

**UNIVERSIDADE FEDERAL DE ALAGOAS - UFAL  
CENTRO DE CIÊNCIAS AGRÁRIAS - CECA  
PROGRAMA DE PÓS-GRADUAÇÃO EM PRODUÇÃO VEGETAL**

**RENATO ÁQUILA SOUZA DA SILVA**

**PLASTOMA DE *Syagrus coronata* (Martius) Beccari E ANÁLISE FILOGENÉTICA  
EM ARECOIDEAE (ARECACEAE)**

**RIO LARGO  
2016**

RENATO ÁQUILA SOUZA DA SILVA

**PLASTOMA DE *Syagrus coronata* (Martius) Beccari E ANÁLISE FILOGENÉTICA  
EM ARECOIDEAE (ARECACEAE)**

Dissertação apresentada ao Programa de Pós-graduação em Produção Vegetal da Universidade Federal de Alagoas – UFAL, para a obtenção do título de Mestre em Agronomia.

Orientador: Prof. Dr. Cícero Carlos de Souza Almeida

RIO LARGO  
2016

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**

Bibliotecária Responsável: Janaina Xisto de Barros Lima

S586p    Silva, Renato Áquila Souza da.  
          Plastoma de *Syagrus coronata* e análise filogenética em arecoideae  
(Arecaceae) / Renato Áquila Souza da Silva. – 2016.  
          38 f. : il.

Orientador: Cícero Carlos de Souza Almeida.  
Dissertação (mestrado em Produção Vegetal) – Universidade Federal  
de Alagoas. Centro de Ciências Agrárias. Rio Largo, 2016.

Bibliografia: f. 33-38.

1. Cloroplasto. 2. Genoma. 3. Ouricuri. I. Título.

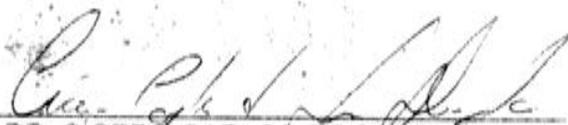
CDU: 575.113

TERMO DE APROVAÇÃO

RENATO ÁQUILA SOUZA DA SILVA

**"PLASTOMA DE *Syagrus coronata* E ANÁLISE FILOGENÉTICA EM ARECOIDEAE  
(ARECACEAE)"**

Dissertação apresentada e avaliada pela banca examinadora em nove de setembro de 2016, como parte dos requisitos para obtenção do título de Mestre em Agronomia, área de concentração em Produção Vegetal do Programa de Pós-Graduação em Agronomia "Produção Vegetal" da Unidade Acadêmica Centro de Ciências Agrárias da UNIVERSIDADE FEDERAL DE ALAGOAS.



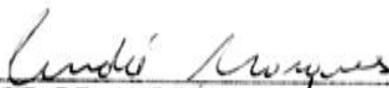
PROF. DR. CÍCERO CARLOS DE SOUZA ALMEIDA

Presidente



PROF.<sup>a</sup> DR.<sup>a</sup> JUREMA ROSA QUEIROZ SILVA

Membro



PROF. DR. ANDRÉ SECO MARQUES DA SILVA

Membro

RIO LARGO – AL  
Setembro/2016

## **Ofereço**

**À minha família, meus amados Pais Maria Betânia Souza da Silva e José Arnaldo da Silva Santos**

**Meu irmão Arnaldo Silas Souza da Silva**

**E a meus Avós Maria Bernadete Souza Santos e Renato Correia dos Santos**

**Por toda educação e carinho,**

**que recebo de vocês desde o dia em que nasci, muito obrigado!**

**À minha linda esposa Elisama Viana Pugliesi Souza,  
pela paciência, dedicação e amor,  
e por sempre estar ao meu lado me apoiando  
e fazendo-me cada dia mais feliz, e meu príncipe Gabriel Áquila,  
o tesouro de DEUS em minha vida!  
Dedico**

À Deus primeiramente, à minha amada família, aos meus amigos de pesquisa e trabalho, e á todos que contribuíram e fazem parte desta vitória.

## **AGRADECIMENTOS**

Agradeço primeiramente ao meu DEUS todo poderoso, que sempre está comigo, guiando meus passos e me abençoando a cada dia, sem ELE eu não sou nada, e tudo em minha vida é para honra e glória do seu nome.

Agradeço a minha família que me incentiva e me ajuda em busca dos meus objetivos, sem a presença deles eu não teria concluído mais essa etapa na minha carreira profissional.

Ao meu brilhante orientador e amigo Prof. Dr. Cícero Carlos por todo conhecimento a mim transmitido de forma clara, por todo acompanhamento e pela paciência no decorrer de todo projeto, além dos conselhos e direcionamento no desenvolver do meu trabalho, sei que cresci muito no período que passei sob sua orientação.

Aos meus companheiros de luta em busca de novas oportunidades, e de crescimento profissional Daniel, Edilma, Meriele, Suzane, Lucília, Beatriz, Cleide, Paulo, e à todos que compõem o laboratório de Recursos Genéticos, e o Programa de Pós-Graduação em Agricultura e Ambiente da UFAL-Arapiraca.

A todos professores que tive o prazer de conviver nesses anos que se passaram durante o Mestrado, muito obrigado por todo ensinamento e dificuldades implantadas, pois tudo contribuiu para o nosso crescimento.

E a todos que direta ou indiretamente contribuíram para que eu conquistasse esta vitória tão importante e especial na minha vida, pessoal e profissional. Obrigado a todos, que Deus vos abençoe !!!

## RESUMO

*Syagrus coronata* classificada como pertencente à família Arecaceae, subfamília Arecoideae, é uma espécie de palmeira nativa do Brasil que possui grande importância econômica, ecológica e social. O gênero *Syagrus* possui 59 espécies, e sua origem é datada em cerca de 27,02 milhões de anos e possui *Cocos* como gênero mais próximo. O objetivo do trabalho foi montar o genoma plastidial de *S. coronata* usando dados de NGS e realizar uma filogenia utilizando genomas completos de cloroplastos para a subfamília Arecoideae. Foram utilizados 77 milhões de *Illumina single end reads* de 100 pb mapeados, para obtenção do genoma de *S. coronata*. O genoma completo de *S. coronata* e outros genomas de Arecoideae disponíveis no NCBI foram usados para obtenção da filogenia para Arecoideae. O genoma de *S. coronata* apresentou 155.053 pb, com IRA possuindo 26.476 pb, IRB 26.647 pb, LSC com 84.410 pb e SSC com 17.520 pb. Foram identificados 90 genes, 8 rRNAs, 39 tRNAs e 30 regiões repetitivas. A comparação do genoma plastidial de *S. coronata* com as espécies de Arecoideae mostrou *Cocos nucifera* como mais próxima e *Chamaedorea seifrizii* como mais divergente, e análise filogenética usando os genomas plastidiais corrobora com filogenias prévias para a subfamília Arecoideae. Conclui-se que o uso de *Illumina single end reads* foi eficiente para sequenciar o genoma plastidial de *S. coronata*, e que filogenia usando genomas completos de cloroplastos apresenta similar topologia quando comparados com filogenia usando poucas regiões plastidiais e genes nucleares.

**Palavras-chave:** Cloroplasto; Genoma; Ouricuri; Sequenciamento de Nova Geração; Bioinformática;

## ABSTRACT

*Syagrus coronata* classified as belonging to the family Arecaceae, subfamily Arecoideae, is a species of palm native to Brazil that has great economic, ecological and social importance. The genus *Syagrus* has 59 species, and its origin is dated to about 27.02 million years and has *Cocos* as its closest genus. The objective of the work was to assemble the plastidial genome of *S. coronata* using NGS data and perform a phylogeny using complete chloroplast genomes for the subfamily Arecoideae. A total of 77 million *Illumina single end reads* of 100 bp were mapped to obtain the *S. coronata* genome. The complete genome of *S. coronata* and other Arecoideae genomes available in NCBI were used to obtain the phylogeny of Arecoideae. The genome of *S. coronata* showed 155,053 bp, with IRA having 26,476 bp, IRB 26,647 bp, LSC with 84,410 bp e SSC with 17,520 bp. We identified 90 genes, 8 rRNAs, 39 tRNAs and 30 repetitive regions. The comparison of the plastidial genome of *S. coronata* with the Arecoideae species showed *Cocos nucifera* as the closest and *Chamaedorea seifrizii* as more divergent, and phylogenetic analysis using the plastid genomes corroborates with previous phylogenies for the subfamily Arecoideae. It is concluded that the use of *Illumina single end reads* was efficient to sequence the plastidial genome of *S. coronata*, and that phylogeny using complete chloroplast genomes presents similar topology when compared to phylogeny using few plastid regions and nuclear genes.

Key words: Chloroplast; Genome; Ouricuri; New Generation Sequencing; Bioinformatics;

## LISTA DE FIGURAS

- Figura 1 – Gráfico de cobertura do genoma de *S. coronata* após mapeamento dos Illumina short reads, obtidos pelo sequenciamento Hiseq 2500. O anel mais interno representa o genoma de *S. coronata*, o segundo representa a cobertura dos reads mapeados (Média 128x) e o terceiro indica as regiões LSC, SSC, IRA e IRB. As anotações em vermelho mostram as regiões usadas para validação usando sequências provenientes do método de Sanger.....24
- Figura 2 – Completo Mapa do genoma Plastidial de *S. coronata*. Genes dentro e fora do círculo são transcritos em direções sentido horário e anti-horário, respectivamente. LSC, região longa de cópia única; SSC, região pequena de cópia única; IR, repetição invertida.....27
- Figura 3 – Comparação de genomas pertencentes a Arecoideae. (A) Círculos dos genomas das espécies: *Cocos nucifera* (roxo), *Elaeis guineensis* (vermelho), *Podococcus barteri* (verde), *Chamaedorea seifrizii* (azul), *Areca vestiaria* (laranja), *Veitchia arecina* (azul claro), *Pseudophoenix vinifera* (amarelo).(B) A análise filogenética usando *neighbor joining*. Os valores suportados são estimados com inicialização.....30

## LISTA DE TABELAS

Tabela 1 – Acessos e referências de Arecoideae e Grupo externo utilizadas para comparação genômica no estudo.....	22
Tabela 2 – Lista de genes presentes no plastoma de <i>S. coronata</i> .....	26
Tabela 3 – Sequências repetitivas no genoma do cloroplasto do <i>S. coronata</i> .....	28
Tabela 4 – Comparação dos genomas plastidiais das espécies de Arecoideae e Out group, alinhados com Genoma de <i>S. coronata</i> .....	29

