectados permitiram amplificar com sucesso as regiões microssatélite das cultivares de *C. illinoensis*. Apenas o locus *cpCil1* não amplificou para a cultivar Choctaw (Tabela 5). Não houve amplificação para os loci *cpCil8* e *cpCil9* na cultivar Chickasaw (Tabela 5). Esse polimorfismo no tamanho dos alelos é importante, pois determina a capacidade informativa desses marcadores. Contudo, a utilização de métodos eletroforéticos mais precisos, como a eletroforese capilar em sequenciadores de DNA, é recomendada para a devida caracterização dos alelos presentes em cada cultivar.

Não há estudos com prospecção e validação de cpSSR no gênero *Carya*, pois a maioria está baseado em marcadores plastidiais universais. Cheng et al. (2006), Weising e Gardner (1999) e Bryan et al. (1999) desenvolveram marcadores plastidiais universais a partir de *Nicotiana tabacum* e *Arabidopsis thaliana* que foram capazes de amplificar microssatélite plastidiais em uma ampla gama de espécies de dicotiledôneas frutíferas.

Grauke et al. (2010) utilizaram os marcadores plastidiais universais desenvolvidos por Weising and Gardner (1999) e Bryan et al. (1999) em *C. illinoensis* e compararam as informações fornecidas por esses marcadores com as fornecidas pelos SSR nucleares espécie-específicos. Os marcadores de cpDNA exibem uma taxa evolutiva mais lenta, sendo assim, alguns marcadores com uma taxa evolutiva mais rápida, como os microssatélite, podem agregar informações aos resultados com marcadores citoplasmáticos e ajudar a elucidar a

história evolutiva mais recente de espécies relacionadas. Nesse estudo, três dos oito marcadores de cloroplastos produziram polimorfismos reprodutíveis e informativos: *ccmp2*, *ntcp40* e *ntcp9* em *C. illinoensis*. Os alelos variaram em apenas um par de bases, mas puderam ser distinguidos de forma confiável.

O uso de marcadores universais tem benefícios relativos aos custos de desenvolvimento, no entanto, como estratégia para acessar a diferença do cpSSR em espécie nova, é bastante limitado. A alternativa para não depender destes marcadores universais para análises de diversidade genética na espécie *C. illinoensis* foi desenvolver novos marcadores cpSSR espécie-específicos.

## 5 CONCLUSÃO

Os marcadores microssatélite recém-desenvolvidos para *C. illinoensis* são úteis em estudos genéticos, sendo possível validar estes marcadores e gerar informações a respeito da diversidade genética das cultivares de noqueira. Portanto, com a caracterização e validação destes marcadores SSRs espécie-específicos, outros estudos genéticos poderão ser realizados utilizando essas marcas e assim avançar na caracterização genética da espécie como subsídio a programas de melhoramento.

Observou-se um alto potencial de transferibilidade dos marcadores microssatélite plastidiais desenvolvidos para *C. illinoensis* em outras espécies no mesmo gênero (91%) e entre gêneros mais distantes (de 65,5 a 70%) da família Juglandaceae. O número de *loci* amplificados por espécie apresentou ampla variação sendo que os 10 marcadores amplificaram nas cultivares de *C. illinoensis*, nove em espécies do gênero *Carya*, e sete no gênero *Juglans*, sugerindo seu potencial para aplicação em estudos genéticos dessas espécies.

A utilização de ferramentas de bioinformática facilitou a realização de tarefas laboratoriais, economizando reagentes, confrontando sequências de DNA e gerando resultados potencialmente significativos. Através dos resultados obtidos pode-se concluir que há possibilidade de utilização dos marcadores SSR plastidiais nas cultivares de *C. illinoensis*, tendo a validação *in silico* de marcadores SSR como uma alternativa de barateamento de extremo interesse, permitindo excluir previamente marcadores que não seriam ideais.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, M. C.; PEREK, M.; PEREIRA, F. B.; MORO, M.; TAMBARUSSI, E. V. Quantity, organization, and distribution of chloroplast microsatellites in all species of *Eucalyptus* with available plastome sequence. **Crop Breeding and Applied Biotechnology**, v. 18, p. 97-102, 2018.

AZEVEDO, V. C. R. **Desenvolvimento e aplicações de microsatélites, análise de cpDNA e modelagem computacional para estudos da estrutura e dinâmica genética de maçaranduba - *Manilkara huberi* (Ducke) Chev. Sapotaceae.** Universidade de Brasília. 2007.

BACKES, P.; IRGANG, B. Árvores do Sul. **Guia de identificação & interesse Ecológico.** As principais espécies nativas Sul-Brasileiras. Santa Cruz do Sul: Instituto Souza Cruz, 204 p. 2004.

BARBARÁ, T., PALMA-SILVA, C., PAGGI, G.M., BERED, F., FAY, M.F. & LEXER, C. 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. **Molecular Ecology** 16: 3759-3767.

CAO, Y.; WANG, L.; XU, K.; KOU, C.; ZHANG, Y.; WEI, G.; HE, J.; WANG, Y.; ZHAO, L. Information theory-based algorithm for *in silico* prediction of PCR products with whole genomic sequences as templates. **BMC Bioinformatics**, v.6, n.190, 2005.

COLLEVATTI, R.G.; GRATTAPAGLIA, D.; HAY, J.D. Evidences for multiple maternal lineages of *Caryocar basiliense* populations in the Brazilian Cerrado based on the analysis of chloroplast DNA sequences and microsatellite haplotype variation. **Molecular Ecology**, v. 12, p. 105-115, 2003.

CARON, H.; DUMAS, S.; MARQUE, G.; MESSIER, C.; BANDO, E.; PETIT, R.J.; KREMER, A. Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. **Molecular Ecology**, v. 9, p. 1089-1098, 2000.

CHEN, X.; TEMNYKH, S.; XU, Y.; CHO, Y. G.; MCCOUCH, S. R. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). **Theoretical and Applied Genetics**, v. 95, n. 4, p. 553–567, 1997.

CHEN, M.; TAN, Z.; ZENG, G.; PENG, J. Comprehensive analysis of simple sequence repeats in Pre-miRNAs. **Molecular Biology and Evolution**, v. 27, n. 10, p. 2227–2232. 2010.

CHRISTENHUSZ, M. J. M.; BYNG, J. W. The number of known plants species in the world and its annual increase. **Phytotaxa**. 261 (3): 201–217. doi:10.11646/phytotaxa.261.3.1. 2016.

CLOUTIER, D.; PÓVOA, J.S.R.; PROCÓPIO, L.C.; LEÃO, N.V.M.; WADT, L.H. de O; CIAMPI, A.Y.; SCHOEN, D.J. Chloroplast DNA variation of *Carapa guianensis* in the Amazon basin. **Silva e Genética**, v. 54, n. 6, p. 257-300, 2005.

DANG, MENG & LIU, ZHI-XI & CHEN, XI & ZHANG, TIAN & ZHOU, HUI-JUAN &

YIHENG, HU & ZHAO, PENG. Identification, development, and application of 12 polymorphic EST-SSR markers for an endemic Chinese walnut (*Juglans cathayensis* L.) using next-generation sequencing technology. **Biochemical Systematics and Ecology**. 60. 74-80. 10.1016/j.bse.2015.04.004. 2015.

DE OLIVEIRA L.O.; BEISE D.C.; DOS SANTOS D.D.; NAGEL, J.C.; POLETTO, T.; POLETTO, I.; STEFENON, V.M. Molecular markers in *Carya illinoensis* (Juglandaceae): from genetic characterization to molecular breeding, **The Journal of Horticultural Science and Biotechnology**, DOI: 10.1080 / 14620316.2021.1892534. 2021.

DOYLE, J.J.; DOYLE, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin**, v.19, p.11-15, 1987.

DUARTE, V.; ORTIZ, E. R. N. Podridão de *Phytophthora* da amêndoa e da casca da noqueira pecan. In: LUZ, E. D. M. N. et al. (Ed.). Doenças causadas por *Phytophthora* no Brasil. Campinas: Rural, 2001. p. 493-508.

EBERT D, PEAKALL R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. **Mol Ecol Resour**. 2009 May;9(3):673-90. doi: 10.1111/j.1755-0998.2008.02319.x. Epub 2009 Jan 28. PMID: 21564725.

EBRAHIMI, AZIZ & MATHUR, SAMARTH & LAWSON, SHANEKA & LABONTE, NICHOLAS & LORCH, ADAM & COGGESHALL, MARK & WOESTE, KEITH. Microsatellite Borders and Micro-sequence Conservation in *Juglans*. **Scientific Reports**. 9. 3748. 10.1038/s41598-019-39793-z. 2019.

FAN, D. M. , YE L. J., LUO Y., HU W., TIAN S., AND ZHANG Z. Y.. Development of microsatellite loci for *Cyclocarya paliurus* (Juglandaceae), a monotypic species in subtropical China. **Applications in Plant Sciences** 1: 1200524. 2013.

FERREIRA, M. E.; FALEIRO, F. G. Biotecnologia: avanços e aplicações no melhoramento genético vegetal. In: FALEIRO, F. G.; FARIAS NETO, A. L. (Eds.). **Savanas: desafios e estratégias para o equilíbrio entre sociedade, agronegócio e recursos naturais**. Planaltina, DF: Embrapa Cerrados. p. 765–792. 2008.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3o ed. Brasília, DF: EMBRAPA-CENARGEN, 1998.

FJELLHEIM, S. et al. Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. **Journal of Biogeography**, v. 33, n. 8, p. 1470-1478, 2006.

GALVÃO, K. S. C. et al. Functional molecular markers (EST-SSR) in the full-sib reciprocal recurrent selection program of maize (*Zea mays* L.). **Genetics and Molecular Research**, n. 14, p. 7344-7355, 2015.

GAO, C.; REN, X.; MASON, A. S.; et al. Revisiting an important component of plant genomes: Microsatellites. **Functional Plant Biology**, v. 40, n. 7, p. 645–661, 2013.

GATTO, D. A. et al. Características tecnológicas das madeiras de *Luehea divaricata*, *Carya*

*illinoensis* e *Platanus x acerifolia* quando submetidas ao vergamento. **Ciência Florestal**, Santa Maria, v. 18, n. 1, p. 121-131, 2008.

GEORGE, B.; BHATT, B. S.; AWASTHI, M.; GEORGE, B.; SINGH, A. K. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. **Currents Genetics**, v. 61, n. 4, p. 665-677, 2015.

GRAUKE, L.J., MENDOZA-HERRERA, M.A. AND BINZEL, M.L. Plastid microsatellite markers in *Carya*. **Acta Hortic.** 859, 237-246 DOI: 10.17660/ActaHortic.2010.859.27 <https://doi.org/10.17660/ActaHortic.2010.859.27>. 2010.

HALE M.L., BORLAND A.M., WOLF K. High degree of conservation of nuclear microsatellite loci in the genus *Clusia*. **Genome** 48:946–950. 2005.

HU, Y. & ZHAO, PENG & ZHANG, QIANG & WANG, YANG & GAO, X.-X & ZHANG, TIAN & ZHOU, H.-J & DANG, MENG & WOESTE, KEITH. De novo assembly and characterization of transcriptome using Illumina sequencing and development of twenty five microsatellite markers for an endemic tree *Juglans hopeiensis* Hu in China. **Biochemical Systematics and Ecology**. 63. 201-211. 10.1016/j.bse.2015.10.011. 2015.

JOY, N.; SONIYA, E. V. Identification of an miRNA candidate reflects the possible significance of transcribed microsatellites in the hairpin precursors of black pepper. **Functional and Integrative Genomics**, v. 12, n. 2, p. 387–395, 2012.

LIAN, C.; OISHI, R.; MIYACHITA, N.; NARA, K.; NAKAYA, H.; WU, B.; ZHOU, Z.; HOGETSU, T. Genetic structure and reproduction dynamics of *Salix reinii* during primary succession on Mount Fuji, as revealed by nuclear and chloroplast microsatellite analysis. **Molecular Ecology**, v. 12, p. 609-618, 2003.

LIRA, C.F.; CARDOSO, S.R.S.; FERREIRA, P.C.G.; CARDOSO, M.A.; PROVAN, J.. Long-term population isolation in the endangered tropical tree species *Caesalpinia echinata* Lam. revealed by chloroplast microsatellites. **Molecular Ecology**, v. 12, p. 3219-3225, 2003.

LITT, M.; LUTY, J. A. A Hypervariable Microsatellite Revealed by In Vitro Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. **American Journal of Human Genetics**, v. 44, p. 397–401, 1989.

