

OPEN

The morpho-genetic and ecological niche analyses reveal the existence of climatically restricted *Cycas zeylanica* complex in Sri Lanka

Asanka Mudannayake^{1,2}, Lahiru Ranaweera³, Preminda Samaraweera^{2,3},
Suneth Sooriyapathirana^{2,3} & Anoma Perera^{1,2*}

Taxonomy and phylogenesis of Sri Lankan cycad species of the subsection *Rumphiae* has not been fully resolved and therefore, we conducted an island-wide survey of cycads of the subsection to assess their morphological or genetic variations while exploring the phylogenetic relationship between Sri Lankan *Rumphiae* and other world cycad species. Further, we assessed the possible distribution of the species in the region through climatic profiling, using maximum entropy modeling approach. We analyzed 21 variable morphological features in collected specimens and used the polymorphism of *trnH-psbA* locus to understand the phylogeny. The distance tree drawn from the principal component analysis revealed a significant variation in female reproductive structures. The maximum likelihood tree separated Sri Lankan *Cycas zeylanica* to a well-supported unigeneric clade (bootstrap = 96, posterior probability = 100) with shallow divergence. Ecological niche modeling supported the existence of *Cycas zeylanica* in South East Asia and in southern Western Ghats in India in addition to the Wet Zone of Sri Lanka. We rename the taxa as *Cycas zeylanica* complex based on the observed high morphological diversity of female reproductive structures which might have ascended due to multiple introductions of South East Asian cycads by long distance dispersal of seeds through sea currents.

The genus *Cycas* which belongs to the unigeneric family Cycadaceae, is considered as the basal lineage of the order Cycadales¹. To date, 115 species are reported under the genus *Cycas* worldwide², and of these, around 40 species occur in the Indo-Chinese region³. Inherent characters of cycads such as slow-growth, a heavy natural variability of morphological traits⁴ and dioecious nature made the identification of individuals to the species level difficult in some instances, but identifying to the section or subsection levels merely based on morphological features is not challenging. Thus, six sections of the genus *Cycas* have been recognized and often used in taxonomic descriptions. Four of these sections viz., *Cycas*, *Asiorientales*, *Indosinenses* and *Stangerioides*⁵ had been identified initially, but later two other sections, i.e. *Wadeae*⁶ and *Panzhihuaenses*⁷ have been added to the classification system. Of these, the section *Cycas* is further divided into three subsections: *Cycas*^{8,9}, *Endemicae*^{9,10} and *Rumphiae*^{8,9}.

The combined feature of the presence of spongy tissues in the endotesta or the presence of a fibrous layer in sarcotesta of the seed is a primary attribute of taxonomic significance in differentiating the individuals of the three subsections of the section *Cycas*^{8,11}. Thus, this feature allows the differentiation of Sri Lankan cycads into two subsections, *Cycas* and *Rumphiae*^{12,13}. Owing to the fact that there are many similarities in morphological features among different Sri Lankan taxa, identifying these to the species level is practically difficult¹³. Reproductive morphological characteristics of individuals of the subsection *Rumphiae* in Sri Lanka vary dramatically though these show a restricted distribution within the country, reflecting complex evolutionary relationships among Asian cycads.

Presence of *C. rumphii* Miq. (subsection *Rumphiae*) in Sri Lanka has initially been recorded in the *Revised Handbook to the Flora of Ceylon*¹⁴, but later, the occurrence of *Cycas zeylanica* (J. Schuster) A. Lindstrom and K. D. Hill in Sri Lanka was confirmed by Lindstrom and Hill¹². However, there are no records available to negate the presence of *C. rumphii* in Sri Lanka. Therefore, by an island-wide survey of individuals of the genus *Cycas*

¹Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka. ²Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka. ³Department of Molecular Biology and Biotechnology, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka. *email: anoma29@gmail.com

in Sri Lanka, Mudannayake *et al.*¹³ have studied 26 reproductive and 23 vegetative features of individuals of the subsection *Rumphiae* and observed the presence of substantial continuous variation in many of the important morphological features among the individuals, which led to describe the group as *C. rumphii* complex. In general, the features of the megasporophyll which is crucial in delimiting the subsection *Rumphiae* species^{8,13,15,16} show a considerable variation. The existence of such a substantial variation of morphological features within the subsection *Rumphiae* in Sri Lanka could be attributed to many factors including their ecological niches or dispersal pathways, and need to be further investigated.

The spongy tissues present in seeds of the subsection *Rumphiae* plants make them buoyant in seawater, allowing their seeds to disperse over different localities. As many researchers have explained^{17–19}, seed dispersal through seawater may lead to the successful recolonization of plants in new regions and to form genetically distinct forms in colonized locations. As such, hybridization among the members of the subsection *Rumphiae* could have occurred in introduced locations^{10,20}. Due to the genetic mixing caused by hybridization, the taxonomy of the subsection *Rumphiae* based on morphological features has become very complicated, leaving molecular characterization as the sole means of identification and characterization of species. Thus, it is vital to revisit the subsection *Rumphiae* in Sri Lanka along with other cycads using molecular systematics.

Molecular phylogenetics is a reliable tool in taxonomic and phylogeographic studies. Nucleotide polymorphism of conserved gene regions is a key component in explaining the genealogies. In plants, most of the taxonomically essential gene regions are found in the chloroplast. The *trnH-psbA* locus, a non-coding intergenic spacer region in the plastid genome, is considered as the most appropriate single locus barcoding region for plants²¹ due to the presence of both SNP (Single nucleotide polymorphism) and INDELs (Insertions and deletions)^{22,23}. Moreover, *trnH-psbA* exhibits higher mean percent sequence divergence at generic level enabling a good resolution power²¹. The application of *trnH-psbA* sequence polymorphism to identify species is a routine practice in plant systematics²³. The *trnH-psbA* has become a popular barcoding locus for Cycadaceae as well^{24–26}. The length of *trnH-psbA* sequences in Gymnosperms ranges from 283–1006 bp²³, and with good priming sites, short read lengths (~ 600 bp) can be efficiently sequenced across many taxa.

In the present study, we aimed to reveal the presence of morphological or genetic variations within the subsection *Rumphiae* in Sri Lanka, explore the phylogenetic relationship between Sri Lankan *Cycas zeylanica* and other world cycads, and explain the possible distribution of the species using climatic profiling through Ecological Niche Modeling (ENM) analysis.

Materials and Methods

Sampling. We conducted an island-wide survey by examining about 200 individuals of Sri Lankan cycads using the morphological features explained by Mudannayake *et al.*¹³. We considered a total of 49 morphological features (Supplementary Table S1) and recorded only the features available at the time of sampling. We identified that 53 individuals belong to the subsection *Rumphiae* and all of these except two cultivated plants (R1 and R2) inhabit the lowland Wet Zone of Sri Lanka (Fig. 1) where the elevation is below 1,067 m (3,500 ft). Also, we noted that the frequency of occurrence of individuals is comparatively higher near the coast than inland. Evidence proves that the above mentioned two individuals sampled in the highland Wet Zone are cultivated plants and have not occurred due to natural regeneration. We photographed and made line-diagrams of female and male reproductive structures available in the individuals assessed.

Morphological data analysis. Out of the 53 individuals found, we examined morphological (vegetative, female and male reproductive) features of 40 individuals which had attained the reproductive maturity. Out of these 40 individuals, female reproductive structures were available only in 25 individuals and male reproductive structures were available only in one individual. We did not find reproductive structures in the rest of the individuals. Therefore, we only considered 25 female plants with observable reproductive structures for the morphological data analysis. After initial examination of 49 morphological features (Supplementary Table S1), we removed less variable and monomorphic characters of the individuals studied, and thus, 21 quantitative and qualitative morphological features (11 vegetative and 10 female reproductive) were considered for the detailed analysis (Supplementary Table S1). We analyzed the data by variable reduction multivariate techniques²⁷, Pearson's correlation analysis for quantitative data, Spearman rank correlation analysis for qualitative data and Principal Component Analysis (PCA) for both qualitative and quantitative data by PCAmix function in PCAmixdata Package²⁸ using R Statistical Software²⁹. The presence of different groups was revealed by the morphological analysis based on a similarity matrix developed for the Principal Components (PCs) by dendrogram generated by Ward linkage method based on Euclidean distances.