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	11
<b>2</b>	<b>REVISÃO DE LITERATURA</b> .....	13
2.1	Biomás Brasileiros .....	13
2.2	Ordem Arecales .....	15
2.3	<i>Syagrus coronata</i> .....	15
2.3.1	Taxonomia, Caracterização Morfológica e Distribuição .....	15
2.3.2	Importância Socioeconômica e Ambiental .....	16
2.4	Genoma Plastidial .....	18
2.5	Sequenciamento de Nova Geração (NGS) .....	19
<b>3</b>	<b>OBJETIVOS</b> .....	20
3.1	Geral .....	20
3.2	Específicos .....	20
<b>4</b>	<b>MATERIAL E MÉTODOS</b> .....	21
<b>5</b>	<b>RESULTADOS</b> .....	23
<b>6</b>	<b>DISCUSSÃO</b> .....	31
<b>7</b>	<b>CONCLUSÃO</b> .....	32
<b>8</b>	<b>REFERÊNCIAS</b> .....	33

## 1 INTRODUÇÃO

O gênero *Syagrus* é de ocorrência neotropical, sendo classificado como pertencente à família Arecaceae, subfamília Arecoideae, tribo Cocoseae, subtribo Atellaineae e existem cerca de 59 espécies no gênero (DRANSFIELD et al., 2005; LEITMAN et al., 2015). Análises filogenéticas indicam que *Syagrus* possui próximo relacionamento com os gêneros *Cocos*, *Attalea*, *Butiae* *Allagoptera* (BAKER et al., 2011; ASMUSSEN et al., 2006; MEEROW et al., 2009) e análise citogenética sugere que o gênero possui número básico  $n = 16$  cromossomos (MORO et al., 1999). Estudo com datação indica que o gênero se originou a cerca de 27 milhões de anos e o ancestral que compartilha como gênero *Cocos* ocorreu a cerca de 34 milhões de anos (MEEROW et al., 2009).

*Syagrus coronata* (Mart.) Becc. é uma espécie de palmeira também conhecida por licuri, ouricuri, aricuri, coqueiro cabeçudo, coqueiro dicori, licurizeiro, nicuri, urucuri, alicuri e baba-de-boi. (DRUMOND, 2007; LUGARINI et al., 2012). É uma palmeira nativa do Brasil, que possui alta importância econômica, ecológica e social (CREPALDI et al., 2001) tendo sua distribuição geográfica nos biomas Caatinga e Floresta Atlântica. Seus frutos possuem boa composição nutricional com alto valor energético (CREPALDI et al., 2001), que são consumidos na alimentação humana ou extraído o óleo para uso na culinária e na fabricação de biodiesel (LA SALLES et al., 2010), além das folhas que também são aproveitadas para confecção de diversos produtos (BONDAR, 1938). Devido à relevância econômica das plantas, como as palmeiras, diversos estudos são desenvolvidos para se compreender as características fenotípicas e genotípicas destas espécies. O desenvolvimento de inúmeras pesquisas ligadas a genética de plantas, e em particular de organelas celulares, que carregam o código genético, permite um vasto estudo, com diversas implicações e aplicações.

Os cloroplastos são organelas que possuem papel essencial no funcionamento das células das plantas, pois se envolvem em processos como fotossíntese e fixação de carbono. Essa organela citoplasmática possui genoma independente do genoma nuclear e herança unicamente do genitor feminino. Na maioria das plantas o genoma é organizado na forma circular, com uma variação de tamanho de 100-200 kpb, com uma estrutura “quadripartida” constituída de dois grandes *repeats* invertidos (IRs), que inclui genes ribossomais e outros genes plastidiais. Geralmente possui os genes 16S, 23S, 5S e 27-31 genes tRNA, suficientes para a tradução de todos os aminoácidos; três genes para a subunidade da RNA polimerase (similar procariota) e a maioria dos genes para fotossistema I, fotossistema II, citocromo e

síntese de ATP (GREEN et al., 2011), totalizando aproximadamente 80 proteínas (HUANG et al., 2013).

Os genomas plastidiais são amplamente utilizados em estudos de taxonomia, filogenia, filogeografia e identificação molecular de plantas usando os genes *rbcL* e *matK* e o espaçador intergênico *trnH-psbA*, denominados de DNA barcoding (CBOL, 2009; HUANG et al., 2013; JANSEN et al., 2007; JUNG et al., 2014). Por serem haplóides, e de alta conservação no conteúdo genético e estrutura genômica, os genomas dos cloroplastos têm sido muito utilizados estudos de relações evolutivas, em diferentes níveis taxonômicos das plantas. Com o advento das tecnologias de sequenciamento de DNA de nova geração (NGS - New Generation sequence), o número de genomas de cloroplastos tem aumentado, devido a facilidade de sequenciar DNA, com custo menor, otimização do tempo e alta qualidade dos dados gerados, em comparação com outras metodologias. Atualmente existem 753 genomas plastidiais sequenciados (NCBI: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), no entanto, dentro da subfamília Arecoideae existem apenas seis genomas plastidiais sequenciados, indicando um número relativamente baixo para a subfamília que contém 107 gêneros e cerca de 1300 espécies (DRANSFIELD et al., 2005, 2008). O estudo permitiu a montagem e anotação completa do genoma plastidial de *S. coronata*, e a compreensão da filogenia da espécie dentro da subfamília Arecoideae, quando comparada com as espécies da subfamília que possuem o genoma plastidial sequenciados.

## 2 REVISÃO DE LITERATURA

### 2.1 Biomas Brasileiros de ocorrência de *S. coronata*

#### 2.1.1 Caatinga e Mata Atlântica

A caatinga é um bioma genuinamente brasileiro que abrange uma área de aproximadamente 734.478 Km<sup>2</sup>, o que equivale a 11% do território brasileiro, sendo Alagoas, Bahia, Ceará, Paraíba, Pernambuco, Piauí, Minas Gerais, Rio Grande do Norte e Sergipe os estados com ocorrência desse bioma. A caatinga possui uma grande biodiversidade de espécies vegetais, principalmente espécies endêmicas onde das 932 espécies registradas 318 são endêmicas (TABARELLI; VICENTE, 2004), sendo o mesmo de grande importância na conservação ambiental de espécies propensas a extinção. Por apresentar essa grande biodiversidade esse bioma possui também uma importante função social devido à utilização de diversos recursos principalmente para famílias que utilizam do potencial de espécies para alimentação, fins medicinais, artesanato entre outros. Devido a sua relevância sócio-ambiental a UNESCO em 2001 incluiu o bioma Caatinga na Rede Mundial de Reservas da Biosfera, através do Programa *Man e Biosfera* – MaB (PAES; DIAS, 2008).

A Mata Atlântica é um importante bioma com elevada degradação devido à ação antrópica. Atualmente ocupa uma área de 98.800 Km<sup>2</sup>, o que representa apenas 7,6% de sua extensão original. Este complexo bioma contém uma biodiversidade de espécies mais elevada do que a maioria das florestas, e é caracterizada também por altos níveis de endemismo (média de quase 50% do total da biodiversidade) (BROWN & BROWN, 1992).

Entre as centenas de espécies nativas que ocorrem nesses biomas as palmeiras se apresentam com grande importância, sendo as espécies da família Arecaceae uma das mais representativas, devido principalmente às multiformas de exploração dessas espécies, e pela ocorrência das palmeiras em quase todo território da caatinga e Mata Atlântica. As palmeiras possuem cerca de 2000 representantes sendo um dos grupos de maior importância ecológica, principalmente por sua abundância em diversos ecossistemas com sua interação com outros organismos (SOUZA; LORENZI, 2008).

### 2.2 Ordem Arecales

#### 2.2.1 Família Arecaceae

A família Arecaceae antigamente conhecida como Palmae é a única representante da Ordem Arecales, sendo considerado um dos grupos de vegetais mais antigos da terra, e seus vestígios estimados em 120 milhões de anos (LORENZI et al., 2004). A família Arecaceae é a terceira família economicamente mais importante de plantas, seguida das leguminosas (Fabaceae) e gramíneas (Poaceae) (HUANG et al., 2013).

As espécies de Arecaceae são conhecidas como palmeiras, e existem divergências quanto à precisão da quantidade de espécies dentro da família, estima-se que existam entre 2.000 a 2.600 espécies reunidas em 200 a 240 gêneros (JUDD et al., 1999; LORENZI et al., 2004; UICN, 1996). Segundo Dransfield et al. (2008) a família possui 183 gêneros e aproximadamente 2600 espécies distribuídas por todo o mundo. Cerca de 800 espécies ocorrem nas Américas, das quais 437 distribuídas em 50 gêneros estão presentes na América do Sul, sendo 18 gêneros endêmicos (BALSLEV et al., 2011; DRANSFIELD et al., 2008; PINTAUD et al., 2008).

É uma das maiores famílias do mundo, que devido a sua forma e aspecto é a mais característica da flora tropical, e está amplamente distribuída em diversos habitats. As palmeiras caracterizam a maior parte do ecossistema das florestas tropicais concentrando mais de 90% da diversidade de espécies (COUVREUR et al., 2011), se destacando a América do Sul, como um dos principais centros de riqueza e diversidade de Arecaceae (PINTAUD et al., 2008). Possui grande importância ecológica e socioeconômica (ARAÚJO; LOPES, 2012), devido sua alta distribuição e abundância, produtividade e diversidade de usos, destacando a alimentação, uso medicinal, sociocultural e econômica de populações locais (ARAÚJO; LOPES, 2012; MACÍA, et al., 2011; MARISELVAM et al., 2014; ZAMBRANA et al., 2007).

### 2.2.2 Subfamília Arecoideae

A subfamília Arecoideae é considerada a maior e a mais diversificada dentro da família Arecaceae (DRANSFIELD et al., 2008). Estão incluídas neste grupo 60% dos gêneros de palmeiras e 50% das espécies (BAKER et al., 2011). As palmeiras representantes dessa subfamília ocorrem principalmente em florestas tropicais e em alguns habitats úmidos e secos. No Brasil há ocorrência de 200 espécies distribuídas em cerca de 43 gêneros (LORENZI et al., 2004), sendo encontrada grande variedade de espécies na região Nordeste do País com cerca de 70 espécies distribuídas em 16 gêneros, sendo o gênero *Syagrus* o que apresenta maior

representatividade em relação ao número de espécies (HEDERSON; MEDEIROS-COSTA, 2006)

### 2.2.3 Gênero *Syagrus*

O gênero *Syagrus* é composto por 59 espécies descritas, das quais 51 são endêmicas do Brasil. Ocorre em sua maior parte na América do Sul, possuindo como centro de dispersão de espécies as regiões central e leste (LEITMAN et al., 2015).

