New Phytologist

Research

Congruence between ocean-dispersal modelling and phylogeography explains recent evolutionary history of *Cycas* species with buoyant seeds

Jian Liu[1,2], Anders J. Lindstrom[3], Yong-Sheng Chen[4], Ran Nathan[5] and Xun Gong[1,2,6]

[1]CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China; [2]Department of Economic Plants and Biotechnology, Yunnan Key Laboratory for Wild Plant Resources, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China; [3]Global Biodiversity Conservancy, 144/124 Moo3, Soi Bua Thong, Bangsalae, Sattahip, Chonburi 20250, Thailand; [4]Institute of Ecology and Key Laboratory for Earth Surface Processes of the Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China; [5]Movement Ecology Laboratory, Department of Ecology, Evolution and Behavior, Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel; [6]University of Chinese Academy of Sciences, Beijing 100049, China

Authors for correspondence:
Jian Liu
Email: liujian@mail.kib.ac.cn

Xun Gong
Email: gongxun@mail.kib.ac.cn

Received: *8 June 2021*
Accepted: *29 July 2021*

New Phytologist (2021) **232:** 1863–1875
doi: 10.1111/nph.17663

Key words: Buoyancy, cycads, long-distance dispersal, ocean drift modelling, phylogenomics, phylogeography.

Summary

- Ocean currents play a significant role in driving the long-distance dispersal (LDD), spatial distribution and phylogeographic patterns of many organisms. Integrating phylogeographic analyses and mechanistic ocean current modelling can provide novel insights into the evolutionary history of terrestrial littoral species but has been rarely applied in this context.
- We focused on a group of *Cycas* that have buoyant seeds and occupy coastal habitats. By integrating evidence from mechanistic simulations and whole plastomic data, we examined the role of ocean circulation in shaping the phylogeography of these *Cycas* species.
- Plastomes of the studied *Cycas* species showed extreme conservatism, following a post-Pleistocene divergence. Phylogenies revealed three subclades, corresponding to the Pacific Ocean, Sunda Shelf and Indian Ocean. The ocean modelling results indicate that hotspots of seed stranding coincide well with the contemporary distribution of the *Cycas* species and that drifting trajectories from the three subclades are largely confined to separate regions.
- These findings suggest that ocean current systems, by driving long-distance dispersal, have shaped the distribution and phylogeography for *Cycas* with buoyant seeds. This study highlights how the combination of genomic data and ocean drift modelling can help explain phylogeographic patterns and diversity in terrestrial littoral ecosystems.

Introduction

Ocean currents, driven by global wind systems, play a key role in mediating long-distance dispersal (LDD) and maintaining connectivity in many terrestrial organisms (Renner, 2004; Ali & Huber, 2010). Transoceanic LDD enables colonization of unoccupied habitats and large-scale exchange of individuals between remote populations (Nathan *et al.*, 2008), and thereby is considered a key driver of the genetic and spatial structure of populations and communities across multiple scales (Levin *et al.*, 2003; Excoffier *et al.*, 2009; Gillespie *et al.*, 2012; Kremer *et al.*, 2012; Nathan, 2013; Gallaher *et al.*, 2017). Based on the classic isolation-by-distance (IBD) model, higher levels of long- vs. short-distance gene flow act to reduce or increase local genetic correlation, respectively (Wright, 1969), offering a framework for analyzing how phylogeographic patterns develop at different scales.

Recent studies that integrated phylogeographic inference, numerical ocean circulation models, and Lagrangian simulation of individual particle trajectories (van Sebille *et al.*, 2018) have greatly improved our knowledge of the ecological factors driving evolution in the oceans. These include studies of genetic structure (Bertola *et al.*, 2020; Sefc *et al.*, 2020), connectivity and establishment (Jensen *et al.*, 2020; Nikolic *et al.*, 2020) of marine organisms. For terrestrial plants, mechanistic dispersal simulations also have been used to predict propagule transoceanic dispersal of mangroves (Van der Stocken *et al.*, 2019), seagrasses (Smith *et al.*, 2018) and Urticaceae (Wu *et al.*, 2018), and to confirm the origin of pre-Columbian bottle gourds (Kistler *et al.*, 2014). This mode of transport is particularly relevant for littoral plants that tend to be widely distributed and often are equipped with buoyant seeds, fruits, or vegetative propagules. Although the role of transoceanic LDD as the driving mechanism for shaping phylogeography of these taxa is well-recognized (Miryeganeh *et al.*, 2014; Gallaher *et al.*, 2017; Guo *et al.*, 2018; Banerjee *et al.*, 2020; Wee *et al.*, 2020), only a few previous studies (e.g. Wee *et al.*, 2014) have integrated comprehensive phylogenetic and ocean-circulation modelling into explanations of the evolutionary history of littoral plants. Therefore, mechanistic assessment of the genetic consequences of LDD by ocean currents (e.g.

metacommunity-scale phylogeographic pattern) in terrestrial littoral ecosystems is still rather limited.

Cycads are the most ancient living seed plant lineage (Brenner *et al.*, 2003), considered to be living fossils that have undergone recent radiation (Nagalingum *et al.*, 2011). Extant cycads include approximately 364 species, and the genus *Cycas* (Cycadaceae) is the most diversified group within cycads, containing 120 species (Calonje *et al.*, 2021). Cycads occur in various ecosystems, including wet (rainforest, littoral plants), arid and semi-arid conditions (xerophytes). Most cycad species are distributed inland, often having heavy seeds that sink in water, implying limited water-dispersal ability. However, 10 *Cycas* species possess buoyant seeds (hereafter referred to as buoyant *Cycas*) and are classified as section *Cycas* subsection *Rumphiae*, which is characterized anatomically by the presence of a distinct spongy layer in the seed (Hill, 1994). Facilitated by this structure, viable seeds of the buoyant *Cycas* float vertically with the hilum always oriented upwards (Fig. 1, inset). Some species such as *C. seemanii* also have developed an alate flattened extension at the micropyle end, which serves as a keel and balances the seed perfectly in the water, thus creating a stable environment for embryo development (Lindstrom, pers. obs.).

Buoyant *Cycas* represents a complex of closely related widely distributed taxa, ranging from Indochina and Malesia (an area that includes the Malay Peninsula, Indonesia, New Guinea and the Philippines), north to coastal South Indochina, south to New Guinea, west to East Africa, and as far east as Fiji and Tonga (Figs 1, S1). Previous studies involving this group used allozyme (Keppel *et al.*, 2008) and multilocus (Liu *et al.*, 2018) data to reveal high genetic similarity within this subsection, which echoed an earlier casted hypothesis that the distribution of buoyant *Cycas* resulted from LDD by ocean currents (Dehgan & Yuen, 1983; Hill, 1996). Given its distinct propagules and extensive ranges, buoyant *Cycas* are highly suitable to test the hypothesis that phylogeographic pattern follows ocean currents among regions. However, quantitative evidence based on the integration of phylogeographic analyses and mechanistic oceanic transport is still lacking.

In order to better understand the role of ocean currents in maintaining connectivity and shaping the phylogeographic pattern of terrestrial littoral plants, we focused on the buoyant *Cycas* group and addressed three major questions: (1) Is the phylogeographic pattern of buoyant *Cycas* congruent with predictions from ocean current modelling? (2) Can ocean current simulations predict the contemporary distribution of buoyant *Cycas*? and (3) Did LDD and past climate oscillations leave a genetic signature on plastomes, as *Cycas* plastids are maternally inherited (Zhong *et al.*, 2011) and likely infer seed-mediated gene flow? To address

