



## O USO DE FERRAMENTAS DE BIOINFORMÁTICA PARA ANÁLISE DA ILEX PARAGUARIENSIS (ERVA-MATE): SIMILARIDADES E FILOGENIA COM A CAMELLIA SINENSIS (CHÁ-VERDE), COFFEA ARABICA, COFFEA CANEPHORA (CAFÉ) E CITRUS SINENSIS (LARANJA)

ROBSON SOARES SILVA; ESTER SANTIAGO DA COSTA; FERNANDO AUGUSTO COSTA PEREIRA

### RESUMO

Nesta pesquisa foi utilizada técnicas de análise em sílica de seqüências genéticas da erva- mate, chá verde, café arábica, café canephora e laranja para estudar a diversidade genética e similaridade dessas espécies, bem como suas relações filogenéticas. foram utilizadas as ferramentas de bioinformática, como o NCBI/BLAST e o MEGA, para identificar as similaridades de seqüências e criação da árvore filogenética das espécies analisadas. A espécie utilizada como base é a erva-mate para valorizar esta espécie que representa um papel importante para o Brasil pois possui muitos benefícios e oportunidades econômicas para quem a planta e atua na sua produção para o chá, o tereré e o chimarrão além de varias outras possibilidades de utilização. Foi escolhida uma sequênciã alvo que no caso foi escolhida de forma aleatória que foi a “carboxylase” (gene *rbcL*) da *Ilex Paraguariensis* e das outras espécies para poder analisar a similaridade e proximidade entre elas usando Filogenia. A análise em sílica de seqüências genéticas é uma técnica importante para a pesquisa em biologia molecular e pode fornecer informações valiosas sobre a diversidade genética e evolução das espécies. Esse trabalho trata da importância do uso das ferramentas da bioinformática para a análise de dados genéticos, bioquímicos e de biologia molecular. A bioinformática é uma linha de pesquisa que envolve aspectos multidisciplinares e que surgiu a partir da utilização de ferramentas computacionais para a análise de dados genéticos. Ela envolve a união de diversas linhas de conhecimento, como a ciência da computação, a engenharia de softwares, a matemática, a estatística e a biologia molecular, e tem como finalidade principal desvendar a grande quantidade de dados que vem sendo obtida através de seqüências de DNA e proteínas. A bioinformática tem um papel fundamental na compreensão da evolução e diversidade genética de uma espécie e de suas relações filogenéticas com outras espécies, contribuindo para a pesquisa em biologia molecular e para a valorização de culturas importantes como a da erva-mate nome científico “*Ilex Paraguariensis*”, chá verde “*Camellia sinensis*”, café “*Coffea arabica*” e “*Coffea canephora*” e laranja “*Citrus sinensis*”.

**Palavras-chave:** bioinformática; árvore filogenética; similaridade biotecnologia; erva-mate

### 1 INTRODUÇÃO

A introdução deve ser breve e conter. Justificar o problema estuda de seqüências; de orma clara, utilizando-se fontes bibliográficas. O último parágrafo deve conter os objetivos do

trabalho realizado. As citações do resumo expandido devem seguir o modelo de ABNT.

A *Ilex Paraguariensis*, mais conhecida como erva-mate, é uma espécie vegetal de grande importância econômica e cultural em diversas regiões do mundo, especialmente na América do Sul. Seu consumo é associado a diversos benefícios para a saúde, como ação antioxidante, aumento da capacidade cognitiva e melhora da função cardiovascular (SOUZA, 2013). Através do estudo de sua filogenia, propriedades genéticas e similaridades com outras espécies, é possível ampliar o conhecimento sobre suas propriedades e as espécies que tem mais proximidade, e buscar novas possibilidades de benefícios e cultivos. Para isso, a bioinformática se apresenta como uma importante ferramenta de pesquisa, permitindo analisar dados genéticos e moleculares e realizar estudos comparativos entre diferentes espécies.

A justificativa do trabalho é que a *Ilex Paraguariensis*, mais conhecida como erva-mate, é uma espécie vegetal de grande importância econômica e cultural em diversas regiões do mundo, especialmente na América do Sul com muitas possibilidades de estudos genéticos (DEBAT et al., 2014). Através do estudo de sua filogenia, propriedades genéticas e similaridades com outras espécies, como a *Camellia sinensis* e a *Coffea arábica*, é possível ampliar o conhecimento sobre suas propriedades e buscar novas possibilidades de benefícios e cultivos. Para isso, a bioinformática se apresenta como uma importante ferramenta de pesquisa, permitindo analisar dados genéticos e moleculares e realizar estudos comparativos entre diferentes espécies.

