

**ODYONE NASCIMENTO DA SILVA**

**SEQUENCIAMENTO, CARACTERIZAÇÃO E COMPARAÇÃO EVOLUTIVA DOS  
PLASTOMAS DE *Euterpe edulis* E *Euterpe Oleracea***

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fisiologia Vegetal, para obtenção do título de *Magister Scientiae*.

Orientador: Marcelo Rogalski

**VIÇOSA - MINAS GERAIS  
2017**

**Ficha catalográfica elaborada pela Biblioteca Central da  
Universidade Federal de Viçosa - Campus Viçosa**

T

S586s  
2017

Silva, Odyone Nascimento da, 1990-  
Sequenciamento, caracterização e comparação  
evolutiva dos plastomas de *Euterpe edulis* e *Euterpe  
oleracea* / Odyone Nascimento da Silva. - Viçosa, MG,  
2017.  
1 dissertação eletrônica (56 f.): il.

Inclui apêndice.  
Orientador: Marcelo Rogalski.  
Dissertação (mestrado) - Universidade Federal de  
Viçosa, Departamento de Biologia Vegetal, 2017.  
Referências bibliográficas: f. 38-45.  
DOI: <https://doi.org/10.47328/ufvbbt.2022.192>  
Modo de acesso: World Wide Web.

1. Arecaceae - Filogenia. 2. Açai - Genética.  
3. Cloroplastos. 4. Diversidade genética. I. Rogalski,  
Marcelo, 1975-. II. Universidade Federal de Viçosa.  
Departamento de Biologia Vegetal. Programa de Pós-  
Graduação em Fisiologia Vegetal. III. Título.

CDD 22. ed. 584.5

Bibliotecário(a) responsável: Alice Regina Pinto CRB6 2523


**ODYONE NASCIMENTO DA SILVA**

**SEQUENCIAMENTO, CARACTERIZAÇÃO E COMPARAÇÃO EVOLUTIVA DOS  
PLASTOMAS DE *Euterpe edulis* E *Euterpe Oleracea***

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fisiologia Vegetal, para obtenção do título de *Magister Scientiae*.


APROVADA: 25 de julho de 2017

Assentimento:

Documento assinado digitalmente  
 ODYONE NASCIMENTO DA SILVA  
Data: 11/01/2023 08:26:33-0300  
Verifique em <https://verificador.iti.br>

---

Odyone Nascimento da Silva  
Autora

Documento assinado digitalmente  
 MARCELO ROGALSKI  
Data: 11/01/2023 11:28:29-0300  
Verifique em <https://verificador.iti.br>

---

Marcelo Rogalski  
Orientador

**Dedico aos meus Pais**  
**Luiz Miguel Borges Silva (*in memoriam*)**  
**& Odinéia Nascimento Silva**

## AGRADECIMENTOS

A DEUS que em seu infinito amor não me desamparou ao longo da minha vida e principalmente nesses dois anos onde sua ajuda foi diariamente necessitada.

Aos meus pais, pelo apoio incondicional antes, e durante a minha jornada, certeza se não fosse pelo apoio que me deram eu não teria conseguido.

Ao meu esposo Renato Bandeira pelo companheirismo e compreensão, obrigada por não medir esforços para me ajudar e apoiar, mesmo sabendo o quão difícil seria o tempo que ficaríamos longe, você tem uma importância indescritível neste trabalho e na minha vida.

A minha irmã Paula Francyneth que foi um pilar de apoio incondicional, que me ouviu e aconselhou em muitos momentos difíceis, sem sua ajuda com certeza teria sido muito mais difícil.

Aos meus irmãos Lywelthton, Lywyney e Raiane, o amor de vocês e a confiança que depositaram em mim quando eu mesma duvidei, o meu muito obrigado.

A Amanda Lopes que foi um dos meus anjos da guarda nesta etapa, só posso agradecer muito a Deus pela sua vida, sua ajuda é imensurável e desejo tudo de melhor que possa existir nessa vida para você.

A Tulio Pacheco pela paciência e amizade, sua ajuda foi de fundamental importância para a realização deste trabalho.

A Gélia Viana uma amiga e companheira nessa jornada, momentos triste de dissiparam por você está presente, muito obrigada por me apoiar e aconselhar.

A Suzy pela amizade e companheirismo você também fez parte dessa etapa e tem lugar garantido nas minhas melhores lembranças.

Ao meu cunhado Marcelo Lima que pelo carinho e apoio com o cartão de crédito para as incontáveis passagens aéreas que precisei comprar.

A todos os meus sobrinhos pelo amor e carinho que foram de grande ajuda.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Fisiologia Vegetal por todo suporte e auxílio acadêmico e também pela oportunidade.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

Ao Prof. Marcelo Rogalski, pela orientação, atenção, paciência que foi de grande valia para o desenvolvimento desta pesquisa me ajudando e dando força para superar as minhas dificuldades de iniciante.

Aos professores Samuel Martins e Camilo Vital por aceitarem integrar a banca examinadora deste trabalho.

A todos os professores da Fisiologia Vegetal pelo conhecimento compartilhado. A todos os técnicos do Departamento de Biologia Vegetal pelo auxílio e suporte.

Aos amigos do laboratório de fisiologia Molecular (Jessenia Moncaleano, Emanuel Naves, Juliene Moreira e João Vitor) que ajudaram diretamente neste trabalho, jamais me esquecerei do quanto aprendi e evolui convivendo com você, quero leva-los comigo, na vida e no coração para sempre.

A Jonas Vargas, pela amizade e momentos incríveis que compartilhamos, conhecer você foi muito importante e reconfortante.

Aos meus amigos do Grupo de Convivência, (Adassa, Luna, Jaque, Caito, Marcelo, Evandro, Léo, Elder e Douglas), minha família espiritual, nossas orações e estudos da Palavra foram minha força, levarei cada um no meu coração.

A todos que ajudaram direta ou indiretamente para a realização deste trabalho, muito obrigado.

## RESUMO

SILVA, Odyone Nascimento, M.Sc., Universidade Federal de Viçosa, julho de 2017. **Sequenciamento, caracterização e comparação evolutiva dos plastomas de *Euterpe edulis* e *Euterpe Oleracea***. Orientador: Prof. Marcelo Rogalski.

As espécies *E. edulis* e *E. oleracea* pertencem à família Arecaceae e são as espécies mais importantes desse gênero, pela multiplicidade de usos e com destaque para comercialização de seus frutos e palmito. Neste sentido, reportamos a sequência completa e caracterização do genoma plastidial de ambas as espécies, que são as primeiras da tribo Euterpeae a terem seu plastoma sequenciado. Os plastomas das duas espécies apresentam uma estrutura quadripartida circular, que é típica na maioria das angiospermas. A estrutura do plastoma, a ordem dos genes e o conteúdo gênico são semelhantes às outras espécies da família Arecaceae. Exceções foram observadas nos genes *yef1* e o *rps16* que aparecem como pseudogenes em algumas espécies da família Arecaceae e nesta pesquisa aparecem na forma funcional em ambas as espécies. A análise de divergência de genes mostra que o gene *yef1* é um dos genes mais divergentes na família Arecaceae. As espécies do gênero *Euterpe* mostraram uma alta conservação de sítios de edição de mRNA e a presença de 23 predições de sítios mRNA não observados em outras palmeiras. As duas espécies apresentam distribuição similar de SSRs com algumas regiões contendo um número mais elevado dos localizados principalmente nas regiões de cópia única. Nos plastomas das duas espécies foram localizados e caracterizados seis *hotspots* de divergência de nucleotídeos entre as espécies em estudo, todos localizados na região de cópia única. Baseadas no plastoma de 36 espécies pertencentes à família Arecaceae, as árvores filogenômicas obtidas por duas análises estatísticas mostraram um alto valor de suporte para as relações entre as tribos amostradas, incluindo o gênero *Euterpe* dentro da tribo *Euterpeae*. Por tanto, o sequenciamento permitiu caracterizar o genoma plastidial, identificar sequências repetidas e microssatélites, detectar pontos de divergência gênica, analisar divergência dos genes e previsão de edição de mRNA nas duas espécies do gênero *Euterpe*, que serão úteis para seleção de dados em futuros estudos filogenéticos e evolutivos da tribo e dentro da família Arecaceae.

Palavras-Chave: Arecaceae. Filogenia. Açaí. Genética de cloroplastos. Diversidade genética.

## ABSTRACT

SILVA, Odyone Nascimento, M.Sc., Universidade Federal de Viçosa, July, 2017. **Sequencing, characterization and evolutionary comparison of the plastomas of *Euterpe edulis* and *Euterpe Oleracea*.** Adviser: Prof. Marcelo Rogalski.

The species *E. edulis* and *E. oleracea* belong to the Arecaceae family and are the most important species of this genus, due to the multiplicity of uses and especially for the commercialization of their fruits and palm hearts. Here, we reported the complete sequence and characterization of the plastid genome of both species, which are the first species of the Euterpeae tribe to have their plastome sequenced. The plastomes of both species have a circular quadripartite structure, which is conserved between most angiosperms. The plastome structure, gene order and gene content are similar to other species of the family Arecaceae. Exceptions were observed in the *yef1* and *rps16* genes that appear as pseudogenes and in this research they are functional in both species. The gene divergence analysis shows that the *yef1* gene is one of the most divergent genes in the family Arecaceae. These species of the genus *Euterpe* showed a high conservation of mRNA editing sites and the presence of 23 predictions of RNA editing sites, which was not observed before in other palm trees. The two species show similar distribution of SSRs with some regions containing a higher number of those located mainly in the single copy regions. In the plastomes of the two species, six nucleotide divergence hotspots between the species under study were located and characterized, all located in the single copy region. Based on the plastomes of 36 species belonging to the family Arecaceae, the phylogenomic trees obtained by two statistical analyzes showed a high support value for the relationships between the sampled tribes, including the genus *Euterpe* within the tribe *Euterpeae*. Therefore, sequencing the plastome of these species allowed the characterizing, the identification of repeated sequences and microsatellites, the detecting genetic divergence. This information will be useful data selection in further studies, mainly in phylogenetic and evolutionary areas of the tribe *Euterpeae* and whole family Arecaceae.

Keywords: Arecaceae. Phylogeny. Palm genetics. Chloroplasts. Genetic diversity.

## SUMÁRIO

1. INTRODUÇÃO .....	9
2. MATERIAL E MÉTODOS .....	12
2.1 Isolamento do DNA do cloroplasto, sequenciamento, montagem e anotação .....	12
2.2 Análise da estrutura do genoma e identificação das sequências repetidas .....	14
2.3 Análise de <i>sliding window</i> e predição de sítios de edição de RNA .....	15
2.4 Inferências filogenômicas .....	15
2.5 Estimativa da divergência gênica .....	16
3. RESULTADOS .....	17
3.1 Tamanho, conteúdo gênico e análise estrutural do plastoma de <i>E. edulis</i> e <i>E. oleracea</i> .....	17
3.2 Análise da divergência dos genes nos plastomas de <i>E. edulis</i> e <i>E. oleracea</i> .....	19
3.3 Sítios de edição de mRNAs previstos nos plastomas de <i>E. edulis</i> e <i>E. oleracea</i> .....	19
3.4 Identificação de sequências repetidas e análise de <i>sliding window</i> em <i>E. edulis</i> e <i>E. oleracea</i> .....	27
3.5 Inferência filogenômica .....	31
4. DISCUSSÃO .....	33
4.1 Tamanho e análise estrutural dos plastomas de <i>E. edulis</i> e <i>E. oleracea</i> .....	33
4.2 Divergência de genes em <i>E. edulis</i> e <i>E. oleracea</i> .....	33
4.3 Previsão de sítios de edição de RNA .....	34
4.4 Sequências repetidas e <i>hotspots</i> de divergência de nucleotídeos .....	35
4.5 As relações entre as tribos dentro da família <i>Arecaceae</i> com base em genomas plastidiais inteiros .....	36
5. CONSIDERAÇÕES FINAIS .....	37
REFERÊNCIAS BIBLIOGRÁFICAS .....	38
APÊNDICE .....	45

## 1. INTRODUÇÃO

As palmeiras pertencem à família Arecaceae a qual possui a maior diversidade dentro das monocotiledôneas, com cerca de 2700 espécies reunidas em mais de 240 gêneros (Lorenzi, 2010). O gênero *Euterpe* Mart. contém sete espécies distribuídas na América do Sul e Central, das quais cinco são encontradas no Brasil e apenas uma está fora da floresta Amazônica, o *E. edulis* (Henderson e Galeano, 1996). No entanto, o número de espécies pertencentes a esse gênero já foi motivo de discussão e tem sido objeto de constantes revisões. Os primeiros trabalhos englobavam 49 espécies (Glassman, 1972), na revisão proposta por Uhl e Dransfield (1986) o gênero foi representado por 28 espécies. Representam a terceira família mais importante economicamente para o uso humano e do ponto de vista agroindustrial e principalmente pela multiplicidade de usos. *E. oleracea* e *E. edulis* são as espécies mais importantes desse gênero, sendo que *E. oleracea* é conhecida popularmente como Açaí ou Açaí-do-Pará e a *E. edulis* como Palmiteiro, destacando-se atualmente nesse setor como a principal fonte de extração de palmito e comercialização de seus frutos, na forma de polpa para a fabricação de derivados, de forma expressiva pela população amazônica brasileira (Yamaguchi *et al.*, 2015).

A localização geográfica onde a espécie é encontrada e desenvolve suas características únicas é chamada centro de diversidade (Esquinas-Alcazar, 2004), o gênero *Euterpe* possui dois centros de diversidade, o primeiro está localizado no noroeste da Colômbia e o segundo na região compreendida pelo escudo das Guianas e alto rio Negro em áreas bastante úmidas (Oliveira, 1995). A região do estuário do rio Amazonas também é considerada como centro de origem, além de constituir o centro de maior diversidade genética de *E. oleracea*, pois nessa região são encontradas numerosas populações com variações intra e interpopulacional bem acentuadas, no que concerne às características morfológicas, fenológicas, fisiológicas e agrônômicas das plantas (Oliveira, 2000). A espécie *E. oleracea* foi enquadrada na lista das espécies mais abundantes encontradas na região da floresta Amazônica (Steege *et al.*, 2013). Quanto à reprodução, essas espécies possuem assincronia nas fases de floração na mesma inflorescência (protândria), por isso são consideradas espécies predominantemente alógamas, no entanto, pode também ocorrer autofecundação (Souza, 2002). O alto valor comercial dos produtos oriundos dessas palmeiras incentivou o desenvolvimento de programas de melhoramento genético para produzir cultivares com maior rendimento e melhor qualidade do fruto, além de favorecer o cultivo e a conservação dessas

espécies, o que é urgente no caso de *E. edulis* uma vez que é uma espécie ameaçada (Oliveira *et al.*, 2016).

A palmeira *E. edulis* ocorre na Mata Atlântica brasileira e foi intensamente explorada nas suas áreas de ocorrência natural, por constituir a principal fonte de extração natural de palmito no Brasil (Brancalion *et al.*, 2012). Estudos populacionais de *E. edulis* em fragmentos florestais da Mata Atlântica, utilizando marcadores microssatélites nucleares e dados de dispersão de pólen, indicaram que essa espécie, apesar de fragmentada, possui uma variabilidade genética alta. No entanto, esta variabilidade genética da espécie estaria ameaçada atualmente pelos processos antropológicos de desmatamento, degradação e fragmentação do meio ambiente (Carvalho *et al.*, 2015). O uso de marcadores moleculares em estudos da diversidade genética dentro e entre populações tem apontado perdas de diversidade, aumento dos níveis de endogamia e alterações na organização genômica das populações exploradas (Conte *et al.*, 2006; 2008).

O gênero *Euterpe* pertence à tribo Euterpeae enquadrada dentro da subfamília Arecoideae, que abriga aproximadamente 1300 espécies (Baker *et al.*, 2015). Vários estudos usando sequências de genes plastidiais reclassificaram muitas espécies da subfamília Arecoideae, que foi circunscrita para incluir as tribos Areceae, Chamaedoreae (anteriormente Hyophorbeae, Ceroxyloideae), Cocoseae, Euterpeae, Geonomateae, Iriarteeae, Leopoldinieae, Manicarieae, Oranieae, Pelagodoxeae, Podococceae, Reinhardtieae, Roystoneae e Sclerospermeae. Caryoteae foi removida da subfamília Arecoideae e adicionada na subfamília Coryphoideae, essas reclassificações só foram possíveis graças à utilização de sequências de genes plastidiais (Baker *et al.*, 2009).

Os plastídios são organelas centrais no metabolismo vegetal, e possuem seu próprio genoma (plastoma) e processo de replicação independente do ciclo celular (Pyke, 2007). Os plastomas das plantas superiores são, em geral, bem conservados, apresentando variações em tamanho de 70-220 kb e no número de genes de 25-131 (Rogalski *et al.*, 2015; Daniell *et al.*, 2016; Lee *et al.*, 2016). Esse pequeno número de genes são os remanescentes do genoma de seus ancestrais, as cianobactérias, cujo genoma atual abriga aproximadamente 3200 genes (Kaneko *et al.*, 1996). Contudo, o plastoma retém inúmeras características procarióticas incluindo um genoma circular, agrupamento dos genomas em nucleoides, organização dos genes em operons (complexos genéticos com Ativador, promotor e operador) e maquinaria de expressão gênica similar à procariótica, porém com complexa expressão gênica com interação entre polimerases nuclear e plastidial. Os genes presentes no genoma plastidial estão divididos

em diferentes grupos: (i) maquinaria da fotossíntese (subunidades do aparato fotoquímico e a subunidade maior da rubisco), (ii) genes relacionados à expressão gênica (subunidades da RNA-polimerase plastidial, proteínas ribossomais, tRNAs e rRNAs) e outros genes relacionados à síntese de ácidos graxos e homeostase de proteínas (Bock, 2007). O primeiro grupo de genes que codificam componentes da maquinaria fotossintética (i) não é essencial para a sobrevivência celular e pode ser mutado individualmente sob condições heterotróficas, como a cultura *in vitro*, podendo ser desligados (*knockouts*) sem causar nenhum efeito no crescimento da planta, ao exemplo do que é feito com *Nicotiana tabacum* para o estudo das *ycf* e de outros genes fotossintéticos plastidiais (Hager *et al.*, 1999; Bock, 2001). Por outro lado, genes que compõe a maquinaria de expressão gênica plastidial são, na grande maioria, essenciais para a viabilidade celular e sua mutação leva à letalidade (Rogalski *et al.*, 2008; Fleischman *et al.*, 2011; Alkatib *et al.*, 2012).

Diversos estudos utilizam marcadores moleculares plastidiais como ferramentas altamente eficientes para estudos de divergência genética, fluxo gênico, evolução e genética de populações (Wu *et al.*, 2007; Hirao *et al.*, 2008; Lin *et al.*, 2010; Rogalski *et al.* 2015). Contudo observa-se que apesar da importância do gênero *Euterpe*, estudos utilizando dados genéticos moleculares são limitados para esse gênero, sendo que nenhuma espécie desta tribo (Euterpeae) teve seu plastoma sequenciado até o presente momento. Apenas a sequência de poucos genes plastidiais utilizados como marcadores moleculares (*rbcL*, *rps16*) embasou análises filogenéticas e a relação entre várias tribos e subfamílias dentro da família Arecaceae (Comer *et al.*, 2015; Faurby *et al.*, 2016). A classificação com base em uma análise abrangente de sequências do gene *matK* removeu tribos da subfamília Arecoideae que foram transferidas para subfamília Coryphoideae (Dransfield *et al.*, 2008). A sequência plastidial de apenas um gene (*psbA*) possibilitou a resolução filogenética relacionada ao grau de parentesco das tribos pertencentes à família Arecaceae (Comer *et al.*, 2015).

