Ancient Polyploidy and Genome Evolution in Palms

Craig F. Barrett
Michael R. McKain
Brandon T. Sinn
Xue Jun Ge
Yuqu Zhang

See next page for additional authors

Follow this and additional works at: https://researchrepository.wvu.edu/faculty_publications

Part of the Biology Commons
Authors
Craig F. Barrett, Michael R. McKain, Brandon T. Sinn, Xue Jun Ge, Yuqu Zhang, Alexandre Antonelli, and Christine D. Bacon
Ancient Polyploidy and Genome Evolution in Palms

Craig F. Barrett, Michael R. McKain, Brandon T. Sinn, Xue-Jun Ge, Yuqu Zhang, Alexandre Antonelli, and Christine D. Bacon

1Department of Biology, West Virginia University
2Department of Biological Sciences, University of Alabama
3Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, PR China
4Department of Biological and Environmental Sciences, University of Gothenburg, Sweden
5Gothenburg Global Biodiversity Centre, Göteborg, Sweden
6Royal Botanical Gardens Kew, Richmond, United Kingdom

*Corresponding author: E-mail: cfb0001@mail.wvu.edu.

Accepted: April 19, 2019

Data deposition: This project has been deposited at the NCBI Sequence Read Archive under the accession BioProject (PRJNA313089).

Abstract

Mechanisms of genome evolution are fundamental to our understanding of adaptation and the generation and maintenance of biodiversity, yet genome dynamics are still poorly characterized in many clades. Strong correlations between variation in genomic attributes and species diversity across the plant tree of life suggest that polyploidy or other mechanisms of genome size change confer selective advantages due to the introduction of genomic novelty. Palms (order Arecales, family Arecaceae) are diverse, widespread, and dominant in tropical ecosystems, yet little is known about genome evolution in this ecologically and economically important clade. Here, we take a phylogenetic comparative approach to investigate palm genome dynamics using genomic and transcriptomic data in combination with a recent, densely sampled, phylogenetic tree. We find conclusive evidence of a paleopolyploid event shared by the ancestor of palms but not with the sister clade, Dasypogonales. We find evidence of incremental chromosome number change in the palms as opposed to one of recurrent polyploidy. We find strong phylogenetic signal in chromosome number, but no signal in genome size, and further no correlation between the two when correcting for phylogenetic relationships. Palms thus add to a growing number of diverse, ecologically successful clades with evidence of whole-genome duplication, sister to a species-poor clade with no evidence of such an event. Disentangling the causes of genome size variation in palms moves us closer to understanding the genomic conditions facilitating adaptive radiation and ecological dominance in an evolutionarily successful, emblematic tropical clade.

Key words: Arecaceae, Arecales, chromosome, dysploidy, genome duplication, genome size.

Introduction

Genomic studies across the eukaryotic tree of life reveal that genome size is not indicative of organismal complexity, known as the “C-value paradox,” or “C-value enigma” (e.g., Thomas 1971; Cavalier-Smith 1978; Lewin 1983; Gregory 2001, 2005). For example, the genomes of some simple chlorophyte algae are orders of magnitude larger than the genomes of many flowering plants, despite multicellularity and the extensive differentiation of tissues found in the latter. Genome size and complexity have been hypothesized to correlate with or even drive rates of speciation, but evidence is equivocal (reviewed by Kraaijeveld [2010]). Instead, polyploidy and genome size variation may be more strongly correlated with species richness among major plant clades (e.g., Soltis et al. 2009; Wood et al. 2009; Jiao et al. 2011; Puttick et al. 2015). Plant genomes vary immensely in size (2,400-fold), from 61 megabases (Mb) in the carnivorous Genlisea (Fleischmann et al. 2014) to the lilioid species Paris japonica, at 148.8 gigabases (Gb; Pellicer et al. 2010). What causes such drastic genome size variation in plants? Genome expansion in plants occurs by well-characterized mechanisms, and polyploidy, including both autoploidy...
and allopolyploidy, is often implicated (e.g., Hawkins et al. 2008; Soltis et al. 2009; Grover and Wendel 2010; Wendel 2015; Kellogg 2016). Genome expansion may also occur via tandem or segmental duplication of chromosomal regions (e.g., Zhang 2003). Genome size reduction is less well understood. From a mechanistic perspective, genomes can decrease in size via fractionation and diploidization following a polyploidy event, wherein chromosomes undergo purging of many duplicated regions and structural rearrangements; illegitimate recombination between chromosomes, where misalignment during synapsis leads to large chromosomal deletions; intrastrand recombination, where misalignments occur within a single chromosome leading to large deletions; homologous recombination during meiosis; and chromosomal inversions, particularly those that expose formerly pericentric regions to the distal, telomeric ends of chromosomes where they can be more easily be deleted (Devos et al. 2002; Bennetzen et al. 2005; Hawkins et al. 2008; Zenil-Ferguson et al. 2016; Ren et al. 2018). Transposable elements also play a major role in both genomic growth via bursts of, for example, copy-paste transposition, and in genomic downsizing by causing misalignments during synapsis (e.g., Bennetzen et al. 2005).