MAIA, L.C.; PALMIERI, D.A.; DE SOUZA, V.Q.; KOPP, M.M.; DE CARVALHO, F.I.; COSTA DE OLIVEIRA, A. **SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation**. International Journal of Plant Genomics, v. 2008, n. 41 p. 26-96, 2008.

LILIAN DE OLIVEIRA MACHADO L. O., VALDIR MARCOS STEFENON V.M., LEILA DO NASCIMENTO VIEIRA L. N., RUBENS ONOFRE NODARI R. O. Structural and evolutive features of the *Plinia phitrantha* and *P. cauliflora* plastid genomes and evolutionary relationships within tribe Myrteae (Myrtaceae). **Genetics and Molecular Biology**, 45,1, e20210193. 2022.

MARCHIORI, J.N.C. **Dendrologia das angiospermas: das magnoliáceas às flacourtiáceas**. Santa Maria: Ed. da UFSM. 271 p. 1997.

MARTINS, K.; CHAVES, L.J.; BUSO, G.S.C.; KAGEYAMA, P.Y. Mating system and finescale spatial genetic structure of *Solanum lycocarpum* St.Hil. (Solanaceae) in the Brazilian Cerrado. **Conservation Genetics**, DOI 10.1007/s10592-006-9140-y, 2006.

MCCAULEY DE. The use of chloroplast DNA polymorphism in studies of gene flow in plants. **Trends in Ecology and Evolution** 10: 198-202. 1995.

MCPHERSON, H. et al. Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. **BMC ecology**, v. 13, n. 1, p. 8, 2013.

MCWILLIAMS, J. The pecan: A history of America's native nut. **Austin: Hardcover**. 178 p. 2013.

MORO, MICHELE e BORTOLANZA PEREIRA, FERNANDA e FERREIRA FILHO, JAIRE e PEREK, MATHEUS e PERES, FABIANA e DE ASSIS FURQUIM, GABRIEL e TAMBARUSSI, EVANDRO. *In silico* analysis of *Pinus L.* chloroplast DNA to microsatellites regions. **Scientia Forestalis**. 47. 10.18671/scifor.v47n123.16. 2019.

NAGEL, J. C. et al. Structural, evolutionary and phylogenomic features of the plastid genome of *Carya illinoensis* cv. Imperial. **Annals of Forest Research**, [S.l.], p. 3-18. ISSN 20652445. Available at: <<https://www.afrjournal.org/index.php/afr/article/view/1413>>. Date accessed: 07 Aug. 2021. doi:10.15287/afr.2019.1413. 2020.

NEWTON AC, ALLNUT TR, GILLIES ACM, LOWE AJ, ENNOS RA. Molecular phylogeography, intraspecific variation and the conservation of tree species. **Trends Ecology & Evolution** 14: 140-145. 1999.

NOCK, C. J. et al. Chloroplast genome sequences from total DNA for plant identification. **Plant biotechnology journal**, v. 9, n. 3, p. 328-333, 2011.

PRENTICE, H.C.; MALM, J.U.; MATEU-ANDRÉS, I.; SEGARRA-MORAGUES, J. Allozyme and chloroplast DNA variation and mainland populations of the rare Spanish endemic, *Silene hifacensis* (Cayophyllaceae). **Conservation Genetics**, v. 4, p. 543-555, 2003.

POWELL, W.; MORGANTE, M.; ANDRE, C.; MCNICOL, J.W.; MACHRAY, G.C.; DOYLE, J.J.; TINGEYT, S.V.; RAFALSKI, J.A. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. **Current Biology**, v. 5, n. 9, p. 1023-1029, 1995.

POWELL, W.; MACHRAY, G.C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends in Plant Science**, v. 1, n. 7, p. 215-222, 1996.

PROVAN, J.; POWELL, W.; HOLLINGSWORTH, P. M. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. **Trends in Ecology & Evolution**, v. 16, n. 3, p. 142-147, 2001.

RASEIRA, A. **A cultura da noqueira pecã (*Carya illinoensis*)**. Embrapa Clima Temperado, Pelotas. Comunicado Técnico nº 63. 3p. 1990.

SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**.

v. 27. p. 379-623. 1948.

SMITH, D. N.; DEVEY, M. E. Occurrence and inheritance of microsatellites in *Pinus radiata*. **Genome**, v. 37, n. 6, p. 977-983, 1994.

STEFENON V.M., KLABUNDE G., LEMOS R.P.M., ROGALSKI M. e NODARI R.O. Phylogeography of plastid DNA sequences suggests postglacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. *Sci Rep* 9:2752. 2019.

OLIVEIRA, E.J., PÁDUA, J.G., ZUCCHI, M.I., VENCOVSKY, R. & VIEIRA, M.L.C. 2006. Origin, evolution and genome distribution of microsatellites. **Genetics and Molecular Biology** 29: 294- 307. 2006.

PEAKALL, R., GILMORE, S., KEYS, W., MORGANTE, M. & RAFALSKI, A. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. **Molecular Biology and Evolution** 15: 1275-1287. 1998.

PEAKALL, R.; SMOUSE P.E. **GenAlEx 6.5**: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28, 2537-2539, 2012.

PETIT RJ, PINEAU E, DEMESURE B, BACILIERI R, DUCOUSSO A. et al. 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. **Proceedings National Academy of Sciences** 94: 9996–10001.

ROSS-DAVIS, AMY & WOESTE, KEITH. Microsatellite markers for *Juglans cinerea* L. and their utility in other Juglandaceae species. **Conservation Genetics**. 9. 10.1007/s10592-007-9337-8. 2008.

SAHA, M.C., COOPER, J.D., MIAN, M.A.R., CHEKHOVSKIY, K. & MAY, G.D. 2006. Tall fescue genomic SSR markers: development and transferability across multiple grass species. **Theoretical and Applied Genetics** 113: 1449-1458.

VARSHNEY, R. K; THIEL, T.; STEIN, N.; LANGRIDGE, P.; GRANER, A. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. **Cellular And Molecular Biology Letters**, v. 7, n. 2, p. 537-546, 2002.

VARSHNEY, R.K.; GRANER, A.; SORRELLS, M.E. Genic microsatellite markers in plants: features and applications. **TRENDS in Biotechnology**, v. 23, n. 1, p. 48-55, 2005.

VIEIRA, M. L. C.; SANTINI, L.; DINIZ, A. L.; MUNHOZ, C. DE F. Microsatellite markers: What they mean and why they are so useful. **Genetics and Molecular Biology**, v. 39, n. 3, p. 312–328, 2016.

YU, K., PARK, S.J., POYSA, V. (1999). **Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*)**. *Genome*, 42: 27-34.

XIANG XG, WANG W, LI RQ, LIN L, LIU Y, ZHOU ZK, LI ZY, CHEN ZD. Large-scale phylogenetic analyses reveal fagalean diversification promoted by the interplay of diaspores and environments in the Paleogene. **Perspectives in Plant Ecology, Evolution and**

**Systematics**. 16 (3): 101–110. doi:10.1016/j.ppees.2014.03.001. 2014.

WANG, M. L., BARKLEY N.A. e JENKINS M. T.. “Microsatellite Markers in Plants and Insects. Part I: **Applications of Biotechnology**.”2009.

WANG, PENG-FEI & LI, YONG & QIAN, ZHI-HAO & LI, JIA-XIN & GE, XUE-JUN. Isolation and characterization of microsatellite loci from *Pterocarya stenoptera* (Juglandaceae). **Applications in Plant Sciences**. 6. 10.1002/aps3.1205. 2018.

WEBER, J. L. **Informativeness of human (dC- dA)n.(dG- dT)n polymorphisms**. Genomics, 7: 524-530. 1990.

WHEELER, G. L. et al. A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. **Applications in plant sciences**, v. 2, n. 12, p. 1400059, 2014.

**CAPÍTULO 3. Análise da divergência genética entre cultivares de *Carya illinoensis*  
(Juglandaceae) através de marcadores microssatélite plastidiais e nucleares**

## RESUMO

A *Carya illinoensis* (nogueira-pecã) é uma espécie frutífera florestal, nativa dos Estados Unidos. O interesse pelos seus frutos tem apresentado um crescimento exponencial no Brasil e outros países, principalmente por seu elevado valor nutricional. Além disso, é uma alternativa de renda para agricultores familiares, haja vista sua potencialidade de uso em áreas de reserva legal, tornando-se uma espécie de interesse econômico e social, principalmente no sul do Brasil. Entretanto, a espécie é carente em informações genéticas e, em decorrência da falta de controle de polinização e de registro de coleta de propágulos nos pomares, os materiais genéticos em cultivo, muitas vezes não têm origem conhecida. A caracterização molecular a partir de marcadores de DNA pode solucionar este problema, identificando e diferenciando cultivares, e propiciando, simultaneamente, as bases para o planejamento de um eventual programa de melhoramento genético. Além disso, contribuiria, também para a melhoria do sistema de manejo dos pomares, com repercussão direta na produtividade. O presente trabalho objetivou avaliar a divergência genética entre cultivares de *C. illinoensis* em desenvolvimento em condições de campo, por meio de marcadores SSR plastidiais e nucleares. Para tanto, foi realizado o isolamento do DNA genômico em folhas coletadas em 17 amostras (13 cultivares) de Anta Gorda, Don Feliciano e Estrela, no Rio Grande do Sul. Foram testados 10 marcadores microssatélites nucleares (SSR) e 13 plastidiais (cpSSR), sendo 10 prospectados do genoma plastidial de *C. illinoensis* cv. Imperial. Através da análise de coordenadas principais (PcoA) dos nSSR, os dois componentes explicaram, no total, 21,39% da variabilidade total observada, enquanto para os cpSSR o total de 84,65%. Os 17 indivíduos avaliados apresentaram uma grande amplitude de variação em relação à correlação cofenética, mesmo o valor cofenético sendo alto (cpSSR= 0,9931 e nSSR= 0,9791) os valores de *bootstrap* nos agrupamentos maiores foram baixos. O PCoA demonstrou coerência com as estimativas de similaridade obtidas e as análises sugerem que há correlação positiva entre os indivíduos de *C. illinoensis* estudados. Os resultados obtidos neste estudo evidenciaram alta diversidade genética tanto para marcadores nucleares ( $I= 2,365$ ) e plastidiais ( $I= 2,183$ ), esta estimativa de diversidade genética é bem significativa, e este resultado pode ser utilizado para determinar o padrão genético das cultivares analisadas.

**Palavras-chave:** Nogueira-pecã, marcadores SSR, diversidade genética.

## 1 INTRODUÇÃO

A noz-pecã [*Carya illinoensis* (Wangenh.) K. Koch, Juglandaceae] é uma espécie de frutífera nativa dos Estados Unidos e do México (GRAUKE et al., 2016). Dentro de sua faixa de ocorrência natural, o clima varia de ameno a invernos severos, e de condições muito úmidas a semiáridas, conferindo grande variação entre genótipos de diferentes regiões climáticas (SPARKS, 2005). Essa distribuição propicia maior adaptabilidade às cultivares, permitindo que cresça fora de sua área de ocorrência natural, sobretudo em regiões de clima temperado e subtropical. A noqueira-pecã difundiu-se para várias localizações e atualmente é cultivada em vários países como Uruguai, Argentina, Brasil, Chile, Peru, China, Índia, Egito, África do Sul e Austrália (DE OLIVEIRA et al., 2021).

Os Estados Unidos lideram a produção global de pecã, representando aproximadamente 53% do total, com 200,3 mil toneladas produzidas (em casca) na safra 2019 (INC, 2021). O México é o segundo maior produtor do mundo, com 150.000 toneladas produzidas em 2019 (INC, 2021). A China, por sua vez, é o maior importador mundial de nozes, exigindo mais de um terço da produção de nozes dos EUA (ZHANG et al., 2015).

A pecanicultura está em expansão e é considerada uma alternativa atrativa para diversificação da atividade econômica para propriedades rurais (FILIPPIN, 2011). Além disso, existem aspectos agronômicos favoráveis como a possibilidade de inserção de sistemas consorciados com pastagens e pecuária. E, também a possibilidade de consorciação com culturas anuais e outras espécies frutíferas devido à adaptação edafoclimática, a longevidade da produção, e o fácil armazenamento de frutas após a colheita (DUARTE et al., 2001).

