

Variation in palm tree plastidial simple sequence repeats, characterization, and potential use

Tatieli Silveira¹, Karine Elise Janner de Freitas², Railson Schreinert dos Santos³, Antonio Costa de Oliveira^{2*} and Rosa Lía Barbieri³

Abstract: Palm trees are the third most important botanical family for humans because of their potential use in oils, drugs, cosmetics, food, and feed. Despite their importance, little information on their genetics and molecular variations exists, and a better understanding could contribute to breeding programs. This study aimed to determine the amount, distribution, and organization of plastid simple sequence repeats (cpSSRs) and their potential use in breeding in 52 species belonging to the order Arecales. Plastid genomes were analyzed to identify cpSSRs according to their nature, position, and presence in genic or intergenic regions. Primer pairs were validated *in silico* for amplification and polymorphisms in these SSRs and their dissimilarities were evaluated. The results showed a high frequency of mononucleotide repeats in the intergenic regions. Approximately 76 primer pairs were generated and are suggested for further studies. The dissimilarity analysis of cpSSRs showed that mono- and trinucleotides were highly abundant in plastid SSRs.

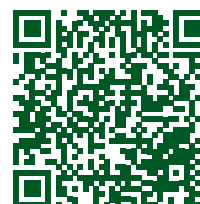
Keywords: *In silico*, chloroplast DNA, markers, primers

INTRODUCTION

Arecaceae includes most palm trees and represents the third most important botanical family for humans (Johnson 1998). This family is part of an order of flowering plants called Arecales. This order comprises the Arecaceae and, more recently, the family Dasypogonaceae (The Angiosperm Phylogeny Group 2016). For Arecales, 192 genera and 2,603 species have been described (Missouri Botanic Garden 2021, NCBI 2021).


Among its species, most are economically important, given their potential for use in human and animal food, and production of inputs, oils, drugs, cosmetics, and decorative utensils, as well as for the ornamentation of gardens and public roads. Some species, such as *Cocos nucifera* (coconut), *Phoenix dactylifera* (date palm), and *Elaeis guineensis* (oil palm), deserve special attention because of their local economic importance and high potential for use. Coconut cultivation reached an area of 12.3 million hectares in 2017, with production of 60.7 million tons of coconut fruits in 92 countries. The Philippines, Indonesia, and India account for 72.7% of the total area. Brazil has a planted area of 215,683 hectares and produces approximately 2.4 million tons of fruit per year (FAO 2017). The Northeast region is responsible for 74% of national production and

Crop Breeding and Applied Biotechnology
22(4): e41812241, 2022
Brazilian Society of Plant Breeding.
Printed in Brazil
<http://dx.doi.org/10.1590/1984-70332022v22n4a36>



*Corresponding author:

E-mail: acostol@terra.com.br

 ORCID: 0000-0001-8835-8071

Received: 12 April 2022

Accepted: 26 September 2022

Published: 14 October 2022

¹ Embrapa Clima Temperado, Estação Experimental Terras Baixas, Avenida Eliseu Maciel, s/n, Campus Universitário, 96010-971, Capão do Leão, RS, Brazil

² Universidade Federal de Pelotas, Rua Gomes Cerneiro, 1, Centro, 96010-610, Pelotas, RS, Brazil

³ Embrapa Clima Temperado, Rodovia BR 392, km 78, 9º Distrito, Monte Bonito, 96010-971, Pelotas, RS, Brazil

is concentrated in the states of Bahia, Sergipe, Rio Grande do Norte, and Pernambuco (IBGE 2015). Despite the progress achieved with coconuts, poor understanding of the genetic features of palm species slows down their economic spread, and their full potential is yet to be exploited. In addition, genome sequences and physiological data remain scarce, as seen in *Butia* spp. (Nazareno et al. 2011).