O genoma plastidial da maioria das angiospermas possui herança uniparental materna (Bock, 2007). Tal aspecto, associado com a baixa taxa de evolução dos genes, faz dos genes plastidiais excelentes marcadores para estudos filogenéticos (Jansen *et al.*, 2007; Moore *et al.*, 2010; Vieira *et al.*, 2014a; Rogalski *et al.*, 2015). Sequências simples repetidas (SSRs) presentes no genoma plastidial também podem ser utilizadas em estudos de variabilidade inter e intraespecífica para monitorar o fluxo de genes como no trabalho realizado com a espécie *Scutellaria baicalensis* (Jiang *et al.*, 2017), divergência genética populacional, diversidade citoplasmática e filogeografia. Além disso, o sequenciamento do genoma plastidial permite a

realização de estudos evolutivos como eventos recombinantes, rearranjos, deleções e perda de genes em nível de espécies, subtribos, tribos, famílias, ordem e grupos de plantas (Rogalski *et al.*, 2015).

Dada a importância comercial dessas palmeiras bem como as escassas pesquisas de classificação taxonômica e processos evolutivos dessas espécies, o sequenciamento permitiu caracterizar o genoma plastidial, identificar sequências repetidas e microssatélites, determinar pontos de divergência gênica, analisar a divergência de genes e prever sítios de edição de RNA mensageiro (mRNA) nas duas espécies do gênero *Euterpe*, realizando também o estudo da posição filogenética. Todos estes dados serão úteis para estudos evolutivos dentro da tribo Euterpeae da família Arecaceae.

## **2. MATERIAL E MÉTODOS**

### **2.1 Isolamento do DNA do cloroplasto, sequenciamento, montagem e anotação**

Folhas novas e frescas de *E. oleracea* foram coletadas na Universidade Federal de Viçosa (UFV), MG, Brasil, enquanto que o material vegetal de *E. edulis* foi coletado no Centro de Ciências da Universidade Federal de Santa Catarina (UFSC), as mesmas foram mantidas por 96 horas a 4 °C para diminuir o teor de amido. O isolamento de cloroplasto foi realizado através de gradiente salino de Percoll, com adaptações específicas para as duas espécies de *Euterpe*. Com isso, 50 g de folhas foram desinfestadas com solução de 1,5% de Tween 20, e em seguida homogeneizadas em 600 ml de tampão de isolamento (Tabela 1) por 30 s em um liquidificador. Em seguida o homogeneizado foi filtrado primeiramente em duas camadas de gaze e após, em duas camadas de Miracloth (Merck, Darmstadt, Germany). O filtrado foi centrifugado por 15 min a 200 g. O pellet resultante, contendo núcleo e restos celulares, foi descartado, enquanto o supernadante contendo os cloroplastos foi centrifugado por 20 min a 3000 g. O pellet resultante, contendo os cloroplastos e alguns contaminantes, foi ressuspensionado cuidadosamente em 12 ml de tampão de lavagem (Tabela 1) com o auxílio de um pincel. Posteriormente, a suspensão de cloroplastos foi distribuída em 6 tubos (50 ml) contendo, em cada um, 20 ml de gradiente 70%-30% de Percoll (GE Healthcare, Uppsala, Sweden) (Tabela 1) e submetida à centrifugação por 25 min a 5000 g (Vieira *et al.*, 2014b).

Os cloroplastos intactos ficam estacionados na interface 70%-30%, os quais foram coletados e lavados duas vezes com 100 ml de tampão de lavagem, seguido por centrifugação de 20 min a 3000 g. O pellet resultante dos cloroplastos purificados foi armazenado a

temperatura de -80 °C até posterior uso para a extração do DNA plastidial. Todo o procedimento de extração foi realizado utilizando vidrarias, tampões e centrifugas previamente resfriados (0-4°C).

Tabela 1 – Componentes das soluções utilizadas para a extração de cloroplastos

<b>Solução</b>	<b>Componentes</b>
<b>Tampão de isolamento (pH 3.8)</b>	NaCl 1,25 M; ácido ascórbico 0,25 M; metabissulfito de sódio 10 mM; bórax 0,0125 M; Tris-HCl 50 mM (pH 8.0); EDTA 7 mM; PVP-40 1% (m/v); BSA 0,1 % (m/v)
<b>Tampão de lavagem (pH 8.0)</b>	NaCl 1,25 M; bórax 0,0125 M; Tris-HCl 50 mM (pH 8.0); EDTA 25 mM; PVP-40 1% (m/v); BSA 0,1 % (m/v)
<b>Gradiente 70%-30% de Percoll</b>	Consiste de tampão de lavagem com Percoll à concentração final de 70% (v/v) e 30% (v/v)

A partir do pellet de cloroplastos purificados que foi obtido inicia-se a lise dos cloroplastos mediante ressuspensão do pellet em 8 ml de tampão de isolamento de DNA (NaCl 0,1 M; Tris-HCl 0,1 M (pH 8.0); EDTA 50 mM; DTT 1 mM) acrescido de 1,5 ml de SDS 20%, 2 µl de 2-Mercaptoetanol e 30 µl de Proteinase K (10 mg/ml). A solução foi incubada por 4 h a 55°C e posteriormente, incubado no gelo por 5 min. Em seguida, foi adicionado aos cloroplastos lisados, 1,5 ml de KAc 5 M (pH 5.2) e incubado novamente no gelo por mais 30 min, sendo que ao término deste período, o lisado foi centrifugado por 15 min a 10000 g. O pellet resultante foi descartado, enquanto o supernadante foi extraído 2 vezes com um volume igual de fenol saturado e clorofórmio: álcool isoamílico (24:1) e centrifugado por 20 min a 10.000 g. A fase aquosa foi retirada e à mesma foi adicionado um volume igual de isopropanol e incubada por 12 h à -20°C. Em seguida, a solução incubada foi centrifugada por 20 min a 10.000 g, e o pellet de DNA resultante foi lavado sequencialmente com etanol 70% e etanol 96%, e por fim, ressuspendido em tampão TE (Tris-HCl 10mM; EDTA 1mM; pH 8.0).

As amostras de DNA foram tratadas com RNase e as bandas de DNA foram visualizadas em gel de agarose 0,8% (Vieira *et al.*, 2014b). Posteriormente, a concentração e a pureza das amostras foram avaliadas em espectrofotômetro QIAxpert System (Qiagen). Foi utilizado aproximadamente 50 ng do DNA, quantificado pelo kit Qubit dsDNA HS Assay para preparar bibliotecas de sequenciamento com kit Nextera DNA Sample Prep (Illumina

Inc., San Diego, CA, EUA) de acordo com as instruções do fabricante. A biblioteca obtida foi sequenciada usando a plataforma Illumina MiSeq (Illumina Inc., San Diego, CA, EUA) no Departamento de Bioquímica e Biologia Molecular, da Universidade Federal do Paraná (UFPR) Curitiba-PR, Brasil (Vieira *et al.*, 2014b).

Os programas Dual Organellar GenoMe Annotator (DOGMA) (Wyman *et al.*, 2004) e BLAST foram utilizados para anotação preliminar dos genes, com o objetivo de obter o genoma completo bem como as posições dos genes. A partir desta anotação inicial, foi realizada uma curadoria do genoma com a correta anotação dos prováveis codons de iniciação, os codons de parada e as posições do introns foram determinados com base em comparações com genes homólogos pertencentes a genomas plastidiais que já foram descritos e curados na literatura encontrados na base de dados *GenBank* (*Cocos nucifera* e *Mauritia flexuosa*). Todos os genes de RNAs transportadores (tRNA) foram verificados através do uso do servidor tRNAscan-SE (Lowe e Eddy, 1997). Os mapas físicos do genoma plastidial das duas espécies de *Euterpe* foram desenhados utilizando o programa Organellar Genome DRAW (OGDRAW) (Lohse *et al.*, 2013). Os genomas plastidiais de *E. edulis* e *E. oleracea* foram depositados no banco de dados do GenBank no NCBI.

## 2.2 Análise da estrutura do genoma e identificação das sequências repetidas

O script Perl de Nucleotídeo MUMmer (NUCMER) no MUMmer 3.0 (Kurtz *et al.*, 2004) e o programa MAUVE v. 2.4.0 (Darling *et al.*, 2004) foi utilizado para alinhar, visualizar e comparar as estruturas dos genomas plastidiais entre *E. edulis* e *E. oleracea*. Sequências simples repetidas (SSRs) foram detectadas usando o script Perl de MicroSatellite-MISA (Thiel *et al.*, 2003). Os limites estabelecidos foram oito unidades repetidas para SSRs mononucleotídios, quatro unidades repetidas para SSRs di e trinucleotídios e três unidades repetidas para SSRs de tetra, penta e hexanucleotídios. As repetições em tandem foram identificadas usando o programa Tandem Repeats Finder (TRF) (Benson, 1999). As configurações de parâmetros utilizadas foram 2, 7 e 7 para correspondência, incompatibilidade e indel, respectivamente. O escore de alinhamento mínimo para reportar as repetições e o tamanho máximo do período foi definido como 80 e 500, respectivamente. Posteriormente, as repetições encontradas foram verificadas manualmente e os resultados redundantes foram removidos. O REPuter (Kurtz *et al.*, 2001) foi usado para localizar repetições dispersas diretas e invertidas. O tamanho de repetição mínima foi ajustado para 30 pb e a identidade de repetições  $\geq 90\%$  (distanciamento = 3).

### 2.3 Análise de *sliding window* e predição de sítios de edição de RNA

Os pontos críticos de divergência de sequência entre *E. edulis* e *E. oleracea* foram caracterizados pela análise de *sliding window*. Os genomas plastidiais completos das 34 espécies de Arecaceae disponíveis no banco de dados e os plastomas das espécies *E. edulis* e *E. oleracea* foram alinhados usando ClustalW implementado no Mega 7.0 (Tamura *et al.*, 2013). Além disso, a análise de *sliding window* foi realizada com auxílio do programa DnaSP v.5 (Librado e Rozas, 2009). O comprimento da janela e o tamanho do passo foram definidos como 800 pb e 400 pb, respectivamente.

Os sítios de edição dos RNAs mensageiros plastidiais de espécies da família Arecaceae foram preditos pelo programa Predictive RNA Editor for Plants (PREP) suite (Mower, 2009). O programa PREP usa 35 genes de referência para detectar sites de edição de RNA em genomas plastidiais. O valor de corte foi ajustado para 0,8. Os seguintes genes foram utilizados como referência: *accD*, *atpA*, *atpB*, *atpF*, *atpI*, *ccsA*, *clpP*, *matK*, *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *petB*, *petD*, *petG*, *petL*, *psaB*, *psaI*, *psbB*, *psbE*, *psbF*, *psbL*, *rpl2*, *rpl20*, *rpl23*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*, *rps8*, *rps14*, *rps16*, e *yecf3*.

### 2.4 Inferências filogenômicas

Para a análise filogenômica foram selecionados 36 genomas da família Arecaceae do banco de dados de plastomas do Genbank (<https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/Arecaceae>) (Tabela 2). Como grupos externos, foram utilizados os plastomas das espécies *Bacteria australis* e *Dasyogon bromellifolius*. Os plastomas foram alinhados com as das espécies *E. edulis* e *E. oleraceae* utilizando o programa Mafft v 7 (Kato, 2013), implementado dentro do programa Aliview v.1.17.1 (Larsson, 2014). Para a reconstrução filogenômica foram utilizados dois métodos: a inferência bayesiana utilizando o programa Beast v.1.8 (Drummond e Rambaut, 2007), com 1 milhão de gerações utilizando 3 cadeias de Markov e a análise por máxima verossimilhança realizada no programa IQ-Tree v.1.0.1 (Nguyen *et al.*, 2015) com 1000 réplicas de bootstrap. Os logs e as árvores geradas foram combinados no LogCombiner e as árvores anotadas no TreeAnnotator, estes programas fazem parte do pacote de programas do BEAST v.1.8. Os resultados da corrida bayesiana foram analisadas no programa Tracer v.1.6 (Rambaut, 2009), sendo considerados valores de Effective Sample Size (ESS) ótimos acima de 200. Os modelos evolutivos para os conjuntos de dados foram determinados pelo programa jModelTest v.2.1.7 (Darriba *et al.*, 2012), o qual

indicou o melhor modelo sendo o GTR para a árvore MV e o modelo HKY+I para árvore BI. As árvores foram visualizadas e editadas no programa FigTree v.1.4.2 (Rambaut, 2009).

## 2.5 Estimativa da divergência gênica

Os 79 genes codificadores de proteínas presentes nos plastomas de *E. edulis* e *E. oleracea* e os plastomas das espécies *Bacteria australis* e *Dasyogon bromellifolius* (como grupo externo) foram alinhados individualmente utilizando o programa Mafft v 7 (Katoh, 2013), implementado dentro do programa Aliview v.1.17.1 (Larsson, 2014). Os modelos de substituição para cada gene foram selecionados usando jModelTest v.2.1.7 (Darriba *et al.*, 2012). A divergência dos genes foi estimada pela soma do comprimento total dos ramos.

Tabela 2 – Espécies utilizadas na inferência filogenética da posição de *E. edulis* e *E. oleracea*

<b>Espécies</b>	<b>Tamanho plastoma/pb</b>	<b>Nº acesso</b>
<i>Acoelorrhapha wrightii</i>	158,503	NC_029973.1
<i>Areca vestiaria</i>	130,917	NC_029972.1
<i>Arenga caudate</i>	159,974	NC_029971.1
<i>Bismarckia nobilis</i>	158,210	NC_020366.1
<i>Borassodendron machadonis</i>	158,144	NC_029969.1
<i>Brahea brandegeei</i>	158,733	NC_029968.1
<i>Calamus caryotoides</i>	157,270	NC_020365.1
<i>Caryota mitis</i>	159,819	NC_029948.1
<i>Chamaerops humilis</i>	159,553	NC_029967.1
<i>Chuniophoenix nana</i>	153,806	NC_029966.1
<i>Cocos nucifera</i>	154,731	NC_022417.1
<i>Colpotherinax cookii</i>	157,867	NC_028026.1
<i>Corypha lecomtei</i>	154,342	NC_029965.1
<i>Elaeis guineensis</i>	156,973	NC_017602.1
<i>Eremospatha macrocarpa</i>	154,031	NC_029964.1
<i>Eugeissona tristis</i>	153,304	NC_029963.1
<i>Leucothrinax morrisii</i>	158,432	NC_029961.1
<i>Lodoicea maldivica</i>	159,010	NC_029960.1
<i>Mauritia flexuosa</i>	156,367	NC_029947.1
<i>Metroxylon warburgii</i>	157,516	NC_029959.1
<i>Nypa fruticans</i>	158,391	NC_029958.1
<i>Phoenix dactylifera</i>	158,462	NC_013991.2
<i>Phytelephas aequatorialis</i>	159,075	NC_029957.1
<i>Pigafetta elata</i>	157,708	NC_029956.1
<i>Podococcus barteri</i>	157,668	NC_027276.1
<i>Pritchardia thurstonii</i>	157,909	NC_029955.1

<i>Pseudophoenix vinifera</i>	157,829	NC_020364.1
<i>Sabal domingensis</i>	157,835	NC_026444.1
<i>Salacca ramosiana</i>	157,047	NC_029954.1
<i>Serenoa repens</i>	158,952	NC_029953.1
<i>Syagrus coronata</i>	155,053	NC_029241.1
<i>Tahina spectabilis</i>	126,251	NC_029952.1
<i>Trithrinax brasiliensis</i>	158,487	NC_029951.1
<i>Veitchia arecina</i>	157,170	NC_029950.1
<i>Wallichia densiflora</i>	159,717	NC_029949.1
<i>Washingtonia robusta</i>	157,866	NC_029974.1
<i>Dasyopogon bromeliifolius</i> *	157,858	NC_020367.1
<i>Baxteria australis</i> *	158,858	NC_029970.1

\*Grupo externo.

### 3. RESULTADOS

#### 3.1 Tamanho, conteúdo gênico e análise estrutural do plastoma de *E. edulis* e *E. oleracea*

Os plastomas de *E. edulis* e *E. oleracea* apresentam uma estrutura quadripartida circular (Figuras 1 e 2), que é típica na maioria das angiospermas. O tamanho do genoma plastidial da espécie *E. edulis* é de 158.397 pb (Tabela 2) sendo que o par de regiões invertidas (IRs) é de 27.139 pb, separadas por uma grande cópia única (LSC) e uma pequena cópia única (SSC) de 86.716 pb e 17.403 pb, respectivamente (Tabela 3). Já o plastoma de *E. oleracea* é 405 pb maior quando comparado com a palmeira da Mata Atlântica, onde suas regiões LSC e SSC detém 87.117 e 17.453 pb de comprimento, respectivamente (Tabela 3), e a IRs com tamanho de 27,116 pb de comprimento cada uma. O conteúdo genético dos plastomas de *E. edulis* e *E. oleracea* não diferiram entre si, sendo constituídos de 113 genes únicos, sendo 79 genes codificadores de proteínas, 30 genes de tRNA e 4 genes rRNA (Tabela 4). No total, 19 genes são duplicados e localizados nas IRs (IRA e IRB), e o gene *ycf1* é parcialmente duplicado no limite das IRs com a região SSC.

A estrutura do genoma plastidial das duas espécies do gênero *Euterpe* foi comparada com plastomas de outras cinco espécies da família Arecaceae, através da análise de alinhamento do genoma múltiplo produzida pelo software MAUVE (Figura 3), não sendo encontrada nenhuma alteração estrutural relevante quando comparado o plastoma de *E. edulis* e *E. oleracea* com plastomas de outras três espécies da família Arecaceae. Do mesmo modo, o dot-plot foi realizado utilizando o programa MUMmer v3.0 (Kurtz *et al.*, 2004), para realizar

o teste foi usado o plastoma *E. edulis* contra o plastoma de *E. oleracea* para uma exploração adicional revelando uma estrutura e ordem de genes conservada típico da família Arecaceae (Figura 4).