A morfologia das espécies do gênero *Syagrus* é muito variável. Sendo palmeiras monóicas, polícarpas, de pequeno ou grande porte, solitárias e entouceiradas, com estipe subterrâneo ou elevado, raramente estolonífero, caule liso ou cobertos de tricomas remanescentes das bainhas e das folhas senescentes. Na região semiárida ou de cerrado predomina as espécies acaulescentes e de pequeno porte, enquanto um número menor de espécies de porte arbóreo é encontrado nas áreas tropicais ou subtropicais úmidas (DRANSFIELD et al., 2008).

As espécies que compõem o gênero são em sua maioria plantas altas, com folhas pinadas, cujas pinas estão geralmente agrupadas e dispostas em mais de um plano. As inflorescências interfolares, primariamente ramificadas ou podem ser, em algumas vezes, espigadas, com poucos ou numerosos ramos florais. Apresentam bráctea peduncular quase sempre lenhosa e persistente, geralmente estriada na superfície externa. As flores unissexuais e dispostas em tríades, sendo uma flor central pistilada e as duas flores laterais estaminadas, com seis estames cada (LORENZI et al., 2004).

Os frutos são globosos, ovóides ou elipsoides, com mesocarpo suculento e fibroso. O endocarpo é rígido e coloração marrom-escuro, contendo três poros basais e seu interior possui somente de uma a duas sementes. O endosperma da semente é geralmente homogêneo, podendo ser ruminado em algumas espécies (LORENZI et al., 2010).

## 2.3 *Syagrus coronata* (Mart.) Becc.

### 2.3.1 Taxonomia, Caracterização Morfológica e Distribuição

*Syagrus coronata* (Mart.) Becc. é uma espécie de palmeira também conhecida por licuri, ouricuri, aricuri, coqueiro cabeçudo, coqueiro dicori, licurizeiro, nicuri, urucuri, alicuri e baba-de-boi. (DRUMOND, 2007; LUGARINI et al., 2012). É uma monocotiledônea da

ordem Arecales, família Arecaceae (Palmae), subfamília Arecoideae, tribo Cocoseae, subtribo Attaleinae e gênero *Syagrus* (DRANSFIELD, 2005). Foi descrita inicialmente como *Cocos coronata* Mart., 1826; *Cocos coronata* var. *todari* Beccari, 1886; *Calappa coronata* (Mart.) Kuntze, 1891; *Cocos quinquefaria* Barb. Rodr, 1900; *Glaziova treubiana* Becc., 1910; *Syagrus cononata* var. *todari* (Becc.) Becc., 1916; *Syagrus quinquefaria* (Barb Rodr) Becc., 1916; *Syagrus treubiana* (Becc.) Becc., 1916; no mesmo ano Beccari fez nova modificação, em que à inseriu no gênero *Syagrus* e denominou como *Syagrus coronata* (Mart.) Becc. (GOVAERTS, 2015).

O Ouricuri é uma espécie de palmeira diplóide contendo 16 pares de cromossomos ( $2n=32$ ) (LAM, 1998). Possui tronco ereto de porte médio, podendo chegar a 10 metros de altura, com folhas grandes, de dois a três metros de comprimento, dispostas geralmente em fileiras de cinco no ápice do tronco, formando uma “coroa foliar”, característica que denominou a planta com o acréscimo específico *coronata* (DRUMOND, 2007; LUGARINI et al., 2012). Flores pequenas, de cor amarela e formam cachos de aproximadamente 60 centímetros (cm) de comprimento, que surgem entre dezembro e março (LUGARINI et al., 2012). O fruto é tipo drupas, medindo aproximadamente 1,9 cm de comprimento e 2,3 cm de diâmetro, de cor amarela ao atingir a maturidade fisiológica, distribuídos em cachos com aproximadamente 1070 frutos (LUGARINI et al., 2012). Frutifica todos os meses do ano, sendo nos meses de junho e julho o maior nível de produção (CREPALDI et al., 2006). Os frutos maduros têm polpa com aspecto pegajoso e sabor adocicado, e as sementes são de cor escura quando secas, com tegumento duro que reveste a amêndoa, rica em óleo (DRUMOND, 2007).

Palmeira nativa do semiárido brasileiro (DRUMOND, 2007), predominantemente em regiões secas e áridas da Caatinga, distribuída em parte dos estados de Minas Gerais, Bahia, Sergipe, Alagoas e Pernambuco (NOBLICK, 1991). No estado da Bahia se encontra maior concentração de ouricuri, principalmente nos municípios de Itiúba, Maracás, Milagres, Monte Santo, Santa Teresinha e Senhor do Bonfim (BONDAR, 1942), e a mesma possui grande importância socioeconômica e ecológica.

### 2.3.2 Importância Socioeconômica e Ambiental

A palmeira Ouricuri (*Syagrus coronata* (Mart.) Becc.), é uma das espécies de maior importância ecológica e socioeconômica da região semiárida do Brasil. É muito utilizada pelas populações tradicionais do Brasil, fazendo parte da economia de subsistência (MANSO,

2009) e também é indicada para recuperação de áreas degradadas em risco de desertificação, pelo seu potencial para facilitar a sucessão ecológica em áreas impactadas (LIMA, 2012), sendo considerada uma das palmeiras mais importantes da região Nordeste (ROCHA, 2009).

Na região do semiárido é importante para a subsistência do sertanejo, como fonte de alimento para o gado, aves e animais silvestres (BONDAR, 1938) sendo o principal alimento de animais silvestres como a arara-azul-de-Lear (*Anodorhynchus leari* Bonaparte, 1856). Entretanto, as populações nativas de *S. coronata* têm sofrido redução drástica devido ao desmatamento, às queimadas e à sobre-exploração (DRUMOND et al., 2004), que podem ter como consequência a modificação da estrutura populacional e espacial dessas espécies.

Suas folhas são utilizadas na confecção de sacolas, chapéus, vassouras, espanadores, etc. Assim como para retirada da cera do ouricuri utilizada na fabricação de papel carbono, graxa para sapatos, móveis e pintura de automóveis, sendo considerada equivalente a da carnaubeira. A amêndoa é consumida *in natura*, é também utilizada para fabricação de cocadas, licores, e o “leite” de licuri, muito utilizado na culinária baiana. Da amêndoa pode também ser extraído o óleo usado em culinária típica da população na região do semiárido (BONDAR, 1938).

Outra aplicação industrial com a utilização do óleo de ouricuri é na produção de saponáceos (sabão em pó, detergentes, sabão em barra e sabonetes finos) considerados de alta qualidade, visto que o ouricuri é considerado o melhor óleo brasileiro para a produção de sabão (SANTOS; SANTOS, 2002).

O resíduo obtido com a extração do óleo dá origem a uma torta que pode ser comercializada para fins de alimentação animal. Esta torta apresenta 41% de substâncias não anotadas, 19% de proteínas, 16% de celulose e 11% a 12% de óleo, sendo ótima ração adicional para vacas leiteiras de bom padrão racial, e para o desenvolvimento precoce de animais de corte e para reprodutoras (GOMES, 1977).

Na composição nutricional do fruto do ouricuri se destacam o teor de lipídios (49,2%) e proteínas (11,5%), da amêndoa e o teor de carboidratos totais da polpa dos frutos (13,2%). O teor de lipídios é elevado e similar ao padrão encontrado para outras espécies de palmeiras, porém, o teor de proteínas é superior ao encontrado em frutos de espécies de palmeiras amazônicas. O teor de carboidratos totais da polpa do ouricuri é semelhante ao de espécies amazônicas, exceto *Euterpe oleracea* Mart., o açaí ou juçara (57,4%) e *Bactrisgasipaes* H. B. K., e a pupunha (14,5 a 84%). O b-caroteno é a principal vitamina encontrada na polpa dos frutos do ouricuri, e mesmo sendo o seu teor menor (2,6  $\mu\text{g.g}^{-1}$ ) que o de outras espécies de

palmeiras, ainda é uma boa fonte dessa vitamina, sobretudo porque em períodos de seca severa constitui-se o único alimento disponível na vegetação (CREPALDI et al., 2001).

Devido a sua diversidade climática, solo e flora, o Brasil possui mais de 200 espécies oleaginosas. Essas espécies são utilizadas na produção de biodiesel (BELTRÃO, 2006). O ouricuri é uma oleaginosa com grande potencial para produção de biodiesel, despertando o interesse dos produtores rurais do nordeste em utilizar de forma sustentável, com manejo adequado, as populações naturais de ouricuri para fins de produção de biodiesel (DRUMOND, 2007).

## 2.4 Genoma Plastidial

Os cloroplastos possuem papel essencial no funcionamento das células das plantas, envolvendo a fotossíntese e fixação de carbono. São altamente especializadas e possuem um complexo de proteínas essenciais para a fotossíntese. Eles têm seu próprio DNA genômico (cpDNA), presente em muitas cópias. O DNA de cloroplastos está entre as estruturas mais estudadas do genoma da planta, sendo geralmente o formato desse genoma uma molécula circular (KOLODNER; TEWARI, 1972; 1979).

Essa organela citoplasmática possui genoma independente do genoma nuclear e herança unicamente pelo genitor feminino (XIONG et al., 2009). Entretanto, estudos recentes sugerem habilidade de transferência de genes dos cloroplastos para o núcleo (GREEN et al., 2011). Sua variação de tamanho (100-200 kpb) numa estrutura “quadripartida”, é constituída de dois grandes *repeats* invertidos (IRs), que inclui genes ribossomais e outros genes plastidiais. Geralmente possui os genes 16S, 23S, 5S e 27-31 genes tRNA, suficientes para a tradução de todos os aminoácidos; três genes para a subunidade da RNA polimerase (similar procariota) e a maioria dos genes para fotossistema I, fotossistema II, citocromo e síntese de ATP (GREEN et al., 2011), totalizando aproximadamente 80 proteínas (HUANG et al., 2013)

As diferenças nos genomas plastidiais estão relacionadas com a presença/ausência de introns *repeats*, fatores de transcrição, organização física do genoma, incluindo espaços intergênicos e rearranjos. Entre as plantas terrestres, incluindo as vasculares, a variação no tamanho está entre 120 e 160 kpb, por outro lado as algas vermelhas possuem genomas maiores (232-252 kpb), entretanto, seus genomas são mais compactos, com nenhum IRS, introns e pouco espaço intergênico, mas possuem elevado número de genes em comparação com os genomas plastidiais de plantas terrestres.