Simple sequence repeats (SSRs), also called microsatellites, are highly informative because of their size variation and are widely used for breeding and genetic studies in many plant species (Weber 1990, Kashi and King 2006, Palliyarakkal et al. 2012). The development of molecular markers for palm species has expanded after the complete sequencing of the nuclear genomes of the oil palm (Singh et al. 2013) and coconut (Aljohi et al. 2016). Currently, SSR markers are used in Arecales to study ecology, systematics, conservation, and phylogeny, and in breeding programs (Elshibli and Korpelainen 2008, Aljohi et al. 2016, Zhao et al. 2017, Xiao et al. 2017, Bai et al. 2017, Khan et al. 2018, Babu et al. 2019, Bhagya et al. 2020, Kpatènou et al. 2020). Some nuclear SSRs and plastid SSRs (cpSSRs) have been developed for the most important representative species of this order, such as *P. dactylifera*, *C. nucifera*, *Calamus simplicifolius*, *E. oleifera*, and *E. guineensis*.

The cpSSRs have great potential for use in phylogeography and DNA fingerprint analyses of Arecaceae (Lopes et al. 2018, Lopes et al. 2019) and many researchers have sought to use hypervariable regions of chloroplast or plastidial DNA (cpDNA) to characterize the genetic variation in species (Shaw et al. 2005). Therefore, the aim of this study was to detect the amount, distribution, and organization of cpSSRs in 52 species of the order Arecales, and to identify the SSRs that contribute the most to the differences found among the plastidial genomes.

MATERIAL AND METHODS

All available plastid genomes from Arecales were downloaded from the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=plastid&taxid=2759>). We assessed the SSRs for coding (genic) and non-coding (intergenic) regions according to the information available in the NCBI database for each species. A total of 52 complete cpDNAs and their coding regions (Supplementary Material 1) were inserted in *SSRlocator* (Maia et al. 2008). The minimum repeat size was set to ≥ 8 -mono, ≥ 6 -di, ≥ 3 -tri, tetra, penta, and hexanucleotides (hexamers).

Primer pairs for sequences flanking each SSR were also designed using *SSRLocator* according to the following parameters: GC content 40-60%, melting temperature (T_m) range 57-60 °C, primer size range 0.022-0.027 kb, and Polymerase chain reaction (PCR) product size range 101-300 bp, as suggested by Preethi et al. (2020). Common primers for some species were validated in silico using the Blastn resource available on the NCBI website. For the SSR comparative analysis, a heat map based on dissimilarity was generated using CIMminer (<https://discover.nci.nih.gov/cimminer/>). Organellar Genome Draw (OGDRAW) was used to represent the SSR annotation (Lohse et al. 2007). A phylogeny based on RuBisCo sequences was built and the RuBisCo sequence of each species of the order Arecales was extracted and aligned with ClustalW. Based on the identification of the best nucleotide substitution model, a phylogenetic tree was constructed based on the Bayesian model using 1 million bootstrap replicates. Principal component analysis was used to highlight the vector component that contributed the most to the differentiation of the phylogenetically defined groups. For this analysis, we used the R statistical software v. 4.1.3 (R Core Team 2020).

RESULTS AND DISCUSSION

The cpSSRs from the 52 Arecales species reported in the NCBI database are listed in Supplementary Material 1. The highest number of cpSSRs was observed in *Wallichia densiflora* (302), whereas the lowest was found in *Dasyopogon bromeliifolius* (200). The largest variation in cpSSR size was found between *Tahina spectabilis* (126,251 bp) (Barrett et al. 2016) and *Caryota obtusa* (159,882 bp) (Gao et al. 2020).

Although we identified mono-, di-, tri-, tetra-, penta-, and hexamers in the majority of Arecales species (Supplementary Material 2), monomers exhibited the highest abundance. Currently, no studies are available that evaluate the distribution and organization of cpSSRs across all Arecales; however, Palliyarakkal et al. (2012), Al-Faifi et al. (2017), Lopes et al. (2018), and Khan et al. (2018) identified a high number of monomers in intergenic regions of palm trees (for the species *Acrocomia aculeata*, *Phoenix dactylifera*, and other species). However, monomers are rarely used as markers in palms,

whereas dimers, trimers, tetramers, pentamers, and hexamers are more useful (Preethi et al. 2020). Monomers may deteriorate over short periods of time as they have a high replacement rate (Ceplitis et al. 2005). Previous studies on cpDNAs from palm species (Magnabosco et al. 2020, Zou et al. 2021) have laid a foundation to perform genetic studies using cpSSRs to analyze populations of these plants.