Diante dessa expansão, investir em programas de melhoramento da pecã é um fator condescendente para dotar o setor produtivo de cultivares produtivas e saudáveis. No entanto, a noz-pecã tem sido considerada uma cultura com atenção científica limitada (CONNER, 2015). Devido ao longo ciclo de vida desta espécie, programas de melhoramento genético demoram um longo tempo e necessitam da ajuda de ferramentas biotecnológicas como a seleção assistida por marcadores.

Limitações quanto à falta de contribuições significativas de ferramentas moleculares para o cultivo de pecã foi compartilhada por Conner (2015). Foi evidenciado que os criadores de pecã não rejeitaram ferramentas biotecnológicas, nem que os marcadores moleculares são ineficientes. Conner (2015) conclui que algumas barreiras existem, e intrigam a aplicação de marcadores moleculares para ajudar a resolver as necessidades da pecanicultura.

O número de estudos genéticos em noqueira-pecã empregando marcadores moleculares baseados em DNA tem aumentado, com o objetivo de caracterizar coleções e pomares de germoplasma, acessando padrões de diversidade genética, e mais recentemente, caracterizando traços genômicos. Esses estudos, por sua vez, demonstram vários aspectos da diversidade genética e reprodução de noz-pecã, tanto em sua distribuição nativa, bem como nos países onde foi introduzido (DE OLIVEIRA et al. 2021).

Microssatélites são marcadores ideais para análises de diversidade genética em genética da conservação (POWELL et al., 1995, RAJORA et al., 2000), construção de mapas genéticos de alta resolução, identificação de *locus* controladores de características de interesse e seleção assistida por marcadores (DEVEY et al., 1996). *Locus* específicos também são usados frequentemente para o *fingerprinting* de DNA (HAYDEN et al., 2001).

Os microssatélites plastidiais, também conhecidos como *chloroplast simple sequence repeats* ou pela sigla “cpSSR”, foram primeiramente desenvolvidos por Powell et al. (1995) e são geralmente repetições curtas em tandem de mononucleotídeos. Caracterizam-se como importante fonte de informações na biologia vegetal, principalmente para estudos a níveis populacionais. O polimorfismo dos *locus* microssatélite plastidiais varia entre as espécies, elevando a variabilidade intraespecífica destes marcadores cpSSR (EBERT e PEAKALL, 2009), mas normalmente apresentam de 8 a 15 repetições de motivos mononucleotídeos.

Os plastídios (*loci* cpSSR) são herdados de apenas um dos genitores (geralmente a árvore-mãe nas angiospermas), enquanto o genoma nuclear (*loci* nSSR) é herdado tanto do pólen quanto do óvulo genitores. Além disso, os genomas de plastídios carecem de recombinação e apresentam uma menor taxa de mutação em comparação com genomas nucleares (STEFENON et al., 2019).

Marcadores cpSSR e nSSR são ferramentas muito importantes para delinear estratégias efetivas de conservação e manejo de espécies arbóreas tropicais (TARAZI et al., 2009) proporcionando à espécie e ao ecossistema a expressão de seu potencial evolutivo. Os marcadores cpSSR foram indicados como ferramentas úteis para iniciativas de identificação e conservação de uma amostra representativa de germoplasma de noz-pecã em coleções *ex situ* (GRAUKE et al., 2011).

Estudos relacionados à caracterização genética de cultivares de *Carya illinoensis* utilizando marcadores moleculares podem auxiliar tanto na conservação da variabilidade genética como na melhoria das técnicas de manejo da produção dessa espécie. E, podem constituir um subsídio importante para técnicos e produtores, fornecendo informações precisas acerca das cultivares e, como consequência, aumentando a qualidade do produto a ser

comercializado e, desta forma, sua rentabilidade. Este estudo tem por objetivo avaliar a diversidade genética existente entre as cultivares de *Carya illinoensis* por meio de marcadores microssatélite plastidiais (cpSSR) e nucleares (nSSR). Para isso, foram utilizados marcadores SSR plastidiais prospectados do plastoma de *C. illinoensis* cv Imperial, marcadores plastidiais universais e marcadores microssatélite nucleares já desenvolvidos para a espécie.

## 2 MATERIAL E MÉTODOS

O DNA de 17 amostras (13 cultivares) de *C. illinoensis* foram cedidas pelo engenheiro florestal Tales Poletto, no âmbito do projeto “Caracterização genético-molecular de cultivares de noqueira-pecã do Brasil” (Tabela 1).

**Tabela 1.** Número da amostra e as referentes cultivares de *C. illinoensis* para cada numeração.

Amostra	Cultivar
1	Moneymaker
2	Barton
3	Desirable
4	Desirable
5	Jackson
6	Farley
7	Imperial
8	Imperial
9	Imperial
10	Sucess
11	Barton
12	Chickasaw
13	Shawnee
14	Stuart
15	Importada
16	Choctaw
17	Mahan

### 2.1 Material vegetal

As amostras foliares coletadas que se encontraram em desenvolvimento em condições de campo nos municípios de Anta Gorda, Don Feliciano e Estrela, no Rio Grande do Sul,

foram acondicionadas em tubos de ensaio contendo PVP-40 2% (polivinilpirrolidona), posteriormente, encaminhadas ao Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal (LFDGV) da Universidade Federal de Santa Catarina (UFSC), para a realização dos procedimentos e análises genéticas. As amostras foliares foram maceradas e a extração do DNA total foi realizada por meio do método CTAB (DOYLE e DOYLE, 1987), com modificações. A quantidade e a qualidade de DNA extraído foram verificadas por meio de eletroforese horizontal em gel de agarose (concentração de 2,0%), corados com GelRed (Biotium, Hayward, California, USA) e contendo um padrão Ladder 100pb (Ludwig). A quantidade e a qualidade de DNA extraído também foram verificadas em espectrofotômetro NanoDrop®1000 (Thermo Fisher Scientific, Wilmington, DE, USA).

## 2.2 Amplificação e Genotipagem

Foram utilizados 13 marcadores microssatélite plastidiais, sendo 10 prospectados do genoma plastidial de *C. illinoensis* cv. Imperial (capítulo 2), além dos marcadores desenvolvidos para o genoma plastidial de tabaco *ccmp2* (WEISING e GARDNER, 1999), *ntcp9* e *ntcp40* (BRYAN et al., 1999) (Tabela 2). Além desses, foram utilizados 10 marcadores SSR nucleares desenvolvidos para *C. illinoensis* por Grauke et al. (2003) e por Grusak et al. (2007) (Tabela 3).

Para as reações de amplificação via PCR para todos os marcadores foram realizadas para um volume final de 15 µL, contendo 2,0 µl de DNA, 0,9 µl de dNTP (Sinapse Inc, Hollywood, Florida, USA), 0,2 µl de Taq Polimerase DNA (Ludwig), 1,5 µl tampão 10X (Ludwig), 1,20 µL de BSA, 0,75 µl de MgCl<sub>2</sub> – free reaction buffer (Ludwig), 1,0 µM de iniciador forward, 1,20 µM de iniciador reverse. A programação no Termociclador Veriti® foi realizada da seguinte maneira: 95°C por 3 min; seguido de 30 ciclos de 95°C por 15 s, temperatura de 60° de anelamento por 30 s, 72°C por 30 s; seguido de oito ciclos de 94°C por 30s, temperatura de 53°C por 45 s, 72°C por 45s; e a etapa de extensão final a 72°C por 10 min, sendo mantidos a 4°C.

A caracterização dos genótipos foi realizada por meio de eletroforese em gel de agarose 2,5%. Os géis foram corados com GelRed e visualizados em transiluminador de luz UV.

**Tabela 2.** Loci microssatélite prospectados e universais testados nas variedades de *Carya illinoensis*.

<i>Locus</i>	Primer sequence (5'->3')	RepeatMotif	Product Size
<i>cpCil1</i> <sup>A</sup>	GATCGGATTATATTTGTTTG TAAATGAAGTAGTTGGGTCT	(AAT) <sub>4</sub>	280
<i>cpCil2</i> <sup>A</sup>	AGGAGATTCCCATATCTATC AATAAACTACGCAGAAAATG	(TA) <sub>10</sub>	237
<i>cpCil4</i> <sup>A</sup>	AAATAAAGTAAGAAGGGCTC GACGAAATCAAAGAGGTACT	(AT) <sub>6</sub> -(AT) <sub>7</sub>	244
<i>cpCil5</i> <sup>A</sup>	TGTACCCTCTTCAAATTCTA AATAACCGATTAGTCCTTCT	(ATT) <sub>6</sub>	267
<i>cpCil6</i> <sup>A</sup>	CCTGAATATGATTCTACCAA GTGAAAAGGAGAACACATAG	(AATA) <sub>3</sub>	190
<i>cpCil7</i> <sup>A</sup>	TCTTATCAATAAAAATTTCCG GGAATAGTTCCTTTCATTCT	(ATT) <sub>5</sub>	270
<i>cpCil8</i> <sup>A</sup>	AACGTTTTTATTGTTGTGTT TTTCGTAAAATCTCGTCTAC	(TTA) <sub>4</sub>	210
<i>cpCil9</i> <sup>A</sup>	ATTACTCAAAAGAATCGACA ATTTGTTTATTCCTTATCCC	(AT) <sub>6</sub>	257
<i>cpCil10</i> <sup>A</sup>	CTTTCTGGATGTAGATGATG CGACGTAATTCATAGAGTC	(TA) <sub>6</sub>	239
<i>cpCil12</i> <sup>A</sup>	CGACGTAATTCATAGAGTC CTTTCTGGATGTAGATGATG	(AT) <sub>6</sub>	239
<i>ccmp2</i> <sup>B</sup>	GATCCCGGACGTAATCCTG ATCGTACCGAGGGTTCGAAT	(A) <sub>11</sub>	207 – 210
<i>ntcp9</i> <sup>C</sup>	CTTCCAAGCTAACGATGC CTGTCCTATCCATTAGACAATG	(A) <sub>14</sub>	348 – 379
<i>ntcp40</i> <sup>C</sup>	GATGTAGCCAAGTGGATCA TAATTTGATTCTTCGTCGC	(T) <sub>10</sub>	185 – 209

<sup>A</sup> Prospectados e validados neste estudo (capítulo 2), <sup>B</sup> Weising e Gardner, 1999, <sup>C</sup> Bryan et al., 1999

**Tabela 3.** *Loci* microssatélite nucleares espécie-específicos para *Carya illinoensis*.

<i>Locus</i>	Primer sequence (5'->3')	RepeatMotif	Product size
<i>pm-ca10</i> <sup>A</sup>	GCAAATCAACCCTGTAGCATAC GCTCAGACATGCAAACGTACC	(TG) <sub>2</sub> (TG) <sub>4</sub>	115
<i>pm-ga38</i> <sup>A</sup>	CCAATGTTTCTTCGAAAGGACA GTAAAGCCTACAACCTACAACAGTCTATG	(CT) <sub>11</sub>	109 – 125
<i>pm-ga39</i> <sup>A</sup>	TGTAAATGCGTGCTATTGCTGAT GAATAGACAAAGAAACGAAACTCATTGA	(CT) <sub>9</sub>	86 – 106
<i>pm-ga41</i> <sup>A</sup>	GGCGAAACCGTAAACCTG AACACCAGCACCCAAATC	(CT) <sub>9</sub>	104
<i>pm-cin4</i> <sup>A</sup>	GGCATCAGAGAAGGCTCCT CTCACCCGTCTCTAGGGCTA	(CTT) <sub>12</sub>	93 – 141
<i>pm-cin13</i> <sup>A</sup>	CAAAAAGCATCAAAGCCATC ACAAATTCCTCACTCCGGAG	(GAA) <sub>8</sub>	132-135
<i>pm-cin20</i> <sup>A</sup>	G TTCACGACAGATGAGTGCGA GCTGGTACACTTCTGCTTCTTCT	(GAA) <sub>4</sub>	138
<i>pm-cin22</i> <sup>A</sup>	GGTTATGAAATTTGATATTACTTTTGG CCAACAAGGGAAGCCAACCTT	(CTT) <sub>10</sub>	114 – 120
<i>pm-cin23</i> <sup>A</sup>	GGAGTTGTGGAAGCAGTGGA GGACCATAAGAGTTTTGACCCTT	(GAT) <sub>6</sub>	80 – 83
<i>1-A05</i> <sup>B</sup>	AAAGGTGTCTGTGGAAAGG GTCAATCATCCCAATCAAA	(GTTTT) <sub>5</sub>	160 – 165

<sup>A</sup> Grauke et al. (2003), <sup>B</sup> Grauke et al. (2012).

