New Phytologist

Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes

Haomin Lyu[1], Ziwen He[1], Chung-I Wu[1,2] and Suhua Shi[1]

[1]State Key Laboratory of Biocontrol and Guangdong Key Laboratory of Plant Resources, Sun Yat-sen University, Guangzhou 510275, China; [2]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

Author for correspondence:
Suhua Shi
Tel: +86 20 84113677
Email: lssssh@mail.sysu.edu.cn

Received: *17 May 2017*
Accepted: *10 August 2017*

New Phytologist (2018) **217:** 428–438
doi: 10.1111/nph.14784

Key words: adaptive evolution, convergent evolution, genomic parasite, mangrove, marginal habitat, transposon.

Summary

- Several clades of mangrove trees independently invade the interface between land and sea at the margin of woody plant distribution. As phenotypic convergence among mangroves is common, the possibility of convergent adaptation in their genomes is quite intriguing.
- To study this molecular convergence, we sequenced multiple mangrove genomes. In this study, we focused on the evolution of transposable elements (TEs) in relation to the genome size evolution. TEs, generally considered genomic parasites, are the most common components of woody plant genomes. Analyzing the long terminal repeat-retrotransposon (LTR-RT) type of TE, we estimated their death rates by counting solo-LTRs and truncated elements.
- We found that all lineages of mangroves massively and convergently reduce TE loads in comparison to their nonmangrove relatives; as a consequence, genome size reduction happens independently in all six mangrove lineages; TE load reduction in mangroves can be attributed to the paucity of young elements; the rarity of young LTR-RTs is a consequence of fewer births rather than access death.
- In conclusion, mangrove genomes employ a convergent strategy of TE load reduction by suppressing element origination in their independent adaptation to a new environment.

Introduction

Convergent adaptation of phylogenetically distinct organisms to the same environment is a popular study subject in evolutionary biology because it provides unique mechanistic insights into the operation of natural selection (Morris, 2003). While phenotypic convergence has been extensively documented (Losos, 2011), a long-standing question has been whether convergence is typically driven by the same underlying molecular mechanisms. Occasionally, genes involved in highly specialized traits, such as echo-location and cardenolide resistance in insects, are identified, often exciting great interest (Teeling, 2009; Dobler *et al.*, 2012; Zhen *et al.*, 2012). A more ambitious task is to ask whether molecular convergence at the genomic scale may also exist. With many whole genome sequences becoming available, there have been many efforts to carry out such studies (Parker *et al.*, 2013; Foote *et al.*, 2015).

Adaptive convergence at the genomic level, however, is a noisy process as a result of the vast number of sites that can potentially evolve convergently by chance (Thomas & Hahn, 2015; Zou & Zhang, 2015). Foote *et al.* (2015) found that the convergence signal was no weaker in the genomes of a control group (terrestrial mammals) than in marine mammals independently invading the ocean. These studies thus collectively point to a crucial requirement for detecting convergent adaptation: a high degree of environmental similarity that confers strong directional selection. Perhaps whales, manatees, and seals inhabit environments that are sufficiently disparate to produce high amounts of noise mimicking convergent evolution (Foote *et al.*, 2015). For this reason, mangroves appear to be an ideal system for studying genome-wide convergence.

Mangrove trees colonize the margin of woody plant distribution. They comprise several lineages that independently invade tropical intertidal zones, at the interface between terrestrial and marine ecosystems (Tomlinson, 1986). Mangrove species thus grow in similar environments characterized by high salinity, hypoxia, tidal fluctuations, strong ultraviolet (UV) light, and high temperature (Rothschild & Mancinelli, 2001; Giri *et al.*, 2011). To adapt to such marginal environments for woody plants, mangroves have evolved many shared adaptive traits, including salt secretion, vivipary, and aerial rooting systems (Tomlinson, 1986; Parida & Jha, 2010). The shared phenotypes reflect comparable selective pressures that might also drive genetic and genomic convergence.

To investigate the molecular basis of convergent adaptation, we sequenced or resequenced the genomes of eight mangrove species or subspecies (Xu *et al.*, 2017b; Z. He *et al.*, unpublished). At the genomic level, convergence can be manifested in many

ways. Xu *et al.* (2017a) focus on amino acid substitutions by using the newly developed convergence at conservative sites (CCS) method to suppress background noise. In this study, we focus on the composition of repetitive sequences, the mechanisms for removing them, and their potential effects on adaptation. Repetitive sequences are the main determinant of plant genome size (Petrov, 2001; Bennetzen & Wang, 2014) and transposable elements (TEs) are the key components of repetitive elements.

Transposable elements are often considered to be 'selfish' or 'parasitic' because of their ability to multiply within the host genome without providing any direct fitness benefit (Orel & Crick, 1980; Levin & Moran, 2011; Fedoroff, 2012). Indeed, previous studies on multiple lines of *Drosophila melanogaster* have demonstrated that fitness is negatively correlated with TE number (Houle & Nuzhdin, 2004; Pasyukova *et al.*, 2004; Papaceit *et al.*, 2007). In plants, Vinogradov (2003, 2004a,b) found a positive correlation between genome size increase associated with TE activity and the extinction probability. TEs often proliferate faster than they can be removed, thus contributing to genome size growth (Brookfield, 2005; Lisch, 2013). Following earlier studies (Ibarra-Laclette *et al.*, 2011, 2013; Kelley *et al.*, 2014; Zhang *et al.*, 2014), we hypothesize that the trend might be reversed when the host is under long-term stress – for example, when it moves into a new habitat. We therefore analyze changes of TE load in mangroves as compared with their nonmangrove relatives.