DNA extraction and sequencing. We extracted genomic DNA from the immature leaf samples of 17 individuals of the subsection *Rumphiae* representing different localities using DNeasy plant mini kit (Qiagen, Hilden, Germany). Polymerase Chain Reaction (PCR) was carried out in 40 µl reactions using 2 × GoTaq Green Master Mix (Promega Co, Madison, WI, USA) according to manufacturer's protocol using the primers *psbA-trnH* (P_F: 5'/GTT ATG CAT GAA CGT AAT GCTC 3'³⁰; P_R: 5'/CGC GCA TGG TGG ATT CAC AA TCC 3'³¹). PCR was carried out under the following conditions: 5 min denaturation at 94 °C, followed by 30 cycles of 94 °C for 1 min, primer annealing at 55 °C for 1 min and synthesis at 72 °C for 1 min and 30 s, with a final extension at 72 °C for 7 min²¹. We purified PCR products and subjected to the 3x cycle sequencing using a 3100 sequencer (Applied Biosystems).

Phylogenetic analysis. We retrieved the sequence data of 30 individuals (in addition to the 17 sequences generated in the present study) for the locus *trnH-psbA* from GenBank to represent the total variation of the six subsections in the genus *Cycas* (Supplementary Table S2). To align 47 sequences, we implemented CLUSTAL

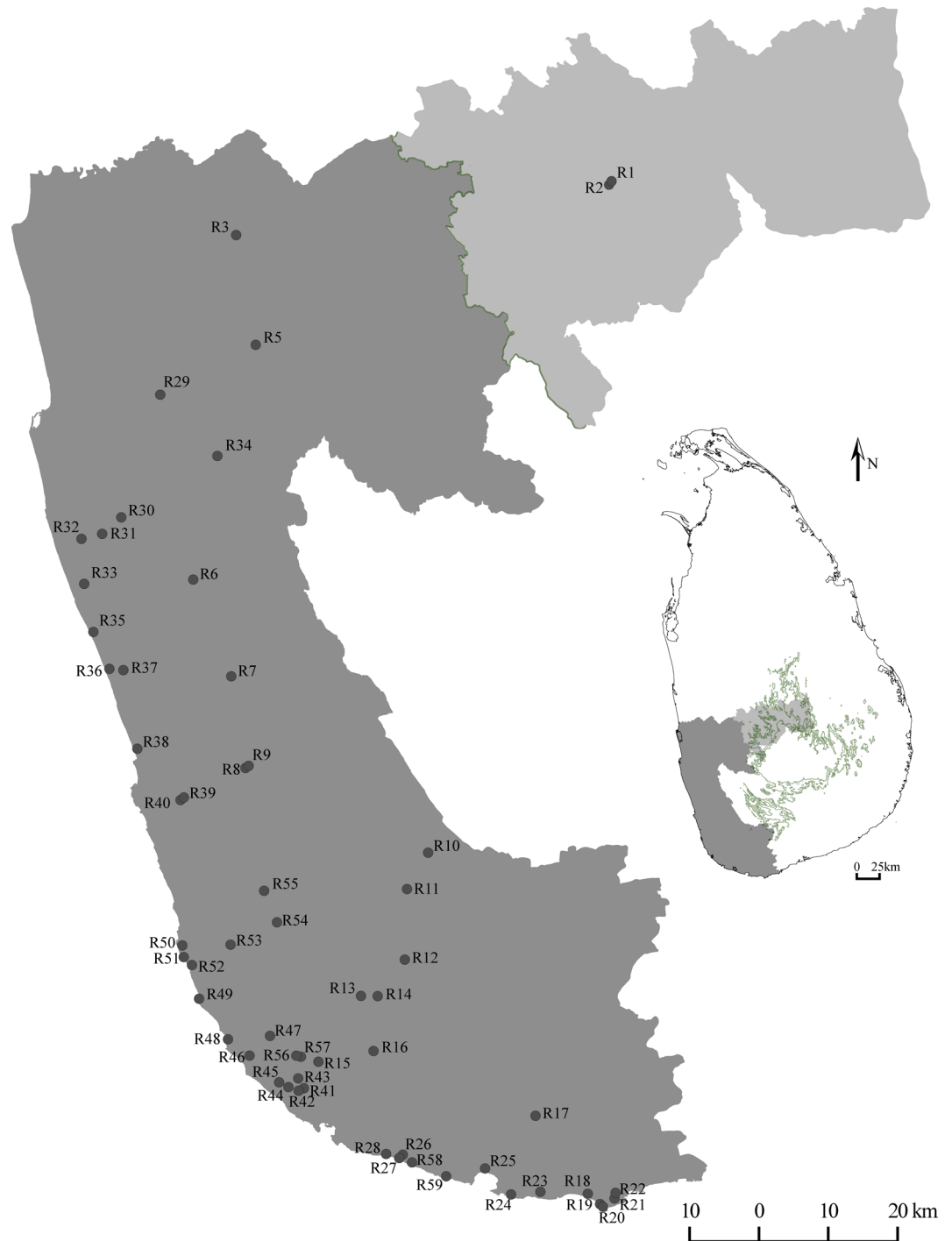


Figure 1. The locations of the individuals examined, of subsection *Rumphiae* in Sri Lanka. The contour at 1,067 m (3,500 ft) is shown in green lines in the map of Sri Lanka.

W algorithm³² in MEGA V7³³. To select the precise evolutionary model to fit the data, we implemented jModelTest V.2³⁴ in CIPRES supercomputer³⁵. We avoided oversimplification of the model selection by assessing across 88 models using more than 95 parameters. To choose the best model in Akaike Information Criteria (AIC)³⁶, we used AIC weights and negative log-likelihood values. We carried out a tree search in RAxML³⁷ using the rapid + bootstrap algorithm for 1000 iterations in CIPRES science gateway³⁵. The best scoring Maximum Likelihood (ML) tree was selected using log Likelihood values. Then all the bootstrap (bs) trees generated were implemented to produce one majority rule consensus tree. We used the best ML tree topology as the primary tree structure to interpret the bs values given in the majority rule consensus tree. To support our analysis further, we ran a Bayesian tree search using MrBayes 3.1³⁸ in CIPRES Science gateway³⁵. We implemented four Markov Chain Monte Carlo (MCMC) chains (two hot and two cold chains) for 50 million generations. The MCMC chains

sampled trees in every 5000 cycles, and the analysis was set to discard the first 10% of the trees as burn-in. We used substitution parameter, the proportion of invariable sites and gamma shape parameters given by the evolutionary model to run this analysis. We assessed the Effective Sample Size (ESS) of the target distribution to measure the chain convergence to avoid the poor mixing and for the enhancement of the independent sampling of the sampled trees. The trees probed after maximum chain convergence were used to draw the 50% majority rule consensus tree. We used FigTree v1.4.3³⁹ to edit the trees further.

Ecological Niche Modeling (ENM). We used 53 ‘presence-locations’ (i.e., total sampling effort) to model the predicted distribution of *C. rumphii* complex. We used geographic extant including most of the regions of South Asia and South East Asia where closely related cycads are known to occur. As the environmental covariates, we downloaded 19 bioclimatic variables (Supplementary Table S3) in Global Climate Databases⁴⁰. Using ArcGIS Version 10.4⁴¹, we clipped 19 climatic layers into chosen geography. We used Maxent version 3.3.3k⁴² to carry out maximum entropy modeling. Here we did not use a Bias file as we are modeling the species distribution in full extent of selected geography, thus over fitting of the model could be avoided. The model fitting was carried out by using the Threshold Independent Method⁴³ where the different β regulation parameters were changed ($\beta = 0.1, 1, 2, 3, 4, 5$) while assessing the Area Under Curve (AUC) parameters. We kept most of the other parameters default, as the Maxent chooses the best parameters for the analysis^{44–46}. We ran the analysis for 5000 iterations using 1000 background points to increase the robustness of the model. We created the predicted distribution map using 10% threshold^{47–49}.

Results

Morphology. Out of the 49 morphological features considered (23 vegetative, 20 female reproductive and six male reproductive traits) (Supplementary Table S1), 21 were found to be polymorphic, two were polymorphic but depend on the position of attachment on the tree, 17 were monomorphic, three were less variable, while six were not considered due to data deficiency. We considered only the 21 polymorphic traits of the female plants for the analysis which consists of 17 quantitative and four qualitative traits.

Variability of the vegetative traits. The three vegetative traits, % spines on the petiole (PSP), leaflet width (LW) and cataphyll length (CL) (Supplementary Table S1) showed non-normal distributions as revealed by the significant Anderson Darling (AD) coefficient [$(p \leq 0.05)$ (Fig. 2A,E,J)] whereas the features circumference of the petiole (CP), length of the leaf lamina (LLL), leaflet length (LL), ratio of leaflet length to leaflet width (LL/LW), number of leaflets (NL), ratio of NL to LLL (NL/LLL) and leaflet spacing (LSP) were normally distributed [$(p > 0.05)$ (Fig. 2B–D,F–I)]. The PSP of 90 was more common in the plants studied and fell within the mid-point of the range observed (Fig. 2A). The variation of the CP displayed quantitative and normal distribution with AD of 0.30. The parameters LLL, LL, LL/LW, NL, LSP and CL, displayed continuous quantitative distributions with very high variability. The only selected variable morphological feature of insertion angle to rachis (IAR) showed strict dimorphism with 3:2 observed ratio (Fig. 2K) and fitting into the dihybrid epistatic ratio of 9:7 (complementary gene action) ($\chi^2 = 0.143, p = 0.0705$).