### 2.3 Análise dos dados

O *software* TotalLab TL120 1D v2009 (Nonlinear Dynamics Ltd., Newcastle, Inglaterra) foi usado para análise das bandas no gel de agarose 2,5%. Primeiramente, fez-se a seleção das linhas/colunas, definindo-se o local de cada poço em que foi pipetada a amostra, após fez-se a detecção das bandas manualmente e, em seguida, a calibração do marcador de peso molecular 100 pb. Posteriormente, os tamanhos dos produtos amplificados (em pares de bases) foram fornecidos, por coluna, sendo que, a partir de então, o programa calcula uma série de algoritmos para se verificar os picos de cada banda/coluna no gel, assim como os limites das bandas, sendo que esses limites podem ser ajustados manualmente. Ao lado da imagem do gel, o programa fornece um gráfico com a posição da intensidade (em pixels) da

banda. Em seguida, foi selecionada cada coluna, sendo que o programa disponibiliza uma tabela informando os valores dos pares de bases.

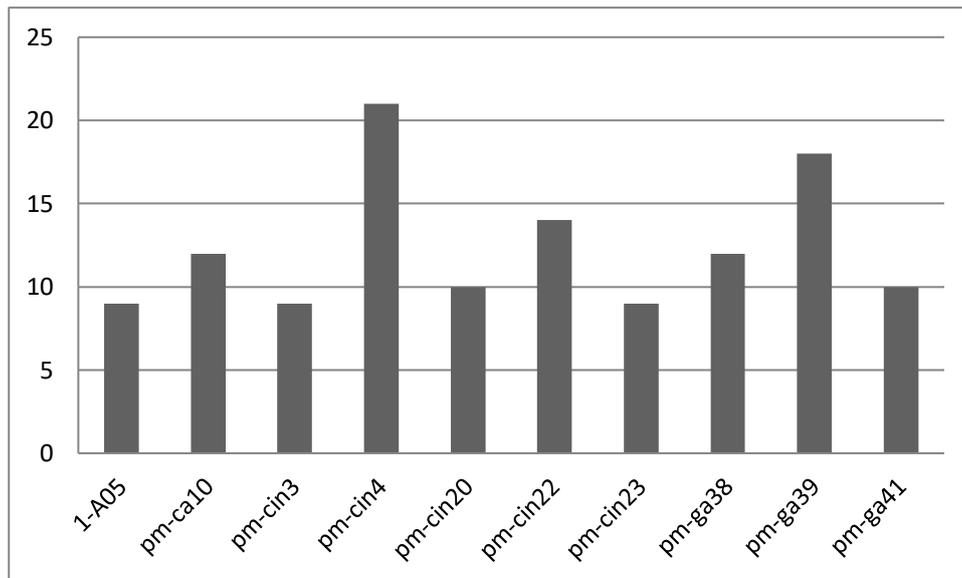
Os alelos exclusivos foram identificados utilizando o programa GenAlEx versão 6.5 (PEAKALL; SMOUSE, 2012) e o número de alelos raros foi calculado com auxílio do Excel, onde foram considerados alelos raros aqueles cuja frequência alélica foi menor que 0,05 para cada cultivar. A partir da análise das bandas, foram estimadas as frequências alélicas e os índices de diversidade genética, o número de alelos ( $A$ ), o número efetivo de alelos ( $A_e$ ), a riqueza alélica ( $R_s$ ), a similaridade genética detectada pelos marcadores microssatélite foi explorada com Análises de Coordenadas Principais (PCoA) e a mensuração da diversidade genética molecular foi usado o Índice de Diversidade de Shannon (SHANNON, 1948) por meio do programa GenAlEx versão 6.5 (PEAKALL e SMOUSE, 2012).

Para os marcadores SSR plastidiais, foram estimadas as frequências alélicas, o número de alelos ( $A$ ), o Índice de Shannon ( $I$ ) e a diversidade de haplótipos ( $h$ ), diversidade não-viesada ( $uh$ ), e PCoA por meio do programa GenAlEx versão 6.5, considerando as análises para dado haplóide (PEAKALL; SMOUSE, 2012).

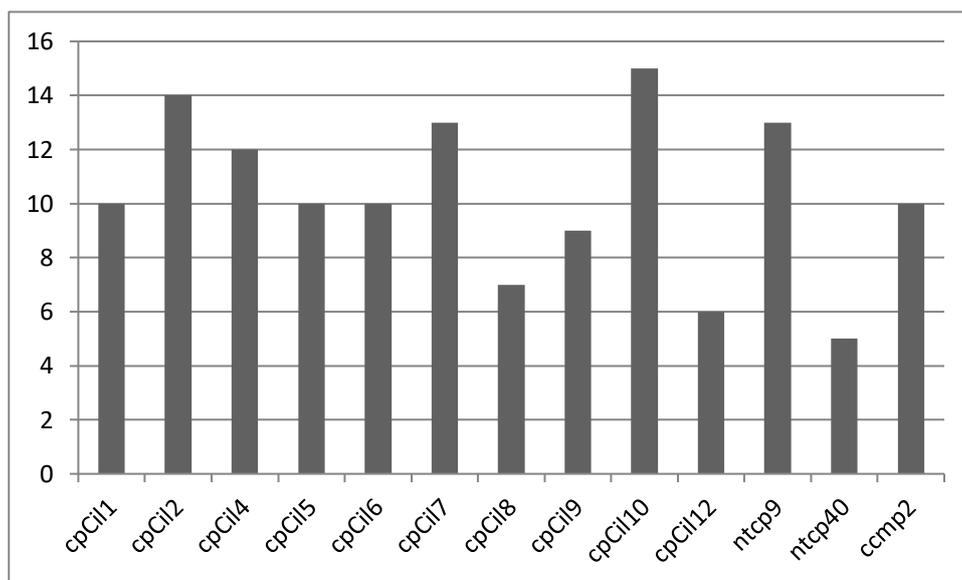
O software Past 4.03 (HAMMER et al., 2001) foi utilizado para elaborar um dendrograma baseado na distância Euclidiana e algoritmo UPGMA.

### 3 RESULTADOS E DISCUSSÃO

Em relação à estimativa da diversidade genética, as frequências alélicas obtidas com a utilização dos 10 *loci* microssatélite nucleares e dos 13 *loci* plastidiais estão apresentadas nas figuras 1 e 2. O *locus pm-cin4*, com 21 alelos, foi o que apresentou o maior número de alelos dentre os microssatélites nucleares. E os *loci* 1-A05, *pm-cin3* e *pm-cin23* apresentaram o menor número de alelos, no total de 9 cada (Figura 1). Quanto aos cpSSR, o *locus cpCil10* apresentou o maior número de alelos (15) e, o *locus ntcp40* foi o com menor número, apresentando apenas cinco alelos (Figura 2). Considerando o total de 23 *loci* microssatélite em 17 amostras (13 cultivares), foram observados 104 alelos plastidiais e 79 nucleares diferentes.



**Figura 1.** Número de alelos por *locus* para cada um dos 10 nSSR analisados nas cultivares de *C. illinoensis*.



**Figura 2.** Número de alelos por *locus* para cada um dos 13 cpSSR analisados nas cultivares de *C. illinoensis*.

Foram identificados 123 alelos exclusivos, sendo 45 nucleares e 78 plastidiais. No total foram identificados 34 alelos raros (com frequência alélica menor que 0,05), presença de alelos raros, que também indica a perda da diversidade genética (KAGEYAMA et al., 1998) foi observada utilizando os SSR nucleares (*pm-cin4*, *pm-ga39*, *pm-cin22* e *1-A05*). Com o uso dos marcadores plastidiais, não houve alelos raros.

A genotipagem realizada com as cultivares de *C. illinoensis*, utilizando 23 *loci* polimórficos obteve uma média de 12,4 alelos por locos SSR nucleares e 10,3 alelos por *locus*

SSR plastidiais (Tabela 4 e 5). A porcentagem de *locus* polimórficos foi de 100% considerando todas as cultivares de *C. illinoensis*, tanto para nSSR quanto cpSSR. Os altos níveis de polimorfismo em regiões microssatélite são naturais, devido à sua elevada taxa de mutação, que altera o número de repetições (ZANELLA et al., 2017). Essa alta taxa de mutação ocorre devido ao pareamento não-homólogo e/ou ao processo de *slippage* durante a replicação (GAO et al., 2013). A grande quantidade de alelos encontrados em cultivares de *C. illinoensis* também pode estar relacionada com o tamanho das regiões microssatélite analisadas nesses *locus*. Geralmente os *locus* que apresentam maior número de repetições tendem a apresentar maior número de alelos (WEBER, 1990), sugerindo-se que as regiões microssatélite mais longas estão sujeitas a maiores taxas de mutações, em relação a regiões de menor tamanho (ZANELLA et al., 2017).

Os resultados obtidos não evidenciaram ligação entre os *locus*, portanto os 10 SSR nucleares foram utilizados para estimar o número de alelos, a heterozigosidade observada ( $H_O$ ) e esperada ( $H_E$ ) para cada um dos *locus* analisados nas cultivares (Tabela 5). Ao analisar a  $H_O$  por marcador SSR nuclear, os valores apresentaram grande variação entre as cultivares, sendo que o maior valor observado foi obtido com o *locus pm-ga39* ( $H_O = 0,846$ ). A  $H_E$  também apresentou grande variação, onde o maior valor esperado foi observado no *locus pm-ga39* ( $H_E = 0,935$ ). O valor de  $H_E$  pode ser utilizado para estimar a diversidade genética nas cultivares da espécie, sendo considerado excelente, especialmente levando em conta o elevado número de alelos por *locus*. Esses valores demonstram a ampla variabilidade genética encontrada nas cultivares de *C. illinoensis*. Em um estudo de isolamento e caracterização de 20 marcadores microssatélite polimórficos para *Juglans mandshurica* (Juglandaceae), o número de alelos variou de um a 17 (média = 6), enquanto os níveis de heterozigosidade esperados e observados foram 0 a 0,925 e 0 a 1,000 (média = 0,358 e 0,627), respectivamente. O grande número de alelos por *locus* indicou a potencial utilidade desses marcadores para caracterizar a estrutura genética populacional da espécie (CHEN et al., 2013).

A média do índice de Shannon foi  $I = 2,365$  com desvio de padrão de 0,097. O índice de Shannon é um índice de diversidade genética em que, quanto menor o valor, menor o grau de incerteza e, portanto, a diversidade da amostra é baixa. A diversidade tende a ser mais alta quanto maior o valor do índice (LEWONTIN, 1972). A análise de marcadores microssatélite nucleares revelou um nível alto de diversidade genética nas cultivares estudadas, evidenciando a importância destes para o desenvolvimento de programas de melhoramento genético participativo desta espécie.

Fan et al. (2013) trabalhando com a espécie *Cyclocarya paliurus* (Juglandaceae)

observaram que o número de alelos por *locus* variou de dois a oito, com média de 3,3. A heterozigidade esperada e observada variou de 0,153 a 0,802 e de 0 a 0,750, respectivamente, dentre os 28 marcadores microssatélite desenvolvidos neste estudo observou-se a eficácia para a detecção da diversidade genética em *C. paliurus* e elucidação da dinâmica populacional da espécie. A diversidade genética é matéria prima para a evolução adaptativa das espécies. O estudo desta diversidade e sua distribuição dentro e entre populações naturais e seus parentes cultivados é fundamental para gerar programas de conservação e exploração de espécies com importância econômica e/ou em perigo de extinção (VARSHNEY *et al.* 2005).

**Tabela 4.** Número de alelos por *locus* (*A*), Número efetivo de alelos por *locus* (*A<sub>e</sub>*), Índice de Shannon (*I*), heterozigidade observada (*H<sub>O</sub>*) e esperada (*H<sub>E</sub>*), Índice de fixação (*F*), para os 10 *locus* microssatélite nucleares em 11 cultivares de *C. illinoensis*.

