JUKES-CANTOR EVOLUTIONARY MODEL BASED PHYLOGENETIC RELATIONSHIP OF ECONOMICALLY IMPORTANT ORNAMENTAL PALMS USING MAXIMUM LIKELIHOOD APPROACH

IJAZ, S.[1*] – HAQ, I. U.[2] – BABAR, M.[1]

[1]Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, University Road, Faisalabad, Pakistan

[2]Department of Plant Pathology, University of Agriculture, University Road, Faisalabad, Pakistan

*Corresponding author
e-mail: siddraijazkhan@yahoo.com

Abstract. Ornamental palms are famous for their patterned leaves, which make them quite suitable for indoor and outdoor decors. Hitherto, palms were identified morphologically resulting in ambiguous identification. Hence, to stitch this loophole in phylogenetic studies, DNA barcoding is a remarkable way for correct identification and evolutionary study of ornamental palms. We investigated phylogenetic relationship of commonly used ornamental palms in Pakistan. We focused on two genetic regions, rbcL and matK, which had been proved as potential DNA barcodes for species identification and phylogenetic studies. In this study, fifteen ornamental palms were investigated for their evolutionary relationship and molecular identification. This study resolved their evolutionary history using Maximum Likelihood method based on Jukes-Cantor evolutionary model. Among them, nine ornamental palm shared evolutionary history, though four showed evolutionary distinctness from the rest. *Chamaedorea seifrizii* and *Chamaedorea costaricana* revealed close residence in the evolutionary tree with 70% bootstrap support. *Syagrus romanzoffiana* resolved as independent lineage with distinct and highly diverse evolutionary history. Furthermore, *Wodyetia bifurcata* and *Dypsis lutescens* nested together with 61% bootstrap though, they belong to different genera. This showed that the genus Wodyetia and Dypsis had very close evolutionary history. The evolutionary forces either geographical isolation or genetic drift could be the reason for the divergence of these palms into two distinct genera, though being the lineages of the same parent.
Keywords: *ornamental palm, DNA barcoding, Bayesian information criterion, maximum likelihood method, Jukes-Cantor evolutionary model*

Introduction

Arecaceae family, also known as Palmae family, is well known for its ornamental palm members. It consists of 240 genera and 27000 species (Lorenzi et al., 2010). It is a monotypic family of Arecales order. Palms are thought to be originated in Australasia and their fossil records reveal their distribution in early Cenzoic era. In present era, palms are mostly distributed in tropical regions and some species are reported in subtropical regions as well (Henderson et al., 1995). Palmae family comprised of commercially and horticulturally famous landscape ornamental palms which are famous for their palmate patterned evergreen leaves and bright, colorful flowers. Palms are noticeably used in outdoor and indoor house decors. Since many decades, the taxonomy of these ornamental palms is poorly understood due to lack of information about their character identification. These problems are caused because of their morphological

Ijaz et al.: Jukes-Cantor evolutionary model based phylogenetic relationship of economically important ornamental palms using
maximum likelihood approach
- 14860 -

similarities with each other and difficulties in hybrid discrimination (Pintaud, 2008). High evolutionary rate has been observed in palm trees and the evaluation of genetic evolution through botanical characters is quite difficult task as morphological traits distinguish the plant at specific life stage. The discrimination based upon such characters results in either ambiguous or wrong identification of palm (Sedra et al., 1996) which made them pseudo palms.

To overcome these problems, botanists have moved towards molecular approach for species identification and discrimination. For this purpose, DNA barcodes are used suggested by CBOL (Consortium of Barcode of Life) in 2009. DNA barcodes consists of conserved short DNA sequences which are unique for every species. In case of animals, well known barcode in COI (Cytochrome oxidase gene) while for plants, many plastid sequences have been studied and confirmed as barcode loci (Kress et al., 2005; Ford et al., 2009). Out of many barcodes, two plastid genes namely maturase K (*matK*) and Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) have been recommended and standardized as potential plant barcodes by PWG (International body of Plant Working Group) of CBOL (2009). The barcodes suggested by CBOL group have great potential for discrimination, authentication and identification for international/foreign as well as native plant species (Maloukh et al., 2017). The discriminatory power of *rbcL* is lower than *matK* but it offers many advantages like its unambiguous and clear alignment power, high sequencing frequency and its universality (Dong et al., 2014). In this study, 15 ornamental palms were investigated for their phylogenetic relationship to unravel their evolutionary history and ancestral inference for assessing their divergence as distinct lineages.