Monocots, comprising nearly one fourth of all flowering plant species, display the largest range of genome size variation among flowering plants (Leitch et al. 2010). Among monocots, the palms (family Arecaceae, with >2,500 accepted species) represent a diverse and ancient clade >100-Myr old (Couvreur et al. 2011; Givnish et al. 2018) and comprise major ecological components of all tropical ecosystems on Earth, especially in Southeast Asia and the Neotropics, where they are particularly diverse and abundant (Uhl and Dransfield 1987; Dransfield et al. 2008; ter Steege et al. 2013; Baker and Dransfield 2016; Balslev et al. 2016). Palms are of immense economic importance as ornamentals, in oil production, and in many tropical areas such as Amazonia they are nearly as important as members of the grass family for human nutrition and shelter (Câmara-Leret 2014; Baker and Dransfield 2016; Balslev et al. 2016). Palms are divided into five subfamilies: Arecioidea (111 genera/1,390 species), Coryxylloideae (8/47), Coryphoideae (47/505), Nypoideae (1/1; Nypa fruticans); and Calamoideae (21/645) (Asmussen et al. 2006; Dransfield et al. 2008; Baker and Dransfield 2016). Most recent analyses based on complete sets of plastid genes support placement of Arecaceae as sister to a small family, Dasyypogonaceae. In contrast to the diverse and pantropical palms, this family contains only four genera and 18 species, and is restricted to Mediterranean habitats of southern and western Australia (Givnish et al. 2010; Barrett et al. 2013, 2016; Givnish et al. 2018). Both families have been placed in the order Arecales (The Angiosperm Phylogeny Group 2016), but recent studies have revealed that these ancient lineages should be recognized as distinct, as together they lack a uniquely definitive synapomorphy and diverged >100 Ma (e.g., Givnish et al. 2018).

The evolutionary dynamics of genomes are poorly understood in the Arecales (sensu stricto, i.e., the palms) and even more so for the Dasyypogonales (Leitch et al. 2010). There is a 33-fold range in genome size across the palms, which typically harbor from 2n = 26–36 chromosomes, though Voaonioala is a remarkable outlier with 2n = 596 (Johnson et al. 1989; Röser 1997). Leitch et al. (2010) compared genome sizes for each chromosome number class among palms, from 2n = 26–36 and concluded that “changes in genome size can occur with no alteration of chromosome number leading to related species having significantly different sized chromosomes.” Evidence for polyploidy in the palms is piecemeal, for example, in the arecoid tribe Coccoaceae (Gunn et al. 2015). Instances of allopolyploidy in sympatry may occur more widely, based on putative hybrid introgression in some genera, but detailed genomic studies are lacking to pinpoint causality (e.g., Attalea, Brahea, Coccothrinax, Copernicia, Geonoma, Latania, Phoenix, Pritchardia, and Pzychosperma; Glassman 1999; Dransfield et al. 2008; Ramirez-Rodriguez et al. 2011; Bacon et al. 2012). Observations based on comparing silent substitutions among duplicate gene pairs (Ks plots) suggest at least that oil and date palms (Elaeis guineensis and Phoenix dactylifera, respectively) show evidence of past whole-genome duplications (WGDs) (Al-Mssallem et al. 2013; Singh et al. 2013). The only formal phylogenomic analyses to include more than one palm species are those of D’Hont et al. (2012) and McKain et al. (2016), providing more conclusive evidence of a shared WGD event among the two model palms, which represent subfamilies Arecioidea and Coryphoideae, respectively.