Tabela 3 – Resumo dos dois genomas plastidiais

	<i>E. edulis</i>	<i>E. oleracea</i>
<b>Tamanho total do plastoma</b>	158.397	158.802
<b>Comprimento da região LSC</b>	86.716	87.117
<b>Comprimento da região IR</b>	27.139	27.116
<b>Comprimento da SSC</b>	17.403	17.453
<b>Conteúdo GC (%)</b>	37.35	37.37
<b>Número de genes únicos</b>	113	113
<b>Codificador de proteína</b>	79	79
<b>tRNA</b>	30	30
<b>rRNA</b>	4	4

Tabela 4 – Lista dos genes identificados no genoma plastidial de *Euterpe edulis* e *Euterpe oleracea*

<b>.Grupo de genes</b>	<b>Nome do gene</b>
<b>Expressão gênica</b>	
Gene de RNA Ribossomal	<i>rrn16<sup>b</sup>; rrn23<sup>b</sup>; rrn5<sup>b</sup>; rrn4.5<sup>b</sup></i>
Gene de RNA Transportador	<i>trnA</i> –UGC <sup>ab</sup> ; <i>trnC</i> –GCA; <i>trnD</i> –GUC; <i>trnE</i> –UUC; <i>trnFM</i> –CAU; <i>trnG</i> –UCC; <i>trnG</i> –GCC <sup>a</sup> ; <i>trnH</i> –GUG; <i>trnI</i> –CAU <sup>b</sup> ; <i>trnI</i> –GAU <sup>ab</sup> ; <i>trnK</i> –UUU <sup>a</sup> ; <i>trnL</i> –CAA <sup>b</sup> ; <i>trnL</i> –UAA <sup>a</sup> ; <i>trnL</i> –UAG; <i>trnM</i> –CAU; <i>trnN</i> –GUU <sup>b</sup> ; <i>trnP</i> –UGG; <i>trnQ</i> –UUG; <i>trnR</i> –ACG; <i>trnR</i> –UCU; <i>trnS</i> –GCU; <i>trnS</i> –UGA; <i>trnS</i> –GCA; <i>trnT</i> –UGU; <i>trnT</i> –GGU; <i>trnV</i> –GAC; <i>trnV</i> –UAC <sup>a</sup> ; <i>trnW</i> –CCA; <i>trnY</i> –GUA; <i>trnF</i> –GAA
Subunidade menor do ribossomo	<i>rps2; rps3; rps4; rps7<sup>b</sup>; rps8; rps11; rps12<sup>ab</sup>; rps14; rps15; rps16<sup>a</sup>; rps18; rps19</i>
Subunidade maior do ribossomo	<i>rpl2<sup>a</sup>; rpl14; rpl16<sup>a</sup>; rpl20; rpl22; rpl23; rpl32; rpl33; rpl36</i>
Subunidades da RNA polimerase plastidial	<i>rpoA; rpoB; rpoC1<sup>a</sup>; rpoC2</i>
Fator de iniciação da tradução	<i>infA</i>
<b>Genes fotossintéticos</b>	
Subunidades do fotossistema I (PSI)	<i>psaA; psaB; psaC; psaI; psaJ; ycf3<sup>a</sup>; ycf4</i>
Subunidades do fotossistema II (PSII)	<i>psbA; psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ</i>
Subunidades do citocromo	<i>petA; petB<sup>a</sup>; petD<sup>a</sup>; petG; petL; petN</i>
Subunidades da ATP sintase	<i>atpA; atpB; atpE; atpF<sup>a</sup>; atpH; atpI</i>
Subunidades de NADH desidrogenase	<i>ndhA<sup>a</sup>; ndhB<sup>ab</sup>; ndhC; ndhD; ndhE; ndhF; ndhG; ndhH; ndhI; ndhJ; ndhK</i>
Subunidade maior da Rubisco	<i>rbcL</i>
<b>Outros genes</b>	
Maturase	<i>matK</i>

Proteína da membrana do envelope	<i>cemA</i>
Subunidades da acetil-CoA carboxilase	<i>accD</i>
Gene Tipo-C da síntese do citocromo	<i>ccsA</i>
Protease	<i>clpP<sup>a</sup></i>
Componente do complexo TIC	<i>ycf1<sup>c</sup></i>
Genes de função desconhecida	<i>ycf2</i>

### 3.2 Análise da divergência dos genes nos plastomas de *E. edulis* e *E. oleracea*

Árvores baseadas em sequências de genes que codificam proteínas plastidiais permitiram a soma dos comprimentos de ramos totais para cada espécie/gene que foi utilizada para avaliar a divergência dos genes das espécies do gênero *Euterpe* a partir dos genes das espécies da família Dasypogonaceae (*Dasypogon bromellifolius* e *Baxteria australis*) que foram utilizados para a raiz das árvores (Figura 5 e 6). Os 79 genes codificadores de proteínas nos plastomas das espécies *E. edulis* e *E. oleracea* mostraram-se bastante conservados e a divergência mais significativa foi encontrada nos genes *rps16* e *ycf1* de ambas espécies (Figura 5 e 6).

### 3.3 Sítios de edição de mRNAs previstos nos plastomas de *E. edulis* e *E. oleracea*

Os sítios de edição de mRNA previstos em ambas espécies ocorreram na primeira ou segunda posição do codon (Tabela 5), e todas as mudanças de base observadas foram de citidina (C) para uridina (U). Entre os 35 genes plastidiais analisados pelo programa cp-PREP, 25 teriam a presença de 78 sítios de edição nos mRNAs de *E. edulis* e 76 sítios para *E. oleracea*. Apenas os sítios preditos para gene *ycf3* (1535; 1859) não se mostraram conservado para ambas as espécies (Tabela 5). Os genes com sítios previstos de edição de mRNA são: com 11 sítios: *ndhB*; com 7 sítios: *ndhF*; com 6 sítios: *rpoB*; com 5 sítios: *ndhD*; com 4 sítios: *accD*, *ndhA* e *petB*; com 3 sítios: *ccsA*; *matK*; *rpoA*; *rpoC1*, *rpoC2* e *ycf3*; com 2 sítios: *atpA*, *atppF*, *atpI*, *clpP*, *rpl2* e *rps4*; *atpB*, *ndhG*, *psaI*, *rps2* e *rps8* com apenas um sítio.

Figura 1 – Mapa dos genes do plastoma de *E. edulis*. Os genes representados dentro do círculo são transcritos no sentido horário e os genes representados fora são transcritos na direção anti-horária. Diferentes grupos funcionais de genes são codificados por cores. O cinza escuro no círculo interno corresponde ao conteúdo GC e o cinza claro ao conteúdo AT. Região de cópia única maior (LSC). Região de cópia única menor (SSC). Repetição invertida A/B (IRA/B)

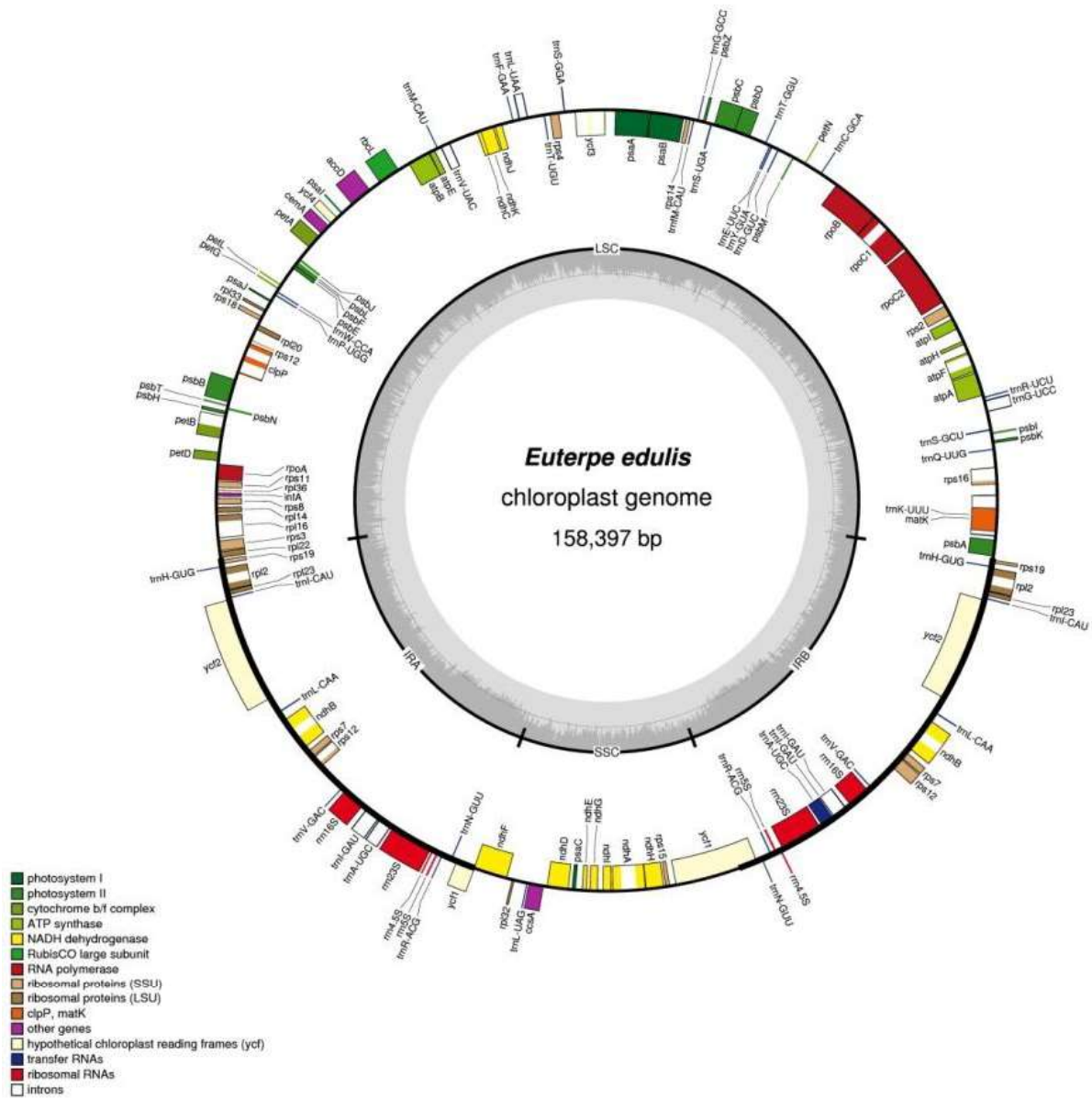


Figura 2 – Mapa dos genes do plastoma de *E. oleracea*. Os genes representados dentro do círculo são transcritos no sentido horário e os genes representados fora são transcritos na direção anti-horária. Diferentes grupos funcionais de genes são codificados por cores. O cinza escuro no círculo interno corresponde ao conteúdo GC e o cinza claro ao conteúdo AT. Cópia única maior (LSC). Cópia única menor (SSC). Repetição invertida A/B (IRA/B)

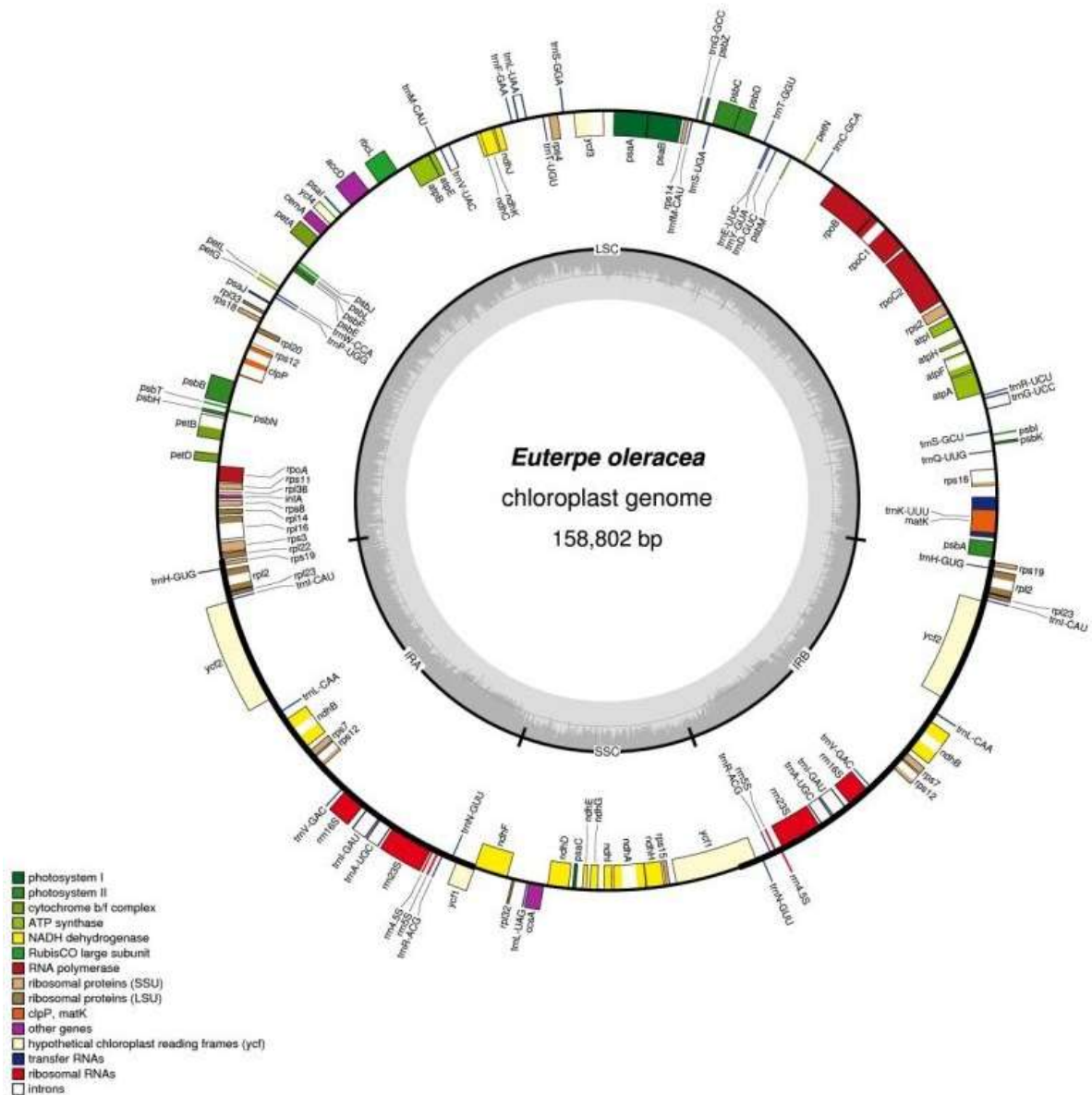


Figura 3 – Comparação da estrutura do genoma plastidial entre cinco espécies da família Areceae, produzido pelo programa MAUVE. Em branco CDS; em verde tRNA; em vermelho rRNA

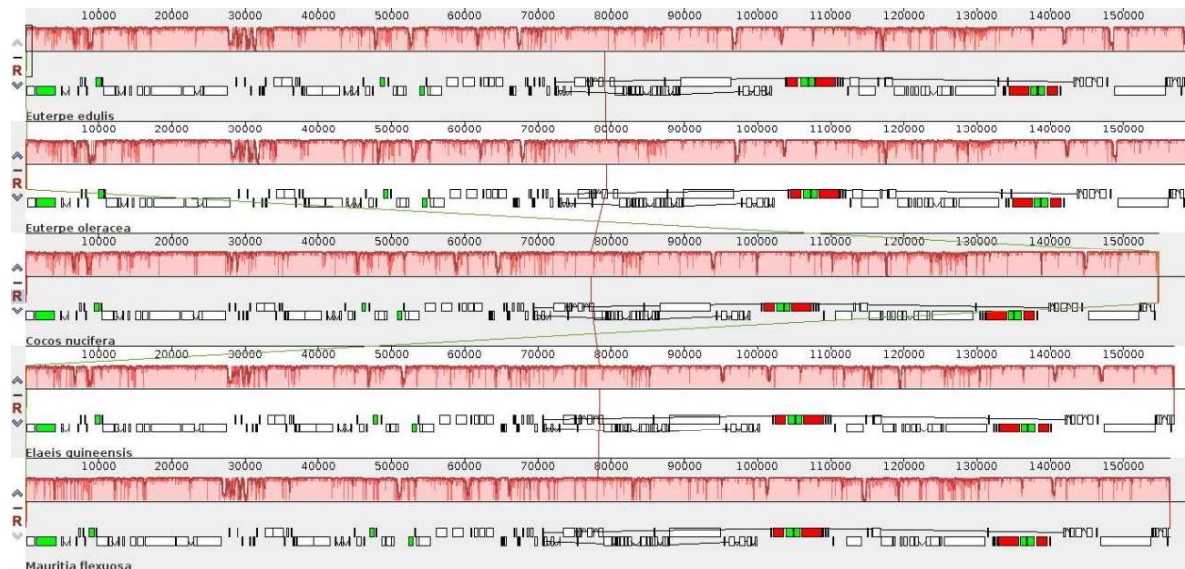


Figura 4 – Análise de Dot-plot de *Euterpe edulis* (Eixo-Y) e *Euterpe oleracea* (Eixo-X). Uma inclinação positiva indica que os pares de seqüências comparadas estão na mesma orientação. Uma inclinação negativa indica que os pares de seqüências comparadas podem estar alinhados, mas a sua orientação é oposta. Seqüências na mesma direção são vermelhas e inversões são azuis

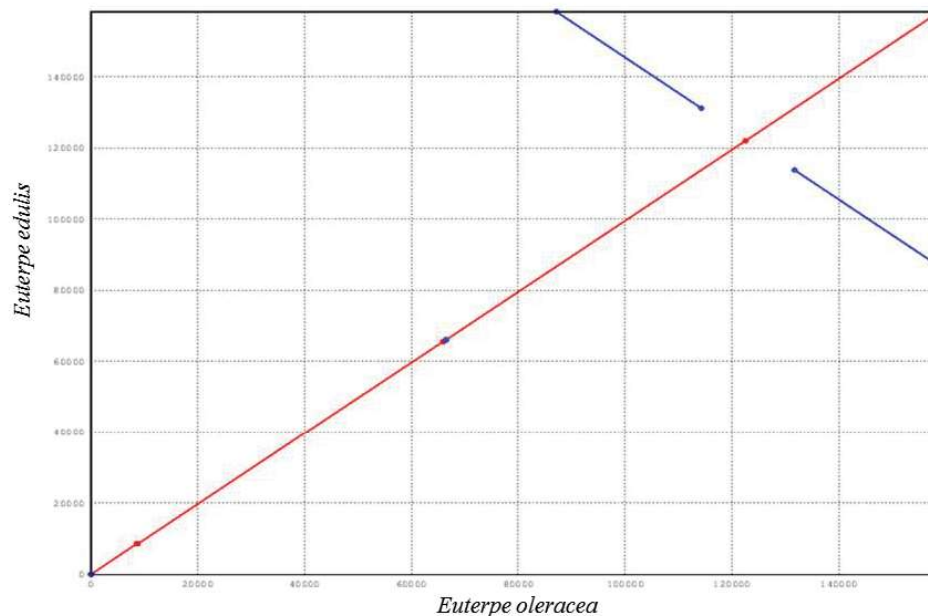


Figura 5 – Divergência nos genes que codificam as proteínas plastidiais das espécies de *Euterpe edulis* comparadas com o grupo externo da família Dasypogonaceae. A divergência nos genes foi estimada pela soma do comprimento total dos ramos em cada árvore de gene inferida

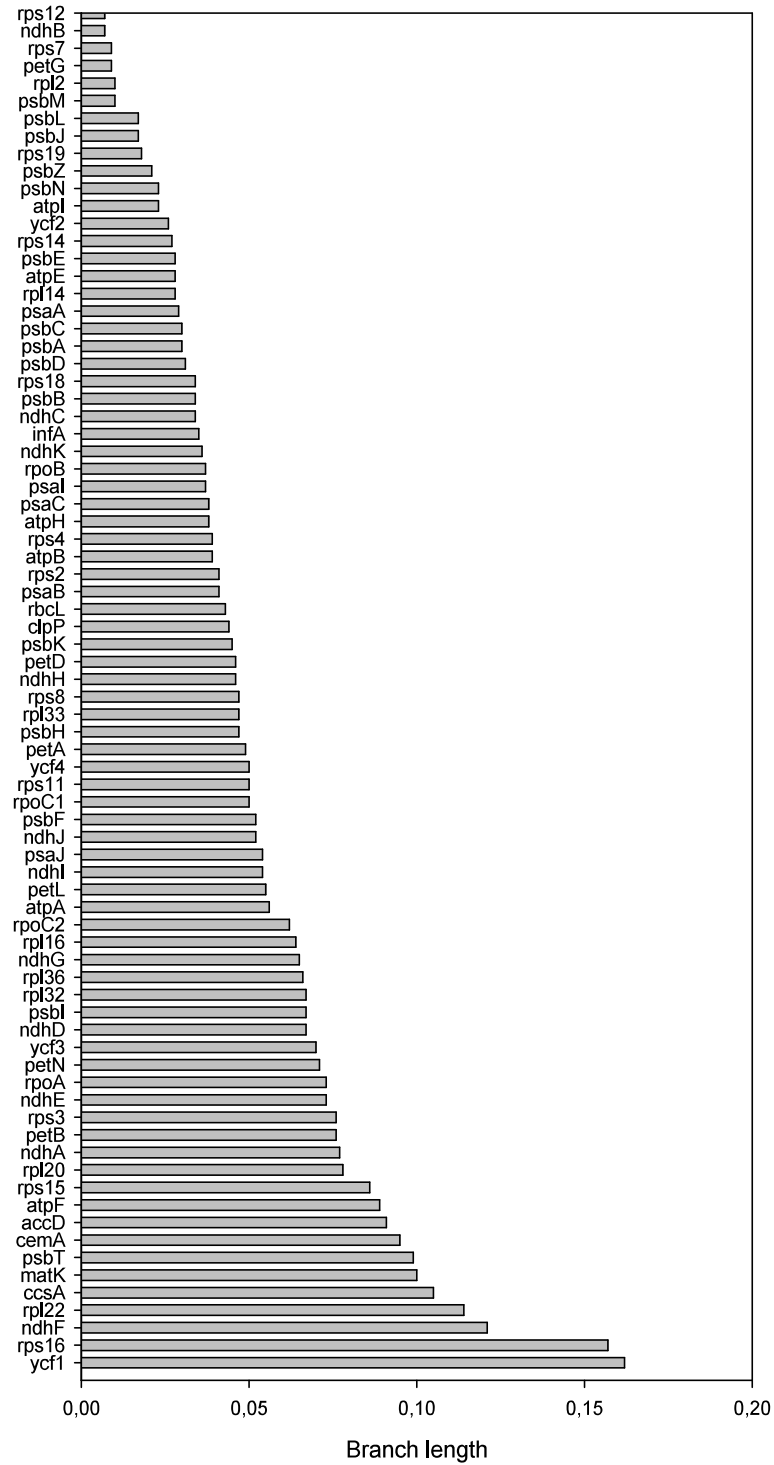


Figura 6 – Divergência nos genes que codificam as proteínas plastidiais das espécies de *Euterpe oleracea* comparadas com o grupo externo da família Dasypogonaceae. A divergência nos genes foi estimada pela soma do comprimento total dos ramos em cada árvore de gene inferida