Materials and Methods

Data mining of long terminal repeat retrotransposons

We collected whole genome sequences of four true mangroves (Xu *et al.*, 2017b; Z. He *et al.*, unpublished) and 29 nonmangrove plants (The Arabidopsis Genome Initiative, 2000; Tuskan *et al.*, 2006; Jaillon *et al.*, 2007; Ming *et al.*, 2008; Huang *et al.*, 2009, 2013; Chan *et al.*, 2010; Velasco *et al.*, 2010; Argout *et al.*, 2011; Dassanayake *et al.*, 2011; Hu *et al.*, 2011; Shulaev *et al.*, 2011; Wang *et al.*, 2011, 2012a,b, 2014; Xu *et al.*, 2011, 2013; Young *et al.*, 2011; Prochnik *et al.*, 2012; The Tomato Genome Consortium, 2012; Wu *et al.*, 2012; Chamala *et al.*, 2013; Hellsten *et al.*, 2013; Ibarra-Laclette *et al.*, 2013; Verde *et al.*, 2013; Denoeud *et al.*, 2014; Myburg *et al.*, 2014; Schmutz *et al.*, 2014). The four true mangroves are *Avicennia marina* (Forssk.) Vierh., *Rhizophora apiculata* Blume, *Sonneratia alba* Sm., and *Sonneratia caseolaris* (L.) Engl. Their genomic sequences are accessible from http://evolution.sysu.edu.cn/Sequences.html, or by contacting the authors. [Correction added after online publication 27 September 2017: the text in the previous sentence has been corrected.] Genome sequences of the 29 nonmangrove plants, representing the relatives of the four true mangroves, were downloaded from Phytozome v10.1 and other databases (Supporting Information Table S1). We used the LTRharvest work flow (Ellinghaus *et al.*, 2008) and LTRdigest (Steinbiss *et al.*, 2009) to *de novo* predict long terminal repeat retrotransposons (LTR-RTs) (Fig. S1). We required that a candidate LTR-RT be separated by 1 to 15 kb from other candidates and flanked by a pair of putative LTRs, which ranged from 100 bp to 3000 bp, with similarity > 80%. We then annotated internal sequences of candidate LTR-RTs using *LTRdigest* and analyzed profile Hidden Markov Models of protein domains in the GyDB 2.0 database (Llorens *et al.*, 2011). If an LTR-RT candidate possessed a complete *Gag-Pol* protein sequence, it was retained as an intact LTR-RT (Fig. S1).

Unpaired LTRs, including solo-LTRs and truncated LTR-RTs, were identified based on sequence similarity to LTRs of intact LTR-RTs in each genome (Fig. S1). We saved LTR paralogous sequences from a Blastn analysis if they showed a 90% overlap in length and a 90% identity with any true LTR sequences, in addition to an e-value cutoff of 1e–10. We extracted 3 kb sequences both upstream and downstream of every detected LTR paralog to compare with the *Gag-Pol* protein sequences in the GyDB 2.0 database using tBlastn (Fig. S1). If at least 50% of any *Gag-Pol* sequence was covered by the flanking sequences with an identity > 30% and an e-value cutoff of 1e–8, the corresponding LTR was excluded from the solo-LTR list. The LTR paralogs that lacked any *Gag-Pol* homologs in both the upstream and downstream sequences were considered to be solo-LTRs, probably the remnants of unequal and illegitimate homologous recombination (Ma *et al.*, 2004; Hawkins *et al.*, 2009). In addition, those LTRs with *Gag-Pol* sequences on one side of flanking sequences were retained as truncated LTR-RTs. They may have been generated from partial deletion caused by nonhomologous end joining following double-stranded DNA breaks (Fawcett *et al.*, 2012; Chen *et al.*, 2013).

Dating of LTR-RT insertion events

The timing of LTR-RT insertion can be estimated based on the divergence between the 5′-LTR and 3′-LTR of the same transposon because LTR pairs are strictly identical immediately after insertion has occurred. Subsequent accumulation of mutations can be used as a measure of time since insertion (Sanmiguel *et al.*, 1998). Each LTR pair was aligned using MUSCLE (v3.8.31) with default options (Edgar, 2004). We employed the Kimura two-parameter method (Kimura, 1980) with the mutation rate of 1.3e–8 substitutions yr^{-1} per site to calculate insertion time (Ma & Bennetzen, 2004).

Estimation of the proportion of transposable elements

The proportion of TEs in each collected genome was estimated using the same methods and TE databases across all species to perform fair comparisons without subjective preferences. A comparative analysis of all whole genome sequences was performed based on the sequences of known TEs available in the MIPS Repeat Element Database (Mewes *et al.*, 2008), the TIGR Plant Repeat Databases (Ouyang & Buell, 2004) and Repbase (Jurka, 2000) using Blastn (Fig. S1). Genome segments were annotated as TEs when comparative output met the following criteria: matching length ≥ 80 bp; identity ≥ 80%; e-value ≤ 1e–10. In addition, we used a new database consisting of all *de novo* predicted LTR-RTs from 33 plant genomes to search the LTR sequences in each genome, employing the same rules as earlier. All putative TEs were identified to mask the genomes and the proportions of TEs in each

plant genome were obtained. We performed *t*-tests to examine the significance of TE content differences between our four true mangroves and 29 nonmangrove relatives.

We then conducted an ancestral state reconstruction to further examine TE loads along the phylogenetic tree. The phylogenetic tree reflecting the taxonomic relationships among our sample of species was obtained from the NCBI taxonomy tool, with branch lengths among species based on estimates from TimeTree (Hedges *et al.*, 2006). The R package phytools was used to reconstruct and visualize TE loads at ancestral nodes in the tree (Revell, 2012).

LTR clustering and *S* : *I* value estimation

The 5′-LTR sequences of all *de novo* predicted LTR-RTs in each species were compared against each other with Blastn. Two sequences were included in the same cluster if one LTR covered at least 70% of the length of the other, with an identity of at least 60%; this analysis was performed using the Silix software package (Miele *et al.*, 2011). Solo-LTRs (*S*) and truncated LTR-RTs (*T*) were subjected to the same cluster as the most similar intact LTR-RTs (*I*). Thus, LTR-RTs, solo-LTRs or truncated LTR-RTs with homologous LTR sequences would be assigned to the same 'cluster', which is similar to the concept of gene family and does not refer to physical proximity.