Materials and methods

Sample collection and molecular processing

The leave samples of 15 economically important ornamental palms in Pakistan (*Table 1*) were collected for DNA extraction. Genomic DNA was extracted using Plant DNA extraction kit (Favorgen Biotech Corp.). For the amplification of *matK* and *rbcL* regions, their universal primer pairs suggested by Consortium of Barcode of Life (CBOL, 2009) were used in PCR analysis. The total volume of a reaction mixture for PCR was 25 µl containing 50 ng/µl DNA template, 10 µM each primer in a pair, 2 µl dNTPs, 2.5 µl of HF buffer, 0.25 µl of high fidelity polymerase and nuclease free water for volume makeup. PCR product was resolved on 2% high resolution agarose with gel electrophoresis. The required amplicons were eluted by gel purification mini kit (Favorgen Biotech Corp.) and then sequenced through Eurofins Genomics DNA sequencing services, USA. The sequences were trimmed with BioEdit v7.2.6.1. The success frequency of PCR amplification of each target regions (rbcl and matK) was determined as the proportion of total No. of amplified individuals to the total No. of individuals analyzed in study while successful sequencing was evaluated as the percentage of number of good quality sequences to the total number of sequences of studied individuals.

Data analysis

The good quality trimmed sequences were *in silico* characterized and submitted to NCBI database. The homology search of each sequence was performed using BLASTn

Ijaz et al.: Jukes-Cantor evolutionary model based phylogenetic relationship of economically important ornamental palms using
maximum likelihood approach
- 14861 -

tool. Each sequence was identified based on the 100% identity with the reference sequences of type materials in the NCBI database. The obtained sequences were aligned using ClustalW and then were submitted to NCBI database. For phylogenetic analysis, version 7.0 (Molecular Evolutionary Genetics Analysis) was used to employing Maximum Likelihood (ML) method based on best-fit evolutionary model to generate optimal topology in an evolutionary tree.

Table 1. *Ornamental palm with their native and scientific name along with GenBank accession numbers*

Serial#	Common name	Scientific name	GenBank accession no.	
			rbcL	*matK*
1	Alexandra palm	*Archontophoenix alexandrae*	MK947439	------------
2	Bamboo palm	*Chamaedorea seifrizii*	MK639602	MK639603
3	Bismarckia palm	*Bismarckia nobilis*	MK947440	------------
4	Bottle palm	*Hyophorbe lagenicaulis*	MK947437	------------
5	Cluster palm	*Chamaedorea costaricana*	MK947436	MN047253
6	Clustering fishtail palm	*Caryota mitis*	MK947435	------------
7	Foxtail palm	*Wodyetia bifurcata*	MK947442	------------
8	Golden palm	*Dypsis lutescens*	MK947441	------------
9	Lady palm	*Rhapis excelsa*	MK947443	------------
10	Queen palm	*Syagrus romanzoffiana*	MK650287	MN047252
11	Dwarf Majesty palm	*Ravenea hildebrandtii*	MK947438	MN047254
12	Royal palm	*Roystonea regia*	MK947434	------------
13	Washingtonia palm	*Washingtonia robusta*	MK940273	------------
14	Dioon palm	*Dioon spinulosum*	------------	------------
15	Kangi palm	*Cycas revolute*	------------	------------

Results

Ornamental palms are the pillars of horticultural industry. However, their accurate taxonomic identification is challenging. Hence, in this study, *matK* and *rbcL* regions were targeted to taxonomic investigation of 15 economically important ornamental palm in Pakistan for resolving their phylogenetic topology unequivocally. The universal barcode primers of both genetic regions were used in DNA barcoding analysis.

