

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

BÁRBARA SCORSIM ARJONA

**UTILIZAÇÃO DE TÉCNICAS E FERRAMENTAS DA BIOINFORMÁTICA PARA
CLASSIFICAÇÃO DOS RETROTRANSPÓSONS DE *Manihot esculenta* Crantz**

DOIS VIZINHOS, PR

2022

BÁRBARA SCORSIM ARJONA

**UTILIZAÇÃO DE TÉCNICAS E FERRAMENTAS DA BIOINFORMÁTICA PARA
CLASSIFICAÇÃO DOS RETROTRANSPOSONS DE *Manihot esculenta* Crantz**

**Use of Bioinformatics Techniques and Tools to Classification the
Retrotransposons of *Manihot esculenta* Crantz**

Trabalho de conclusão de curso de Especialização apresentado como requisito para obtenção do título de Especialista em Biologia Molecular – Habilitação Bioinformática da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientadora: Betty Cristiane Kuhn

DOIS VIZINHOS, PR

2022



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Esta licença permite download e compartilhamento do trabalho desde que sejam atribuídos créditos ao(s) autor(es), sem a possibilidade de alterá-lo ou utilizá-lo para fins comerciais. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

BÁRBARA SCORSIM ARJONA

**UTILIZAÇÃO DE TÉCNICAS E FERRAMENTAS DA BIOINFORMÁTICA PARA
CLASSIFICAÇÃO DOS RETROTRANSPÓSONS DE *Manihot esculenta* Crantz**

Trabalho de conclusão de curso de
Especialização apresentado como requisito
para obtenção do título de Especialista em
Biologia Molecular – Habilitação
Bioinformática da Universidade Tecnológica
Federal do Paraná (UTFPR).

Data de aprovação: 10 de janeiro de 2023

Betty Cristiane Kuhn

Doutora

Universidade Tecnológica Federal do Paraná – Dois Vizinhos

Fernanda Gatto de Almeida

Doutora

Universidade Tecnológica Federal do Paraná – Dois Vizinhos

Francisco Menino Destefanis Vitola

Doutor

Universidade Tecnológica Federal do Paraná – Dois Vizinhos

DOIS VIZINHOS, PR

2022

Dedico à minha rede de apoio e amor.

AGRADECIMENTOS

A Deus.

À minha família, meus pais, Sheila Scorsim e Marcos Arjona, minha irmã, Isadora Scorsim por todo apoio ao longo de todo o curso.

Ao meu namorado Cesar Augusto Machado por toda ajuda na fase final do curso.

À minha orientadora Prof^a Dra Betty Cristiane Kuhn, pela sabedoria com que me guiou nesta trajetória.

A todos que me auxiliaram de perto ao longo da realização deste trabalho.

A todos os colegas de curso, que mesmo a distância, sempre estiveram dispostos a ajudar.

A Secretaria do Curso, pela cooperação.

Os autores gostariam de agradecer ao Laboratório de Análises Biológicas e Biologia Molecular (BioMol) da Universidade Tecnológica Federal do Paraná – Câmpus Dois Vizinhos

RESUMO

O cultivo de *Manihot esculenta* Crantz no território brasileiro apresenta alta relevância nutricional e cultural, além de ser matéria prima para diversos produtos. A partir da aplicação de técnicas *in vitro* algumas variações podem surgir, acarretando variabilidade. Em muitos casos, os retrotransposons vêm sendo utilizados com sucesso para identificá-las; além disso, suas aplicações são as mais variadas, indo de investigação de ativação, mobilidade para estudos de biodiversidade para mapeamento de genes, bem como a análise da variabilidade genética. Dessa forma, o presente manuscrito teve como objetivo geral a prospecção dos retrotransposons de *Manihot esculenta* disponíveis no banco de dados primário e secundário, e como objetivos específicos o agrupamento das sequências de LT_LTR, bem como a identificação dos cromossomos nos quais esses elementos estão disponíveis. Para isso, foi realizado o levantamento de registros para a espécie utilizando banco de dados para se obter informações como o país de origem da sequência de proteínas, do código de identificação do cultivar, o tamanho da sequência e cromossomo em que se encontra. Posteriormente, foram selecionadas apenas as sequências referentes a retrotransposons de longa repetição terminal (RT_LTR), devido a sua alta taxa de presença nos genomas. Os dados obtidos, além de identificar a presença desse elemento em 17 cromossomos distintos, permitiram a identificação de uma migração não constante entre os cromossomos, mostrando que essa inserção não segue um padrão previsível para a espécie estudada.

Palavras-chave: cromossomos; LTR; mandioca; migração cromossomal.

ABSTRACT

The cultivation of *Manihot esculenta* Crantz in the Brazilian territory has high nutritional and cultural relevance, in addition to being raw material for several products. From the application of in vitro techniques some variations may arise, causing variability. In many cases, retrotransposons have been successfully used to identify them; moreover, its applications are the most varied, ranging from activation investigation, mobility to biodiversity studies for gene mapping, as well as the analysis of genetic variability. Thus, the present manuscript had the general objective of prospecting the *Manihot esculenta* retrotransposons available in the primary and secondary database, and the specific objectives of grouping the LT_LTR sequences, as well as identifying the chromosomes in which these elements are available. For this, a survey of records for the species was carried out using a database to obtain information such as the country of origin of the protein sequence, the cultivar identification code, the size of the sequence and the chromosome in which it is located. Subsequently, only sequences referring to long terminal repeat retrotransposons (RT_LTR) were selected, due to their high presence rate in genomes. The data obtained, in addition to identifying the presence of this element in 17 different chromosomes, allowed the identification of a non-constant migration between the chromosomes, showing that this insertion does not follow a predictable pattern for the studied species.

Keywords: cassava; chromosomal migration; LTR

SUMÁRIO

1 INTRODUÇÃO.....	09
1.1 OBJETIVOS.....	10
2 REVISÃO DE LITERATURA.....	11
2.1 CLASSIFICAÇÃO E DISTRIBUIÇÃO DE <i>Manihot esculenta</i>	11
2.1.1 Genética de <i>Manihot esculenta</i>	12
2.2 RETROTRANSPOSONS.....	14
2.3 BIOINFORMÁTICA.....	15
2.3.1 Alinhamento de sequências e algoritmo Clustal W.....	16
2.4 FILOGENÉTICA.....	18
3 MATERIAL E MÉTODOS.....	20
4 RESULTADOS E DISCUSSÃO.....	22
5 CONSIDERAÇÕES FINAIS.....	28
REFERÊNCIAS.....	29
APÊNDICE A - Relação das sequências utilizadas nas análises do presente trabalho, após as triagens realizadas.....	38
APÊNDICE B - Distância p referente as sequências classificadas como “<i>uncharacterized protein</i>” no banco de dados.....	57
APÊNDICE C - Distância p referente as sequências classificadas como “<i>hypothetical protein</i>” e “<i>mucin-17</i>” no banco de dados.....	60
APÊNDICE D - Distância p referente as sequências agrupadas de acordo com os cromossomos de origem.....	62

1 INTRODUÇÃO

A mandioca (*Manihot esculenta* Crantz), também conhecida como aipim, macaxeira ou maniva, é uma espécie pertencente à família Euphorbiaceae, originária da floresta amazônica (ALLEM, 1994; OLSEN; SCHAAL, 1999), e que se encontra dispersa por praticamente todo o globo, sendo a América do Sul o continente com maior número de registros (GBIF.org). O cultivo em território brasileiro apresenta alta relevância nutricional e cultural, além de ser matéria prima para diversos produtos (GUIMARÃES et al., 2022). Devido ao fato de ser uma espécie altamente heterozigótica e poliploide ($2n=4x=36$), sua propagação geralmente é feita por estacas e não por sementes (FAO, 2000).

Diante da possibilidade de se obter plantas completas por meio do cultivo de células, tecidos ou órgãos, as técnicas *in vitro* tornaram-se ferramentas utilizadas na conservação de recursos, sendo uma valiosa alternativa para a manutenção das coleções de plantas que podem ser incorporadas em programas de melhoramento vegetal (VIEIRA, 2000).

Durante as últimas décadas a utilização de marcadores moleculares vem desempenhando um papel fundamental nos estudos genéticos referentes a biotecnologia vegetal, bem como no campo da genética de melhoramento para seleção de plantas e melhoramento de culturas (HAYWARD et al., 2015). Em muitos casos, os retrotransposons vêm sendo utilizados com sucesso e as aplicações são as mais variáveis, indo de investigação de ativação, mobilidade para estudos de biodiversidade para mapeamento de genes, bem como a análise da variabilidade genética (KALENDAR et al., 2011).

Esses elementos transponíveis foram descobertos na década de 40 (MCCLINTOCK, 1931), constituem grande parte do genoma dos eucariotos, e por definição podem ser descritos como fragmentos de DNA que podem se inserir em novas regiões cromossômicas, fazendo cópias duplicadas de si mesmos, sendo um dos principais componentes dos genomas das plantas, podendo vir a constituir mais de 70% do genoma em algumas plantas (FESCHOTTE et al., 2002; VENNER et al., 2009).

Estudos recentes apresentam por base a identificação desses elementos, sendo por sequenciamento ou por utilização de marcadores moleculares

desenvolvidos com base em sequências de longa repetição terminal (LTRs) dos retrotransposons, desses estudos, diversos foram realizados tentando elucidar a importância e o impacto desses elementos na evolução dos genomas, considerando as suas inserções e/ou deleções em diferentes locais do genoma, causando alterações, podendo ser por transferência vertical ou horizontal (SCHAACK et al., 2010).

Com base no exposto, é inequívoca a importância dos estudos envolvendo elementos transponíveis com o intuito de compreensão dos genomas, e conseqüentemente, sua evolução; além também da potencial utilização na distinção de genótipos, servindo como uma ferramenta para futuros estudos de biologia molecular de plantas.

1.1. Objetivos

Prospectar os retrotransposons de *Manihot esculenta* disponíveis no banco de dados primário e secundário. E, como objetivos específicos:

- Agrupar as sequências de codificadas por LT_LTR;
- Identificar os cromossomos no quais os LTRs estão inseridos;
- Analisar o padrão de migração dos retrotransposons nos cromossomos.

2 REVISÃO DE LITERATURA

Abaixo estão relacionados os itens que versam em torno da teoria e conceitos, bem como a revisão literária acerca do assunto a ser tratado neste trabalho.

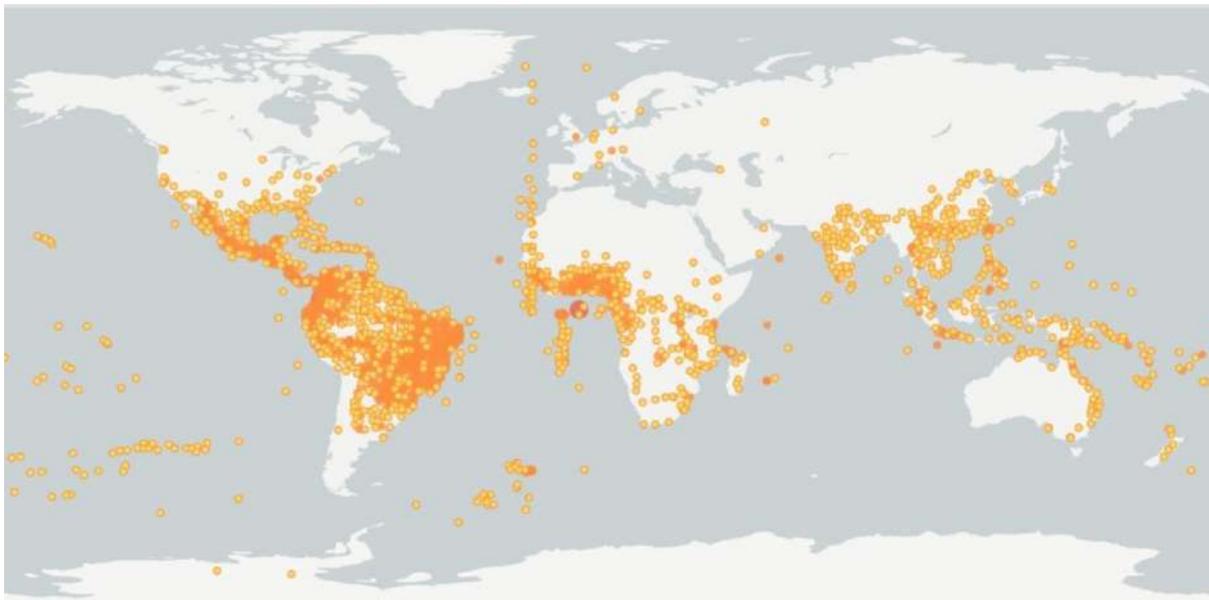
2.1 Classificação e distribuição de *Manihot esculenta*

Euphorbiaceae s.l. é uma família que apresenta destaque entre as Angiospermas pelo fato de abrigar 8.382 espécies reunidas em 403 gêneros, além de apresentar mais de 3.530 ocorrências registradas (GBIF.org, 2022), um aumento de mais de 2.000 descrições de espécies para o ano de 2000 (GOVAERTS et al., 2021). Encontram-se distribuídas principalmente nas regiões tropical e subtropical, especialmente nos continentes americano e africano, com exceção para poucos gêneros extratropicais (WEBSTER, 1994). Compreende um dos grupos taxonômicos mais complexos e morfológicamente diversos entre as Eudicotiledôneas (JUDD et al., 1999; SOLTIS et al., 2005; SIMPSON, 2006).