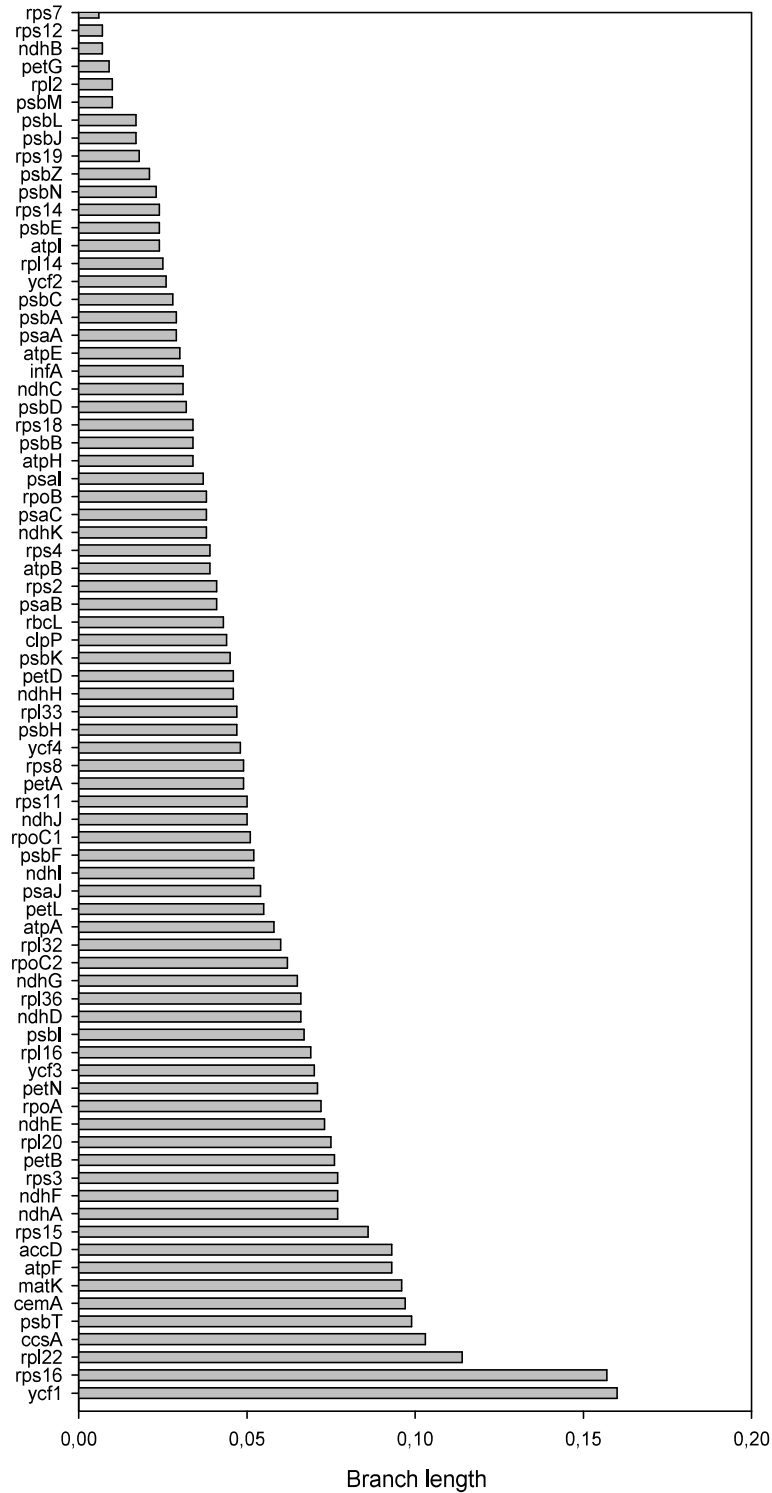


Tabela 5 – Edição de RNA predito pelo programa PREP-cp no genoma plastidial de *E. edulis* e *E. oleracea*

Gene	Posição no Alinhamento	Posição do		Efeito	Pontuação	<i>E.edulis</i>	<i>E.oleracea</i>	
		Codón						
<i>accD</i>	154	52		CGG (R) => TGG(W)	1.00	X	X	
	794*	265		TCG (S) => TTG (L)	0.80	X	X	
	1157*	386		TCA (S) => TTA (L)	1.00	X	X	
	1403	468		CCT (P) => CTT (L)	1.00	X	X	
<i>atpA</i>	914*	305		TCA (S) => TTA (L)	1.00	X	X	
	1148*	383		TCA (S) => TTA (L)	1.00	X	X	
<i>atpB</i>	1184**	395		TCA (S) => TTA (L)	1.00	X	X	
<i>atpF</i>	92	31		CCA (P) => CTA (L)	0.86	X	X	
	305	102		TCG (S) => TTG (L)	1.00	X	X	
<i>atpI</i>	428*	143		CCC (P) => CTC (L)	1.00	X	X	
	629*	210		TCA (S) => TTA (L)	1.00	X	X	
<i>ccsA</i>	647	216		ACT (T) => ATT (I)	0.86	X	X	
	670	224		CAT (H) => TAT (Y)	0.86	X	X	
	818	273		TCA (S) => TTA (L)	1.00	X	X	
<i>clpP</i>	82**	28		CAT (H) => TAT (Y)	1.00	X	X	
	559**	187		CAT (H) => TAT (Y)	1.00	X	X	
<i>matK</i>	188	63		TCA (S) => TTA (L)	1.00	X	X	
	922	308		CAC (H) => TAC (Y)	1.00	X	X	
	1270	424		CAC (H) => TAC (Y)	1.00	X	X	
<i>ndhA</i>	50*	17		TCG (S) => TTG (L)	1.00	X	X	
	476*	159		TCA (S) => TTA (L)	1.00	X	X	
	566*	189		TCA (S) => TTA (L)	1.00	X	X	
<i>ndhB</i>	1073	358		TCC (S) => TTC (F)	1.00	X	X	
	149*	50		TCA (S) => TTA (L)	1.00	X	X	
	467*	156		CCA (P) => CTA (L)	1.00	X	X	
	542*	181		ACG (T) => ATG (M)	1.00	X	X	
	586*	196		CAT (H) => TAT (Y)	1.00	X	X	
	704*	235		TCC (S) => TTC (F)	1.00	X	X	
	737**	246		CCA (P) => CTA (L)	1.00	X	X	
	1145	382		ACC (T) => ATC (I)	1.00	X	X	
	1285	429		CCT (P) => TCT (S)	0.80	X	X	
	1387	463		CTT (L) => TTT (F)	1.00	X	X	
<i>ndhD</i>	1496	499		TCG (S) => TTG (L)	1.00	X	X	
	2140	714		CCA (P) => TCA (S)	1.00	X	X	
	293	98		TCA (S) => TTA (L)	1.00	X	X	
	584	195		TCG (S) => TTG (L)	1.00	X	X	
	857	286		ACA (T) => ATA (I)	1.00	X	X	
	1103	368		TCA (S) => TTA (L)	0.80	X	X	
	1220	407		TCA (S) => TTA (L)	0.80	X	X	
	<i>ndhF</i>	62*	21		TCA (S) => TTA (L)	1.00	X	X

	290*	97	TCA (S) => TTA (L)	1.00	X	X
	392*	131	TCC (S) => TTC (F)	1.00	X	X
	442*	148	CAT (H) => TAT (Y)	1.00	X	X
	586	196	CTT (L) => TTT (F)	0.80	X	X
	1393	465	CAC (H) => TAC (Y)	1.00	X	X
	2093	698	TCC (S) => TTC (F)	1.00	X	X
<i>ndhG</i>	314	105	ACA (T) => ATA (I)	0.80	X	X
<i>petB</i>	691	231	CAT (H) => TAT (Y)	1.00	X	X
	884	295	ACA (T) => ATA (I)	1.00	X	X
	1130	377	CCC (P) => CTC (L)	1.00	X	X
	1141	381	CTT (L) => TTT (F)	1.00	X	X
<i>psaI</i>	80*	27	TCT (S) => TTT (F)	0.86	X	X
<i>rpl2</i>	2**	1	ACG (T) => ATG (M)	1.00	X	X
	1399	467	CCC (P) => TCC (S)	1.00	X	X
<i>rpl20</i>	26*	9	ACA (T) => ATA (I)	0.86	X	X
	308*	103	TCA (S) => TTA (L)	0.86	X	X
<i>rpoA</i>	368*	123	TCA (S) => TTA (L)	1.00	X	X
	830*	277	TCA (S) => TTA (L)	1.00	X	X
	887	296	TCG (S) => TTG (L)	1.00	X	X
<i>rpoB</i>	467*	156	TCG (S) => TTG (L)	0.86	X	X
	545*	182	TCA (S) => TTA (L)	1.00	X	X
	560*	187	TCG (S) => TTG (L)	1.00	X	X
	617*	206	CCG (P) => CTG (L)	0.86	X	X
	1994*	665	TCT (S) => TTT (F)	1.00	X	X
	2420*	807	TCA (S) => TTA (L)	0.86	X	X
<i>rpoC1</i>	41*	14	CCA (P) => CTA (L)	1.00	X	X
	617*	206	TCA (S) => TTA (L)	1.00	X	X
	1663	555	CAT (H) => TAT (Y)	1.00	X	X
<i>rpoC2</i>	1381	461	CAT (H) => TAT (Y)	0.86	X	X
			CGG (R) => TGG			
	2275	759	(W)	1.00	X	X
	2309*	770	TCG (S) => TTG (L)	1.00	X	X
<i>rps2</i>	248*	83	TCA (S) => TTA (L)	1.00	X	X
<i>rps8</i>	182	61	TCA (S) => TTA (L)	0.86	X	X
<i>rps14</i>	80*	27	TCA (S) => TTA (L)	1.00	X	X
	149*	50	CCA (P) => CTA (L)	1.00	X	X
<i>yef3</i>	44*	15	TCT (S) => TTT (F)	1.00	X	X
	1535	512	TCT (S) => TTT (F)	0.86	X	
	1859	620	TCC (S) => TTC (F)	1.00	X	

\*Validados em *Cocos nucifera*; \*\*Validados em *Elaeis guineenses*

### 3.4 Identificação de sequências repetidas e análise de *sliding window* em *E. edulis* e *E. oleracea*

A ocorrência e distribuição de sequências simples repetidas (SSRs) nos plastomas foram analisadas nas duas espécies do gênero *Eutерpe*. No total, 265 SSRs foram identificados para *E. edulis* e *E. oleracea* (Tabela 6 e 7). Os monopolímeros e os dipolímeros constituem a maior parte dos SSRs identificados (Tabela S10 e S11), representando mais de 93,5% do total com 67% constituídos por sequências A/T. A sequência, o tamanho e a localização de todos os SSRs dos plastomas de *E. edulis* e *E. oleracea* são exibidos nas tabelas suplementares S11 e S12. Entre os 265 SSRs identificados em *E. edulis*, 154 estão localizados nos espaçadores intergênicos (IGSs), 73 em sequências codificantes (CDSs) e 38 nos introns. Já nos 265 SSRs identificados em *E. oleracea*, 156 estão localizados nos IGSs, 69 CDSs e 40 nos introns.

Os 73 SSRs identificados em CDSs estão distribuídos entre 29 genes, dos quais se destacam os genes *ycf1* (19 SSRs), *ycf2* (8 SSRs), *rpoC2* (06 SSRs) e *accD* (6 SSRs) com o maior número de SSRs para espécie *E. edulis*. Para os 69 SSRs identificados em CDSs na espécie *E. oleracea*, a distribuição ocorreu entre 20 genes, dos quais o *ycf1* (21 SSRs), *ycf2* (7 SSRs) e *rpoC2* (6 SSRs), são os genes com maior número de SSRs. Os introns que abrigam SSRs em *E. edulis* estão localizados em 8 genes (Tabela 6), dos quais o gene *clpP* com 9 SSRs possui o maior número de marcadores. Na espécie *E. oleracea* os introns que possuem SSRs estão distribuídos em 12 genes (Tabela S11), sendo o gene *clpP* com maior número contendo 9 SSRs.

Além disso, através da análise de *sliding window* foi encontrado seis *hotspots* no plastomas das duas espécies, todos localizados dentro da região LSC (Figura 7). Com critério acima de 80% de correspondência em cópias repetidas, o *Tandem Repeats Finder* (TRF) mostrou que das 13 repetições em tandem 11 foram encontradas na LSC e 2 nas IRs do plastoma de *E. edulis* (Tabela 8). No genoma plastidial de *E. oleracea* das 21 repetições em tandem 16 estavam localizadas na LSC, três nas IRs e duas na SSC (Tabela 9). Das 13 repetições em tandem no plastoma de *E. edulis*, 10 estão localizadas em IGSs e 3 em CDS (Tabela 8).



Figura 7 – Análise Sliding-window do alinhamento dos plastomas completo das espécies *E. edulis* e *E. oleracea*. As regiões com alta variabilidade nucleotídica ( $Pi > 0.01$ ) são indicadas. Pi diversidade nucleotídica por cada window. Tamanho do window, 800 pb. Tamanho do passo, 400 pb

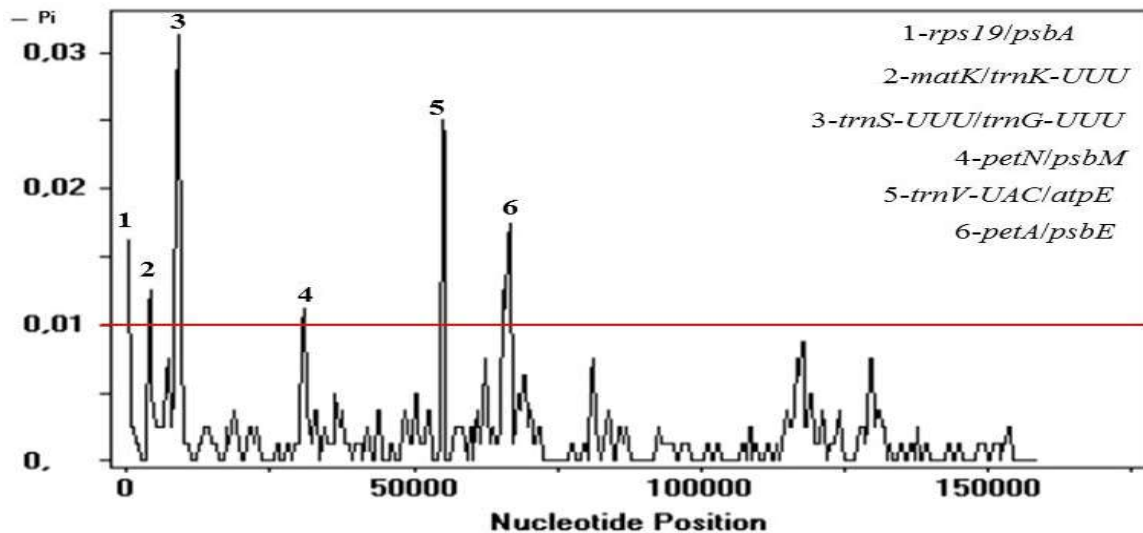


Tabela 8 – Distribuição das repetições em tandem no genoma plastidial de *Euterpe edulis*

Numero de cópias	Tamanho do consenso	Início	Fim	Pontuação	Localização
2	30	8872	8932	88	<i>trnS-GCU/trnG-GCC</i> (IGS)
2	30	14044	14101	100	<i>atpF/atpH</i> (IGS)
2	30	17068	17125	93	<i>rps2/rpoC2</i> (IGS)
2	20	33627	36667	82	<i>trnT-GGU/psbD</i> (IGS)
2	41	47576	47685	136	<i>rps4/trnT-GGU</i> (IGS)
3	27	52480	52533	81	<i>ndhC/trnV-UAC</i> (IGS)
3	33	69849	69921	80	<i>psaJ/rpl23</i> (IGS)
3	31	70355	70414	95	<i>rpl33/rps18</i> (IGS)
3	20	78982	79027	85	<i>petB/petD</i> (IGS)
3	34	82034	82082	89	<i>rps11</i> (CDS)
3	20	85572	85611	80	<i>rpl16/rps3</i> (IGS)
3	23	92036	92097	92	<i>ycf2</i> (CDS)
3	24	94455	94555	202	<i>ycf2</i> (CDS)

No plastoma de *E. oleracea* das 21 repetições em tandem encontradas, 14 foram em IGSs e 6 em CDSs (Tabela 9). Em relação às repetições dispersas (SDRs)  $\geq 30$  pb também foram identificadas 9 SDRs distribuídas entre IGSs com quatro, três em CDSs e duas nos introns dos genes, das quais 5 são invertidas e 4 são diretas para o genoma plastidial de *E. edulis*, respectivamente (Tabela 10). No plastoma de *E. oleracea* 11 SDRs foram localizadas,

cinco invertidas e seis diretas localizadas em regiões codificantes e intergênicas bem como nos introns dos genes e regiões codificantes (Tabela 11).