Most of the SSRs found were monomers in the intergenic regions. Tri and tetramers were also observed, which were frequent in both genic and intergenic regions (Supplementary Material 2). However, 18 species of the family Arecaceae and one of Dasypogonaceae presented a high abundance of trimer repeats in genic regions, which presents immense potential for the development of molecular markers.

Primer pairs for the amplification of all SSRs from the 52 analyzed palm species were designed and are available online (Supplementary Material 3). In species of high economic interest, such as *Cocos nucifera*, *Elaeis guineensis*, *Phoenix canariensis*, and *Syagrus coronata*, 74, 83, 75, and 79 cpSSRs, respectively, were detected that could be amplified, whereas in species of the family Dasypogonaceae, such as *Bacteria australis* and *Dasypogon bromeliifolius*, 66 and 75 cpSSR, respectively, were detected. Primers were developed based on previous reports (Preethi et al. 2020), for *Cocos nucifera*. The melting temperature ranged from 57 °C to 60 °C, primer GC content ranged from 40% to 60%, primer size ranged from 22 bp to 27 bp, and product size ranged from 101 bp to 300 bp.

The palm species *Phoenix dactylifera* was reported to have 93 polymorphic nuclear SSRs (Al-Faifi et al. 2017). However, with the parameters used in this study, *Phoenix dactylifera* presented 76 cpSSRs. Markers based on cpSSRs are more effective indicators of population subdivision and differentiation than nuclear markers in plants, even with taxonomic variations, and are generally abundant in most plants (Powell et al. 1995a, Powell et al. 1995b, Provan et al. 2001, Petit et al. 2005, Ebert and Peakall 2009), as observed in this study. We provide a list of primers validated in silico and common to some species that can be used in transfer studies between palm species (Supplementary Material 4). Despite the observation of conserved plastid genomes between sister species (Ebert and Peakall 2009), there is some divergence between phylogenetically close genomes that does not allow great sharing of microsatellite primers for transfer studies. This was also observed in our dissimilarity analysis (Figure 1), which indicated divergence between closely related species. The application of cpSSR markers is especially useful because they can amplify homologous regions in several taxa owing to their uniparental inheritance and slow accumulation of mutations through the ages, which makes evolutionary studies easy and effective. The regions that are proposed in this study can be used to explore species with fewer studies, such as species of *Butia* (Mistura et al. 2012) or species of *Areca*, *Arenga*, *Astrocaryum*, *Brahea*, and *Phoenix* (Supplementary Material 4). However, a high divergence could be observed when evaluating the molecular evolution of plastids within the Arecaceae family using every gene that encodes plastid proteins. This suggests that the degeneration process may occur within Arecaceae at the genus or species level (Lopes et al. 2018). The distribution of positive signatures across the phylogenomics of Arecaceae suggests convergent evolution at most sites, including genes involved in photosynthesis. Therefore, researchers have sought to use non-coding regions, including introns and intergenic spacers, to characterize genetic variation (Shaw et al. 2005, Shaw et al. 2007). However, the use of SSRs as markers can aid in the discovery of genetic and molecular features that have not been extensively explored in Arecaceae.

In the *Cocos nucifera* plastid genome map, the most frequent cpSSRs in the 52 species studied in this study were *trnH-psbA*, *matK*, *rcbL*, and *rps19* (Supplementary Materials 5 and 6). The *rcbL* and *matK* genes are used and standardized for DNA barcoding in terrestrial plants (CBOL Plant Working Group 2009, Hollingsworth et al. 2011, Magnabosco 2020) given their high levels of high-quality sequences and acceptable levels of species differentiation and identification (Burgess et al. 2011).