These two barcode genes *rbcL* and *matK* showed 100% PCR amplification. The sequencing frequency showed that the sequencing success of *rbcL* was comparatively higher than *matK*. The low sequencing rate of *matK* is due to high heterozygosity in the respective regions, which may result in polymerase slippage during Sanger sequencing method. The sequencing rate of *rbcL* was recorded as 86.66% while for *matK* it was recorded as 26.66%. However, the species discriminating power of *matK* is much greater than *rbcL*. The phylogenetic tree for rbcl region of 13 ornamental palm was constructed through Mega 6 software (Tamura et al., 2013) by using maximum likelihood algorithm, which was based upon Jukes-Cantor evolutionary model with Gamma distribution (JC + G) 1 (*Fig. 1*). The best substitution pattern is described the models with lowest Bayesian Information Criterion scores. Akaike Information Criterion (AICc), No. of parameters along with branch lengths and Maximum Likelihood value (lnL) along with estimated values of transversion bias (R), rate of nucleotide substitution (r) and nucleotide frequencies for each nucleotide pair is shown (*Table 2*).

The constructed tree showed highest log likelihood of -387.1215 with trees percentage is showed next to the branches. Through neighbor joining method, the pairwise distance matrix was assessed by MCL (Maximum Composite Likelihood) approach. The level of evolutionary differences was estimated by discrete Gamma distribution as + G 0.1000 parameter. The branch lengths of the constructed tree were calculated with the No. of substitutions per sites. The analyses were based upon 13 nucleotide sequences with 1st, 2nd, 3rd and noncoding codon positions. A total of 225 positions were included in final dataset with all missing gaps eliminated (*Table 2*).

Table 2. *Maximum likelihood fits of 24 different nucleotide substitution models on 13 ornamental palms using rbcL barcode*

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
JC+G	24	965.423	822.292	-386.939	n/a	0.05	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G+I	25	969.110	820.032	-384.792	0.46	0.05	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC	23	969.704	832.520	-393.070	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2+I	25	969.768	820.690	-385.121	0.84	n/a	1.20	0.250	0.250	0.250	0.250	0.057	0.057	0.136	0.057	0.136	0.057	0.057	0.136	0.057	0.136	0.057	0.057
JC+I	24	970.073	826.941	-389.264	0.62	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2+G	25	971.320	822.242	-385.897	n/a	0.05	1.19	0.250	0.250	0.250	0.250	0.057	0.057	0.136	0.057	0.136	0.057	0.057	0.136	0.057	0.136	0.057	0.057
T92+G	26	975.573	820.550	-384.033	n/a	0.05	1.19	0.290	0.290	0.210	0.210	0.065	0.047	0.116	0.065	0.116	0.047	0.065	0.160	0.047	0.160	0.065	0.047
K2	24	975.685	832.554	-392.070	n/a	n/a	1.18	0.250	0.250	0.250	0.250	0.057	0.057	0.135	0.057	0.135	0.057	0.057	0.135	0.057	0.135	0.057	0.057
K2+G+I	26	978.378	823.355	-385.435	0.12	0.05	1.20	0.250	0.250	0.250	0.250	0.057	0.057	0.136	0.057	0.136	0.057	0.057	0.136	0.057	0.136	0.057	0.057
T92	25	979.778	830.700	-390.126	n/a	n/a	1.18	0.290	0.290	0.210	0.210	0.066	0.048	0.115	0.066	0.115	0.048	0.066	0.159	0.048	0.159	0.066	0.048
T92+G+I	27	980.977	820.011	-382.744	0.29	0.05	1.20	0.290	0.290	0.210	0.210	0.065	0.047	0.116	0.065	0.116	0.047	0.065	0.160	0.047	0.160	0.065	0.047
HKY+I	28	982.633	815.724	-379.582	0.95	n/a	1.21	0.255	0.324	0.193	0.228	0.072	0.043	0.127	0.057	0.107	0.051	0.057	0.180	0.051	0.142	0.072	0.043
T92+I	26	983.797	828.774	-388.145	0.40	n/a	1.18	0.290	0.290	0.210	0.210	0.065	0.047	0.115	0.065	0.115	0.047	0.065	0.159	0.047	0.159	0.065	0.047
HKY+G	28	989.541	822.632	-383.036	n/a	0.05	1.19	0.255	0.324	0.193	0.228	0.073	0.043	0.126	0.057	0.106	0.051	0.057	0.179	0.051	0.141	0.073	0.043
TN93+I	29	990.587	817.738	-379.568	0.95	n/a	1.21	0.255	0.324	0.193	0.228	0.072	0.043	0.130	0.057	0.105	0.051	0.057	0.176	0.051	0.145	0.072	0.043
HKY	27	993.797	832.831	-389.155	n/a	n/a	1.18	0.255	0.324	0.193	0.228	0.073	0.043	0.125	0.057	0.106	0.051	0.057	0.178	0.051	0.140	0.073	0.043
HKY+G+I	29	997.540	824.691	-383.045	0.00	0.05	1.19	0.255	0.324	0.193	0.228	0.073	0.043	0.126	0.057	0.106	0.051	0.057	0.179	0.051	0.141	0.073	0.043
TN93+G	29	997.552	824.703	-383.051	n/a	0.05	1.19	0.255	0.324	0.193	0.228	0.073	0.043	0.129	0.057	0.104	0.051	0.057	0.174	0.051	0.145	0.073	0.043
TN93	28	1001.80	834.891	-389.165	n/a	n/a	1.18	0.255	0.324	0.193	0.228	0.073	0.043	0.129	0.057	0.103	0.051	0.057	0.174	0.051	0.144	0.073	0.043
TN93+G+I	30	1005.53	826.745	-383.051	0.00	0.05	1.19	0.255	0.324	0.193	0.228	0.073	0.043	0.129	0.057	0.104	0.051	0.057	0.174	0.051	0.145	0.073	0.043
GTR+G	32	1014.97	824.312	-379.791	n/a	0.05	0.66	0.255	0.324	0.193	0.228	0.035	0.169	0.054	0.027	0.101	0.024	0.224	0.169	0.055	0.060	0.035	0.047
GTR+G+I	33	1017.30	820.709	-376.966	0.73	0.07	0.71	0.255	0.324	0.193	0.228	0.036	0.159	0.056	0.028	0.107	0.025	0.210	0.179	0.055	0.062	0.036	0.047
GTR	31	1019.22	834.493	-385.904	n/a	n/a	0.65	0.255	0.324	0.193	0.228	0.037	0.167	0.055	0.029	0.098	0.026	0.221	0.166	0.055	0.062	0.037	0.047
GTR+I	32	1019.76	829.102	-382.186	0.62	n/a	0.65	0.255	0.324	0.193	0.228	0.036	0.168	0.054	0.028	0.099	0.025	0.223	0.167	0.055	0.061	0.036	0.047