Tabela 9 – Distribuição das repetições em tandem no genoma plastidial de *Euterpe oleracea*

Numero de cópias	Tamanho do consenso	Início	Fim	Pontuação	Localização
3	39	8511	8635	101	<i>trnS-GCU/trnG-GCC</i> (IGS)
2	22	8666	8711	92	<i>trnS-GCU/trnG-GCC</i> (IGS)
2	25	9120	9169	94	<i>trnS-GCU/trnG-GCC</i> (IGS)
2	30	9158	9218	88	<i>trnS-GCU/trnG-GCC</i> (IGS)
2	30	14334	14391	100	<i>atpH/atpI</i> (IGS)
2	30	17360	17417	93	<i>rpoC2</i> (CDS)
2	24	30481	30537	114	<i>petN/psbM</i> (IGS)
4	16	30808	30866	95	<i>petN/psbM</i> (IGS)
2	20	33930	33970	82	<i>trnF-GGU/psbD</i> (IGS)
3	41	47895	48004	136	<i>rps4/trnT-UGU</i> (IGS)
2	27	52798	52851	81	<i>ndhC/trnV-UAC</i> (IGS)
2	33	70256	70328	80	<i>rpl33/rps18</i> (IGS)
2	31	70762	70821	95	<i>rps18/rpl20</i> (IGS)
2	20	79392	79433	84	<i>petB/petD</i> (IGS)
2	24	82438	82486	89	<i>infA</i> (CDS)
3	20	85974	86013	80	<i>rps3</i> (CDS)
3	22	92438	92499	90	<i>yef2</i> (CDS)
6	17	94850	94940	86	<i>yef2</i> (CDS)
3	24	98857	94933	154	<i>yef2</i> (CDS)
2	29	116524	116582	84	<i>rpl32/trnL-UAG</i> (IGS)
2	45	122477	122552	189	<i>ndhI</i> (CDS)

Tabela 10 – Distribuição dos locos das sequências dispersas com repetições diretas (D) e invertidas (I) do genoma plastidial de *Euterpe edulis*

Tipo	Tamanho (pb)	Repetição 1 (início)	Repetição 2 (início)	Repetição 1 (localização)	Repetição 2 (localização)
D	37	39698	41922	<i>psaB</i> (CDS)	<i>psaA</i> (CDS)
D	34	44628	101932	<i>yef3</i> (intron)	<i>rps12/trnV-GAC</i> (IGS)
D	32	8315	36480	<i>Yef2</i> (CDS)	<i>Yef2</i> (CDS)
D	30	8993	47914	<i>trnS-GCU/trnG-GCC</i> (IGS)	<i>psbC/trnS-UGA</i> (IGS)
I	34	8313	46114	<i>psbI/trnS-GCU</i> (IGS)	<i>trnT-UGU//trnL-UAA</i> (IGS)
I	37	44628	143144	<i>yef3</i> (intron)	<i>rps12/rps7</i> (IGS)
I	32	36480	46114	<i>psbC/trnS-UGA</i> (IGS)	<i>rps12/trnV-GAC</i> (IGS)
I	30	92055	153007	<i>yef2</i> (CDS)	<i>yef2</i> (CDS)
I	30	33	72	<i>rps19/psbA</i> (IGS)	<i>rps19/psbA</i> (CDS)

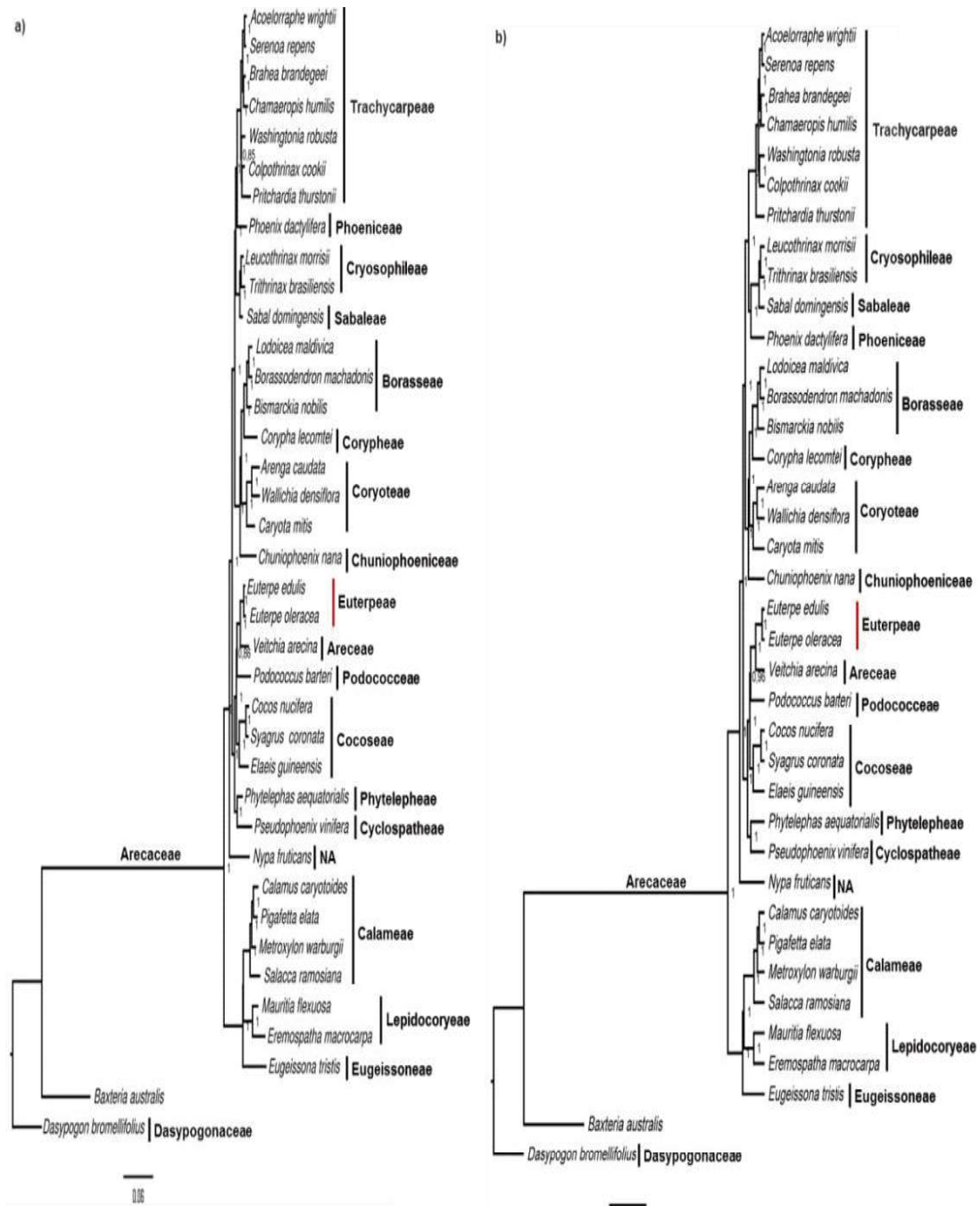
Tabela 11 – Distribuição dos locos das sequências dispersas com repetições diretas (D) e invertidas (I) do genoma plastidial de *Euterpe oleracea*

Tipo	Tamanho (pb)	Repetição 1 (início)	Repetição 2 (início)	Repetição 1 (localização)	Repeat 2 (localização)
D	34	39999	42223	<i>psaB</i> (CDS)	<i>psaA</i> (CDS) <i>rps12/trnV-GAC</i> (IGS)
D	37	44931	102310	<i>ycf3</i> (intron)	<i>psaC/ndhE</i> (IGS)
D	36	9270	121244	<i>trnS-GCU/trnG-GCC</i> (IGS)	<i>trnS-GCU</i> (CDS)
D	32	8328	32782	<i>trnS-GCU</i> (CDS)	<i>trnT-UGU//trnL-UAA</i> (IGS)
D	31	47929	47968	<i>trnT-UGU//trnL-UAA</i> (IGS)	<i>ndhF/rpl32</i> (IGS)
D	30	70802	116546	<i>rps18</i> (CDS)	<i>trnS-GCU/trnG-GCC</i> (IGS)
I	37	9119	9133	<i>trnS-GCU/trnG-GCC</i> (IGS)	<i>trnS-GGA</i> (CDS)
I	34	8326	46429	<i>trnS-GCU</i> (CDS)	<i>psaC/ndhE</i> (IGS)
I	31	9778	121249	<i>trnS-GCU/trnG-GCC</i> (IGS)	<i>rps19/psba</i> (IGS)
I	30	47	86	<i>rps19/psba</i> (IGS)	<i>trnS-GGA</i> (CDS)
I	32	36782	46429	<i>trnS-UGA</i> (CDS)	

### 3.5 Inferência filogenômica

As árvores filogenômicas foram realizadas utilizando tanto o método de máxima verossimilhança (MV) quanto à inferência bayesiana (BI). A árvore filogenética pela MV apresentou valores de suporte do *bootstrap* significativos (100%) para todos os nós com exceção da relação entre *Washingtonia robusta* e o clado formado por *Colpothrinax cookii* e *Pritchardia thurstonii* com suporte de 85% e entre Podococceae e Areceae possuindo suporte de 86% (Figura 8a). Os valores de probabilidade à posteriori de BI foram de 100% para todos os clados, exceto para a relação entre Podococceae e Areceae que foi de 96% (Figura 8b). Assim, ambos os métodos apresentaram altos valores de suporte para todos os taxóns analisados com valores de suporte acima de 0,7. Nas análises realizadas pelos dois métodos, ambos classificaram corretamente as espécies dentro de suas respectivas tribos taxonômicas. As espécies *E. edulis* e *E. oleracea* agruparam dentro da tribo Euterpeae tendo como grupo irmão a tribo Areceae da espécie *Veitchia arecina*, sendo a tribo Podococceae (espécie *Podococcus barteri*) clado irmão e a tribo Cocoseae mais basal. Ambas as espécies de *Euterpe* se encontram em ramos separados, com valores de suporte altos, validando sua classificação em espécies distintas (Figura 8a/b).

Figura 8 – Reconstrução filogenômica das espécies *E. edulis* e *E. oleracea* utilizando 34 plastomas de Arecaceae. a) Árvore calculada pelo método de máxima verossimilhança, com 1000 replicas de bootstrap; b) Árvore inferida pelo método bayesiano com 3 cadeias de Markov e 1 milhão de gerações. Como grupo externo, foram usados os genomas de *Bacteria australis* e *Dasypogon bromellifolius*. Em ambas as análises, foram descartados valores abaixo de 0,7.



## 4. DISCUSSÃO

### 4.1 Tamanho e análise estrutural dos plastomas de *E. edulis* e *E. oleracea*

O tamanho, a ordem dos genes e o conteúdo gênico nas regiões LSC, SSC e IRs de *E. edulis* e *E. oleracea* são altamente conservados, características marcantes em diferentes espécies da família Arecaceae (Barret *et al.*, 2016). Comparando o plastoma de *E. edulis* e *E. oleracea* com os plastomas de 36 espécies de palmeiras disponíveis no banco de dados, foi observada uma variação no tamanho, variando de 126,251 pb (*Tahina spectabilis*) a 159,974 pb (*Arenga caudata*) o que não reflete no arranjo filogenômico sendo que ambos os genoma de *E. edulis* e *E. oleracea* possuem aproximadamente 158.000 pb, implicando em relações mais genéticas que estruturais do seu genoma. O alinhamento múltiplo do genoma produzido pelo software MAUVE apenas confirmou a estrutura conservada com ausência de rearranjos em ambos plastomas. No alinhamento foi possível observar que os plastomas de ambas as espécies são conservados a nível estrutural, sem inversões, deleções ou rearranjos para ambas as espécies analisadas. O gene *infA*, que está ausente em algumas dicotiledôneas e monocotiledôneas (Mardanov *et al.*, 2008), está presente no genoma plastidial das duas *Euterpes* bem como no plastoma de outras espécies da família, gerando a sequência proteica completa, indicando que este gene está produzindo proteína funcional.

### 4.2 Divergência de genes em *E. edulis* e *E. oleracea*

A comparação dos genes que codificam proteínas entre os plastomas de *E. edulis* e *E. oleracea* revelou genes altamente conservados, sendo os genes *ycf1* e *rps16* os mais divergentes. O gene *ycf1* é essencial para a viabilidade celular (Drescher *et al.*, 2000) e sua função foi recentemente caracterizada como uma subunidade (Tic214) do complexo TIC, um complexo protéico relacionado a importação de proteínas localizado na membrana interna do plastídio para o interior da célula vegetal (Kikuchi *et al.*, 2013). No entanto, este gene é descrito como um dos genes mais divergentes presente em plastomas de angiospermas com vários relatos de perdas, pseudogenização e/ou transferência para o núcleo (Wicke *et al.*, 2011; Vries *et al.*, 2015), sendo anotado como pseudogene na maioria das espécies de palmeiras depositadas no NCBI (Barret *et al.*, 2015), encontra-lo na forma aparentemente completa e funcional reforça uma evidência de espécies bem adaptadas evolutivamente, haja vista que o TIC é um sistema altamente complexo e dinâmico (Kovács-Bogdán *et al.*, 2010).

O gene *rps16* possui diversos trabalhos relatando perdas e/ou transferência para o núcleo em várias espécies da ordem Malpighiales investigadas, incluindo a perda da funcionalidade. Contudo, embora seja um gene altamente divergente, os estudos mostram que o mesmo encontra-se funcional nas espécies da família Arecaceae até aqui estudadas, visto que este gene codifica a proteína ribossômica (S16) localizada na subunidade menor (30S) do ribossomo plastidial, demonstra ser essencial para a viabilidade celular (Fleischmann *et al.*, 2011).

### 4.3 Previsão de sítios de edição de RNA

A edição de RNA mensageiro é um processo de pós-transcrição muito observado no genoma plastidial podendo resultar em alternância de tradução por induzir a ocorrência de substituições (Zanduetta-Criado e Bock, 2004). A edição de mRNA pode criar códons de iniciação e terminação, bem como restaurar aminoácidos conservados modificando a sequência dos mRNAs a serem traduzidos, produzindo assim uma proteína funcional (Tsudzuki *et al.*, 2001; Takenaka *et al.*, 2013). Dos 55 sítios previstos para *E. edulis* e *E. oleracea* conservados em relação a espécie *Cocos nucifera*, 41 foram validados por Huang *et al.* (2013) e cinco validados em *Elaeis guineensis* por Uthaipaisanwong *et al.* (2012) (Tabela 5). O gene *ndhB* que contém o maior número de sítios preditos, apenas seis foram validados, já o gene *rpoB* teve todos os seus seis sítios previstos validados (Huang *et al.*, 2013). Nas angiospermas, a edição de RNA ocorre quase que exclusivamente com a substituição de C para U (Bock, 2000; Covello e Gray, 1993). Se os outros 37 sítios previstos nas *Euterpes* foram validados, esse número pode se aproximar de 75 sítios de edição de mRNA já validados para *Cocos nucifera* e distante dos sítios presentes em *Elaeis guineensis* (32 sítios validados); *Arabidopsis thaliana* (34 sítios validados); *Nicotiana tabacum* (37 sítios validados); *Zea mays* (26 sítios validados); *Oryza sativa* (21 sítios validados), evidenciando que embora as características de edição sejam semelhantes, o número total de sítios de edição podem variar entre diferentes espécies (Tillich *et al.*, 2006). Além disso, devido à conservação evolutiva da edição do mRNA, espécies relacionadas geralmente compartilham mais sítios de edição (Chen *et al.*, 2011). Adicionalmente, existem mais sítios de edição compartilhados dentro da família Poaceae do que aqueles compartilhados entre as gramíneas e as dicotiledôneas (Guzowska-Nowowiejska *et al.*, 2009). Takenaka *et al.* (2013) sugerem que o mecanismo de edição de mRNA tem o papel de controlar moléculas funcionais de RNA maduro disponíveis para a tradução ou como controle pós transcricional sob diferentes

condições ambientais. A palmeira *E. edulis* apresentou 23 predições de sítios mRNA que ainda não foram observados em outras palmeiras, dos quais apenas dois não são compartilhados com a espécie *E. oleracea* que apresentou 21 sítios de mRNA igualmente inéditos. Estudos subsequentes são necessários para possíveis validações desses novos sítios.

#### 4.4 Sequências repetidas e hotspots de divergência de nucleotídeos

O uso de SSRs plastidiais como marcadores moleculares foi introduzido por Powell *et al.*, (1995). Essas sequências exibem variação de comprimento e polimorfismo, que podem ser uma ferramenta útil para analisar a diversidade genética, e já foram identificados em vários plastomas de plantas terrestres (Marshall *et al.*, 2001; Provan *et al.*, 2001; Raubeson *et al.*, 2007; Vieira *et al.*, 2014a, 2016a). Os SSRs identificados neste estudo podem ser avaliados para análises intraspecíficas e interespecíficas de polimorfismos (Vieira *et al.*, 2014a, Rogalski *et al.*, 2015). O padrão de repetições em tandem aqui encontrado segue o padrão encontrado em outras palmeiras como *Cocos nucifera* (13 repetições em tandem), *Elaeis guineensis* (7 repetições em tandem) e a *Phoenix dactylifera* (11 repetições), embora na literatura existam dados que mostram que o número possa diferir entre espécies da mesma família (Huang *et al.*, 2013; Uthaisaisanwong *et al.*, 2012). Quando comparando com outras monocotiledôneas, como a *Bamboo emeiensis* (Poaceae) que possuem 39 repetições em tandem, deixa evidente que as palmeiras possuem menos dessas repetições (Zhang *et al.*, 2011). As sequências repetitivas em genomas plastidiais podem produzir recombinações e induzir a rearranjos (Rogalski *et al.*, 2006; Guisinger *et al.*, 2010), com papel crucial para a estabilização do plastoma ou mesmo na sua viabilidade (Marechal *et al.*, 2009; Vieira *et al.*, 2016). Por isso, o conteúdo relativamente baixo do DNA repetitivo nos genomas plastidiais em espécies da família das palmeiras sugere um grau relativamente maior de estabilidade e conservação nas diferentes espécies de palmeiras até aqui analisadas (Huang *et al.*, 2013).

Sequências de plastomas são uma fonte valiosa de marcadores moleculares para decifrar as relações filogenéticas entre táxons intimamente relacionados (Rogalski *et al.*, 2015, Daniell *et al.*, 2016). Estudos anteriores mostraram maior conservação de genes em IRs quando comparados com as regiões LSC e SSC (Zhu *et al.*, 2016), e as taxas de substituições sinônimas e não sinônimas são muito maiores na LSC quando comparadas com as IRs (Kim *et al.*, 2004; Hu *et al.*, 2015). Esses resultados corroboram com o encontrado nesta pesquisa, pois todos os pontos de maior polimorfismo foram encontrados na região LSC. A característica estrutural conservada das regiões IRs é resistente a alterações devido à alta taxa

de recombinação homóloga (Perry e Wolfe, 2002). A presença das regiões IRs pode ajudar a estabilizar o genoma plastidial. A evidência mais direta para essa sugestão é que mais rearranjos ocorreram dentro do grupo de vegetais que perderam uma cópia das IRs do que naqueles que mantêm as duas cópias estáveis (Raubeson e Jansen, 2005; Chang *et al.*, 2006; Hansen *et al.*, 2007; Zhu *et al.*, 2016). Consequentemente, podemos inferir que as taxas evolutivas do plastoma na família das palmeiras são relativamente leves, a julgar pela menor flutuação comparativamente das regiões IRs com outras famílias de angiospermas. Alterações nas regiões IRs ocorreram sutilmente durante a história evolutiva das angiospermas (Guisinger *et al.*, 2010; Yang *et al.*, 2010; Wu *et al.*, 2010). O caso mais extremo foi o encontrado em *Pelargonium hortorum* (Geraniaceae) onde uma das cópias das IRs expandiu para 76 kb cerca de três vezes maior que a repetição inversa (Palmer *et al.*, 1987). Dos seis pontos de divergência de nucleotídeos plastidiais identificados neste estudo com base na análise de *sliding window*, apenas o *matK* já foi usado anteriormente para estudos filogenéticos dentro família Arecaceae (Faurby *et al.*, 2016; Barret *et al.*, 2016; Baker *et al.*, 2009). Além disso, os genes plastidiais mais divergentes identificados aqui, como *ycf1*, *rps16*, *ndhF*, *rpl22* e *ccsA*, são boas fontes de informação genética úteis para melhorar a resolução filogenética na própria tribo Euterpeae que ainda não tem todas suas relações filogenéticas resolvidas.

