Gene 800 (2021) 145845

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Genome skims analysis of betel palms (*Areca* spp., Arecaceae) and development of a profiling method to assess their plastome diversity

Pauline Raimondeau [a], Sophie Manzi [a], Nicolas Brucato [a], Christopher Kinipi [b], Matthew Leavesley [b,c], François-Xavier Ricaut [a,1], Guillaume Besnard [a,1,*]

[a] Laboratoire Évolution & Diversité Biologique (EDB, UMR 5174), Université de Toulouse, CNRS-IRD-UPS, 118 route de Narbonne, Bât. 4R1, 31062 Toulouse, France
[b] Strand of Anthropology, Sociology and Archaeology, School of Humanities and Social Sciences, University of Papua New Guinea, PO Box 320, University 134, National Capital District, Papua New Guinea
[c] CABAH & College of Arts, Society and Education, James Cook University, PO Box 6811, Cairns, QLD 4870, Australia

ARTICLE INFO

Keywords:
Betel nut domestication
Herbarium
Indo-Pacific
Mitogenome
Phylogeography
Plastome

ABSTRACT

The betel nut (*Areca catechu* L., Arecaceae) is a monoecious cultivated palm tree that is widespread in tropical regions. It is mainly cultivated for producing areca nuts, from which seeds are extracted and chewed by local populations principally in the Indo-Pacific region. Seeds contain alkaloids which are central nervous system stimulants and are highly addictive. Wild relatives of the betel nut are distributed in South Asia and Australasia, with ca. 40–50 *Areca* species currently recognized. The geographic origin(s) of the betel nut and its subsequent diffusion and diversification remains poorly documented. Here, a genome skimming approach was applied to screen nucleotidic variation in the most abundant genomic regions. Low coverage sequencing data allowed us to assemble full plastomes, mitochondrial regions (either full mitogenomes or the full set of mitochondrial genes) and the nuclear ribosomal DNA cluster for nine representatives of the *Areca* genus collected in the field and herbarium collections (including a 182-years old specimen collected during the Dumont d'Urville's expedition). These three genomic compartments provided similar phylogenetic signals, and revealed very low genomic diversity in our sample of cultivated betel nut. We finally developed a genotyping method targeting 34 plastid DNA microsatellites. This plastome profiling approach is useful for tracing the spread of matrilineages, and in combination with nuclear genomic data, can resolve the history of the betel nut. Our method also proves to be efficient for analyzing herbarium specimens, even those collected >100 years ago.

1. Introduction

The betel nut (also called arecanut; *Areca catechu* L., Arecaceae) is a monoecious cultivated palm tree that is widespread in tropical regions (Staples and Bevacqua, 2006). Archeological and ethnobotanical evidence suggests that this crop originated during the Neolithic either from Mainland Asia (Oxenham et al., 2002; Zumbroich, 2008) or the Malesian region (Whitmore, 1977; Valdes, 2014). *Areca catechu* is mainly cultivated for producing areca nuts, from which seeds are extracted and chewed by local populations, as well as for religious rituals (Staples and Bevacqua, 2006; Valdes, 2014). Nowadays, as many as 600 million people worldwide, concentrated in the Indo-Pacific region, regularly chew betel nut. Seeds contain four major alkaloids (Gheddar et al., 2020) which are central nervous system stimulants and are highly addictive (acting on the same receptor proteins in the brain as nicotine). They have psychoactive effects and are appreciated for their appetite suppressant and antimycobacterial properties (Gautam et al., 2007). However, they are carcinogenic and intoxicating, and their consumption can spread major diseases, because users spit pathogenic saliva which increases the spread of tuberculosis (e.g., Thomas and MacLennan, 1992; Norton, 1998; Amarasinghe et al., 2010; Sharan et al., 2012; Chen et al., 2017). The tree is also frequently exploited for floor material, natural fibers, and fishing stakes, as well as ornamentals (Koonlin, 1979; Heatubun et al., 2012; Sunny and Rajan, 2021). While the origins of the

Abbreviations: CDS, coding sequence; cpDNA, Plastid DNA; cpSSR, Plastid Simple Sequence Repeats; ETS, External Transcribed Sequence; ITS, Internal Transcribed Spacers; mtDNA, Mitochondrial DNA; nrDNA, Nuclear ribosomal DNA; PNG, Papua New Guinea; SNP, Single Nucleotide Polymorphism.

* Corresponding author.
E-mail address: guillaume.besnard@univ-tlse3.fr (G. Besnard).
[1] Co-senior authors

https://doi.org/10.1016/j.gene.2021.145845
Received 3 April 2021; Accepted 13 July 2021
Available online 15 July 2021

betel nut cultivation in Southeast Asia and the Malesian region remains unclear (Oxenham et al., 2002; Zumbroich, 2008), there is archaeological evidence that the crop has been spread in southeasternmost regions [e.g., Papua New Guinea (PNG) and surrounding archipelagos and Micronesia] over the last three millennia (Yen, 1993; Bourke, 2009; Denham, 2010; Fitzpatrick et al., 2013), before it was introduced on different continents during recent historical times (e.g., Tropical Africa, Madagascar, Caribbean Islands; Oreca et al., 2009). All wild relatives of the betel nut are distributed in South Asia and Australasia, with ca. 40–50 *Areca* species currently recognized, but no comprehensive phylogenetic work has been done on this group until now (Loo et al., 2006; Dransfield et al., 2008; Henderson, 2009; Heatubun, 2011). Several of these species are frequently used as inferior substitutes for betel chewing, such as *Areca macrocalyx* Zipp in PNG and *Areca vestiaria* Giseke in North Sulawesi and North Maluku (Murthy and Pillai, 1982; Heatubun et al., 2012; FXR and ML, pers. observ.). Both wild and cultivated forms may locally coexist in the native distribution area of the *Areca* genus, and spontaneous populations are also exploited (Whitmore, 1977). Because cultivated trees are exclusively propagated via seedlings (Staples and Bevacqua, 2006), such a mix of trees with putatively different origins may have led to frequent admixture during the crop diffusion. This could have promoted diversification of the cultivated betel nut, and at the same time blurred phylogeographic patterns, as has already been shown in several perennial crops (e.g., Myles et al., 2011; Cornille et al., 2012; Besnard et al., 2013; Roullier et al., 2013; Flowers et al., 2019). Therefore, documenting the geographic origin(s) of the betel nut as well as its subsequent diffusion and diversification in southern Asia and Australasia is still necessary, and a phylogeographic approach may bring new insights on this topic.