Dissimilarity analysis of cpSSRs in palms (Figure 1) showed highly dissimilar mono- and trimers, reflecting the great abundance of these SSR types. In addition, we observed that the 52 species evaluated in this study formed two groups based on SSR frequency. The first group included *Dasypogon bromeliifolius* and *Bacteria australis*, both of which belong to Dasypogonaceae (Liliopsida), whereas the second was a large group, subdivided into four subgroups, all within Magnoliophyta. The genetic relationship of the second group has been described previously (Meerow et al. 2009) when analyzing seven WRKY genes between *Cocos nucifera* and *Syagrus coronata*. Nevertheless, we observed the proximity between *Astrocaryum* sp. and *Acrocomia* sp., already reported as sibling genera, when evaluating six WRKY genes (Meerow et al. 2015).

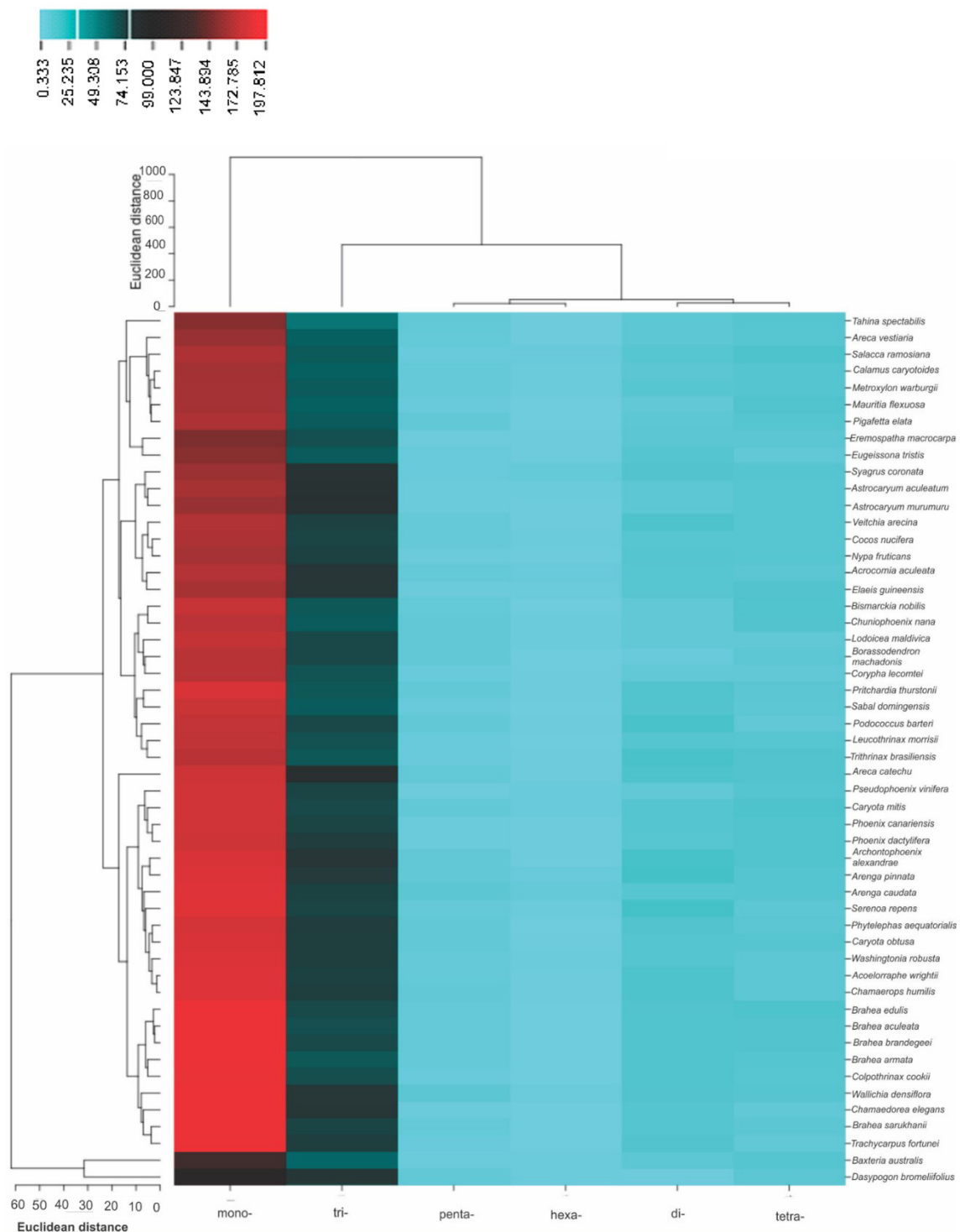


Figure 1. Dissimilarity analysis of 52 species of the order Arecales based on plastid simple sequence repeats (cpSSRs).