O objetivo geral do trabalho é analisar a similaridade de uma sequência alvo que no caso foi escolhida de forma aleatória que foi a “carboxylase” (gene *rbcL*) da *Ilex Paraguariensis* com o mesmo gene de algumas espécies conhecidas por seus benefícios como o café, o chá verde e a laranja e analisar as proximidades entre elas indicando com quais a erva-mate é mais próxima com relações as outras. Os objetivos específicos são:

- Realizar uma revisão bibliográfica sobre as qualidades e benefícios da *Ilex Paraguariensis*, *Camellia sinensis*, *Coffea arábica*, *Coffea canephora* e *Citrus sinensis*.
- Investigar a filogenia entre as espécies analisadas.
- Utilizar ferramentas de bioinformática para explorar possíveis semelhanças e diferenças entre as espécies analisadas.

## 2 MATERIAIS E MÉTODOS

Os Materiais e Métodos (ou metodologia) deve ser conciso, mas coeso e coerente, de modo que o leitor entenda e possa reproduzir os procedimentos utilizados.

O método utilizado para buscar e analisar as seqüências genéticas de proteínas das espécies estudadas envolveu o uso de ferramentas de bioinformática, como o banco de dados online UniProt e NCBI/BLAST, e para uma pesquisa mais específica foi inserido os nomes das espécies e uma função química, no caso, “carboxylase”. A ferramenta NCBI/BLAST foi usada para encontrar seqüências similares e cruzamento entre elas. A ferramenta MEGA-X foi usada para fazer a análise da similaridade de seqüências usando o algoritmo Clustalw e também a criação da árvore filogenética usando *Maximum Likelihood Estimation* (MLE), que corresponde ao algoritmo de máxima verossimilhança, que é um método para estimar os parâmetros de um modelo estatístico. Assim, a partir de um conjunto de dados e dado um modelo estatístico, a estimativa por máxima verossimilhança estima valores para os diferentes parâmetros do modelo.

As buscas pelas seqüências foram feitas no site do NCBI (*National Center for Biotechnology Information*) localizado em <https://www.ncbi.nlm.nih.gov/>. O NCBI é o Centro Nacional de Informações sobre Biotecnologia promove o avanço da ciência e da saúde, fornecendo acesso a informações biomédicas e genômicas (NCBI, 2024).

O BLAST (*Basic Local Alignment Search Tool*) é uma Ferramenta básica de pesquisa de alinhamento local, que encontra regiões de similaridade entre sequências biológicas. O programa compara sequências de nucleotídeos ou proteínas com bancos de dados de sequências e calcula a significância estatística (BLAST, 2024).

O software de Análise Genética Evolutiva Molecular (MEGA) amadureceu para conter uma grande coleção de métodos e ferramentas de evolução molecular computacional. Aqui, descrevemos novas adições que tornam o MEGA um produto mais abrangente ferramenta para construir árvores temporais de espécies, patógenos e famílias de genes usando métodos rápidos (*Molecular Evolutionary Genetics Analysis* (MEGA), 2024). O MEGA possui métodos para estimar tempos de divergência e intervalos de confiança são implementados para usar densidades de probabilidade para calibração restrições para datação de nós e datas de amostragem de sequência para análises de datação de dicas (TAMURA; STECHER; KUMAR, 2021)

De acordo com Barnes (2007), a função da bioinformática é essencial para a interrogação eficaz de dados genéticos e genômicos, bem como da maioria dos outros dados biológicos. Isso torna a experiência em bioinformática um pré-requisito para a eficácia da genética. A especialização em bioinformática não é mistério; as ferramentas de bioinformática certas, juntamente com uma mente curiosa e vontade de experimentar (requisitos fundamentais para qualquer cientista, bioinformático ou não), podem gerar confiança e competência no tratamento de dados de bioinformática num espaço de tempo muito curto.

As árvores filogenéticas nada mais são que a representação da história evolutiva de uma espécie. Essas representações devem ser lidas da base para as pontas, sendo a base a história mais antiga e as pontas a história mais recente daquele táxon (SANTOS, 2024).

A erva-mate é um chá de ervas feito a partir das folhas e galhos da planta *Ilex paraguariensis*. As folhas são desidratadas no fogo e depois mergulhadas em água quente ou fria para fazer a infusão. A iguaria pode proporcionar diversos benefícios à saúde, como melhorar o desempenho atlético e o foco e ainda ajudar a perder peso e diminuir o risco de doença cardíaca (ZANIN,2022).