Polymorphisms of the organellar genomes (i.e., plastome and mitogenome) have been often screened for inferring the diffusion of plants, and particularly of crops (e.g., Carreel et al., 2002; Petit et al., 2002; Besnard et al., 2013). Plastid DNA has, however, been more frequently used than mitochondrial DNA due to its higher substitution rate (Duminil and Besnard, 2021). The common use of organellar DNA data is due to six main reasons: (*i*) these genomes are usually uniparentally inherited, mostly from the mother, and in this case, gene dispersal by seeds without the recombination effect can be investigated (Schaal and Olsen, 2000); (*ii*) their haploid nature facilitates their sequencing; (*iii*) such genomes are more prone to stochastic events because their effective population size is half that of diploid genomes, allowing a more accurate detection of evolutionary events such as a long persistence of relict populations in refuge zones during last glaciations (Petit et al., 1993). In addition, the dispersion of maternally inherited genomes by seed dissemination occurs at shorter geographic distances than for nuclear genomes. The consequence of a reduced gene dispersal and high genetic drift in organelle genomes is a generally pronounced geographic structure, which facilitates phylogeographic analyses and tracing the origins of crops or invasive populations (Schaal and Olsen, 2000); (*iv*) they exhibit a high number of copies per cell (Bendich, 1987), which may represent a significant advantage for forensic or ancient DNA analyses; (*v*) they are circular and protected by a double-membrane envelope, which makes them resistant to exonucleases and less prone to endonuclease degradation (another advantage for forensics and archaeogenomics); and finally (*vi*) they exhibit a lower mutation rate than nuclear genomes (Wolfe et al., 1987), and such stability is generally required for traceability analyses. Due to this low mutation rate, it may be useful to target the most variable regions of the plastome when studying relatively recent demographic process. For this reason, several studies have targeted plastid microsatellites (or Simple Sequence Repeats; hereafter referred to cpSSRs) in plants for which extremely low plastid mutation rates have been recorded (e.g., Provan et al., 2001; Besnard and Bervillé, 2002; Navascués and Emerson, 2005; Gryta et al., 2017). Such extremely low substitution rates have also been observed in plastomes of palm trees (Wilson et al., 1990; Comer et al., 2015; Barrett et al., 2016) that may call for the use of cpSSRs in population genetic studies. These markers have found applications in various fields, such as phylogeography, taxonomy, conservation and agronomy (e.g., Li et al., 2020; Sharma et al., 2020; Gorrilliot et al., 2021;Monteiro et al., 2021). The combined use of organellar DNA data with nuclear genome information remains, however, necessary to contrast phylogeographic patterns due to seed- and seed + pollen-mediated gene dispersals, and to improve understanding of population demographic events (e.g., Magri et al., 2006; Tollefsrud et al., 2009).

In this study, we aimed to generate new genomic data of the *Areca* genus that will help to reconstruct its phylogeography and document the domestication of the betel nut (*A. catechu*), which will be essential information to manage genetic resources of the crop. A genome skimming approach was applied to screen variation in the most abundant genomic regions. Low coverage sequencing data thus allowed us to assemble full plastomes (cpDNA), mitochondrial DNA regions (mtDNA; including all coding genes) and the nuclear ribosomal DNA cluster (nrDNA) on a few representatives of the *Areca* genus collected in the field (in PNG) and from herbarium collections. Phylogenetic signals from these three genomic compartments were compared and discussed. Finally, to track the diffusion of betel nut, we targeted the most variable regions of the plastid genome (cpSSRs) and developed a set of loci to characterize cpDNA variation from fresh and herbarium samples.

2. Material and methods

2.1. Plant sampling and DNA extraction

Our study was conducted on six *Areca* taxa from Southeast Asia to Australasia (Table 1), including the cultivated betel nut (*A. catechu*) and wild relatives frequently used as inferior betel substitutes. Five fresh samples were collected by ML and FXR during 2019 field seasons in PNG. These samples consisted of three specimens of *A. catechu* and two specimens of *A. macrocalyx* (commonly named 'Kavivi' in PNG). Permission to conduct research was granted by the National Research Institute of PNG (permit 99902292358), with full support from the School of Humanities and Social Sciences, University of PNG. Five additional herbarium specimens (two of them were + 100 years old) were provided by the Museum d'Histoire Naturelle de Paris (MNHNP). In particular, P02146304 (*A. vestiaria*) was collected in North Maluku during the Dumont d'Urville's expedition ("*Voyage de l'Astrolabe et de la Zélée, 1838-1840*"), 182 years ago (stop in Ternate from January 31st to February 1st, 1839; Dumont d'Urville and Jacquinot, 1842–1853).

For all samples, DNA extractions were performed from 20 mg of dried leaves with the BioSprint method (Qiagen Inc.), in a dedicated facility for samples with low DNA quantity (e.g., Staats et al., 2013).

2.2. Sequence assembly of plastome, mitochondrial genes and nuclear ribosomal DNA cluster

We used a genome skimming approach (i.e., Straub et al., 2012; Besnard et al., 2014; Zedane et al., 2016) to attempt the assembly of the plastome (cpDNA), the mitogenome or mitochondrial genes (mtDNA), and a nearly complete nuclear ribosomal DNA unit (nrDNA) on each sample. Fourteen (for 'BORNEO') to 200 ng of double-stranded DNA (according to the Qubit quantification; Table S1) were used for library preparation followed by shotgun sequencing with the Illumina technology (HiSeq 3000, Illumina Inc.) as described in Olofsson et al. (2019). For herbarium samples, DNA was highly fragmented, and a short sonication (20 s instead of 40 s) was carried out before library construction. Plastomes and nrDNA clusters were assembled using the approach described by Olofsson et al. (2019). The nrDNA cluster is composed of three genes (18S, 5.8S, and 26S), internal transcribed spacers (ITS1 and ITS2), and an incomplete intergenic spacer that includes external transcribed sequences (ETS). Two mitogenomes were also assembled on one accession of *A. catechu* ('PNG-1') and *A. macrocalyx* ('PNG-5') following the method described by Van de Paer et al. (2018). Thirty-seven

Table 1
Plant material used in the present study. Ten trees belonging to six *Areca* species were characterized with a genome skimming approach. Two accessions from GenBank (full plastome) were added to our study for comparison. Five samples are from herbarium specimens (vouchered at MNHN Paris) collected between 1839 and 2010.

Species	Sampling locality, country	Yr of collect	Accession / Voucher [CODE]	GPS coordinates
Areca caliso Becc.	Mt Balusan, Philippines[h]	1916	A.D.E. Elmer 14,839 (P00742110) [PHIL]	12°46′07″N, 124°03′18″E
A. catechu L.	Iles Torres, Vanuatu[h]	2007	Y. Pillon 1136 (P01815581) [VANUATU]	13°15′00″S, 166°37′00″E
	Kavieng, New Ireland, PNG	2019	M. Leavesley [PNG-1]	2°34′40″S, 150°48′26″E
	Manim, Western Highland, PNG	2019	M. Leavesley & F.X. Ricaut [PNG-2]	5°54′18″S, 144°18′05″E
	Awim, East Sepik, PNG	2019	M. Leavesley & F.X. Ricaut [PNG-3]	4°45′08″S, 143°34′48″E
	South Kanara, Karnataka, India*	2020	M.K. Rajesh et al. [INDIA]	12°46′20″N, 75°06′58″E
A. macrocalyx Zipp	Kundiman, East Sepik, PNG	2019	M. Leavesley & F.X. Ricaut [PNG-4]	4°36′07″S, 143°30′13″E
	Manim, Western Highland, PNG	2019	M. Leavesley & F.X. Ricaut [PNG-5]	5°54′18″S, 144°18′05″E
A. mogeana Heatubun	E. Kalimantan, Borneo, Indonesia[h]	1992	J. VV 1192B (P01675050) [BORNEO]	1°40′N, 115°00′E
A. aff. *triandra* Roxb. ex Buch.-Ham.	Thong Phaphum, Thailand[h]	2010	M. Jeanson 69 (P00889502) [THAI]	7°51′49″N, 99°05′40″E
A. vestiaria Giseke	Ternate, N. Maluku, Indonesia[h]	1839	M. Le Guillou 1841 (P02146304) [MOL-1]	0°47′45″N, 127°21′41″E
	Sulawesi-Maluku, Indonesia*	2008	W.B. Zomlefer 2310 (FTG, NY) [MOL-2]	na (Fairchild Tropical Botanic Garden, Miami)

[h] Herbarium specimen; *Accessions from GenBank (complete plastome; Comer et al., 2015; Rajesh et al., 2020); na = data not available.

mitochondrial genes were then assembled for other samples using one generated reference ('PNG-1') following the method described in Dupin et al. (2020). Reads were mapped onto the different regions using Bowtie 2 with default parameters (Langmead and Salzberg, 2012) and the average depth of coverage was estimated from the generated bam files using Mosdepth (Pedersen and Quinlan, 2018).