O conhecimento de regiões do cloroplasto tem sido de extrema importância em estudos de filogenia e ou filogeografia, permitindo inferir sobre os relacionamentos entre espécies e as relações demográficas da biodiversidade nos mais diversos biomas do mundo. Os genomas plastidiais têm sido amplamente utilizados para inferências intraespecíficas e interespecíficas, como os estudos de filogeografia (análise espacial da variação genética usando árvores de genes) usando genes e espaçadores intergênicos (SCARCELLI et al., 2011), filogenias para determinação das relações entre espécies e determinação da identificação molecular de plantas usando os genes *rbcL* e *matK* e o espaçador intergênico *trnH-psbA*, denominados de DNA *barcoding* (CBOL, 2009).

## 2.5 Sequenciamento de Nova Geração (NGS)

As técnicas de sequenciamento tradicional requerem muitos esforços metodológicos que demandam tempo, como a clonagem dos fragmentos de DNA por vetores bacterianos para amplificação.

A técnica utilizada até 2005 para obter a sequência de DNA era o método de Sanger, que foi usado para obter o genoma humano e o primeiro genoma de planta (*Arabidopsis thaliana*), entretanto esse método possui elevado custo por genoma. Essa limitação mudou após 2005 com o advento da tecnologia de sequenciamento de nova geração ou NGS (Next Generation Sequencing). NGS é uma técnica que permite obter informações genômicas, incluindo os genes e a fração repetitiva do genoma como os elementos de transposição, sequências de DNA satélites e microssatélites. Esse método consiste de várias técnicas que basicamente dependem da preparação do DNA (Biblioteca), sequenciamento massivo em paralelo e método de alinhamento das sequências, permitindo produzir milhões de pequenas sequências de DNA (chamadas de *reads*) com variação de 25 a 700 pb de comprimento (METZKER, 2010). Existem várias plataformas para sequenciamento NGS, como Roche/454, Illumina/Solexa, ABI/SOLiD e Helicos Biosciences.

As aplicações das informações obtidas por NGS são variadas, incluindo sequenciamento de organelas citoplasmáticas; variantes de genes em populações, análise de transcriptomas em células, tecidos e organismos (RNA-seq); estudos epigenéticos e estrutura da cromatina e descoberta de novos genes em estudos de metagenômicas. Para organismos não modelos ou negligenciados no âmbito científico, NGS permite obter dados para exploração dos genomas, principalmente em genomas grandes como os de plantas (KELLY, LEITH, 2011).

### **3 OBJETIVOS**

#### **3.1 Geral**

- a) Obter o genoma completo de cloroplasto de Ouricuri (*S. coronata*), com a utilização da técnica de sequenciamento de nova geração (NGS), e realizar a análise filogenética dentro da subfamília Arecoideae.

#### **3.2 Específicos**

- a) Montar o genoma completo para o cloroplasto da espécie *S. coronata*;
- b) Obter anotações do genoma, identificando os genes, repeats, rRNA e tRNA;
- c) Realizar uma filogenia utilizando genomas completos de cloroplastos para a subfamília Arecoideae.

## 4 MATERIAL E MÉTODOS

### 4.1 Coleta do material vegetal, extração de DNA e sequenciamento

O material vegetal de *S. coronata* foi coletado no Estado de Alagoas, Brasil, e a extração do DNA foi realizada em material vegetal vivo usando aproximadamente 2 cm<sup>2</sup> de folhas seguindo o método de extração de Cetyl Trimethylammonium Bromide - CTAB (Doyle & Doyle, 1987). A qualidade e quantidade do DNA total extraído foi verificada por meio de visualização em gel de agarose 1% e de espectrofotômetro, respectivamente. A amostra de DNA foi fragmentada em 200-500 pb para construir a biblioteca de sequenciamento. Os fragmentos foram ligados com adaptadores utilizando o kit “*Nextera DNA Sample Preparation*” (Illumina) e sequenciados os *single reads* com tamanho de 100 nt através da plataforma Illumina HiSeq2500. O sequenciamento foi realizado no Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LacTad) na Universidade Estadual de Campinas – UNICAMP, São Paulo.

### 4.2 Obtenção do genoma, validação e anotação

Para obtenção do genoma plastidial parcial (*Draft Genome*), foi realizado um mapeamento usando 77 milhões de reads (segmentos de leituras), tendo como referência o genoma plastidial de *C. nucifera*. O mapeamento dos reads foi realizado usando o software Geneious (<http://www.geneious.com>). O draft genome foi corrigido utilizando contigs de novo. Os contigs de novo foram contruídos usando 77 milhões de reads nos softwares Ray (BOISVERT et al., 2012) e abyss (SIMPSON et al., 2009), com tamanho mínimo de 500 nt e cobertura de 8x. Os contigs foram então mapeados usando o draft genome como referência e corrigidas as regiões com erros na identificação dos nucleotídeos.

A validação foi realizada utilizando sequências obtidas por meio do sequenciamento de Sanger, usando as regiões *trnH-psbA*, *rps16*, *trnD-trnT*, *atpB-rbcL*, *rbcL*, *petB-petD* e *rpl2*, usando *primers* descritos por Scarcelli et al., (2011). As reações de PCR (Reação em Cadeia da Polimerase), foram realizadas com um volume de 50 µl, contendo 5 µl de tampão de reação; 1,5 mM MgCl<sub>2</sub>; 0,2 mM dNTP; 1,25 U Taq DNA *polimerase*; 0,5 µM de cada *primer* e 200 ng do DNA. A amplificação foi realizada com uma desnaturação inicial a 94 °C por 3 min, seguidos de 40 ciclos de 94 °C por 30 s, anelamento 55° - 60 °C por 30 s e extensão final de 72 °C por 10 min. As reações da PCR foram conduzidas em termociclador

*BioCycler* e os produtos da PCR foram submetidos à eletroforese em gel de agarose a 1% para a confirmação da amplificação. Os produtos da PCR foram sequenciados usando *BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems®)* com eletroforese em um 3500 *Genetic Analyzers (Applied Biosystems Inc., Foster City, Ca, USA)*.

A anotação do genoma foi realizada usando o plastoma de *C. nucifera* (Genbank: KF285453) como referência usando o software Geneious e conferida com a anotação realizada usando o software DOGMA, (Dual Organellar Genome Annotation - <http://dogma.ccbb.utexas.edu>). As anotações foram verificadas e quando necessário foram manualmente corrigidas. Uma representação gráfica do plastoma foi realizada usando Geneious. Para análise de *repeats*, estes foram identificados usando Phobos software (MAYER, 2010).

### 4.3 Comparação entre genomas e filogenia

Foram utilizados seis genomas plastidiais da subfamília Arecoideae disponíveis no NCBI (Tabela 1); o genoma de *S. coronata* desenvolvido neste estudo e o genoma de *Pseudophoenix vinifera*, pertencente à subfamília Ceroxyloideae, como grupo externo a Arecoideae. Todos os genomas de cloroplastos foram comparados usando BLAST Ring Image Generator (BRIG) assumindo os ajustes padrões do software (ALIKHAN et al., 2011). Para análise filogenética, todos os genomas foram alinhados usando MAFFT (KATO E STELEY, 2013) e uma árvore de neighbor joining foi obtida usando o modelo HKY para substituição de nucleotídeos.

Tabela 1- Acessos e referências de Arecoideae e Grupo externo utilizadas para comparação genômica no estudo.

Espécies	Tribo	GenBank n°.	Referência
<i>Syagrus coronata</i>	Cocoseae	KT896548	Produced in thisstudy
<i>Cocos nucifera</i>	Cocoseae	KF285453	Huang et al. 2013
<i>Elaeisguineensis</i>	Cocoseae	JF274081	Uthaipasanwong et al. 2012
<i>Podococcusbarteri</i>	Podococceae	KR347117	Scarcelliet al. 2015*
<i>Areca vestiaria</i> Giseke	Areceae	KT312940	Barrett et al. 2015
<i>Veitchia spiralis</i> Becc	Areceae	KT312917	Barrett et al. 2015
<i>Chamaedorea seifrizii</i>	Chamaedoreae	JX088667	Barrett et al. 2013
<i>Pseudophoenix vinifera</i>	Cyclospatheae	JX088662	Barrett et al. 2013