We used our collection of solo-LTRs and truncated LTRs to study removal rates of LTR-RTs over the past several million yr. The values of $S+T$ and $I+S+T$, as well as the numbers of intact LTR-RTs, were compared between mangroves and nonmangroves using a *t*-test. Genome assembly quality can influence the precision of $S:I$ and $T:I$ estimates because an intact LTR-RT could be separated into two different scaffolds and misidentified as two independent LTRs. We progressively filtered out solo-LTRs, truncated LTR-RTs, and intact LTR-RTs anchored in short scaffolds until the $S:I$ values stopped changing with subsequent scaffold removals to obtain accurate $S:I$ and $T:I$ values. When all intact LTR-RTs in one cluster were filtered, that cluster was no longer considered. The $S:I$ or $T:I$ ratios were calculated using the number of remaining intact LTR-RTs and solo-LTRs or truncated LTR-RTs, respectively. To evaluate the effect of LTR-RT deletions we calculated the proportions of clusters with $S:I$ values > 3 in each species, as well as proportions of LTR-RTs contained in these clusters. A *t*-test was used between the four true mangroves and 29 nonmangrove relatives to test the difference between the $S:I$ and $T:I$ ratios, including $(S+T):I$ ratios, and the proportions of LTR-RTs with high removal rates.

Genome size comparison between additional mangroves and nonmangroves

Ttransposable element lengths and gene numbers were regressed on assembled genome sizes for all 33 collected genomes to test whether TE content could explain genome size variation. In addition, we performed a regression analysis of the proportion of TEs and actual genome sizes estimated using flow cytometry.

We used flow cytometry to estimate DNA content (Doležel *et al.*, 2007). *Lycopersicon esculentum* (*Solanum lycopersicum*) cv 'Stupické polni tyckove rané' (1*C* = 958 Mb) (Doležel *et al.*, 1992) and *Oryza sativa* ssp. *japonica* cv 'Nipponbare' (1*C* = 442 Mb) (Arumuganathan & Earle, 1991) were employed as reference standards to measure genome sizes of two true mangroves and seven mangrove associates. *C*-values of all mangroves sampled were independently estimated three times, and the mean was used to determine the genome size of the sample.

Estimated genome sizes of five Rhizophoraceae species, as well as three *Sonneratia* species, *Avicennia marina*, and *Nypa fruticans* were obtained from previous studies (Röeser *et al.*, 1997; Z. He *et al.*, unpublished). The mangroves included in this study belong to eight orders according to Angiosperm Phylogeny Group III (APG III, 2009). We also collected genome size information for all species from these orders recorded in the *C*-value database (Bennett & Leitch, 2012). The family was considered the smallest statistical unit for data analysis in this study. *C*-value data for 58 families were collected with at least one species from each family, but only families containing more than three species were available. One-tailed Wilcoxon rank sum test with continuity correction was performed to compare genome sizes of all true mangroves and their nonmangrove relatives.

Results

In this study, we analyzed multiple genomes from the three main mangrove taxa (*Avicennia marina*, *Rhizophora apiculata*, *Sonneratia alba*, and *S. caseolaris*) based on our own work (Xu *et al.*, 2017b; Z. He *et al.*, unpublished), as well as the genomes of 29 nonmangrove species from the literature (for a complete list of citations, see the Materials and Methods section) (Table S1). We sought to understand how the TE load changes in different mangrove genomes; the potential mechanisms of TE load reduction in mangroves; and whether convergent evolution has happened with respect to the TE dynamics. To answer these questions, we compared TE prevalence and age structure, as well as the birth and death rates of TEs between nonmangrove and mangrove genomes.

TE prevalence in true mangroves and nonmangroves

To compare TE loads between true mangrove and nonmangrove genomes, we collected repetitive sequences and proportions of TEs from published genome studies (Table S2). Approximately 52 nonmangrove species (mainly angiosperms) with reported annotations of repetitive sequences are available in public databases (http://www.phytozome.net; http://genomevolution.org/wiki/index.php/Sequenced_plant_genomes; up to December 2016). Comparative analyses of these data show that genomes of true mangroves harbor relatively few TE sequences (Table S2). However, these available data on repeated elements are derived using different repetitive sequence databases and quite different strategies for TE prediction. We thus performed our own prediction of TE contents in genomes of true mangroves and nonmangroves.

Transposable element sequences were identified in the genomes of four true mangroves and 29 nonmangroves using repetitive sequence databases, combined with *de novo* predicted LTR retrotransposons (Fig. S1; Table S1). The estimated genome

proportions covered by TEs range from 1.19% in *Cucumis sativus* (1*C* = 367 Mb) to 44.5% in *Gossypium raimondii* (1*C* = 880 Mb) (Fig. 1; Table S3). Among the mangrove genomes, 3.14% of the *A. marina* sequence (1*C* = 507 Mb) consists of TEs. The fraction is 7.83% in *R. apiculata* (1*C* = 274 Mb), 8.16% in *S. alba* (1*C* = 284 Mb), and 3.49% in *S. caseolaris* (1*C* = 259 Mb). Although the genome size of *A. marina* is the largest among the four *de novo* sequenced mangroves, it contains the smallest proportion of transposons. The fraction of the *S. caseolaris* genome occupied by TEs is less than half that of the closely related *S. alba*. TE contents of these true mangrove genomes are among the lowest in their respective clades (*R. apiculata* in the Rosids I clade, *S. caseolaris* and *S. alba* in the Rosids II clade, and *A. marina* in the Asterid clade) (Fig. 1). In addition, a phylogenetic analysis shows that the TE loss occurred on the terminal branches leading to mangrove species (Fig. S2). Overall, TE contents are significantly lower in mangrove genomes than in the nonmangroves examined in the present study (*t*-test, *P*-value = 4.26e–06; Fig. 1). In contrast to most higher plant genomes, TE load reductions have occurred in all the true mangrove genomes we examined.