Several questions remain with respect to genome evolution in the palms. Did WGD events influence genome evolution across the palms and close relatives, and if so, how and at what point in their evolutionary history? Does variation in genome size and chromosome number carry phylogenetic signal across palms and relatives? Here, we use publicly available and newly generated transcriptomic and genomic data, a densely sampled phylogenetic tree, and published data on genome size and chromosome number to address the above questions. Our specific objectives are to 1) reconstruct the evolution of genome size and chromosome number and 2) detect and place the hypothesized WGD event(s), both within a phylogenetic context.

Palm genomes may be model lineages in which to test relationships among trait evolution, biogeography, paleoenvironments, and tropical biodiversity (e.g., Eiserhardt et al. 2011, 2013; Kissling, Baker, et al. 2012; Kissling, Eiserhardt, et al. 2012; Baker and Couvreur 2013; Couvreur and Baker 2013; Bacon et al. 2018). Analyses in palms will help to elucidate patterns of genome size evolution in long-lived monocots, which are typically understudied in the world of evolutionary genomics. Ultimately, our aim is to generate a framework in which to
integrate genome evolutionary dynamics, biogeography, and trait evolution to elucidate the drivers of palm biodiversity.

Materials and Methods

Phylogenetic Trees

Two recently published trees include dense taxon sampling for the palms (Fairby et al. 2016; Antonelli et al. 2017). The “SUPERSMART” tree (Antonelli et al. 2017) was chosen because it has the best taxonomic representation that matches the available genome size, chromosome number, and genome skim data (see below). The tree contains 733 species and 293 genera and is based on all publicly available data from 37 loci (see Antonelli et al. 2017 for details).

Transcriptomic Data

Data were compiled from previously published RNA-seq data sets across monocots from the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra; last accessed December 12, 2018), and the OneKP Project (https://sites.google.com/a/ualberta.ca/onekp/; last accessed December 12, 2018). Complete genomes were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank; last accessed December 12, 2018). Additional RNA-seq data sets were generated for Chamaedorea seifrizii (Arecaceae: Arecoideae) and one representative species from four genera in family Dasypogonaceae: Baxeria australis, Calectasia narragara, Dasypogon bromeliifolius, and Kingia australis (supplementary table S1, Supplementary Material online).

Genome Size and Chromosome Numbers

Genome sizes and chromosome numbers were obtained from the Royal Botanic Gardens, Kew Angiosperm DNA C-values database (Bennett and Leitch 2012; http://data.kew.org/cvalues) and Dransfield et al. (2008), using only “prime” estimates (i.e., excluding those with low confidence). Data and trees were pruned in the “APE” package of R (Paradis and Schliep 2019) to match sampled tips from the SUPERSMART tree at the species level.

Data and Tree Articulation

We attempted to maximize the match of each data set (tree, chromosome number, and genome size) at the species level (supplementary fig. S1 and table S2, Supplementary Material online). In cases where genome size, chromosome number, or genome skim data did not match at the species level, and there were multiple genome size estimates represented by different species within a genus, we used another species of the same genus for the genome size estimate. Although ideally, we would prefer only data from the same species for genome size (further, even from the same individuals per species), using a congener is unlikely to bias our results, because the focus of this analysis is on large-scale relationships among repeat fractions and genome sizes.