2.3. Development of a plastome profiling approach in arecanuts

Based on complete plastome sequences of *Areca* (including two accessions from GenBank), we searched for length polymorphisms that were due to a variable number of repeats in microsatellite motifs (or cpSSRs; Table S1). Several criteria were used to select cpSSR loci among the eleven available plastomes: *i*) first, primers need to be defined in conserved regions to ensure the transferability of loci among close relatives of *A. catechu*. We targeted short segments (generally inferior to 250 bp, except for locus *Ac*-cp16; Table S1) to facilitate the PCR amplification on degraded DNA; the absence of highly conserved regions flanking microsatellite motif led us to exclude the region. Finally, two non-exclusive criteria were considered: *ii*) the longest microsatellite stretch should contain at least 12 repeats, or *iii*) at least three length variants among *Areca* species, or at least two variants among the four *A. catechu* accessions. With these three criteria, we selected 34 variable microsatellite regions: 24, three, and one loci with one type of mono-, di-, or penta-nucleotide stretches, respectively, and the six remaining being composite with at least two different types of repeats (loci *Ac*-cp3, 4, 5, 16, 31, and 32; Table S1).

We then amplified 33 loci in nine PCR multiplexes, while *Ac*-cp16 was amplified separately. For each multiplex (including three to five loci; see Table S1), the annealing temperature of all primers was the same, while the size of PCR products of each locus was as different as possible. The cost-effective labeling method of Schuelke (2000) was used for PCR reactions. An 18-bp tail of M13 was added to each forward primer (Table S1). All primer pairs and specific characteristics of generated fragments are given in Table S1. Each PCR reaction (25 µL) contained 10-ng DNA template, 1 × reaction buffer, 5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µmol of one universal fluorescent labeled M13(–21) primer (5'-TGTAAAACGACGGCCAGT-3'; labeled with one of the four following fluorochromes: YAK, 6-FAM, AT565, or AT550), 0.04 µmol of the reverse primer, 0.01 µmol of the forward primer, and 0.5 U of *Taq* DNA polymerase (Promega). The reaction mixtures were incubated in a T1 thermocycler for 2 min at 95 °C, followed by 28 cycles of 30 s at 95 °C, 30 s at 57 °C, and 1 min at 72 °C, and then by 9 cycles of 30 s at 95 °C, 30 s at 51.5 °C, and 1 min at 72 °C. The last cycle was followed by a 20-min extension at 72 °C. Two series of PCR multiplexes were finally mixed together with GeneScan-600 LIZ as internal standard to run 17 loci at the same time (Table S1): Mix I included PCR multiplexes I-A, I-B, I-C, and I-D, while Mix II included II-A, II-B, II-C, II-D, II-E, and II-F. PCR products were separated on an ABI Prism 3730 DNA analyzer (Applied Biosystems), and the size of amplified fragments was determined with Geneious v9.1.2 (Kearse et al., 2012).

2.4. Data analyses

We first annotated and aligned our genomic data: *i*) Plastomes were annotated using PGA (Qu et al., 2019) and aligned using Mauve to detect potential rearrangements. They were then aligned using Mafft v7.313 (Katoh and Standley, 2013). The resulting alignment was manually curated in some short-inverted regions (less than 30 bp) and finally cleaned using TrimAl (parameters -gt 0.8 -st 0.001 -cons 60; Capella-Gutierrez et al., 2009). For each plastid gene, coding sequence (cds) was aligned as codons using Decipher (Wright, 2016); *ii*) Secondly, we annotated mtDNA regions. The assembled mitogenome of *A. catechu* ('PNG-1') was not circular and composed of a single contig, whereas the mitogenome of *A. macrocalyx* ('PNG-5') was composed of two contigs (see below). We annotated these contigs by transferring annotations from the *Cocos nucifera* mitogenome (KX028885) in Geneious. As some genes were split between the two contigs in 'PNG-5' and unannotated, the annotation was completed using blastn searches against *C. nucifera* annotations. Mitochondrial genes of these two accessions were extracted and aligned with those assembled from other accessions using Mafft, and their cds was aligned as codons with Decipher (Wright, 2016); *iii*) Finally, the nrDNA region composed of ETS, 18S, ITS1, 5.8S, ITS2 and 26S was aligned with Mafft.

Phylogenetic analyses were conducted using IQ-Tree2 (Minh et al., 2020b). *Cocos nucifera* (from GenBank) was used as an outgroup in all analyses. For each dataset (ptDNA, mtDNA and nrDNA), we used a concatenation approach with an edge-linked proportional partition model and assessed branch support with 1000 ultrafast bootstrap (UFB) replicates with UFBoot2 (Hoang et al., 2018). The best partition scheme (gene, codon positions or gene and codons positions) for each dataset was determined with PartitionFinder v2.1.1 (Lanfear et al., 2017) and the best fitted evolutionary model for each partition was selected according to the best BIC score with ModelFinder, as implemented in IQ-Tree2, using the maximum likelihood (ML).

We also estimated an ML phylogeny for nine *Areca* accessions and the

outgroup (*C. nucifera*) combining nuclear and organellar information (full plastome matrix, mtDNA genes and the nrDNA cluster). We quantified the concordance between this phylogeny and each dataset by calculating the gene concordance factor (gCF) and the site concordance factor (sCF) for each branch of the reference tree (Minh et al., 2020a). The gCF indicates the fraction of individual trees (here, topology built from each compartment) concordant with a given branch, and the sCF, the proportion of sites that support that branch in the alignment. It thus described the presence of sites inside each dataset supporting the combined topology, even if it is different from the one obtained with a given individual dataset.

Finally, we reconstructed a haplotype network based on cpSSR data. A reduced-median network was reconstructed with Network v10 (Bandelt et al., 1999). Multi-state microsatellites were treated as ordered alleles and coded by the number of repeated motifs for each allele (e.g., number of T or A). Basically, this coding strategy assumes that variation at cpDNA microsatellites is mainly due to single-step mutations (Gryta et al., 2017). *Areca* aff. *triandra* from Thailand was used as outgroup. When comparing distantly related taxa, length variation at a locus can be due to a mix of microsatellite polymorphisms plus other indels in flanking regions. For such loci (loci *Ac*-cp16, 18, 27, 31), the allele was treated as missing data in the outgroup.