Comparisons of mangrove genome sizes

Mangroves can be divided into true mangroves and mangrove associates, with the former being restricted to intertidal habitats and the latter growing in both intertidal and terrestrial zones (Tomlinson, 1986). We collected and estimated genome sizes of 12 true mangroves and seven mangrove associates from the main mangrove lineages which represent the majority of mangrove taxa across eight orders (Arecales, Ericales, Fabales, Lamiales, Malpighiales, Malvales, Myrtales and Sapindales) (Table S4). The overall distribution of nonmangrove genome sizes we collected is shown in Fig. 2(a). The mean size is

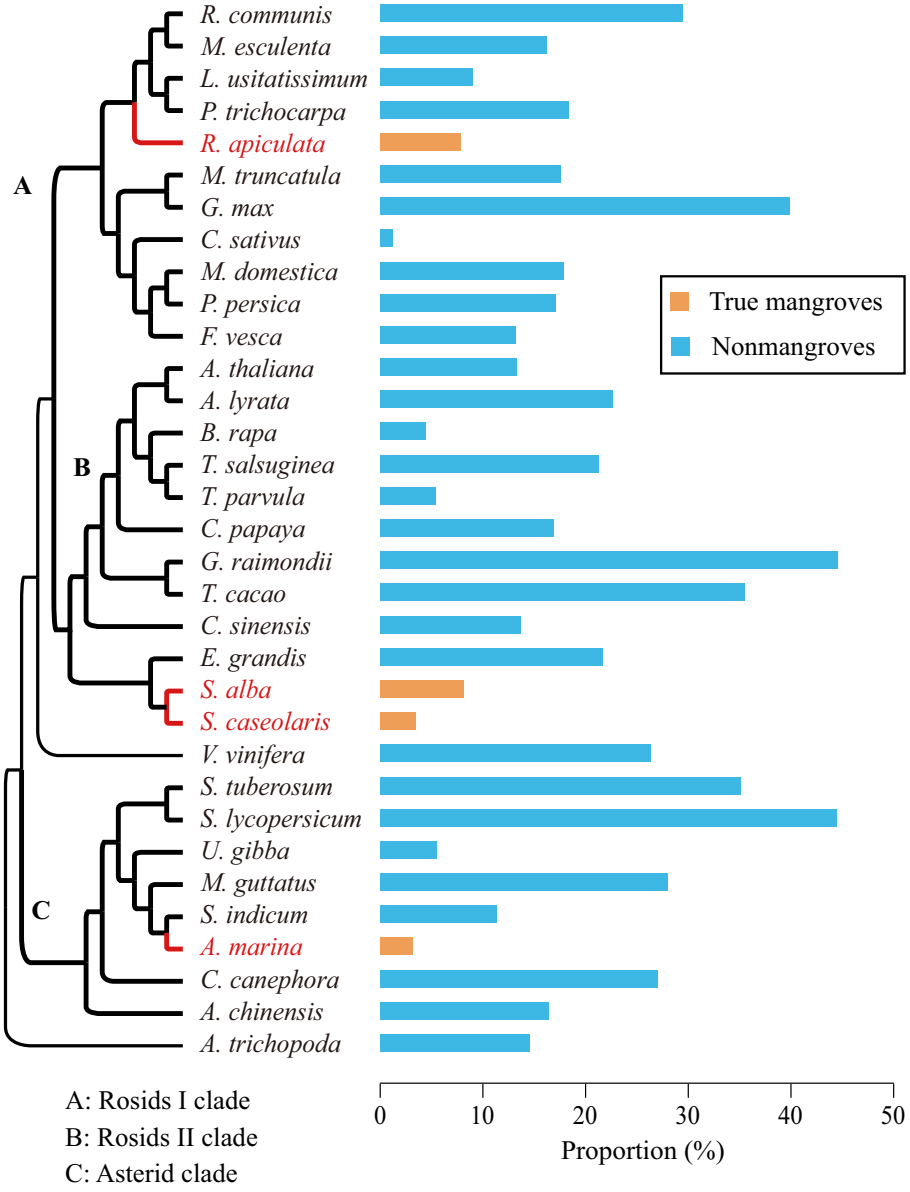

Fig. 1 Proportions of transposable elements (TEs) in the genomes of four true mangroves and 29 nonmangroves. Comparison of the proportions of TEs in 33 whole genomes based on their phylogenetic relationships. The four true mangroves, *Avicennia marina*, *Rhizophora apiculata*, *Sonneratia alba*, and *Sonneratia caseolaris*, belong to three main clades, Asterid (C), Rosids I (A), and Rosids II (B), respectively. The TE proportions shown in the right panel are compared among true mangroves and nonmangroves.

1830 Mb and the median is 1100 Mb. True mangrove genomes are significantly smaller than genomes of close nonmangrove relatives (Wilcoxon rank sum test, *P*-value = 7.596e–06; Fig. 2b). Although *Nypa fruticans* (1*C* = 1149 Mb) and *Aegiceras corniculatum* (1*C* = 841 Mb) appear to be outliers among true mangroves, they still represent smaller genomes in comparison with their close nonmangrove relatives in Primulaceae or Arecaceae (Fig. S3; Table S5). By contrast, there is no significant difference between genome sizes of the seven mangrove associates and those of closely related nonmangrove species (Wilcoxon rank sum test, *P*-value = 0.6767; Figs 2b, S4; Table S5). We conclude that there is a uniform reduction in genome size specifically among true mangroves.

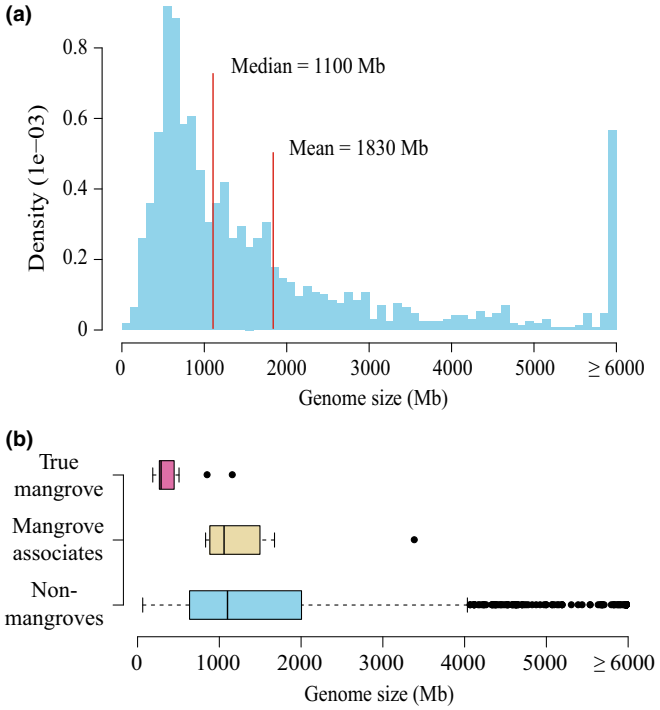

Fig. 2 Comparison of genome size among true mangroves, mangrove associates, and closely related nonmangroves species. (a) Distribution of genome sizes of the nonmangrove relatives used in this study. The last bar represents the species with a genome size ≥ 6000 Mb. (b) Genome size distribution among true mangroves, mangrove associates, and closely related nonmangroves species. The error bars present the extremes of whiskers, indicating 1.5 times the interquartile range. The points denote outlier data which lie beyond the extremes of the whiskers. The true mangrove outliers are *Nypa fruticans* (1*C* = 1149 Mb) and *Aegiceras corniculatum* (1*C* = 841 Mb).