\*Não publicado

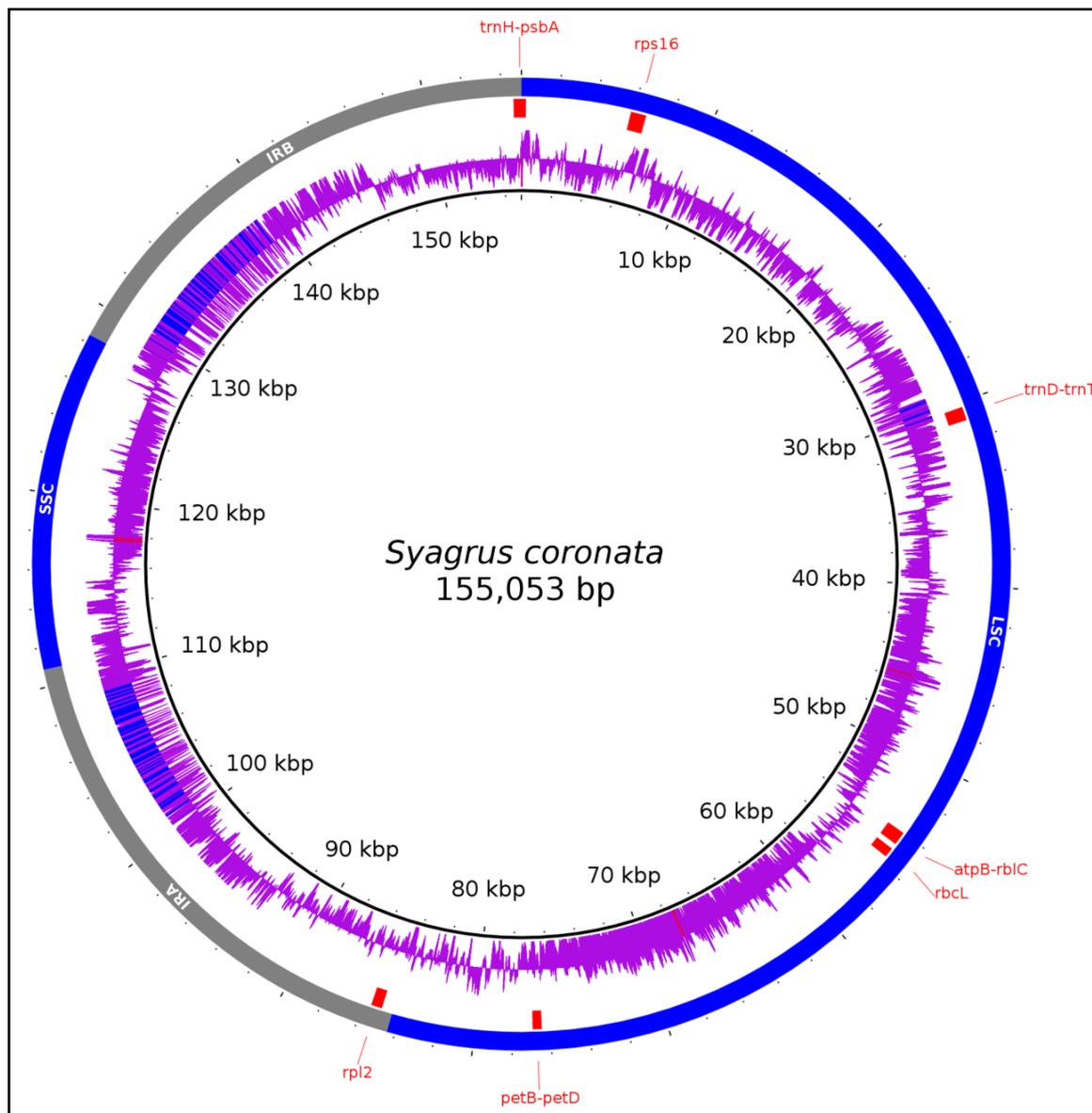
## 5 RESULTADOS

### 5.1 Sequenciamento do genoma de cloroplasto: montagem com genoma de referência

Foram utilizados 77 milhões de reads com tamanho de 100 pb, apresentando mais 90 % das leituras menos de 1% de probabilidade de erro de identificação de bases. Estas leituras de *S. coronata* foram mapeadas utilizando o genoma de *C. nucifera* como referência, para obtenção do genoma parcial de *S. coronata*. Após o mapeamento, o draft genoma gerado, apresentou um tamanho de 160.103 pb, sendo utilizadas 316.161 reads mapeados na montagem. Foram identificadas 28 regiões com falhas, no total 123 nucleotídeos não identificados, correspondendo a 0,079% do genoma final. A maior parte das falhas foram identificadas na região de repetição invertida IRA, sendo 10 regiões com erros, mais precisamente na região entre 60-70 kbp. As falhas corrigidas manualmente através de alinhamentos com reads mapeados e alinhados com o Draft genoma, resultaram em sequências com 100% de identidade.

. Paralelamente, foram usados os 77 milhões de reads para obter contigs *de novo* com tamanho mínimo de 500 nt. Esses contigs foram mapeados no draft genoma de *S. coronata*, permitindo corrigir os erros facilmente. Após a obtenção do genoma final de *S. coronata*, foi realizado um mapeamento com short reads, possibilitando mapear os reads com ausência de erros e 100% de identidade, resultando em uma cobertura média de 128x, com mínimo de 16x e máximo de 299x (Figura 1). Adicionalmente, as regiões *trnH-psbA*, *rps16*, *trnD-trnT*, *atpB-rbcL*, *rbcL*, *petB-petD* e *rpl2* foram sequenciadas usando o método Sanger e alinhadas com o genoma, resultando em 100% de identidade nos alinhamentos (Figura 1).

Figura 1 – Gráfico de cobertura do genoma de *S. coronata* após mapeamento dos Illumina short reads, obtidos pelo sequenciamento Hiseq 2500. O anel mais interno representa o genoma de *S. coronata*, o segundo representa a cobertura dos reads mapeados (Média 128x) e o terceiro indica as regiões LSC, SSC, IRA e IRB. As anotações em vermelho mostram as regiões usadas para validação usando sequências provenientes do método de Sanger.



Fonte: produzido pelo Autor

### 5.1 Genoma de cloroplasto de *S. coronata*

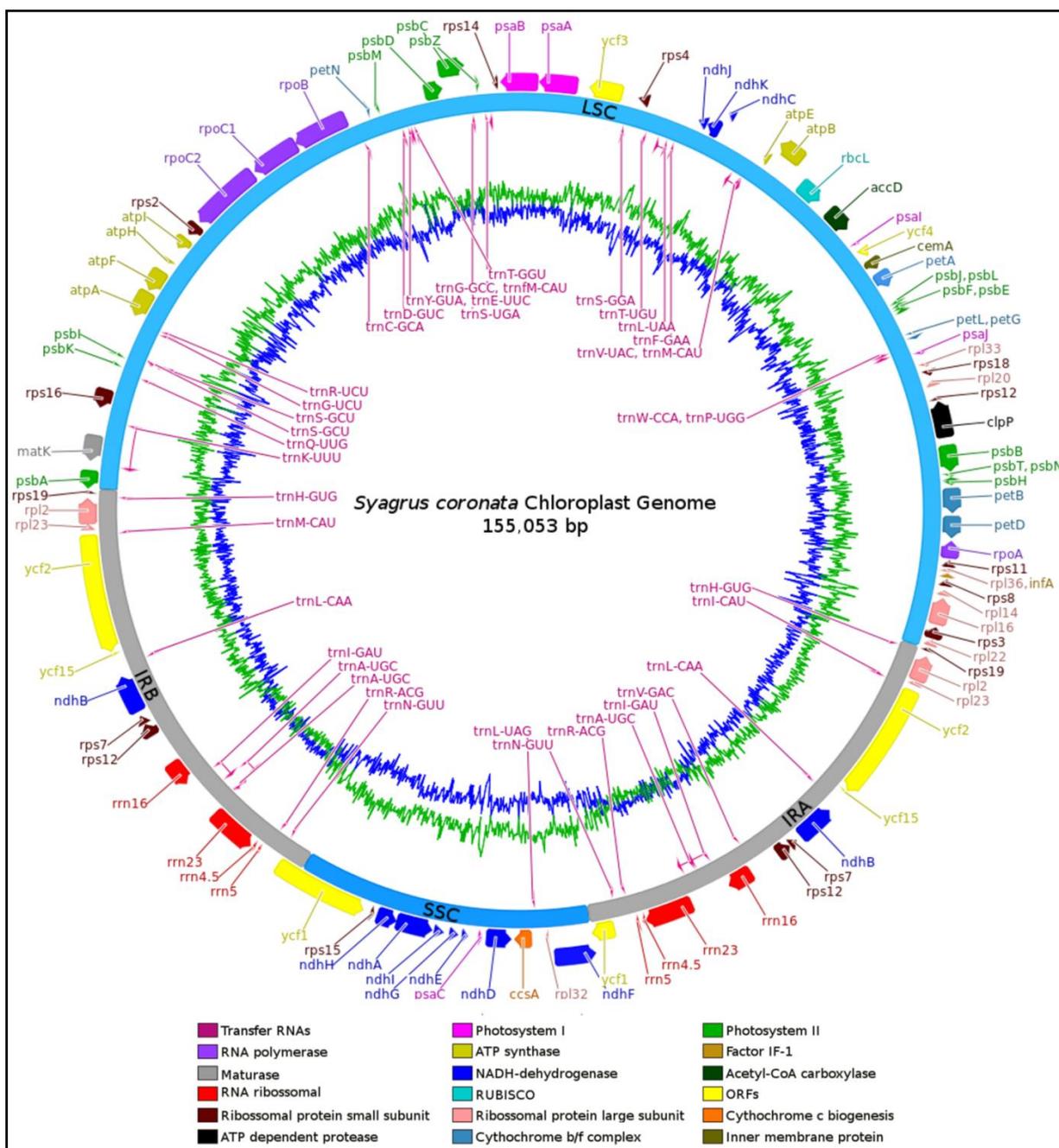
O plastoma de *S. coronata* apresentou 155,053 pb (Genbank: KT896548) e possui uma conservada estrutura quadripartida, com um par de repetições invertidas IRA e IRB com 26.476 e 26.647 pb, respectivamente, que separam a região LSC de 84.410 pb e a região SSC de 17.520 pb (Figura 2). O genoma é constituído de 90 genes, 39 *tRNAs* e 8 *rRNAs* (Tabela 2).

A análise de sequências repetitivas mostrou regiões com tamanho superior a 10 pb, incluindo 25 repeats em tandem, e cinco repeats invertidos (Tabela 3). Dezenove repeats foram encontrados em LSC, cinco repeats em IRB, cinco repeats em IRA e apenas um em SSC. A maioria dos repeats está localizada em espaçadores intergênicos, com apenas 12 repeats presentes em regiões codificantes. Dentre das regiões codificantes, *rpl16* revela o mais longo repeat (28 pb) (Tabela 3).

Tabela 2 - Lista de genes presentes no plastoma de *S. coronata*

	Grupo de genes	Nome dos genes
Síntese de proteínas e replicação de DNA	RNAs de transferência	trnH-GUG (2x), trnM-CAU (2x), trnL-CAA (2x), trnI-GAU (2x), trnA-UGC (3x), trnR-ACG (2x), trnN-GUU (2x), trnL-UAG, trnV-GAC, trnI-CAU, trnP-UGG, trnW-CCA, trnV-UAC, trnF-GAA, trnL-UAA, trnT-UGU, trnS-GGA, trnM-CAU, trnG-GCC, trnS-UGA, trnT-GGU, trnE-UUC, trnY-GUA, trnD-GUC, trnC-GCA, trnR-UCU, trnG-UCC, trnS-GCU (2x), trnQ-UUG, trnK-UUU
	RNAs ribossômicos (16S, 23S, 4.5S, 5S)	rnr16 (2x), rnr23 (2x), rnr4.5 (2x), rnr5 (2x)
	Subunidade pequena da proteína ribossomal	rps2, rps3, rps4, rps7 (2x), rps8, rps11, rps12 (2x), rps14, rps15, rps16, rps18, rps19 (2x)
	Subunidade gree da proteína Ribossomal	rpl2 (2x), rpl14, rpl16, rpl20, rpl22, rpl23 (2x), rpl32, rpl33, rpl36
	Subunidades ( $\alpha$ , $\beta$ , $\beta'$ , $\beta''$ ) da RNA polimerase dependente de DNA	rpoA, rpoB, rpoC1, rpoC2
Fotossíntese	Fotossistema I	psaA, psaB, psaC, psaI, psaJ
	Fotossistema II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ
	Complexo Citocromo b/f	petA, petB, petD, petG, petL, petN
	ATP sintase	atpA, atpB, atpE, atpF, atpH, atpI
	NADH-dehydrogenase	ndhA, ndhB (2x), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK
Grupos Diversos	Subunidade gree da RUBISCO	RbcL
	Fator de iniciação da tradução IF-1	InfA
	Acetyl-CoA carboxylase	AccD
	Citocromo c biogênese	CcsA
	Maturase	MatK
	ATP-dependente de protease	ClpP
Pseudogene ou com função desconhecida	Proteína da membrana interna	CemA
	ORFs hipotéticos de cloroplastos conservados	ycf1 (2x), ycf15 (2x), ycf2 (2x), ycf3, ycf4

Figura 2 - Completo Mapa do genoma Plastidial de *S. coronata*. Genes dentro e fora do círculo são transcritos em direções sentido horário e anti-horário, respectivamente. LSC, região longa de cópia única; SSC, região pequena de cópia única; IR, repetição invertida.