Variability of the female reproductive traits. Except number of ovules (NO) (Fig. 3B), length of the fertile region (LFR) (Fig. 3D) and, ratio of length of the infertile region (LIR) to maximum width of the infertile region (LIR/MWIR) (Fig. 3F), the other quantitative traits exhibited normal distributions (Fig. 3A,C,E and G). The feature MWIR and LAS (Fig. 2A,C) showed higher variation across the individuals assessed. Out of the three quantitative polymorphic variables, the margin of the infertile region (MIR) showed trimorphic distribution (9:6:1, $\chi^2 = 1.24, p = 0.538$) (Fig. 4A) whereas shape of the lamina of the infertile region (SLIR) displayed dimorphic distribution (15:1, $\chi^2 = 0.216, p = 0.642$) (Fig. 4B).

Correlation analyses. To ascertain the correlation among the quantitative vegetative and reproductive features the detailed Pearson Correlation Coefficient (PCC) values with the associated probability levels were calculated, and are given in the Supplementary Tables S4 and S5. The number of leaflets (NL) and the length of the leaf lamina (LLL) were strongly and significantly correlated (PCC = 0.71). However, CL and LLL were negatively correlated, and the strength of correlation was less (PCC = -0.42). The length of the apical spine (LAS) and LFR were correlated with NO (PCC = 0.59 and PCC = 0.53 respectively). A solitary significant correlation was observed between the qualitative traits MIR and NLS (PCC = -0.67).

Cluster analysis based on the morphological diversity. The selected polymorphic morphological features subjected to PCA, detected 12 PCs, which explained 91% of the total variability (Supplementary Table S6). The calculated PCs, explained variance and Eigen values are given in Supplementary Table S6. Based on all the PCs, we drew a distance tree and obtained six clear clusters at a morphological divergence of 30% (Fig. 5). The Cluster A included nine individuals, and the Cluster B included only one individual. The Clusters C and F contained three individuals in each and Clusters D and E contained five and four individuals respectively. The occurrence of extreme heterozygosity through hybridization is apparent from Fig. 5 as no two individuals were identical to each other.

Phylogenetic analysis. The sequenced gene region consisted of 611 base pairs. Of these 31 bases were polymorphic and only 16 sites were parsimony-informative while 14 sites were singleton. The details of the sequence polymorphism of *trnH psbA* are given in Supplementary Table S7. The TPMuf + I sub model was selected as the best evolutionary model to describe the dataset, during the model selection (negative log likelihood = 1063.24, AIC weight = 0.076). We used rate parameters (AC = 0.302, AG = 1.017, AT = -1.000, CG = 0.302, CT = 1.017 and GT = 1.000), proportion of invariable sites (0.66) and nucleotide frequencies [$f(A) = 0.27, f(C) = 0.18,$

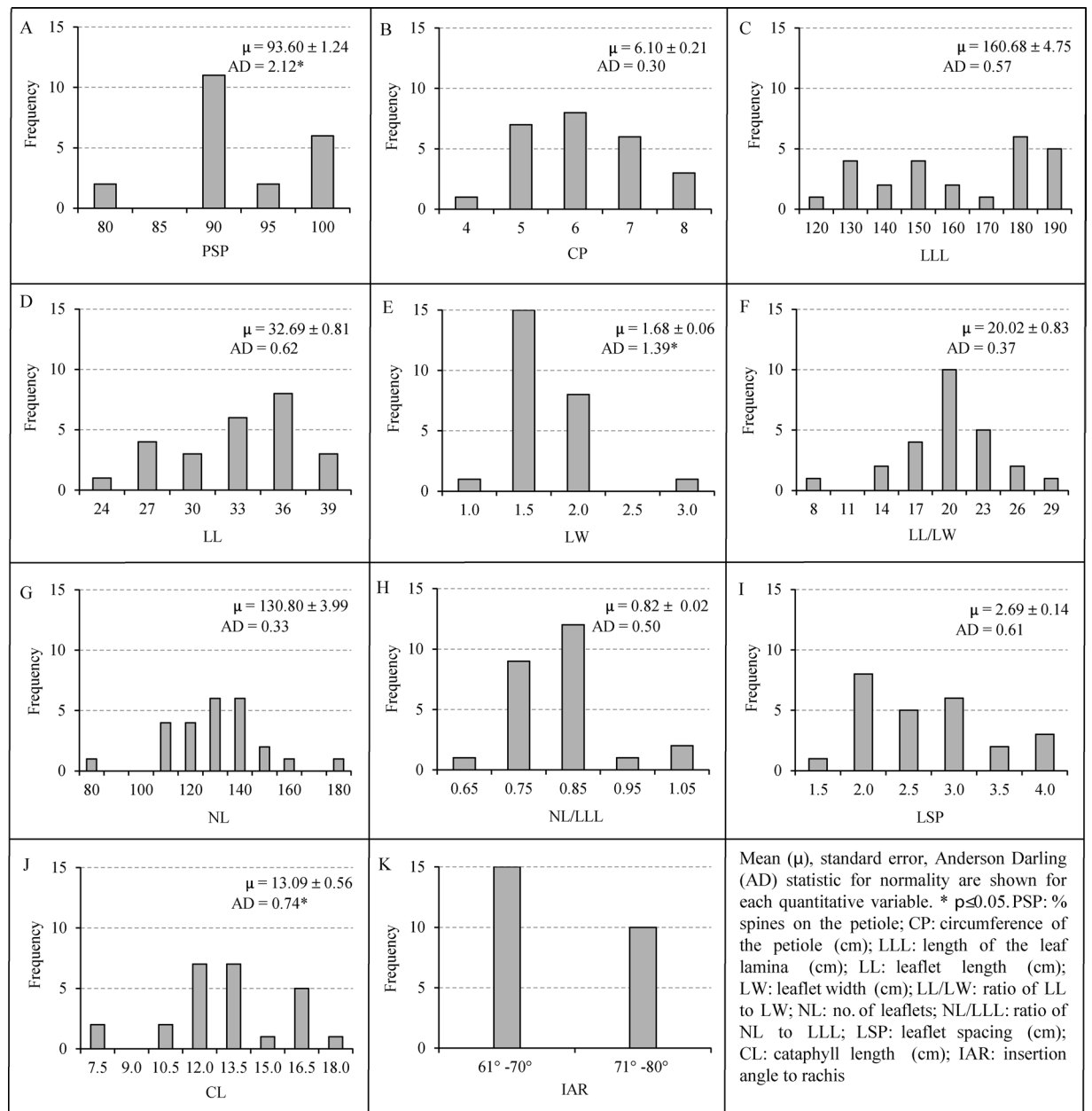


Figure 2. Variability of the vegetative features of the subsection *Rumphiae* in Sri Lanka. (A–J) quantitative traits, (K) qualitative trait.

$f(G) = 0.21$ and $f(T) = 0.34$] of the evolutionary model to carry out the further analysis. Our tree search in Maximum Likelihood (ML) framework resulted in best scoring ML tree at a negative log likelihood value of 1062.74. The bootstrap values are indicated on the best scoring ML tree (Fig. 6). The MCMC chains converged maximum in the first 5000 runs which were discarded as burn-in. The ESS for sampled priors were above 200 indicating independent sampling from posterior distribution and reliable tree topology. The best scoring ML tree and the 50% majority rule consensus tree built in Bayesian framework had almost similar branching patterns. For the ease of representation, we only show the ML tree with imprinted node support values (Fig. 6). The Sri Lankan *Cycas Rumphii* complex (*Cycas zeylanica* complex nov.) formed a well-supported unigeneric clade (bootstrap = 96, posterior probability = 100) with almost identical individuals. However, the accession MF348799²⁶ was also clustered within this clade, but its specimen has been identified as *C. circinalis* of the subsection *Cycas* and it is very unlikely to have members of subsection *Cycas* within this clade. It is possible that this could be a misidentification. The GenBank accession of *C. zeylanica* (KX182329) was also diverged out from the *C. zeylanica* complex clade nov., and it could also be a misidentification and must be revisited (Fig. 6). We submitted the sequences generated in the present study in GenBank under the accession numbers MH458253–MH458269.