Segundo Prosdocimi (2007), podemos considerar a bioinformática como uma linha de pesquisa que envolve aspectos multidisciplinares e que surgiu a partir do momento em que se iniciou a utilização de ferramentas computacionais para a análise de dados genéticos, bioquímicos e de biologia molecular.

A bioinformática envolve a união de diversas linhas de conhecimento – a ciência da computação, a engenharia de softwares, a matemática, a estatística e a biologia molecular – e tem como finalidade principal desvendar a grande quantidade de dados que vem sendo obtida através de seqüências de DNA e proteínas. Para o desenvolvimento de genomas completos, a informática é imprescindível e a biologia molecular moderna não estaria tão avançada hoje, não fossem os recursos computacionais existentes.

### 3 RESULTADOS E DISCUSSÃO

Foram coletadas 5 Seqüências selecionadas como busca específica para função química da “carboxylase”, gene *rbcL*. O RuBisCO (abreviatura de ribulose-1,5- bisfosfato carboxilase oxigenase) é a enzima mais abundante nas plantas e, por conseguinte, a proteína mais abundante no planeta (ROBINSON; PORTIS, 1989).

Esta enzima capta o dióxido de carbono procedente do ar e um açúcar existente na célula chamado RuDP (ribulose 1,5-difosfato ou RuBP - ribulose bis-fosfato). A reacção entre estes dois reagentes dá origem a duas moléculas do açúcar PGA (fosfoglicerato). A RuBisCO é assim

responsável pelo importante primeiro passo do ciclo de Calvin e em concreto pela fixação do dióxido de carbono na sua forma orgânica.

**Tabela 1: Sequências analisadas na pesquisa**

ESPÉCIE	SEQUÊNCIA
ERVA MATE	>D0Q0J8_rbcL_Ilex_paraguariensis MSPQTETKASVGFKAGVKDYKLYYTPDYDTKDTDILAAFRVSPQPG VPP EEAGAAVA AESSTGTWTTVWTDGLTSLDRYKGRCYQIEPVAGEENQY IA YVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPTAYTKTF QG PPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLD FT KDDENVNSQPFMRWRDRFLFCAEALYKAQAETGEIKGHYLNATAGTC EE MMKRAVFARELAVPIVMHDYLTGGFTANTTLAHYCRDNGLLLHIHRA M HAVIDRQKNHGMHFRVLAKALRLSGGDHIHAGTVVGKLEGERDITLG FV LLRDDVIEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALTEIFG D DSVLQFGGGT LGHPWGNAPGAVANRVALEACVQARNEGRD
CHÁ VERDE	>L0E907_rbcL_Camellia_sinensis MSPQTETKASVGFKAGVKDYKLNYYTPDYETKDTDILAAFRVTPQPG VP PEEAGAAVA AESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVAGEESQF IA YVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPTAYVKTF Q GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGL DF TKDDENVNSQPFMRWRDRFVFC AEAIYKAQAETGEIKGHYLNATAGT CE EMMKRAVFARELGVPIVMHDYLTGGFTANTSLAHYCRDNGLLLHIHR A MHAVIDRQKNHGMHFRVLAKALRMSGGDHVHAGTVVGKLEGEREIT LG FVDLLRDDYVEKDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMPALT EIF GDDSVLQFGGGT LGHPWGNAPGAVANRVALEACVQARNEGRDLARE GNEIIREASKWSPELAAACEVWKEIKFEFPAMDTL

<p>CAFÉ ARÁBICA</p>	<p>&gt;P48694_rbcL_Coffea_arabica MSPQTETKASVGFKAGVKEYKLYYTPEYETKDTDILAAFRVTPQPGV PP EEAGAAVAAESSTGTWTAVWTDGLTSLDRYKGRCYHIEPVPGEENQY IA YVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRVPPAYIKTF Q GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGL DF TKDDENVNSQPFMRWRDRFCFCAEALFKAQAETGEIKGHYLNATAGT CE EMIKRAVFARELGVPIVMHDYLTGGFTANTSLAHYCRDNGLLLHIHRA M HAVIDRQKNHGMHFRVLAKGLRMSGGDHIIHAGTVVGKLEGERDITLG FV DLLRDDFIEKDRSRGIYFTQDWVSLPGVIPVASGGIHWHPALTEIFG DDSVLQFGGGT LGHPWGNAPGAVANRVALEACVKARNEGRDLAAEG NEIIR EASKWSPELAAACEVWKEIRFNFEAMDKLDKEKDL</p>
<p>CAFÉ CANEPHORA</p>	<p>&gt;YP_009251133.1 rbcL [organism=Coffea canephora] [GeneID=27439022] MSPQTETKASVGFKAGVKEYKLYYTPEYETKDTDILAAFRVTPQPGV PP EEAGAAVAAESSTGTWTCVWTDGLTSLDRYKGRCYHIEPVPGEEDQYI A YVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRV PPAYIKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGR AVY</p>
	<p>ECLRGGLDFTKDDENVNSQPFMRWRDRFCFCAEALYKAQAETGEIKG HY LNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLAHYCRD NG LLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHVHAGTVVGK LE GERDITLGFVDLLRDDFIEKDRSRGVYFTQDWVSLPGVIPVASGGIHW H MPALTEIFGDDSVLQFGGGT LGHPWGNAPGAVANRVALEACVKARNE G RDLAAEGNEIIREASKWSPELAAACEVWKEIRFNFEAVDKLDKEKE L</p>