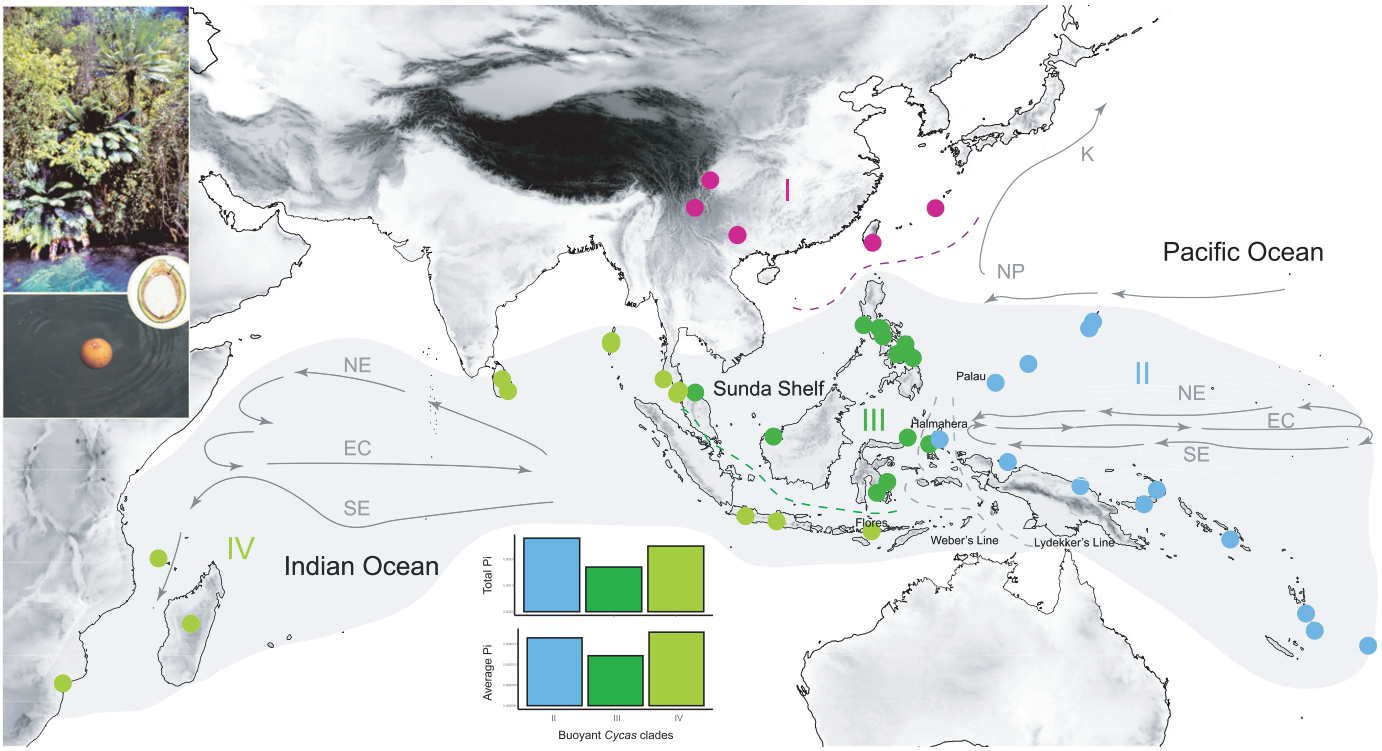

Fig. 1 The phylogeographic pattern of buoyant *Cycas*. Purple dots represent inland *Cycas* that are unable to float on water (refer to clade I in Fig. 2). Blue, dark green, light green dots denote the distribution of lineages from three subclades II, III, IV, respectively, as revealed by phylogenetic reconstruction of buoyant *Cycas* (Fig. 2). Note that the *C. seemanii* from New Caledonia (Voucher 19450) is not shown because the source specimen probably is of mixed origin. The shaded area indicates the approximate distribution range of buoyant *Cycas* lineages. The inset bar plot represents the nucleotide diversity (π) level of three clades (II–IV) of buoyant *Cycas*. Insets at the top left show the coastal habitat of buoyant *Cycas*, and the buoyancy and vertical section of its seed. The spongy layer is indicated by an arrow in the inset. The major ocean currents in the map are modified from Gordon & Cenedese (2018). NE, North Equatorial Current; EC, Equatorial Countercurrent; SE, South Equatorial Current; NP, North Pacific Current; K, Kuroshio Current.

these questions, we first sequenced plastomes and evaluated differentiation at the genomic scale of all extant buoyant *Cycas*. Then we investigated the phylogeographic pattern and evolutionary history of buoyant *Cycas* and used ocean drift modelling to evaluate the effect of ocean currents in mediating the contemporary distribution and divergence of buoyant *Cycas*. Last, we used plastomic data to estimate the genetic diversity and selective pressure among different phylogeographical lineages.

Materials and Methods

Sampling

Taxonomy follows the accepted names of the World Cycad List (Calonje *et al.*, 2021) and sampling included all species in the section *Cycas* subsection *Rumphiae* (Hill, 1994). A total of 45 accessions from different localities, representing 10 species and the entire range of buoyant *Cycas*, were sampled (Table S1). Notably, there are a couple of *Cycas* species (Table S2) that possess spongy tissue that is not extensive enough to make the seed float on water. These species may represent transitional lineages or the result of hybridization with spongy layer-deficient subsection *Rumphiae* (Lindstrom *et al.*, 2009); thus we did not include these taxa in the sampling. We added 17 published plastomes, including six nonbuoyant *Cycas* representing three *Cycas* sections (*Panzhihuaensis*, *Asiorientales* and *Stangerioides*), the other nine genera (Zamiaceae) from Cycadales, and two *Ginkgo* accessions into our sampling.

DNA extraction, plastome sequencing, assembly and annotation

Total genomic DNA was extracted from silica gel-dried materials by the modified CTAB method (Doyle, 1991). A total of 2G of sequencing data from the Illumina Hiseq Platform (Novogene, Beijing, China) were filtered and used for plastome assembly in the *get_organelle* pipeline (Jin *et al.*, 2020) by using *Cycas szechuanensis* (NC042668) as reference. The resulting contigs were visualized, trimmed and edited further in Bandage v.0.7.1 (Wick *et al.*, 2015) to obtain the quadrantal structure contigs. We applied both the GeSeq (Tillich *et al.*, 2017) and the PGA pipelines (Qu *et al.*, 2019) to annotate the plastid genomes using *C. szechuanensis* as reference. The annotations were compared, double-checked and adjusted in Geneious prime v.2020 (Kearse *et al.*, 2012). The plastome graph was visualized in OGDraw (Greiner *et al.*, 2019).

Plastome structural variation analyses

In order to identify regions with substantial variability within buoyant *Cycas* species, we chose one accession from each of the 10 buoyant *Cycas* species based on the phylogenetic results. We first compared the global alignment of the complete chloroplast genomes using mVista (Frazer *et al.*, 2004), with *C. szechuanensis* (NC042668) as a reference. We also used the IRscope script (Amiryousefi *et al.*, 2018) in R v.3.6.3 (R Core Team, 2020) to generate and compare the variation of inverted-repeat (IR) and single-copy (SC) borders of the 10 surveyed buoyant *Cycas* species.

Phylogenetic analyses

Phylogenetic reconstructions were implemented in IQtree v.2.1.1 (Minh *et al.*, 2020) to infer the maximum-likelihood (ML) tree using the ultrafast bootstrap approximation method (Hoang *et al.*, 2018) with 1000 replicates. The TIM+F+R2 model was determined to be the best substitution model in ModelFinder (Kalyaanamoorthy *et al.*, 2017) under the default Bayesian information criterion (BIC). To validate the result of IQtree, we also performed a rapid bootstrap analysis in RAxML v.8.2.12 with 1000 bootstraps under the Gtrgamma substitution model (Stamatakis, 2014). For both phylogenetic inferences, we ran the analyses based on two datasets: protein-coding regions and the whole plastomic dataset.

Divergence time estimation

Age estimation was implemented in the Beast package v.2.6.1 (Bouckaert *et al.*, 2019) based on the protein-coding region dataset. We employed two fossil calibrations from a comprehensive dating analysis of all cycads based on six fossils (Condamine *et al.*, 2015). We did not incorporate all six fossils because our study focused only on *Cycas* and not all cycads. Two of these fossils, *Crossozamia chinensis* (Gao & Thomas, 1989) and *Antarcticycas schopfii* (Hermsen *et al.*, 2006), are the oldest known Cycadophyta fossils and the closest relatives to extant cycads based on phylogenetic studies (Martínez *et al.*, 2012). We applied these two fossils to the stem and crown nodes of Cycadales using uniform prior distributions, following Condamine *et al.* (2015). For the stem node, we used a range of 265.1–364.7 Ma (Myr ago), in which 265.1 Ma is the minimum age of Shihhotse Formation, Lower Permian where *Crossozamia* was found, and 364.7 Ma is the upper boundary of VCo (versabilis-cornuta) spore Biozone Formation where the first known record of seed in the form of preovule (*Elkinsia polymorpha*) was discovered (Rothwell *et al.*, 1989), and was used to constrain the origin of gymnosperms. For the crown node, we used a range of 235.0–364.7 Ma, in which 235.0 Ma is the conservative age of the early Middle Triassic and is the age ascribed to the Fremouw Formation where *A. schopfii* occurs (Gradstein *et al.*, 2012), and 364.7 Ma is again the upper boundary of the origin of gymnosperms.

The recommended substitution model in Beast analyses was determined by PartitionFinder2 (Lanfear *et al.*, 2017), yielding GTR + G + I. We chose the uncorrelated lognormal relaxed-clock model as the clock model because it was favored by the prior nested sampling (NS) model selection test in Beast. The choice of the branching process prior can have a drastic influence on the ages of clades, especially for lineages with a history of high extinction, so we adopted a birth–death prior here, as employed and recommended by previous analyses of cycads (Nagalingum *et al.*, 2011; Condamine *et al.*, 2015). Twenty independent

searches were run for 100 million generations each and the log files were subsequently combined by LOGCOMBINER v.2.6.1 (Bouckaert *et al.*, 2019) in BEAST to reach an effective sample size (ESS > 200). The log output files were evaluated in TRACER v.1.7 (Rambaut *et al.*, 2018). In total, two billion iterations were run in BEAST, and it generated 40 000 trees by sampling the log and tree every 50 000 generations. The first 25% of trees (10 000 trees) were discarded as a burn-in before generating maximum clade credibility (MCC) consensus species tree in TREEANNOTATOR (Bouckaert *et al.*, 2019). The consensus MCC tree with estimated ages was visualized in FIGTREE v.1.4.4 (Rambaut *et al.*, 2018).

Estimation of substitution rates and nucleotide diversity

Genes in the plastome encode proteins and several types of RNA molecules, which impact plant metabolism and can consequently undergo selective pressures. To estimate the signatures of selective pressure among different groups, we concatenated the protein-coding genes shared by Cycadaceae, Zamiaceae and Ginkgoaceae to generate a dataset with 57 456 characters after deleting gaps. We then estimated the synonymous (d*s*) and nonsynonymous (d*n*) substitution rates of the examined taxa using the codeml program of PAML v.4.8 (Yang, 2007) by using *Zamia furfuracea* as a reference. The pairwise d*n* and d*s* substitution rates between different taxa were calculated based on the custom selection model by setting the *CodonFreq* prior as *codon table*. The d*n*/d*s* value (ratio of nonsynonymous to synonymous substitution rates) then was calculated for each accession and further compared between pairwise groups. We also used a Student's *t*-test to detect if there were significant differences in d*n*/d*s* rates between different groups (i.e. all buoyant *Cycas* with nonbuoyant *Cycas*, and between three buoyant subclades). To compare the genetic diversity level of buoyant *Cycas* from different regions, we used DNASP v.5 (Librado & Rozas, 2009) to calculate the Nei's nucleotide diversity (π) of the buoyant subclades revealed by phylogenetic reconstruction that is based on the complete plastomic alignment.

Mantel test between geographical and genetic distance

In order to test the association between physical factors (i.e. geographical distance) and phylogeographic pattern, termed the isolation-by-distance (IBD) effect, in the process of transoceanic dispersal of buoyant *Cycas*, we performed a Mantel test to quantify the correlation between pairwise genetic distance and geographical distance of all 45 buoyant *Cycas* in this study. All analyses were performed in R v.3.6.3 (R Core Team, 2020). First, we estimated the relative geographical distance between each accession based on their sampling locality (Table S1) using the STATS package in R. Then we employed the package APE v.5.4 (Paradis & Schliep, 2018) to calculate the pairwise genetic distance of each accession based on the complete plastomic sequences under the default F80 model. Lastly, the Mantel test was performed by the package ADE4 v.1.7 (Dray & Dufour, 2007) with 10 000 permutations to generate the observation *r*-value as well as simulated *P*-value, and we used GGPMISC v.0.3.8 (Aphalo, 2020) and GGPLOT2 v.3.3.3 (Wickham, 2016) for simple linear regression and plotting.