Fonte: produzido pelo Autor

Tabela 3 - Sequências repetitivas no genoma do cloroplasto do *S. coronata*.

Nº	Posição inicial	Tipo	Tamanho (pb)	Número de repats	Unidade de repetição	Região
1	152335, 152347	T	12	2	ATCTCTCTCAAT	IRB: ycf2 gene
2	152042, 152057	T	15	2	TAGGTATAGGAAGAA	IRB: ycf2 gene
3	149738, 149759	T	21	2	GAAGTGACTTGGACAAAAA GA	IRB: ycf2 gene
4	147295, 147319, 147343	T	24	3	ATATCGTCACTATCATCAAT ATCG	IRB: ycf2 gene
5	134549, 134560	T	11	2	AGTATCAGCTA	IRB: entre o espaçador rrn23-rRNA e trnA- UGC
6	118457, 118468	T	11	2	ATAAAAAACCTT	SSC: entre o espaçador ndhE e ndhG
7	104767, 104778	I	11	2	TAGCTGATACT	IRA: entre o espaçador trnA-UGC e rrn23-Rrna
8	91971, 91995, 92019	I	24	3	CGATATTGATGATAGTGAC GATAT	IRA: entre o espaçador trnA-UGC e rrn23-Rrna
9	89558, 89579	I	21	2	TCTTTTTGTCCAAGTCACTT C	IRA: ycf2 gene
10	87266, 87281	I	15	2	TTCTTCCTATACCTA	IRA: ycf2 gene
11	86979, 86991	I	12	2	ATTGAGAGAGAT	IRA: ycf2 gene
12	82472, 82486	T	14	2	TTTATTATTTTTTT	LSC: rpl16 gene
13	81932, 81960	T	28	2	TGAGTTAATAATATAAGTA TATATAAGT	LSC: rpl16 gene
14	81888, 81901	T	13	2	ATAATAATAATA	LSC: rpl16 gene
15	67363, 67378	T	15	2	TTATTAATTATTACT	LSC: entre o espaçador rpl33 e psaJ
16	64878, 64892, 64906	T	14	3	AATAAATATACTAT	LSC: entre o espaçador psbE e petL
17	59366, 57377	T	11	2	ATATTTATATT	LSC: entre o espaçador accD e psaI; SSC
18	59286, 59300	T	14	2	TAAATATTCATTTTC	LSC: entre o espaçador accD e psaI
19	59259, 59270	T	11	2	ACTTATTAACT	LSC: entre o espaçador accD e psaI
20	56614, 56630	T	16	2	AACTCAATACTTCTAT	LSC: entre o espaçador rbcL e accD
21	47149, 47166	T	17	2	AATACTTTACTTTATTT	LSC: entre o espaçador trnL-UAA e trnF-GAA
22	45736, 45748	T	12	2	TATATATAATTA	LSC: entre o espaçador trnT-UGU e trnL-UAA
23	43681, 43693	T	12	2	ATTTTCATCTAAG	LSC: entre o espaçador ycf3 e trnS-GGA
24	31848, 31868	T	20	2	TTAAAAGATATACTCTGGA A	LSC: entre o espaçador trnT-GGU e psbD
25	31492, 31507	T	15	2	AATTCGATATAGATT	LSC: entre o espaçador trnT-GGU e psbD
26	28826, 28845	T	19	2	TATATATATTATATATAAA	LSC: entre o espaçador petN e psbM
27	28458, 28469	T	11	2	TAGTTCATCAA	LSC: entre o espaçador trnC-GCA e petN
28	16973, 16988	T	15	2	TGTCTCATGGAAATT	LSC: rpoC2 gene
29	12707, 12724	T	17	2	TTCTTTATTTGTATTTG	LSC: atpF gene
30	3771, 3783	T	12	2	AATAAATTAAT	LSC: entre o espaçador matK gene e rps16 gene

T representa Tandem, IRA, IRB, LSC e SSC representa região invertida A, região invertida B, região longa de cópia única e região curta de cópia única, respectivamente.

## 5.2 Comparação entre os genomas de cloroplasto na subfamília Arecoideae

A comparação dos genomas da subfamília Arecoideae com o genoma de *S. coronata* mostrou que a espécie *C. nucifera* é a mais próxima com identidade de 98,9% (153.863 pb), 379 SNPs e 1360 Indels. As espécies *Elaeis guineensis*, *Podococcus barberi*, *Pseudophoenix vinifera*, *Veitchia spiralis*, e *Areca vestiaria*, apresentaram uma identidade variado de 94,9 - 95,9% quando comparadas a *S. coronata*, caracterizeo uma alta similaridades entre os

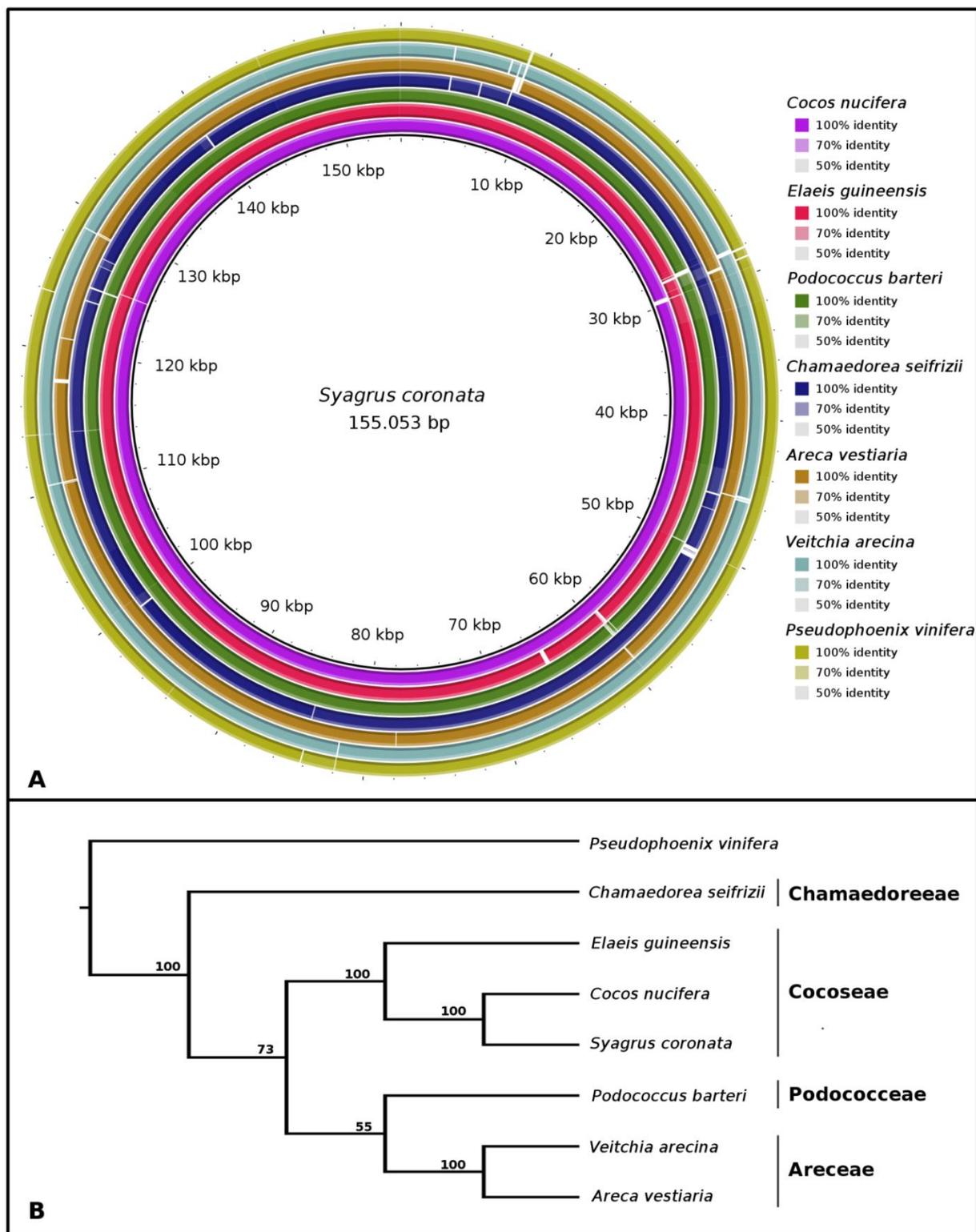
genomas. Por outro lado, a espécie *Chamaedorea seifrizii* apresentou 94,0% (149.911pb) de identidade, 2.214 SPNs e 7.170 Indels (Tabela 4).

Tabela 4 – Comparação dos genomas plastidiais das espécies de Arecoideae e Grupo externo\*, alinhados com Genoma de *S. coronata*

Espécies	Características		Tipo de Polimorfismo	
	Identidade de pareamento	Nucleotídeos idênticos	SNPs	Indels
<i>Cocos nucifera</i>	98.9	153,833	379	1360
<i>Elaeis guineensis</i>	95.9	152,231	989	5462
<i>Podococcus barteri</i>	95.3	151,627	1735	5652
<i>Pseudophoenix vinifera</i> *	95.2	151,710	1563	6164
<i>Veitchia spiralis</i> Becc	95.1	151,414	1349	6419
<i>Areca vestiaria</i> Giseke	94.9	126,275	1379	5252
<i>Chamaedorea seifrizii</i>	94.0	149,624	2214	7170

Os SNPs e Indels estão distribuídos ao longo dos genomas, com exceção da região de 30 kpb, que mostrou maior frequência dos SNPs e Indels (Figura 3A). A análise filogenética mostrou que a tribo Areaceae é mais derivada, Chamaedoreae a mais basal e Coccoceae em um clado intermediário (Figura 3B).