#### **4.5 As relações entre as tribos dentro da família Arecaceae com base em genomas plastidiais inteiros**

Estudos evolutivos utilizam árvores filogenéticas como fonte de informação sobre processos macroevolutivos que possibilitam a reconstrução das relações evolutivas em nível de espécies, permitindo testar hipóteses que somente os dados morfológicos e taxonômicos não conseguem (Comer *et al.*, 2015). Várias filogenias baseadas em sequências nucleotídicas nucleares, mitocôndriais e plastidiais têm sido usadas para resolver as relações entre subfamílias, tribos, gêneros e espécies dentro da família Arecaceae (Asmusse *et al.*, 2006; Lin *et al.*, 2010; Kissling *et al.*, 2012; Baker e Couvreur, 2013; Couvreur *et al.*, 2015; Comer *et al.*, 2016; Faurby *et al.*, 2016; Daniell *et al.*, 2016). No entanto, as árvores filogenéticas são, muitas vezes, altamente incompletas a este nível, principalmente pela ausência de genes que possam mostrar não só a evolução de um determinado marcador, mas sim a história evolutiva da espécie (Meerow *et al.*, 2015). A análise do plastoma completo permite um nível de informação muito mais preciso e robusto para a reconstrução genética em nível de espécie

(Faurby *et al.*, 2016). A reconstrução da filogenômica permitiu classificar as espécies *Euterpe edulis* e *Euterpe oleracea*, dentro da tribo Euterpeae o que corroborou com as análises morfológicas de suas estruturas (Oliveira *et al.*, 2016).

Em análises filogenéticas utilizando grupos de genes plastidiais, a tribo Areceae e Euterpeae formou um clado fortemente apoiado. No entanto, as relações entre Areceae + Euterpeae com Leopoldinieae, Geonomateae, Manicarieae, e Pelagodoxeae não foram bem suportados (Barrett *et al.*, 2016). Com isso, evidenciou-se que a filogenômica com plastomas é mais eficiente para traçar a história evolutiva, uma vez que todos os nós nas árvores obtidas a partir plastomas completos tiveram altos valores de suporte.

A filogenômica permitiu também a inferência evolutiva acurada da família Arecaceae, o resultado chamou a atenção pelo posicionamento taxonômico da tribo Areceae em relação às tribos Podococcoseae e Cocoseae. O trabalho de Comer *et al.*, (2016) utilizando marcadores nucleares, posicionava Areceae como um clado mais relacionado a Cocoseae e tendo Podococcoseae mais basal a esses dois clados. Contudo o trabalho de Baker *et al.*, (2009) utilizando alguns marcadores plastidiais (*atpB*; *matK*; *ndhF*; *rbcL*; *rps16*; *trnD-trnT*; *trnQ-rps16*; *trnL-trnF*), além de dados morfológicos e de RFLP, mostrava que Areceae é um grupo mais relacionado com Podococcoseae sendo Cocoseae mais basal. O trabalho de Faurby *et al.*, (2016) que utilizando sequências plastidiais, também corroboram com a topologia encontrada no presente estudo. O arranjo filogenético também é corroborado pelo trabalho de Barret *et al.*, (2016) que utilizou o genoma plastidial para obter a reconstrução evolutiva mais profunda.

## 5. CONSIDERAÇÕES FINAIS

*E. edulis* e *E. oleracea* são as primeiras espécies do gênero *Euterpe* com plastoma completamente sequenciado, montadas, anotadas, e caracterizadas em detalhes. Ao comparar esses dois plastomas, observou-se que ambos compartilham a mesma organização estrutural e o conteúdo gênico é mantido nas outras espécies de palmeiras. Não foram observadas perdas em genes essenciais para a viabilidade celular. Além disso, estes plastomas de ambas as *Euterpes* mostraram que os genes mais divergentes nessas espécies foram o *ycf1* e o *rps16*. Vários SSRs foram detectados neste plastoma e podem ser usados para estudos genéticos intraespecíficos e interespecíficos. A análise de mRNAs revelou um número elevado de sítios de edição para as duas espécies. A palmeira *E. edulis* apresentou 23 predições de sítios mRNA que ainda não foram observados em outras palmeiras, dos quais apenas dois não são

compartilhados com a espécie *E. oleracea*. Estudos subsequentes pelo método de PCR são necessários para possíveis validações experimentais desses novos sítios. Os resultados revelam que as árvores filogenômicas possuem topologias consistentes com o descrito na literatura e com valores de suportes altos nos dois métodos distintos utilizados, o que reforça a importância do sequenciamento completo de plastomas para a correta inferência evolutiva das espécies.

## REFERÊNCIAS

- Alkatib S., Scharff L.B., Rogalski M., Fleischmann T.T., Matthes A., Seeger S., Schöttler M.A., Ruf S., Bock R. (2012) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genetics*, v. 8, n. 11, p. e1003076.
- Asmussen, C.B., Dransfield, J., Deickmann V., Barfod A.S., Pintaud, J.C., Baker, W.J. (2006) A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. *Botanical Journal of the Linnean Society*, v. 151, n. 1, p. 15-38.
- Baker, W.J. (2015) A revised delimitation of the rattan genus *Calamus* (Arecaceae). *Phytotaxa*, v. 197, n. 2, p. 139-152.
- Baker, W.J., Couvreur, T.L.P. (2013) Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *Journal of Biogeography*, v. 40, n. 2, p. 274-285.
- Baker, W.J., Savolainen, V., Asmussen-Lange, C.B., Chase, M.W., Dransfield, J., Forest, F., Harley, M.M., Uhl, N.W. & Wilkinson, M. (2009) Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Systematic Biology*, v. 58, n. 2, p. 240-256.
- Barrett, C.F., Baker, W.J., Comer, J.R., Conran, J.G., Lahmeyer, S.C., Leebens-Mack, J.H., Li, J., Lim, G.S., Mayfield-Jones, D.R., Perez, L., et al. (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist*, v. 209, n. 2, p. 855-870.
- Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, v. 27, n. 2, p. 573.
- Bock, R. (2000) Sense from nonsense: how the genetic information of chloroplasts is altered by RNA editing. *Biochimie*, v. 82, n. 6-7, p. 549-557.
- Bock, R. (2001) Transgenic plastids in basic research and plant biotechnology. *Journal of Molecular Biology*, v. 312, n. 3, p. 425-438.
- Bock, R. (2007) Structure, function, and inheritance of plastid genomes. In: *Cell and Molecular Biology of Plastids*. Bock, R. (Ed). Springer Berlin Heidelberg. p. 29-63.
- Brancalion, P.H.S., Vidal, E., Lavorenti, N.A., Batista, J.L.F., Rodrigues, R.R. (2012) Soil-mediated effects on potential *Euterpe edulis* (Arecaceae) fruit and palm heart sustainable

management in the Brazilian Atlantic Forest. *Forest Ecology and Management*, v. 284, p. 78-85.

Carvalho, C.S., Ribeiro, M.C., Côrtes, M.C., Galetti, M., Collevatti, R.G. (2015) Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. *Heredity*, v. 115, n. 3, p. 216.

Chang, C.C., Lin, H.C., Lin, I.P., Chow, T.Y., Chen, H.H. et al. (2006) The Chloroplast Genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative Analysis of Evolutionary Rate with that of Grasses and Its Phylogenetic Implications. *Molecular Biology and Evolution*, v. 23, n. 2, p. 279-291.

Chen H., Deng L., Jiang Y., Lu P., Yu, J. (2011) RNA editing sites exist in protein-coding genes in the chloroplast genome of *Cycas taitungensis*. *Journal of Integrative Plant Biology*, v. 53, n. 12, p. 961-970.

Comer, J.R., Zomlefer, W.B., Barrett, C.F., Stevenson, D.W., Heyduk, K., Leebens-Mack J.H. (2015) Resolving relationships within the palm subfamily Arecoideae (Areaceae) using plastid sequences derived from next-generation sequencing. *American Journal of Botany*, v. 102, n. 6, p. 888-899.

Conte, R., Reis, M.S., Mantovani, A., Vencovsky, R. (2006) Effects of management on the genetic structure of *Euterpe edulis* Mart. populations based on microsatellites. *Scientia Forestalis*, v. 72, p. 81-88.

Conte, R., Reis, M.S., Mantovani, A., Vencovsky, R. (2008) Genetic structure and mating system of *Euterpe edulis* Mart. populations: A comparative analysis using microsatellite and allozyme markers. *Journal of Heredity*, v. 99, n. 5, p. 476-482.

Couvreur, T.L., Kissling, W.D., Condamine, F.L., Svenning, J.C., Rowe, N.P., Baker, W.J. (2015) Global diversification of a tropical plant growth form: environmental correlates and historical contingencies in climbing palms. *Frontiers in Genetics*, v. 5, p. 452.

Covello, P.S. and Gray, M.W. (1993) On the evolution of RNA editing. *Trends in Genetics*, v. 9, n. 8, p. 265-268.

Daniell, H., Lin, C.S., Yu, M., and Chang, W.J. (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*, v. 17, n. 1, p. 134.

Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, v. 14, n. 7, p. 1394-1403.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, v. 9, n. 8, p. 772.

Dransfield, J., Uhl, N.W., Asmussen, C.B., Baker, W.J., Harley, M.M., Lewis, C.E. (2008) *Genera Palmarum - the evolution and classification of palms*. Kew Publishing, Royal Botanical Gardens, Kew, UK. p. 732.

Drescher, A., Ruf, S., Calsa, T. Jr, Carrer, H., Bock, R. (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *The Plant Journal*, v. 22, n. 2, p. 97-104.

- Drummond, A. J., and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, v. 7, n. 1, p. 214.
- Esquinas-Alcazar, J. (2004) International treaty on plant genetic resources for food and agriculture. *Plant Genetic Resources Newsletter*.
- Faurby, S., Eiserhardt, W.L., Baker, W.J., Svenning J.C. (2016) An all-evidence species-level supertree for the palms (Arecaceae). *Molecular Phylogenetics and Evolution*, v. 100, p. 57-69.
- Fleischmann, T.T., Scharff, L.B., Alkatib, S., Hasdorf, S., Schottler, M.A., and Bock, R. (2011) Nonessential Plastid-Encoded Ribosomal Proteins in Tobacco: A Developmental Role for Plastid Translation and Implications for Reductive Genome Evolution. *The Plant Cell*, v. 23, n. 9, p. 3137-3155.
- Glassman, S.F. (1972) A revision of B.E. Dahlgren's index of american palms. *Phanerogamarum monographiae*. J. Cramer, Germany, p. 294.
- Guisinger, M.M., Chumley, T.W., Kuehl, J.V., Boore, J.L., Jansen, R.K. (2010) Implications of the Plastid Genome Sequence of *Typha* (Typhaceae, Poales) for Understanding Genome Evolution in Poaceae. *Journal of Molecular Evolution*, v. 70, n. 2, p. 149-166.
- Guzowska-Nowowiejska M, Fiedorowicz E, Pla der W. (2009) Cucumber, melon, pumpkin, and squash: Are rules of editing in flowering plants chloroplast genes so well known indeed?. *Gene*, v. 434, n. 1, p. 1-8.
- Hager, M., Biehler, K., Illerhaus, J., Ruf, S., Bock, R. (1999) Targeted inactivation of the smallest plastid genome-encoded open reading frame reveals a novel and essential subunit of the cytochrome b(6)f complex. *The EMBO Journal*, v. 18, n. 21, p. 5834-5842.
- Hansen, D.R., Dastidar, S.G., Cai, Z., Penaflor, C., Kuehl, J.V. (2007) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Molecular Phylogenetics and Evolution*, v. 45, n. 2, p. 547-563.
- Henderson, A., and Galeano, G. (1996) *Euterpe*, *Prestoea*, and *Neonicholsonia* (Palmae: Euterpeinae). *The New York Botanical Garden*, p. 1-90.
- Hirao, T., Watanabe, A., Kurita, M., Kondo, T., Takata, K. (2008) Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biology*, v. 8, n. 1, p. 70.
- Hu, S., Sablok, G., Wang, B., Qu, D., Barbaro, E., Viola, R., Li, M., Varotto, C. (2015) Plastome organization and evolution of chloroplast genes in Cardamine species adapted to contrasting habitats. *BMC Genomics*, v. 16, n. 1, p. 306.
- Huang, Y.Y., Matzke, A.J.M., Matzke, M. (2013) Complete Sequence and Comparative Analysis of the Chloroplast Genome of Coconut Palm (*Cocos nucifera*). *PLoS One*, v. 8, n. 8, p. e74736.
- Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies

genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences*, v. 104, n. 49, p. 19369-19374.

Jardim, M.A.G. Aspectos da biologia reprodutiva de uma população natural de açazeiro (*Euterpe oleracea* Mart.) no estuário amazônico. Piracicaba, ESALQ-USP, p. 90. Dissertação de mestrado – Escola Superior de Agricultura Luiz de Queiroz – Piracicaba – SP, 1991.

Jiang, D., Zhao, Z., Zhang, T., Zhong, W., Liu, C., Yuan, Q., & Huang, L. (2017) The Chloroplast Genome Sequence of *Scutellaria baicalensis* Provides Insight into Intraspecific and Interspecific Chloroplast Genome Diversity in *Scutellaria*. *Genes*, v. 8, n. 9, p. 227.

Katoh, K., Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, v. 30, n. 4, p. 772-780.

Kikuchi, S., Bédard, J., Hirano, M., Hirabayashi, Y., Oishi, M., Imai, M., Takase, M., Ide, T., Nakai, M. (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science*, v. 339, n. 6119, p. 571-574.

Kim, K.J., Lee, H.L. (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research*, v. 11, n. 4, p. 247-261.

Kissling, W.D., Eiserhardt, W.L., Baker, W.J., Borchsenius, F., Couvreur, T.L.P., Balslev, H., Svenning, J.C. (2012) Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide. *Proceedings of the National Academy of Sciences*, v. 109, n. 19, p. 7379-7384.

Kovács-Bogdán, E., Soll, J., & Bölter, B. (2010) Protein import into chloroplasts: the Tic complex and its regulation. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, v. 1803, n. 6, p. 740-747.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*, v. 29, n. 22, p. 4633-4642.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biology*, v. 5, n. 2, p. R12.

Larsson, A.A. (2014) View: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, v. 30, n. 22, p. 3276-3278.

Lee, J., Cho, C.H., Park, S.I. et al. (2016) Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biology*, v. 14, n. 1, p. 75.

Librado, P., and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinforma. Bioinformatics*, v. 25, n. 11, p. 1451-1452.

Lin, C.P., Huang, J.P., Wu, C.S., Hsu, C.Y., Chaw, S.M. (2010) Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biology and Evolution*, v. 2, p. 504-517.

- Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013) OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research*, v. 41, n. W1, p. W575-W581.
- Lorenzi, H., Kahn, F., Noblick, L.R., Ferreira E. (2010) *Flora Brasileira: Arecaceae* . (Palmeiras). Nova Odessa: Instituto Plantarum, p. 367.
- Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, v. 25, n. 5, p. 955.
- Mardanov, A., Ravin, N., Kuznetsov, B., Samigullin, T., Antonov, A. et al. (2008) Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other Angiosperms. *Journal of Molecular Evolution*, v. 66, n. 6, p. 555-564.
- Marechal, A., Parent, J.S., Veronneau-Lafortune, F., Joyeux, A., Lang, B.F., et al. (2009) Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, v. 106, n. 34, p. 14693-14698.
- Marshall, H.D., Newton, C., Ritland, K. (2001) Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. *Molecular Biology and Evolution*, v. 18, n. 11, p. 2136-2138.
- Meerow, A.W., Noblick, L., Leiva, D.E.S. et al. (2014) Phylogeny and historical biogeography of the coccosoid palms (Arecaceae, Arecoideae, Cocoseae) inferred from sequences of six WRKY gene family loci. *Cladistics*, v. 31, n. 5, p. 509-534.
- Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E. (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences*, v. 107, n. 10, p. 4623-4628.
- Mower, J.P. (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Research*, v. 37, n. 2, p. W253-W259.
- Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, v. 32, n. 1, p. 268-274.
- Oliveira, L.C., Oliveira, M.S.P., Davide, L.C., Torres, G.A. (2016) Karyotype and genome size in *Euterpe Mart.* (Arecaceae) species. *Comparative Cytogenetics*, v. 10, n. 1, p. 17.
- Oliveira, M. S. P. Avaliação do modo de reprodução e de caracteres quantitativos em 20 acessos de açazeiro (*Euterpe oleracea Mart.* ARECACEAE) em Belém-PA. Recife-PE, Brasil, 146p. Tese de Mestrado em Botânica da Universidade Federal Rural de Pernambuco, 1995.
- Palmer, J.D., Osorio, B., Aldrich, J., Thompson, W.F. (1987) Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. *Current Genetics*, v. 11, n. 4, p. 275-286.
- Perry, A.S. and Wolfe, K.H. (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution*, v. 55, n. 5, p. 501-508.

- Powell, W., Morgante, M., Andre, C., McNicol, J.W., Machray, G.C., Doyle, J.J., Tingey, S.V., Rafalski, J.A. (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Current Biology*, v. 5, n. 9, p. 1023-1029.
- Provan, J., Powell, W., and Hollingsworth, P.M. (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution*, v. 16, n. 3, p. 142-147.
- Pyke, K. (2007) Plastid biogenesis and differentiation. In: *Cell and Molecular Biology of Plastids*. Bock, R. (Ed). Topics in Current Genetics, p. 524.
- Rambaut, A. (2009) FigTree – Molecular Evolution, phylogenetics and Epidemiology (2009). Available: <<http://tree.bio.ed.ac.uk/software/figtree/>>. Acessado: 29 de Junho de 2017.
- Raubeson, L.A., Jansen, R.K. (2005) Chloroplast genomes of plants. In: Henry RJ, editor. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. Cambridge: CAB International, p. 45–68.
- Raubeson, L.A., Peery, R., Chumley, T.W., Dziubek, C., Fourcade, H.M., Boore, J.L., and Jansen, R.K. (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics*, v. 8, n. 1, p. 174.
- Rogalski, M., Ruf, S., Bock, R. (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Research*, v. 34, n. 16, p. 4537-4545.
- Rogalski, M., Schoettler, M.A., Thiele, W., Schulze, W.X., Bock, R. (2008) Rpl33, a nonessential plastid encoded ribosomal protein in tobacco, is required under cold stress conditions. *The Plant Cell*, v. 20, n. 8, p. 2221-2237.
- Rogalski, M., Vieira, L.N., Fraga, H.P., Guerra, M.P. (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Frontiers in Plant Science*, v. 6, p. 586.
- Souza, P.C.A. Aspectos ecológicos e genéticos de uma população natural de *Euterpe oleracea* Mart. no estuário Amazônico, p. 60. Dissertação (Mestrado em Recursos Florestais) Escola Superior de Agricultura Luiz de Queiroz – Piracicaba – SP, 2002.
- Steege, H. et al. (2013) Hyperdominance in the Amazonian Tree Flora. *Science*, v. 342, n. 6156, p. 1243092.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., and Brennicke, A. (2013) RNA Editing in Plants and Its Evolution. *Annual Review of Genetics*, v. 47, p. 335-352.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, v. 30, n. 12, p. 2725-2729.
- Thiel, T., Michalek, W., Varshney, R., Graner, A. (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, v. 106, n. 3, p. 411-422.