Ocean drift modelling

We used an individual-based Lagrangian particle model (Bennett, 2006) to predict the possible trajectories and stranding events of passively drifting individuals (*Cycas* seeds herein). The modelled surface current fields we used here were from the Global Ocean Ensemble Physics Reanalysis (GOEPR) which was developed by the Copernicus Marine Environment Monitoring Service (CMEMS: https://marine.copernicus.eu/). A high spatial and temporal resolution of the ocean model resolves mesoscale eddies, which can strongly affect the dispersal and connectivity of drifting organisms. The GOEPR model has a high horizontal resolution of 0.25° and 75 levels of vertical coverage, starting from 1 January 1993 and providing daily-mean/monthly-mean, creating a close-to-reality, 3D simulation of World Ocean dynamics as it assimilates satellite altimetry, temperature, salinity measurements and *in situ* observations. Model-based and observational studies indicate that major surface current directions in Indo-Pacific remained relatively unchanged in the Pliocene compared to the modern (Haywood & Valdes, 2004), and most continents were essentially at their present positions during that period (Hall, 2009). We used the contemporary GREPV2-DAILY ocean current dataset of 2010 as our simulation input; the option of the year 2010 was arbitrary. For the wind factor in our passive ocean surface drift modelling, we applied the global atmospheric model to obtain ocean-wind variables from National Centers for Environmental Prediction (NCEP, source from https://pae-paha.pacioos.hawaii.edu).

The passive drift trajectories of individuals were computed using the *OceanDrift* model implemented in the Python-based Lagrangian trajectory simulation framework OPENDRIFT (Dagestad *et al.*, 2018) (available on https://github.com/OpenDrift). We seeded a 2% fraction as the factor of wind drift (wind speed at which elements will be advected) which is the default setting in OPENDRIFT. Three subregions were predefined based on the whole range of buoyant *Cycas* according to the phylogenetic results: the Pacific Ocean (Pacific), the Sunda Shelf (Sunda) and the Indian Ocean (Indian). We first ran the general simulations based on a total of 10 000 particles (buoyant seeds) from 25 sites representing the global occurrence across all three subregions based on the present distribution of buoyant *Cycas* (nine, seven and nine from Pacific, Sunda and Indian subregions, respectively; Table S3). Then the passive drift trajectories of individuals from the three subregions were simulated. For all simulations described above, 400–500 particles were released from each locality, representing the rough seed reproduction from one mature female cone. Cycads are rather variable in their coning season (Griffith *et al.*, 2012). The fruiting seasons and the centralized months of seed maturation for *Cycas* are from June to February (based on our field observation). Hence, in this study, we set two releasing dates for passive drift simulation as 1 July and 1 January. Evidence shows that it takes 10–18 months after release from cones

for sufficient maturity before germination in most species of the genus *Cycas* subsection *Rumphiae* (Lindstrom AJ, unpublished data), and early observations and experiments on the buoyant *Cycas* in Fiji showed they could float on water for months (Guppy, 1906). To make the simulation computationally inexpensive, we simulated each drifting process for a conservative 180-d period and recorded the trajectories every 6 h with a final output time step of 12 h. Although our simulation parameters are idealized, the objective of this study was to provide a representation of the potential dispersal for buoyant *Cycas* under contemporary ocean currents.

Additionally, to evaluate the sensitivity of current velocity uncertainty in influencing the seed drifting, we used the constant current module in OpenDrift by setting the seawater velocity as zero and simulated the trajectories of particles under two uncertainty levels: a low threshold value of 0.5 m s^{-1} and a high one of 2 m s^{-1}. The releasing sites, numbers and simulating time of particles were set the same as described above.

Results

Plastome assembly and structural characterization

The plastomes of buoyant *Cycas* displayed identical plastomic features in annotation, with a total of 133 genes annotated (87 protein-coding genes, 37 tRNA, eight rRNA and a nonfunctional *tufA* gene). All plastomes were highly conserved and displayed typical quadrantal structure within *Cycas*, with neither large structural variation (inversion, deletion) nor gene loss detected (Figs S2,S3). Notably, we found a 317-bp deletion around position 118k, and this indel was shared by all samples from the Indian Ocean coast (see *C. zeylanica* and *C. thouarsii* in Fig. S3). A slight IR expansion (> 50 bp) to LSC was detected across all buoyant *Cycas* compared with the inland species *C. szechuanensis* (Fig. S4), yet this expansion was conserved within the buoyant group. The Pacific *Cycas* species – namely *C. seemanii*, *C. micronesica* and *C. bougainvilleana* – shared an 8-bp sequence in IRb from the *trn*L gene, whereas the remaining buoyant *Cycas* shared a length of 9 bp (Fig. S4).

Phylogenies and divergence time estimation

IQtree generated identical topologies based on the complete plastomic dataset and protein-coding gene dataset, but the complete dataset yielded greater bootstrap values in many nodes (Fig. S5a,b). The best tree inferred by RAxML based on the complete plastomic dataset was consistent with IQtree results (Fig. S5c), whereas the RAxML topology based on the protein-coding gene dataset differed from the above three in several clades (Fig. S5d). Thus, the ML topology inferred by IQtree based on a complete plastomic dataset is used as the favored ML tree in the subsequent discussion. In this phylogenetic tree, all six accessions from three *Cycas* sections are resolved as sister to the buoyant *Cycas* clade. Within the buoyant *Cycas* clade, many accessions from the same morphologically recognized species did not cluster together, because of the use of the uniparentally inherited marker. For example, *C. seemanii* from New Caledonia (Voucher 19450) is not close to other *C. seemanii* lineages (Fig. S5), which probably is caused by a cultivation mixture. Nevertheless, we obtained three strongly supported subclades within the buoyant *Cycas* clade, which correspond to the geographical subregions of Pacific Ocean (Pacific), Sunda Shelf (Sunda) and Indian Ocean (Indian) (Fig. 1).

The MCC tree generated by Beast analysis resolves most of the deep nodes of buoyant *Cycas* (Fig. 2), whereas it conflicts with the ML tree in several shallow nodes where the PP and BP values are < 0.9 and 90, respectively (Fig. S6). Despite this, all three subclades of buoyant *Cycas* generated by the ML tree are strongly supported in the MCC tree (Figs 2, S6). Based on the divergence time estimation, the *Cycas* crown was estimated as *c.* 10 Ma (95% HPD (highest posterior density) 16.60–4.51 Ma), after a long branch divergence with its sister Zamiaceae (Fig. 2). The divergence of three subclades of buoyant *Cycas* was estimated to be in the early Pliocene (4.81 Ma, 95%HPD 7.94–2.00 Ma), and the dispersal and diversification of buoyant *Cycas* within each region occurred in the Quaternary (2.44 Ma (95%HPD 4.33–0.49 Ma), 1.34 Ma (95%HPD 2.30–0.50 Ma) and 1.98 Ma (95%HPD 3.46–0.54 Ma) for clades II (Pacific), III (Sunda) and IV (Indian) respectively]. It is noteworthy that there are two independent lineages from both clades II and IV (Fig. 2, *C. micronesica* from Palau and *C. sundaica* from Flores, respectively), which makes the clade age estimation more than two-fold greater than that by excluding the two independent taxa (2.44 vs 1.12 Ma and 1.98 vs 0.84 Ma, respectively).

Estimation of d*n*/d*s* substitution rates, nucleotide diversity and Mantel test

All *Cycas* species are extremely conserved in evolutionary rates and shared very close d*n* and d*s* substitution rates (Fig. 3a). Specifically, the inland nonbuoyant *Cycas* are more scattered than buoyant *Cycas* in the d*n* vs d*s* plot. Samples from three subclades of buoyant *Cycas* aggregate separately, and subclades III and IV are closer to each other than to subclade II, which is consistent with the phylogeny. We found no significant difference in d*n*/d*s* between nonbuoyant *Cycas* and the buoyant clade, whereas a significantly greater d*n*/d*s* value was detected in buoyant subclade III (Sunda Shelf) than in subclades IV and II. By contrast, the nucleotide diversity was lowest in subclade III compared to subclades II and IV, regardless of the total or average diversity level (Table 1; Fig. 1). The average nucleotide diversity level remained the lowest in the Sunda Shelf clade compared to the core Pacific Ocean clade and core Indian Ocean clade (Table S4; Fig. 2). A relatively strong positive correlation between genetic and geographical distance was detected ($r = 0.43$; $P < 0.0001$; Fig. 4).