Manihot esculenta Crantz, pertencente à família *Euphorbiaceae*, é originária da floresta amazônica (ALLEM, 1994; OLSEN & SCHAAL, 1999), também conhecida como aipim, macaxeira ou maniva se encontra dispersa por praticamente todo o globo, e de acordo com os registros disponíveis no site GBIF.org, a América do Sul é o continente que apresenta o maior número de registros, seguido pela Europa e América do Norte. O Brasil é o país com maior incidência (Figura 1).

Na revisão realizada por Guimarães e colaboradores (2022) concluiu-se que o cultivo da mandioca produzida no território brasileiro apresenta grande relevância, seja nutricional ou cultural, além de indicar que a mandioca é a matéria prima para o desenvolvimento de inúmeros produtos, como por exemplo o álcool, plásticos biodegradáveis, cosméticos, bem como ração animal e matéria da indústria têxtil. Dada a importância, a mandioca pode ser considerada um componente da cultura brasileira.

Figura 1. Distribuição referente aos registros de *Manihot esculenta*. Fonte: GBIF.org



A mandioca é considerada o alimento predominante na dieta diária da população de muitos países (CEBALLOS et al., 2006) e, de acordo com o último levantamento da Organização das Nações Unidas para a Alimentação e Agricultura, o maior produtor mundial de mandioca é a Nigéria, país que no ano de 2016 computou 57,13 milhões de toneladas, seguido por Tailândia e Indonésia. O Brasil é o 4º maior produtor mundial (AJAYI; OLUTUMISE, 2018) com 21,08 milhões de toneladas de raiz de mandioca. Na safra referente ao ano de 2022 o Brasil, na variável de produção, teve 18.129.427 t, sendo o Norte a região com maior produção compreendendo 35,61%, seguido pela região Sul com 22,61%, região Nordeste (21,05%), Sudeste (12,51%) e Centro-Oeste (8,22%) (IBGE, 2022).

2.1.1. Genética de *Manihot esculenta*

A *M. esculenta* é uma angiosperma que se apresenta com 18 cromossomos, como mostram os dados disponíveis no GenBank, provenientes do BioProjeto com o código de acesso “PRJNA234389” (Disponível em <<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA394209>>). Além disso, de acordo com Leitch e Bennet (1997), em torno de 80% das angiospermas teriam origem poliploide; é entendido por poliploidia a existência, em um mesmo núcleo celular, de dois ou mais genomas. Tal condição pode ocorrer naturalmente ou ser induzida

(HAHN et al., 1990; HANN et al., 1992). A produção dos poliploides da mandioca por vias induzidas tem sido relatada por meio de tratamento *in vivo* de segmentos nodais (NASSAR, 2004) e por tratamento *in vitro* de botões apicais e axilares (AWOLEYE et al., 1994; CARVALHO et al., 2016), sendo destacada a associação realizada por Awoleye e colaboradores (1994) entre as diferentes características fenotípicas à poliploidia durante a regeneração *in vitro* de plantas de mandioca.

No estudo de Qi e colaboradores (2021), por meio do uso da estratégia de sequenciamento HiFi combinada com Hi-C, foi possível a reconstrução de dois genomas haploides em escala cromossômica; o genoma haplótipo-resolvido permitiu a realização da primeira visão sistemática da organização heterozigótica do genoma diploide da mandioca. A identificação de indivíduos heterozigotos diploides também foi possível a partir do uso de marcadores microssatélites (PAREDES et al., 2021).

Algumas variações podem surgir com a aplicação de técnicas *in vitro*, acarretando variabilidade. Com o uso desse tipo de cultivo para germinação de sementes tem-se um aumento na sua diversidade por meio das variações somaclonais, as quais não podem ser evitadas por esse meio (COSTA, 2013).

A variação somaclonal pode ser utilizada como uma ferramenta para gerar variabilidade e consiste na variação originária das culturas de células e tecidos (LARKIN & SCOWCROFT, 1981). Tais variações são úteis para se obter espécimes mais resistentes à alguns tipos de estresse, sejam esse bióticos ou abióticos (BARROSO et al., 2003), como por exemplo pragas, que causam perdas de até 40% na produção (BULL et al., 2011). Todavia, dela também podem resultar haplótipos diferentes e hereditários (KAEPPLER et al., 2000; BAIRU et al., 2011), resultando em prejuízos aos produtores que tenham interesse em uma produção geneticamente uniformes (CELÉDON et al., 2000; SILVA et al., 2010). O estudo, bem como o acompanhamento, em gerações seguintes da variação somaclonal é de suma importância (MARSHALL & COURDURIES, 1992).

Dentre as ferramentas utilizadas para o estudo da variação somaclonal, os marcadores moleculares estão em destaque. Existem diferentes tipos de marcadores, tais como morfológicos, bioquímicos e moleculares baseados no DNA. Xiong e colaboradores (2022) desenvolveram uma técnica nova e universal de marcador molecular EBAP (*exon based amplified polymorphism*) que não requer informação da

sequência. Dos parâmetros obtidos no trabalho, a mandioca apresentou 5 genótipos, 21 primers polimórficos, 123 bandas totais, sendo 67 bandas polimórficas.

Dentre os marcadores usados atualmente, os locos de sequências simples de DNA repetidas (locos SSR, também denominados locos microssatélites) têm sido apontados como um dos mais adequados para estudos referentes ao genoma de plantas. Essa classe apresenta vantagens em relação ao elevado polimorfismo, característica que é útil para estudos em cultivares com baixa variabilidade genética por exemplo (RAMOS et al., 2006; SILVA et al., 2009; WHANKAEW et al., 2011), além da sua alta taxa de mutação no genoma (JARNE & LAGODA, 1996). No trabalho de Paredes e colaboradores (2021), a utilização de marcadores microssatélites possibilitou a determinação da diversidade genética. Além disso, o marcador foi capaz de gerar uma quantidade crítica de informação genética devido sua capacidade de cobrir todo o genoma.

2.2 Retrotransposons

Os retrotransposons podem ser definidos como fragmentos de DNA que podem se inserir em novas regiões cromossômicas, fazendo cópias duplicadas de si mesmos (FESCHOTTE et al., 2002). Em muitos casos eles compreendem mais de 50% do conteúdo do DNA nuclear (KUMAR & BENNETZEN, 1999) e em alguns casos chegam a valores próximos de 78%, como é o caso do milho (KUHN et al., 2014). Devido a estas propriedades, essa família de marcadores tem sido utilizada com sucesso para análise da variabilidade genética (KALENDAR et al., 2011).

Kumar e Bennetzen (1999) apontaram os retrotransposons como elementos que apresentam um potencial excepcional para serem utilizados como ferramentas genéticas com o intuito de análises do genoma vegetal, mapeamento de ligações, estudos de filogenia e biodiversidade genética, marcação de genes, análise funcional de genes, avaliação de variantes somaclonais e nos processos de transferência de genes. Tais *insights* certamente foram a base para os diversos estudos que existem atualmente.

Resultados positivos por meio do uso desses elementos como marcadores também foram encontrados em outros tipos de cultivares, como no trabalho realizado por Santos e colaboradores (2019), no qual os autores pontuaram a eficiência do uso de marcadores moleculares SSR e IRAP (Inter-retrotransposon amplified

polymorphism) em cultivares de bananas, uma vez que eles foram eficientes para quantificar a variabilidade genética resultante da poliploidização via colchicina de diploides de banana selvagem.

Em cultivares de uvas, a análise realizada a partir do polimorfismo dos retrotransposons IRAP e REMAP (retrotransposon-microsatellite amplified polymorphism) foi essencial para mostrar que a propagação clonal de mutações somáticas pode levar à formação de cultivares geneticamente divergentes, além de mostrar a mistura de genomas dentro de cada cultivar (STRIOTO et al., 2019).

Os retrotransposons vem sendo utilizados em estudos referentes aos cultivares de mandioca com sucesso. No trabalho realizado por Kuhn e colaboradores (2016), por meio do uso de LTRs, foram identificadas doze famílias de retrotransposons LTR no banco de dados; ademais, os autores apontaram que marcador baseado em retrotransposon se mostrou eficiente tanto na diferenciação de quatro cultivares de mandioca quanto para determinar a diversidade dentro das amostras.

Em outro estudo de Kuhn e colaboradores (2017), mostrou-se que, o uso de marcadores baseados em retrotransposons LTRs (marcadores IRAP e REMAP) é indicado para monitoramento de programas de cultivo *in vitro*, analisando plantas cultivadas com as mesmas concentrações de reguladores de crescimento o marcador baseado em retrotransposon revelou maior polimorfismo que o marcador baseado em sequencias EST-SSR.

2.3 Bioinformática

A bioinformática pode ser definida como um campo da ciência computacional que tem relação com análise de sequências de moléculas biológicas, podendo essa ser pensada como a parte linguística da genética (AUSTIN, C.P., 2022). Dessa maneira, trata-se da gestão e análise computacional da informação biológica: genes, proteínas e células (HUANG & LEMPICKI, 2009; RAMHARACK & SOLIMAN, 2018).

Devido aos avanços nas técnicas de manipulação de DNA, nos estudos referentes aos genes estruturais e/ou funcionais, nas análises de genomas, de variabilidade genética e na identificação por meio da biologia molecular, o entendimento e consequente estudo das biomoléculas se mostrou uma ferramenta poderosa (LESK, 2008). Com isso, esse campo assume um papel interdisciplinar por

meio da fusão das ciências computacionais, biológica e estatística (RAMHARACK & SOLIMAN, 2018).

Das competências da bioinformática, uma refere-se à criação e manutenção dos bancos de dados biológicos (BDB), os quais, dentre os diversos objetivos, pode-se citar o armazenamento de milhões de sequências de biomoléculas, tais como: DNA, RNA e proteínas (LESK, 2008). Um banco de dados pode ser definido como uma coleção de dados interrelacionados, no qual as informações estão organizadas e estruturadas. De acordo com o conteúdo por eles armazenados, os bancos de dados podem ser subdivididos em três grupos: primários, secundários e especializados (ROB, 2014).

O banco de dados primário diz respeito aos dados originais de sequências de nucleotídeos ou proteínas. Um grande exemplo desse grupo é o GenBank <<https://www.ncbi.nlm.nih.gov/genbank/>>, o qual é uma base de dados pública de sequências nucleotídicas e de anotações bibliográficas e biológicas de apoio. Essa base é construída e distribuída pelo *National Center for Biotechnology Information* (NCBI), uma divisão da *National Library of Medicine* (NLM), localizada no campus dos Institutos Nacionais de Saúde dos EUA (NIH) em Bethesda, MD, EUA (BENSON et al., 2018).

Por sua vez, os bancos de dados secundários armazenam resultados de análises feitas a partir de dados primários. Pode-se citar, como exemplo, o Phytozome <phytozome.jgi.doe.gov>, que é um grande portal genômico de plantas mantido pelo Departamento de Energia dos EUA (DOE). Além das ferramentas BLAST e Gbrowse, Phytozome também fornece Biomart, que permite aos utilizadores anotar famílias de genes vegetais, estudar a evolução desses e exibir genes no contexto genômico (GOODSTEIN et al., 2012), o que é valioso para elucidação acerca da evolução das famílias de genes. Por fim, os bancos de dados especializados, como por exemplo, o Maize Genomics Resource <maize.uga.edu>, tendem a contemplar dados que atendem a um interesse particular.

2.3.1 Alinhamento de sequências e algoritmo Clustal W

Como forma de organizar sequências para identificar regiões similares, usa-se o alinhamento, o qual se dá por meio da comparação de duas ou mais sequências procurando caracteres ou padrões de caracteres que se encontram na mesma ordem,

objetivando organizar estruturas primárias de DNA, RNA ou proteínas e identificar regiões similares e divergentes. Esses alinhamentos podem ser classificados em global e local. O alinhamento global refere-se ao alinhamento que se estende por toda a sequência, sendo o algoritmo Needleman-Wunsch o mais conhecido para realizar esse tipo de alinhamento (THOMPSON et al., 1994). Em contrapartida, o alinhamento local acontece quando a comparação entre as sequências é feita por meio de pequenas regiões, não importando as sequências adjacentes a estas regiões, e o principal programa utilizado para o alinhamento local de sequências é o BLAST (*Basic Local Alignment Search Tool*) (PROSDOCIMI et al., 2012).