3. Results

3.1. Assembled genomic data

In the present study, we generated genome skims of ten *Areca* samples (Table S2). The number of paired-end reads generated on each sample ranged from 9′608′805 to 22′981′975 (Table S2). Based on these data, we first assembled and annotated nine complete plastid genomes (ranging from 157′582 to 158′692 bp), excluding *A. mogeana* for which very low sequencing coverage (< 1 ×) hampered the assembly of organellar genomic regions. Sequencing depth of the nine plastomes was superior to 30 ×, except for *A. vestiaria* (23 ×). The mean insert size of reads mapped on this genome was ca. 380–400 bp in libraries reconstructed from recent silica dried samples, but inferior to 250 bp for herbarium samples (Table S2), confirming that DNAs extracted from old specimens were initially degraded. Gene content and order were identical across all nine *Areca* plastomes, which were also collinear with the *C. nucifera* plastome and the *A. catechu* plastome recently released (Rajesh et al., 2020). In addition, we produced the first mitogenome assemblies (766 kb for 'PNG-1', and 731 kb for 'PNG-5') for the *Areca* genus and 37 mitochondrial protein-coding sequences (ca. 59 kb) for six other individuals (excluding *A. mogeana*). Sequencing depth of these mtDNA regions varied from 40 to 50 × for 'PNG-1' and 'PNG-5' to 5–6 × in 'PNG-4' and 'THAI'. Very short segments were not covered in some genes of 'PNG-4', 'THAI', 'PHIL' and 'MOL-1' (1,606, 1,030, 444 and 43 bp of missing data, respectively), whereas mtDNA sequences of other accessions were complete. All genes annotated in the *C. nucifera* mitogenome were annotated in our two mitogenome assemblies, which further supports their completeness. Finally, we obtained nearly complete nrDNA clusters (ca. 6.7 to 7.3 kb; sequencing depth > 36 ×) for the same nine samples. For the *A. mogeana* specimen, sequencing data allowed us to assemble only full ITS and ETS regions of the nrDNA cluster (ca. 2 kb; sequencing depth of 9 ×). Sequence alignments are provided in Supplementary Information (Supplementary Materials S1 to S3).

3.2. Comparative phylogenetic analyses of Areca based on three genomic regions

Similar phylogenetic topologies were revealed on plastid, mitochondrial, and nuclear ribosomal DNAs (Fig. 1). Compared to the topologies based on organellar genomes, only the placement of *A.* aff. *triandra* differs on the nrDNA cluster, but this placement remains poorly supported. In our combined analysis, the topology sustained the organellar scenario. However, concordance factors convey some incongruence (sCF = 60%) meaning that a large proportion of sites do not support this branching. Based on our limited species sampling, we did not reveal any clear phylogeographic structure since *A. vestiaria* from Moluccas (with *A. mogeana* from Borneo on nrDNA) is sister to a clade including other wild *Areca* species collected in PNG (*A. macrocalyx*), Philippines (*A. caliso*), and Thailand (*A.* aff. *triandra*). In contrast, phylogenetics of all three genomes supports one clade including commercial betel nuts from India, PNG, and Vanuatu. The nucleotide diversity revealed in this group is, however, very low, with only 12 SNPs in the full plastome (of which five were detected in genes, and three were non synonymous), four SNPs in mitochondrial genes, and seven variable sites in the nrDNA cluster (Table S3). In addition, accessions 'PNG-2' and 'PNG-3' were identical based on SNPs from organellar genomes and nrDNA, and these two accessions were also not distinguished from accession 'INDIA' based on the plastome sequence deposited in GenBank (Table S3). Such a low variation calls for the use of the most variable regions of the organellar genome to track back the diffusion of matrilineages.

3.3. Plastid genome variation of the betel nut revealed with a cpSSR profiling approach

We finally developed a method targeting cpSSRs, expected to be the most variable regions of the plastome. All 34 tested cpSSR loci were successfully amplified on the five DNAs extracted from recent samples desiccated in silica (i.e., *A. catechu* and *A. macrocalyx* from PNG; Table S4). In contrast, frequent failures were recorded on the three tested herbarium DNAs for the longest loci (>180 bp) when used in PCR multiplexes, as expected on highly fragmented DNA (see above). Nonetheless, their separated amplification was usually successful, except for loci *Ac*–cp10, 16, 19, 20, 30 and 31, which were not amplified on two or three samples (Table S4). For all successful cpSSR amplification, allele sizes were fully congruent with genome sequences. The combination of cpSSR polymorphisms allowed defining a chlorotype for each tree. Each tree analyzed in our study harbors a distinct chlorotype. Nine cpSSR loci revealed polymorphisms in our *A. catechu* samples. Phylogenetic relationships deduced from a reduced-median network (Fig. 3) of betel nut chlorotypes show that the closely related accessions from India and PNG mentioned above ('PNG-2', 'PNG-3', and 'INDIA') were distinguished from each other by one or two loci (*Ac*-cp1 and 23) and form a distinct lineage compared with the accession from Vanuatu and the third PNG sample ('PNG-1' from Kavieng).

4. Discussion

4.1. New genomic datasets for investigating Areca phylogeography and the origins of betel nut

In this study, we generated new genomic datasets for the betel nut and wild relatives. In particular, the first mitogenomes of the *Areca* genus were assembled and annotated, and we also released nine plastomes, mitochondrial genes, and nearly complete nrDNA clusters for five taxa. A comparative phylogenetic analysis proved their utility for distinguishing *Areca* taxa and investigating their phylogeographic history. Overall, phylogenetic information provided by organellar genomes and the nrDNA cluster were congruent (except for the poorly supported position of *A.* aff. *triandra* 'THAI'; Figs. 1 and 2), but our sample is far from being comprehensive. Our phylogenetic trees already constitute backbones for more exhaustive studies of the *Areca* genus, but considering the relatively small number of wild taxa (ca. 40–50 species; Dransfield et al., 2008; Henderson, 2009; Heatubun, 2011), a genome-skimming analysis of all species is a reachable objective. Here, we also analyzed one of the oldest herbarium specimens of *Areca* collected in Maluku during the Dumont d'Urville's expedition in 1839, and another