Ocean drift modelling

According to the ocean drift modelling results when starting from different months (Figs 5, S7), no great difference was found between the two sets of simulations except the stranding on Taiwan and Madagascar islands. Generally, the July-initiated

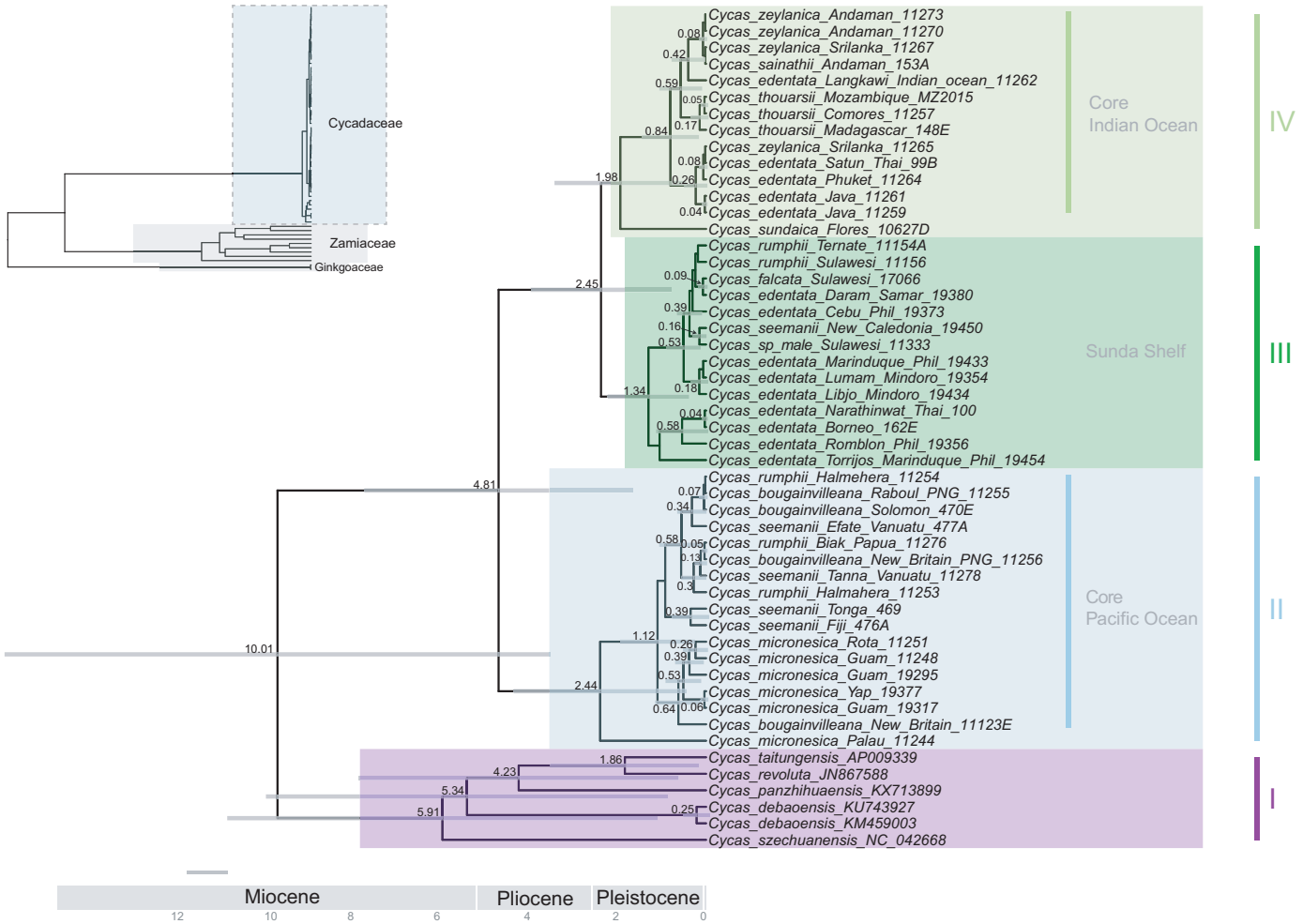

Fig. 2 Chronogram of *Cycas* reconstructed by Beast based on protein-coding regions. Clade I denotes the inland *Cycas* that are unable to float on water. Clades II, III and IV represent the buoyant *Cycas* group. The colors of different clades are consistent with those in Figs 1, 3. Inset is the chronogram including all outgroups. The numbers beside the nodes and below the timescale bar refer to age time points (Myr ago (Ma)). The last geological period, the Quaternary, is not indicated in the timescale.

simulation of buoyant *Cycas* better corresponded to its contemporary distribution (Fig. S1) as a consequence of simulated colonization events occurring in Madagascar but not in Taiwan (Fig. 5), unlike in the January-initiated simulations (Fig. S7). Therefore, we will focus on the July-initiated simulation results in the following interpretations.

After a half-year simulation, *c.* 72% of particles had met the coastline, mostly in the Philippines, Papua New Guinea, Java, Sumatra and East Africa (Fig. 5a). This simulation, although based on a set of simplified priors, matches well with the observed occurrence of buoyant *Cycas* (Fig. 1). Significantly, we found there were very few overlapping trajectories among the three predefined subregions, and the stranded events mostly were restricted in each subregion (especially in the Sunda region; Fig. 5). Few exceptions were the trajectories around Palau (Fig. 5b, from Sunda to Pacific) and eastern Borneo (Fig. 5c, from Pacific to Sunda). Additionally, strong westward trajectories were detected from Sri Lanka/Sumatra to East Africa (Fig. 5a,b). The passive drift trajectories in the Pacific Ocean based on July-initiated simulations displayed a general westward tendency (Fig. 5d) but were weaker than those initiated in January (Fig. S7d).

The simulation result using a low level of ocean current velocity uncertainty shows most seeds drifting around the release sites with connectivity rarely established within a region (0.5 m s^{-1} scheme; Fig. S8a), whereas using a higher velocity uncertainty produced densely stranded particles along the coasts with few particles randomly drifting in the ocean (2 m s^{-1} scheme; Fig. S8b).

Discussion

In this study, we revealed low plastomic variation but a clear genetic structure of the buoyant *Cycas*, and congruence between ocean current modelling and both the phylogeographic patterns and the contemporary distribution of these species. Furthermore, the significantly lower selective pressure but the higher genetic diversity detected in the Pacific and Indian Ocean clades than the Sunda Shelf clade suggests adaptation to new niches after transoceanic dispersal. Although there are limitations to the two approaches used (e.g. idealized parameters in ocean drift

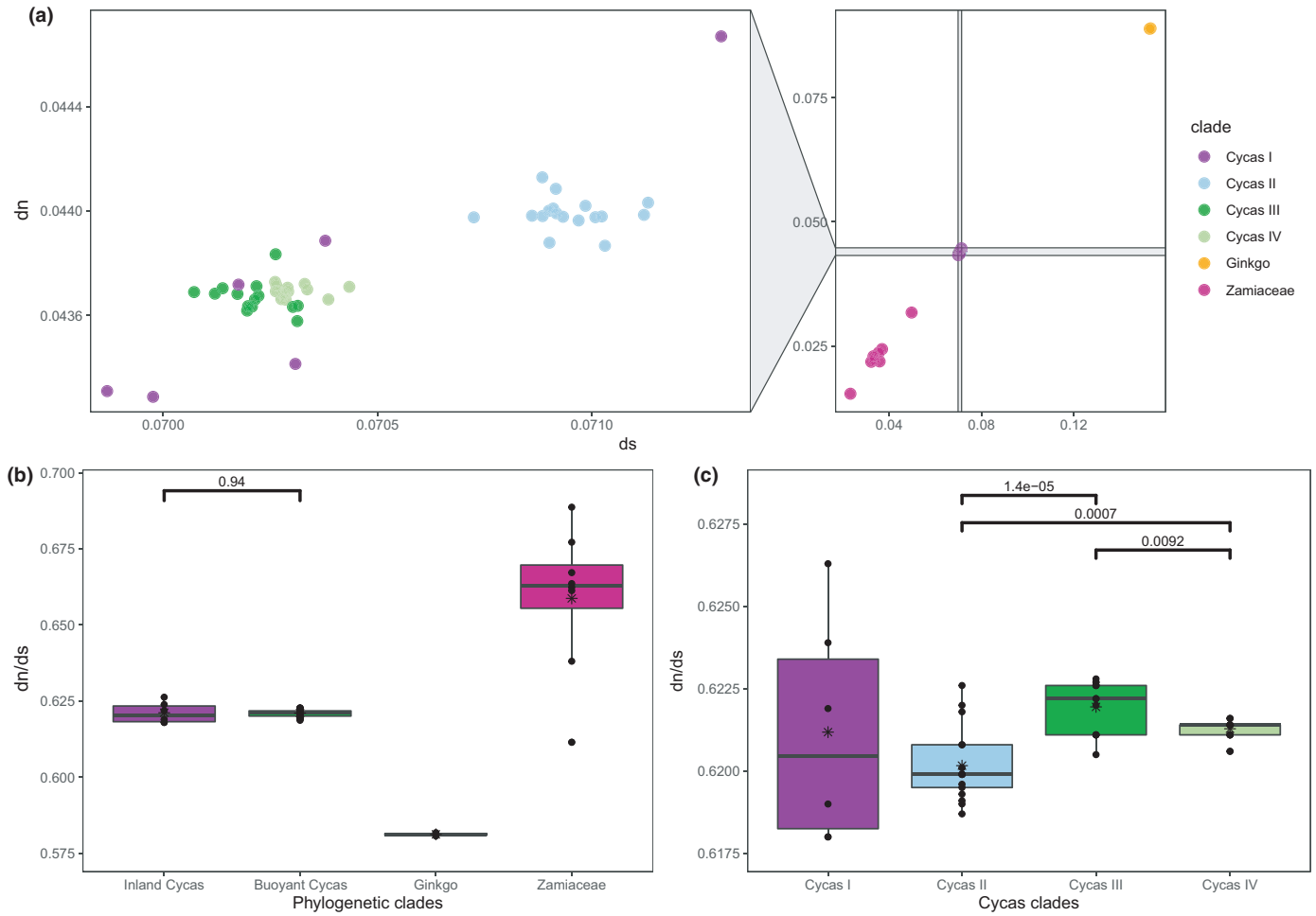

Fig. 3 Substitution rates and selection pressure analyses among different groups. The *Cycas* clades I–IV shown here are consistent with those revealed by the phylogeny in Fig. 2. (a) Comparison of synonymous (d*s*) and nonsynonymous (d*n*) substitution rates among Cycadaceae and its sister clades. Each group is color-coded. It is of note that all *Cycas* evolve extremely slow substitution rates in their plastome protein-coding sequences. (b) Comparisons of d*n*/d*s* ratios between different groups (Inland *Cycas*, buoyant *Cycas*, Zamiaceae, and Ginkgoaceae). (c) Comparisons of d*n*/d*s* between different phylogenetic clades of buoyant *Cycas*. (b–c) Box-and-whisker plots indicate the median (horizontal line), 25th and 75th percentiles (bottom and top of the box), and limits of the 95% confidence intervals (lower and upper whiskers) of the d*n*/d*s* rates. Dots beyond the 95% confidence intervals are outliers.

Table 1 Summarized information of the three subclades from buoyant *Cycas*.