A number of previously described phylogenetically close species are shown in Figure 1, where the clusters were divided by the families analyzed. Furthermore, the highest frequency of monomers was found in the order Arecales. Therefore, studies on the development of cpSSR markers in these species would provide a better understanding of the genetic relationships within this order (Zhang et al. 2017).

We detected 14,017 SSRs across all the analyzed palm species. A total of 9,087 SSRs consisted of monomers, and 502, 3,676, 483, 204, and 65 consisted of dimers, trimers, tetramers, pentamers, and hexamers, respectively. The most abundant repetitive motifs found in the 52 species studied were AT and TTC/AAC/TAT for the di- and trimers, respectively. A total of 38,086 SSRs were found in the Arecaceae family in public domains (Palliyarakkal et al. 2012), whereas 1,563 sequences were found in *Syagrus romanzoffiana* alone (Laindorf et al. 2019). The most frequent motifs in these studies were di- and trimers. In contrast, our study showed a higher number of SSRs comprising monomers. The discrepancy in the values found in different studies can be attributed to the type of data filtering used, as well as the increase in the amount of information available in databases, or even changes in the pattern of the strings (Laindorf et al. 2019).

The evolution of SSRs has occurred under different pressures, even at the level of closely related species. This was evident when we compared the dissimilarity analysis based on the number of SSRs and phylogeny based on the conserved RuBisCo gene (Supplementary Material 7). A difference was observed in the positioning of species of specific groups between the two analyses, in which we can highlight that, through phylogeny, all species of the same genus were side by side, as was the case with *Arenga* sp., *Astrocaryum* sp., and *Areca* sp. It is worth noting that the proximity of *Cocos nucifera* and *Syagrus coronata* has been widely discussed in the literature and used in several studies to infer phylogenies with representatives of the Arecaceae family (Meerow et al. 2009, Meerow et al. 2015). Thus, the distribution of SSRs can show differences, even between closely related species, as the positioning of the species observed in the dissimilarity does not follow the same organization as that of the phylogenetic analysis.

Principal component analysis (Figure 2) showed that the motifs that most affect the grouping of species owing to both their presence × absence and intron × exon positioning are TAT3, GAA3, AAG3, and AGA3, which occur in three or