Correlation between TE content and genome size

Regression of assembled genome size on a number of genomic features could help to identify attributes contributing to size variation (Hou & Lin, 2009; Tenaillon *et al.*, 2010). We estimated correlations between genome size and TE content and compared them with correlations between genome size and gene number. TE length exhibits an extremely high correlation coefficient with assembled genome size ($R^2 = 0.712$; *F*-test, *P*-value = 2.29e–10; Fig. 3a), while gene number shows a lower correlation ($R^2 = 0.345$; *F* test, *P*-value = 1.54e–4; Fig. 3b). In addition, the positive correlation between the proportion of TEs and actual genome size measured using flow cytometry indicates that large genomes bear higher proportions of TE sequences ($R^2 = 0.379$; *F* test, *P*-value = 8.16e–5; Fig. 3c). This suggests that TE load reduction is the most important driver of genome diminution in true mangroves.

Age structure of intact LTR retrotransposons

The LTR-RTs are class I TEs with quite strictly defined characteristics. They are the most prevalent TEs in most plant genomes (Michael, 2014). We examined insertion time distributions of intact LTR-RTs to probe the effect of LTR-RTs on genome size change over the past several million yr. Three of the four mangrove species in our dataset exhibit a deficit in recent insertion

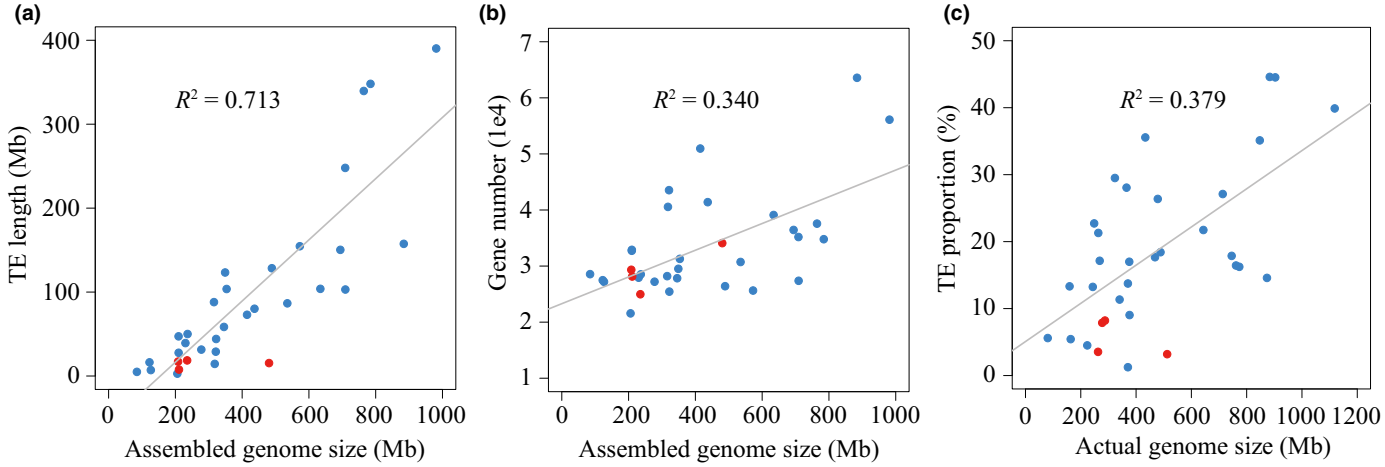

Fig. 3 Regression analyses of genome size on genome features. (a) Regression analysis of transposable element (TE) lengths in the genomes of four true mangroves and 29 nonmangroves against assembled genome sizes. The red points denote true mangroves, and the blue ones refer to the nonmangroves. (b) Regression of assembled genome size on gene number. (c) Regression of the proportion of TEs in the assembled genomes on their actual genome sizes obtained using flow cytometry.

events. The insertion time distribution of intact LTR-RTs peaks *c.* 6 million yr ago in the genome of *A. marina* (Fig. 4a) and *c.* 4 million yr ago in *R. apiculata* and *S. alba* (Fig. 4b,c). The *S. caseolaris* genome does not present such a decrease, but the steady rate of insertion has been extremely low, even for mangroves (Fig. 4c). By contrast, nonmangrove relatives undergo much higher rates of LTR-RT insertion (Figs 4a–c, S5). Furthermore, the pace has steadily increased in the past 9 million yr in most species. Where it has decreased, the rate reduction has been more recent and TE levels have not come close to even the peak estimated historical values for mangroves (Fig. S5).

Removal of LTR-RTs

Transposable element copy number change is a function of the difference between their birth and death rates. Thus, TE load reduction can occur either through the removal of extant TEs after they transpose (i.e. increasing death) or through the suppression of their transposition (i.e. decreasing birth), or a combination of the two (Tenaillon *et al.*, 2010; Fedoroff, 2012). To explore how TE load reduction occurred in the genomes of true mangroves, we analyzed the proliferation and removal of LTR-RTs. First, we compared the numbers of intact LTR-RTs that remain after proliferation in the preceding 9 million yr. The results show that the intact LTR-RTs (marked *I* in Fig. 5) are maintained at a lower level in the genomes of true mangroves than in those of nonmangrove relatives (*t*-test, *P*-value = 2.23e-05; Fig. 5a; Table S6). Partial deletion (generating truncated LTR-RTs) and unequal or illegitimate homologous recombination (resulting in solo-LTRs) are the main known mechanisms of LTR-RT copy number reduction (Vitte & Bennetzen, 2006; Hawkins *et al.*, 2009). Control of birth rate can be achieved by limiting transposon mobility through epigenetic regulation (Law & Jacobsen, 2010; Blumenstiel, 2011). To estimate LTR-RT death rates, we compared the numbers of genome-wide solo-LTRs (*S*; remnants of transposons deleted via recombination) and truncated LTR-RTs (*T*; generated through partial deletion). In almost all the genomes in this study, solo-LTRs are far more prevalent than truncated LTR-RTs (Table S6). The $S+T$ values for mangroves are significantly smaller than those of all the 29 inland relatives (*t*-test, *P*-value = 5.26e–04; Fig. 5b; Table S6). In addition, the sums of intact LTR-RTs, solo-LTRs and truncated LTRs ($I+S+T$) in the genomes of true mangroves are significantly smaller than those in nonmangroves (*t*-test, *P*-value = 3.35e–04; Fig. 5c; Table S6).