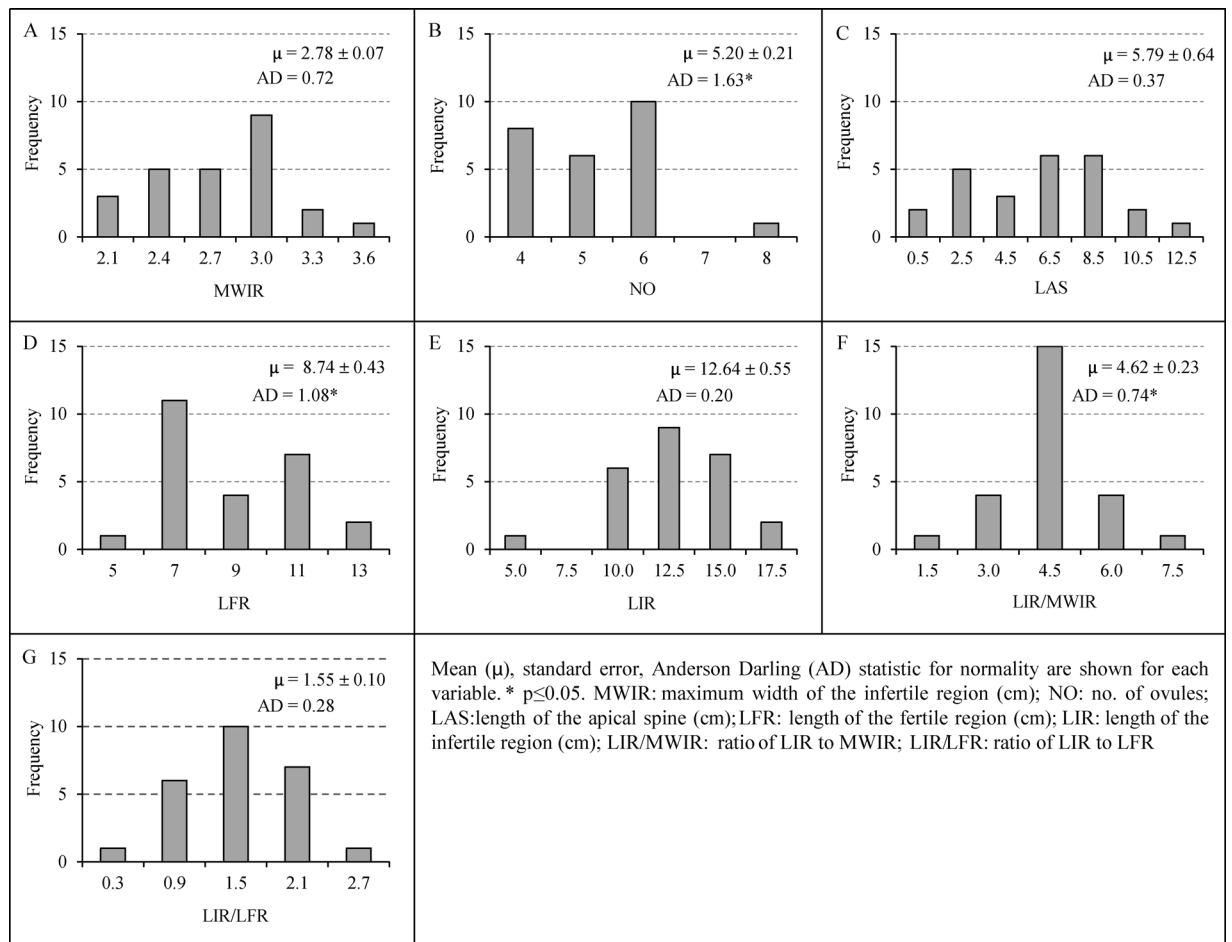


Figure 3. Variability of the quantitative traits of the female reproductive features of the subsection *Rumphiae* in Sri Lanka.

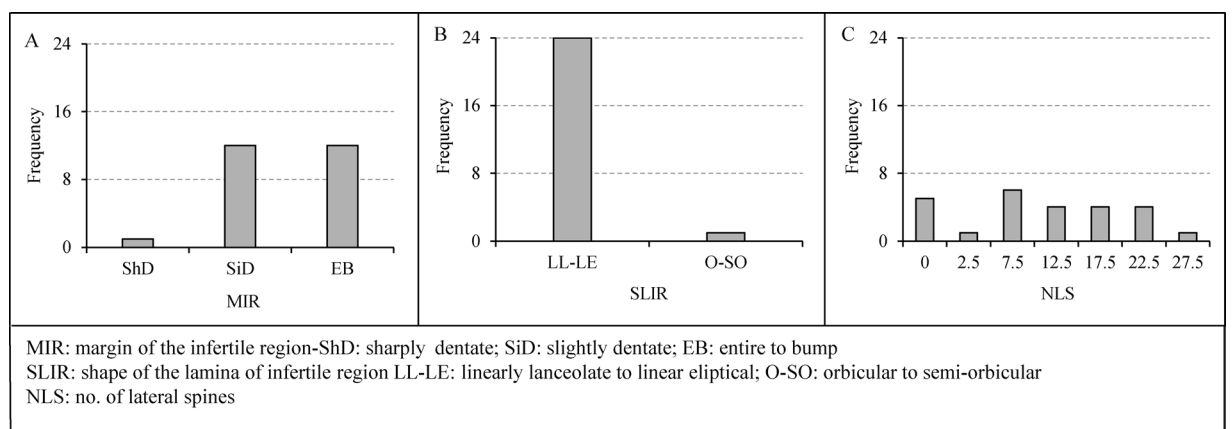


Figure 4. Variability of the qualitative traits of the female reproductive features of the subsection *Rumphiae* in Sri Lanka.

Ecological niche model. We obtained the best AUC value of 0.975 for Receiver Operating Characteristic (ROC) Curve and the best Omission/Commission curves at the β regulation parameter of 5. The training omission rate was slightly higher than the test omission rate (ROC and AUC curves in Supplementary Fig. S1). Upon the total predicted distribution, Mean Diurnal Range had 53.9% of variable importance. Precipitation of the driest month and isothermality had 16.8% and 10.8% variable importance respectively. The predicted distribution derived from ENM shows Andaman and Nicobar Islands, Northern part of the Sumatra Island, Southern part of