Clades	Defined region	Number of Accessions	Number of Species	Total nucleotide diversity	Average nucleotide diversity	Average d*n*/d*s*	Divergence time (Ma)
II	Pacific Ocean	17	4	0.00028	1.65E-5	0.6202	2.44
III	Sunda Shelf	14	4	0.00017	1.21E-5	0.6220	1.34
IV	Indian Ocean	14	5	0.00025	1.79E-5	0.6213	1.98

D*n*, nonsynonymous substitution rates; d*s*, synonymous substitution rates; Ma, Myr ago.

modelling and phylogeography inferred based only on plastomic data), our results suggest a primary role for ocean currents in shaping the distribution and phylogeographic pattern of this terrestrial plant group. In the following, we further discuss the probability and processes of transoceanic long-distance dispersal (LDD) in buoyant *Cycas*, the level of congruence in inferring from phylogenetics vs. mechanistic ocean-drift simulations, the genomic signature on buoyant *Cycas*, and the main implications of ocean modelling.

Transoceanic LDD and connectivity establishment in buoyant Cycas

The low genetic variation, highly conserved plastome structure, and post-Pliocene divergence for the *Cycas* between the Pacific and the Indian Oceans further confirmed the importance of LDD in this group by previous studies (Dehgan & Yuen, 1983; Hill, 1996; Keppel *et al.*, 2008; Xiao & Möller, 2015; Mankga *et al.*, 2020). Because *Cycas* seeds are poisonous and heavy

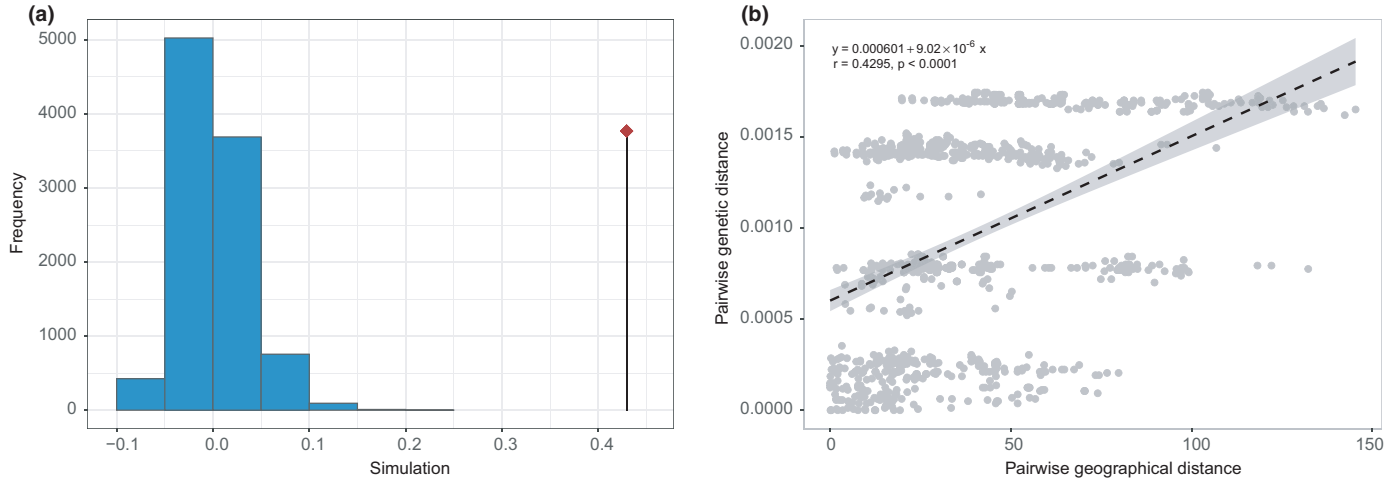

Fig. 4 Mantel test analyses of buoyant *Cycas*. (a) Mantel test correlation result, the original value of the correlation between the distance matrices is represented by the red dot, whereas histograms represent permuted values (i.e. under the absence of spatial structure; here the isolation-by-distance is significant). (b) Plot of geographical distance against genetic distance for 45 buoyant *Cycas*.

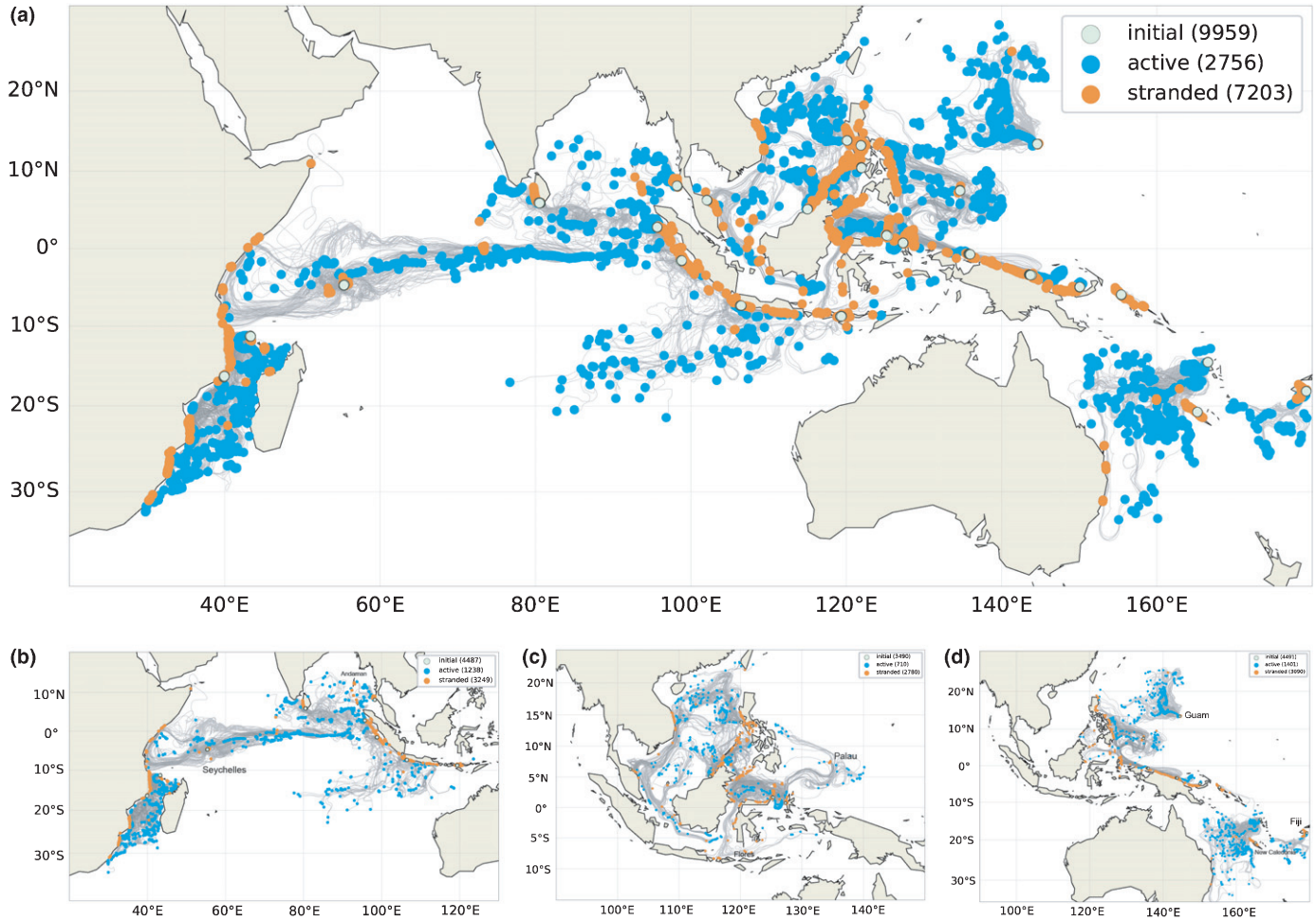

Fig. 5 Simulated stranding (orange) and active (blue) distribution of buoyant *Cycas* propagules across the Indo-Pacific. Trajectories are represented as gray lines and were generated using velocity fields from a high-resolution GOEPR ocean model simulation. Particles were released hourly for 180 d (1 July 2010 to 28 December 2010) from (a) a total of 25 sites globally; (b) nine sites from the Indian Ocean coast; (c) seven sites around the Sunda Shelf; (d) nine sites from the Pacific Ocean coast.

(Bradley & Mash, 2009; Marler *et al.*, 2010), birds and winds are very unlikely to disperse them over such long distances, rendering transoceanic drift the most plausible LDD mechanism.

Although a single event of a sufficiently prolonged seed transport by ocean currents over such large geographical scales is extremely rare, producing numerous propagules makes transoceanic LDD feasible (Nathan, 2006; Nathan *et al.*, 2008; Smith *et al.*, 2018). Our simulations indicated that, with transoceanic drift for up to six months and a sufficient number of propagules, connectivity among established populations of buoyant *Cycas* (>72% stranded seeds) is feasible with the assistance of stepping stones in the Pacific and Indian Ocean (Fig. 5a). Based on the ocean modelling results, westward dispersal of buoyant *Cycas* is driven primarily by the Equatorial Current (Hu *et al.*, 2015). This current facilitates the spread into the Indian and Pacific regions, transporting propagules from Java and Sumatra to Sri Lanka, Seychelles and East Africa (Fig. 5b), and from Guam, Fiji and New Caledonia to the west (Fig. 5d). Potential evidence for the effectiveness of dispersal in buoyant *Cycas* across the oceans is provided by the successful recolonization of Krakatau Island (Indonesia) by *Cycas rumphii* in 1918, only 25 yr after a volcanic eruption sterilized the island (Whittaker *et al.*, 1989).