Two confounding factors can complicate these results. First, most genomes in our dataset are assembled into scaffolds smaller than whole chromosomes. Second, there is a strong correlation between the numbers of LTR-RT removal signatures (truncated LTR-RTs and solo-LTRs) and the number of intact elements. We corrected for the potential influence of these factors by filtering out short scaffolds (Table S6) and then calculating ratios of truncated LTR-RTs or solo-LTRs to intact elements ($T:I$ and $S:I$ ratios). The genome of *R. apiculata* possesses the highest $S:I$ ratio (estimated 3.19) among mangroves, while the calibrated $S:I$ ratio is 2.14 in *A. marina*, 1.6932 in *S. alba*, and 1.50 in *S. caseolaris* (Table S6). Filtered $S:I$ ratios are likewise significantly lower in the genomes of mangroves than in those of nonmangroves (*t*-test, *P*-value = 1.17e–02; Fig. 5d; Table S6). The patterns of calibrated $T:I$ (truncated to intact) ratios in four mangrove genomes are consistent with corresponding $S:I$ ratios. The $T:I$ ratio in *R. apiculata* (2.74) is higher than in the other three mangroves (0.97 in *A. marina*, 1.24 in *S. alba*, and 0.78 in *S. caseolaris*). The calibrated $T:I$ ratios reflect a different mechanism of TE removal and show no difference between mangroves and nonmangroves (*t*-test, *P*-value = 0.627; Table S6). As a result, the $(S+T):I$ ratios of true mangroves are slightly smaller than those of nonmangroves (*t*-test, *P*-value = 9.47e–02; Table S6).

In addition to studying individual insertion/deletion events, we clustered LTR-RT sequences by similarity (irrespective of

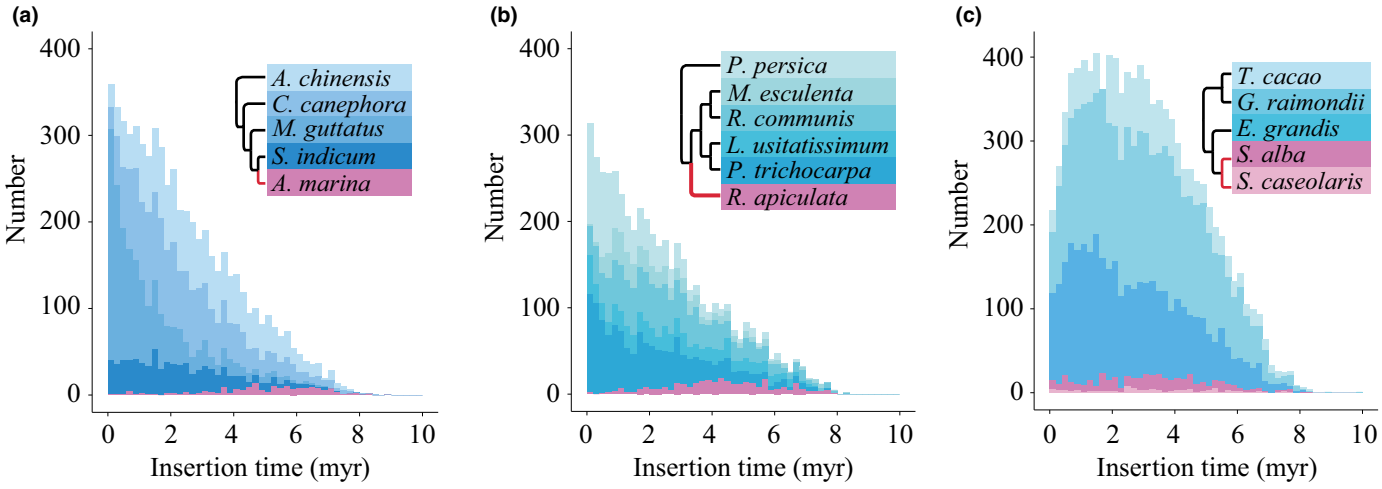

Fig. 4 Age structures of long terminal repeat-retrotransposons (LTR-RTs) in true mangroves and closely related nonmangroves. (a) Insertion time distribution of all intact LTR-RTs in *Avicennia marina* and four closely related nonmangroves, *Sesamum indicum*, *Mimulus guttatus*, *Coffea canephora*, and *Actinidia chinensis*. (b) Insertion time distribution of all intact LTR-RTs in *Rhizophora apiculata* and five related nonmangroves, *Populus trichocarpa*, *Linum usitatissimum*, *Ricinus communis*, *Manihot esculenta*, and *Prunus persica*. (c) Insertion time distribution of all intact LTR-RTs in *Sonneratia alba*, *Sonneratia caseolaris* and three related nonmangroves, *Eucalyptus grandis*, *Gossypium raimondii*, and *Theobroma cacao*. myr, millions of yr.