The phylogenetic topology of the constructed tree revealed the evolutionary history of 13 ornamental palms. Of these thirteen ornamental palms, nine showed evolutionary relatedness. These nine ornamental palms viz., *Chamaedorea seifrizii*, *Chamaedorea costaricana*, *Roystonea regia*, *Caryota mitis*, *Ravenea hildebrandtii Wodyetia bifurcata*, *Dypsis lutescens*, *Archontophoenix alexandrae* and *Hyophorbe lagenicaulis* revealed divergence from the same parent in the evolutionary tree (*Fig. 1*) as nested in a clade. This phylogenetic tree showed that they are independent lineages that had been evolutionary diverged. *Chamaedorea seifrizii* and *Chamaedorea costaricana* showed closeness to each other with 70% bootstrap support. They resolved as two different species of the same genus. However, the other ornamental palms resolved as diverse evolutionary history. *Caryota mitis* and Ravenea *hildebrandtii* showed the divergence from the same parent but longer branch length of *Caryota mitis* supported that it is more evolutionary diverse than Ravenea *hildebrandtii* from their parents. Moreover, *Wodyetia bifurcata* and *Dypsis lutescens* nested closely with 61% bootstrap though, they belong to different genera. This showed genus Wodyetia and Dypsis had very close evolutionary history. The geographical isolation and genetic drift during evolution could be the reason of divergence of these palms into two distinct genera, though being lineages of the same parent. Although, *Syagrus romanzoffiana*, *Bismarckia nobilis*, *Washingtonia robusta* and *Rhapis excelsa* resolved different evolutionary history from nine ornamental palm of a clade. *Washingtonia robusta* and *Rhapis excelsa* showed evolutionary divergence from the same parent by 53% bootstrap support. However, *Bismarckia nobilis* and parent lineage of *Washingtonia robusta* and *Rhapis excelsa* showed close evolutionary history with 44% bootstrap support.
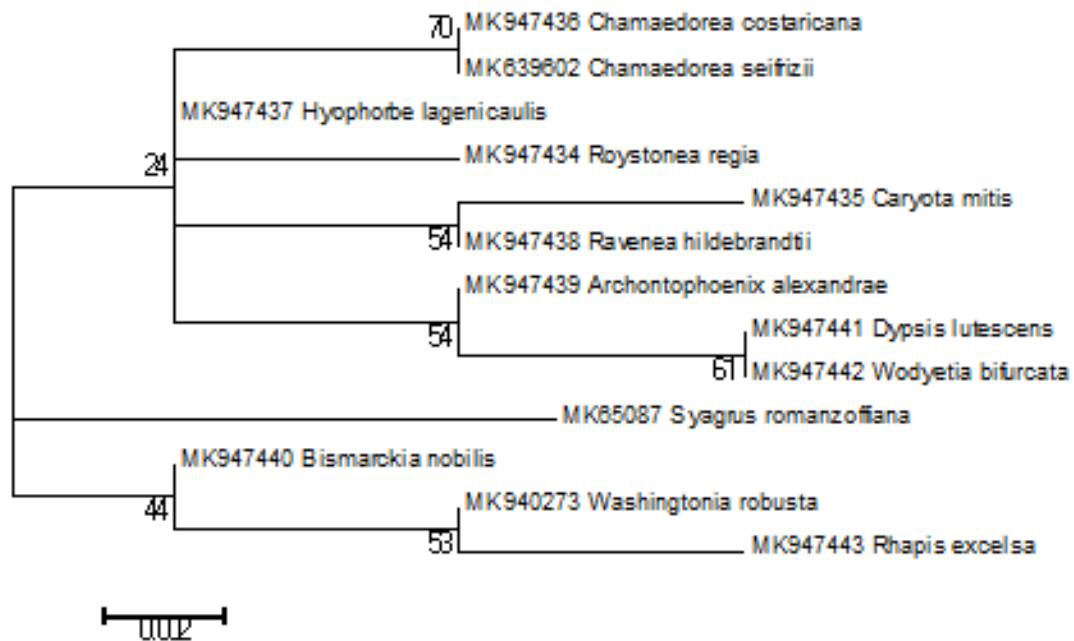
Figure 1. Chloroplast region, rbcL, based phylogenetic tree of 13 ornamental palms