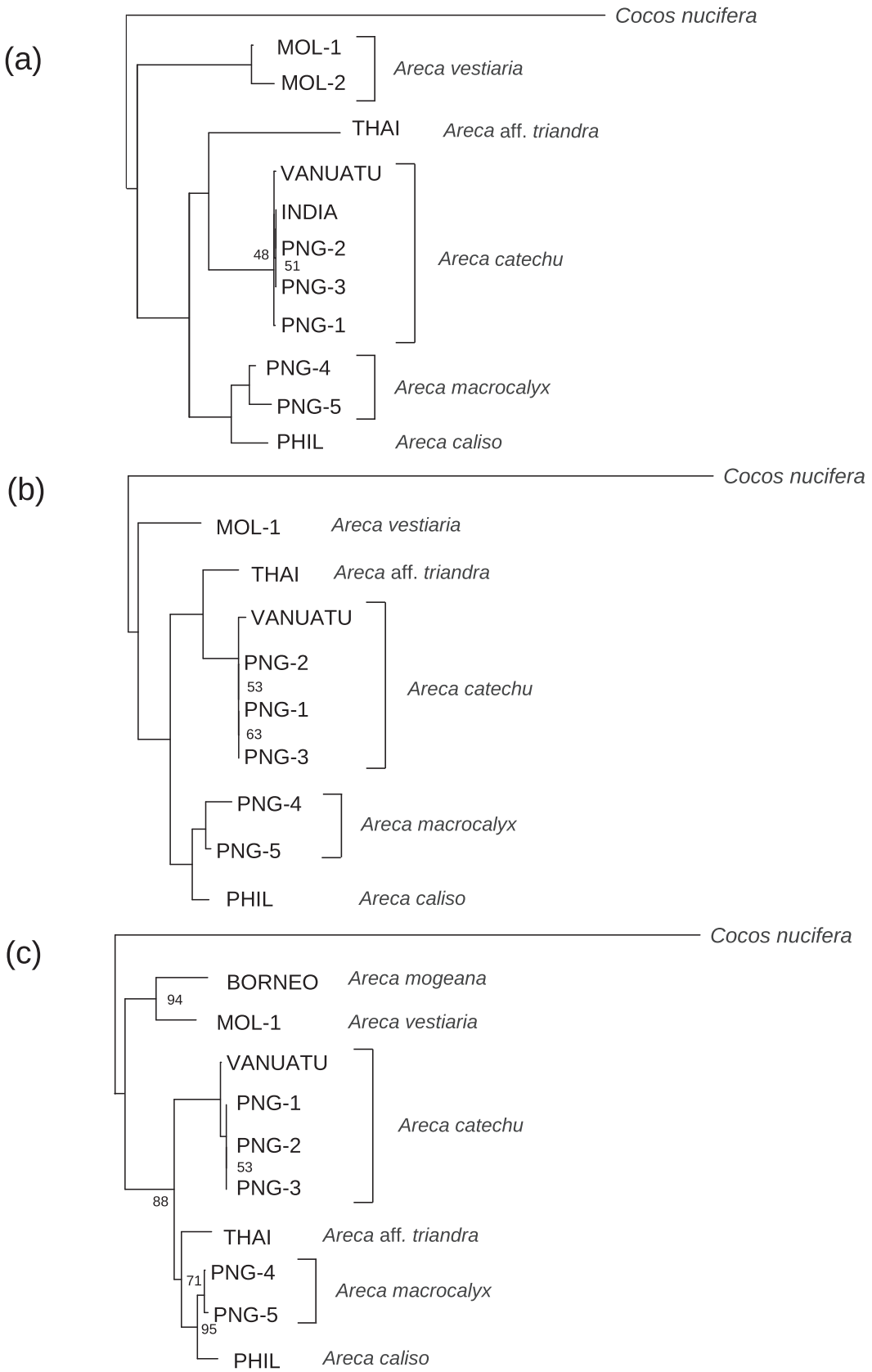

Fig. 1. Maximum likelihood phylogenies obtained with (a) full plastome matrix, (b) mitochondrial protein-coding sequences, and (c) nuclear ribosomal DNA (nrDNA) data. All trees were rooted on *Cocos nucifera*. The scale is in substitution per site. Ultrafast bootstrap support values are indicated on nodes when inferior to 100.

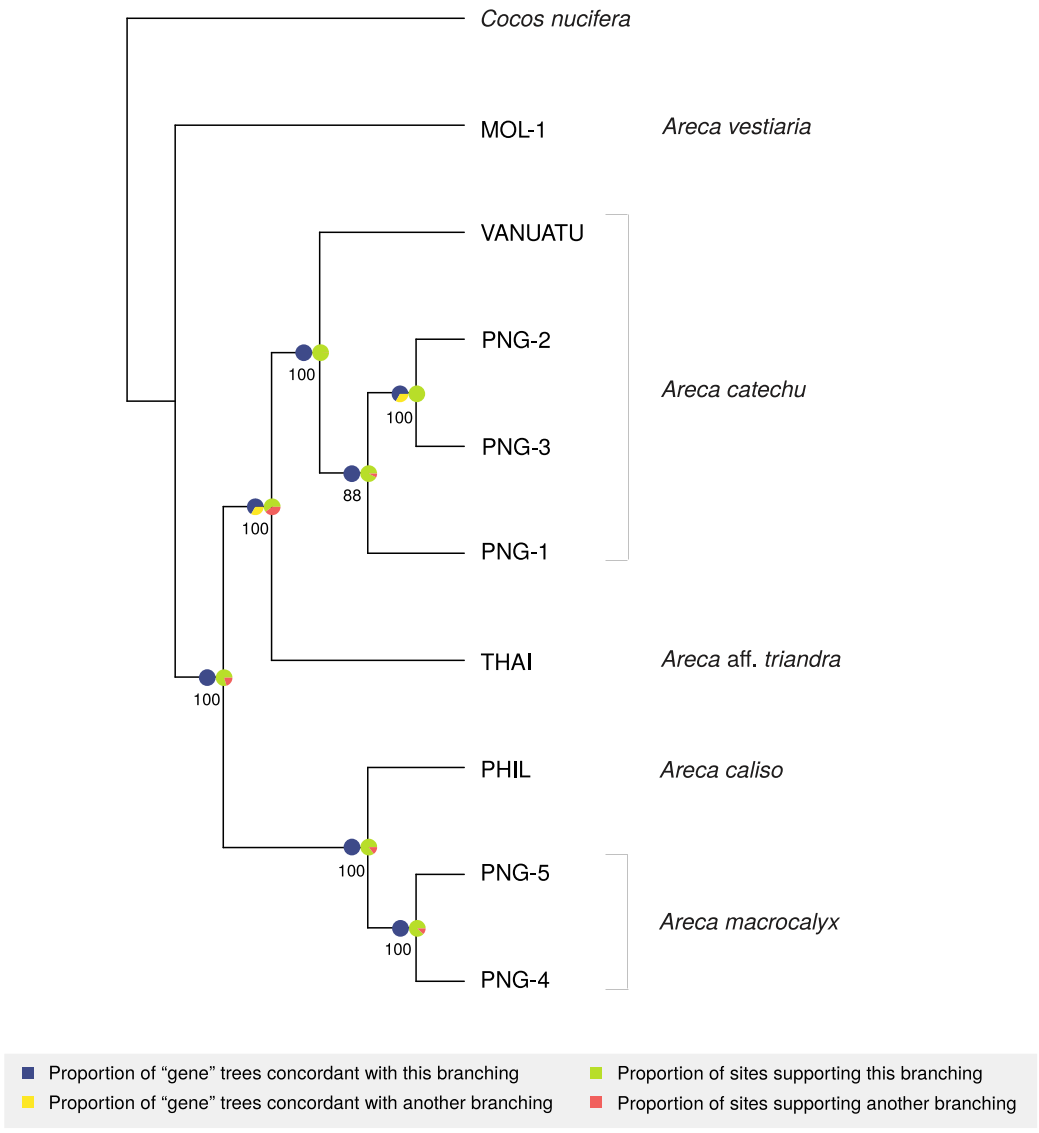

Fig. 2. Maximum likelihood topology estimated from the complete partitioned concatenation of the full plastome matrix, 37 mitochondrial genes, and the nuclear ribosomal DNA cluster. Concordance factors were calculated in relation to the species trees inferred for each partitioned dataset. Gene concordance factors are represented by the blue/yellow pie charts (left), and site concordance factors by the green/orange ones (right). UFB support values are indicated near their respective nodes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