Transcriptome Assembly and Gene Tree Reconstruction

RNA-seq data were assembled using Trinity v.2.2.0 (Grabherr et al. 2011; Haas et al. 2013) as described in McKain et al. (2016). Reads were cleaned using Trimmomatic v.0.32 (Bogler et al. 2014) with adapter trimming for TruSeq adapter sequence using one seed mismatch, a palindrome threshold of 30, and a simple clip threshold of 10. After adapter trimming, a sliding window of 10 base pairs a minimum threshold average Phred score of 20 was used to trim reads based on quality. Finally, reads <40 bp in length were discarded. Once assembled, reads were mapped back to transcripts using bowtie v.1.0.0 (Langmead et al. 2009), and read abundance per transcript was estimated using RSEM v.1.2.29 (Li and Dewey 2011) using the “align_and_estimate_abundance.pl” script packaged with Trinity. FPKM (fragments per kilobase of exon per million fragments mapped) was estimated for each gene identified by Trinity. The percentage of mapped fragments per isoform was estimated and transcripts with a value of <1% were removed from further analysis. FPKM filtered transcripts were translated using the RefTrans pipeline (McKain et al. 2016, https://github.com/mrmckain/RefTrans). Transcripts were aligned to gene models from the Ananas comosum v.1.0 (Ming et al. 2015), Asparagus officinalis v.1.0 (Harkess et al. 2017), E. guineensis v.1.0 (Singh et al. 2013), Oryza sativa v.7.0 (Kawahara et al. 2013), Phalaenopsis equestris v.1.0 (Cai et al. 2015), and P. dactylifera v.1.0 (Al-Mssallem et al. 2013) genomes using BlastX with an e-value cutoff of 1.0 × e−10 (Camacho et al. 2009). BLAST results were filtered to identify best hits as defined by transcript and gene model pairs with the lowest e-value and at least 85% bidirectional overlap. Best hit gene models were used to translate transcripts using GeneWise 2.2.0. (Binney et al. 2004). The longest translation for each transcript were used, and if internal stop codons were identified, they were removed from assemblies.

OrthoFinder v.2.2.1 (Emms and Kelly 2015) under default settings was used to circumscribe putative gene families. Diamond v.0.9.19 (Buchfink et al. 2014) with an e-value cut off of 0.001 and the BlastP algorithm was used to align sequences to each other for the initial steps of OrthoFinder. In addition to transcriptomes, gene models from genome sequences for P. dactylifera v.1.0, E. guineensis v.1.0, Musa acuminate v.1.0 (D’Hont et al. 2012), A. comosum v.1.0, O. sativa v.7.0, Pha. equestris v.1.0, and As. officinalis v.1.0 were used in gene family estimation. Orthogroups were filtered to remove those with sequences from <12 taxa. Amino acid sequences for each orthogroup were aligned using MAFFT v.7.313 with
Identification and Phylogenetic Placement of WGD Events

After filtering for a minimum number of 12 taxa per tree, a total of 6,242 gene trees were used to identify and phylogenetically place putative WGD events. The software PUG (McKain et al. 2016) was used to identify putative gene duplications that coincide with the topology of the reconstructed coalescence-based species phylogeny. We ran PUG with the “estimate_paralogs” parameter flag, which has PUG identify all possible paralogs in a given gene tree by identifying all possible transcript pairs derived from the same taxon in a single gene tree. Each multilabeled gene tree was rerooted to a non-Arecaceae and Dasypogonaceae outgroup with preference given in the order: Acorales, As. officinalis, Pha. equestris, Typha latifolia, A. comosum, Neoregalia carolinae, O. sativa, Hanguana malayana, Costus pulverulentus, Musa acuminate, and Tradescantia paludosa. With PUG, each putative paralog pair was queried to identify the most recent common ancestor node in the gene tree. The taxonomic composition of the subtree identified by the most recent common ancestor node was used to identify the equivalent node in the species tree. A placement of the duplication on the species tree was considered acceptable if taxa above the node match those in the gene tree and at least one species sister to this clade in the species tree was found sister to the equivalent clade in the gene tree. For all gene trees and paralogs, we ran PUG to identify both unique duplications (the default) and all duplications (flag “all_pairs”) to identify support for putative WGD events.