- Tillich, M., Lehwark, P., Morton B.R., Maier, U.G. (2006) The evolution of chloroplast RNA editing. *Molecular Biology and Evolution*, v. 23, n. 10, p. 1912-1921.
- Uhl, N.W and Dransfield, J. (1986) *Genera Palmarum*. A new classification of palms and its implications. In: BALICK, M.J., ed. *The palm- tree of life: biology, utilization and conservation*. *Advances in Economical Botany*, New York, v.6, p.1-19.
- Uthaipaisanwong, P., Chanprasert, J., Shearman, J.R., Sangsrakru, D., Yoocha, T. et al. (2012) Characterization of the chloroplast genome sequence of oil palm (*Elaeis guineensis* Jacq.). *Gene*, v. 500, n. 2, p. 172-180.
- Vieira, L.N., Faoro, H., Fraga, H.P. de F., Rogalski, M., de Souza, E.M., de Oliveira Pedrosa, F., Nodari, R.O., and Guerra, M.P. (2014b) An Improved Protocol for Intact Chloroplasts and cpDNA Isolation in Conifers. *PLoS One*, v. 9, n. 1, p. e84792.
- Vieira, L.N., Faoro, H., Rogalski, M., Fraga, H.P. de F., Cardoso, R.L.A., de Souza, E.M., de Oliveira Pedrosa, F., Nodari, R.O., and Guerra, M.P. (2014a) The Complete Chloroplast Genome Sequence of *Podocarpus lambertii*: Genome Structure, Evolutionary Aspects, Gene Content and SSR Detection. *PLoS One*, v. 9, n. 3, p. e90618.
- Vries, J., Sousa, F.L., Bolter, B., Soll, J., and Gould, S.B. (2015) YCF1: A Green TIC?. *The Plant Cell*, v. 27, n. 7, p. 1827-1833.
- Wicke, S, Schneeweiss, G.M., Muller, K.F., Quandt., D. (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*, v. 76, n. 3-5, p. 273-297.
- Wu, C.S., Wang, Y.N., Liu, S.M. et al. (2007) Chloroplast Genome (cpDNA) of *Cycas taitungensis* and 56 cp Protein-Coding Genes of *Gnetum parvifolium*: Insights into cp DNA Evolution and Phylogeny of Extant Seed Plants. *Molecular Biology and Evolution*, v. 24, n. 6, p. 1366-1379.
- Wu, F.H., Chan, M.T., Liao, D.C., Hsu, C.T., Lee, Y.W. et al. (2010) Complete chloroplast genome of *Oncidium Gower Ramsey* and evaluation of molecular markers for identification and breeding in *Oncidiinae*. *BMC Plant Biology*, v. 10, n. 1, p. 68.
- Wyman, S.K., Jansen, R.K., and Boore., J.L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, v. 20, n. 17, p. 3252-3255.
- Yamaguchi, K.K.L., Pereira, L.F.R., Lamarão, C.V., Lima E.S., Veiga-Junior, V.F. (2015) Amazon acai: Chemistry and biological activities: A review. *Food Chemistry*, v. 179, p. 137-151.
- Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K. et al. (2010) The Complete Chloroplast Genome Sequence of Date Palm (*Phoenix dactylifera* L.). *PloS One*, v. 5, n. 9, p. e12762.
- Zanduetta-Criado, A., Bock, R. (2004) Surprising features of plastid *ndhD* transcripts: addition of non-encoded nucleotides and polysome association of mRNAs with an unedited start codon. *Nucleic Acids Research*, v. 32, n. 2, p. 542-550.
- Zhang, Y.J, Ma, P.F., Li, D.Z. (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PloS One*, v. 6, n. 5, p. e20596.

Zhu, A., Guo, W., Gupta, S., Fan, W., and Mower, J.P. (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist*, v. 209, n. 4, p. 1747-1756.

## APÊNDICE

**Tabela S11:** Distribuição dos locos das sequências com repetições simples (SSR) no genoma plastidial de *Euterpe edulis*

<b>Tipo de SSR</b>	<b>SSR</b>	<b>Início</b>	<b>Fim</b>	<b>Localização</b>
di	(AT)4	1326	1333	psbA/trnk-UUU(IGS)
mono	(A)12	2637	2648	matk (CDS)
mono	(T)9	2886	2894	matK(CDS)
di	(AT)4	3382	3389	matK/trnk-UUU(IGS)
mono	(A)13	3609	3621	matK/trnk-UUU(IGS)
mono	(T)10	3813	3822	matK/trnk-UUU(IGS)
mono	(A)10	3836	3845	matK/trnk-UUU(IGS)
tri	(AAT)4	3928	3939	matK/trnk-UUU(IGS)
mono	(C)9	4786	4794	trnK-UUU/rps16(IGS)
mono	(T)10	5604	5613	rps16(INTRON)
di	(GT)4	6018	6025	rps16/trnQ-UUG(IGS)
tetra	(TCTA)4	6052	6067	rps16/trnQ-UUG(IGS)
mono	(T)10	6856	6865	rps16/trnQ-UUG(IGS)
mono	(A)8	6871	6878	rps16/trnQ-UUG(IGS)
tetra	(AATA)3	7196	7207	trnQ-UUG/psbK(IGS)
mono	(A)13	7369	7381	trnQ-UUG/psbK(IGS)
mono	(T)9	7402	7410	trnQ-UUG/psbK(IGS)
mono	(A)10	7418	7427	trnQ-UUG/psbK(IGS)
mono	(T)8	7610	7617	trnQ-UUG/psbK(IGS)
mono	(T)8	7964	7971	psbK/psbI(IGS)
mono	(A)10	8041	8050	psbK/psbI(IGS)
mono	(A)8	8178	8185	psbI/trnS-GCU(IGS)
mono	(T)9	8217	8225	psbI/trnS-GCU(IGS)
di	(GA)4	8321	8328	trnS-GCU(CDS)
mono	(T)10	8510	8519	trnS-GCU/trnG-GCC(IGS)
di	(AT)4	8520	8527	trnS-GCU/trnG-GCC(IGS)
di	(AT)6	8564	8575	trnS-GCU/trnG-GCC(IGS)
mono	(A)8	8591	8598	trnS-GCU/trnG-GCC(IGS)
di	(AT)6	8599	8610	trnS-GCU/trnG-GCC(IGS)
penta	(ATAAG)3	8611	8625	trnS-GCU/trnG-GCC(IGS)
mono	(T)12	8929	8940	trnS-GCU/trnG-GCC(IGS)
di	(AT)10	9000	9019	trnS-GCU/trnG-GCC(IGS)
di	(AT)4	9031	9038	trnS-GCU/trnG-GCC(IGS)
mono	(A)13	9079	9091	trnS-GCU/trnG-GCC(IGS)
penta	(TATAA)3	9128	9142	trnS-GCU/trnG-GCC(IGS)

mono	(A)11	9246	9256	trnS-GCU/trnG-GCC(IGS)
mono	(T)9	9876	9884	trnG-GCC(INTRON)
mono	(T)8	12184	12191	atpA/atpF(IGS)
mono	(T)10	12797	12806	atpF(INTRON)
mono	(T)8	12990	12997	atpF(INTRON)
mono	(T)9	13032	13040	atpF(INTRON)
mono	(A)8	13469	13476	atpF(CDS)
mono	(T)9	13708	13716	atpF/atpH(IGS)
mono	(A)8	13870	13877	atpF/atpH(IGS)
penta	(AGTAT)5	14095	14119	atpF/atpI(IGS)
mono	(A)11	14410	14420	atpH/atpI(IGS)
mono	(T)8	14463	14470	atpH/atpI(IGS)
mono	(T)9	14475	14483	atpH/atpI(IGS)
di	(AT)7	14847	14860	atpH/atpI(IGS)
di	(AT)7	14874	14887	atpH/atpI(IGS)
mono	(T)13	15065	15077	atpH/atpI(IGS)
mono	(T)11	16961	16971	rps2(CDS)
mono	(A)8	18356	18363	rpoC2(CDS)
mono	(T)10	19069	19078	rpoC2(CDS)
mono	(T)11	19175	19185	rpoC2(CDS)
mono	(A)8	19318	19325	rpoC2(CDS)
di	(AT)4	20449	20456	rpoC2(CDS)
di	(AT)5	20539	20548	rpoC2(CDS)
mono	(A)8	22991	22998	rpoC1(CDS)
mono	(A)10	23182	23191	rpoC1(INTRON)
mono	(T)11	23301	23311	rpoC1(INTRON)
di	(TA)4	23545	23552	rpoC1(INTRON)
di	(TC)4	23572	23579	rpoC1(INTRON)
mono	(T)10	23616	23625	rpoC1(INTRON)
mono	(T)8	26868	26875	rpoB(CDS)
mono	(A)8	27760	27767	rpoB/trnC-GCA(IGS)
mono	(A)10	28631	28640	rpoB/trnC-GCA (IGS)
mono	(A)9	28892	28900	trnC-GCA/petN (IGS)
mono	(T)10	28908	28917	trnC-GCA/petN (IGS)
mono	(A)11	29424	29434	trnC-GCA/petN (IGS)
mono	(A)8	29450	29457	trnC-GCA/petN (IGS)
di	(GT)4	29681	29688	trnC-GCA/petN (IGS)
mono	(A)14	30060	30073	petN/psbM (IGS)
hexa	(TTTTAT)3	30094	30111	petN/psbM (IGS)
mono	(T)8	30112	30119	petN/psbM (IGS)
di	(TA)7	30189	30202	petN/psbM (IGS)
di	(TA)4	30238	30245	petN/psbM (IGS)
tetra	(AAAT)3	30246	30257	petN/psbM (IGS)
di	(AT)3	30258	30263	petN/psbM (IGS)
di	(TA)4	30264	30271	petN/psbM (IGS)
mono	(A)10	30391	30400	petN/psbM (IGS)
mono	(T)8	30861	30868	psbM/trnD (IGS)
mono	(T)9	32900	32908	trnT-GGU/psbD (IGS)

mono	(T)8	33314	33321	trnT-GGU/psbD (IGS)
mono	(T)10	33602	33611	trnT-GGU/psbD (IGS)
di	(TA)4	33701	33708	trnT-GGU/psbD (IGS)
mono	(A)10	33769	33778	trnT-GGU/psbD (IGS)
mono	(G)8	35227	35234	psbD/psbC (IGS)
mono	(G)10	35498	35507	psbC CDS)
di	(GA)4	36486	36493	trnS(CDS)
mono	(A)8	37195	37202	psbZ/trnG-GCC (IGS)
mono	(A)9	37582	37590	psbZ/trnG-GCC (IGS)
mono	(A)10	37752	37761	trnfM-CAU/rps14 (IGS)
mono	(T)11	38103	38113	rps14 (CDS)
mono	(A)8	38323	38330	psaB (CDS)
di	(AG)4	42192	42199	psaA (CDS)
penta	(TATTT)3	42944	42958	psaA/ycf3 (IGS)
mono	(A)10	43979	43988	ycf3 (INTRON)
mono	(T)9	44292	44300	ycf3 (INTRON)
mono	(A)8	46319	46326	trnS-GGA/rps4 (IGS)
mono	(G)9	46586	46594	rps4 (CDS)
mono	(T)10	47066	47075	rps4/trnT (IGS)
di	(TA)4	47216	47223	rps4/trnT (IGS)
di	(TA)6	47602	47613	trnT/ trnL (IGS)
di	(TA)4	47904	47911	trnT/ trnL (IGS)
di	(TA)7	47926	47939	trnT/ trnL (IGS)
tetra	(ATTT)3	48002	48013	trnT/ trnL (IGS)
mono	(A)10	48158	48167	trnT/ trnL (IGS)
di	(AT)6	48234	48245	trnT/ trnL (IGS)
mono	(A)8	48468	48475	trnT/ trnL (IGS)
di	(AT)7	48825	48838	trnL (INTRON)
di	(AG)4	48973	48980	trnL (INTRON)
di	(AT)4	49554	49561	trnF-GAA/ndhJ (IGS)
di	(AT)7	49764	49777	trnF-GAA/ndhJ (IGS)
mono	(A)10	49779	49788	trnF-GAA/ndhJ (IGS)
di	(AT)5	50563	50572	ndhJ (CDS)
mono	(T)8	52123	52130	ndhC/trnV-UAC (IGS)
mono	(A)8	52259	52266	ndhC/trnV-UAC (IGS)
mono	(A)9	52306	52314	ndhC/trnV-UAC (IGS)
di	(AT)6	52457	52468	ndhC/trnV-UAC (IGS)
mono	(A)10	52555	52564	ndhC/trnV-UAC (IGS)
mono	(T)12	52641	52652	ndhC/trnV-UAC (IGS)
mono	(A)8	52849	52856	ndhC/trnV-UAC (IGS)
mono	(A)11	52943	52953	ndhC/trnV-UAC (IGS)
di	(CA)4	53386	53393	ndhC/trnV-UAC (IGS)
mono	(T)11	54753	54763	trnM-CAU/atpE (IGS)
mono	(T)8	56798	56805	atpB/rbcL (IGS)
mono	(A)8	56871	56878	atpB/rbcL (IGS)
mono	(T)8	57129	57136	atpB/rbcL (IGS)
di	(AT)4	57334	57341	atpB/rbcL (IGS)
mono	(A)9	59077	59085	rbcL/accD (IGS)

mono	(C)12	59599	59610	rbcL/accD (IGS)
mono	(A)8	59611	59618	rbcL/accD (IGS)
mono	(T)8	60137	60144	accD (CDS)
mono	(A)10	60440	60449	accD (CDS)
di	(TG)4	60869	60876	accD (CDS)
mono	(A)8	61325	61332	accD (CDS)
mono	(A)10	61532	61541	accD (CDS)
mono	(A)8	61576	61583	accD (CDS)
di	(TA)4	61954	61961	accD/psaI (IGS)
mono	(T)8	62732	62739	ycf4 (CDS)
mono	(A)11	63324	63334	cemA (CDS)
di	(TC)5	63391	63400	cemA (CDS)
tetra	(AATG)3	64001	64012	cemA (CDS)
mono	(C)8	64591	64598	petA (CDS)
mono	(A)8	65643	65650	petA/psbJ (IGS)
mono	(T)8	67348	67355	psbE/petL (IGS)
penta	(TTTCA)3	67365	67379	psbE/petL (IGS)
mono	(A)8	67777	67784	psbE/petL (IGS)
mono	(T)9	68588	68596	petG/trnW-CCA (IGS)
mono	(T)8	68644	68651	petG/trnW-CCA (IGS)
mono	(A)9	69151	69159	trnP-UGG/psaJ (IGS)
mono	(A)8	69189	69196	trnP-UGG/psaJ (IGS)
mono	(A)10	69320	69329	trnP-UGG/psaJ (IGS)
mono	(T)8	69518	69525	psaJ (CDS)
mono	(T)12	69608	69619	psaJ/rpl33 (IGS)
di	(AT)6	70455	70466	rpl33/rps18 (IGS)
mono	(T)9	71020	71028	rps18/rpl20 (IGS)
mono	(T)8	71071	71078	rps18/rpl20 (IGS)
mono	(A)8	71726	71733	rpl20/rps12 (IGS)
mono	(T)12	71762	71773	rpl20/rps12 (IGS)
di	(TA)4	72352	72359	rps12/clpP (IGS)
mono	(T)9	72422	72430	rps12/clpP (IGS)
mono	(A)8	72797	72804	clpP (INTRON)
mono	(A)8	72979	72986	clpP (INTRON)
mono	(T)12	73085	73096	clpP (INTRON)
mono	(T)10	73351	73360	clpP (INTRON)
tetra	(ATAA)3	73549	73560	clpP (INTRON)
mono	(A)9	73737	73745	clpP (INTRON)
mono	(T)8	73788	73795	clpP (INTRON)
mono	(A)11	73885	73895	clpP (INTRON)
mono	(T)8	74283	74290	clpP (INTRON)
mono	(A)9	74572	74580	clpP/psbB (IGS)
mono	(T)8	75753	75760	psbB (CDS)
mono	(T)8	76578	76585	psbB/psbT(IGS)
mono	(A)9	77679	77687	petB (INTRON)
mono	(T)9	80515	80523	rpoA/rps11 (IGS)
mono	(A)10	80531	80540	rpoA/rps11 (IGS)
mono	(T)11	82416	82426	rpl36/infA (IGS)

mono	(T)13	83291	83303	rps8/rpl14 (IGS)
tetra	(AATA)4	84677	84692	rpl16 (INTRON)
mono	(T)11	84977	84987	rpl16 (INTRON)
mono	(T)8	84989	84996	rpl16 (INTRON)
tetra	(TTTA)3	85001	85012	rpl16 (INTRON)
mono	(T)12	85351	85362	rpl16/rps3 (IGS)
mono	(T)9	87077	87085	rps19/trnH-GUG (IGS)
mono	(T)11	87113	87123	rps19/trnH-GUG (IGS)
mono	(A)8	87331	87338	trnH-GUG/rpl2 (IGS)
di	(GA)4	89480	89487	ycf2 (CDS)
di	(GA)4	89492	89499	ycf2 (CDS)
di	(GA)4	90494	90501	ycf2 (CDS)
mono	(A)8	91203	91210	ycf2 (CDS)
mono	(A)8	91396	91403	ycf2 (CDS)
mono	(A)9	92658	92666	ycf2 (CDS)
di	(TA)4	96068	96075	ycf2 (CDS)
di	(TA)4	97618	97625	trnL-CAA/ndhB (IGS)
di	(AG)4	98363	98370	ndhB (INTRON)
mono	(T)8	99349	99356	ndhB (INTRON)
mono	(T)8	102303	102310	rps7/trnV-GAC (IGS)
mono	(A)8	102469	102476	rps7/trnV-GAC (IGS)
mono	(T)8	106330	106337	trnI-CAU (INTRON)
di	(CT)4	109652	109659	rrn23 (CDS)
mono	(A)8	112580	112587	trnN-GUU/ycf1 (IGS)
mono	(T)8	113682	113689	ycf1 (CDS)
mono	(A)8	113846	113853	ycf1 (CDS)
mono	(A)8	113874	113881	ndhF (CDS)
mono	(T)9	116059	116067	ndhF/rpl32 (IGS)
mono	(A)11	116137	116147	ndhF/rpl32 (IGS)
di	(AT)7	116161	116174	ndhF/rpl32 (IGS)
di	(AT)4	116176	116183	ndhF/rpl32 (IGS)
di	(AT)7	116185	116198	ndhF/rpl32 (IGS)
mono	(A)12	116248	116259	ndhF/rpl32 (IGS)
mono	(A)8	116294	116301	ndhF/rpl32 (IGS)
mono	(T)8	116971	116978	rpl32/trnL-UAG (IGS)
mono	(A)9	116983	116991	rpl32/trnL-UAG (IGS)
mono	(A)9	117047	117055	rpl32/trnL-UAG (IGS)
mono	(T)8	117996	118003	ccsA (CDS)
tetra	(AATA)3	118873	118884	ndhD (CDS)
di	(AT)6	120862	120873	ndhE/ndhG (IGS)
tetra	(TTTA)3	121402	121413	ndhE/ndhG (IGS)
mono	(T)8	122049	122056	ndhG/ndhI (IGS)
mono	(T)9	122096	122104	ndhG/ndhI (IGS)
mono	(T)9	122571	122579	ndhI (CDS)
mono	(A)8	123667	123674	ndhA (INTRON)
di	(AT)8	123767	123782	ndhA (INTRON)
mono	(T)8	123861	123868	ndhA (INTRON)
mono	(A)11	123963	123973	ndhA (INTRON)

mono	(A)13	124442	124454	ndhA (INTRON)
di	(TC)5	125855	125864	ndhA/ndhH (IGS)
mono	(A)10	126353	126362	ndhH/rps15 (IGS)
di	(AT)4	127062	127069	yef1 (CDS)
mono	(T)8	127453	127460	yef1 (CDS)
mono	(T)11	127875	127885	yef1 (CDS)
mono	(T)11	127998	128008	yef1 (CDS)
mono	(A)8	128180	128187	yef1 (CDS)
mono	(T)9	128243	128251	yef1 (CDS)
mono	(T)8	129196	129203	yef1 (CDS)
mono	(T)9	129475	129483	yef1 (CDS)
mono	(T)11	129680	129690	yef1 (CDS)
mono	(T)9	129739	129747	yef1 (CDS)
mono	(T)10	129859	129868	yef1 (CDS)
mono	(T)8	130242	130249	yef1 (CDS)
di	(TC)4	130299	130306	yef1 (CDS)
mono	(A)10	130357	130366	yef1 (CDS)
mono	(T)8	131204	131211	yef1 (CDS)
mono	(T)8	131261	131268	yef1 (CDS)
mono	(A)8	131425	131432	yef1 (CDS)
mono	(T)8	132527	132534	yef1/trnN-GUU (IGS)
di	(AG)4	135455	135462	rrn23 (CDS)
mono	(A)8	138777	138784	trnI (INTRON)
mono	(T)8	142638	142645	trnV-GAC/rps7 (IGS)
mono	(A)8	142804	142811	trnV-GAC/rps7 (IGS)
mono	(A)8	145758	145765	ndhB (INTRON)
di	(CT)4	146744	146751	ndhB (INTRON)
di	(AT)4	147488	147495	ndhB/trnL-CAA (IGS)
di	(TA)4	149039	149046	yef2 (CDS)
mono	(T)9	152448	152456	yef2 (CDS)
mono	(T)8	153711	153718	yef2 (CDS)
mono	(T)8	153904	153911	yef2 (CDS)
di	(TC)4	154613	154620	yef2 (CDS)
di	(TC)4	155615	155622	yef2 (CDS)
di	(TC)4	155627	155634	yef2 (CDS)
mono	(T)8	157776	157783	rpl2/trnH-GUG (IGS)
mono	(A)11	157991	158001	trnH-GUG/rps19 (IGS)
mono	(A)9	158029	158037	trnH-GUG/rps19 (IGS)