LARANJA	>Q09MH0_1_rbcL_Citrus_sinensis MSPQTETKASVGFKAGVKDYKLYYTPDYVTKDTELDILAAFRVTPQPG VP PEEAGAAVAAESSTGTWTAVWTDGLTSLDRYKGRYCYNIEPVAGEENQ YI CYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYTKT FQ GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGL DF TKDDENVNSQPFMRWRDRFLFCAEALYKAQAETGEIKGHYLNATAGT CE EMLKRAVFARELGVPIVMHDYLTGGFTANTTLAHYCRDNGLLLHIHR AM HAVIDRQKNHGMHFRVLAKALRLSGGDHIIHAGTVVGKLEGERDITLG FV DLLRDDFVEKDRSRGIYFTQDWVSIPGVIPVASGGIHWHPALTEIFG D DSVLQFGGGTLGHPWGNAPGAVANRVALEACVQARNEGRDLAREGN EII REASKWSPELAAACEVWKSIFFEFAAMDTL
---------	--

Foi usado o Software Mega para fazer o alinhamento e a criação da árvore filogenética para poder fazer o teste in silico da proximidade entre as espécies para analisar suas similaridades de sequências genéticas. Para o alinhamento usou-se o algoritmo do ClustalW disponível no MEGA. O CLUSTAL é uma série de programas de computador comumente usados para alinhamento de múltiplas sequências. Diversas modificações são incorporadas a este novo programa, CLUSTAL W, que está disponível gratuitamente para uso (THOMPSON, HIGGINS, GIBSON, 1994).

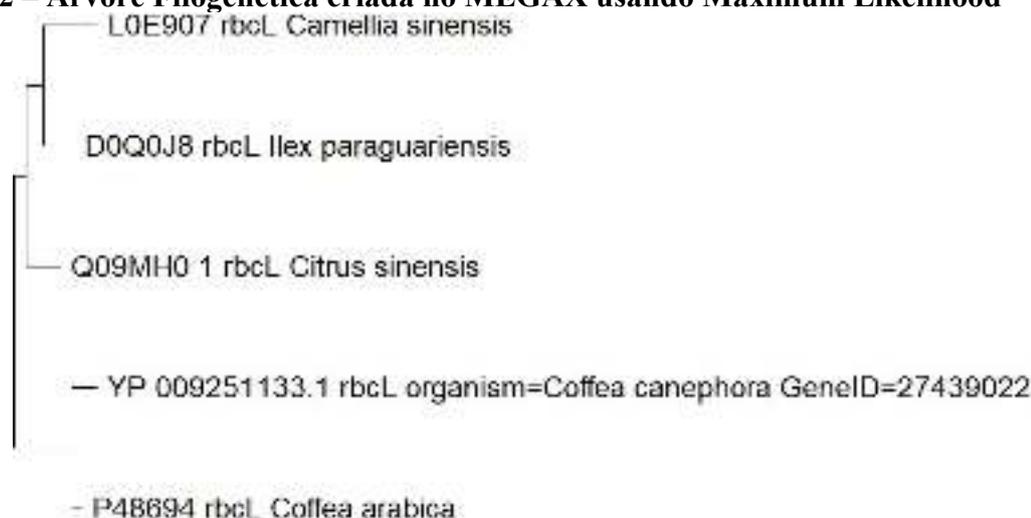
A Figura 1 a seguir mostra o alinhamento das sequências realizado no Software MEGA 11.

**Figura 1** – Sequências alinhadas no MEGAX usando o algoritmo Clustalw



A Figura 2 a seguir mostra o Árvore Filogenética realizado no Software MEGA 11 também , que foi criada após ter sido feito o alinhamento das sequências.