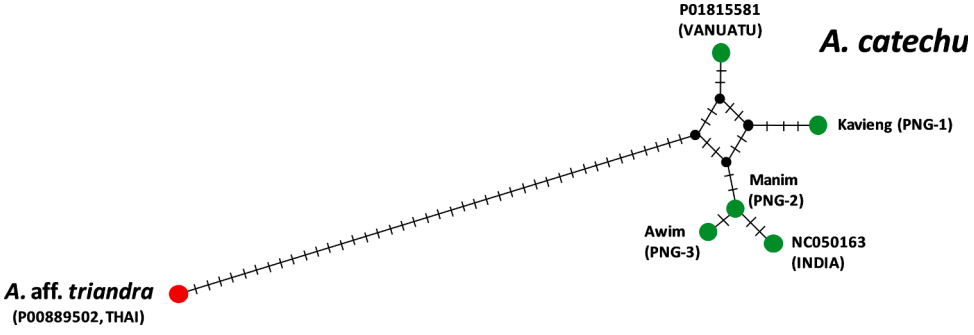

Fig. 3. Network of *Areca catechu* chlorotypes based on cpSSR data (Table S4). A reduced-median network was reconstructed with Network v10 (Bandelt et al., 1999). *Areca* aff. *triandra* (Thailand) was used as an outgroup. The traits on branches represent each individual change. Each haplotype is represented by a green or red color (for *A. catechu* and *A.* aff. *triandra*, respectively). The missing, intermediate nodes are indicated by small black points. PNG = Papua New Guinea. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

collected in Philippines in 1916. This shows how herbarium collections can provide material for phylogenomics studies, and facilitate an exhaustive taxa sampling (e.g., Zedane et al. 2016; Hardion et al., 2021), especially in remote islands of Wallacea (Strijk et al., 2020).

Generating genomic references is also an important prerequisite before investigating betel nut history and characterizing its genetic resources. Such data are key to developing molecular tools (e.g., Chapman, 2019), so we can define the best approaches for screening the genetic diversity of a crop. Organellar genome sequences could also represent appropriate references for archaeogenomic analyses (due to their repeated nature facilitating their analysis). Here, our screening reveals very low nucleotidic variation in plastomes and mitogenomes of the investigated betel nut accessions, which were nevertheless sampled from distant locations (i.e., India, PNG, and Vanuatu). This could result from a genetic bottleneck associated with domestication and its subsequent spread toward Southeast Asia and Australasia, as shown in other

woody crops (Arroyo-García et al., 2006; Besnard et al., 2013). A wide genomic analysis may provide important insights on betel nut domestication, but we recommend that organellar genome variation is screened with appropriate tools to detect genetic variants that may have accumulated in the most recent times during human history. Plastid DNA microsatellites may provide such information (e.g., Besnard et al., 2013).

4.2. Plastid genome profiling to investigate the spread of betel nut matrilineages

A cpSSR profiling method was developed to investigate the spread of maternal lineages of the betel nut and wild relatives. The method proves to work on degraded herbarium DNAs, although the longest loci (>180 bp) were generally more difficult to amplify on such samples. Besides, the transferability of loci among *Areca* samples may allow generating a plastome profiling database on geo-referenced samples (including herbarium specimens), to assist the identification of species and develop a provenance test. A rapid screening of *Areca* species may also be useful to target relevant taxa involved in *A. catechu* domestication. The profiling method on the betel nut revealed a distinct plastid haplotype for each investigated tree, which was not possible based on the 12 SNPs detected in full plastome and mitochondrial gene sequences. This means that plastid DNA diversity may be efficiently screened with such markers, which is also far cheaper than a genome skimming approach. Our network analysis does not reveal any geographic clustering of the five investigated accessions, calling for an extended sampling of betel nut over its whole distribution area to investigate phylogeographic patterns in *A. catechu* and closely related taxa.

4.3. Further prospects

We showed that cultivated betel nut trees from India, Vanuatu, and PNG are very closely related according to organellar DNA, indicating a probable human-mediated spread of the crop in most of its cultivated range as suggested by archaeological evidence (Yen, 1993; Bourke, 2009; Denham, 2010; Fitzpatrick et al., 2013). Yet, wild substitutes may have been used before the introduction of the betel nut. Wild and cultivated arecanuts (both exploited) thus co-exist in numerous places, such as in Indonesia and PNG (Murthy and Pillai, 1982; Heatubun et al., 2012). Documenting betel nut origins may thus address different topics such as the identification of its domestication cradle, the consequences of its multiplication mode on genetic diversity, and the putative role of interspecific hybridizations on its diversification. A combination of nuclear and cytoplasmic genomic data may be very useful to tackle these open questions, and document patterns of gene flow mediated by seeds and pollen. Our cpSSR profiling method can investigate the spread of cultivated matrilineages and contrasted patterns of diversity due to various practices of propagation.

5. Data accessibility

All genomic sequences assembled and analyzed in this study have been deposited in GenBank (accession numbers are given in Table S2). Sequence alignments are also provided in Supplemental Materials.

Funding

This work was supported by the French Ministry of Foreign and European Affairs (French Prehistoric Mission in PNG to FXR), the French Embassy in Papua New Guinea, and by the grant GeneRes (Occitanie-France Olive). We also thank the LABEX TULIP (ANR-10-LABX-0041) and CEBA (ANR-10-LABX-25-01).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank John Muke for assistance in collecting PNG samples and Marc Jeanson (MNHN Paris) for providing herbarium samples. We acknowledge support from the GenoToul bioinformatics facility of Genopole Toulouse Midi-Pyrénées, France. Uxue Suescun and Sandra Fourre gave assistance for cpSSR genotyping.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2021.145845.

References

Amarasinghe, H.K., Usgodaarachchi, U.S., Johnson, N.W., Lalloo, R., Warnakulasuriya, S., 2010. Betel-quid chewing with or without tobacco is a major risk factor for oral potentially malignant disorders in Sri Lanka: A case-control study. Oral Oncol. 46 (4), 297–301.

Arroyo-García, R., Ruiz-García, L., Bolling, L., Ocete, R., López, M.A., Arnold, C., Ergul, A., Söylemezoğlu, G., Uzun, H.I., Cabello, F., Ibáñez, J., Aradhya, M.K., Atanassov, A., Atanassov, I., Balint, S., Cenis, J.L., Costantini, L., Goris-Lavets, S., Grando, M.S., Klein, B.Y., McGovern, P.E., Merdinoglu, D., Pejic, I., Pelsy, F., Primikirios, N., Risovannaya, V., Roubelakis-Angelakis, K.A., Snoussi, H., Sotiri, P., Tamhankar, S., This, P., Troshin, L., Malpica, J.M., Lefort, F., Martinez-Zapater, J.M., 2006. Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. Mol. Ecol. 15 (12), 3707–3714.

Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16 (1), 37–48.

Barrett, C.F., Baker, W.J., Comer, J.R., Conran, J.G., Lahmeyer, S.C., Leebens-Mack, J.H., Li, J., Lim, G.S., Mayfield-Jones, D.R., Perez, L., Medina, J., Pires, J.C., Santos, C., Wm. Stevenson, D., Zomlefer, W.B., Davis, J.I., 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytol. 209 (2), 855–870.

Bendich, A.J., 1987. Why do chloroplasts and mitochondria contain so many copies of their genome? BioEssays 6 (6), 279–282.