<i>Carya illinoensis</i>							
<i>Locus</i>	<i>N</i>	<i>A</i>	<i>A<sub>e</sub></i>	<i>I</i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>F</i>
<i>I-A05</i>	11	9	7,1	2,071	0,182	0,860	0,788
<i>pm-ca10</i>	15	12	10,7	2,431	0,000	0,907	1,000
<i>pm-cin3</i>	13	9	8,0	2,138	0,000	0,876	1,000
<i>pm-cin4</i>	15	21	15,0	2,893	0,733	0,933	1,000
<i>pm-cin20</i>	15	10	7,2	2,154	0,000	0,862	0,214
<i>pm-cin22</i>	12	14	11,0	2,521	0,833	0,910	0,084
<i>pm-cin23</i>	14	9	7,0	2,069	0,000	0,857	1,000
<i>pm-ga38</i>	14	12	10,9	2,441	0,000	0,908	0,095
<i>pm-ga39</i>	13	18	15,4	2,811	0,846	0,935	1,000
<i>pm-ga41</i>	15	10	6,8	2,119	0,000	0,853	1,000
<b>Média</b>	13,7	12,4	10	2,365	0,259	0,890	0,718
<b>Desvio Padrão</b>	0,448	1,310	1,027	0,097	0,121	0,010	0,130

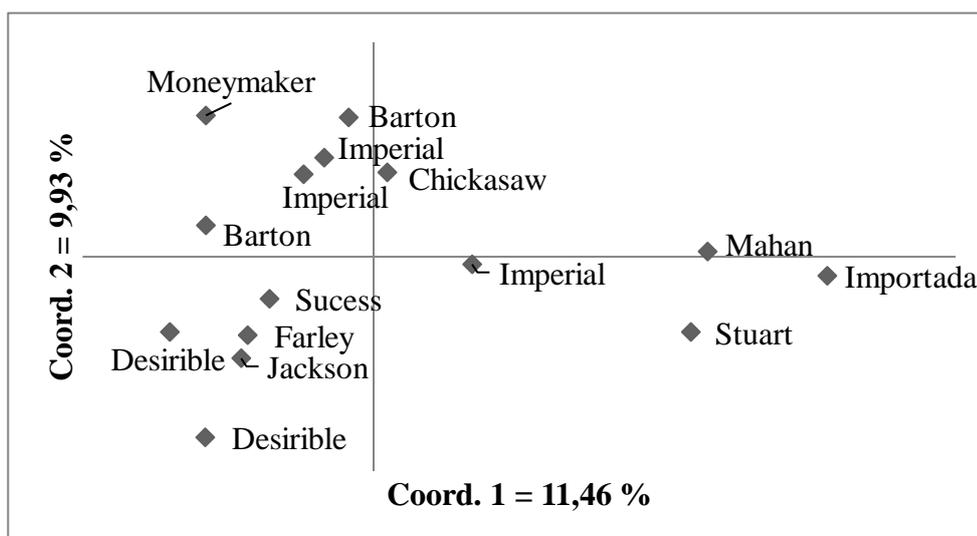
*Nota:* *N* = tamanho amostral;

Com base nos índices de heterozigidade observada e esperada, foi calculado o índice de fixação (*F*) (Tabela 5). Este índice variou de 0,214 a 1,000 com média de 0,718, com um intervalo de confiança de 95%. Estimativas de índices de fixação maiores que zero indicam a ocorrência de endogamia entre indivíduos. Entretanto, índices próximos ou menores do que zero indicam a ocorrência de cruzamento entre indivíduos não aparentados (BARROS, 2009).

Através da análise de coordenadas principais (PCoA) dos nSSR, a partir da matriz de distâncias genéticas, é possível observar a clara separação entre as 15 amostras (11 cultivares)

(Figura 3). Dessa forma, os marcadores selecionados foram eficientes em discriminar as cultivares nessas amostras. A Coordenada Principal (CP) 1 foi responsável pelo máximo da variabilidade revelada pelos marcadores SSR nucleares (11,46%), enquanto o CP 2, por 9,93%, dessa variância, que não está correlacionada àquela do CP 1 anterior. Os dois componentes explicaram, no total, 21,39% da variabilidade total observada. Essa variância foi baixa para marcadores moleculares que geralmente explicam por volta de 30-40% da variabilidade dos dados. Isto quer dizer que apenas 21,39% da variância total foi vista em duas dimensões. Esse total é reduzido, e explica uma parte pequena da variância total, haja vista que pelo menos 70% da variância total devem ser explicadas pelos dois primeiros componentes principais (RENCHER, 2002).

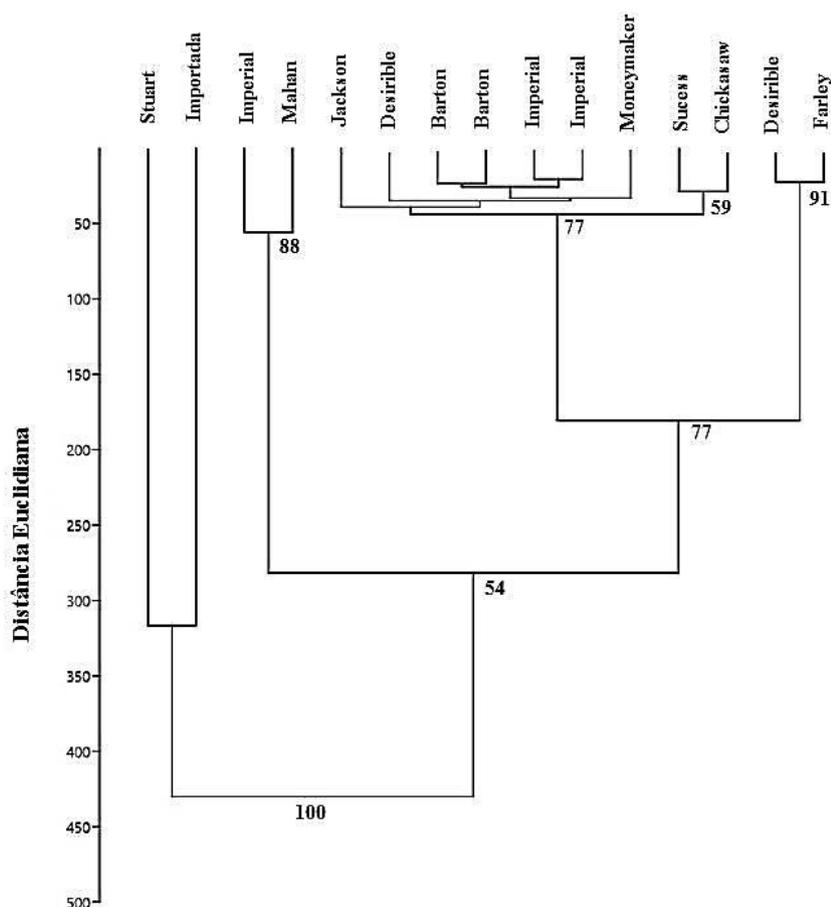
Observa-se, também, na Figura 3, a formação de três grupos: o primeiro, de maior similaridade genética, formado por três indivíduos (Mahan, Importada e Stuart), os quais estão em desenvolvimento na mesma propriedade de Anta Gorda e podem ter uma origem comum, com exceção da cultivar Stuart, localizada em Estrela. O segundo, com menor similaridade genética, constituído por seis outros indivíduos (uma Moneymaker, duas Barton, duas Imperial e uma Chickasaw), os quais encontram-se localizados em três localidades diferentes, Anta Gorda (Moneymaker e Barton), Don Feliciano (Imperial) e Estrela (Chickasaw), sendo que, na qual se encontram as cultivares Moneymaker e Chickasaw é a mesma em que estão os três membros do primeiro grupo, aqueles que são mais similares entre si. As cinco cultivares do terceiro grupo encontram-se em uma única propriedade, Don Feliciano. A cultivar Imperial (o qual se encontra localizado na propriedade em Don Feliciano), ficou isolado entre os três grupos, apresentando similaridade intermediária. Estudos de análises morfométricas, químicas e genéticas, foram feitas com noz-pecã, e os dados morfométricos apresentaram 98,88% de variação total, contrastando com os dados das análises químicas que revelaram uma homogeneidade relativa entre os acessos avaliados. Já os dados genéticos mostraram uma distribuição de grupos distintos entre os acessos, porém não foi observado relação entre esses grupos com os dados morfológicos ou químicos, embora os marcadores AFLP possam estar ligados a genes de interesse agrônomo (POLETTO et al. 2020). Um estudo feito com o fruto *Bertholletia excelsa* (castanha do Brasil), foi observada expressiva diferença na largura da semente, no comprimento dos frutos e em sua massa seca. Por outro lado, os marcadores moleculares não mostraram uma diferenciação genética entre as cultivares analisadas, e mesmo assim o resultado melhorou a forma de avaliar o rendimento da produção (SUJII et al, 2013).



**Figura 3.** Análise de coordenadas principais das cultivares de Nogueira-pecã, com base nos dados genéticos moleculares nucleares.

Dang et al. (2015) usando 12 marcadores polimórficos EST-SSR para avaliar a estrutura genética e PCoA de 96 nogueiras chinesas em onze populações mostraram que essas populações se agruparam em quatro grupos, enquanto a investigação da estrutura filogeográfica de Bai et al. (2014) relataram dois *clusters* (oeste e leste) de *J. cathayensis* usando microssatélite nuclear (nSSR). Em um estudo com *Juglans cinerea* e em outras espécies da família Juglandaceae observou-se que a análise usando o princípio de componentes e plotagem dos dois primeiros escores de componentes principais, que explicaram 41% da variância, mostraram um agrupamento claro dos alelos, mostrando que os alelos de *J. cinerea* foram localizados muito mais próximos de *J. nigra*. Corroborando que todos os táxons poderiam ser significativamente separados uns dos outros usando análise discriminante canônica, com exceção de *J. mandschurica* e *J. ailanthifolia* (dados não mostrados no estudo) (ROSS-DAVIS e WOESTE, 2007).

A partir da análise integral dos dados feita com todos os nSSR, obteve-se uma correlação cofenética de valor 0,9791 possibilitando a construção de um dendrograma (Figura 4) onde pode se observar a maior distância genética entre três grupos.



**Figura 4.** Dendrograma de UPGMA baseada na distância euclidiana de 15 cultivares de noqueira-pecã com marcadores nucleares.

O primeiro grupo contém as duas cultivares Stuart e Importada, pertencentes à mesma propriedade no Município de Anta Gorda. São duas cultivares com dicogamia do tipo protogínica (HAMANN et al., 2018). O segundo, com as cultivares Mahan e Imperial permaneceram próximas ao agrupamento das cultivares Stuart e Importada no dendrograma, se mostram mais próximas geneticamente com *bootstrap* de 88%. O terceiro grupo observado contempla a maior parte das cultivares comerciais de noqueira-pecã, com *bootstrap* de 77%. Agrupando cultivares como Desirable e Farley (Município de Don Feliciano), com *bootstrap* de 91%. O grupo subsequente, com suporte de 77%, observou-se as cultivares q Jackson, Desirable, Barton (com dicogamia do tipo protândrica) e Imperial (dicogamia protogínica) (HAMANN et al., 2018) com maior proximidade no dendrograma. Moneymaker, entre dois grupos, o citado posteriormente e com Sucess e Chickasaw que formam um grupo com maiores semelhanças genéticas com *bootstrap* de 59%. Um fator que também pode explicar as discrepâncias entre o genótipo e a identificação de cada indivíduo é a ocorrência de troca

de amostras durante o manejo das cultivares no Estado do Rio Grande do Sul ou até mesmo a perda ou troca de sua identificação original, acarretando falhas tanto no manejo quanto no controle de qualidade das cultivares comercializadas.

Foi possível identificar no dendrograma grupos genéticos bem definidos, consistentemente representados nos *bootstraps*, o que indica subpopulações bem estruturadas dentro da população das cultivares de *C. illinoensis*. A avaliação da confiabilidade dos nós obtidos, através dos resultados dos *bootstraps*, mostrou a consistência entre alguns grupos formados no dendrograma. Apenas alguns nós do dendrograma apresentaram valores de *bootstrap* maiores que 50%, estando estes localizados geralmente entre indivíduos com alta similaridade genética no final dos ramos ou nós que agrupam maior quantidade de indivíduos. O fato de amostras das mesmas cultivares serem encontradas tanto no grupo dois quanto no grupo três mostra a variabilidade intra e interespecífica contida entre as cultivares.

Estimar as distâncias genéticas entre as cultivares de *C. illinoensis* é importante, pois a partir destes dados é possível sugerir a realização de cruzamentos ou acompanhar possíveis cruzamentos potenciais, a fim de explorar a heterose (vigor híbrido) entre as progênies em futuros programas de melhoramento genético (FURLAN et al., 2007; SOUZA et al., 2010). Além disso, esta análise permite avaliar possíveis redundâncias no conjunto de indivíduos. Os cruzamentos entre indivíduos mais distantes possibilitam que haja maior segregação em futuros ciclos de melhoramento e, portanto, enriquecimento de sua base genética (SOUZA et al., 2010; CRUZ et al., 2011). Dentre os indivíduos presentes na amostra estudada com marcadores microssatélite nucleares, as cultivares Stuart e Farley são os que apresentam maior distanciamento genético.