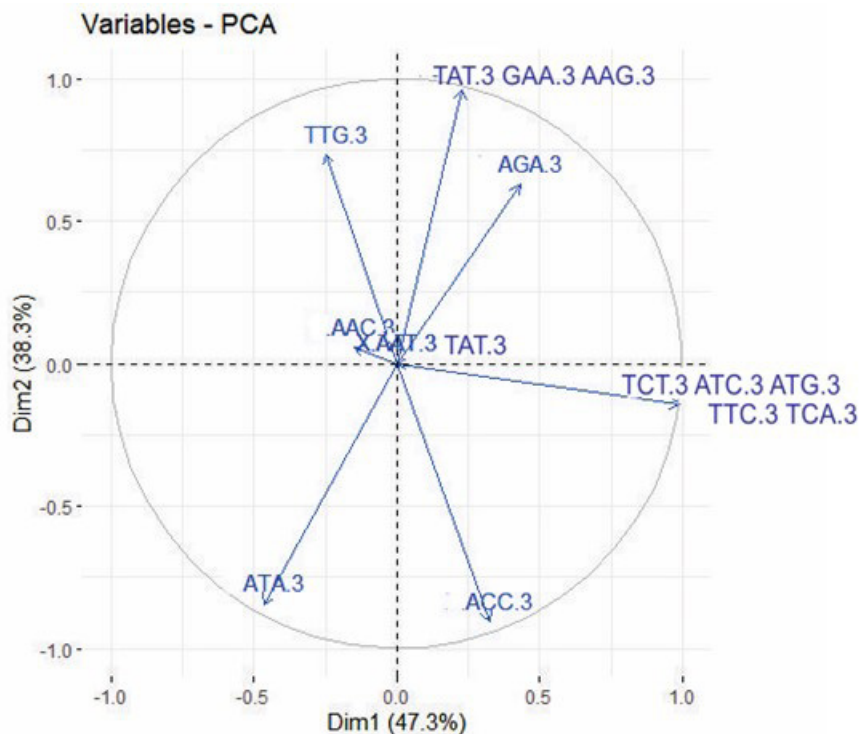


Figure 2. Ordering diagram of the principal component analysis (PCA) performed with the five groups from the dissimilarity analysis of 52 species from the order Arecales based on plastid simple sequence repeats (cpSSRs).

four groups formed in the dissimilarity analysis, showing persistence and stability. These data can be useful for diversity analysis because of their capacity to differentiate between closely related species and possibly even populations of the same species. Motif ATA3, however, exhibited nonlinear behavior and was not selected for further tests.

CONCLUSION

In this study, a high frequency of mononucleotide repeats in the intergenic regions of chloroplast genomes was observed, and 76 pairs of primers that could be used in future studies are reported. Through cpSSR dissimilarity analyses, mono- and trinucleotides were found to be highly different from each other and abundant in the plastids. In the phylogeny analysis, comparison of the Rubisco SSRs of the 52 species studied, presented a distribution pattern of the species that was very different from that found in the dissimilarity analysis. Furthermore, we identified that the SSRs that contributed the most to the grouping of species were TAT3, GAA3, AAG3, and AGA3. In summary, studies of this nature are important because they can provide direction to phylogenies and help in the conservation of germplasm and genetic improvement of palm trees.

ACKNOWLEDGMENTS

The authors are thankful to Editage for the English revision. Supplementary files are available upon request to the corresponding author (acostol@terra.com.br).

REFERENCES

- Al-Faifi SA, Migdadi HM, Algamdi SS, Khan MA, Al-Obeed RS, Ammar MH and Jakse J (2017) Development of genomic simple sequence repeats (SSR) by enrichment libraries in date palm. **Methods in Molecular Biology** **1638**: 283-313.
- Aljohi HA, Liu W, Lin Q, Zhao Y, Zeng J, Alamer A, Alanazi IO, Alawad AO, Al-Sadi AM, Hu S and Yu J (2016) Complete sequence and analysis of coconut palm (*Cocos nucifera*) mitochondrial genome. **PLoS One** **11**: 1-18.
- Babu KB, Mary Rani KL, Sarika Sahu, Mathur RK, Naveen Kumar P, Ravichandran G, Anitha P and Bhagya HP (2019) Development and validation of whole genome-wide and genic microsatellite markers in oil palm (*Elaeis guineensis* Jacq.): First microsatellite database (OpSatdb). **Scientific Reports** **9**: 1-8.