Remarkably, clades at the edge of the Sunda Shelf (Flores and Palau islands) display relatively early divergence from their sister lineages. Neither phylogenetic data nor seed dispersal trajectories suggest connectivity between Pacific and Indian Ocean taxa, possibly resulting from the Sunda Shelf barrier (Briggs, 1974; Gaither *et al.*, 2010). Furthermore, the Sunda lineage has an earlier divergence than the core Pacific and Indian clades (Fig. 2), suggesting that the ancestors of buoyant *Cycas* arose in the Sunda region and that the two independent taxa at the eastern and western edges of the Sunda Shelf may originate from early dispersal and subsequently gave rise to other taxa in the Pacific and Indian clades. Simulations revealed frequent trajectories from the Sunda region (Sulawesi and Halmahera) that can reach the Palau and Flores islands (Fig. 5c) or adjacent regions (Fig. S7c). Additionally, plastomic alignments showed that the 317-bp (118k position) indel, which was absent in the Indian clade, was preserved in the *C. sundaica* plastome from Flores as well as all Sunda clade lineages (Fig. S3). Analogously, a shared loss of 17 bp (119k position) in the Palau lineage and the Sunda clade was present in the Pacific clade (not shown in the figure). These significant characteristics shared between Sunda lineages and the two taxa in the Pacific and Indian suggest the retention of ancestral plastomic sequences, making Palau and Flores the likely 'stepping stones' in *Cycas* dispersal to the Pacific and the Indian Oceans.

Dispersal and other movements of humans also may play a role in dispersing buoyant *Cycas*, given the significance of cycads to certain ethnic groups as important cultural and food plants (Bonta *et al.*, 2019). For example, *C. seemannii* has been of great cultural importance and could have been an important food source for the first colonizers throughout the Pacific islands (Keppel, 2009). Considering the active trade that existed between the various Pacific islands in pre-European times (Kirch & Hunt, 1988; Cann & Lunn, 1996), the exchange of culturally important plant materials may have occurred. However, our phylogenetic reconstruction does not support recent transregional dispersal events by humans, as the buoyant *Cycas* lineages among the three regions all diverged in the Pleistocene (Figs 1,2).

Although ocean currents can disperse *Cycas* over long distances, their buoyant seeds must tolerate the hazards of remaining long periods at sea, commonly through a dormant phase (Gillespie *et al.*, 2012). Many cycad species have developed seeds that need a pre-germination latency after detaching from the megasporophylls, which consists of the growth and elongation of the embryo into the megagametophyte tissue (Norstog & Nicholls, 1997). This process varies in different cycad taxa (Calonje *et al.*, 2011) but is particularly slow in most species of the *Rumphiae* group (10 to 18 months). In addition, the sarcotesta of *Cycas* remains intact for weeks, repelling the water with its thick, waxy, resinous epidermis, and provides sufficient time for the embryo to reach maturity while crossing the ocean (Dehgan & Yuen, 1983). Within buoyant *Cycas*, the species with the longest pre-germination latencies are *C. thouarsii*, *C. micronesica* and *C. seemannii* from the more remote Indian and Pacific regions, whereas *C. edentata* and *C. rumphii* from the ancestral Sunda region have shorter latency phases (A. J. Lindstrom, unpublished data). The observed variation in pre-germination latency potentially reinforces reproductive isolation among species based on different dispersal distances (see Fig. S1).

Phylogeography of buoyant Cycas and the association with ocean currents

Spatiotemporal patterns of water dispersal in plants are likely to be nonrandom in the context of both recent invasions (Horvitz *et al.*, 2017) and long-term plant evolution and phylogeny (Kudoh *et al.*, 2006). Our study provides clear evidence for the important role of ocean gyres in shaping the divergence and the phylogeographic pattern of buoyant *Cycas* (Fig. 1), as the passive drifts of propagules from the three clades remained largely confined to their respective regions (Fig. 5b,c,d). This confinement driven by ocean currents also is in line with the significant correlation between genetic distance and geographical distance result, which implies an isolation-by-distance (IBD) pattern in buoyant *Cycas* (Fig. 4). Therefore, the isolation and genetic discontinuity of buoyant *Cycas* may be maintained by directional ocean currents that impede mixing, thereby hampering frequent gene flow between Pacific, Sunda and Indian regions via LDD.

The genetic discontinuity of buoyant *Cycas* maintained by ocean circulation in Indian and Sunda clades also conforms to the realm-scale boundary between the Western and Central Indo-Pacific, which is associated with discontinuous habitat types, geomorphological and oceanographic features of coastal and shelf areas in the two realms (Spalding *et al.*, 2007; Crandall *et al.*, 2019). The genetic discontinuity of Pacific and Indian provinces has always been explained by the existence of substantial archipelagos in the Sunda Shelf, a long-recognized barrier between the Pacific and the Indian Ocean in separating faunal distributions (Briggs, 1974). Historical and contemporary dispersal barriers between the Pacific and Indian oceans are indicated

by the confinement of many marine species primarily to one ocean or the other (Briggs, 1999; Gaither *et al.*, 2010). Our results also indicate the overall strength of the Sunda Shelf barrier in shaping regional species distribution and the genetic pattern of buoyant *Cycas*. This is supported by three lines of evidence: (1) the absence of shared species in the Pacific and Indian subclades; (2) the relative distant relationship between the Pacific and Indian subclades; and (3) the significant distance effects across regions inferred by IBD. It is widely accepted that glacial cycles throughout the Pleistocene were accompanied by lower sea levels (as low as 120 m below present), making the Sunda Shelf landmass a nearly complete barrier between the two oceans (Voris, 2000). Although our study indicates that the Pliocene divergence of buoyant *Cycas* from Indian and Pacific clades (4.81 Myr ago (Ma), 95%HPD (highest posterior density) 7.94–2.00 Ma) was not associated with that epoch, the emergence of this barrier could have enhanced the genetic discontinuity of buoyant *Cycas* in the Indian and Pacific Oceans.

At a regional scale, the classic biogeographical barriers within the Sunda Shelf (i.e. Wallace's Line and Huxley's Line) did not affect the observed genetic patterns of buoyant *Cycas*, suggesting that these barriers are porous for some plants and supporting the previous finding for *Begonia* (Thomas *et al.*, 2012) but not for *Phalaenopsis* (Tsai *et al.*, 2015). This is not surprising, considering the dispersal potential of buoyant *Cycas* seeds. Instead, Weber's and Lydekker's lines appear to be more relevant in impacting the genetic pattern, as they separate most Pacific and Sunda lineages but not the heterogeneous lineages from Halmahera (Fig. 1).

Genetic signature shaped by LDD and past climate oscillations

Past climate oscillations are expected to leave signatures in plant genomes. Plastomes, which evolve in their entirety and interact with nuclear genomes, leading to intertwined coevolution (Rousseau-Gueutin *et al.*, 2018), will also mutate to respond to the pressure from shifted niches and constantly changing environments. Selection pressure shows no significant differences between buoyant and nonfloating *Cycas*, but significant differences among different biogeographical regions (Fig. 3b,c). This indicates that *Cycas* lineages from the Indian and the Pacific Ocean are undergoing faster and greater negative (purifying) selection than those in the Sunda region, suggesting that selection is purging changes that cause deleterious impacts (Wagner, 2002) on the fitness of *Cycas* in the Indian and Pacific Oceans. These changes, together with more recent colonization events from Sunda to Pacific and Indian regions, imply the ongoing innovation of buoyant *Cycas* plastomes to adapt to new niches after colonizing the two oceanic coasts.

Climatic oscillations in the Sunda Shelf since the Miocene could shape the genetic diversity pattern. Sea level has markedly changed in the Sunda Shelf over the Pleistocene, resulting in frequent emergence and submergence cycles of part of the Sunda landmass (Voris, 2000; Woodruff, 2010; Hanebuth *et al.*, 2011). These cycles of exposure and inundation may have facilitated the second contact (i.e. two allopatrically distributed populations to be geographically reunited), and tend to homogenize the distribution of genetic variation on continental shelves (Benzie, 1999; Crandall *et al.*, 2019). Our results reflected this homogenization as indicated by the low average nucleotide diversity of buoyant *Cycas* from the Sunda region (Tables 1, S4; Fig. 1). Evidence of the second contact also could be found in the five unsampled *Cycas* species with transitional characters of limited spongy layers, all from the Philippines (Table S2), as they could be results of present or past hybridization with the subsection *Rumphiae* (Lindstrom *et al.*, 2009). However, future studies with more extensive sampling are needed to confirm these suggestions.

Implications of ocean drifting modelling

This study illustrates the importance of ocean drifting modelling in simulating connectivity establishment and predicting species distribution, especially for underexplored or undiscovered areas. The Lagrangian method, used widely in marine organisms (Brischoux *et al.*, 2016; Gaspar & Lalire, 2017), seems applicable to estimate the distribution range of littoral plants that produce propagules with transoceanic dispersal potential, such as buoyant *Cycas*. Buoyant *Cycas* species are absent from some regions that should be receiving propagules based on our simulations, such as the Maldives archipelagos and the continental margins of Somalia and East Australia. This mismatch, may be the result of (1) lack of successful post-colonization adaptations or local extinction as a consequence of geological and climate change (Urban, 2015); (2) deviated simulations compared to the true LDD process because the modelling is sensitive to the uncertainty of the ocean velocity (Fig. S8); or (3) undiscovered populations of buoyant *Cycas* being present in the regions. Additionally, the better correspondence of July-started (vs January-started) simulations (Fig. 5) with the current distribution of *Cycas* (Fig. S1) highlights the importance of the actual timing of seed release and fits growing evidence that the temporal dynamics of seed production and release affect LDD patterns (Nathan & Katul, 2005; Wright *et al.*, 2008). Therefore, future studies should take the actual seed release period into account when using ocean drift modelling for species range prediction.

Conclusion

In order to test the hypothesis that phylogeographic patterns of littoral plants with buoyant propagules are related to ocean currents, we used plastomic data to help reveal the genetic structure of buoyant *Cycas* and simulated oceanographic drift to estimate seed dispersal patterns. Our results highlight the importance of combining these two approaches to explain the phylogeography of ocean-dispersed littoral plants. Furthermore, the utility of the plastid genome to track seed dispersal and assess genomic footprints under climate oscillations in recently diverged plant lineages is demonstrated by (1) the conformation of three major taxa of buoyant *Cycas* to prevalent ocean currents; (2) the identification of two critical stepping stones (Flores and Palau) for dispersal; and (3) the detection of comparative levels of negative

selection in different phylogeographic regions. The integration of ocean current modelling and phylogeography therefore potentially has broad applications to other littoral taxa with buoyant seeds (e.g. *Cocos nucifera* and *Barringtonia asiatica*), providing an opportunity for novel insights into the evolution of terrestrial coastal ecosystems globally.