Ancient State Reconstruction, Shifts, and Phylogenetic Signal

We reconstructed ancestral genome sizes and chromosome numbers initially in the “APE” and “PHYTOOLS” (“contmap” function, Revell 2012) under a Brownian Motion Model. We further applied an Ornstein–Uhlenbeck model to investigate evidence of significant shifts in trait values over time and across the tree using the R package “l1ou” (Khabbazian et al. 2016), which requires no a priori assumptions on the locations of trait shifts. We additionally analyzed evolutionary changes in chromosome number across the tree with ChromEvol (Glick and Mayrose 2014). This software compares explicit models of chromosome evolution by parameterizing ascending and descending dysploidy (where the current number of chromosomes, \( j = i + 1 \) or \( i - 1 \), respectively, where “i” represents the ancestral chromosome number); WGD (\( j = 2i \)); demipolyploidy (\( j = 1.5i \)); chromosome number changes involving a “base” haploid chromosome number (\( x \)) and linear versus constant rates of change, where linear changes in chromosome number are dependent upon the current chromosome number. We removed Voanioala gerardii (2n = 596) from the analysis because the sampling in that clade is inadequate to reconstruct such a drastic change in chromosome number. We tested ten models of chromosome evolution under the same set of dysploidy parameters as above. We compared the fit of alternative models via the Akaike Information Criterion (Akaike 1974) and Akaike Weights (Wagenmakers and Farrell 2003). We tested for correlation between log-transformed genome size and chromosome number using phylogenetically independent contrasts (Felsenstein 1985; Garland et al. 1992).

Results

Evidence of WGD in Palms

We found unequivocal evidence for an ancient WGD event shared by all representatives of the palms included here, but not shared with the sister clade, Dasypogonales (fig. 1). Coalescent analysis of relationships based on 1,102 single copy nuclear loci yields a tree with representative Arecoidae sister to Coryphoidae, which together are sister to the monotypic Nypoideae, with Maurita, representing the Calamoideae, the subfamily sister to rest of Arecaceae (fig. 1). Ceroxyloideae were not sampled here. We analyzed a total of 6,242 gene families and detected 2,685 unique gene duplications supporting the species tree topology with a minimal
The bootstrap value of 80, representing 31.5% of all sampled gene families. The palms shared 278 unique gene duplications (3,321 paralog pairs), representing 4.6% of all gene families analyzed.

**Genome Size**

We found a lack of phylogenetic signal for genome size (fig. 2, n = 54 species; Pagel’s $\lambda = 7.97 \times 10^{-10}$, $P = 1.0$; Blomberg’s $K = 0.47$, $P = 0.6$). The ancestral genome size for palms under a BM model is $3.6$ Gb (95% confidence interval, or CI = 0.74 to 8.0 Gb). We found limited evidence for significant shifts in genome size; all of these shifts are increases relative to inferred ancestral values, in *Borassus*, *Coccothrinax*, *Pinanga*, *Iriartea*, and *Voanioala* (fig. 2). Comparison of chromosome number and genome size via phylogenetically independent contrasts yields no significant correlation ($F = 0.3832$, $P = 0.54$).