- Bai B, Le W, Lee M, Zhang Y, Rahmadsyah YA, Ye BQ, Zi YW, Lim CH, Suwanto A, Chua NH and Yue GH (2017) Genome-wide identification of markers for selecting higher oil content in oil palm. **BMC Plant Biology** **17**: 1-11.
- Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack J, Li J, Lim GS, Mayfield-Jones DR, Perez L, Medina J, Pires JC, Santos C, Stevenson DWM, Zomlefer DB and Davis JJ (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. **New Phytologist** **209**: 855-870.
- Bhagya HP, Babu BK, Gangadharappa PM, Naika MBN, Satish D and Mathur RK (2020) Identification of QTLs in oil palm (*Elaeis guineensis* Jacq.) using SSR markers through association mapping. **Journal of Genetics** **99**: 1-10.
- Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei M and Barrett SCH (2011) Discriminating plant species in a local temperate flora using the *rbcL* plus *matK* DNA barcode. **Methods of Ecology and Evolution** **2**: 333-340.
- CBOL Plant Working Group (2009) A DNA barcode for land plants. **Proceedings of the National Academy of Sciences of the USA** **106**: 12794-12797.
- Ceplitis A, Su Y and Lascoux M (2005) Bayesian inference of evolutionary history from chloroplast microsatellites in the cosmopolitan weed *Capsella bursa pastoris* (Brassicaceae). **Molecular Ecology** **14**: 4221-4233.
- Ebert D and Peakall R (2009) Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. **Molecular Ecology Resources** **9**: 673-690.
- Eshibli S and Korpelainen H (2008) Microsatellite markers reveal high genetic diversity in date palm (*Phoenix dactylifera* L.) germplasm from Sudan. **Genetica** **134**: 251-260.
- FAO - Food and Agriculture Organization of the United Nations (2017) Available at <<http://fenix.fao.org/faostat/beta/en/#data/QC>>. Accessed on 20 October, 2021.
- Gao Y, Lv M, Cui T and Wan X (2020) The complete chloroplast genome of *Caryota obtusa*, an endangered and economically important species. **Mitochondrial DNA Part B Resources** **5**: 2176-2177.
- Hollingsworth PM, Graham SW and Little DP (2011) Choosing and using a plant DNA barcode. **PLoS One** **6**: 1-13.
- IBGE - Instituto Brasileiro de Geografia e Estatística (2015) Produção e área plantada de lavouras permanentes. Sistema IBGE de recuperação automática. Available at <<http://www.sidra.ibge.gov.br/bda/tabela/protabl.asp?c=1613&z=t&o=11&i=P>>. Accessed on 20 October, 2021.
- Johnson DV (1998) Non-wood forest products 10: Tropical palms. [S.l.].