Discussion

Ornamental palms, belong to Arecaceae family, are quite important landscape plants of tropical, subtropical and warm climate regions. These are famous for their palmate patterned leaves, and colorful flowers, which makes these plants noticeable in indoor and outdoor decors. High rate of plastid DNA evolution has been observed in palm trees (Asmussen et al., 2000; Dransfield et al., 2008). The determination of genetic evolution and variation through morphological characters is quite tedious in ornamental palm as morphological traits discriminates the plants at particular life stage which results in either wrong or ambiguous identification (Sedra et al., 1996). Taking these problems in account, genetic variability is determined through molecular approaches. DNA barcoding is an efficient molecular approach for efficient species identification based upon the conserved regions as a barcode/marker. For the evaluation of DNA barcode, the success rate of PCR amplification and sequencing are quite important parameters. In the present study, two conserved regions of chloroplast genome, *matK* and *rbcL* were successfully amplified for 15 ornamental palm plant identification. However, the sequencing rate of matK (26.66%) was comparatively lower than rbcL (86.66%) because of its complex heterozygosity nature. The universality of *rbcL* region have been proved by high amplification and sequencing rate as already explained by Huang et al. (2015) and Kang et al. (2017) while studying the Asian tropical trees and tropical cloud forest trees, respectively. The evolutionary tree of thirteen ornamental palms were constructed using Maximum likelihood algorithm whose nodes were supported by bootstrap values. Sequence evolution model, JC + G model, with lowest BIC value (965.423) was selected to generate a tree under maximum likelihood (ML) algorithm. In the respective tree, *Wodyetia bifurcata* and *Dypsis lutescens* are supported by 61% bootstrap, which make them closer to each other although these belong to different genera. This may be due to the less resolving power of *rbcL* genetic region of these two ornamental palm due to less evolution.