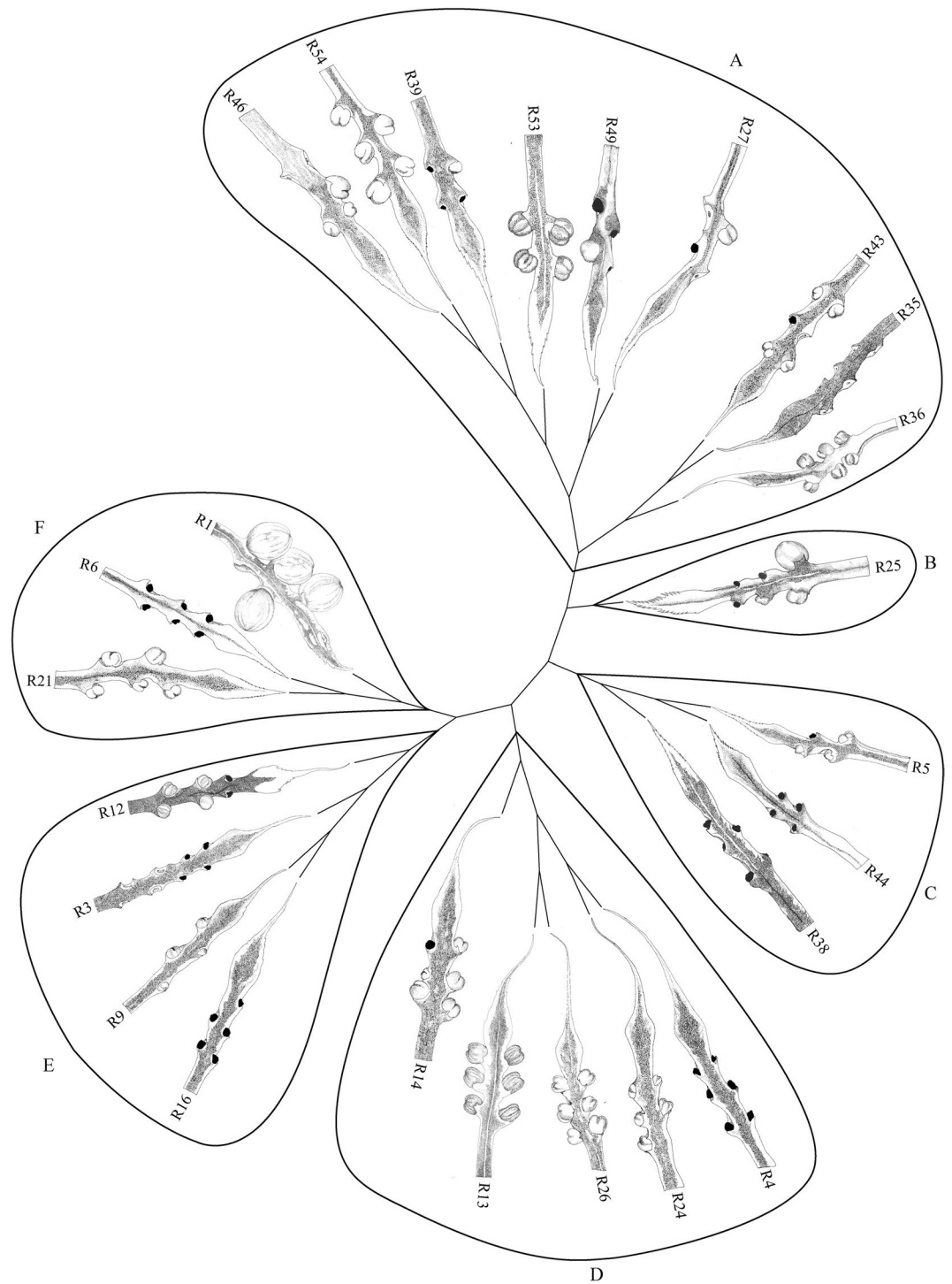


Figure 5. Dendrogram generated by Ward linkage method based on Euclidean distance drawn for all the Principal Components. The similarity among the samples was studied based on morphological traits and six distinct Clusters (A–F) were identified at 70% similarity. The megasporophyll structures of each individual of the subsection *Rumphiae* of Sri Lanka are given at the Operational Taxonomic Units of the dendrogram.

Thailand extending into northern part of Peninsular Malaysia as the predicted favourable niche for *C. zeylanica* complex *nov.* (Fig. 7). Apart from South East Asia, the model predicted a low level of distribution in Southern parts of the Western Ghats in India. The predicted distribution of *C. zeylanica* complex *nov.* is in line with our sampling locations as lowland Wet Zone of the country is the envisaged niche in Sri Lanka. This strengthens the accuracy and reliability of the model applied in the analysis.

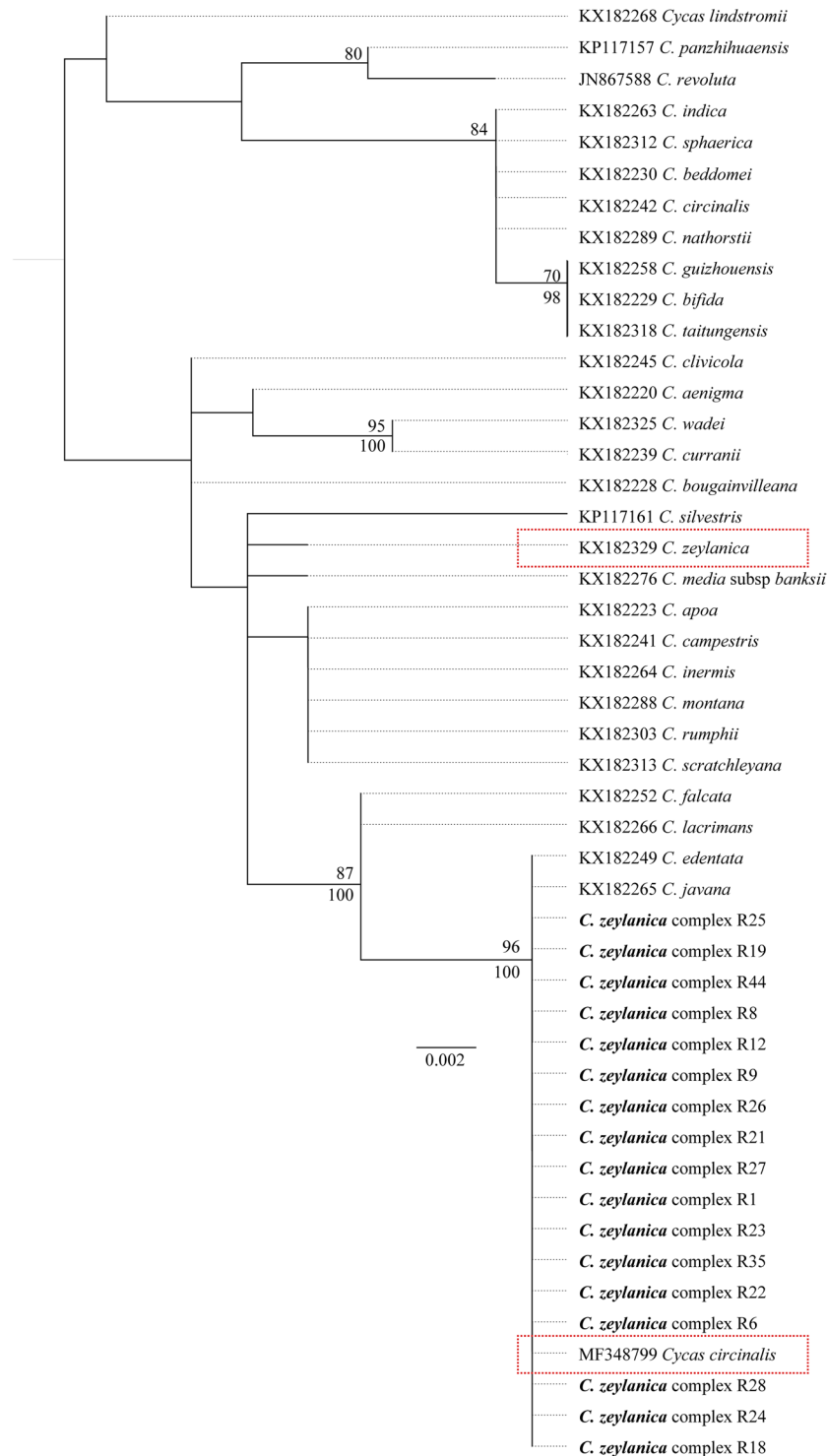


Figure 6. The best scoring Maximum Likelihood tree constructed in RAxML. The bs values are given above the node and PP values are indicated below the node. The bs values less than 70 and PP values less than 90 are not shown. Possibly misidentified individuals are highlighted in red boxes.

Discussion

Based on the morphological, molecular and climatic features of the studied taxa, the present study reveals that the Sri Lankan *Cycas* populations of the subsection *Rumphiae* mostly resemble *Cycas zeylanica*. In the *trnH-psbA* phylogeny (Fig. 6), *C. rumphii* diverged out from the clade containing the Sri Lankan subsection *Rumphiae* species.

We detected a high morphological variation in some vegetative traits as well as in some female reproductive features of *C. zeylanica*. This could be an artifact of hybridization among different congeneric species, and accordingly, we rename the studied taxa as *C. zeylanica* complex. The variation of many of the examined

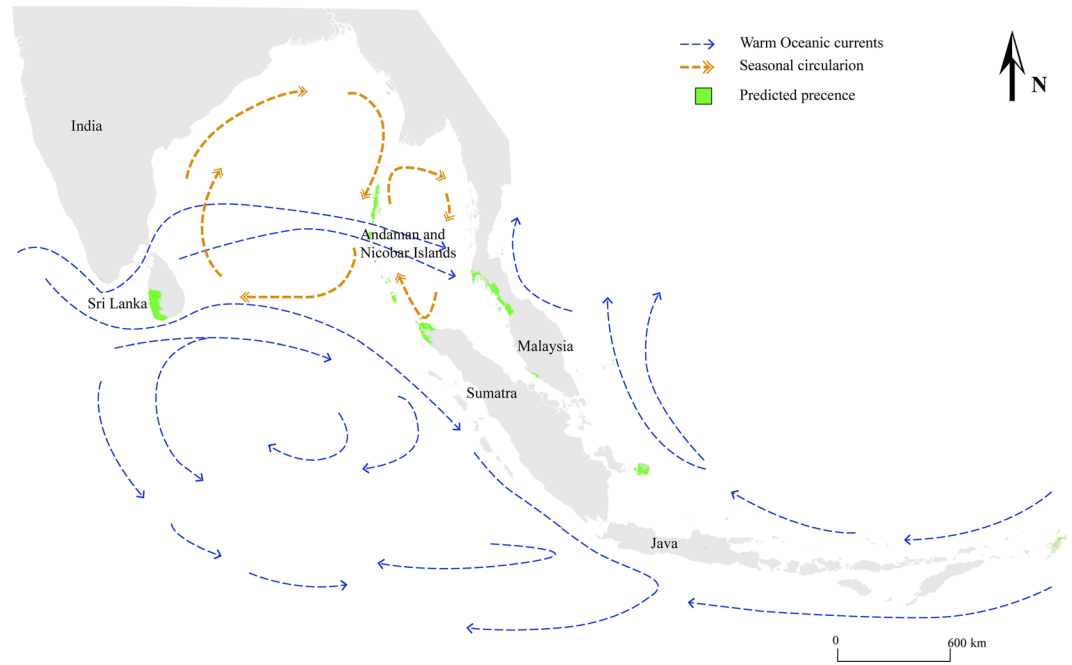


Figure 7. Predicted distribution of *C. zeylanica* derived from maximum entropy modeling. The oceanic currents related to Indian Ocean Gyre are represented in arrows.

vegetative features appears to be so shallow and could not be used to distinguish different morphological groups (Supplementary Tables S4 and S5). In contrast, the female reproductive features considered in our study appear to be exhibiting significant variations. Of these, we identified five discriminating morphological traits that contribute to the variation of female reproductive structures of the *C. zeylanica* complex and these could be used in field identification of the species. The infertile region of the megasporophyll is the key feature in which all the significantly variable traits were detected (ratio of LIR/MWIR, ratio of LIR/LFR, LIR, NLS and MIR) (Supplementary Tables S4 and S5). Absence of apical crest-like structure in the seeds⁸ is also a feature that can be used in differentiating *C. zeylanica* from *C. rumphii*, but we were unable to use this feature as mature seeds were not available in all samples examined.