Figura 3- Comparação de genomas pertencentes a Arecoideae. (A) Círculos dos genomas das espécies: *Cocos nucifera* (roxo), *Elaeis guineensis* (vermelho), *Podococcus barteri* (verde), *Chamaedorea seifrizii* (azul), *Areca vestiaria* (laranja), *Veitchia arecina* (azul claro), *Pseudophoenix vinifera* (amarelo). (B) A análise filogenética usando *neighbor joining*. Os valores suportados são estimados com inicialização.



Fonte: produzido pelo Autor

## 6 DISCUSSÃO

A montagem do draft genome usando 77 milhões single reads resultou em um genoma prévio com apenas 28 regiões com erros de montagem. Esses problemas foram resultantes das diferenças genéticas (variabilidade genética) entre o genoma de *S. coronata* com o genoma de *Cocos nucifera* utilizado como referência, o que é comum no processo de montagem de genomas por referência, entretanto, esses problemas foram facilmente identificados e corrigidos com o alinhamento dos *contigs de novo*, e validação por realinhamento dos reads, com obtenção de 100% de similaridade.

A subfamília Arecoideae é a mais abundante da família Arecaceae, com cerca de 107 gêneros e 1300 espécies. Entre as tribos de Arecoideae, Areceae e Cocoseae são as que contêm mais espécies, 630 e 360, respectivamente (DRANSFIELD et al., 2005). Estudos filogenéticos em Arecaceae têm sido desenvolvidos por Baker et al., (2011) usando genes nuclear PRK e RPB; Comer et al. (2016) usando abordagem filogenômica com genes nucleares e Barrett et al. (2016) usando genomas completos de cloroplastos. Em todas as abordagens os resultados mostram similar topologia para os clados, em que Areceae é a tribo mais derivada. Entretanto, o número de genomas de cloroplastos para a subfamília Arecoideae é bastante reduzido, contendo apenas seis espécies sequenciadas. Nesse sentido, o presente trabalho apresenta um genoma adicional e uma análise filogenética mais detalhada para Arecoideae. Curiosamente, a topologia da filogenia usando genes nucleares (BAKER et al., 2011; COMER et al., 2016) e usando genomas completos de cloroplastos são similares, indicando maior consistência na filogenia da subfamília.

O sequenciamento dos genomas de cloroplastos tem aumentado com o advento das técnicas de sequenciamento de nova geração, permitindo obter genomas com maior facilidade, reduzindo tempo e custo, quando comparados com o método de sequenciamento de Sanger (BARRETT et al., 2016). O genoma nuclear de *S. coronata* possui cerca de 2.2 Gb (dados não publicados) e o uso de 77 milhões de *single reads* de 100 nt sequenciados aleatoriamente (incluindo DNA nuclear, mitocondrial e plastidial) foram suficientes para obter o genoma de cloroplasto de *S. coronata*, com uma cobertura média de 128x. A obtenção do genoma foi facilitada pelo uso do plastoma de *Cocos nucifera* como referência, resultando em poucas regiões para correção manual, devido principalmente a similaridade genética de ambas as espécies, com 98,9% dos nucleotídeos idênticos, justificando a distribuição filogenética quando avaliadas.

*Syagrus coronata* é um gênero que surgiu aproximadamente há 27 milhões de anos e tem o gênero *Cocos* como mais próximo (MEEROW et al. 2009). A comparação do plastoma de *S. coronata* com *C. nucifera* e *E. guineensis* mostrou pouca variação, com similaridade acima de 95% entre os genomas, existindo SNPs e Indels, sugerindo que a tribo possui pouca variação nos genomas plastidiais. Os genomas possuem estrutura “quadripartida”, constituindo de dois grandes *repeats* invertidos (IRs), LSC e SSC, tendo *S. coronata* e *C. nucifera* tamanhos muito similares.

## 7 CONCLUSÃO

É possível montar genomas inteiros de cloroplastos, através da Tecnologia de Sequenciamento de Nova Geração, utilizando bibliotecas do tipo *single read* com 100 pb, obtendo altíssimo nível de confiabilidade dos dados, sendo fator determinante a quantidade e qualidade das leituras, que determinam o nível de cobertura do genoma e confiabilidade dos resultados quanto a espécie estudada, além da proximidade parental do genoma de referência. O alinhamento mapeado das leituras, utilizando um genoma de referência, produz dados consistentes, com erros mínimos na identificação de bases, podendo ser facilmente corrigidos manualmente e validados com realinhamentos de short reads, e/ou alinhamentos de contigs e sequências provenientes no sequenciamento de Sanger, sendo eficiente na montagem de genomas. E que a filogenia usando genomas completos de cloroplastos apresenta similar topologia quando comparados com estudos de filogenia usando poucas regiões plastidiais e genes nucleares.

## 8 REFERÊNCIAS

- ALIKHAN, N.-F., PETTY, N.K., BEN ZAKOUR, N.L., BEATSON, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. **BMC Genomics** 12, 402.doi:10.1186/1471-2164-12-402, 2011.
- ARAÚJO, F. R; LOPES, M. A. Diversity of use e local knowledge of palms (Arecaceae) in eastern Amazonia.**Biodiversity Conservation**, v. 21, p. 487–501, 2012.
- ASMUSSEN, C.B.E.A. A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. **Bot. J. Linn. Soc.** 151, 15–38. doi:10.1111/j.1095-8339.2006.00521.x, 2006.
- BAKER, W. J., NORUP, M. V., CLARKSON, J. J., COUVREUR, T. L. P., DOWE, J. L., LEWIS, C. E., PINTAUDE, J-C., SAVOLAINEN, V., WILMONT, T., CHASE, M. W. Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). **Annals of Botany.**, p. 1-16, 2011.
- 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., MEDINA, J., PIRES, J.C., SANTOS, C., WM. STEVENSON, D., ZOMLEFER, W.B., DAVIS, J.I. Plastid genomes reveal support for deep phylogenetic relationships e extensive rate variation among palms e other commelinid monocots. **New Phytol.** 209, 855–870. doi:10.1111/nph.13617, 2016.
- BALSLEV, H. et al. Species Diversity e Growth Forms in Tropical American Palm Communities.**The Botanical Review**, v. 77, p. 381-425, 2011.
- BELTRÃO, N. E. M. **Considerações gerais sobre o pinhão-manso (*Jatropha curcas* L.) e a necessidade urgente de pesquisas, desenvolvimento e inovações tecnológicas para esta planta nas condições brasileiras.** 2006. Disponível em: <www.mda.gov.br>
- BOISVERT, S. RAYMOND, F. GODZARIDIS, É. LAVIOLETTE, F. CORBEIL, J. Ray Meta: scalable *de novo* metagenome assembly e profiling. **Genome Biology**. DOI: 10.1186/gb-2012-13-12-r122, 2012.
- BONDAR, G. O. **Licurizeiro e suas potencialidades na economia brasileira.** Instituto Central de Fomento Economico da Bahia 2:18. 1938.
- BONDAR, G. **As ceras no Brasil e o licuri (*Cocos coronata* Mart.) na Bahia.** Bahia: Instituto Central de Fomento Econômico da Bahia, 1942. p. 53, Boletim 9.
- BROWN, K. S. JR., E G. G. BROWN.**Habitat alteration e species loss in Brazilian forests.**In T. C. Whitmore e J. A. Sayer (Eds.).Tropical deforestation e species extinction, pp. 119–142.Chapman e Hall, London, Engle, 1992.
- CBOL, Plant Working Group, A DNA barcode for le plants,**PNAS**.vol. 106, no. 31, p.12794–12797, 2009.

COMER, J.R., ZOMLEFER, W.B., BARRETT, C.F., STEVENSON, D.W., HEYDUK, K., LEEBENS-MACK, J.H. Nuclear phylogenomics of the palm subfamily Arecoideae (Arecaceae). **Mol. Phylogenet. Evol.** 97, 32–42. doi:10.1016/j.ympev.2015.12.015, 2016.

CREPALDI, I.C.; ALMEIDA-MUREIAN, L.B.; RIOS, M.D.G.; PENTEADO, M.V.C. & SALATINO, A. Composição nutricional do fruto de licuri (*Syagrus coronata* (Martius) Becari). **Revista Brasileira de Botânica** 24: 155-159, 2001.

CREPALDI, I. C.; SALATINO, A.; LIMA-BRITO, A. Notes about phenology e biometry of two *syagrus* species. **Magistra**, v. 18, n. 3, p. 162-167, 2006.

COUVREUR, T. L. P.; FOREST, F.; BAKER, W. J. Origin e global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. **BMC Biology**, v. 9, p. 44, 2011.

DOYLE JJ E DOYLE JL.A rapid DNA isolation procedure for small quantities of fresh leaf tissue.**Phytochemical Bulletin** 19:11-15, 1987.

DRANSFIELD J. et al. A new phylogenetic classification of the palm family, Arecaceae.**Kew Bulletin**, v. 60, p. 559-569, 2005.

DRANSFIELD, J. UHL, N. W., ASMUSSEN, C. B. BAKER, W. J. HARLEY, M. M., LEWIS, C. E. **Genera Palmarum; the evolution e classification of palms**. Richmond, UK: Royal Botanical Garden, Kew. 2008.

DRUMOND, M. A.; KIILL, L. H. P.; LIMA, P. C. F.; OLIVEIRA, M. C.; OLIVEIRA, V. R.; ALBUQUERQUE, S. G.; NASCIMENTO, C. E. S. Estratégias de uso sustentável da biodiversidade da caatinga. In: SILVA, J. M. C.; TABARELLI, M.; LINS, L. V. (Org.). **Biodiversidade da caatinga: áreas e ações prioritárias**. Brasília, DF: Ministério do Meio Ambiente; UFPE, p. 329-340, 2004.

DRUMOND, M. A. Licuri *Syagrus coronata* (Mart.) Becc. **Documentos online**. Setembro, 2007. Disponível em: <<http://www.cpsa.embrapa.br>> Acesso em 12 de junho de 2016.

GOVAERTS, R. **World Checklist of *Syagrus coronata***.Facilitated by the Royal Botanic Gardens, Kew.Published on the Internet; <<http://apps.kew.org/wcsp/>> Acess in 22 of dez of 2015.

GREEN, B. R.; Chloroplast genomes of photosynthetic eukaryotes, In: THE PLANT GENOME: AN EVOLUTIONARY VIEW ON STRUCTURE E FUNCTION. **The Plant Journal** 66, p. 34–44, 2011.

HENDERSON, A.; MEDEIROS-COSTA, J.T. **Arecaceae**.In: BARBOSA, M.R.V.; SOTHERS, C.; MAYO, S.; GAMARRA-ROJAS, C.F.L.; MESQUITA, A.C. (Org.) Checklist das plantas do nordeste brasileiro: angiospermas e gimnospermas. Brasília: Ministério de Ciência e Tecnologia, p.33-34, 2006.