A importância do alinhamento se dá em diversas vertentes, bem como em diversos organismos, tais como: estabelecer identidade entre as sequências, definição de domínios proteicos conservados, dedução de função de proteínas baseado em similaridade, caracterização de marcadores *barcodes*, estudo evolutivo e filogeográfico (MARI-RIBEIRO et al., 2022; DIAMANTE et al., 2020; MORAIS-SILVA et al., 2018).

O algoritmo básico de alinhamento múltiplo, ou seja, para três ou mais sequências, consiste em três principais fases conforme proposto por Thompson e colaboradores (1994); primeiramente todos os pares de sequências são alinhados separadamente a fim de calcular uma matriz de distância que dá a divergência de cada par de sequências; em seguida, uma árvore guia é calculada a partir da matriz de distância; por fim, as sequências são progressivamente alinhadas de acordo com a ordem de ramificação na árvore guia.

É importante levar em conta que ao realizar um alinhamento, o produto deve ser o melhor possível de ser obtido; em condições normais, o mais indicado é utilizar os algoritmos que produzem resultados precisos. Entretanto, em alguns casos, como na comparação de sequências de tamanhos mais longos, precisa-se de maior rapidez de busca (PROSDOCIMI et al., 2012). Nestes casos, portanto, aceita-se que o resultado obtido não seja “o melhor possível” e, assim, utiliza-se algoritmos que apresentam algum tipo de heurística, ou seja, que não irão realizar o alinhamento preciso entre sequências, uma vez que frequentemente utilizam alguma técnica alternativa para acelerar o resultado da busca por sequências similares (PROSDOCIMI et al., 2012).

Dentre os principais programas que permitem a realização de alinhamento, pode-se citar o Clustal, o qual consiste em uma série de programas de computador comumente utilizados para o alinhamento de sequências múltiplas (THOMPSON et al., 1994). O Clustal W é um tipo de alinhamento global, sem limitação de sequências a serem analisadas e com uma precisão de alinhamento do tipo heurístico, fornecendo, desse modo, ao usuário, uma grande quantidade de parâmetros e de saídas diferentes (THOMPSON et al., 1994; PROSDOCIMI et al., 2012), o qual pode ser realizado por meio do software MEGA (KUMAR et al., 2012).

2.4 Filogenética

Uma filogenia é definida por uma árvore de um grupo familiar de espécies, sendo essas inferidas a partir de análises morfológicas e consiste no estudo das relações entre diferentes grupos de organismos e seu desenvolvimento evolutivo. Com o início da utilização de ferramentas moleculares no ano de 1960, esses estudos tomaram uma grande proporção e por volta da década de 1980 os estudos conhecidos por filogenética molecular sobrevieram (RIDLEY, 2007).

Em estudos filogenéticos em plantas, a maioria das sequências obtidas derivam de plastomas, os quais são genomas dos plastídios, e isso ocorre devido ao elevado número destes por célula (TWYFORD & NESS, 2017). Com isso, os plastomas são muito utilizados para inferir relações a diferentes níveis taxonômicos, desde a filogeografia até as filogenias (FAYE et al., 2016; GITZENDANNER et al., 2018). Em contrapartida, o genoma mitocondrial é negligenciado quando se trata de filogenética vegetal. Isso se dá pela sua baixa taxa de substituição de nucleotídeos, gerando, de certa maneira, uma limitação para esses estudos (MOWER et al., 2012; FONSECA et al., 2019).

No trabalho executado por Neumann e colaboradores (2019), a análise filogenética realizada para os três domínios de poliproteínas mais conservados (RT, RH e INT) proporcionou a divisão de Ty1/copia e Ty3/gypsy retrotransposons em 16 e 14 linhagens, respectivamente. No trabalho, os autores também foram capazes de debitar várias características de sequências de retrotransposons LTR, estabelecendo, desse modo, um sistema e classificação melhorado aplicável aos retrotransposons LTR de uma vasta gama de espécies vegetais.

No trabalho de Biryukov e Ustyantsev (2022), a partir do uso do grupo mais abundante de ETs em retrotransposons de repetição terminal longa (LTR) de plantas (LTR-RTs) foi possível demonstrar que as DARTS (*Domain-Associated Retrotransposon Search*) apresentam uma sensibilidade maior para a identificação, podendo facilitar a descoberta e análise detalhada da diversidade e evolução de retrotransposons, LTR-RTs e outras ETs com codificação de proteínas, bem como possibilitar análises comparativas e filogenéticas.

3 MATERIAL E MÉTODOS

Como demonstrado no resumo fluxograma (Figura 2), a prospecção de registros disponíveis para *Manihot esculenta* Crantz se deu por meio de buscas realizadas em banco de dados primário NCBI (*National Center for Biotechnology Information*) e secundário Phytozome.

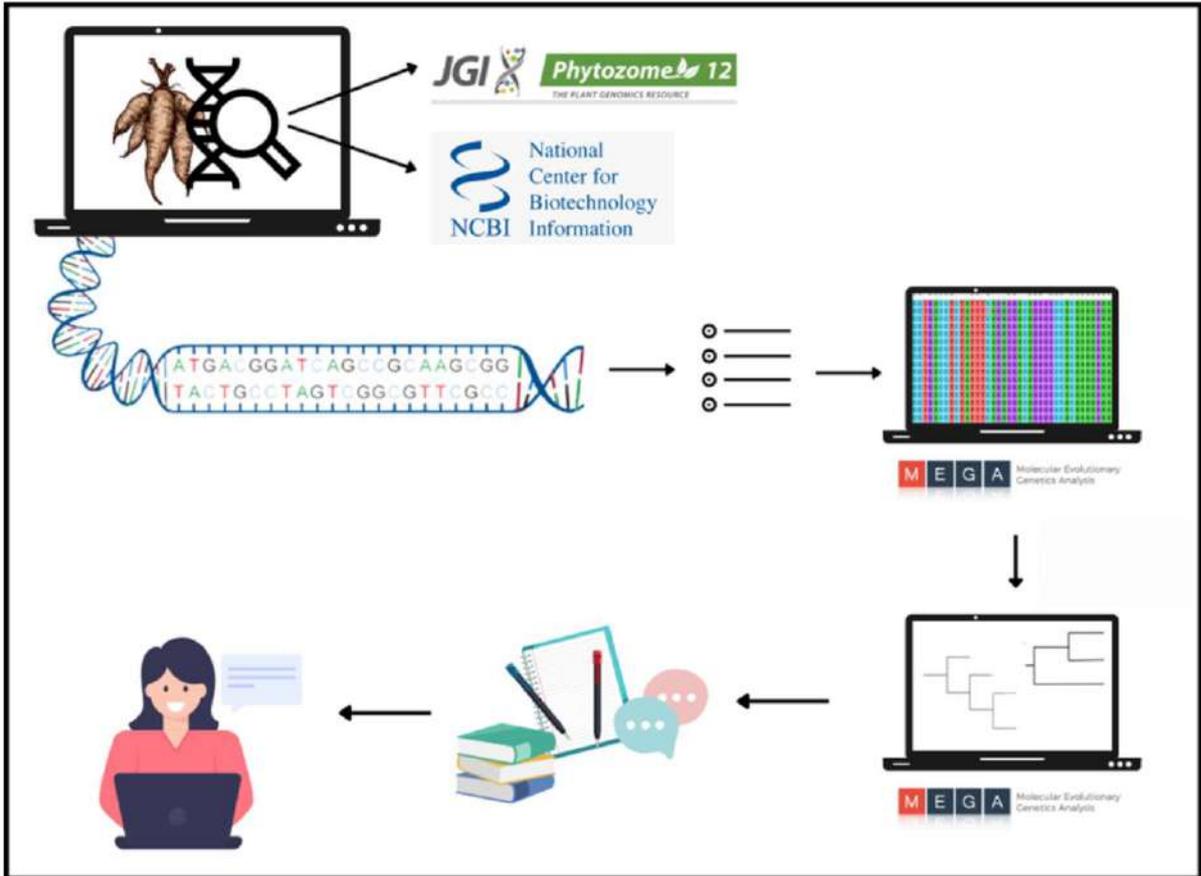
No NCBI foram obtidas 194 sequências de proteínas referente à espécie *Manihot esculenta* Crantz por meio da pesquisa do termo “retrotransposons cassava”, especificando apenas dados referentes a proteínas. Destas, foram coletadas informações como: o país de origem da sequência, o código de identificação do cultivar, o tamanho da sequência, cromossomo em que se encontra, bem como as regiões de interesse identificadas.

Para cada uma das sequências encontradas na pesquisa inicial, foram realizadas buscas por meio da comparação de sequências utilizando a ferramenta blastp, inserida na plataforma previamente utilizada, para realizar a comparação de uma sequência de proteína com uma base de dados de proteínas. Com essa prospecção, o total de sequências obtidas foi de 362 mas que, após análise das regiões apresentadas por elas, foram mantidas apenas as sequências referentes a qualquer região de retrotransposons, totalizando 61 sequências. Junto com aquelas obtidas previamente no NCBI, totalizaram-se 255 sequências, sendo que algumas delas tinham referência para o banco de dados secundário.

Posteriormente, foram selecionadas apenas as sequências referentes a região RT_LTR, totalizando em 69 após essa segunda triagem (vide Apêndice A). Os resultados foram analisados a partir desta última seleção de sequências de proteínas, ressaltando que o próprio banco de dados forneceu os dados de quais cromossomos são portadores das sequências que codificaram as proteínas analisadas.

As sequências obtidas foram alinhadas pelo algoritmo clustal W (THOMPSON et al., 1994) utilizando o software MEGA 7 (KUMAR et al., 2012), a construção das árvores gênicas pelo modelo neighbor-joining com 1000 reamostragens de bootstrap e o cálculo das distâncias p foram realizadas usando o mesmo software, para tais análises, os dados obtidos foram divididos em dois grandes grupos: “*uncharacterized protein*” e “*hypothetical protein*” com “mucin17”.

Figura 2. Resumo das etapas a serem realizadas no projeto, iniciando-se com a prospecção das sequências de *Manihot esculenta* Crantz, seguindo para alinhamento e análise de dados, bem como discussão e apresentação dos dados. Fonte: Canva.com