Our results further revealed that *C. zeylanica* complex in Sri Lanka can be assembled into six morphological groups with respect to female reproductive traits (Fig. 5). The Clusters B, D and F show independently unique features among them (MIR, ratio of LIR to MWIR and LIR). In contrast, the clusters A, C and E have some features derived from B, D and F clusters, indicating a possible mixing of *C. zeylanica* complex within Sri Lanka. Possibly, individuals belonging to Clusters B, D and F could be serving as the maternal populations, and all these findings support the theory of hybridization in *Cycas*^{50–53}. A very low genetic divergence was observed in the phylogeny of the Sri Lankan *C. zeylanica* populations and this could be due to high gene flow among different morphotypes through hybridization. The possibility of hybridization among the congeneric members of subsection *Rumphiae* or their ancestors have also been observed by Keppel *et al.*¹⁹, and there is a possibility that these hybridized seeds may have dispersed over long distances aided by ocean currents or humans.

It appears that *C. zeylanica* complex in Sri Lanka is derived from multiple introductions from South East Asia. As depicted in the shape of the infertile region of the megasporophyll, one individual of the Cluster E (R12) (Fig. 5) resembles those of *Cycas pschannae* and *Cycas andamanica*. Semi-orbicular shape of the infertile region and presence of distinct hook-like structure in the margin of the base of the sterile portion were prominent features of the megasporophylls. Moreover, megasporophylls of Cluster D show a close resemblance to that of *C. edentata* (native range: Indonesia, Malaysia, Myanmar, Philippines, Singapore, Thailand and Vietnam), by the presence of lengthy infertile regions with entire margins. The individual R39 of Cluster A resembles *Cycas javana* (enigmatic taxa⁵⁴ endemic to Java Islands) with respect to the arrangement of lateral spines and size of the infertile region, which is strongly supported by the ML tree where *Cycas edentata* and *C. javana* are clustered together with *C. zeylanica* complex (Fig. 6). Besides, individuals of the Cluster B show morphological similarities to *Cycas falcata*, which grows in Indonesia-Sulawesi Islands, by the presence of prominent lengthy lateral spines in the sterile region of the megasporophyll. In our phylogenetic tree, this is sister to the clade containing *C. zeylanica* complex (Fig. 6), confirming close evolutionary relationships. All these imply that convergence of species of the subsection *Rumphiae* may have taken place in Sri Lanka. This is in parallel with MacArthur and Levins's⁵⁵ finding that two similar species could produce a third intermediate species and ultimately will converge towards the nearer of the pair.

Although the vegetative morphological variations contributed less in forming above groups, PSP showed a continuous distribution among samples observed. Many individuals have PSP of 90 (Fig. 2A) indicating that they are frequently undergoing hybridization. Similarly, the very high variability and continuous nature of the

distributions observed for LLL, LL, LL/LW, NL, LSP and CL also support the hypothesis of frequent mixing by hybridization. The observed traits could be controlled either by major genes or polygenes. However, the feature IAR showed a strict dimorphic nature and the observed ratio fits into modified epistatic ratio of 9:7 (i.e., complete gene-action). This could also be considered as important evidence for hybridization between extreme morphotypes. Similarly, the reproductive morphological feature, SLIR with the goodness of fit in modified epistatic ratio 15:1, strongly supports the hypothesis of hybridization events. These shreds of evidence propose multiple introductions of different South East Asian cycad species to Sri Lanka in the past and further hybridization within the island. As a result, a population with a unique genetic structure has been formed *in situ* (Fig. 6).

In Sri Lanka, previous studies have reported that individuals of the subsection *Rumphiae* have inherited a wide range of morphological diversity and had described these as forming a complex^{8,50}. In light of these findings and in the absence of significant molecular and habitat-wise assessment, Mudannayake *et al.*¹³ have previously named Sri Lankan *Rumphiae* populations as *C. rumphii* complex. However, in agreement with Lindstrom and Hill¹², the present study identified that the members of the Sri Lankan subsection *Rumphiae* are closer to *C. zeylanica* and the study group formed a well-supported unigeneric clade in the ML tree despite the high morphological diversity observed in female reproductive structures. Further, the niche of the study group overlaps with *C. zeylanica* niche¹². Accordingly, we rename Sri Lankan *C. rumphii* complex as *C. zeylanica* complex.

The high morphological variation observed in female reproductive and some vegetative features of *C. zeylanica* complex had led to create taxonomic ambiguities not only in Sri Lanka¹³ but also elsewhere in Asia. According to available literature, several new species have been described in the recent past under the subsection *Rumphiae* from the Andaman Islands and sometimes from adjacent areas, but later, some of these have been reported as synonyms of previously named species or recognized to have many similarities to previously named species (Supplementary Table S8). For instance, *Cycas sainatii* has been reported as allied to *C. zeylanica*⁵⁶, *Cycas darshii* allied to *C. rumphii*⁵⁷, and *C. pschannae* allied to both *C. zeylanica* and *C. sainatii*⁵⁸. Meanwhile, in 2015, another *Cycas* population was recorded from the coast of Middle and North Andaman Islands⁵⁹. However, the species is considered as allied to both *C. zeylanica* and *C. edentata*. More recently, Sing⁶⁰ has described *Cycas dharmarajii* which is also endemic to Andaman Islands. The species is reported as allied to both *C. zeylanica* and *C. pschannae*. However, Calonje *et al.*² have stated that *C. andamanica* and *C. dharmarajii* are synonyms of *C. sainatii*. Such frequent ambiguities associated with cycads, especially under the subsection *Rumphiae* may be due to their high morphological variations, which may be an artifact of hybridization. Therefore, it is necessary to study these species at molecular level also to reveal evolutionary relationships of species in the subsection *Rumphiae* and the possibility of convergence of subsection *Rumphiae* species in other locations. As there are no *trnH-psbA* sequences available for the species *C. sainatii*, *C. dharmarajii*, *C. andamanica*, and *C. darshii*, and only a partial sequence of *trnH-psbA* is available for *C. pschannae* (GenBank accession MF977913) in GenBank database, we did not include these species in our phylogenetic analysis.

Cycas zeylanica is recorded only from Sri Lanka, Andaman and Nicobar Islands (evergreen littoral forests of Havelock Island, secondary littoral forest of West coast and beach vegetation of Kodiaghat in south Andaman)¹². This distribution overlaps with our niche model analysis and it is more disposed to climatic restriction hypothesis. The niche of *C. zeylanica* in Sri Lanka overlaps with that in other native countries¹². On the one hand, due to the favorable niche found in the lowland Wet Zone in Sri Lanka, it is possible for *C. zeylanica* to grow naturally in the area. On the other hand, it is also possible that *C. zeylanica* and other congeneric species of the subsection *Rumphiae* reached Sri Lanka through Long Distance Dispersal (LDD) by oceanic currents (Fig. 7) and colonized especially the coastal areas. Our proposal of multiple introductions of South East Asian cycads of the subsection *Rumphiae* forming *C. zeylanica* complex in Sri Lanka is confirmed by the ENM analysis too. The predicted distribution of this complex in South East Asia overlaps with the distributions of *C. sainatii*, *C. pschannae*, *C. dharmarajii*, *C. andamanica*, *C. darshii* and *C. edentata* and it is possible that these reached the island at different times.

Such LDD is presumed to affect the Sri Lankan flora. For instance, the diversity of some plant taxa of Sri Lanka is reported to have been originated by LDD from India, in the Paleozoic ice ages, through migratory animals⁶¹. Further, LDD has been reported for some other plant species such as mangroves (*Rhizophora apiculata* and *Rhizophora mucronata*⁶²) and coconut (*Cocos nucifera* that grows in Sri Lanka⁶³). However, this is the first record to support contribution of LDD in the floral diversification of cycads in Sri Lanka, and it is well known that cycads of the subsection *Rumphiae* undergo LDD due to the remarkable ability of its seeds to be buoyant in seawater¹⁹.