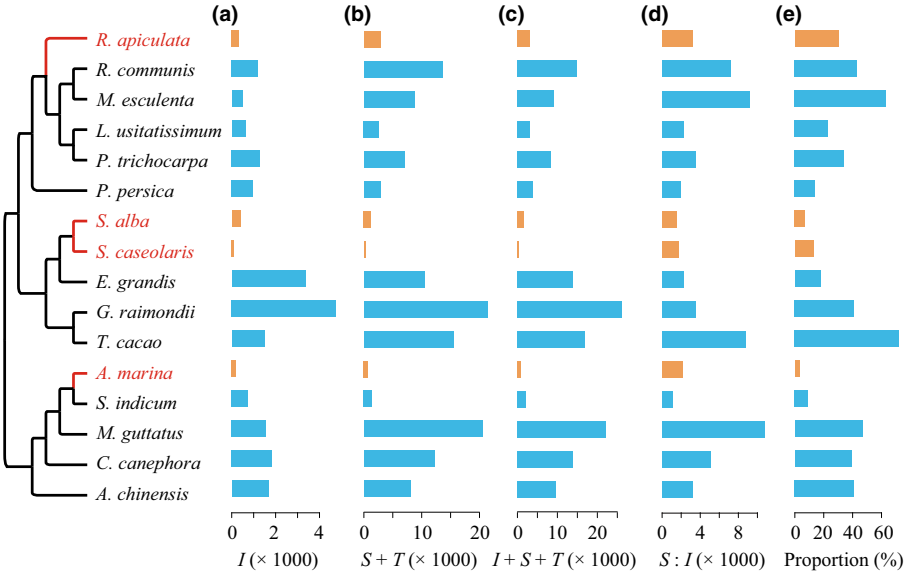

Fig. 5 Birth and death of long terminal repeat-retrotransposons (LTR-RTs) in true mangroves and closely related nonmangroves. (a) Total numbers of LTR-RTs in genomes of true mangroves and their nonmangrove relatives. The red bars denote the four true mangroves. (b) Comparison of $S + T$ values between four true mangroves and nonmangrove relatives. *S*, number of solo-LTRs; *T*, number of truncated LTR-RTs. (c) Total numbers of intact LTR-RTs and traces of LTR-RT deaths. *I*, number of intact LTR-RTs. (d) Ratios of solo-LTR to intact LTR-RT ($S : I$). (e) The proportions of LTR-RTs found in the clusters with high removal rates (filtered $S : I \geq 3$).

genome position; see the Materials and Methods section for details). Cluster-level $S : I$ values reflect removal rates for that family. We considered groups with filtered $S : I$ values higher than three to be particularly prone to removal (i.e. have a high cluster-level death rate). Looking at the proportion of LTR-RT families with high removal rates among all the clusters in a genome gives us a sense of the overall LTR-RT death rate. Although the proportions of highly removed LTR-RTs vary among mangroves (*R. apiculata* has the highest proportion of high-death-rate clusters), true mangroves harbor significantly smaller numbers of these families than the 29 nonmangroves in our dataset (*t*-test, *P*-value = 2.16e–02; Fig. 5e; Table S6).

Birth of LTR retrotransposons

The matrix population model is widely used in ecology to study population dynamics. It summarizes relationships among birth, death, immigration, emigration, and survival of individuals in populations (Caswell, 2001). We applied this model to describe LTR-RT evolution, taking into account the fact that immigration and emigration is negligible for transposons. Intact LTR-RTs (I) can be truncated by partial deletion or eliminated by unequal and illegitimate homologous recombination, leaving truncated LTR-RTs (T) and solo-LTRs (S), respectively (Ma *et al.*, 2004; Zhao & Ma, 2013; Baidouri *et al.*, 2014). Thus, the numeric relationship between the death and birth of retrotransposons can be described by the equation:

$$I = B - D + K \qquad \text{Eqn 1}$$

where B is the number of LTR-RTs generated in the past 9 million yr, D is the number of LTR-RT deaths leading to the appearance of solo-LTRs and truncated LTR-RTs ($S + T$), and K is the number of ancient LTR-RTs. Because few ancient LTR-RTs are maintained in the current genome based on the analyses of age structures of LTR-RTs in each studied genome, K can be considered a constant. According to Eqn 1, the number of transposition events generating new LTR-RTs (B) depends on the number of both intact LTR-RTs (I) and the traces of LTR-RT deaths (D, or $S + T$). Thus, the number of LTR-RT births is equal to the sum $I + S + T$ when K is close to zero. Looking at estimates of this value, we see that LTR retrotransposon birth rates are significantly lower in the host genomes of mangroves compared with nonmangrove relatives (Fig. 5c).

Loss of LTR-RTs clusters in mangrove genomes

It has been known that mobile elements within abundant families are expected to experience high amounts of purifying selection, as a result of an increased opportunity for ectopic recombination (Petrov *et al.*, 2003, 2011). Thus, the distribution of element frequencies in families could inform us about how selection is acting. To gain further insight into the nature of selection for TE repression, we examined relative numbers of LTR-RT clusters (see the Materials and Methods section for details). First, only two of the higher plants in this study, *C. sativus* (37 clusters) and *U. gibba* (41 clusters), have fewer such clusters than our mangrove sample. We found 51 clusters in *A. marina*, 84 in *R. apiculata*, 77 in *S. alba*, and 51 in *S. caseolaris* (Table S6). By contrast, the mean cluster number among the nonmangroves in the same clade as *A. marina* (*Actinidia chinensis*, *C. canephora*, *Mimulus guttatus*, and *Sesamum indicum*) is 359.75; that among nonmangrove relatives of *R. apiculata* (*Linum usitatissimum*, *Manihot esculenta*, *Prunus persica*, *Populus trichocarpa*, and *Ricinus communis*) is 256.2; and that among *S. alba* and *S. caseolaris* relatives (*Eucalyptus grandis*, and *G. raimondii*) is 803 (Table S6; Fig. S6). This indicates that the cluster numbers have been considerably reduced in mangroves, which accounts for the bulk of overall LTR-RT loss. Second, we compared the cumulative frequencies of LTR-RTs in clusters between mangrove and closely related nonmangrove species (Fig. S6). Frequencies of LTR-RTs belonging to the 10 most frequent clusters are higher in mangroves than in nonmangroves. Thus, the reduction in TE prevalence in mangroves is attributable to loss of whole clusters (or most members in these clusters) at the less frequent end of the spectrum (Fig. S6).

Discussion

Genome size reduction in true mangroves is a result of cuts in TE load

All four species examined in this study, *A. marina*, *R. apiculata*, *S. alba*, and *S. caseolaris*, experience major removals of TE sequences during their invasion of intertidal habitats. This indicates that significant TE removals have occurred independently at least three times in true mangroves. In addition, as TE length has a significantly positive correlation with total genome size based on the analysis of 33 plant genomes, the smaller genome sizes in other true mangroves could provide additional evidence for pervasive TE load reductions. It appears that independent TE load reductions may have happened in all six investigated lineages of true mangroves as they invaded intertidal habitats.