4 RESULTADOS E DISCUSSÃO

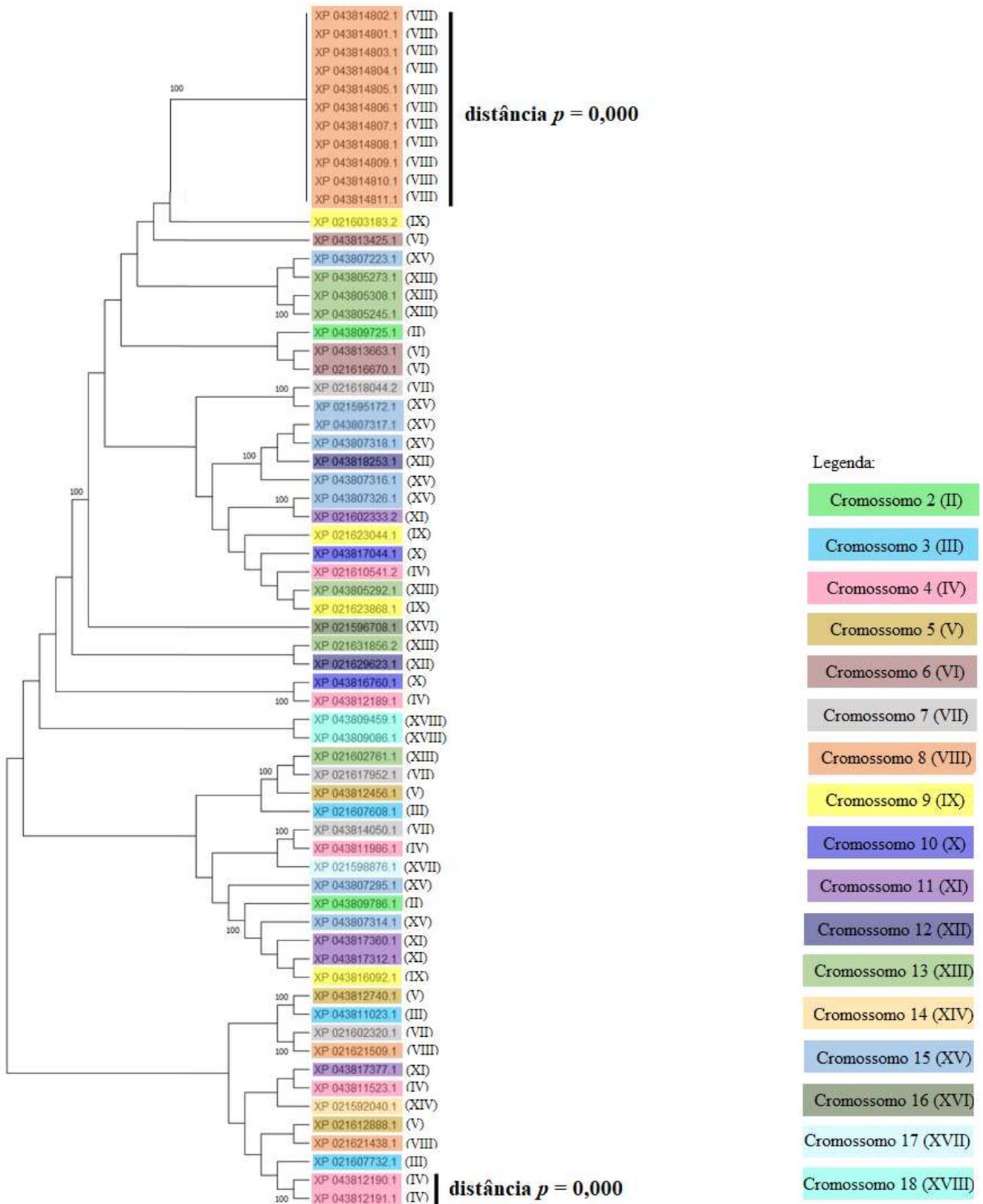
Considerando os 18 cromossomos montados disponíveis no GenBank, provenientes do BioProjeto com código de acesso “PRJNA234389” (<<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA394209>>), as análises do presente estudo resultaram na presença de RT_LTRs em 17 cromossomos distintos, ou seja, em 94,44% dos cromossomos já descritos.

Com exceção das sequências do NCBI KAG8634001.1, KAG8647619.1 e KAG8650611.1 que não possuíam informações quanto a sequência de proteínas (vide Apêndice A), as demais foram registradas a partir de sequências genômicas utilizando o método preditivo de genes Gnomon. As sequências do NCBI XP_043812190.1 e XP_043812191.1, contudo, tiveram seu método preditivo corroborado por evidências EST.

No que se refere aos genes, das 69 sequências de proteínas utilizadas, três delas não apresentaram referências para gene. Com base no que se encontra disponível no banco de dados, todos estão em status de modelo e o tipo é de codificação de proteínas. Entretanto na descrição gênica, podemos classificar todas as sequências em quatro grupos distintos. A classificação de “não caracterizados” se apresenta para 92,42% do total de sequências utilizadas no estudo, por outro lado, as sequências gênicas com algum tipo de descrição compreendem 7,58% do total. As sequências 122721509 e 122721946 apresentam como descrição “uncharacterized protein K02A2.6-like”, a sequência 110608268 apresenta como descrição “transposon Tf2-1 polyprotein”, conforme descrito no UniProt, as retransposições Tf2 pertencem aos elementos gypsy-like (metaviridae).

A árvore obtida para as sequências classificadas como “*uncharacterized protein*” no banco de dados (Figura 3) demonstrou que não houve um padrão correspondendo cromossomos e variações, podendo ser identificados apenas 2 grupos que apresentaram 100% de similaridade levando em consideração a matriz de distância calculada (Apêndice B). O primeiro grupo agrupou 11 sequências, enquanto o segundo grupo contou com apenas duas; nas demais, não pode ser observado nenhum padrão nas sequências.

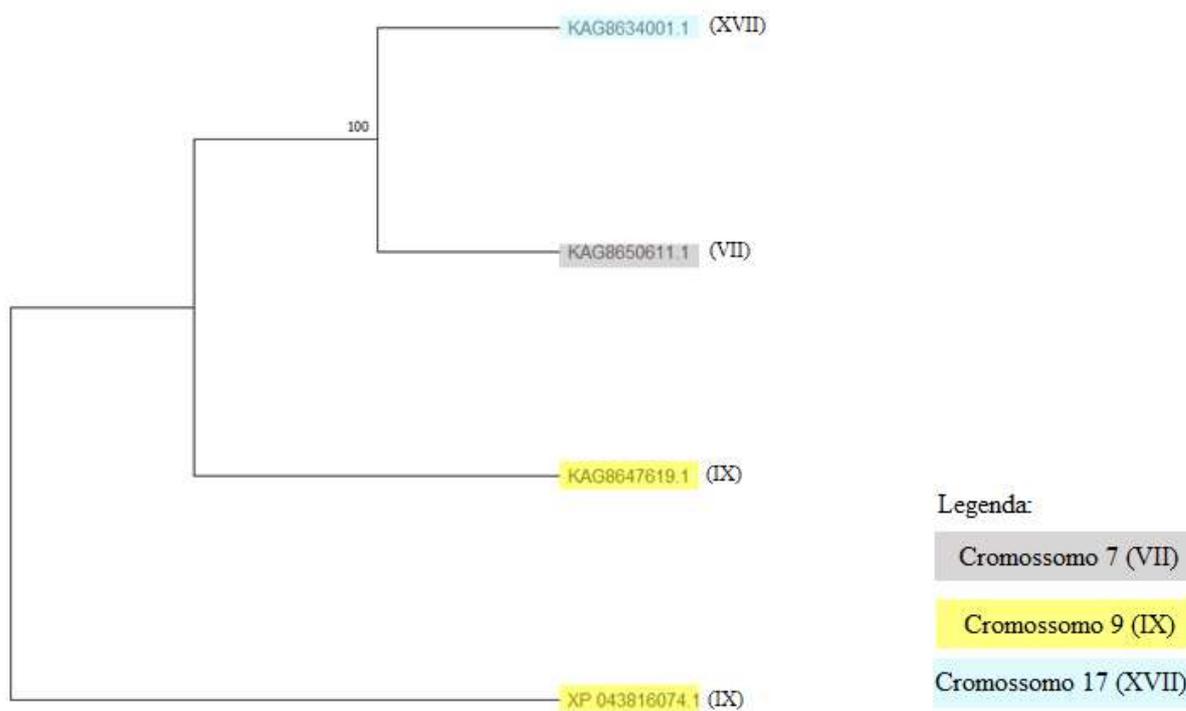
Figura 3. Árvore construída pelo método neighbor-joining com 1000 reamostragens de bootstrap para as sequências classificadas como “*uncharacterized protein*” no banco de dados. Sequências que obtiveram distância p igual a 0,000 estão destacadas na imagem, bem como as cores referentes aos cromossomos a qual cada sequência pertence. Fonte: Autor.



Embora algumas sequências de cromossomos distintos foram agrupadas no mesmo clado, podemos perceber que, para os grupos que não estão destacados na árvore, a distância p encontrada entre elas varia de 0,006 a 0,966, ficando menor que a média de todas as sequências analisadas (distância p média igual a 0,772) em 62,5% dos casos.

A árvore obtida para as sequências classificadas como “*hypothetical protein*” e “mucin-17” (Figura 4) manteve o padrão de não correspondência observado na figura 3, entretanto, a única sequência classificada como “mucin-17” (XP_043816074.1) ficou separada do restante, apesar da sua origem ser o mesmo cromossomo de sequência KAG8647619.1 e apresentou uma distância p que variou de 0,957 a 0,989 em relação às outras sequências analisadas (Apêndice C).

Figura 4. Árvore construída pelo método neighbor-joining com 1000 reamostragens de bootstrap para as sequências classificadas como “*hypothetical protein*” e “mucin-17” no banco de dados, as cores referentes aos cromossomos a qual cada sequência pertence se encontra destacada na imagem. Fonte: Autor.



No que se refere as sequências referentes a codificação de proteínas analisadas e os cromossomos as quais elas pertencem, o cromossomo 8 apresentou 18,84% do total de sequências, sendo o mais representativo nesse parâmetro, todavia, em relação ao tamanho total em pares de base (pb), ele é o segundo menor cromossomo já descrito e o menor que apresentou sequência de RT_LTRs. O cromossomo 15 foi o segundo mais representativo em sequências, totalizando 10,14% do total e em relação ao seu tamanho, ele é o quarto maior cromossomo descrito. Por fim, os cromossomos 4, 9 e 13 apresentaram 8,69% cada do total das sequências analisadas; em relação ao tamanho eles são o 12º, 15º e 13º maiores, respectivamente.

Quando comparadas as distâncias com as sequências agrupadas de acordo com o cromossomo de origem (Apêndice D), pode-se perceber que a menor distância observada foi de 0,406 entre os cromossomos 8 e 6, enquanto a maior distância foi de 0,953 entre os cromossomos 9 e 3. Na distância interna referente a cada cromossomo, não foi possível o cálculo para os cromossomos 14 e 16, devido ao fato de ter uma sequência única. O cromossomo 8 apresentou a menor distância (0,283) e foi o cromossomo com mais sequências analisadas (n=13), o que acabou sendo refletido na árvore construída, agrupando 11 das 13 sequências em um único clado. Em contrapartida, o 5 apresentou maior distância interna (0,947), o que também acabou sendo refletido na árvore, uma vez que seus três representantes ficaram dispersos pelos cladogramas construídos.

Com base na relação quantidade de sequências analisadas/ variação interna, é possível perceber que alguns cromossomos, como o 2, 10, 17 e 18 apresentaram poucas sequências, mas com uma variação muito alta, mostrando a não identidade entre elas. É possível observar também que, de maneira geral, conforme aumenta a quantidade de sequências, as distâncias internas reduzem.

É sabido que novas cópias de retrotransposons são inseridas aleatoriamente em sequências pré-existentes do genoma por meio do sistema *copy-paste*, o que faz com que os números de cópias aumentem (KALENDAR e SCHULMAN, 2006); e por meio das análises realizadas, foi possível perceber que essa migração e consequente inserção descrita pelos autores ocorre de forma aleatória, apresentando uma migração não constantes no genoma, uma vez que são notadas relações

extracromossomais entre os elementos e uma variação interna, sendo registradas poucas sequências idênticas em relação ao total avaliado.

Embora a maioria dos retrotransposons LTRs apresentem características estruturais semelhantes entre si, existe também uma considerável diversidade a ser levada em conta na sua organização, o que acaba por refletir em diferentes estratégias de proliferação dentro dos genomas hospedeiros (HAVECKER et al., 2004).

Entre os LTRs, são caracterizados dois genes responsáveis pelo ciclo de transposição do elemento: gag e pol. O primeiro codifica uma proteína similar ao capsídeo dos retrovírus, enquanto o segundo codifica um complexo poliproteico, o qual inclui os domínios protease (PR), transcriptase reversa (RT), RnaseH (RH) e integrase (INT) (FESCHOTTE et al. 2002; NOVIKOVA, 2008). Apesar da carência de estudos relacionando retrotransposons-proteínas e a alta incidência de proteínas não caracterizadas disponíveis, podemos perceber por meio dos resultados que esses elementos transponíveis estão ligados a possíveis sequências expressas, ademais, considerando a sua ampla distribuição cromossomal, podem estar relacionados com as mais diversas características dentro dos genomas. A ativação transcricional dos retrotransposons nos vegetais pode estar relacionada a diversos fatores; tem sido mostrado que a hibridação, evento que ocorre em muitos casos de forma natural, levam a ativação desses elementos (KASHKUSH et al., 2002). A maioria dos retrotransposons se encontram inativos durante a fase de desenvolvimento, enquanto algumas condições de estresse, como por exemplo, as condições de cultivo *in vitro*, devido ao estresse nutricional, químico, físico e de fotoperíodo, estimulam a atividade desses elementos (BAYRAM et al., 2012), todavia não são todas as famílias que passam a ter a retrotransposição ativada a partir de um mecanismo estressor em um mesmo indivíduo (UNGERER e KAWAKAMI 2013).

Os elementos transponíveis, constituem uma proporção substancial de genomas vegetais e contribuem para a estrutura e evolução (FESCHOTTE et al., 2002), podendo vir a servir para genômica funcional de espécies de plantas modelo e de cultura, e no caso da mandioca, podem vir a ter potencial para auxiliar na caracterização genética ou melhoramento dessa cultura (GBADEGESIN et al., 2008), inclusive Kuhn e colaboradores (2016) tiveram resultados eficazes na quantificação da diversidade da mandioca com a utilização dos *primers* desenvolvidos no estudo. Entretanto, a dinâmica do processo que os RT_LTRs apresentam nos genomas

vegetais ainda é mal compreendida devido à falta de estudos comparativos (EL BAIDOURI e PANAUD, 2013) e essa carência de estudos é ainda mais acentuada no que se trata da espécie *M. esculenta* Crantz.