Besnard, G., Bervillé, A., 2002. On chloroplast DNA variations in the olive (*Olea europaea* L.) complex: comparison of RFLP and PCR polymorphisms. Theor. Appl. Genet. 104 (6), 1157–1163.

Besnard, G., Khadari, B., Navascués, M., Fernández-Mazuecos, M., El Bakkali, A., Arrigo, N., Baali-Cherif, D., Brunini-Bronzini de Caraffa, V., Santoni, S., Vargas, P., Savolainen, V., 2013. The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. Proc. Roy. Soc. Lond. Ser. B 280 (1756), 20122833.

Besnard, G., Christin, P.A., Malé, P.J., Coissac, E., Lhuillier, E., Lauzeral, C., Vorontsova, M.S., 2014. From museums to genomics: old herbarium specimens shed light on a C_3 to C_4 transition. J. Exp. Bot. 65 (22), 6711–6721.

Bourke, R.M., 2009. History of agriculture in Papua New Guinea. In: Bourke, R.M., Harwood, T. (Eds.), Food and Agriculture in Papua New Guinea. ANU Press, Canberra, pp. 10–26.

Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25 (15), 1972–1973.

Carreel, F., Gonzalez de Leon, D., Lagoda, P., Lanaud, C., Jenny, C., Horry, J.P., Tezenas du Montcel, H., 2002. Ascertaining maternal and paternal lineage within *Musa* chloroplast and mitochondrial DNA RFLP analyses. Genome 45 (4), 679–692.

Chapman, M.A., 2019. Optimizing depth and type of high-throughput sequencing data for microsatellite discovery. Appl. Plant Sci. 7 (11), e11298.

Chen, P.-H., Mahmood, Q., Mariottini, G.L., Chiang, T.-A., Lee, K.-W., 2017. Adverse health effects of betel quid and the risk of oral and pharyngeal cancers. BioMed Res. Int. 2017, 3904098.

Comer, J.R., Zomlefer, W.B., Barrett, C.F., Davis, J.I., Stevenson, D.W., Heyduk, K., Leebens-Mack, J.H., 2015. Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. Am. J. Bot. 102 (6), 888–899.

Cornille, A., Gladieux, P., Smulders, M.J.M., Roldán-Ruiz, I., Laurens, F., Le Cam, B., Nersesyan, A., Clavel, J., Olonova, M., Feugey, L., Gabrielyan, I., Zhang, X.-G., Tenaillon, M.I., Giraud, T., Mauricio, R., 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genet. 8 (5), e1002703.

Denham, T., 2010. From domestication histories to regional prehistory: using plants to re-evaluate Early and Mid-Holocene interaction between New Guinea and Southeast Asia. Food Hist. 8 (1), 3–22.

Dransfield, J., Uhl, N.W., Asmussen, C.B., Baker, W.J., Harley, M.M., Lewis, C.E., 2008. Genera Palmarum: the Evolution and Classification of Palms. Kew: Royal Botanic Gardens, International. Palm Society and LH Bailey Hortorium.

Duminil, J., Besnard, G., 2021. Utility of the mitochondrial genome in plant taxonomic studies. In: Besse, P. (Ed.) Molecular Plant Taxonomy. Methods in Molecular Biology, vol 2222, pp. 107–118. Humana, New York, NY.

Dumont d'Urville, J., Jacquinot, C.H., 1842-1853. Voyage au Pole Sud et dans l'Océanie sur les corvettes l'Astrolabe et la Zélée pendant les années 1837-1838-1839-1840. Gide, Baudry, J. (Eds), Paris.

Dupin, J., Raimondeau, P., Hong-Wa, C., Manzi, S., Gaudeul, M., Besnard, G., 2020. Resolving the phylogeny of the olive family (Oleaceae): confronting information from organellar and nuclear genomes. Genes 11 (12), 1508.

Fitzpatrick, S.M., Nelson, G.C., Reeves, R., 2013. The prehistoric chewing of betel nut (*Areca catechu*) in western Micronesia. People Cult. Ocean. 19, 55–65.

Flowers, J.M., Hazzouri, K.M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A., Ferrand, S., Khierallah, H.S.M., Fuller, D.Q., Aberlenc, F., Fournaraki, C., Purugganan, M.D., 2019. Cross-species hybridization and the origin of North African date palms. Proc. Natl. Acad. Sci. USA 116 (5), 1651–1658.

Gautam, R., Saklani, A., Jachak, S.M., 2007. Indian medicinal plants as a source of antimycobacterial agents. J. Ethnopharmacol. 110 (2), 200–234.

Gheddar, L., Ricaut, F.-X., Ameline, A., Brucato, N., Tsang, R., Leavesley, M., Raul, J.S., Kintz, P., 2020. Testing for betel nut alkaloids in hair of Papuans abusers using UPLC-MS/MS and UPLC-Q-Tof-MS. J. Analytic. Toxicol. 44 (1), 41–48.

Gorrilliot, O., Hong-Wa, C., Rakotonasolo, F., Besnard, G., 2021. Microsatellite-assisted identification and comparative population genetics of Malagasy olive species (*Noronhia* spp., Oleaceae). Bot. Lett. https://doi.org/10.1080/23818107.2021.1912636.

Gryta, H., Van de Paer, C., Manzi, S., Holota, H., Roy, M., Besnard, G., 2017. Genome skimming and plastid microsatellite profiling of alder trees (*Alnus* spp., Betulaceae): phylogenetic and phylogeographical prospects. Tree Genet. Genomes 13 (6), 118.

Hardion, L., Verlaque, R., Kaymak, E., Vila, B., Haan-Archipoff, G., Martinez Martin, M., Tournay, F., Migliore, J., 2021. Plastome sequencing of a 167-year-old herbarium specimen and classical morphology resolve the systematics of two potentially extinct grass species. Bot. J. Linn. Soc. 195 (2), 115–123.

Heatubun, C.D., 2011. Seven new species of *Areca* (Arecaceae). Phytotaxa 28, 6–26.

Heatubun, C.D., Dransfield, J., Flynn, T., Tjitrosoedirdjo, S.S., Mogea, J.P., Baker, W.J., 2012. A monograph of the betel nut palms (*Areca*: Arecaceae) of East Malesia. Bot. J. Linn. Soc. 168 (2), 147–173.

Henderson, A., 2009. Field Guide to the Palms of Southern Asia. Princeton University Press, Princeton, NJ.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35 (2), 518–522.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12), 1647–1649.

Koonlin, T., 1979. The palm economy, in particular the Swamp Sago economy in Southeast Asia. J. Agric. Trad. Bot. Appl. 26 (2), 129–146.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34 (3), 772–773.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth. 9 (4), 357–359.

Li, C., Zheng, Y., Huang, P., 2020. Molecular markers from the chloroplast genome of rose provide a complementary tool for variety discrimination and profiling. Sci. Rep. 10 (1), 12188.

Loo, A.H.B., Dransfield, J., Chase, M.W., Baker, W.J., 2006. Low-copy nuclear DNA, phylogeny and the evolution of dichogamy in the betel nut palms and their relatives (Arecinae; Arecaceae). Mol. Phylogenet. Evol. 39 (3), 598–618.

Magri, D., Vendramin, G.G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., Latałowa, M., Litt, T., Paule, L., Roure, J.M., Tantau, I., Van Der Knaap, W.O., Petit, R.J., De Beaulieu, J.-L., 2006. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. New Phytol. 171 (1), 199–221.