According to the current understanding, the dispersion of the subsection *Rumphiae* has occurred exclusively through LDD while South East Asia serves as the center of origin^{17–19}. The warm oceanic currents and continuous seasonal circulations present in the Indian oceanic gyre could aid long distance gene flow frequently (Fig. 7). Besides, there is a possibility that seeds were brought to Sri Lanka from South East Asia by merchants during Portuguese and Dutch colonization periods (16th to 17th century) and planted them in different locations in the lowland wet zone. This may be the reason for having very old cycad plants in historical places such as old catholic churches in the southern Wet Zone of Sri Lanka. However, the establishment of congeneric species within the country could support continuous hybridization so as to maintain a low genetic divergence within *C. zeylanica* populations in the country.

Our entropy modeling exhibits climatic envelop structure of the population indicating that the *C. zeylanica* complex has undergone a major climatic isolation event. Individuals of the *C. zeylanica* complex are clearly restricted to the coastal areas with tropical humid climates. Being an island surrounded by the Indian Ocean, there is a chance for *C. zeylanica* complex to reach other coastal regions of the country but during field explorations, we observed that the species has not established in other climatic regions of the country.

In Sri Lanka, *C. zeylanica* is declared as a Critically Endangered species⁶⁴ and the populations are also declining dramatically. Ramana *et al.*⁶⁵ suggested that when considering the conservation of *C. zeylanica* the samples must be collected from diverse habitats of Sri Lanka. Lindstrom and Hill¹² have reported that any viable regenerating populations of *C. zeylanica* in Sri Lanka is absent. In the current study, we have conducted an intense field

survey spanning the Wet Zone of Sri Lanka and comprehended that no natural populations of the complex exist other than the hybridized and scattered individuals. A couple of Tsunami events had occurred around Sri Lanka in the past, causing severe damages to the coastal regions of the country, and as a result, the natural populations might have been adversely affected. During our field sampling, we noticed that local people deliberately cultivated the majority of these individuals, especially those found in the interior of the island. However, the population size is rather small, with probably less than 200 individuals existing in the entire country at present, highlighting the importance of strict conservation of the taxa.

Conclusions

In the present study we renamed the Sri Lankan cycads of the subsection *Rumphiae* as *Cycas zeylanica* complex based on morphological and molecular sequence data. The higher morphological variation found in female reproductive structures distinguished the *C. zeylanica* complex *nov.* into six clusters; however, the genetic diversity among populations is very low. ENM suggests that the *C. zeylanica* complex prevails in South East Asia and the lowland Wet Zone of Sri Lanka. The morphological clustering suggests that long distance seed dispersal through sea currents and hybridization could have led to the formation *C. zeylanica* complex, and distribution within the region occurred due to human and animal interventions. The presence of a climatic envelop for the establishment of *C. zeylanica* complex in the country is evident. Due to the low population density and threats of extinction, *C. zeylanica* complex in Sri Lanka needs urgent conservation measures.

Data availability

The DNA sequences generated during the current study are available in GenBank® under the accession numbers MH458253-MH458269 and the other datasets produced during and/or analyzed are available from the corresponding author on reasonable request.

Received: 19 July 2019; Accepted: 22 October 2019;

Published online: 14 November 2019

References

1. Stevenson, D. W. A formal classification of the extant cycads. *Brittonia*. **44**, 220–223, <https://doi.org/10.2307/2807191> (1992).
2. Calonje, M., Stevenson, D. W. & Stanberg, L. The world list of cycads, <https://cycadlist.org> (2018).
3. Lindstrom, A. J. & Hill, K. D. The genus *Cycas* (Cycadaceae) in India. *Telopea*. **11**, 463–488 (2007).
4. de Laubenfels, D. J. & Adema, F. A taxonomic revision of the genera *Cycas* and *Epicycas* gen. *nov.* (Cycadaceae). *Blumea*. **43**, 351–400 (1998).
5. Hill, K. D. Infrageneric relationships, phylogeny and biogeography of the genus *Cycas* (Cycadaceae). In Proceedings of the third international conference on cycad biology, held in Pretoria, South Africa. 5–9 (1995).
6. Lindstrom, A. J., Hill, K. D. & Stanberg, L. C. The genus *Cycas* (Cycadaceae) in the Philippines. *Telopea*. **12**, 119–145, <https://doi.org/10.7751/telopea20085805> (2008).
7. Hill, K. D. The genus *Cycas* (Cycadaceae) in China. *Telopea*. **12**, 71–118 (2008).
8. Hill, K. D. The *Cycas rumphii* complex (Cycadaceae) in New Guinea and the western Pacific. *Aust Syst Bot.* **7**, 543–567, <https://doi.org/10.1071/SB9940543> (1994).
9. Lindstrom, A. J. & Hill, K. D. The genus *Cycas* (Cycadaceae) in India. *Telopea*. **11**, 463–488, <https://doi.org/10.7751/telopea20075745> (2007).
10. Hill, K. D. The *Cycas media* group (Cycadaceae) in New Guinea. *Aust Syst Bot.* **7**, 527–541, <https://doi.org/10.1071/SB9940527> (1994).
11. Lindstrom, A. J., Hill, K. D. & Stanberg, L. C. The genus *Cycas* (Cycadaceae) in Indonesia. *Telopea*. **12**, 385–418, <https://doi.org/10.7751/telopea20095826> (2009).
12. Lindstrom, A. J. & Hill, K. D. Notes on the species of *Cycas* (Cycadaceae) from Sri Lanka and Islands of the Andaman Sea. *Novon.* **12**, 237–240, <https://doi.org/10.2307/3392960> (2002).
13. Mudannayake, A., Sooriyapathirana, S., Samaraweera, P. & Perera, A. *Cycas* Taxa in Sri Lanka and their morphological characteristics of taxonomic significance. *Ceylon Journal of Science (Biological Sciences)*. **44**, <https://doi.org/10.4038/cjsbs.v44i1.7337> (2015).
14. Wadhwa, B. M. Cycadaceae. In *A revised handbook to the Flora of Ceylon* (ed. Dassanayake, M. D. & Clayton, W. D.) 300–304 (Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi, India, 2000).
15. Yang, S. L. & Meerow, A. W. The *Cycas pectinata* (Cycadaceae) complex: genetic structure and gene flow. *Int J Plant Sci.* **157**, 468–483, <https://doi.org/10.1086/297364> (1996).
16. Liu, N. The *Cycas taiwaniana* Complex in southeast China. *Proc. Acad. Conf. Cycad China (Gejiu)* (2004).
17. Dehgan, B. & Yuen, C. K. Seed morphology in relation to dispersal, evolution, and propagation of *Cycas* L. *Bot Gaz.* **144**, 412–418, <https://doi.org/10.1086/337391> (1983).
18. Keppel, G. Notes on the natural history of *Cycas seemanii* (Cycadaceae). *The South Pacific Journal of Natural and Applied Sciences.* **19**, 35–41, <https://doi.org/10.1071/SP01007> (2001).
19. Keppel, G., Hodgskiss, P. D. & Plunkett, G. M. Cycads in the insular South-west Pacific: dispersal or vicariance? *J Biogeogr.* **35**, 1004–1015, <https://doi.org/10.1111/j.1365-2699.2007.01869.x> (2008).
20. Lindstrom, A. Morphological characters useful in determining species boundaries in *Cycas* (Cycadaceae) in Cycad classification, concepts and recommendations (ed. Walters, T. Osborne, R.) 45–55 (Trowbridge: CABI Publishing, 2004).
21. Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. & Janzen, D. H. Use of DNA barcodes to identify flowering plants. *PProc Natl Acad Sci USA* **102**, 8369–8374, <https://doi.org/10.1073/pnas.0503123102> (2005).
22. Kress, W. J. & Erickson, D. L. A two-locus global DNA barcode for land plants: the coding *trnH-psbA* gene complements the non-coding *trnH-psbA* spacer region. *PLoS one.* **2**, e508, <https://doi.org/10.1371/journal.pone.0000508> (2007).
23. Pang, X. *et al.* Utility of the *trnH-psbA* intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. *PLoS One.* **7**, e48833, <https://doi.org/10.1371/journal.pone.0048833> (2012).
24. Feng, X., Liu, J. & Gong, X. Species delimitation of the *Cycas segmentifida* complex (Cycadaceae) resolved by phylogenetic and distance analyses of molecular data. *Front Plant Sci.* **7**, 134, <https://doi.org/10.3389/fpls.2016.00134> (2016).
25. Liu, J., Zhou, W. & Gong, X. Species delimitation, genetic diversity and population historical dynamics of *Cycas diannanensis* (Cycadaceae) occurring sympatrically in the Red River region of China. *Front Plant Sci.* **6**, 696, <https://doi.org/10.3389/fpls.2015.00696> (2015).