5 CONSIDERAÇÕES FINAIS

O presente trabalho identificou uma migração não constante entre os cromossomos, mostrando que essa inserção não segue um padrão previsível para a espécie estudada; tais resultados são significativos para área de estudo, uma vez que proporciona conhecimentos necessários para o avanço da presente área, bem como de informações referentes ao genoma dessa espécie tão importante, especialmente no território brasileiro.

Considerando que as sequências analisadas são proteicas, portanto, elas são expressas e isso faz com que a inserção dos retrotransposons gere diferente fenótipos, acreditamos que os resultados obtidos são de grande interesse para a literatura, especialmente devido a carência de estudos para a espécie. Das 69 sequências analisadas, a descrição gênica de quase 93% delas não foi caracterizada.

Por esse mesmo motivo, acreditamos que pesquisas futuras que envolvam a elucidação de questões cromossômicas sejam de grande valia, como a porcentagem de retrotransposons para cada cromossomo no qual já foi identificada sua presença e sua participação na expressão de proteínas, assim como maiores estudos para elucidação dos mecanismos de ativação dos retrotransposons.

REFERÊNCIAS

- AJAYI, C.; OLUTUMISE. Determinants of food security and technical efficiency of cassava farmers in Ondo State, Nigeria. **International Food and Agribusiness Management Review** v.21, p. 915 – 928. 2018
- ALLEM, A.C. The origin of *Manihot esculenta* Crantz (Euphorbiaceae). **Genetic Resources and Crop Evolution** v. 41, p. 133-150. Brasília, DF, 1994.
- AUSTIN, C.P. in National Human Genome Research Institute (NIH). Talking Glossary of Genetic Terms: Bioinformatics. Disponível em <<https://www.genome.gov/genetics-glossary/Bioinformatics>>. Acesso em 24/04/2022.
- AWOLEYE, F.; VAN DUREN, M.; DOLEZEL, J.; NOVAK, F.J.; Nuclear DNA content and *in vitro* induced somatic polyploidization cassava (*Manihot esculenta* Crantz) breeding. **Euphytica**, v. 76, p. 195-202, 1994.
- BAIRU, M.W., AREMU, A.O. & VAN STADEN, J. Somaclonal variation in plants: causes and detection methods. **Plant Growth Regulation**, v.63, p. 147–173, 2011.
- BARROSO, P.A.V.; MOURA, G.E.D.D.; BRITO, L.K.F.; MARTINS, C.P.; MACEDO, C.E.C.; LOPES, D.B.; ALLOUFA, M.A.I. Efeito do cultivo *in vitro* na presença de NaCl em plantas de abacaxizeiro na fase de aclimação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 7, p. 473-477, 2003.
- BAYRAM, E.; YILMAZ, S.; HAMAT-MECBUR, H.; KARTAL-ALACAM, G.; GOZUKIRMIZI, N. Nikita retrotransposon movements in callus cultures of barley (*Hordeum vulgare* L.) **Plant Omics Journal**, 5:211-215, 2012.
- BENSON D.A.; CAVANAUGH M.; CLARK K.; KARSCH-MIZRACHI I.; OSTELL J.; PRUITT K.D.; SAYERS E.W. GenBank. **Nucleic Acids Research** 46:D41–D47, 2018.
- BIRYUKOV, M.; USTYANTSEV, K. DARTS: An Algorithm for Domain-Associated Retrotransposon Search in Genome Assemblies. **Genes**, v. 13, 2022.

BULL, S.E.; NDUNGURU, J.; GRUISSEM, W.; BEECHING, J.R.; VNDERSCHUREN, H. Cassava: constraints to production and the transfer of biotechnology to African laboratories. **Plant Cell Report**, v. 30, p. 779-787, 2011.

CARVALHO, M.J.S.; GOMES, V.B.; SOUZA, A.S.; AUD, F.F.; SANTOS-SEREJO, J.A.; OLIVEIRA, E.J. Inducing autotetraploids in cassava using oryzalin and colchicine and their *in vitro* morphophysiological effects. **Genetics and Molecular Research**, v. 15, 2016.

CEBALLOS, A.H.; SÁNCHEZ, T.; CHÁVES, A.L.; IGLESIAS, C.; DEBOUCKA, D.; MAFLA, G.; TOHMEA, J. Variation in crude protein content in cassava (*Manihot esculenta* Crantz) roots. **Journal of Food Composition and Analysis**, v. 19, p. 589-593, 2006.

CELÉDON, P.A.F. KOBAYASHI, A.K.; VIEIRA, L.G.E. Utilização de nutrientes por células de algodoeiro cultivadas em suspensão. **Scientia Agricola**, v. 57, p. 431-437, 2000.

COSTA, M. A. P. de C.; et al. Micropropagação de orquídea. In: JUNGHANS, T. G.; SOUZA, A. da S. (Ed.). Aspectos Práticos da Micropropagação de Plantas. 2. ed. Brasília -DF: **Embrapa**, p. 373-392, 2013.

DIAMANTE, N.A.; FABRIN, T.M.; SILVEIRA, M.J.; OLIVEIRA, A.V.; THOMAZ, S.M.; PRIOLLI, S.M.A.P.; PRIOLLI, A.J. 2020. Molecular analysis of the invasive populations of *Urochloa* (Poaceae) in a large Neotropical reservoir. **Aquatic Botany**, v. 161, 2020.

EL BAIDOURI, M.; PANAUD, O. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. **Genome Biol. Evol.** V. 5, p. 954–965. 2013.

FAYE, A.; DEBLAUWE, V.; MARIAC, C.; RICHARD, D.; SONKÉ, B.; VIGOUROUX, Y.; COUVREUR, T.L.P. Phylogeography of the genus *Podococcus* (Palmae/Arecaceae) in central African rain forests: Climate stability predicts unique genetic diversity. **Molecular Phylogenetics and Evolution**, v. 105, p. 126–138, 2016.

FESCHOTTE, C.; JIANG, N.; WESSLER, S.R. Plant transposable elements: where genetics meets genomics. **Nature Reviews Genetics** 3:329–341, 2002.

FONSECA, L.H.M.; LOHMANN, L.G. Exploring the potential of nuclear and mitochondrial sequencing data generated through genome-skimming for plant phylogenetics: A case study from a clade of neotropical lianas. **Journal of Systematics and Evolution**, p. 1-15, 2019.

Food And Agriculture Of The United Nations (FAO). 2000. **A new strategy for cassava**. FAO: Rome.

GBADEGESIN, M.A.; WILLS, M.A.; BEECHING, J.R. Diversity of LTR-retrotransposons and Enhancer/Suppressor Mutator-like transposons in cassava (*Manihot esculenta* Crantz). **Mol Genet Genomics**, 280(4), 305–317., 2008

GITZENDANNER, M.A.; SOLTIS, P.S.; WONG, G.K-S.; RUHFEL, B.R.; SOLTIS, D.E. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. **American Journal of Botany**, v. 105, p. 291–301, 2018.

GOODSTEIN, D.M.; SHU, S., HOWSON, R.; NEUPANE, R.; HAYES, R.D.; FAZO, J.; et al. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, v. 40, p. 1178–1186, 2012.

GOVAERTS, R., NIC LUGHADHA, E., BLACK, N. *et al.* The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. **Sci Data** v. 8. 2021.

GUIMARÃES, D.L.F; SILVA, R.N.; ANDRADE, H. M.L. da S.; ANDRADE, L.P. Cassava production chain in the Brazilian territory innovations and technologies: a systematic literature review. **Diversitas Journal**, v. 7, p. 0017-0025. 2022.

HAHN, S.K.; BAI, K.V.; ASIEDU, R. Spontaneous somatic tetraploids in cassava. **Japan Journal of Breeding**, v. 42, p. 303-308, 1992.

HAHN, S.K.; BAI, K.V.; ASIEDU, R. Tetraploids, triploids, and 2n pollen from diploid interspecific crosses with cassava. **Theoretical and Applied Genetics**, v. 79, p. 433-439, 1990.

HAVECKER, E.R., GAO, X.; VOYTAS, D.F. The diversity of LTR retrotransposons. **Genome Biol**, 5, 225, 2004.

HAYWARD, A.C.; TOLLENAERE, R.; DALTON-MORGAN, J.; BATLEY, J. Molecular marker applications in plants. In **Plant Genotyping**. Human Press, 2015.

HUANG, D.W.; SHERMAN, B.T.; LEMPICKI, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, **Nucleic Acids Research**, v. 37, p. 1–13, 2009.

IBGE, Indicadores IBGE, Produção Agrícola Municipal 2020. Rio de Janeiro: **IBGE**, 2021.

IBGE, Indicadores IBGE, Levantamento Sistemático da Produção Agrícola. Rio de Janeiro: **IBGE**, 2022.

JARNE, P.; LAGODA, P.J.L. Microsatellites, from molecules to population and back. **Trends in Ecology and Evolution**, v. 11, p. 424-429, 1996.

JUDD, W.S.; CAMPBELL, C.S.; KELLOGG, E.A.; STEVENS, P.F. Plant Systematics: A phylogenetic approach. **Sinauer Associates**, Sunderland. 1999.

KAEPLER, S.M.; KAEPLER, H.F.; RHEE, Y. Epigenetic aspects of somaclonal variation in plants. In: MATZKE, M.A.; MATZKE, A.J.M. (eds) **Plant Gene Silencing**. Springer, 59-68, 2000.

KALENDAR, R.; SCHULMAN, A.H. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. **Nat. Protoc.** 1: 2478-2484, 2006.

KALENDAR, R.; FLAVELL, A.J.; ELLIS, T.H.N.; SJAKSTE, T.; MOISY, C.; SCHULMAN, A.H. Analysis of plant diversity with retrotransposon-based molecular markers, **Heredity** 106:520–530, 2011.

KASHKUSH, K.; FELDMAN, M.; LEVY, A. A. Gene Loss, Silencing and Activation in a Newly Synthesized Wheat Allotetraploid. **Genetics**. v. 160, p. 1651 – 1659, 2002.

KUHN, B.C.; LÓPEZ-RIBERA, I.; DA SILVA MACHADO, M.D.E.F.; VICIENT, C.M. Genetic diversity of maize germplasm assessed by retrotransposon-based markers. *Electrophoresis*. **Epub**, v. 35, p. 1921-7, 2014.

KUHN, B.C.; MAGOLIN, C.A.; SOUTO, E.R.; VICIENT, C.M.; MACHADO, M.F.P.S. Development of retrotransposon-based markers IRAP and REMAP for cassava (*Manihot esculenta*). **Genetics and Molecular Research** 15 (2), 2016.

KUHN, B.C.; MANZATTI, J.S.; PEPINELLI, A.; CARNELOSSI, P.R.; SOUTO, E.R.; MAGOLIN, C.A.; MACHADO, M.F.P.S. Clonal Micropropagation of New Cassava Varieties (*Manihot esculenta* Crantz) Assisted by Inter-Retrotransposon Amplified Polymorphism and Retrotransposon-Microsatellite Amplified Polymorphism Marker. **International Journal of Current Research**, v. 9 (01), p. 45295-45301, 2017.

KUMAR, A.; BENNETZEN, J.L. Plant Retrotransposons. **Annual Review of Genetics**, v. 33(1), p. 479–532, 1999.

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. **Molecular Biology and Evolution**, 33:1870–1874, 2016.

LARKIN P.; SCOWCROFT W. Somaclonal variation—a novel source of variability from cell cultures for plant improvement. **Theor Appl Genet**, v. 60, p. 197–214, 1981.

LEITCH, I.J.; BENNET, M.D. Polyploidy in angiosperms. **Trends in Plant Science**, Oxford, v. 2, p. 470-476, 1997.

LESK, A. M. Introdução à Bioinformática. 2.ed. **Artmed**, 2008.

MARI-RIBEIRO, I.P; SCORSIM, B.; DE OLIVEIRA, A.V.; PORTELA-CASTRO, A.L.B. Cytogenetic and Molecular Characterization of *Oligosarcus pintoii* (Characidae): A New Record of Supernumerary Chromosome in this Species. **Zebrafish**, v. 19, p. 71-80, 2022.

MARSHALL, G.; COURDURIES, P. An assessment of somaclonal variation in linseed (*Linum usitatissimum*). **Annals of Applied Biology**, v. 120, p. 501-509, 1992.