HUANG ,Y-Y.; MATZKE, A.J.M.; MATZKE, M.; Complete Sequence and Comparative Analysis of the Chloroplast Genome of Coconut Palm (*Cocos nucifera*). **PLoS ONE** 8(8): e74736. doi:10.1371/journal.pone.0074736, 2013.

HUANG, WEI-PING., SUN, H., DENG, T., RAZAFIMEIMBISON, S. G., NIE, ZE-LONG., WEN, J. Molecular phylogenetics and biogeography of the eastern Asian eastern North American disjunct *Mitchella* and its close relative *Damnacanthus* (Rubiaceae, Mitchelleae). **Botanical Journal of the Linnean Society**, v. 171, p. 395-412, 2013.

IUCN – International Union for the Conservation of Nature. Palms: Their Conservation e Sustained Utilization. Cambridge: **IUCN/SSC Palm Specialist Group**, 1996.

JANSEN. R. K. CAI, Z. RAUBESON, L. A. DANIELL, H. PAMPHILIS, C. W. LEEBENS-MACK, J. MULLER, K. F. GUISSINGER-BELLIAN, M. HABERLE, R. C. HANSEN, A. K. CHUMLEY, T. W. LEE, SEUNG-B. PEERY, R. MCNEAL, J. R. KUEHL, J. V. E BOORE, J. L. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms e identifies genome-scale evolutionary patterns. **PNAS**. St. Louis, MO. September, 2007.

JUDD, W.S.; CAMPBELL, C.S.; KELLOGG, E.A.; STEVENS, P.F. Plant Systematics. Sunderland: Sinauer Associates, **Inc.**, p.464, 1999.

JUNG, J. KIM, K. H. YANG, K. BANG, KYONG-H. YANG, TAE-J. Practical application of DNA markers for high-throughput authentication of *Panax ginseng* and *Panax quinquefolius* from commercial ginseng products. **Journal of Ginseng Research**, 38.p.123-129, 2014.

KATOH, K., STELEY, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance e usability. **Mol. Biol. Evol.** 30, 772–780. doi:10.1093/molbev/mst010, 2013.

KOLODNER, R. E TEWARI, K. Molecular size e conformation of chloroplast deoxyribonucleic acid from pea leaves. **J. Biol. Chem.** 247, p.6355- 6364, 1972.

KOLODNER, R. E TEWARI, K. Inverted repeats in chloroplast DNA from higher plants. **Proc. Natl. Acad. Sci. USA** 76, p.41-45, 1979.

LAM, V. **Relações de similaridade genética entre espécies de palmeiras com base no estudo de seus cariótipos**. Dissertação (Mestrado), Jaboticabal, São Paulo, p. 95, 1998.

LEITMAN, P. et al. **Areaceae in Lista de Espécies da Flora do Brasil**. Jardim Botânico do Rio de Janeiro. Disponível em: <<http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB15732>>. Acesso em: 02.jul. 2016.

LONGO, D., STEHMANN, JR., BONATO, S. L., SALZANO, F. M., FREITAS, L. B. Filogeografia e validação taxônômica. In: 51° CONGRESSO BRASILEIRO DE GENÉTICA, 2005, Águas de Lindóia. **Anais...** São Paulo, SP: SBMET, 2005.

LOPES, V. DA S. **Morfologia e Fenologia Reprodutiva do Ariri (*Syagrus vagans* (Bondar) Hawkes) – Areaceae numa área de caatinga do Município de Senhor do Bonfim – BA**. 2007. 70 f. Dissertação (Mestrado em Agronomia) – Universidade Federal da Paraíba, Areia, PB, 2007.

LORENZI, H.; SOUZA, H.M.; CERQUEIRA, L.S.C.; COSTA, J.T.M.; FERREIRA, E. **Palmeiras Brasileiras e Exóticas Cultivadas**. Nova Odessa: **Instituto Plantarum**, p.416, 2004.

LORENZI, H. et al. **Flora Brasileira: Arecaceae (Palmeiras)**. Nova Odessa, São Paulo. Instituto Plantarum, p. 385, 2010.

LUGARINI, C.; BARBOSA, A. E. A.; OLIVEIRA, K. G. (org.) **Plano de Ação Nacional para a Conservação da arara-azul-de-lear**. Brasília: ICMBIO – Instituto Chico Mendes de Conservação da Biodiversidade, 2012. Serie espécies ameaçadas de extinção. p. 144, v. 2.

MADABHUSHI, R.S. Separation of 4-color DNA sequencing extension products in noncovalently coated capillaries using low viscosity polymer solutions. **Electrophoresis** 19, 224–230, 1998.

MACÍA, M. J. et al. Palm Uses in Northwestern South America: A Quantitative Review. **Botanical Review.**, v. 77, p. 462–570, 2011.

MAYER, CHRISTOPH, Phobos 3.3.11, [http://www.rub.de/spezzoo/cm/cm\\_phobos.html](http://www.rub.de/spezzoo/cm/cm_phobos.html), 2006-2010.

MARISELVAM, R. et al. Green synthesis of silver nanoparticles from the extract of the inflorescence of *Cocos nucifera* (Family: Arecaceae) for enhanced antibacterial activity. *Spectrochimica Acta Part A: Molecular e Biomolecular Spectroscopy*, v. 129, p. 537–541, 2014.

MEEROW, A.W., NOBLICK, L., BORRONE, J.W., COUVREUR, T.L.P., MAURO-HERRERA, M., HAHN, W.J., KUHN, D.N., NAKAMURA, K., OLEAS, N.H., SCHNELL, R.J. Phylogenetic analysis of seven WRKY genes across the palm subtribe attaleinae (areceaceae) identifies *Syagrus* as sister group of the coconut. **PLoS One** 4, 10–13. doi:10.1371/journal.pone.0007353, 2009.

METZKER, M. L. Sequencing technologies – the next generation. **Nature Reviews-Genetics**. Houston, v. 11, p. 31-46, January, 2010.

MÔRO J, SILVA, MAS E GERALDO JS. Methodology for kariological study of brazilian palms. **Acta Hortic**. 486, 225-228 DOI: <http://dx.doi.org/10.17660/ActaHortic.1999.486.33>. 1999.

NOBLICK, L. R. **The indigenous palms of the state of Bahia, Brazil**. Tese de Doutorado. Chicago, University of Illinois. p. 52, 1991.

PAES, M.L.N.; DIAS, I.F.O. **Plano de Manejo da Estação Ecológica Raso da Catarina. Brasília**: IBAMA – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis/Diretoria de Ecossistemas/Coordenação Geral de Unidades de Conservação, p.326, 2008.

PINTAUD, J.C. et al. Las palmeras de América del Sur: diversidad, distribución e historia evolutiva. **Revista Peruana de Biología**, v. 15, p. 7-29, 2008.

POTTS, A. J., HEDDERSON, T. A., COWLING, R. M. Testing large-scale conservation corridors designed for patterns e processes: comparative phylogeography of three tree species. **Diversity and Distributions.**,v. 19, p. 1418-1428, 2013.

PROBER, J.M., et al. A system for rapid DNA sequencing with fluorescent chain terminating dideoxynucleotides. **Science** 238, 336–341, 1987.

ROCHA, K. M. R. **Biologia Reprodutiva da Palmeira Licuri (*Syagrus coronata*) (Mart.) Becc. (Arecaceae) na Ecorregião do Raso da Catarina**. 2009. 98 f. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal Rural de Pernambuco, Recife, PE, 2009.

SANGER, F. et al. DNA sequencing with chain-terminating inhibitors. **Proc. Natl. Acad. Sci. U. S. A.** 74, 5463–5467, 1977.

SANTOS, H. M. V.; SANTOS, V. de J. **Estudo etnobotânico do licuri *Syagrus coronata* (Martius) Beccari em Senhor do Bonfim, Bahia**. 2002.

Disponivelem:<http://projetolicuri.ubbihp.com.br/pages/resultados2.html>

SCARCELLI, N.; BARNAUD, A.; EISERHARDT, W.; TREIER, U.A.; SEVENO, M. et al., A Set of 100 Chloroplast DNA Primer Pairs to Study Population Genetics and Phylogeny in Monocotyledons. **PLoS ONE** 6(5): e19954. doi:10.1371/journal.pone.0019954, 2011.

SIMPSON, J.T., WONG, K., JACKMAN, S.D., SIMPSON, J.T., WONG, K., JACKMAN, S.D., SCHEIN, J.E., JONES, S.J.M.. ABySS : A parallel assembler for short read sequence data 1117–1123. doi:10.1101/gr.089532.108, 2009.

SMITH, L.M. et al. Fluorescence detection in automated DNA sequence analysis. **Nature** 321, 674–679, 1986.

SOUZA, V.C.; LORENZI, H.; Botânica sistemática: guia ilustrado para identificação das famílias de fanerógamas nativas e exóticas no Brasil, baseado no APG II. **Nova Odessa: Instituto Plantarum**, 2. ed., p.324, 2008.

TABARELLI, M.; VICENTE, A.; **Conhecimento sobre plantas lenhosas da Caatinga: lacunas geográficas e ecológicas**. In: SILVA, J.M.C.; TABARELLI, M.; FONSECA, M.T.; LINS, L.V. (Orgs.). Biodiversidade da Caatinga: áreas e ações prioritárias para a conservação. Brasília: MMA, p.101-111, 2004.

TEIXEIRA DA SILVA DE LA SALLES, K., MENEGHETTI, S. M. P., FERREIRA DE LA SALLES, W., MENEGHETTI, M. R., DOS SANTOS, I. C. F., DA SILVA, J. P. V., DE CARVALHO, S. H. V., SOLETTI, J. I. Characterization of *Syagrus coronata* (Mart.) Becc. oil and properties of methyl esters for use as biodiesel. **Industrial Crops e Products**., v. 32, p. 518-521, 2010.

XIONG, AI-S.PENG, RI-H.ZHUANG, J. GAO, F. ZHU, B. FU, XIAO-Y. XUE, Y. JIN, XIAO-F.TIAN, YONG-S. ZHAO, W. YAO, QUAN-H.. Gene duplication, transfer, and evolution in the chloroplast genome. **Biotechnology Advances**. Volume 27, p. 340–347, 2009.

ZAMBRANA, N.Y.P. et al. Diversity of palm uses in the western Amazon. **Biodiversity and Conservation**, v.16, p. 2771–2787, 2007.