**Chromosome Number**

Ancestral state reconstruction of diploid chromosome number as a continuous character under a BM model yielded a pattern of phylogenetic signal (supplementary fig. 82A, Supplementary Material online). The ancestral 2$n$ value under BM is $32$ for palms ($n = 195$ species). There is a reduction to $2n = 26$ in *Calamus* (subfamily Calamoideae), and a general increase to $2n = 36$ in subfamily Coryphoideae. Chromosome number is unchanged at the crown nodes of subfamilies Ceroxylioideae and Arecoideae, and a reduction to 2$n = 26$ is again observed in many species of Chamaedorea, for which there is dense sampling relative to other genera. A putative chromosome doubling is observed in *Arenga caudata* relative to all other members of this genus (ancestral 2$n = 32 \rightarrow 64$), but few other such events are observed in our data set. *Voanioala gerardi*, with 596 chromosomes, was removed as an outlier. We found evidence for 77 shifts in chromosome number across the palms sampled (OU model, $BIC = 5,739.041$; supplementary fig. 83B, Supplementary Material online). We also found significant phylogenetic signal for chromosome number (Pagel’s $\lambda = 0.41$, $P = 2.5 \times 10^{-10}$; Blomberg’s $K = 0.29$, $P = 0.001$).

A model of linear dependency had the best fit to our data among ten different models of chromosome evolution in ChromEvol (AIC weight = 0.264; supplementary table S4, Supplementary Material online). The maximum-likelihood estimate for ancestral chromosome number was $2n = 30$, though posterior probability estimates were low for the deepest nodes of the tree (i.e., $PP < 0.7$, fig. 3). ChromEvol detected 34 changes in chromosome number in contrast to the 77 shifts identified under an OU model. Most changes in chromosome number were ascending dysploidy, that is, increases in chromosome number of $n \geq 1$ (fig. 3C), and there was one possible case of WGD in *Arenga caudata* ($2n = 32 \rightarrow 64$).
Discussion

Our principal objective was to investigate the evolutionary history of genome evolution in the palms. We found unequivocal evidence of a WGD event shared by all palm subfamilies but not with the sister clade, the monocot order Dasypogonales. We also found evidence of phylogenetic signal for chromosome numbers, evolving predominantly via a linear model of dysploid change.

Shared WGD in the Palms

We found evidence for a shared WGD event across all palm subfamilies, suggesting that polyploidy likely played a role in the diversification and evolutionary success of this emblematic tropical clade (fig. 1). With our data it may be impossible to infer whether this was the result of auto- versus allo-polyploidy: coupled with extinction, accumulation of substitutions among retained duplicates over long temporal scales has likely saturated any patterns that could be used to distinguish between these two processes. Methods used to detect ancient allopolyploidy mostly center around deep reticulate patterns or preferential paralog retention from one parental species, but all these methods require at least some knowledge of the potential donor lineages (see Clark and Donoghue 2017). The palm WGD event must have occurred between ~119 and 85 Ma, that is, after the estimated split of orders Arecales and Dasypogonales but before the first divergence of subfamily Calamoideae from the rest of the palms (Couvreur et al. 2011; Givnish et al. 2018). Although previous studies have alluded to a palm WGD, the hypothesis was only based on divergence comparisons of paralogs within genomes and limited taxon sampling (Al-Mssallem et al. 2013; Singh et al. 2013; He et al. 2015; but see D’Hont et al. 2012; McKain et al. 2016 for explicit phylogenomic comparisons). Such “Ks” comparisons, in which frequency distributions of divergence among paralogs are compared within individual genomes or transcriptomes, are informative for evidence of WGD within a particular genome, but they do not provide a rigorous, phylogenetic, comparative test of shared WGD among taxa. In the present analysis, we definitively and precisely confirm the phylogenetic placement of a palm WGD event, moreover indicating that the palm event is older than has been recently hypothesized (70–75 Ma; e.g., van de Peer et al. 2017).