MCCLINTOCK, M. The Order of the Genes C, Sh and Wx in Zea Mays with Reference to a Cytologically Known Point in the Chromosome. **Proceedings of the National Academy of Sciences USA** 17:485–491, 1931.

MORAIS-SILVA, J.P., OLIVEIRA, A.V. DE.; FABRIN, T.M.C., et al. Geomorphology influencing the diversification of fish in small-order rivers of neighboring basins. **Zebrafish** 15:389–397, 2018.

MOWER, J.P.; SLOAN, D.B.; ALVERSON, A.J. Plant mitochondrial genome diversity: The genomics revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ eds. **Plant genome diversity**. Viena: Springer, 1:123–144, 2012.

NASSAR, N. Polyploidy, chimera and fertility of interspecific cassava (*Manihot esculenta* Crantz) hybrids. **The Indian Journal of Genetics and Plant Breeding** 64:317-318, 2004.

NCBI (National Center for Biotechnology Information). Disponível em <<https://www.ncbi.nlm.nih.gov/>>

NEUMANN, P.; NOVÁK, P.; HOŠTÁKOVÁ, N. *et al.* Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. **Mobile DNA**, v. 10, 2019.

NOVIKOVA, O. et al. Novel clades of chromodomain-containing Gypsy LTR retrotransposons from mosses (Bryophyta). **The Plant Journal** 56: 562-574, 2008.

OLSEN, K.M. & SCHHAL, B.A. Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. **Proc. Natl. Acad. Sci.** v. 96, p. 5586-5591. USA 1999.

PAREDES, N.; ALULEMA, V.; LIMA, L.; SØRENSEN, M.; MONTEROS-ALTAMIRANO, A. Genetic Diversity of Cassava (*Manihot esculenta* Crantz) in Ecuador by Using SSR Markers. **Preprints**, 2021.

PHYTOZOME: The plant genomics resource. Disponível em < <https://phytozome-next.jgi.doe.gov/> >

PROSDOCIMI, F.; CERQUEIRA, G.; BINNECK, E.; SILVA, A.; NEVES, A.; REIS, D.; CAROLINA, A. JUNQUEIRA, M.; CECÍLIA, A.; DOS SANTOS, A.F.; NHANI, A.; WUST, C.; FILHO, F. C.; KESSEDJIAN, J.; PETRETSKI, J.; DE, R.; FERREIRA, G.; LIMA, R.; PEREIRA, R.; FOLGUERAS-FLATSCHART, Á. *Bioinformática: manual do usuário. Biotecnologia Ciência e Desenvolvimento* 29:12 - 25, 2012.

QI, W.; LIM, Y.; PATRIGNANI, A. SCHLÄPFER, P. BRATUS-NEUENSCHWANDER, A.; GRÜTER, S.; CHANEZ, C.; RODDE, N.; PRAT, E.; VAUTRIN, S.; FUSTIER, M.A.; PRATAS, D.; SCHLAPBACH, R.; GRUISSEM, W. The haplotype-resolved chromosome pairs and transcriptome of a heterozygous diploid African cassava cultivar. *bioRxiv*, 2021.

RAMHARACK, P.; SOLIMAN, M.E.S. Bioinformatics-based tools in drug discovery: the cartography from single gene to integrative biological networks. *Drug Discovery Today*, v. 23, p. 1658-1665, 2018.

RAMOS, A.P.; CASTAÑEDA, F.G.; HERNÁNDEZ, G.N.; PÉRE, L.M. Forage yield and quality of early maize hybrids in response to planting date, nitrogen and plant density. *Revista Fitotecnia Mexicana*, v. 29, p. 207-213, 2006.

RIDLEY, M. *Evolução* [recurso eletrônico]; tradução Henrique Ferreira, Luciane Passaglia, Rivo Fischer. – 3. ed. – Dados eletrônicos. – Porto Alegre, **Artmed**, ISBN 978-85-363-0863-0, 2007.

ROB, P; CORONEL, C. *Sistema de banco de dados: Projeto, implementação e gerenciamento. Cengage Learning*, 2014.

SANTOS, T.T.C., DE OLIVEIRA AMORIM, V.B., DOS SANTOS-SEREJO, J.A. et al. Genetic variability among autotetraploid populations of banana plants derived from wild diploids through chromosome doubling using SSR and molecular markers based on retrotransposons. *Mol Breeding*, v. 39, 2019.

SCHAACK, S.; GILBERT, C.; FESCHOTTE, C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology and Evolution* 25: 537–546, 2010.

SILVA, J.O.C.; LACERDA, G.A.; CASTRO, E.M.; CHALFUN-JÚNIOR, C. Aplicação de ácido giberélico (GA3) e anatomia da epiderme foliar visando à detecção de variantes somaclonais de bananeira *Musa* sp. Colla cv. Prata-anã (Musaceae). **Acta Botanica Brasílica**, v. 24, p. 47-52, 2010.

SILVA, M.V.B.M.; QUEIROZ-SILVA, J.R.; BRESSAN, E.A.; BORGES, A.; PEREIRA, K.J.C.; PINTO, J.G.; VEASEY, E.A. Genetic characterization of cassava (*Manihot esculenta*) landraces in Brazil assessed with simple sequence repeats. **Genetics and Molecular Biology**, v. 32, p. 104-110, 2009.

SIMPSON, M.G. Plant systematics. **Elsevier Academic Press**, Amsterdam. 2006.

SOLTIS, D.E., SOLTIS, P.S., ENDRESS, P.K. & CHASE, M.W. Phylogeny and evolution of Angiosperms. **Sinauer Associates**, Sunderland. 2005.

STRIOTO, D.K.; KUHN, B.C., NAGATA, W.S.L.; MARINELLI, G.; OLIVEIRA-COLLET, S.A.; MANGOLIN, C.A.; MACHADO, M.F.P.S. Development and use of retrotransposons-based markers (IRAP/ REMAP) to assess genetic divergence among table grape cultivars. **Plant Genetic Resources**; 1–8, 2019.

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, **Nucleic Acids Research**, v. 22, p. 4673–4680, 1994.

TWYFORD, A.D.; NESS, R.W. Strategies for complete plastid genome sequencing. **Molecular Ecology Resources**, v. 17, p. 858–868, 2017.

UNGERER, M. K. e KAWAKAMI, T. Transcriptional Dynamics of LTR Retrotransposons in Early Generation and Ancient Sunflower Hybrids. **Genome Biology and Evolution**. v. 5, n. 2, p. 329 – 337, 2013.

UniProt (UniProt: the universal protein knowledgebase). Disponível em: <<https://www.uniprot.org/>>

VENNER, S.; FESCHOTTE, C.; BIÉMONT, C. Dynamics of transpos a community ecology of the genome. **Trends in Genetics** 25: 317-323, 2009.

VIEIRA, M. L. C. Conservação de germoplasma in vitro. *Botecnologia Ciência & Desenvolvimento*, Brasília, DF, v. 14, p. 18-20, 2000.

WEBSTER, G.L. Systematics of the Euphorbiaceae. **Annals of the Missouri Botanical Garden**, v. 81, p. 1-144, 1994.

WHANKAEW, S.; KANJANAWATTANAWONG, S.; PHUMICHA, C.; SMITH, D.R.; NARAGAJAVANA, J.; TRIWITAYAKORN, K. Cross-genera transferability of (simple sequence repeat) SSR markers among cassava (*Manihot esculenta* Crantz), rubber tree (*Hevea brasiliensis* Muell. Arg.) and physic nut (*Jatropha curcas* L.). **African Journal of Biotechnology**, v. 10, p. 1768-1776, 2011.

XIONG, F.; LIU, J.; TANG, R.; YANG, T.; YANG, X.; HE, L.; HAN, Z.; QIU, L.; ZOU, C.; TANG, Z.; LUO, C.; ZHONG, R. JIANG, J.; HUANG, Z.; WU, H., LIU,, J.; HE, X. Exon based amplified polymorphism (EBAP): A novel and universal molecular marker for plants. **Electronic Journal of Biotechnology**, v. 56, p. 65-74, 2022.

APÊNDICE A - Relação das sequências utilizadas nas análises do presente trabalho, após as triagens realizadas.

Gene ID	Anotação referente ao gene	NCBI ref. sequence	Anotação referente a proteína	Cromossomo	Tamanho total	Observação
---	---	KAG8634001.1	---	17	1745 aa	hypothetical protein
---	---	KAG8647619.1	---	9	1833 aa	hypothetical protein
---	---	KAG8650611.1	---	7	1485 aa	hypothetical protein
110602059	Símbolo gênico: LOC110602059 Descrição gênica: uncharacterized LOC110602059	XP_043807317.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035175.2) registrado utilizando o método preditivo de genes Gnomon.	15	631 aa	uncharacterized protein
110606868	Símbolo gênico: LOC110606868 Descrição gênica: uncharacterized LOC110606868	XP_043809459.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035178.2) registrado utilizando o método preditivo de genes Gnomon.	18	532 aa	uncharacterized protein
110606254	Símbolo gênico: LOC110606254	XP_043809086.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência	18	1024 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC110606254		genômica (<u>NC_035178.2</u>) registrado utilizando o método preditivo de genes Gnomon.			
122721970	Símbolo gênico: LOC122721970 Descrição gênica: uncharacterized LOC122721970	XP_043807326.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035175.2</u>) registrado utilizando o método preditivo de genes Gnomon.	15	814 aa	uncharacterized protein
122721963	Símbolo gênico: LOC122721963 Descrição gênica: uncharacterized LOC122721963	XP_043807318.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035175.2</u>) registrado utilizando o método preditivo de genes Gnomon.	15	415 aa	uncharacterized protein
122721962	Símbolo gênico: LOC122721962 Descrição gênica: uncharacterized LOC122721962	XP_043807316.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035175.2</u>) registrado utilizando o método preditivo de genes Gnomon.	15	605 aa	uncharacterized protein
122721960	Símbolo gênico: LOC122721960	XP_043807314.1	O dado é predito por análise computacional. Esse registro	15	996 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC122721960		é derivado de uma sequência genômica (<u>NC_035175.2</u>) registrado utilizando o método preditivo de genes Gnomon.			
122721946	Símbolo gênico: LOC122721946 Descrição gênica: uncharacterized protein K02A2.6-like	XP_043807295.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035175.2</u>) registrado utilizando o método preditivo de genes Gnomon.	15	795 aa	uncharacterized protein
122721939	Símbolo gênico: LOC122721939 Descrição gênica: uncharacterized LOC122721939	XP_043807223.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035175.2</u>) registrado utilizando o método preditivo de genes Gnomon.	15	719 aa	uncharacterized protein
110608243	Símbolo gênico: LOC110608243 Descrição gênica: uncharacterized LOC110608243	XP_043805308.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035173.2</u>) registrado utilizando o método preditivo de genes Gnomon.	13	1353 aa	uncharacterized protein

110629250	Símbolo gênico: LOC110629250 Descrição gênica: uncharacterized LOC110629250	XP_021631856.2	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035173.2</u>) registrado utilizando o método preditivo de genes Gnomon.	13	896 aa	uncharacterized protein
122721509	Símbolo gênico: LOC122721509 Descrição gênica: uncharacterized protein K02A2.6-like	XP_043805292.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035173.2</u>) registrado utilizando o método preditivo de genes Gnomon.	13	933 aa	uncharacterized protein
122721498	Símbolo gênico: LOC122721498 Descrição gênica: uncharacterized LOC122721498	XP_043805273.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035173.2</u>) registrado utilizando o método preditivo de genes Gnomon.	13	1041 aa	uncharacterized protein
122721475	Símbolo gênico: LOC122721475 Descrição gênica: uncharacterized LOC122721475	XP_043805245.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035173.2</u>) registrado utilizando o	13	766 aa	uncharacterized protein

			método preditivo de genes Gnomon.			
122725281	Símbolo gênico: LOC122725281 Descrição gênica: uncharacterized LOC122725281	XP_043818253.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035172.2</u>) registrado utilizando o método preditivo de genes Gnomon.	12	659 aa	uncharacterized protein
110625649	Símbolo gênico: LOC110625649 Descrição gênica: uncharacterized LOC110625649	XP_043817377.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035171.2</u>) registrado utilizando o método preditivo de genes Gnomon.	11	889 aa	uncharacterized protein
110607517	Símbolo gênico: LOC110607517 Descrição gênica: uncharacterized LOC110607517	XP_021602333.2	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035171.2</u>) registrado utilizando o método preditivo de genes Gnomon.	11	978 aa	uncharacterized protein
122725014	Símbolo gênico: LOC122725014	XP_043817360.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035171.2</u>)	11	903 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC122725014		registrado utilizando o método preditivo de genes Gnomon.			
122724981	Símbolo gênico: LOC122724981 Descrição gênica: uncharacterized LOC122724981	XP_043817312.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035171.2</u>) registrado utilizando o método preditivo de genes Gnomon.	11	1393 aa	uncharacterized protein
122724910	Símbolo gênico: LOC122724910 Descrição gênica: uncharacterized LOC122724910	XP_043817044.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica(<u>NC_035170.2</u>) registrado utilizando o método preditivo de genes Gnomon.	10	648 aa	uncharacterized protein
122724880	Símbolo gênico: LOC122724880 Descrição gênica: uncharacterized LOC122724880	XP_043816760.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica(<u>NC_035170.2</u>) registrado utilizando o método preditivo de genes Gnomon.	10	1347 aa	uncharacterized protein
110608278	Símbolo gênico: LOC110608278	XP_021603183.2	O dado é predito por análise computacional. Esse registro é derivado de uma sequência	9	895 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC110608278		genômica(NC_035169.2) registrado utilizando o método preditivo de genes Gnomon.			
122724626	Símbolo gênico: LOC122724626 Descrição gênica: uncharacterized LOC122724626	XP_043816092.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica(NC_035169.2) registrado utilizando o método preditivo de genes Gnomon.	9	1190 aa	uncharacterized protein
110622377	Símbolo gênico: LOC110622377 Descrição gênica: mucin-17	XP_043816074.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica(NC_035169.2) registrado utilizando o método preditivo de genes Gnomon.	9	669 aa	mucin-17
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814811.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035168.2) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein

122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814810.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035168.2</u>) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814809.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035168.2</u>) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814808.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035168.2</u>) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814807.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035168.2</u>) registrado utilizando o	8	848 aa	uncharacterized protein

			método preditivo de genes Gnomon.			
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814806.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035168.2</u>) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814805.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035168.2</u>) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814804.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035168.2</u>) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein
122724336	Símbolo gênico: LOC122724336	XP_043814803.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035168.2</u>)	8	848 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC122724336		registrado utilizando o método preditivo de genes Gnomon.			
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814802.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035168.2</u>) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein
122724336	Símbolo gênico: LOC122724336 Descrição gênica: uncharacterized LOC122724336	XP_043814801.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035168.2</u>) registrado utilizando o método preditivo de genes Gnomon.	8	848 aa	uncharacterized protein
110619100	Símbolo gênico: LOC110619100 Descrição gênica: uncharacterized LOC110619100	XP_021618044.2	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035167.2</u>) registrado utilizando o método preditivo de genes Gnomon.	7	341 aa	uncharacterized protein
122723977	Símbolo gênico: LOC122723977	XP_043814050.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência	7	1174 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC122723977		genômica (<u>NC 035167.2</u>) registrado utilizando o método preditivo de genes Gnomon.			
110612552	Símbolo gênico: LOC110612552 Descrição gênica: uncharacterized LOC110612552	XP_043813663.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035166.2</u>) registrado utilizando o método preditivo de genes Gnomon.	6	905 aa	uncharacterized protein
122721247	Símbolo gênico: LOC122721247 Descrição gênica: uncharacterized LOC122721247	XP_043813425.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035166.2</u>) registrado utilizando o método preditivo de genes Gnomon.	6	649 aa	uncharacterized protein
110608389	Símbolo gênico: LOC110608389 Descrição gênica: uncharacterized LOC110608389	XP_043812740.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035165.2</u>) registrado utilizando o método preditivo de genes Gnomon.	5	682 aa	uncharacterized protein
122723651	Símbolo gênico: LOC122723651	XP_043812456.1	O dado é predito por análise computacional. Esse registro	5	475 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC122723651		é derivado de uma sequência genômica (NC_035165.2) registrado utilizando o método preditivo de genes Gnomon.			
110608268	Símbolo gênico: LOC110608268 Descrição gênica: transposon Tf2-1 polyprotein	XP_043812190.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035164.2) registrado utilizando o método preditivo de genes Gnomon, corroborado por evidência EST	4	718 aa	uncharacterized protein
110608268	Símbolo gênico: LOC110608268 Descrição gênica: transposon Tf2-1 polyprotein	XP_043812191.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035164.2) registrado utilizando o método preditivo de genes Gnomon, corroborado por evidência EST	4	718 aa	uncharacterized protein
122723569	Símbolo gênico: LOC122723569 Descrição gênica: uncharacterized LOC122723569	XP_043812189.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035164.2) registrado utilizando o método preditivo de genes Gnomon.	4	1206 aa	uncharacterized protein

122723547	Símbolo gênico: LOC122723547 Descrição gênica: uncharacterized LOC122723547	XP_043811986.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035164.2</u>) registrado utilizando o método preditivo de genes Gnomon.	4	1146 aa	uncharacterized protein
110613627	Símbolo gênico: LOC110613627 Descrição gênica: uncharacterized LOC110613627	XP_021610541.2	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035164.2</u>) registrado utilizando o método preditivo de genes Gnomon.	4	616 aa	uncharacterized protein
122723442	Símbolo gênico: LOC122723442 Descrição gênica: uncharacterized LOC122723442	XP_043811523.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035164.2</u>) registrado utilizando o método preditivo de genes Gnomon.	4	111 aa	uncharacterized protein
122723300	Símbolo gênico: LOC122723300 Descrição gênica: uncharacterized LOC122723300	XP_043811023.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035163.2</u>) registrado utilizando o	3	593 aa	uncharacterized protein

			método preditivo de genes Gnomon.			
122723094	Símbolo gênico: LOC122723094 Descrição gênica: uncharacterized LOC122723094	XP_043809786.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035162.2</u>) registrado utilizando o método preditivo de genes Gnomon.	2	1028 aa	uncharacterized protein
122723054	Símbolo gênico: LOC122723054 Descrição gênica: uncharacterized LOC122723054	XP_043809725.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035162.2</u>) registrado utilizando o método preditivo de genes Gnomon.	2	684 aa	uncharacterized protein
110604894	Símbolo gênico: LOC110604894 Descrição gênica: uncharacterized LOC110604894	XP_021598876.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035177.2</u>) registrado utilizando o método preditivo de genes Gnomon.	17	1140 aa	uncharacterized protein
110603322	Símbolo gênico: LOC110603322	XP_021596708.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035176.2</u>)	16	276 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC110603322		registrado utilizando o método preditivo de genes Gnomon.			
110611653	Símbolo gênico: LOC110611653 Descrição gênica: uncharacterized LOC110611653	XP_021607732.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035163.2</u>) registrado utilizando o método preditivo de genes Gnomon.	3	590 aa	uncharacterized protein
110611529	Símbolo gênico: LOC110611529 Descrição gênica: uncharacterized LOC110611529	XP_021607608.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035163.2</u>) registrado utilizando o método preditivo de genes Gnomon.	3	796 aa	uncharacterized protein
110607906	Símbolo gênico: LOC110607906 Descrição gênica: uncharacterized LOC110607906	XP_021602761.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035173.2</u>) registrado utilizando o método preditivo de genes Gnomon.	13	733 aa	uncharacterized protein
110607500	Símbolo gênico: LOC110607500	XP_021602320.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência	7	465 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC110607500		genômica (<u>NC 035167.2</u>) registrado utilizando o método preditivo de genes Gnomon.			
110618996	Símbolo gênico: LOC110618996 Descrição gênica: uncharacterized LOC110618996	XP_021617952.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035167.2</u>) registrado utilizando o método preditivo de genes Gnomon.	7	729 aa	uncharacterized protein
110617986	Símbolo gênico: LOC110617986 Descrição gênica: uncharacterized LOC110617986	XP_021616670.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035166.2</u>) registrado utilizando o método preditivo de genes Gnomon.	6	612 aa	uncharacterized protein
110615353	Símbolo gênico: LOC110615353 Descrição gênica: uncharacterized LOC110615353	XP_021612888.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC 035165.2</u>) registrado utilizando o método preditivo de genes Gnomon.	5	172 aa	uncharacterized protein
110621463	Símbolo gênico: LOC110621463	XP_021621438.1	O dado é predito por análise computacional. Esse registro	8	604 aa	uncharacterized protein

	Descrição gênica: uncharacterized LOC110621463		é derivado de uma sequência genômica (NC_035168.2) registrado utilizando o método preditivo de genes Gnomon.			
110621543	Símbolo gênico: LOC110621543 Descrição gênica: uncharacterized LOC110621543	XP_021621509.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035168.2) registrado utilizando o método preditivo de genes Gnomon.	8	219 aa	uncharacterized protein
110622743	Símbolo gênico: LOC110621543 Descrição gênica: uncharacterized LOC110621543	XP_021623044.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035169.2) registrado utilizando o método preditivo de genes Gnomon.	9	730 aa	uncharacterized protein
110623248	Símbolo gênico: LOC110623248 Descrição gênica: uncharacterized LOC110623248	XP_021623868.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (NC_035169.2) registrado utilizando o método preditivo de genes Gnomon.	9	627 aa	uncharacterized protein

110627583	Símbolo gênico: LOC110627583 Descrição gênica: uncharacterized LOC110627583	XP_021629623.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035172.2</u>) registrado utilizando o método preditivo de genes Gnomon.	12	790 aa	uncharacterized protein
110599777	Símbolo gênico: LOC110599777 Descrição gênica: uncharacterized LOC110599777	XP_021592040.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035174.2</u>) registrado utilizando o método preditivo de genes Gnomon.	14	645 aa	uncharacterized protein
110602054	Símbolo gênico: LOC110602054 Descrição gênica: uncharacterized LOC110602054	XP_021595172.1	O dado é predito por análise computacional. Esse registro é derivado de uma sequência genômica (<u>NC_035175.2</u>) registrado utilizando o método preditivo de genes Gnomon.	15	514 aa	uncharacterized protein

*Todas as sequências são provenientes da Colômbia e possuem o mesmo Cultivar AM560-2.

APÊNDICE B. Distância p referente as sequências classificadas como “*uncharacterized protein*” no banco de dados