Em um estudo em quatro populações naturais de noz-pecã em uma área do México utilizando marcadores AFLP, os padrões de diversidade genética acessados foram considerados relativamente baixos. Com 52,1% de fragmentos de DNA polimórfico (196 de 376 fragmentos), a diferenciação genética em pares de árvores variou entre 3% e 17% (CERNA-CORTÉS et al., 2003). Mesmo apresentando baixa diferenciação genética entre as árvores, e sobreposição de indivíduos de diferentes origens geográficas no dendrograma UPGMA foi associado à ocorrência de fluxo de genes através dos rios locais e por ação antropogênica (CERNA-CORTÉS et al., 2003). Além da caracterização da diversidade genética nestas populações, neste estudo seis marcadores AFLP estavam associados à dicogamia. Estes marcadores moleculares foram considerados ferramentas importantes para a previsão desta característica na fase de muda, permitindo a seleção de plantas no

estabelecimento de pomares de nozes, assegurando uma polinização completa e, conseqüentemente, melhorando o rendimento.

Com relação aos cpSSR (Tabela 6), o número de alelos por *locus* variou de cinco a 15. Utilizando 13 *loci* plastidiais, os valores de diversidade de haplótipos variaram de 0,727 (*ntcp40*) a 0,927 (*cpCil10*). Em relação à média ( $\pm$ desvio padrão) na diversidade de haplótipos,  $h = 0,866 (\pm 0,016)$ .

**Tabela 5.** Índices de diversidade para os 13 *locus* microssatélite plastidiais em 13 cultivares de *C. illinoensis*.

<i>Carya illinoensis</i>						
<i>Locus</i>	<i>N</i>	<i>A</i>	<i>A<sub>e</sub></i>	<i>I</i>	<i>h</i>	<i>Uh</i>
<i>cpCil1</i>	16	10	7,1	2,133	0,859	0,917
<i>cpCil2</i>	17	14	12,6	2,589	0,920	0,978
<i>cpCil4</i>	17	12	10,7	2,425	0,907	0,963
<i>cpCil5</i>	17	10	8,7	2,232	0,886	0,941
<i>cpCil6</i>	17	10	7,4	2,150	0,865	0,919
<i>cpCil7</i>	17	13	10,7	2,476	0,907	0,963
<i>cpCil8</i>	16	7	5,6	1,808	0,820	0,875
<i>cpCil9</i>	16	9	7,5	2,101	0,867	0,925
<i>cpCil10</i>	17	15	13,8	2,670	0,927	0,985
<i>cpCil12</i>	17	6	4,6	1,650	0,782	0,831
<i>ntcp9</i>	17	13	10,7	2,476	0,907	0,963
<i>ntcp40</i>	17	5	3,6	1,430	0,727	0,772
<i>ccmp2</i>	14	10	8,9	2,243	0,888	0,956
<b>Média</b>	16,538	10,308	8,613	2,183	0,866	0,922
<b>Desvio padrão</b>	0,243	0,850	0,841	0,103	0,016	0,017

*N*= tamanho amostral; *A*= número de alelos; *A<sub>e</sub>*= Número efetivo de alelos; *h*= diversidade de haplótipos; *uh*= diversidade não-viesada; *I* = Índice de Shannon.

Em um estudo utilizando marcadores, foi relatado um padrão equivalente de diversidade genética numa análise de marcadores SSR nucleares e plastidiais, amostras de 19 populações naturais de *C. illinoensis* em crescimento nos EUA e México (GRAUKE et al., 2011). Também foi encontrado uma diversidade genética relativamente alta ( $h = 0,516$  para três *loci* cpSSR e  $H_o = 0,461$  para 14 *loci* nSSR) mas um número moderado a elevado de alelos (número médio total de alelos  $Na = 4,33$  e número médio efetivo de alelos  $Ne = 2,18$

para *loci* cpSSR, e  $N_a = 10,21$  e  $N_e = 4,37$  para *loci* nSSR). Enquanto as estimativas da diversidade genética refletem os padrões de cruzamento e as forças evolutivas (como deriva genética, migração e seleção), o número de alelos é importante para as respostas das populações às tensões bióticas e abióticas. Além disso, um número elevado de alelos proporciona uma maior oportunidade para encontrar combinações alélicas que são de interesse para o melhoramento genético das espécies de culturas. Esta análise de cpSSR e nSSR revelou também uma estruturação geográfica das árvores amostradas em três grupos principais, com pequenas diferenças relativamente à estrutura baseada em marcadores cpSSR e nSSR (GRAUKE et al., 2011).

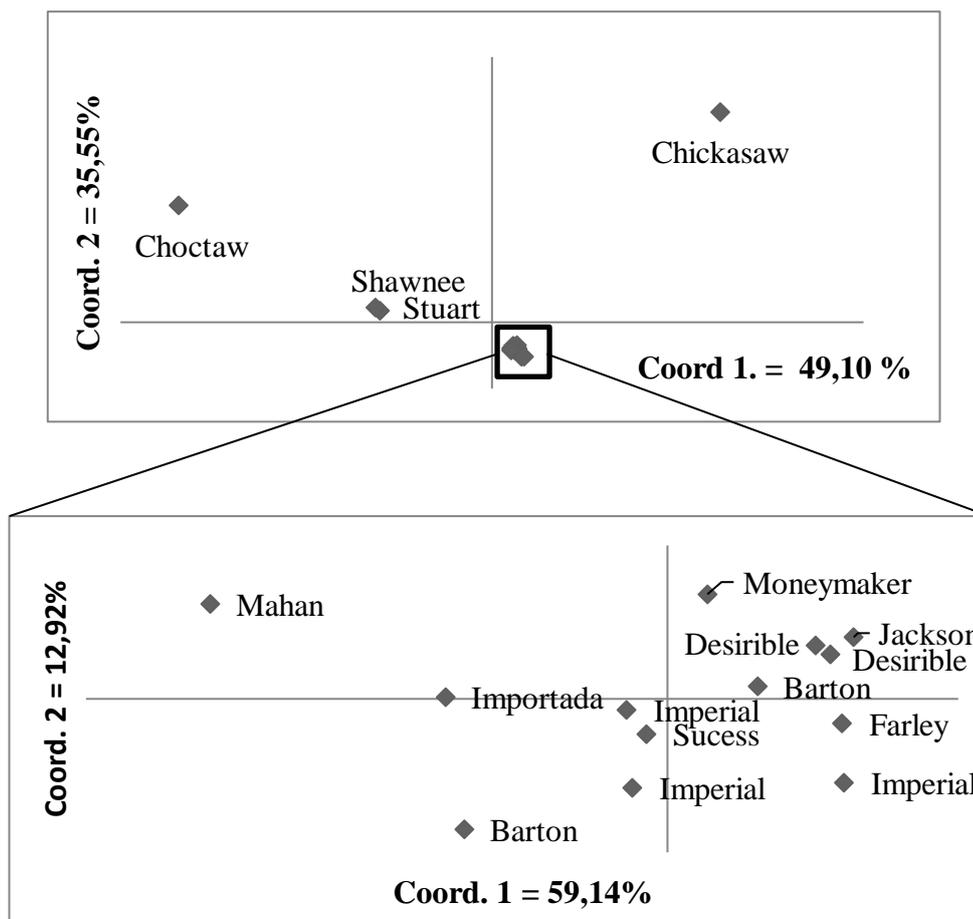
Feng et al (2018) analisando a diversidade *Juglans sigillata* (Juglandaceae; popularmente conhecida como noz de ferro), com cpSSR desenvolvidos para a espécie, observaram que o número de alelos por *locus* variou de 3 a 10. Quatro marcadores polimórficos de cloroplastos desenvolvidos neste estudo revelaram-se úteis para análise genética de populações e estudos de história evolutiva. Foi observado que, esses marcadores podem ser úteis para estudos genéticos de outras espécies de *Juglans*, onde encontraram sete haplótipos em 12 indivíduos de *Juglans cathayensis* com base nos nove *loci* cpSSR examinados no estudo.

Estudos na família Juglandaceae utilizando marcadores microsatélites plastidiais espécie-específicos são escassos. Na literatura, a maioria dos estudos baseados em marcadores plastidiais utilizam cpSSR universais para análises de diversidade genética, tanto em espécies arbóreas cultivadas quanto para as de ocorrência natural, onde encontraram um número baixo de alelos por *locus*. Steane et al. (2005) descreveram um conjunto de 35 *primers* microsatélite para análise de polimorfismos intraespecíficos, conseguindo a identificação de 10 haplótipos entre 16 indivíduos de *Eucalyptus globulus*. Também verificaram que a conservação dos *primers* quando testados em outras quatro espécies de *Eucalyptus* e sete espécies da família Myrtaceae é de 98%, comprovando a variabilidade no DNA plastidial no gênero *Eucalyptus* e na família Myrtaceae. Em espécies arbóreas de ocorrência natural no Brasil, Moreno et al. (2009), empregando marcadores cpSSR universais para estudar a estrutura genética espacial em populações de *Hymenaea stigonacarpa* (Fabaceae; popularmente conhecida como jatobá-do-cerrado) averiguaram um número alelos por *locus* variando de um a seis. No estudo de Moreno et al. (2009), os autores verificaram que uma das populações exibiu um único haplótipo, indicando que a população foi fundada por uma única linhagem materna, evidenciando o efeito fundador.

Em outra população, Moreno et al. (2009) observaram diversidade haplotípica moderada ( $h= 0,667$ ) apresentando seis haplótipos. Em outro estudo utilizando cpSSR universais em análises genéticas em uma população de *Manilkara huberi* (Sapotaceae; popularmente conhecida como maçaranduba), Azevedo et al. (2008) constataram de dois a quatro alelos por *locus* polimórficos e um alto valor de diversidade para adultos ( $h= 0,898$ ) e regenerantes ( $h= 0,986$ ).

No presente estudo, o índice de Shannon variou para os 13 *loci* plastidiais, onde foram observados valores entre 1,430 a 2,670, o maior valor está relacionado ao *locus* *cpCil10* (cpSSR espécie-específico), e o menor a *ntcp40* (cpSSR universal), com média 2,183 (Tabela 6). Dentre os marcadores prospectados o que apresentou menor índice foi *cpCil12*, com 1,650. Ao observar os *loci* plastidiais, os valores do índice de Shannon seguiram a mesma tendência da diversidade de haplótipos observada, indicando alta diversidade genética nas cultivares utilizando marcadores microssatélite plastidiais.

Para avaliar a utilidade dos marcadores microssatélite plastidiais prospectados e universais em separar as espécies de *C. illinoensis*, estimou-se a distância genética através de uma análise de coordenadas principais. A partir da matriz de distâncias genéticas, é possível observar clara separação entre as 13 cultivares (17 amostras) da espécie (Figura 5). Dessa forma, os marcadores selecionados foram eficientes em discriminar as cultivares.



**Figura 5.** Análise dos componentes principais das cultivares de noqueira-pecã, com base nos dados genéticos moleculares plastidiais.