The fact that this WGD event is not shared with the sister clade of palms, order Dasypogonales, is of high significance in terms of potential implications for palm diversification. A growing number of examples like that of Arecales–Dasypogonales is being revealed with the expansion of phylogenomic studies (e.g., see Soltis et al. 2009; Renny-Byfield and Wendel 2014; Panchy et al. 2016). The most comprehensive analysis to date across angiosperms, using RNA-seq data from the 1KP project, revealed that 70 of 99 WGD events are associated with increases in species richness of one clade relative to a species-poor sister clade (Landis et al. 2018). Here, we present a scenario of a species-rich, evolutionarily successful, ecologically dominant, widespread clade with evidence of ancient WGD prior to or coincident with an adaptive
changes (chromosome number), allowing both ascending and descending
diploid in ChromEvol (i.e., chromosome number changes depend on the current
tion of chromosome number based on a model of linear rate dependency
FIG. 3. —Continued
numbers, and numbers in brackets indicate the ML estimate of chromo-
some number for that node. Asterisks refer to the posterior probability for
the highest-likelihood reconstruction of chromosome number. Black dots
 correspond to four of the five palms subfamilies sampled.
verify this WGD event through increased sampling of Arecoidae, and based on the low support values and the putative paraphyly of Howea in the coalescence tree, this may be an artifact. We detected both the sigma (228 unique duplication, Bootstrap = 80) and tau (731 unique duplications, Bootstrap = 80) events described in earlier analyses (McKain et al. 2016), with tau after the divergence of Acorus, and sigma prior to the diversification of Poales. We also confirmed previously identified events in Bromeliaceae (196 unique duplications, Bootstrap = 80; McKain et al. 2016), Commelinales + Zingiberales (283 unique duplications, Bootstrap = 80, D’Hont et al. 2012), and Zingiberales (538 unique duplications, Bootstrap = 80, D’Hont et al. 2012) (supplementary table S3, Supplementary Material online). There was also signal for a commelinid WGD event occurring after the divergence of Asparagales from the remainder of the monocots but is likely an artifact of sampling.

Evolution of Genome Size and Chromosome Number

Genome size is not correlated with chromosome number in the palms when accounting for phylogenetic relationships, nor does it carry phylogenetic signal based on our current sampling. Gene space varies among palms, from over 35,000 genes in oil palm to over 40,000 in date palm (Al-Mssallem et al. 2013; Singh et al. 2013). Further, the oil palm genome reveals evidence for a role for segmental gene duplications in gene space expansion (Singh et al. 2013). An estimate based on a recently published transcriptome of N. fruticans, a monotypic species of mangrove-growing palms, reveals that up to 45,000 genes may be present (>32,000 were identified via BLAST searches), but these numbers carry great uncertainty as only leaf tissue was sampled (He et al. 2014). Moreover, intragenic and even intraspecific variation in genome size can be substantial (e.g., in Dypsis, Phoenix, Pinanga; summarized in Dransfield et al. [2008] with references therein) necessitating population-level sampling.

We identified major trends in chromosome number evolution across the palms, even with only 195 species sampled for chromosome number and phylogenetic tree information. By explicitly modeling chromosome number across the tree, we detected ∼33 changes in chromosome number, which is fewer than the number of significant shifts detected under an OU model (fig. 3). The treatment of chromosome number as a continuous character may be misleading, and thus explicit models of changes in chromosome number are necessary to effectively capture the evolutionary dynamics of changes across the tree. A linear model of chromosome evolution had the best fit out of ten alternative models (supplementary table S4, Supplementary Material online). This is a state-dependent model in which chromosome number changes are dependent upon the current chromosome number (Glick and Mayrose 2014). Although <10% of the ∼2,500 palm species were sampled in this study, this suggests that sampling was enough to track a linear mode of evolution across many clades (fig. 3). Large sampling gaps would be expected to obscure the pattern of chromosome number changes; for example, a linear dysploid transition from 2n = 30 → 32 → 34 → 36 → 34 → 32 within a lineage or clade might be observed as 2n = 30 → 36 → 32 if the taxa with 2n = 32, 34, and 34 are not sampled, respectively.