Minh, B.Q., Hahn, M.W., Lanfear, R., 2020a. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37(9), 2727–2733.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020b. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37(5), 1530–1534.

Monteiro, E., Castro, I., Carvalho, M., Martín, J.P., Rosa, E., Carnide, V., 2021. Iberian Peninsula cowpea diversity: chloroplast, microsatellite and morpho-agronomic variability. Syst. Biodiv. 19 (2), 121–134.

Myles, S., Boyko, A.R., Owens, C.L., Brown, P.J., Grassi, F., Aradhya, M.K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., Bustamante, C.D., Buckler, E.S., 2011. Genetic structure and domestication history of the grape. Proc. Natl. Acad. Sci. USA 108 (9), 3530–3535.

Murthy, K.N., Pillai, R.S.N., 1982. Botany. In: Bavappa, K.V.A., Nair, M.K., Prem Kumar, T. (Eds.), The Arecanut Palm. Central Plantation Crop Research Institute, Kasaragod, pp. 11–49.

Navascués, M., Emerson, B.C., 2005. Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. Mol. Ecol. 14 (5), 1333–1341.

Norton, S.A., 1998. Betel: consumption and consequences. J. Am. Acad. Dermatol. 38 (1), 81–88.

Olofsson, J.K., Cantera, I., Van de Paer, C., Hong-Wa, C., Zedane, L., Dunning, L.T., Alberti, A., Christin, P.-A., Besnard, G., 2019. Phylogenomics using low-depth whole genome sequencing: A case study with the olive tribe. Mol. Ecol. Res. 19 (4), 877–892.

Oreca, C., Mutua, A., Kindt, R., Jamnadass, R., Anthony, S., 2009. Agroforestree Database: a tree reference and selection guide version 4.0. World Agroforestry Centre, Kenya.

Oxenham, M.F., Locher, C., Cuong, N.L., Thuy, N.K., 2002. Identification of *Areca catechu* (betel nut) residues on the dentitions of Bronze Age inhabitants of Nui Nap, Northern Vietnam. J. Archaeol. Sci. 29 (9), 909–915.

Pedersen, B.S., Quinlan, A.R., 2018. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34 (5), 867–868.

Petit, R.J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csaikl, U. M., van Dam, B., Deans, J.D., Espinel, S., Fineschi, S., Finkeldey, R., Glaz, I., Goicoechea, P.G., Jensen, J.S., König, A.O., Lowe, A.J., Madsen, S.F., Mátyás, G., Munro, R.C., Popescu, F., Slade, D., Tabbener, H., de Vries, S.G.M., Ziegenhagen, B., de Beaulieu, J.-L., Kremer, A., 2002. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. For. Ecol. Manag. 156 (1-3), 49–74.

Petit, R.J., Kremer, A., Wagner, D.B., 1993. Finite island model for organelle and nuclear genes in plants. Heredity 71 (6), 630–641.

Qu, X.J., Moore, M.J., Li, D.Z., Yi, T.S., 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Meth. 15, 50.

Provan, J., Powell, W., Hollingsworth, P.M., 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol. Evol. 16 (3), 142–147.

Rajesh, M.K., Gangaraj, K.P., Prabhudas, S.K., Prasad, T.S.K., 2020. The complete chloroplast genome data of *Areca catechu* (Arecaceae). Data in Brief 33, 106444.

Roullier, C., Benoit, L., McKey, D., Lebot, V., 2013. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. Proc. Natl. Acad. Sci. USA 110 (6), 2205–2210.

Schaal, B.A., Olsen, K.M., 2000. Gene genealogies and population variation in plants. Proc. Natl. Acad. Sci. USA 97 (13), 7024–7029.

Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18 (2), 233–234.

Sharan, R.N., Mehrotra, R., Choudhury, Y., Asotra, K., Agarwal, S., 2012. Association of betel nut with carcinogenesis: revisit with a clinical perspective. PLoS One 7 (8), e42759.

Sharma, H., Hyvönen, J., Poczai, P., 2020. Development of chloroplast microsatellite markers for giant ragweed (*Ambrosia trifida*). Appl. Plant Sci. 8 (1), e11313.

Staats, M., Erkens, R.H.J., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J.E., Bakker, F.T., Caramelli, D., 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS One 8 (7), e69189.

Staples, G.W., Bevacqua, R.F., 2006. *Areca catechu* (betel nut palm), ver. I.3. In: Elevitch, C.R. (Ed.), Species Profiles for Pacific Island Agroforestry. Permanent Agriculture Resources (PAR). Hōlualoa, Hawai'i.

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot. 99 (2), 349–364.

Strijk, J.S., Binh, H.T., Ngoc, N.V., Pereira, J.T., Slik, J.W.F., Sukri, R.S., Suyama, Y., Tagane, S., Wieringa, J.J., Yahara, T., Hinsinger, D.D., Robillard, T., 2020. Museomics for reconstructing historical floristic exchanges: Divergence of stone oaks across Wallacea. PLoS ONE 15 (5), e0232936.

Sunny, G., Rajan, T.P., 2021. Review on areca nut fiber and its implementation in sustainable products development. J. Nat. Fibers. https://doi.org/10.1080/15440478.2020.1870623.

Thomas, S.J., MacLennan, R., 1992. Slaked lime and betel nut cancer in Papua New Guinea. Lancet 340 (8819), 577–578.

Tollefsrud, M.M., Sønstebø, J.H., Brochmann, C., Johnsen, Ø., Skrøppa, T., Vendramin, G. G., 2009. Combined analysis of nuclear and mitochondrial markers provide new insight into the genetic structure of North European *Picea abies*. Heredity 102 (6), 549–562.

Valdes, C.O., 2014. Betel chewing in the Philippines. Arts Asia 34 (5), 104–115.

Van de Paer, C., Bouchez, O., Besnard, G., 2018. Prospects on the evolutionary mitogenomics of plants: A case study on the olive family (Oleaceae). Mol. Ecol. Resour. 18 (3), 407–423.

Whitmore, T.C., 1977. Palms of Malaya. Oxford University Press, Oxford, Kuala Lumpur, Singapore.

Wilson, M.A., Gaut, B.S., Clegg, M.T., 1990. Chloroplast DNA evolves slowly in the palm family (Arecaceae). Mol. Biol. Evol. 7 (4), 303–314.

Wolfe, K.H., Li, W.H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84 (24), 9054–9058.

Wright, E.S., 2016. Using DECIPHER v2.0 to analyze big biological sequence data in R. R Journal 8 (1), 352–359.

Yen, D.E., 1993. The origins of subsistence agriculture in Oceania and the potentials for future tropical food crops. Econ. Bot. 47 (1), 3–14.

Zedane, L., Hong-Wa, C., Murienne, J., Jeziorski, C., Baldwin, B.G., Besnard, G., 2016. Museomics illuminate the history of an extinct, paleoendemic plant lineage (*Hesperelaea*, Oleaceae) known from an 1875 collection on Guadalupe Island. Mexico. Biol. J. Linn. Soc. 117 (1), 44–57.

Zumbroich, T.J., 2008. The origin and diffusion of betel chewing: a synthesis of evidence from South Asia, Southeast Asia and beyond. eJ. Indian Med. 1 (3), 87–140.