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. XP_043807317.1																														
2. XP_043809459.1	0,936																													
3. XP_043809086.1	0,944	0,903																												
4. XP_043807326.1	0,435	0,952	0,931																											
5. XP_043807318.1	0,007	0,936	0,944	0,389																										
6. XP_043807316.1	0,115	0,944	0,944	0,403	0,115																									
7. XP_043807314.1	0,910	0,968	0,986	0,921	0,924	0,928																								
8. XP_043807295.1	0,938	0,944	0,931	0,927	0,938	0,928	0,893																							
9. XP_043807223.1	0,443	0,986	0,943	0,343	0,443	0,457	0,900	0,929																						
10. XP_043805308.1	0,656	0,952	0,917	0,613	0,618	0,612	0,939	0,951	0,143																					
11. XP_021631856.2	0,808	0,942	0,958	0,779	0,808	0,817	0,923	0,923	0,657	0,740																				
12. XP_043805292.1	0,475	0,944	0,931	0,345	0,438	0,468	0,910	0,921	0,429	0,656	0,837																			
13. XP_043805273.1	0,605	0,960	0,931	0,554	0,521	0,504	0,907	0,933	0,057	0,485	0,760	0,621																		
14. XP_043805245.1	0,489	0,968	0,931	0,426	0,489	0,500	0,904	0,947	0,143	0,053	0,734	0,457	0,191																	
15. XP_043818253.1	0,027	0,936	0,944	0,399	0,014	0,122	0,919	0,939	0,443	0,628	0,808	0,439	0,534	0,489																
16. XP_043817377.1	0,949	0,928	0,889	0,956	0,949	0,956	0,920	0,934	0,986	0,942	0,962	0,971	0,956	0,957	0,949															
17. XP_021602333.2	0,435	0,952	0,931	0,006	0,389	0,403	0,921	0,927	0,343	0,613	0,779	0,345	0,548	0,426	0,399	0,956														
18. XP_043817360.1	0,904	0,968	0,986	0,915	0,917	0,921	0,015	0,893	0,900	0,933	0,923	0,904	0,907	0,904	0,912	0,912	0,915													
19. XP_043817312.1	0,911	0,968	0,986	0,923	0,924	0,928	0,018	0,888	0,900	0,933	0,923	0,911	0,905	0,904	0,919	0,920	0,923	0,012												
20. XP_043817044.1	0,465	0,952	0,903	0,419	0,465	0,504	0,938	0,930	0,457	0,589	0,827	0,209	0,496	0,489	0,465	0,953	0,411	0,938	0,938											
21. XP_043816760.1	0,911	0,920	0,931	0,911	0,910	0,914	0,943	0,937	0,914	0,918	0,904	0,937	0,905	0,936	0,919	0,956	0,911	0,937	0,937	0,922										
22. XP_021603183.2	0,435	0,960	0,931	0,379	0,438	0,439	0,898	0,960	0,100	0,466	0,740	0,458	0,424	0,149	0,446	0,949	0,379	0,893	0,899	0,488	0,943									
23. XP_043816092.1	0,910	0,968	0,986	0,921	0,924	0,928	0,010	0,893	0,900	0,933	0,923	0,910	0,907	0,904	0,919	0,920	0,921	0,005	0,006	0,938	0,937	0,898								
24. XP_043814811.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904							
25. XP_043814810.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
26. XP_043814809.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
27. XP_043814808.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
28. XP_043814807.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
29. XP_043814806.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
30. XP_043814805.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
31. XP_043814804.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
32. XP_043814803.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
33. XP_043814802.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
34. XP_043814801.1	0,435	0,960	0,944	0,379	0,431	0,432	0,904	0,949	0,014	0,472	0,740	0,441	0,373	0,149	0,439	0,971	0,379	0,898	0,905	0,481	0,930	0,102	0,904	0,000	0,000	0,000	0,000	0,000	0,000	0,000
35. XP_021618044.2	0,505	0,968	0,931	0,453	0,505	0,526	0,926	0,905	0,414	0,568	0,811	0,537	0,558	0,574	0,505	0,968	0,453	0,926	0,926	0,568	0,937	0,568	0,926	0,547	0,547	0,547	0,547	0,547	0,547	0,547
36. XP_043814050.1	0,943	0,967	0,986	0,935	0,943	0,951	0,967	0,927	0,914	0,935	0,942	0,943	0,943	0,926	0,943	0,951	0,935	0,967	0,967	0,927	0,967	0,911	0,967	0,927	0,927	0,927	0,927	0,927	0,927	
37. XP_043813663.1	0,480	0,952	0,889	0,469	0,465	0,482	0,910	0,949	0,314	0,577	0,827	0,469	0,548	0,330	0,473	0,956	0,469	0,904	0,911	0,512	0,905	0,401	0,910	0,384	0,384	0,384	0,384	0,384	0,384	
38. XP_043813425.1	0,435	0,952	0,931	0,401	0,451	0,460	0,887	0,960	0,129	0,417	0,740	0,446	0,435	0,053	0,459	0,949	0,401	0,881	0,888	0,512	0,924	0,175	0,887	0,181	0,181	0,181	0,181	0,181	0,181	
39. XP_043812740.1	0,927	0,939	0,944	0,902	0,927	0,951	0,902	0,951	0,943	0,939	0,939	0,939	0,927	0,939	0,927	0,939	0,902	0,902	0,902	0,939	0,963	0,939	0,902	0,951	0,951	0,951	0,951	0,951	0,951	
40. XP_043812456.1	0,929	0,970	0,972	0,889	0,929	0,939	0,960	0,919	0,900	0,929	0,929	0,899	0,919	0,936	0,929	0,939	0,889	0,960	0,960	0,909	0,949	0,919	0,960	0,919	0,919	0,919	0,919	0,919	0,919	
41. XP_043812190.1	0,898	0,896	0,986	0,915	0,889	0,892	0,911	0,949	0,814	0,908	0,885	0,887	0,916	0,872	0,892	0,927	0,915	0,911	0,911	0,884	0,949	0,881	0,911	0,887	0,887	0,887	0,887	0,887	0,887	
42. XP_043812191.1	0,898	0,896	0,986	0,915	0,889	0,892	0,911	0,949	0,814	0,908	0,885	0,887	0,916	0,872	0,892	0,927	0,915	0,911	0,911	0,884	0,949	0,881	0,911	0,887	0,887	0,887	0,887	0,887	0,887	
43. XP_043812189.1	0,911	0,920	0,931	0,911	0,910	0,914	0,943	0,937	0,914	0,918	0,904	0,937	0,905	0,936	0,919	0,956	0,911	0,937	0,937	0,922	0,013	0,943	0,937	0,930	0,930	0,930	0,930	0,930	0,930	
44. XP_043811986.1	0,951	0,967	0,986	0,943	0,951	0,951	0,959	0,919	0,914	0,943	0,942	0,943	0,935	0,926	0,951	0,959	0,943	0,959	0,959	0,927	0,959	0,927	0,959	0,943	0,943	0,943	0,943	0,943	0,943	
45. XP_021610541.2	0,403	0,960	0,903	0,382	0,403	0,439	0,924	0,917	0,443	0,618	0,817	0,139	0,556	0,479	0,403	0,949	0,382	0,917	0,924	0,163	0,931	0,431	0,924	0,417	0,417	0,417	0,417	0,417	0,417	
46. XP_043811523.1	0,946	0,955	0,958	0,937	0,946	0,946	0,928	0,991	0,986	0,964	0,971	0,955	0,946	0,979																

31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64

0,000
0,000 0,000
0,000 0,000 0,000
0,547 0,547 0,547 0,547
0,927 0,927 0,927 0,927 0,937
0,384 0,384 0,384 0,384 0,589 0,911
0,181 0,181 0,181 0,181 0,558 0,927 0,384
0,951 0,951 0,951 0,951 0,927 0,902 0,927 0,939
0,919 0,919 0,919 0,919 0,905 0,970 0,939 0,929 0,976
0,887 0,887 0,887 0,887 0,853 0,967 0,910 0,887 0,976 0,929
0,887 0,887 0,887 0,887 0,853 0,967 0,910 0,887 0,976 0,929 0,000
0,930 0,930 0,930 0,930 0,937 0,967 0,905 0,924 0,963 0,949 0,949 0,949
0,943 0,943 0,943 0,943 0,937 0,106 0,919 0,935 0,890 0,970 0,959 0,959 0,959
0,417 0,417 0,417 0,417 0,558 0,935 0,458 0,451 0,939 0,919 0,882 0,882 0,931 0,935
0,955 0,955 0,955 0,955 0,979 0,946 0,946 0,955 0,951 0,929 0,937 0,937 0,910 0,955 0,964
0,927 0,927 0,927 0,927 0,927 0,915 0,915 0,939 0,134 0,963 0,968 0,968 0,976 0,902 0,902 0,939
0,904 0,904 0,904 0,904 0,926 0,959 0,910 0,887 0,902 0,960 0,911 0,911 0,937 0,951 0,924 0,928 0,927
0,441 0,441 0,441 0,441 0,613 0,914 0,398 0,473 0,927 0,925 0,892 0,892 0,957 0,914 0,538 0,957 0,915 0,935
0,981 0,981 0,981 0,981 0,958 0,903 0,971 0,981 0,951 0,909 0,942 0,942 0,942 0,893 0,961 0,942 0,939 0,951 0,989
0,598 0,598 0,598 0,598 0,621 0,910 0,516 0,607 0,902 0,949 0,902 0,902 0,885 0,910 0,672 0,946 0,902 0,943 0,484 0,990
0,978 0,978 0,978 0,978 0,958 0,976 0,985 0,985 0,963 0,939 0,867 0,867 0,956 0,967 0,970 0,955 0,951 0,978 0,968 0,961 0,984
0,940 0,940 0,940 0,940 0,947 0,967 0,952 0,928 0,939 0,929 0,946 0,946 0,943 0,959 0,944 0,946 0,927 0,946 0,957 0,913 0,984 0,948
0,930 0,930 0,930 0,930 0,895 0,960 0,950 0,940 0,976 0,131 0,920 0,920 0,940 0,960 0,940 0,960 0,951 0,960 0,935 0,900 0,940 0,930 0,930
0,950 0,950 0,950 0,950 0,979 0,966 0,958 0,950 0,927 0,960 0,966 0,966 0,950 0,966 0,958 0,946 0,927 0,933 0,989 0,932 0,975 0,924 0,933 0,950
0,920 0,920 0,920 0,920 0,884 0,960 0,940 0,930 0,976 0,121 0,910 0,910 0,940 0,960 0,940 0,960 0,976 0,960 0,925 0,890 0,930 0,920 0,930 0,090 0,950
0,350 0,350 0,350 0,350 0,611 0,927 0,288 0,401 0,902 0,919 0,921 0,921 0,899 0,935 0,500 0,937 0,927 0,927 0,312 0,971 0,574 0,963 0,940 0,930 0,958 0,920
0,917 0,917 0,917 0,917 0,917 0,952 0,917 0,952 0,915 0,952 0,929 0,929 0,940 0,940 0,917 0,929 0,902 0,952 0,917 0,976 0,917 0,929 0,964 0,952 0,940 0,952 0,893
0,966 0,966 0,966 0,966 0,958 0,935 0,960 0,960 0,939 0,960 0,944 0,944 0,968 0,943 0,951 0,964 0,939 0,941 0,946 0,971 0,959 0,911 0,970 0,950 0,966 0,960 0,960 0,893
0,959 0,959 0,959 0,959 0,932 0,973 0,959 0,946 0,919 0,959 0,973 0,973 0,946 0,973 0,986 0,946 0,919 0,932 0,973 0,959 0,973 0,892 0,932 0,946 0,230 0,946 0,973 0,946 0,959
0,396 0,396 0,396 0,396 0,547 0,938 0,479 0,417 0,915 0,885 0,844 0,844 0,948 0,938 0,292 0,958 0,902 0,927 0,484 0,969 0,594 0,948 0,948 0,896 0,979 0,906 0,438 0,929 0,958 0,986
0,429 0,429 0,429 0,429 0,476 0,940 0,488 0,464 0,939 0,893 0,845 0,845 0,940 0,940 0,131 0,976 0,915 0,893 0,548 0,952 0,524 0,952 0,976 0,905 0,988 0,905 0,512 0,905 0,952 0,973 0,298
0,839 0,839 0,839 0,839 0,874 0,943 0,890 0,839 0,915 0,919 0,910 0,910 0,916 0,927 0,868 0,955 0,927 0,923 0,828 0,951 0,869 0,941 0,935 0,920 0,941 0,920 0,890 0,905 0,942 0,959 0,823 0,810
0,951 0,951 0,951 0,951 0,947 0,932 0,971 0,951 0,927 0,949 0,932 0,932 0,951 0,932 0,961 0,932 0,927 0,990 0,957 0,932 0,971 0,951 0,942 0,950 0,922 0,950 0,971 0,952 0,951 0,973 0,958 0,988 0,942
0,471 0,471 0,471 0,471 0,171 0,914 0,457 0,457 0,900 0,871 0,800 0,800 0,929 0,914 0,414 0,971 0,900 0,900 0,557 0,971 0,514 0,957 0,957 0,900 0,971 0,900 0,529 0,914 0,957 0,914 0,400 0,414 0,829 0,957

APÊNDICE C - Distância p referente as sequências classificadas como “*hypothetical protein*” e “mucin-17” no banco de dados.

	KAG8647619.1	KAG8634001.1	KAG8650611.1
KAG8647619.1			
KAG8634001.1	0,935		
KAG8650611.1	0,990	0,396	
XP_043816074.1	0,967	0,989	0,957

APÊNDICE D - Distância p referente as sequências agrupadas de acordo com os cromossomos de origem.

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Dentro de cada cromossomo	Quant. de seqüências
2																	0,935	02
3	0,948																0,942	03
4	0,893	0,942															0,873	06
5	0,930	0,852	0,939														0,947	03
6	0,651	0,948	0,844	0,924													0,357	03
7	0,916	0,941	0,901	0,885	0,869												0,946	05
8	0,715	0,945	0,855	0,930	0,406	0,862											0,283	13
9	0,737	0,953	0,877	0,933	0,672	0,902	0,668										0,827	06
10	0,845	0,951	0,787	0,938	0,718	0,908	0,745	0,781									0,922	02
11	0,650	0,947	0,904	0,933	0,802	0,925	0,812	0,782	0,873								0,773	04
12	0,810	0,938	0,874	0,921	0,670	0,893	0,688	0,782	0,786	0,861							0,898	02
13	0,770	0,944	0,869	0,890	0,582	0,858	0,584	0,734	0,757	0,852	0,740						0,678	06
14	0,974	0,940	0,940	0,943	0,964	0,938	0,953	0,973	0,956	0,966	0,961	0,965					n/c	01
15	0,711	0,945	0,857	0,925	0,564	0,863	0,579	0,716	0,756	0,749	0,666	0,673	0,969				0,664	08
16	0,713	0,957	0,869	0,923	0,566	0,877	0,655	0,762	0,770	0,881	0,807	0,687	0,971	0,704			n/c	01
17	0,973	0,933	0,945	0,936	0,967	0,878	0,967	0,956	0,954	0,963	0,952	0,957	0,942	0,949	0,971		0,902	02
18	0,892	0,939	0,908	0,942	0,782	0,918	0,805	0,864	0,847	0,901	0,781	0,852	0,962	0,789	0,866	0,939	0,927	02

*n/c: não foi possível a realização do cálculo