Acknowledgements

We would like to thank three anonymous reviewers for their thoughtful comments and constructive suggestions towards improving our manuscript. This work is supported by the Biodiversity Survey, Observation and Assessment of the Ministry of Ecology and Environment, China (2019HJ2096001006), the National Natural Science Foundation of China (31900184, 31800191 and 32000169), the West Light Foundation of the Chinese Academy of Sciences (Y8246811W1), and Natural Science Foundation of Yunnan Province (202001AT070072). We thank James Shevock and Maia M. Jones for their linguistic assistance during the preparation and revision of this manuscript.

Author contributions

JL and XG conceived the study; JL and AJL did fieldwork; JL conducted laboratory work; JL, RN and YSC analyzed data; JL wrote the first draft and all authors contributed substantially to revisions.

ORCID

Yong-Sheng Chen https://orcid.org/0000-0001-5835-8249
Xun Gong https://orcid.org/0000-0003-1705-6935
Anders J. Lindstrom https://orcid.org/0000-0003-0419-9136
Jian Liu https://orcid.org/0000-0003-1828-4024
Ran Nathan https://orcid.org/0000-0002-5733-6715

Data availability

The data that supports the findings of this study are available in the supplementary material of this article.

References

Ali JR, Huber M. 2010. Mammalian biodiversity on Madagascar controlled by ocean currents. *Nature* **463**: 653–656.

Amiryousefi A, Hyvönen J, Poczai P. 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **34**: 3030–3031.

Aphalo PJ. 2020. *ggpmisc: miscellaneous extensions to 'ggplot2'*. [WWW document] URL: https://cran.rstudio.com/web/packages/ggpmisc/

Banerjee AK, Guo W, Qiao S, Li W, Xing F, Lin Y, Hou Z, Li S, Liu Y, Huang Y. 2020. Land masses and oceanic currents drive population structure of *Heritiera littoralis*, a widespread mangrove in the Indo-West Pacific. *Ecology and Evolution* **10**: 7349–7363.

Bennett AF. 2006. *Lagrangian fluid dynamics*. Cambridge, UK: Cambridge University Press.

Benzie JAH. 1999. Major genetic differences between Crown-of-Thorns starfish (*Acanthaster planci*) populations in the Indian and Pacific Oceans. *Evolution* **53**: 1782.

Bertola LD, Boehm JT, Putman NF, Xue AT, Robinson JD, Harris S, Baldwin CC, Overcast I, Hickerson MJ. 2020. Asymmetrical gene flow in five co-distributed syngnathids explained by ocean currents and rafting propensity. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **287**: 20200657.

Bonta M, Pulido-Silva MT, Diego-Vargas T, Vite-Reyes A, Vovides AP, Cibrián-Jaramillo A. 2019. Ethnobotany of Mexican and northern Central American cycads (Zamiaceae). *Journal of Ethnobiology and Ethnomedicine* **15**: 1–34.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N *et al.* 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **15**: 1–28.

Bradley WG, Mash DC. 2009. Beyond Guam: the cyanobacteria/BMAA hypothesis of the cause of ALS and other neurodegenerative diseases. *Amyotrophic Lateral Sclerosis* **10**: 7–20.

Brenner ED, Stevenson DW, Twigg RW. 2003. Cycads: evolutionary innovations and the role of plant-derived neurotoxins. *Trends in Plant Science* **8**: 446–452.

Briggs JC. 1999. Extinction and replacement in the Indo-West Pacific Ocean. *Journal of Biogeography* **26**: 777–783.

Brischoux F, Cotté C, Lillywhite HB, Bailleul F, Lalire M, Gaspar P. 2016. Oceanic circulation models help to predict global biogeography of pelagic yellow-bellied sea snake. *Biology Letters* **12**: 6–9.

Calonje M, Kay J, Griffith MP. 2011. Propagation of cycad collections from seed. *Sibbaldia: The International Journal of Botanic Garden Horticulture* **9**: 79–96.

Calonje M, Stevenson DW, Osborne R. 2021. *The world list of cycads*. [WWW document] URL http://www.cycadlist.org [accessed 16 January 2021].

Cann RL, Lunn JK. 1996. Mankind and Pacific biogeography. In: Keast A, Miller SE, eds. *The origin and evolution of Pacific island biotas, New Guinea to eastern Polynesia*. Amsterdam, the Netherlands: SPB Academic, 437–443.

Condamine FL, Nagalingum NS, Marshall CR, Morlon H. 2015. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evolutionary Biology* **15**: 1–18.

Crandall ED, Riginos C, Bird CE, Liggins L, Treml E, Beger M, Barber PH, Connolly SR, Cowman PF, DiBattista JD *et al.* 2019. The molecular biogeography of the Indo-Pacific: testing hypotheses with multispecies genetic patterns. *Global Ecology and Biogeography* **28**: 943–960.

Dagestad KF, Röhrs J, Breivik O, Ådlandsvik B. 2018. OpenDrift v1.0: a generic framework for trajectory modelling. *Geoscientific Model Development* **11**: 1405–1420.

Dehgan B, Yuen CKKH. 1983. Seed morphology in relation to dispersal, evolution, and propagation of Cycas L. *Botanical Gazette* **144**: 412–418.

Doyle J. 1991. *DNA protocols for plants-CTAB total DNA isolation*. Berlin, Germany: Springer.

Dray S, Dufour AB. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**: 1–20.

Excoffier L, Foll M, Petit RJ. 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* **40**: 481–501.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Research* **32**: 273–279.

Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW. 2010. Genetic evaluation of marine biogeographical barriers: Perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). *Journal of Biogeography* **37**: 133–147.

Gallaher T, Callmander MW, Buerki S, Setsuko S, Keeley SC. 2017. Navigating the 'broad freeway': ocean currents and inland isolation drive diversification in the *Pandanus tectorius* complex (Pandanaceae). *Journal of Biogeography* **44**: 1598–1611.

Gao Z, Thomas BA. 1989. A review of fossil cycad megasporophylls, with new evidence of *Crossozamia* Pomel and its associated leaves from the lower permian of Taiyuan, China. *Review of Palaeobotany and Palynology* **60**: 205–223.

Gaspar P, Lalire M. 2017. A model for simulating the active dispersal of juvenile sea turtles with a case study on western Pacific leatherback turtles. *PLoS ONE* **12**: 1–30.

Gillespie RG, Baldwin BG, Waters JM, Fraser CI, Nikula R, Roderick GK. 2012. Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology and Evolution* **27**: 47–56.

Gordon AL, Cenedese C 2018. *Ocean current. Encyclopedia britannica.* [WWW document] URL https://www.britannica.com/science/ocean-current [accessed 31 January 2021].

Gradstein F, Ogg J, Schmitz M, Ogg G. 2012. *The geologic time scale*, Volume Set. Boston, MA, USA: Elsevier.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**: W59–W64.

Griffith MP, Calonje MA, Stevenson DW, Husby CE, Little DP. 2012. Time, place, and relationships: cycad phenology in a phylogenetic and biogeographic context. *Memoirs of the New York Botanical Garden* **106**: 59–81.

Guo W, Ng WL, Wu H, Li W, Zhang L, Qiao S, Yang X, Shi X, Huang Y. 2018. Chloroplast phylogeography of a widely distributed mangrove species, *Excoecaria agallocha*, in the Indo-West Pacific region. *Hydrobiologia* **807**: 333–347.

Guppy HB. 1906. *Observations of a naturalist in the Pacific between 1896 and 1899. Volume II: plant dispersal.* London, UK: Macmillan and Co.

Hall R. 2009. Southeast Asia's changing palaeogeography. *Blumea* **54**: 148–161.

Hanebuth TJJ, Voris HK, Yokoyama Y, Saito Y, Okuno J. 2011. Formation and fate of sedimentary depocentres on Southeast Asia's Sunda Shelf over the past sea-level cycle and biogeographic implications. *Earth-Science Reviews* **104**: 92–110.

Haywood AM, Valdes PJ. 2004. Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere. *Earth and Planetary Science Letters* **218**: 363–377.

Hermsen EJ, Taylor TN, Taylor EL, Stevenson DW. 2006. Cataphylls of the Middle Triassic cycad *Antarcticycas schopfii* and new insights into cycad evolution. *American Journal of Botany* **93**: 724–738.

Hill KD. 1994. The *Cycas rumphii* complex (Cycadaceae) in New Guinea and the western Pacific. *Australian Systematic Botany* **7**: 543–567.

Hill KD. 1996. Cycads in the pacific. In: Miller AK, Miller SE, eds. *The origin and evolution of Pacific island biotas, New Guinea to Eastern Polynesia: patterns and processes.* Amsterdam, The Netherlands: SPB Academic Publishing, 267–274.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**: 518–522.

Horvitz N, Wang R, Wan FH, Nathan R. 2017. Pervasive human-mediated large-scale invasion: analysis of spread patterns and their underlying mechanisms in 17 of China's worst invasive plants. *Journal of Ecology* **105**: 85–94.

Hu D, Wu L, Cai W, Gupta AS, Ganachaud A, Qiu Bo, Gordon AL, Lin X, Chen Z, Hu S *et al.* 2015. Pacific western boundary currents and their roles in climate. *Nature* **522**: 299–308.

Hubbs CL, Briggs JC. 1974. *Marine zoogeography.* New York, NY, USA: McGraw-Hill.

Jensen MP, Dalleau M, Gaspar P, Lalire M, Jean C, Ciccione S, Mortimer JA, Quillard M, Taquet C, Wamukota A *et al.* 2020. Seascape genetics and the spatial ecology of juvenile green turtles. *Genes* **11**: 278.

Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**: 241.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al.* 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Keppel G. 2009. Morphological variation, an expanded description and ethnobotanical evaluation of *Cycas seemannii* A. Braun (Cycadaceae). *South Pacific Journal of Natural and Applied Sciences* **27**: 20.

Keppel G, Hodgskiss PD, Plunkett GM. 2008. Cycads in the insular South-west Pacific: dispersal or vicariance? *Journal of Biogeography* **35**: 1004–1015.