Specifically, ascending dysploidy appears to be the predominant mode of chromosomal change in palms based on the data available, suggesting an overall net trend to more chromosomes. The only information on chromosome number available for the sister clade of palms, Dasyponogonaceae, is that of Dasyponogon hookeri, which contains less than half the ancestral chromosome number of palms (2n = 14 vs. 2n = 30, fig. 3; Röser 2000; Leitch et al. 2010). It is plausible that a WGD event in the palm ancestor not shared with Dasyponogonaceae may be responsible for this difference. It would be surprising to observe such a conspicuous pattern of chromosome number “doubling” given the propensity for idiosyncratic chromosomal number change post-WGD; such a doubling after the split of ancestral Arecaceae and Dasyponogonaceae would have had to persist for >100 my of evolution (based on the divergence time estimates in Givnish et al. [2018]). However, just as palms display some of the slowest substitution rates among monocots (see Barrett et al. 2016), plant taxa with relatively longer generation times
generally experience slower rates of postpolyploid diploidization, and perhaps the same is true for descending dysploidy (Mandáková and Lysak 2018).

Our finding of no significant relationship between genome size and chromosome number corroborates an earlier analysis based on comparison of genome size across different categories of chromosome numbers (Leitch et al. 2010). Changes in chromosome number can, however, be an important evolutionary force involved in species diversification, often following a polyploidy event. During post-WGD diploidization and fractionation, dysploid changes in chromosome number can result in reproductive isolation and thus cladogenesis (e.g., Dodsworth et al. 2016; Clark and Donoghue 2017; Mandáková and Lysak 2018). Although WGD (genome doubling or additive fusion) is an important factor in plant diversification in many clades, less study has been devoted to the evolutionary consequences of dysploidy, which appears to be the predominant mode of chromosomal evolution in palms. Additional sampling of both Dasypogonaceae and Arecales is needed, as is a more inclusive, phylogenetically comparative analysis of chromosome number across monocots, for example, including both anagenetic and cladogenetic changes (e.g., chromoSSE; Freyman and Höhna 2018).

Conclusions and Future Directions

Here, we have unequivocally identified an ancient WGD event shared by all palms and characterized the predominant mode of chromosomal change in palms as dysploidy. Remaining questions include the role of repetitive elements in palm genome size evolution and how different genomic attributes have collectively influenced species diversification during the long evolutionary history of this ecologically dominant, evolutionarily successful clade. In the future, it will be critical to obtain whole-genome sequences for multiple representatives of each palm subfamily (including the genome of N. fruticans, the sole member of subfamily Nypoideae), along with each of the four genera of Dasypogonaceae. These genomic resources will allow 1) comparative analyses of genome architecture and synteny, 2) analysis of gene family expansion and contraction with respect to adaptive radiation of the palms, 3) ancestral reconstruction of genome content and architecture (i.e., gene family copy numbers, gene order along chromosomes, and repeat content), and 4) associations of genomic features, important phenotypic traits, ecology, biogeography, and species diversification rates. Such a densely sampled, integrative framework in the palms will advance our understanding of the evolution of tropical biodiversity.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments

This work was supported by the International Partnership Program of Chinese Academy of Sciences, Grant No. 151644KYSB20160005, NSF Awards DEB-0830009 and DEB-0830020, and a W Vu-Program to Stimulate Competitive Research grant. We thank the staff at Montgomery Botanical Center, Fairchild Tropical Botanical Garden, South China Botanical Garden, Xiamen Botanical Garden, and Xishuangbanna Tropical Botanical Garden. We also thank Cecile Ané, William Baker, Sidonie Bellot, John Conran, Jerry Davis, Steven DiFazio, Jeff Doyle, Wolf Eiserhardt, Alejandra Gandolfo, Tom Givnish, Sean Graham, James Leebens-Mack, Ilia Leitch, Claude de Pamphilis, Jaume Pellicer, Chris Pires, Dennis Stevenson, and Wendy Zomlefer for data access and helpful discussions.

Literature Cited


Downloaded from https://academic.oup.com/gbe/article-abstract/11/5/1501/5481000 by guest on 05 May 2020


Associate editor: Shu-Miaw Chaw