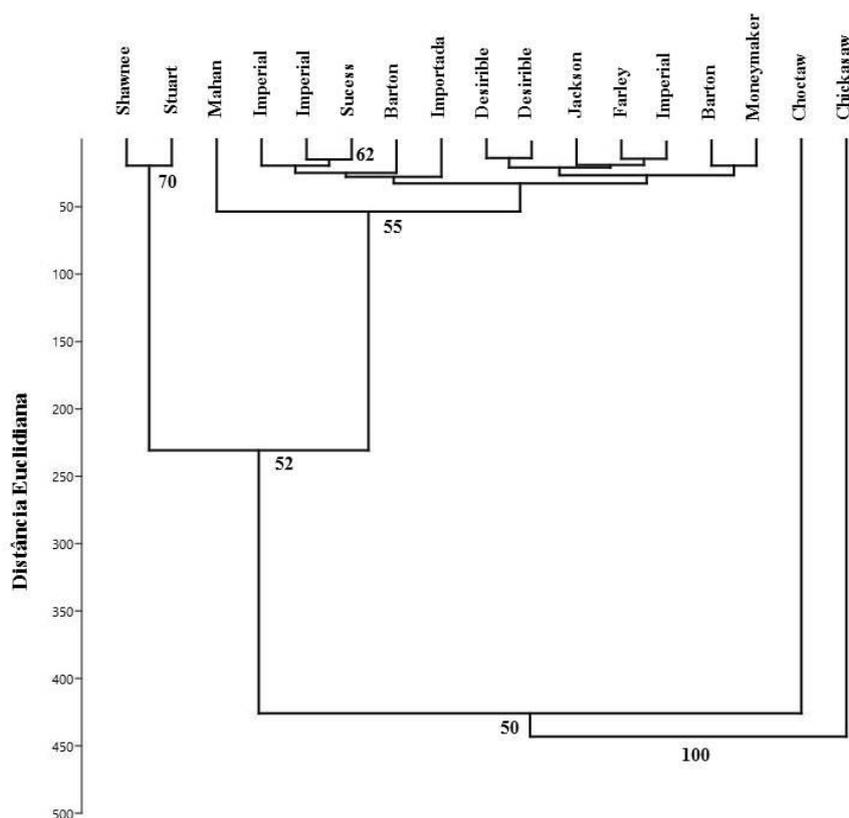
A Coordenada Principal (CP) 1 foi responsável pelo máximo da variabilidade revelada pelos marcadores SSR plastidiais (49,10%), enquanto o CP 2, por 35,55%, dessa variância. Os dois componentes ilustraram, no total, 84,65% da variabilidade total observada. Conforme Cruz et al. (2004), caso as duas primeiras variáveis canônicas permitam estimativas acima de 80% da variação total, pode-se obter uma interpretação satisfatória da variabilidade entre os acessos, tal como ocorreu neste estudo, sendo possível a representação gráfica bidimensional (Figura 4).

Observa-se, também, na Figura 5, a formação de quatro grupos: sendo dois grupos com duas cultivares isoladas (Choctaw e Chickasaw). O primeiro, com a cultivar Chickasaw, o segundo, formado pela cultivar Choctaw, e o terceiro grupo formado por duas cultivares, Shawnee e Stuart. O último, de maior similaridade genética, com 13 cultivares, constituído por Moneymaker, duas Barton, três Imperial, duas Desirable, uma Jackson, uma Farley, uma Sucess, uma Importada e uma Mahan,. O gráfico resultante da Análise de Coordenadas

Principais demonstra consistência com as estimativas de similaridade obtidas nas análises, portanto, revelaram indícios de uma correlação positiva entre dissimilaridade genética e distância geográfica entre as cultivares de *Carya illinoensis* estudados.

Estudos em espécies para fins comerciais como em variedades de oliveira na bacia do Mediterrâneo realizadas por Haddad et al. (2020) utilizando marcadores cpSSR e nSSR demonstraram que a análise de coordenadas principais (PCoA) foi capaz de separar um grande número de variedades ocidentais do grupo centro-oriental. As amostras argelinas mostraram uma distribuição bimodal com um grupo de cultivares da Turquia e do Oriente, e um segundo grupo incorporando principalmente as variedades do Mediterrâneo central (Itália) e da Península Balcânica. As cultivares de oliveira Argelina foram consideradas um valioso recurso genético. Os resultados obtidos tanto com marcadores microsatélite nucleares e plastidiais foram ferramentas úteis para futuros programas de certificação de materiais na Argélia.

Partindo dessas análises, a partir da análise integral dos dados feita com todos os marcadores plastidiais (prospectados e universais), obteve-se uma correlação cofenética de valor 0,9931 possibilitando a construção de um dendrograma (Figura 5). A correlação cofenética mede o grau de ajuste entre a matriz de similaridade original e a matriz resultante da simplificação proporcionada pelo método de agrupamento, sendo assim, quanto mais próxima de 1, menor será a distorção provocada pelo agrupamento dos indivíduos com o método UPGMA (MEYER, 2002). Algumas cultivares coincidiram no dendrograma e no agrupamento gerado pelo GenAlEx e alguns grupos não coincidiram, isto é, foram subdivididos ou englobaram mais cultivares. Uma explicação para isso pode ser devido à baixa consistência dos nós obtidos nos agrupamentos maiores, ou seja, envolvendo mais cultivares, avaliados por 1000 reamostragens *bootstrap*. Mesmo o valor cofenético sendo 0,9931 os valores de *bootstrap* nos agrupamentos maiores foram baixos.



**Figura 6.** Dendrograma de UPGMA baseada na distância euclidiana de 17 amostras de noqueira-pecã com marcadores plastidiais.

O dendrograma (Figura 6) demonstra que pelo menos quatro grupos se formaram o que indica que podem estar ligados a genes sob ação de seleção, mas não necessariamente as características morfológicas. O primeiro agrupamento contém as duas cultivares Shawnee (obtida do cruzamento dirigido entre as cultivares Schley e Barton) e Stuart (origem desconhecida). Ambas com dicogamia do tipo protogínica (HAMANN et al., 2018). E, revelam-se mais próximas geneticamente com *bootstrap* de 70%. O segundo, com as cultivares Mahan, Imperial, Sucess, Barton, Importada, Desirable, Jackson, Farley e Moneymaker permaneceram próximas ao agrupamento das cultivares Stuart e Shawnee no dendrograma, se mostraram mais próximas geneticamente com *bootstrap* de 55%. Todas possuem dicogamia do tipo protogínica, com exceção das cultivares Sucess, Barton, Desirable e Jackson que são protândricas (HAMANN et al., 2018). O terceiro e quarto grupo observado contempla duas cultivares, Choctaw (*bootstrap* 50%) e Chickasaw (*bootstrap* 100%), ambas com dicogamia protogínica (HAMANN et al., 2018). Foi possível observar no dendrograma grupos genéticos bem definidos, representados nos *bootstraps*, o que indica subpopulações

bem estruturadas dentro da população das cultivares de *C. illinoensis*. A avaliação da confiabilidade dos nós obtidos, através dos resultados dos *bootstraps*, mostrou a relação entre alguns grupos formados no dendrograma. Apenas alguns nós do dendrograma apresentaram valores de *bootstrap* maiores que 50%, estando estes localizados geralmente entre indivíduos com alta similaridade genética no final dos ramos ou nós que incorporam maior quantidade de indivíduos.

Feng et al (2018) analisando a diversidade *Juglans sigillata* (Juglandaceae) com marcadores plastidiais observaram que a base para a divergência genética entre as populações da espécie foi provavelmente o isolamento por distância combinado com adaptação regional. As espécies de *Juglans* foram divididas em dois clados correspondentes a duas seções (*Juglans/Dioscaryon* e *Cardiocaryon*) com base na variabilidade de sequência nos seis *loci* cpSSR, com suporte 100% *bootstrap* (com base em sequências inteiras do genoma do cloroplasto de cinco amostras) (HU et al. 2017). O haplótipo denominado como H1 foi encontrado apenas em *J. sigillata* (não em *J. regia*), mas sua posição era intermediária entre H2 e H4, que eram ambos haplótipos de cloroplasto de *J. regia*. O transcriptoma de *J. sigillata* auxiliou em estudos genômicos de *Juglans*, especialmente sobre as relações entre a expressão gênica e o desenvolvimento da noqueira, além disso, poderá ser útil para entender a genômica de *J. regia*, uma espécie importante comercialmente e parente mais próximo de *J. sigillata* (FENG, et al., 2018).

As análises genético-moleculares deste estudo, baseadas nos marcadores microssatélite plastidiais não se correlacionaram com dados morfológicos das cultivares. Os SSR identificam a segregação mendeliana e podem ser isolados em regiões transcritas e intergênicas, sendo geralmente neutros do ponto de vista evolutivo. Estas características permitem sua aplicabilidade em diversos estudos como, dinâmica evolutiva e aplicações na genética de população e da conservação, mapeamento genético, análise de parentesco, perfil molecular, padrões de hibridação, determinação do modo de reprodução, dentre outros (GIUSTINA et al., 2017). No trabalho de Poletto et al. (2020) no qual analisou o comprimento, diâmetro, peso da casca, peso da amêndoa e o peso total conjuntamente utilizando marcadores AFLP. A morfologia apresentou-se variada, assim como a diversidade genética, e não tiveram ligação entre eles. De Oliveira et al. (2021) em análises genético-moleculares com marcadores S-SAP, demonstraram uma diversidade genética considerada moderada a alta nos acessos de *C. illinoensis* estudados no Rio Grande do Sul, onde foi relatado que a diversidade é crucial para a realização de melhoramento genético da espécie.

Uma vez que as árvores utilizadas no estudo foram árvores antigas, todas com mais de 50 anos, concluindo que as cultivares possuíam um padrão genético desejável para fins comerciais.

O conhecimento da diversidade genética é de fundamental importância para a conservação e manutenção de recursos genéticos em programas de melhoramento (COSTA et al., 2012), permitindo entender a relação de parentesco entre os genótipos e identificar os melhores genitores, para a obtenção de maiores ganhos genéticos em populações segregantes (VIANA et al., 2003; GANGA et al., 2004). Este estudo encontrou uma grande variabilidade genética entre as cultivares avaliadas por meio de marcadores microssatélites plastidiais e nucleares. Com o conhecimento da variabilidade genética em *C. illinoensis*, é possível indicar cruzamentos interespecíficos com o objetivo de transferir alelos favoráveis, como o de resistência a doenças, para as cultivares. A diversidade genética encontrada neste trabalho pode ser explorada em um futuro programa de melhoramento genético, e este resultado potencializa os avanços no estudo na seleção de cultivares de noqueira-pecã no Rio Grande do Sul na utilização como fonte de germoplasma. A alta diversidade encontrada neste estudo pode ter ocorrido pelo cruzamento direto e espontâneo entre as diferentes cultivares, que resultou em novas cultivares, pois a forma de seleção das nozes durante muitos anos é baseada na escolha de indivíduos mais promissores, produtivos e resistentes ou tolerantes. Além de a conservação genética ser uma necessidade para as cultivares com características superiores, a confiabilidade da produção e facilidade de seleção para o tipo de uso da pecaneira pode ser facilitado. As análises de divergência genética poderão ser utilizadas para guiar a seleção de genitores a serem cruzados. O cruzamento entre genitores mais divergentes pode, potencialmente, maximizar a heterozigosidade e heterose da progênie, aumentando as chances de seleção de cultivares superiores. Cultivares de *C. illinoensis* selecionados e testados possuem maior chance de adaptação às condições edafoclimáticas do Estado.

## 5 CONCLUSÃO

Os 10 *loci* microssatélite plastidiais desenvolvidos e os índices de variação genética avaliados neste estudo foram úteis para estudar a divergência genética das cultivares. A transferibilidade interespecies desses *loci* microssatélites podem ser úteis em estudos genéticos, sendo possível validar estes marcadores e gerar informações a respeito da diversidade genética das espécies pertencentes à família Juglandaceae.

Nos 23 *loci* microssatélite analisados foram identificados alelos exclusivos utilizando tanto os cpSSR quanto nSSR, indicando a ausência de fluxo alélico, ou seja, há diferenciação entre os indivíduos. A ocorrência de alelos exclusivos pode também ser um indicativo de seleção. Foram identificados alelos raros utilizando os SSR nucleares (com frequência alélica menor que 0,05), que também pode indicar a perda da diversidade genética. Com o uso dos marcadores plastidiais, não houve alelos raros. A média de estimativa do índice de fixação foi acima de zero o que indica a ocorrência de endogamia entre as cultivares.