Kirch PV, Hunt TL. 1988. In: Kirch PV, Hunt TL, eds. *The spatial and temporal boundaries of Lapita.* Seattle, WA, USA: Burke Museum.

Kistler L, Montenegro Á, Smith BD, Gifford JA, Green RE, Newsom LA, Shapiro B. 2014. Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proceedings of the National Academy of Sciences, USA* **111**: 2937–2941.

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K *et al.* 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* **15**: 378–392.

Kudoh H, Shimamura R, Takayama K, Whigham DF. 2006. Consequences of hydrochory in *Hibiscus. Plant Species Biology* **21**: 127–133.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**: 772–773.

Levin SA, Muller-Landau HC, Nathan R, Chave J. 2003. The ecology and evolution of seed dispersal: a theoretical perspective. *Annual Review of Ecology, Evolution, and Systematics* **34**: 575–604.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Lindstrom AJ, Hill KD, Stanberg LC. 2009. The genus *Cycas* (Cycadaceae) in The Philippines. *Telopea* **12**: 119–145.

Liu J, Zhang S, Nagalingum NS, Chiang YC, Lindstrom AJ, Gong X. 2018. Phylogeny of the gymnosperm genus *Cycas* L. (Cycadaceae) as inferred from plastid and nuclear loci based on a large-scale sampling: evolutionary relationships and taxonomical implications. *Molecular Phylogenetics and Evolution* **127**: 87–97.

Mankga LT, Yessoufou K, Mugwena T, Chitakira M. 2020. The cycad genus *Cycas* may have diversified from Indochina and occupied its current ranges through vicariance and dispersal events. *Frontiers in Ecology and Evolution* **8**: 1–13.

Marler TE, Snyder LR, Shaw CA. 2010. *Cycas micronesica* (Cycadales) plants devoid of endophytic cyanobacteria increase in β-methylamino-l-alanine. *Toxicon* **56**: 563–568.

Martínez LCA, Artabe AEE, Bodnar J. 2012. A new cycad stem from the Cretaceous in Argentina and its phylogenetic relationships with other Cycadales. *Botanical Journal of the Linnean Society* **170**: 436–458.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R, Teeling E. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

Miryeganeh M, Takayama K, Tateishi Y, Kajita T. 2014. Long-distance dispersal by sea-drifted seeds has maintained the global distribution of *Ipomoea pes-caprae* subsp. *brasiliensis* (convolvulaceae). *PLoS ONE* **9**: e91836.

Nagalingum NS, Marshall CR, Quental TB, Rai HS, Little DP, Mathews S. 2011. Recent synchronous radiation of a living fossil. *Science* **334**: 796–799.

Nathan R. 2006. Long-distance dispersal of plants. *Science* **25**: 1–4.

Nathan R. 2013. Dispersal biogeography. In: Levin SA, ed. *Encyclopedia of biodiversity, second edition.* New York, NY, USA: Academic Press, 539–561.

Nathan R, Katul GG. 2005. Foliage shedding in deciduous forests lifts up long-distance seed dispersal by wind. *Proceedings of the National Academy of Sciences, USA* **102**: 8251–8256.

Nathan R, Schurr FM, Spiegel O, Steinitz O, Trakhtenbrot A, Tsoar A. 2008. Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution* **23**: 638–647.

Nikolic N, Montes I, Lalire M, Puech A, Bodin N, Arnaud-Haond S, Kerwath S, Corse E, Gaspar P, Hollanda S *et al.* 2020. Connectivity and population structure of albacore tuna across southeast Atlantic and southwest Indian Oceans inferred from multidisciplinary methodology. *Scientific Reports* **10**: 1–17.

Norstog KJ, Nicholls TJ. 1997. *The biology of the cycads.* Ithaca, NY, USA: Cornell University Press.

Paradis E, Schliep K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **15**: 1–12.

R Core Team. 2020. *R: a language and environment for statistical computing.* [WWW document] URL https://www.r-project.org/ [accessed 23 October 2020].

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**: 901–904.

Renner S. 2004. Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences* **165**: S23–S33.

Rothwell GW, Scheckler SE, Gillespie WH. 1989. *Elkinsia* gen. nov., a late devonian gymnosperm with cupulate ovules. *Botanical Gazette* **150**: 170–189.

Rousseau-Gueutin M, Keller J, Ferreira de Carvalho J, Aïnouche A, Martin G. 2018. The intertwined chloroplast and nuclear genome coevolution in plants. In: Ratnadewi D, ed. *Plant growth and regulation – alterations to sustain unfavorable conditions*. London, UK: InTechOpen, 61–84.

van Sebille E, Griffies SM, Abernathey R, Adams TP, Berloff P, Biastoch A, Blanke B, Chassignet EP, Cheng Yu, Cotter CJ *et al.* 2018. Lagrangian ocean analysis: fundamentals and practices. *Ocean Modelling* **121**: 49–75.

Sefc KM, Wagner M, Zangl L, Weiß S, Steinwender B, Arminger P, Weinmaier T, Balkic N, Kohler T, Inthal S *et al.* 2020. Phylogeographic structure and population connectivity of a small benthic fish (*Tripterygion tripteronotum*) in the Adriatic Sea. *Journal of Biogeography* **47**: 2502–2517.

Smith TM, York PH, Broitman BR, Thiel M, Hays GC, van Sebille E, Putman NF, Macreadie PI, Sherman CDH. 2018. Rare long-distance dispersal of a marine angiosperm across the Pacific Ocean. *Global Ecology and Biogeography* **27**: 487–496.

Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson M, Halpern BS, Jorge MA, Lombana Al, Lourie SA *et al.* 2007. Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *BioScience* **57**: 573–583.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Thomas DC, Hughes M, Phutthai T, Ardi WH, Rajbhandary S, Rubite R, Twyford AD, Richardson JE. 2012. West to east dispersal and subsequent rapid diversification of the mega-diverse genus *Begonia* (Begoniaceae) in the Malesian archipelago. *Journal of Biogeography* **39**: 98–113.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**: W6–W11.

Tsai CC, Chou CH, Wang HV, Ko YZ, Chiang TY, Chiang YC. 2015. Biogeography of the *Phalaenopsis amabilis* species complex inferred from nuclear and plastid DNAs. *BMC Plant Biology* **15**: 202.

Urban MC. 2015. Accelerating extinction risk from climate change. *Science* **348**: 571–573.

Van der Stocken T, Carroll D, Menemenlis D, Simard M, Koedam N. 2019. Global-scale dispersal and connectivity in mangroves. *Proceedings of the National Academy of Sciences, USA* **116**: 915–922.

Voris HK. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* **27**: 1153–1167.

Wagner A. 2002. Selection and gene duplication: a view from the genome. *Genome Biology* **3**: 3–5.

Wee AKS, Noreen AME, Ono J, Takayama K, Kumar PP, Tan HTW, Saleh MN, Kajita T, Webb EL. 2020. Genetic structures across a biogeographical barrier reflect dispersal potential of four Southeast Asian mangrove plant species. *Journal of Biogeography* **47**: 1258–1271.

Wee AKS, Takayama K, Asakawa T, Thompson B, Onrizal SS, Tung NX, Nazre M, Soe KK, Tan HTW *et al.* 2014. Oceanic currents, not land masses, maintain the genetic structure of the mangrove *Rhizophora mucronata* Lam. (Rhizophoraceae) in Southeast Asia. *Journal of Biogeography* **41**: 954–964.

Whittaker RJ, Bush MB, Richards K. 1989. Plant recolonization and vegetation succession on the Krakatau Islands, Indonesia. *Ecological Monographs* **59**: 59–123.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**: 3350–3352.

Wickham H. 2016. *ggplot2: Elegant graphics for data analysis*. New York, NY, USA: Springer.

Woodruff DS. 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodiversity and Conservation* **19**: 919–941.

Wright S. 1969. The theory of gene frequencies. *Evolution and the genetics of populations*. Chicago, IL, USA: University of Chicago Press.

Wright SJ, Trakhtenbrot A, Bohrer G, Detto M, Katul GG, Horvitz N, Muller-Landau HC, Jones FA, Nathan R. 2008. Understanding strategies for seed dispersal by wind under contrasting atmospheric conditions. *Proceedings of the National Academy of Sciences, USA* **105**: 19084–19089.

Wu ZY, Liu J, Provan J, Wang H, Chen CJ, Cadotte MW, Luo YH, Amorim BS, Li DZ, Milne RI. 2018. Testing Darwin's transoceanic dispersal hypothesis for the inland nettle family (Urticaceae). *Ecology Letters* **21**: 1515–1529.

Xiao LQ, Möller M. 2015. Nuclear ribosomal ITS functional paralogs resolve the phylogenetic relationships of a late-Miocene radiation cycad *Cycas* (Cycadaceae). *PLoS ONE* **10**: 1–14.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Zhong ZR, Li N, Qian D, Jin JH, Chen T. 2011. Maternal inheritance of plastids and mitochondria in *Cycas* L. (Cycadaceae). *Molecular Genetics and Genomics* **286**: 411–416.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 The global distribution of the buoyant *Cycas*.

Fig. S2 Chloroplast genome graph of *Cycas edentata*.

Fig. S3 Global alignment of 10 buoyant *Cycas* genomes based on mVISTA.

Fig. S4 Comparison IR and SC borders among 10 buoyant *Cycas* chloroplast genomes.

Fig. S5 Cladograms of ML trees of *Cycas*.

Fig. S6 Tanglegram of ML tree and the MCC tree of *Cycas*.

Fig. S7 January-initiated ocean drift simulation of buoyant *Cycas* propagules across the Indo-Pacific.

Fig. S8 Simulated trajectories across the Indo-Pacific by the constant current model in OpenDrift.

Table S1 Collection information, vouchers and plastome NCBI accessions of the buoyant *Cycas* samples.

Table S2 Information of *Cycas* species that occupy spongy tissues but are unable to float on water.

Table S3 GIS data of seed releasing sites in ocean drifting modelling by OpenDrift.

Table S4 Comparison of nucleotide diversity for the Sunda Shelf, the core Pacific Ocean and the core Indian Ocean subclades.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.