Variation in palm tree plastidial simple sequence repeats, characterization, and potential use

- Food and Agriculture Organization of the United Nations (FAO). Available at <<http://www.fao.org/docrep/x0451e/x0451e00.HTM>>. Accessed on 20 April, 2021.
- Kashi Y and King DG (2006) Simple sequence repeats as advantageous mutators in evolution. **Trends in Genetics** **22**: 253-259.
- Khan AL, Asaf S, Lee IJ, Al-Harrasi A and Al- Rawahi A (2018) First chloroplast genomics study of *Phoenix dactylifera* (var. Naghal and Khanezi): A comparative analysis. **Plos One** **13**: 1-20.
- Kpatènón MJ, Salako KV, Santoni S, Zakraoui L, Latreille M, Tollon-Cordet C, Mariac C, Jaligot E, Beulé T and Adéoti K (2020) Transferability, development of simple sequence repeat (SSR) markers and application to the analysis of genetic diversity and population structure of the African fan palm (*Borassus aethiopum* Mart.) in Benin. **BMC Genetics** **21**: 1-23.
- Laindorf BL, Metz GF, Kuster MCT, Lucini F, Freitas KEJ, Victoria FC and Pereira AB (2019) Transfer of microsatellite markers from other *Arecaceae* species to *Syagrus romanzoffiana* (*Arecaceae*). **Genetics and Molecular Research** **18**: 1-10.
- Lopes AS, Pacheco TG, Nimz T, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO and Rogalski M (2018) The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in *Arecaceae*. **Planta** **247**: 1011-1030.
- Lopes AS, Pacheco TG, Silva ON, Cruz LM, Balsanelli E, Souza EM, Pedrosa FO and Rogalski M (2019) The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. **Planta** **250**: 1229-1246.
- Lohse M, Drechsel O and Bock R (2007) Organellar genome DRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. **Current Genetics** **5-6**: 267-274.
- Maia LC, Palmieri DA, Souza VQ, Kopp MM, Carvalho FIF and Oliveira AC (2008) SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. **International Journal of Plant Genomics** **2008**: 412696.
- Magnabosco JWS, Fraga HPF, Silva RS, Rogalski M, Souza EM, Guerra MP and Vieira LN (2020) Characterization of the complete plastid genome of *Butia eriospatha* (*Arecaceae*). **Genetics and Molecular Biology** **43**: 1-5.
- Meerow AW, Noblick L, Borrone JW, Couvreur TLP, Mauro-Herrera M, Hahn WJ, Kuhn DN, Nakamura K, Oleas NH and Schnell RJ (2009) Correction: Phylogenetic analysis of seven WRKY genes across the palm subtribe Attaleinae (*Arecaceae*) identifies *Syagrus* as sister group of the coconut. **Plos One** **4**: 1-17.
- Meerow AW, Noblick L, Salas-Leiva DE, Sanchez V, Francisco-Ortega J, Jestrow B and Nakamura K (2015) Phylogeny and historical biogeography of the cocosoid palms (*Arecaceae*, *Arecoideae*, *Cocoseae*) inferred from sequences of six WRKY gene family loci. **Cladistics** **31**: 509-534.
- Missouri Botanic Garden (2021) Main Tree. *Arecales*. Available at <<http://www.mobot.org/MOBOT/Research/APweb/treeapweb2map.html>>. Accessed 20 April, 2021.
- Mistura CC, Barbieri RL, Castro CM and Priori B (2012) Transferibilidade de marcadores microsatélites de coco (*Cocos nucifera*) para butiá (*Butia odorata*). **Embrapa Clima Temperado** **24**: 360-369.
- Nazareno AG, Zucchi MI and Reis MS (2011) Microsatellite markers for *Butia eriospatha* (*Arecaceae*), a vulnerable palm species from the atlantic rainforest of Brazil. **American Journal of Botany**: 198-200.
- NCBI - National Center for Biotechnology Information (2021) *Arecales*. Available at <<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>>. Accessed on 25 April, 2021.
- Palliyarakkal MK, Ramaswamy M and Vadivel A (2012) Microsatellites in palm (*Arecaceae*) sequences. **Bioinformation** **7**: 47-51.
- Petit RJ, Dumnil J and Fineschi S (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. **Molecular Ecology** **14**: 689-701.
- Powell W, Morgante M and Andre C (1995a) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. **Current Biology** **5**: 1023-1029.
- Powell W, Morgante M, McDevitt R, Vendramin GG and Rafalski JA (1995b) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. **Proceedings of the National Academy of Sciences of the USA** **92**: 7759-7763.
- Preethi P, Rahman S, Naganeeswaran S, Sabana1 AA, Gangaraj KP, Jerard BA, Niral1 V and Rajesh MK (2020) Development of EST-SSR markers for genetic diversity analysis in coconut (*Cocos nucifera* L.). **Molecular Biology Reports** **47**: 9385-9397.
- Provan J, Powell W and Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. **Trends in Ecology & Evolution** **16**: 142-147.
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <<http://www.r-project.org/>>. Accessed on 13 September, 2021.
- Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE and Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. **American Journal of Botany** **92**: 142-166.
- Shaw J, Lickey EB, Schilling EE and Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. **American Journal of Botany** **94**: 275-288.
- Singh R, Ong-Abdullah M, Low ET, Manaf MA, Rosli R and Nookiah R (2013) Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. **Nature** **500**: 335-339.
- The Angiosperm Phylogeny Group (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering

- plants: APG IV. **Botanical Journal of the Linnean Society** **181**: 1-20.
- Weber JL (1990) Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. **Genomics** **7**: 524-530.
- Xiao Y, Xu P, Fan H, Baudouin L, Xia W and Bocs S (2017) The genome draft of coconut (*Cocos nucifera*). **Gigascience** **6**: 1-11.
- Zhang H, Hall N, McElroy JS, Lowe EK and Goertzen LR (2017) Complete plastid genome sequence of goosegrass (*Eleusine indica*) and comparison with other Poaceae. **Gene** **600**: 36-43.
- Zhao Y, Keremane M, Prakash CS and He G (2017) Characterization and amplification of gene-based simple sequence repeat (SSR) markers in date palm. **Methods in Molecular Biology** **1638**: 259-271.
- Zou B, Long W and Wu LHY (2021) The complete plastid genome of *Phoenix canariensis* Chabaud (Arecaceae) and phylogenetic analysis. **Mitochondrial DNA Part B Resources** **6**: 140-142.