26. Zúñiga, J. D. *et al.* Data Release: DNA barcodes of plant species collected for the Global Genome Initiative for Gardens Program, National Museum of Natural History, Smithsonian Institution. *PhytoKeys*. **88**, 119, <https://doi.org/10.3897/phytokeys.88.14607> (2017).
27. Cooley, W. W. & Lohnes, P. R. *Multivariate data analysis*. (John Wiley and Sons Inc, New York, 1971).
28. Chavent, M., Kuentz-Simonet, V., Labenne, A. & Saracco, J. Multivariate analysis of mixed data: The PCAmixdata R package. arXiv e-prints arXiv:1411.4911v1 1–30, <https://arxiv.org/abs/1411.4911v1> (2019).
29. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org> (2017).
30. Sang, T., Crawford, D. J. & Stuessy, T. F. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). *Am J Bot.* **84**, 1120–1136, <https://doi.org/10.2307/2446155> (1997).
31. Tate, J. A. & Simpson, B. B. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany*. **28**, 723–737, <https://doi.org/10.1043/02-64.1> (2003).
32. Thompson, J. D., Higgins, D. G., Gibson, T. J. & Clustal, W. improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680, <https://doi.org/10.1093/nar/22.22.4673> (1994).
33. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol.* **33**, 1870–1874, <https://doi.org/10.1093/molbev/msw054> (2016).
34. Posada, D. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* **25**, 1253–1256, <https://doi.org/10.1093/molbev/msn083> (2008).
35. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 2010. 1–8. <http://www.ieeexplore.ieee.org/abstract/document/5676129> (2010).
36. Akaike, H. A new look at the statistical model identification. *IEEE transactions on automatic control*. **19**, 716–723, <https://doi.org/10.1109/TAC.1974.1100705> (1974).
37. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. **22**, 2688–2690, <https://doi.org/10.1093/bioinformatics/btl446> (2006).
38. Huelsenbeck, J. P. & Ronquist, F. MrBayes: Bayesian inference of phylogeny. *Bioinformatics*. **17**, 754–755, <http://brahms.biology.rochester.edu/software.html> (2001).
39. Rambaut, A. FigTree, a graphical viewer of a phylogenetic trees. <http://tree.bio.ed.ac.uk/software/figtree> (2014).
40. World climate, <http://www.worldclim.org/bioclim> (2016).
41. ESRI, R. ArcGIS desktop: release 10 [software]. Environmental Systems Research Institute, CA (2011).
42. Phillips, S. J., Dudík, M. & Schapire, R. E. A maximum entropy approach to species distribution modeling In *Proceedings of the twenty-first international conference on Machine learning* 83. (ACM, 2004).
43. Anderson, R. P. & Gonzalez, I. Jr. Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecol Modell.* **222**, 2796–2811, <https://doi.org/10.1016/j.ecolmodel.2011.04.011> (2011).
44. Mohammadi, S., Ebrahimi, E., Moghadam, M. S. & Bosso, L. Modelling current and future potential distributions of two desert jerboas under climate change in Iran. *Ecological Informatics*. **52**, 7–13 (2019).
45. Simões, M. V. & Peterson, A. T. Importance of biotic predictors in estimation of potential invasive areas: the example of the tortoise beetle *Eurypedus nigrosignatus*, in Hispaniola. *PeerJ* **6**, e6052 (2018).
46. Gerstner, B. E. *et al.* Revised distributional estimates for the recently discovered olinguito (*Bassaricyon neblina*), with comments on natural and taxonomic history. *Journal of Mammalogy* **99**, 321–332 (2018).
47. Bosso, L. *et al.* Loss of potential bat habitat following a severe wildfire: a model based rapid assessment. *International Journal of Wildland Fire* **27**, 756–769 (2018).
48. Bradley, B. A. *et al.* Cheatgrass (*Bromus tectorum*) distribution in the intermountain Western United States and its relationship to fire frequency, seasonality, and ignitions. *Biological Invasions* **20**, 1493–1506 (2018).
49. Franco, J. N. *et al.* The ‘golden kelp’ *Laminaria ochroleuca* under global change: Integrating multiple eco-physiological responses with species distribution models. *Journal of Ecology* **106**(1), 47–58 (2018).
50. Hill, K. D. A taxonomic revision of the genus *Cycas* (Cycadaceae) in Australia. *Telopea*. **7**, 1–64 (1996).
51. Hill, K. D. A preliminary account of *Cycas* (Cycadaceae) in Queensland. *Telopea*. **5**, 177–206, <https://doi.org/10.7751/telopea19924964> (1992).
52. Chiang, Y. C. *et al.* Asymmetric introgression in the horticultural living fossil *Cycas* Sect. *Asiorientales* using a genome-wide scanning approach. *Int J Mol Sci.* **14**, 8228–8251, <https://doi.org/10.3390/ijms14048228> (2013).
53. Wang, D. Y. Systematic classification and taxonomy in Cycads in China (ed. Wang, F. X. & Liang, H. B.) 9–142 (1996).
54. Hill, K. D. Character evolution, species recognition and classification concepts in the Cycadaceae in Cycad classification, concepts and recommendations (ed. Walters, T. & Osborne, R.). 23–44 (Trowbridge: CABI Publishing, 2004).
55. MacArthur, R. & Levins, R. The limiting similarity, convergence, and divergence of coexisting species. *Am Nat.* **101**, 377–385, <https://doi.org/10.1086/282505> (1967).
56. Srivastava, R. C. A new species of *Cycas* from India. *Indian Journal of Plant Sciences*. **3**, 109–110 (2014).
57. Srivastava, R. C. Notes on green diamonds of AJC Bose Indian Botanic Garden Howrah. *Indian Journal of Plant Sciences*. **3**, 151–153 (2014).
58. Srivastava, R. C. & Singh, L. A New Species of Indian *Cycas*. *Int. J. Curr. Res. Biosci. Plant Biol.* **2**, 35–37 (2015).
59. Prasad, K., Ramana, M. V., Rao, B. R. P. & Sanjappa, M. *Cycas andamanica* (Cycadaceae): a new species from Andaman Islands, India. *International Journal of Innovation Sciences and Research*. **4**, 473–476 (2015).
60. Singh, L. J. *Cycas dharmrajii* sp. nov. (Cycadaceae) from Andaman and Nicobar Islands, India. *Nordic Journal of Botany*. **35**, 69–76, <https://doi.org/10.1111/njb.01284> (2017).
61. Ranaweera, L. T., Hancock, J. F., Weebadde, C. K. & Sooriyapathirana, S. D. Phylogeographic and phylogenetic analyses of selected set of wild and naturalized *Solanum* spp. in Sri Lanka. *Ceylon Journal of Science*. **47**, 85–93 (2018).
62. Lo, E. Y., Duke, N. C. & Sun, M. Phylogeographic pattern of *Rhizophora* (Rhizophoraceae) reveals the importance of both vicariance and long-distance oceanic dispersal to modern mangrove distribution. *BMC Evol Biol.* **14**, 83, <https://doi.org/10.1186/1471-2148-14-83> (2014).
63. Gunn, B. F., Baudouin, L. & Olsen, K. M. Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. *Plos one*. **6**, e21143, <https://doi.org/10.1371/journal.pone.0021143> (2011).
64. MOE. The National Red List 2012 of Sri Lanka: Conservation status of the Fauna and Flora. The Biodiversity Secretariat and The Department of National Botanic Gardens, Sri Lanka (2012).
65. Ramana, M. V., Venu, P. & Sanjappa, M. Name confusions in Indian cycads. *Current Science*. **114**, 269 (2018).

Acknowledgements

We acknowledged Prof. Dhammika Magana-Arachchi, Dr. Sumudu Rubasinghe, Dr. Kanishka Ukuwela, Dr. Chandima Ariyaratna and Dr. Dimanthi Jayatilake for their valuable comments and suggestions given during the study. University of Peradeniya Research Grant (No. RG/AF 2013/72/S) is gratefully acknowledged for funding for the research.

Author contributions

The authors comply with the COPE guidelines, according to the journal rules. A.P., S.S. and P.S. conceived the study. A.M. and A.P. collected field samples. A.M. conducted laboratory work. A.M. and L.R. performed bioinformatic analyses. A.P. supervised the analyses of herbarium specimens and provided details about morphology and taxonomic descriptions. L.R. produced maps and graphical artworks. All authors wrote the manuscript with significant contributions. This study was supported by A.P.'s research funds.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-53011-w>.

Correspondence and requests for materials should be addressed to A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019