Ijaz et al.: Jukes-Cantor evolutionary model based phylogenetic relationship of economically important ornamental palms using
maximum likelihood approach
- 14864 -

Conclusion

The present evolutionary study helped in identifying the ornamental palms of Pakistan through molecular approaches as morphological characterization was not enough for correct identification due their high similarity nature. This phylogenetic study showed that nine ornamental palm shared evolutionary history, though four showed evolutionary distinctness from the rest. *Chamaedorea seifrizii* and *Chamaedorea costaricana* revealed close residence in the evolutionary tree. However, *Syagrus romanzoffiana* resolved as independent lineage with distinct and highly diverse evolutionary history. Furthermore, *Wodyetia bifurcata* and *Dypsis lutescens* showed cladding though they belong to different genera. This showed that the genus Wodyetia and Dypsis had very close evolutionary history.

REFERENCES

[1]   Asmussen, C. B., Baker, W. J., Dransfield, J. (2000): Phylogeny of the Palm Family (Arecaceae) Based on rps16 Intron and trnL-trnF Plastid DNA Sequences. – In: Wilson, K. L., Morrison, D. A. (eds.) Monocuts: Systematics and Evolution: Proceedings of the 2. International Conference on the Comparative Biology of the Monocuts, Sydney. CSIRO, Collingwood.

[2]   Dong, W., Cheng, T., Li, C., Xu, C., Long, P., Chen, C., Zhou, S. (2014): Discriminating plants using the DNA barcode rbc L b: an appraisal based on a large data set. – Mol. Ecol. Resour. 14: 336-43. DOI: 10.1111/1755-0998.12185.

[3]   Dransfield, J., Uhl, N. W., Asmussen, C. B., Baker, J. W., Harley, M. M., Lewis, C. E. (2008): Genera Palmarum: The Evolution and Classification of Palms. – Kew Publishing, Richmond, UK.

[4]   Ford, C. S., Ayres, K. L., Toomey, N., Haider, N., Van, A. S. J., Kelly, L. J., Wikström, N., Hollingsworth, P. M., Duff, R. J., Hoot, S. B., Cowan, R. S. (2009): Selection of candidate coding DNA barcoding regions for use on land plants. – Botanical Journal of the Linnean Society 159(1): 1–11.

[5]   Huang, X. C., Ci, X. Q., Conran, J. G., Li, J. (2015): Application of DNA barcodes in Asian tropical trees–a case study from Xishuangbanna Nature Reserve, Southwest China. – PLoS One 10. DOI: 10.1371/journal.pone.0129295.

[6]   Henderson, A., Galeano, A. G., Bernal, R. (1995): Field Guide to the Palms of the Americas. – Princeton University Press, Princeton, NJ.

[7]   Kang, Y., Deng, Z., Zang, R., Long, W. (2017): DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. – Scientific Reports 7. DOI: 10.1038/s41598-017-13057-0.

[8]   Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., Janzen, D. H. (2005): Use of DNA barcodes to identify flowering plants. – Proc. Natl. Acad. Sci. 102: 8369-74. DOI: 10.1073/pnas.0503123102.

[9]   Lorenzi, H., Larry, R., Noblick, Kahn, F., Ferreira, E. (2010): Brazilian Flora: Arecaceae (Palms). – Instituto Plantarum de Estudos da Flora, Nova Odessa, Brazil.

[10]   Maloukh, L., Kumarappan, A., Jarrar, M., Salehi, J., El-Wakil, H., Lakshmi, T. R. (2017): Discriminatory power of *rbcL* barcode locus for authentication of some of United Arab Emirates (UAE) native plants. – 3 Biotech 7: 144. DOI: 10.1007/s13205-017-0746-1.

[11]   Pintaud, J. C. (2008): An overview of the taxonomy of Attalea (Arecaceae). – Revista Peruana de Biología 15: 53-62.

[12]   Sedra, M., El, F. H., Benzine, A., Allaoui, M., et al. (1996): La palmeraie dattiere marocaine: evaluation du patrimoine phenicicole. – Fruits 1: 247-259.

[13] Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013): MEGA6: molecular evolutionary genetics analysis version 6.0. – Mol. Biol. Evol. 30: 2725-9. DOI: 10.1093/molbev/mst197.