**Figura 2 – Árvore Filogenética criada no MEGAX usando Maximum Likelihood**



#### 4 CONCLUSÃO

Com base na análise filogenética do gene rbcL na pesquisa foi possível observar uma relação filogenética próxima entre as espécies *Ilex paraguariensis* (Erva Mate) e a *Camellia sinensis*, e em seguida com a *Citrus sinensis* (Laranja), sugerindo uma sequência mais similar entre elas. A *Coffea arabica* tem uma proximidade filogenética com a *Coffea canephora* confirmando o que já era esperado por serem espécies próximas por ambas serem café, e depois uma proximidade com a parte em comum com as espécies da *Ilex Paraguariensis*, *Camellia sinensis* e da *Citrus sinensis*.

Além disso, a bioinformática se mostrou uma ferramenta fundamental para a análise de dados genéticos, permitindo a compreensão da evolução e diversidade genética de uma espécie e de suas relações filogenéticas com outras espécies.

O conhecimento obtido a partir da análise de dados genéticos e moleculares pode ser aplicado em diversas áreas, como a agronomia, a farmacologia e a indústria alimentícia, contribuindo para o avanço da ciência e para a melhoria da qualidade de vida da população, desenvolvimento econômico e preservação do meio ambiente.

No caso específico da análise das sequências do gene rbcL em plantas com potencial econômico, como *Ilex paraguariensis*, *Camellia sinensis*, *Coffea arabica* e *Citrus sinensis*, essas informações podem ser úteis para a produção e melhoramento dessas plantas em sistemas agrícolas e silviculturais. Isso pode levar a cultivos mais eficientes e sustentáveis, com impactos positivos na saúde pública e na economia.

Como trabalho futuro pretende-se analisar outras sequências genéticas das mesmas espécies para analisar como se comporta a árvore e incluir espécies novas também para ver se mantém a mesma árvore ou varia de acordo com as sequências a serem analisadas.

#### REFERÊNCIAS

BARNES, Michael. **Bioinformatics for Geneticists: A Bioinformatics Primer for the Analysis of Genetic Data**. Second Edition. John Wiley & Sons Ltd: England. 2007.

BLAST. Disponível em: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Acesso em: 07 março 2024.

DEBAT HJ, GRABIELE M, AGUILERA PM, BUBILLO RE, OTEGUI MB, DUCASSE DA, ZAPATA PD, MARTI DA. **Exploring the genes of yerba mate (*Ilex paraguariensis* A. St.-Hil.) by NGS and de novo transcriptome assembly.** PLoS One. 2014 Oct 16;9(10):e109835. doi: 10.1371/journal.pone.0109835. PMID: 25330175; PMCID: PMC4199719.

MOLECULAR EVOLUTIONARY GENETICS ANALYSIS (MEGA). Disponível em: [https://www.megasoftware.net/web\\_help\\_11/index.htm#t=Introduction.htm](https://www.megasoftware.net/web_help_11/index.htm#t=Introduction.htm). Acesso em: 07 março de 2024.

NCBI. Disponível em: <https://www.ncbi.nlm.nih.gov/>. Acesso em: 28 março 2023.

PROSDOCIMI, Francisco. CURSO ON LINE INTRODUÇÃO À BIOINFORMÁTICA. 2007.

ROBINSON, S.P., PORTIS, A. R. **Ribulose-1,5-bisphosphate carboxylase/oxygenase activase protein prevents the in vitro decline in activity of ribulose-1,5-bisphosphate carboxylase/oxygenase.** Plant Physiol. 1989 Jul;90(3):968-71. doi: 10.1104/pp.90.3.968. PMID: 16666906; PMCID: PMC1061829.

SANTOS, Vanessa Sardinha dos. "Filogenia"; Brasil Escola. Disponível em: <https://brasilecola.uol.com.br/biologia/filogenia-que-isto.htm>. Acesso em 13 de abril de 2024.

SOUZA, Claucia Fernanda Volken de. Importância da Erva-Mate para a Saúde. Disponível em: <https://www.univates.br/noticia/11736-importancia-da-erva-mate-par-a-a-saude>. UNIVATES - Mestrado em Biotecnologia. Postado em: 02/08/2013.

**THOMPSON, J; HIGGINS, D; GIBSON, T.** Clustal W: Improving The Sensitivity Of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties And Weight Matrix Choice. 1994, Nucleic Acids Res 22: 4673–4690.

TAMURA, Koichiro, STECHER, Glen, KUMAR, Sudhir. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol. Biol. Evol. 38(7):3022–3027 doi:10.1093/molbev/msab120 Advance Access publication April 23, 2021.

ZANIN, Tatiana. Erva-mate: o que é, benefícios e como preparar. TuaSaude, junho de 2022. Disponível em: <https://www.tuasaude.com/erva-mate/> Acessado em: 16 de nov. de 2022.