Os gráficos resultantes da Análise de Coordenadas Principais para cpSSR e nSSR demonstraram coerência com as estimativas de similaridade obtidas em ambas as análises. Sendo assim, pelo fato de a espécie apresentar uso comercial, tornam-se necessárias estratégias de um contexto de manejo para a conservação genética da espécie. Ações voltadas para aumentar o fluxo gênico entre as cultivares também são indicadas. Estas ações de manejo devem considerar os resultados observados, sob o risco de aumentar ainda mais o índice de fixação nas populações, à perda de alelos e a estruturação das populações. Assim, no decorrer das gerações, poderá ocorrer perda da diversidade genética e alteração da estrutura genética destas cultivares, o que implica em menor capacidade genética para cultivares com genótipos superiores para fins comerciais.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

- AZEVEDO, V.C.R., KANASHIRO, M., GRATTAPAGLIA, D.; CIAMPI, A. Y. Variabilidade no cpDNA em *Manilkara huberi*, espécie sob manejo sustentável na Amazônia brasileira. **Pesquisa Agropecuária Brasileira**, v. 43, n. 7, p. 859-867, 2008.
- BARROS, E. A. **Estrutura populacional e variabilidade genética do núcleo de conservação da raça marota no Piauí**. Universidade Federal Rural de Pernambuco, Recife. 2009.
- BRYAN, G.J.; MCNICOLL, J.; RAMSAY, G.; MEYER, R.C.; JONG, W.S.D. **Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants**. *Theor. Appl. Genet.* 99:859-867. 1999.
- CONNER, P.J. ROADBLOCKS AND HINDRANCES TO THE DEVELOPMENT OF MOLECULAR TOOLS IN PECAN (*Carya illinoensis*): A Breeder's PERSPECTIVE. **Acta Horticulturae**, 1070, 83–88. doi:10.17660/ActaHortic.2015.1070.9. 2015.
- CRUZ, C.D., REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV. 2ª edição revisada, 390p. 2004.
- CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Viçosa: Suprema, 2011. 620 p.
- DE OLIVEIRA L.O.; BEISE D.C.; DOS SANTOS D.D.; NAGEL, J.C.; POLETTO, T.; POLETTO, I.; STEFENON, V.M. Molecular markers in *Carya illinoensis* (Juglandaceae): from genetic characterization to molecular breeding, **The Journal of Horticultural Science and Biotechnology**, DOI: 10.1080 / 14620316.2021.1892534. 2021.
- DE OLIVEIRA L.O. **Caracterização molecular de acessos de noqueira-pecã *Carya illinoensis* (wangenh.) k. koch no Rio Grande do Sul usando marcadores S-SAP**. (Dissertação: Mestrado). Universidade Federal do Pampa, Rio Grande do Sul. 67 f. 2021.
- DEVEY, M.E.; BELL J.C.; SMITH D.N.; NEALE D.B.; MORAN, G.F. **A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers**. *Theor Appl Genet* 92 : 673—679. 1996.
- DOYLE, J.J.; DOYLE, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, v.19, p.11-15, 1987.
- DUARTE, V.; ORTIZ, E.R.N. Podridão de *Phytophthora* amêndoa e da casca da noqueira pecan. In: dos Santos, A.F., Luz, E.D.M.N., & Bezerra, J.L. (Ed.), **Doenças causadas por *Phytophthora* Brasil** (pp. 493–508). Campinas: Rural. 2001.
- EBERT, D.; PEAKALL, R. Chloroplast simple sequence repeats (cpSSR): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. **Molecular Ecology Resources**, Oxford, US, v. 9, p.673–690, 2009.
- FENG, X., YUAN X., YIHENG, H., ZULFIQAR, S., OUYANG, X., DANG, M., ZHOU, H., WOESTE, K., ZHAO, P. Resources for studies of iron walnut (*Juglans sigillata*) gene expression, genetic diversity, and evolution. **Tree Genetics & Genomes**. 14.

10.1007/s11295-018-1263-z. 2018.

FILIPPIN, I.L. **Viabilidade econômica do cultivo de noqueira pecã em áreas de reserva legal e de preservação permanente.** 74 f (MSc. Dissertation). Universidade Federal de Pelotas, Pelotas, Brazil. 2011.

FURLAN, R. A.; MORI, E. S. TAMBARUSSI, E. V.; MORAES, C. B.; JESUS, F. A.; ZIMBACK, L. Estrutura genética de populações de melhoramento de *Pinus caribaea* Var. *Hondurensis* por meio de marcadores microssatélites. **Bragantia**. Campinas, v. 66, n. 4, p. 553-563, 2007.

GAO, C.; REN, X.; MASON, A. S.; et al. Revisiting an important component of plant genomes: Microsatellites. **Functional Plant Biology**, v. 40, n. 7, p. 645–661, 2013.

GIUSTINA, L. D; ROSSI, A. B; VIEIRA, F. S; TARDIN, F. D; NEVES, L. G; PEREIRA, T. N. S. Variabilidade Genética em genótipos de Teca (*Tectona grandis* Linn. F). **Ciências Florestais**, v. 27, n. 4, p. 1311–1324, 2017.

GRAUKE, L.J.; IQBAL, M.J.; REDDY, A.S; THOMPSON, T.E. Development of 242 microsatellite DNA markers in pecan. **J. Amer. Soc. Hort. Sci.** 128:374-380. 2003.

GRAUKE, L.J.; MENDOZA-HERRERA, M.A.; MILLER, A.J.; WOOD, B.W. **Geographic patterns of genetic variation in native pecans.** *Tree Genetics & Genomes*, 7, 917–932. doi:10.1007/s11295-011-0384-4. 2011.

GRAUKE, L.J.; MENDOZA-HERRERA, M.A. POPULATION STRUCTURE IN THE GENUS *CARYA*. **Acta Hort.** 948, 143-158. DOI: 10.17660/ActaHortic.2012.948.16. <https://doi.org/10.17660/ActaHortic.2012.948.16>. 2012.

GRAUKE, L.J.; WOOD, B.W.; HARRIS, M.K. **Crop vulnerability: *Carya*.** *HortScience*, 51, 653–663. <https://doi.org/10.21273/HORTSCI.51.6.653>. 2016.

HADDAD, B.; GRISTINA, AS; MERCATI, F.; SAAFI, AE; AITER, N.; MARTORANA, A.; SHARAF, A.; CARIMI, F. Molecular Analysis of the Official Algerian Olive Collection Highlighted a Hotspot of Biodiversity in the Central Mediterranean Basin. **Genes**, 11, 303. <https://doi.org/10.3390/genes110303033>. 2020.

HAMANN, J. J.; BILHARVA, M. G.; BARROS, J. de; MARCO, R. de; MARTINS, C. R. (ed.). Cultivares de noqueira-pecã no Brasil. **Embrapa Clima Temperado**. p. 43. <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/190468/1/DOCUMENTO-478.pdf>. 2018.

HAMMER, Oyvind et al. PAST: Paleontological statistics software package for education and data analysis. **Palaeontologia electronica**, v. 4, n. 1, p. 9, 2001

HAYDEN, M.J.; SHARP, P.J. Targeted development of informative microsatellite (SSR) markers. **Nucleic Acids Research**, London, V. 29, n. 8, p. 1-6, 2001.

HU Y. H., WOESTE K. E., ZHAO P. Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. **Front Plant Sci** 7:1955. 2017.

INC – International Nut and Dried Fruit Council. Global statistical review: Crop forecast report. **Nutfruit**, 77, 65–78. 2019.

LEWONTIN, R.C. The apportionment of human diversity. In: **Evolutionary biology**.

Springer, Boston, MA, 1972. p. 381-398.

MEYER, A. S. Comparação de coeficientes de similaridade usados em análises de agrupamento com dados de marcadores moleculares dominantes. Dissertação - Escola Superior de Agricultura Luiz de Queiroz, 2002. Disponível em: Acessado em janeiro de 2022.

MORENO, M.A.; TARAZI, R.; FERRAZ, E.M.; GANDARA, F.B.; KAGEYAMA, P.Y. Estrutura genética espacial em populações de *Hymenaea stigonacarpa* Mart. ex Hayne mediante a utilização de marcadores microssatélites cloroplastidais. **Scientia Forestalis**, v. 37, p. 513-523, 2009.

PEAKALL, R.; SMOUSE P.E. **GenAlEx 6.5**: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28, 2537-2539, 2012.

POWELL, W.; MORGANTE, M.; ANDRE, C.; MCNICOL, J.W.; MACHRAY, G.C.; DOYLE, J.J.; TINGEY, S.V.; RAFALSKI, J.A. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. **Current Biology**, v. 5, n. 9, p. 1023-1029, 1995.

RAJORA, O. P. et al. Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. **Molecular Ecology**, v. 9, p. 339-348, 2000.

RENCHER, A.C. *Methods of Multivariate Analysis*. A John Wiley & Sons, inc. Publication.p.727. 2 ed. 2002.

SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**. v. 27. p. 379-623. 1948.

SPARKS, D. **Adaptability of pecan as a species**. *HortScience*, 40, 1175–1189. doi:10.21273/HORTSCI.40.5.1175. 2005.

STEFENON, V.M.; KLABUNDE, G.; LEMOS, R.P.M.; ROGALSKI, M.; NODARI, R.O. **Phylogeography of plastid DNA sequences suggests post-glacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia***. *Scientific Reports*, 9, 2752. doi:10.1038/s41598-019-39308-w. 2019.

SOUZA, H. G.; DORIA, K. M. A. B. V. S.; BASSETO, M. A.; ROSA, D. D.; FURTADO, E. L.; MARINO, C. L. Diversidade genética em populações-núcleo de *Eucalyptus grandis*. **Acta Scientiarum Agronomy**. Maringá, v. 32, n. 4, p. 621-625, 2010

SUJII, P.S., FERNANDES, E.T.M.B., AZEVEDO, V.C.R., CIAMPI, A.Y., MARTINS, K., WADT, L.H.O. Morphological and molecular characteristics do not confirm popular classification of the Brazil nut tree in Acre, Brazil. **Genet. Mol. Res.** 12 (3), 4018–4027. 2013.

STEANE, D. Complete nucleotide sequence of the chloroplast genome from the tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). **DNA research**, Tóquio, v. 12, p. 215-220, 2005.

TARAZI, R.; MANTOVANI, A.; REIS, M.S. Fine-scale spatial genetic structure and allozymic diversity in natural populations of *Ocotea catharinensis* Mez. (Lauraceae). **Conservation Genetics**, Dordrecht, v. 11, p. 965-976, 2009.

VARSHNEY, R.K., GRANER, A. & SORRELLS, M.E. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23: 48-55. 2005.

WEBER, J. L. Informativeness of human (dC-dA)<sub>n</sub> · (dG-dT)<sub>n</sub> polymorphisms. **Genomics**, v. 7, n. 4, p. 524–530, 1990.

WEISING K.; GARDNER R.C. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. **Genome**, 42:9-19. 1999.

ZANELLA, C. M; TURCHETTO, C; PALMA-SILVA, C. Microsatélites: Metodologias de identificação e Análise. **In: Marcadores Moleculares na Era Genômica: Metodologias e Aplicações**. São Paulo: Sociedade Brasileira de Genética, 2017. CAP 6, P.95-117.

ZHANG, R.; PENG, F.; LI, Y. **Pecan production in China**. *Scientia Horticulturae*, 197, 719–727. doi:10.1016/j.scienta.2015.10.035. 2015.

## 7 CONSIDERAÇÕES FINAIS

As pesquisas com noqueira-pecã estão em crescimento e são de suma importância para o setor de pecanicultura, embora a escassez de estrutura e mão de obra especializada, a possibilidade de novos estudos é bastante promissora e necessária, principalmente no Rio Grande do Sul onde o crescimento da produção é crescente. Esse resultado corrobora com a importância do avanço em estudos genéticos para a espécie, contribuindo dessa maneira para a seleção de cultivares superiores e assim gerar novos bancos de germoplasma ligados a produtividade no sul do Brasil.

Com base nos dados obtidos nessa Tese, os marcadores microssatélite plastidiais recém-desenvolvidos para *C. illinoensis* são úteis em estudos genéticos, onde foi possível gerar informações a respeito da diversidade genética das cultivares plantadas no Rio Grande do Sul. Com a caracterização e validação dos marcadores cpSSR espécie-específicos, outros estudos genéticos poderão ser realizados utilizando essas marcas, e assim avançar na caracterização genética da espécie que se apresenta como um recurso de importância econômica para o Brasil e como subsídio a programas de melhoramento. Além disso, este é o primeiro estudo de desenvolvimento, caracterização e validação de marcadores microssatélite plastidiais espécie-específico para *C. illinoensis*, com potencial aplicabilidade para outros gêneros de Juglandaceae (*Juglans*, *Carya* e *Cyclocarya*).

Os dados obtidos até o momento nos permitem auxiliar no aprofundamento de análises da diversidade genética, pois por meio de estudos recentes a polinização aberta promove alta segregação genética, e a propagação através de sementes pode ter introduzido centenas de variedades de noqueira-pecã pelo mundo. Essa grande diversidade na forma dos frutos, das árvores, a qualidade das nozes e suas diferentes formas de se reproduzir, ainda causa uma grande confusão quanto à nomenclatura e a identificação. Os resultados obtidos neste estudo evidenciaram alta diversidade genética tanto para marcadores nucleares ( $I= 2,365$ ) e plastidiais ( $I= 2,183$ ), esta estimativa de diversidade genética é bem significativa, e este resultado pode ser utilizado para determinar o padrão genético das cultivares analisadas.