---

**Tabela S12.** Distribuição dos locos das sequências com repetições simples (SSR) no genoma plastidial de *Euterpe oleracea*

<b>Tipo SSR</b>	<b>SSR</b>	<b>Início</b>	<b>Fim</b>	<b>Location</b>
di	(AT)4	1340	1347	trnk-UUU(CDS)
mono	(A)12	2651	2662	matK (CDS)
mono	(T)9	2900	2908	matK (CDS)
di	(AT)4	3396	3403	matK/trnk-UUU(IGS)
mono	(A)12	3624	3635	matK/trnk-UUU(IGS)
mono	(T)10	3827	3836	matK/trnk-UUU(IGS)
mono	(A)10	3850	3859	matK/trnk-UUU(IGS)
tri	(AAT)4	3935	3946	matK/trnk-UUU(IGS)
mono	(C)11	4793	4803	trnk-UUU/rps16(IGS)
mono	(T)11	4804	4814	trnk-UUU/rps16(IGS)
mono	(T)13	5618	5630	rps16(INTRON)
di	(GT)4	6035	6042	rps16/trnQ-UUG(IGS)
tetra	(TCTA)3	6069	6080	rps16/trnQ-UUG(IGS)
mono	(T)9	6869	6877	rps16/trnQ-UUG(IGS)
mono	(A)8	6884	6891	rps16/trnQ-UUG(IGS)
tetra	(AATA)3	7209	7220	trnQ-UUG/psbK(IGS)
mono	(A)12	7382	7393	trnQ-UUG/psbK(IGS)
mono	(T)8	7414	7421	trnQ-UUG/psbK(IGS)
mono	(A)10	7429	7438	trnQ-UUG/psbK(IGS)
mono	(T)8	7621	7628	psbK(CDS)
mono	(A)9	8051	8059	psbK/psbI(IGS)
mono	(A)8	8192	8199	psbI/trnS-GCU(IGS)
mono	(T)8	8231	8238	psbI/trnS-GCU(IGS)
di	(GA)4	8334	8341	trnS-GCU(CDS)
mono	(T)10	8523	8532	trnS-GCU/trnG-GCC(IGS)
di	(AT)4	8533	8540	trnS-GCU/trnG-GCC(IGS)
di	(AT)5	8578	8587	trnS-GCU/trnG-GCC(IGS)
mono	(A)12	8651	8662	trnS-GCU/trnG-GCC(IGS)
di	(AT)4	8814	8821	trnS-GCU/trnG-GCC(IGS)
di	(AT)4	8831	8838	trnS-GCU/trnG-GCC(IGS)
di	(TA)4	8840	8847	trnS-GCU/trnG-GCC(IGS)
di	(AT)8	8857	8872	trnS-GCU/trnG-GCC(IGS)
mono	(T)12	9215	9226	trnS-GCU/trnG-GCC(IGS)
di	(AT)10	9287	9306	trnS-GCU/trnG-GCC(IGS)
di	(AT)4	9319	9326	trnS-GCU/trnG-GCC(IGS)
mono	(A)10	9367	9376	trnS-GCU/trnG-GCC(IGS)
penta	(TATAA)3	9413	9427	trnS-GCU/trnG-GCC(IGS)
mono	(A)9	9531	9539	trnS-GCU/trnG-GCC(IGS)
mono	(T)9	10158	10166	trnG-GCC(INTRON)
mono	(T)8	12466	12473	atpA/atpF(IGS)

mono	(T)12	13079	13090	atpF(INTRON)
mono	(T)8	13279	13286	atpF(INTRON)
mono	(T)9	13321	13329	atpF(INTRON)
mono	(A)8	13758	13765	atpF(INTRON)
mono	(T)10	13997	14006	atpF/atpH(IGS)
penta	(AGTAT)4	14385	14404	atpF/atpH(IGS)
mono	(A)11	14695	14705	atpH/atpI(IGS)
mono	(T)8	14748	14755	atpH/atpI(IGS)
mono	(T)12	14760	14771	atpH/atpI(IGS)
di	(AT)5	15145	15154	atpH/atpI(IGS)
di	(AT)7	15168	15181	atpH/atpI(IGS)
mono	(T)8	15359	15366	atpH/atpI(IGS)
mono	(T)10	17254	17263	rps2/rpoC2(IGS)
mono	(A)8	18648	18655	rpoC2(CDS)
mono	(T)10	19361	19370	rpoC2(CDS)
mono	(T)11	19467	19477	rpoC2(CDS)
mono	(A)8	19610	19617	rpoC2(CDS)
di	(AT)4	20741	20748	rpoC2(CDS)
di	(AT)5	20831	20840	rpoC2(CDS)
mono	(A)8	23283	23290	rpoC1(INTRON)
mono	(A)9	23474	23482	rpoC1(INTRON)
mono	(T)13	23592	23604	rpoC1(INTRON)
di	(TA)4	23838	23845	rpoC1(INTRON)
di	(TC)4	23865	23872	rpoC1(INTRON)
mono	(T)11	23909	23919	rpoC1(INTRON)
mono	(T)8	27162	27169	rpoB(CDS)
mono	(A)9	28054	28062	rpoB/trnC-GCA(IGS)
mono	(A)9	28926	28934	rpoB/trnC-GCA(IGS)
mono	(A)8	29186	29193	trnC-GCA/petN (IGS)
mono	(T)10	29201	29210	trnC-GCA/petN (IGS)
mono	(A)11	29717	29727	trnC-GCA/petN (IGS)
mono	(A)8	29743	29750	trnC-GCA/petN (IGS)
di	(GT)4	29974	29981	trnC-GCA/petN (IGS)
mono	(A)14	30353	30366	petN/psbM (IGS)
mono	(T)11	30392	30402	petN/psbM (IGS)
mono	(T)8	30405	30412	petN/psbM (IGS)
di	(TA)7	30482	30495	petN/psbM (IGS)
di	(TA)7	30506	30519	petN/psbM (IGS)
di	(TA)4	30530	30537	petN/psbM (IGS)
tetra	(AAAT)3	30538	30549	petN/psbM (IGS)
di	(AT)4	30549	30557	petN/psbM (IGS)
di	(TA)4	30556	30563	petN/psbM (IGS)
mono	(A)10	30683	30692	petN/psbM (IGS)
di	(TA)4	30907	30914	petN/psbM (IGS)
mono	(T)8	31166	31173	psbM/trnD (IGS)
mono	(T)9	33200	33208	trnT-GGU/psbD (IGS)

mono	(T)8	33614	33621	trnT-GGU/psbD (IGS)
mono	(T)13	33902	33914	trnT-GGU/psbD (IGS)
di	(TA)4	34004	34011	trnT-GGU/psbD (IGS)
mono	(A)9	34072	34080	trnT-GGU/psbD (IGS)
mono	(G)8	35529	35536	psbC (CDS)
mono	(G)10	35800	35809	psbC (CDS)
di	(GA)4	36788	36795	psbC/trnS (IGS)
mono	(A)8	37497	37504	psbZ/trnG-GCC (IGS)
mono	(A)9	37884	37892	TrnG-GCC/ trnfM-CAU(IGS)
mono	(A)9	38054	38062	trnfM-CAU/rps14 (IGS)
mono	(T)11	38404	38414	rps14 (CDS)
mono	(A)8	38624	38631	rps14/psaB (IGS)
di	(AG)4	42493	42500	psaA (CDS)
penta	(TATTT)3	43245	43259	psaA/ycf3 (IGS)
mono	(A)10	44280	44289	ycf3 (INTRON)
mono	(T)11	44593	44603	ycf3 (INTRON)
mono	(A)8	46634	46641	trnS-GGA/rps4 (IGS)
mono	(G)9	46901	46909	trnS-GGA/rps4 (IGS)
mono	(T)14	47381	47394	rps4/trnT-UGU (IGS)
di	(TA)4	47535	47542	rps4/trnT-UGU (IGS)
di	(TA)6	47921	47932	TrnT-UGU/ trnL-UAA (IGS)
di	(TA)4	48223	48230	TrnT-UGU/ trnL-UAA(IGS)
di	(TA)8	48245	48260	TrnT-UGU/ trnL-UAA (IGS)
tetra	(ATTT)3	48323	48334	TrnT-UGU/ trnL-UAA (IGS)
mono	(A)9	48479	48487	TrnT-UGU/ trnL-UAA (IGS)
di	(AT)6	48554	48565	TrnT-UGU/ trnL-UAA (IGS)
mono	(A)8	48788	48795	TrnT-UGU/ trnL-UAA (IGS)
di	(AT)7	49145	49158	TrnL-UAA (INTRON)
di	(AG)4	49293	49300	TrnL-UAA (INTRON)
di	(AT)4	49874	49881	trnF-GAA/ndhJ (IGS)
di	(AT)7	50084	50097	trnF-GAA/ndhJ (IGS)
mono	(A)9	50099	50107	trnF-GAA/ndhJ (IGS)
mono	(T)8	52442	52449	ndhC/trnV-UAC (IGS)
mono	(A)8	52625	52632	ndhC/trnV-UAC (IGS)
di	(AT)6	52775	52786	ndhC/trnV-UAC (IGS)
mono	(A)10	52873	52882	ndhC/trnV-UAC (IGS)
mono	(T)16	52959	52974	ndhC/trnV-UAC (IGS)
penta	(TACAT)3	53023	53037	ndhC/trnV-UAC (IGS)
mono	(A)8	53176	53183	ndhC/trnV-UAC (IGS)
mono	(A)9	53270	53278	ndhC/trnV-UAC (IGS)
di	(CA)4	53706	53713	ndhC/trnV-UAC (IGS)
mono	(T)13	55073	55085	trnM-CAU/atpE (IGS)
mono	(T)9	57129	57137	atpB/rbcL (IGS)
mono	(A)8	57202	57209	atpB/rbcL (IGS)
mono	(T)8	57460	57467	atpB/rbcL (IGS)
di	(AT)4	57665	57672	atpB/rbcL (IGS)

mono	(A)9	59408	59416	rbcL/accD (IGS)
mono	(C)9	59930	59938	rbcL/accD (IGS)
mono	(A)10	59939	59948	rbcL/accD (IGS)
mono	(T)8	60467	60474	accD (CDS)
mono	(A)10	60770	60779	accD (CDS)
di	(TG)4	61199	61206	accD (CDS)
mono	(A)8	61655	61662	accD/psaI (IGS)
mono	(A)10	61862	61871	accD/psaI (IGS)
mono	(A)8	61906	61913	accD/psaI (IGS)
di	(TA)4	62285	62292	accD/psaI (IGS)
mono	(T)8	63059	63066	ycf4 (CDS)
mono	(A)11	63651	63661	cemA (CDS)
di	(TC)5	63718	63727	cemA (CDS)
tetra	(AATG)3	64328	64339	cemA (CDS)
mono	(C)8	64918	64925	petA (CDS)
mono	(A)8	66048	66055	petA/psbJ (IGS)
mono	(T)8	67754	67761	psbE/petL (IGS)
penta	(TTTCA)3	67771	67785	psbE/petL (IGS)
mono	(A)8	68183	68190	psbE/petL (IGS)
mono	(T)9	68992	69000	petG/trnW-CCA (IGS)
mono	(A)8	69560	69567	trnP-UGG/psaJ (IGS)
mono	(A)8	69597	69604	trnP-UGG/psaJ (IGS)
mono	(A)11	69728	69738	trnP-UGG/psaJ (IGS)
mono	(T)8	69927	69934	psaJ (CDS)
mono	(T)10	70017	70026	psaJ/rpl33 (IGS)
di	(AT)6	70862	70873	rpl33/rps18 (IGS)
mono	(T)11	71427	71437	rps18/rpl20 (IGS)
mono	(T)8	71480	71487	rps18/rpl20 (IGS)
mono	(A)8	72135	72142	rpl20/rps12 (IGS)
mono	(T)12	72171	72182	rpl20/rps12 (IGS)
di	(TA)4	72761	72768	rps12/clpP (IGS)
mono	(T)9	72831	72839	rps12/clpP (IGS)
mono	(A)8	73206	73213	clpP (INTRON)
mono	(A)8	73388	73395	clpP (INTRON)
mono	(T)12	73494	73505	clpP (INTRON)
mono	(T)12	73760	73771	clpP (INTRON)
tetra	(ATAA)3	73960	73971	clpP (INTRON)
mono	(A)9	74148	74156	clpP (INTRON)
mono	(T)8	74199	74206	clpP (INTRON)
mono	(A)10	74296	74305	clpP (INTRON)
mono	(T)8	74693	74700	clpP (INTRON)
mono	(A)9	74982	74990	clpP/psbB (IGS)
mono	(T)8	76163	76170	psbB (CDS)
mono	(T)8	76988	76995	psbB (CDS)
mono	(A)9	78089	78097	petB (INTRON)
mono	(T)12	80921	80932	petD/rpoA (IGS)

mono	(T)12	82820	82831	rpl36/infA (IGS)
mono	(T)10	83696	83705	rps8/rpl14 (IGS)
tetra	(AATA)4	85079	85094	rpl16 (INTRON)
mono	(T)9	85379	85387	rpl16 (INTRON)
mono	(T)8	85389	85396	rpl16 (INTRON)
tetra	(TTTA)3	85401	85412	rpl16 (INTRON)
mono	(T)15	85750	85764	rpl16 (INTRON)
mono	(T)9	87479	87487	rps19/trnH-GUG (IGS)
mono	(T)11	87515	87525	rps19/trnH-GUG (IGS)
mono	(A)8	87733	87740	trnH-GUG/rpl2 (IGS)
di	(GA)4	89882	89889	ycf2 (CDS)
di	(GA)4	89894	89901	ycf2 (CDS)
di	(GA)4	90896	90903	ycf2 (CDS)
mono	(A)8	91605	91612	ycf2 (CDS)
mono	(A)8	91798	91805	ycf2 (CDS)
mono	(A)9	93060	93068	ycf2 (CDS)
di	(TA)4	96446	96453	ycf2 (CDS)
di	(TA)4	97996	98003	trnL-CAA/ndhB (IGS)
di	(AG)4	98741	98748	ndhB (INTRON)
mono	(T)8	99727	99734	ndhB (INTRON)
mono	(T)8	102681	102688	rps12/trnV-GAC (IGS)
mono	(A)8	102847	102854	rps12/trnV-GAC (IGS)
mono	(T)8	106708	106715	trnI (INTRON)
di	(CT)4	110030	110037	rrn23 (CDS)
mono	(A)8	112958	112965	trnN-GUU/ycf1 (IGS)
mono	(T)8	114060	114067	ycf1 (CDS)
mono	(A)8	114224	114231	ycf1 (CDS)
mono	(A)8	114252	114259	ycf1/ndhF(IGS)
mono	(T)11	116437	116447	ndhF/rpl32 (IGS)
mono	(A)11	116515	116525	ndhF/rpl32 (IGS)
di	(AT)6	116541	116552	ndhF/rpl32 (IGS)
di	(AT)4	116561	116568	ndhF/rpl32 (IGS)
di	(AT)7	116570	116583	ndhF/rpl32 (IGS)
mono	(A)15	116633	116647	ndhF/rpl32 (IGS)
mono	(A)8	116682	116689	ndhF/rpl32 (IGS)
mono	(T)10	117359	117368	rpl32/trnL-UAG (IGS)
mono	(A)8	117373	117380	rpl32/trnL-UAG (IGS)
mono	(A)11	117436	117446	rpl32/trnL-UAG (IGS)
mono	(T)8	118387	118394	ccsA (CDS)
tetra	(AATA)3	119264	119275	ndhD (CDS)
di	(AT)10	121253	121272	psaC/ndhE (IGS)
di	(TA)4	121274	121281	psaC/ndhE (IGS)
tetra	(TTTA)3	121801	121812	ndhE/ndhG (IGS)
mono	(T)9	122539	122547	ndhG/ndhI (IGS)
mono	(T)9	123014	123022	ndhI (CDS)
di	(AT)8	124208	124223	ndhA (INTRON)

mono	(T)8	124302	124309	ndhA (INTRON)
tetra	(ATTC)3	124327	124338	ndhA (INTRON)
mono	(A)9	124397	124405	ndhA (INTRON)
mono	(A)9	124874	124882	ndhA (INTRON)
di	(TC)5	126283	126292	ndhH (CDS)
mono	(A)9	126781	126789	ndhH/rps15 (IGS)
di	(AT)4	127490	127497	yef1 (CDS)
mono	(T)8	127881	127888	yef1 (CDS)
mono	(T)11	128303	128313	yef1 (CDS)
mono	(T)11	128426	128436	yef1 (CDS)
mono	(A)8	128608	128615	yef1 (CDS)
mono	(T)9	128671	128679	yef1 (CDS)
mono	(T)8	129624	129631	yef1 (CDS)
mono	(T)9	129903	129911	yef1 (CDS)
mono	(T)11	130108	130118	yef1 (CDS)
mono	(T)9	130167	130175	yef1 (CDS)
mono	(T)10	130287	130296	yef1 (CDS)
mono	(T)8	130670	130677	yef1 (CDS)
di	(TC)4	130727	130734	yef1 (CDS)
mono	(A)10	130785	130794	yef1 (CDS)
mono	(T)8	131632	131639	yef1 (CDS)
mono	(T)8	131689	131696	yef1 (CDS)
mono	(A)8	131853	131860	yef1 (CDS)
mono	(T)8	132955	132962	yef1 (CDS)
di	(AG)4	135883	135890	rrn23 (CDS)
mono	(A)8	139205	139212	yef1 (CDS)
mono	(T)8	143066	143073	yef1 (CDS)
mono	(A)8	143232	143239	trnI (INTRON)
mono	(A)8	146186	146193	ndhB (INTRON)
di	(CT)4	147172	147179	ndhB (INTRON)
di	(AT)4	147916	147923	ndhB/trnL-CAA (IGS)
di	(TA)4	149467	149474	yef2 (CDS)
mono	(T)9	152852	152860	yef2 (CDS)
mono	(T)8	154115	154122	yef2 (CDS)
mono	(T)8	154308	154315	yef2 (CDS)
di	(TC)4	155017	155024	yef2 (CDS)
di	(TC)4	156019	156026	yef2 (CDS)
di	(TC)4	156031	156038	yef2 (CDS)
mono	(T)8	158180	158187	rpl2/trnH-GUG (IGS)
mono	(A)11	158395	158405	trnH-GUG/rps19 (IGS)
mono	(A)9	158433	158441	trnH-GUG/rps19 (IGS)

---