Many studies have reported that transposons could be under purifying selection, leading to a compact genome when the host invades stressful environments (Hu *et al.*, 2011; Ibarra-Laclette *et al.*, 2013; Kelley *et al.*, 2014; Zhang *et al.*, 2014). Even within the same species, maize landraces at high altitude tend to possess smaller genome sizes and lower TE contents than those at low altitude, probably driven by natural selection (Díez *et al.*, 2013; Bilinski *et al.*, 2017). Three mechanisms underlying deleterious effects have been discussed to explain purifying selection acting against TE insertions: direct disruptive effects of insertions (Hollister & Gaut, 2009; Naito *et al.*, 2009); deleterious TE product expression (McDonald *et al.*, 1997; Wang *et al.*, 2013); and chromosomal aberrations arising from ectopic recombination among TEs (Chan & Kolodner, 2011; Robberecht *et al.*, 2013). It is generally believed that TE-mediated genome instability would be accelerated in the face of environmental insults such as high salinity and strong UV radiation (Pfeiffer *et al.*, 2000; Schuermann *et al.*, 2005; Argueso *et al.*, 2008), leading to enhanced long-term selection pressures on the TE load.

According to our observations, elimination of most mobile elements leading to a reduced TE load and genome size may be a convergent strategy employed by mangrove hosts adapting to new stressful environments. This is in addition to convergent change in amino acid composition in mangrove genomes (Xu *et al.*, 2017a). Furthermore, this pattern reveals that although the activities of a few TEs could have been induced by abiotic and biotic stress (Grandbastien, 1998; Feschotte *et al.*, 2002), long-term selection pressures seem to have counteracted this effect. While the initial stress of adapting to the new environment might have precipitated a drive towards a smaller genome in mangroves via selective pressure, the pressure to maintain a small genome persists despite the adaptation of mangroves to their new habitat.

Reduction of transposition during mangroves' transition to new habitats

As TE load represents a balance between transposition and removal, two mechanisms can operate to reduce TE load: removal of TEs already present in the genome (increasing death rates) or suppression of proliferation before it happens (decreasing birth rates) (Hawkins *et al.*, 2009; Blumenstiel, 2011; Fedoroff, 2012). In most cases, hosts increase the death rates to counteract mobile element proliferation (Vitte & Panaud, 2003; Pereira, 2004; Baidouri & Panaud, 2013). By contrast, according to our analyses and Eqn 1, significantly fewer deaths and births of LTR-RTs are observed in mangrove genomes. Reduced S: I and D: I ratios in true mangrove genomes strongly suggest a reduction in death rates regardless of the abundance of intact LTR-RTs, suggestive of weakened mechanisms of TE removal by partial deletion or unequal and illegitimate homologous recombination. Based on these results, reductions of both death and birth rates have occurred in true mangrove genomes starting *c.* 9 million yr before present. On the other hand, as indicated by age structure, the continuously diminishing numbers of intact LTR-RTs indicate that lower total LTR-RT length is mainly attributable to a deficit in young LTR-RTs. Thus, the trend of TE load shrinkage has continued up to the very recent past. In conclusion, after a massive deletion of mobile elements during invasion of stressful environments, mangroves evolved a convergent strategy of reducing transposition rate to keep TE loads low.

Abundant clusters are more likely to have been subjected to selection (Petrov *et al.*, 2003, 2011). Thus, if the decrease in TE birth rate is driven by adaptation, we would see a reduction of cluster sizes in these abundant clusters (Barrón *et al.*, 2014). We observe the opposite: it is the small rare clusters that are missing in the mangrove genomes. This suggests that nonadaptive processes, such as genetic drift and mutation, might have played a role in the repressions of LTR-RT activities over the last 9 million yr. Furthermore, because several molecular processes are associated with the frequency distribution, the possibility that the repressions of LTR-RTs are adaptive remains to be further confirmed.

Genome size of mangrove associates

Although reductions in total TE length and genome size have occurred in the genomes of true mangroves, this trend is not observed in the mangrove associates we examined. This difference might be explained by the fact that the mangrove associates exhibit two types of plant growth in distinct habitats. Gene flow between terrestrial and intertidal populations might counteract purifying selection and prevent TE load reduction in the intertidal populations of mangrove associates (Jian *et al.*, 2004). In addition, mangrove associates are less adapted to highly saline environments compared with true mangroves (Yang *et al.*, 2011; Liang *et al.*, 2012), perhaps resulting in a different pattern of TE load in these taxa. Deeper sampling of mangrove associates for genomic studies should enable us to investigate these possibilities. Whole genome sequencing of additional true mangroves and population genomic analyses will provide further direct opportunities to study how host genomes act on TEs and elucidate molecular mechanisms underlying host–TE interaction.

Acknowledgements

We thank Gerald J. Wyckoff, Zixiao Guo, David E. Boufford and Renchao Zhou for their insightful comments on, and language editing of, the manuscript. Zhengzhen Wang and Lishi

Huang are appreciated for their help with collection of the materials and in the flow cytometry experiments. This work was supported by the National Natural Science Foundation of China (91731301 and 91331202 to S.S.; 31600182 to Z.H.); the National Key Research and Development Plan (2017FY100705 to S.S.); the Fundamental Research Funds for the Central Universities (17lgpy99 to Z.H.); and the Chang Hungta Science Foundation of Sun Yat-Sen University.

Author contributions

S.S., C-I.W. and H.L. planned and designed the research. H.L. performed experiments and conducted fieldwork. H.L. and Z.H. analyzed the data. H.L. wrote the manuscript.

References

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**: 105–121.

Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN *et al.* 2011. The genome of *Theobroma cacao*. *Nature Genetics* **43**: 101–108.

Argueso JL, Westmoreland J, Mieczkowski PA, Gawel M, Petes TD, Resnick M. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. *Proceedings of the National Academy of Sciences, USA* **105**: 11845–11850.

Arumuganathan K, Earle ED. 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* **9**: 208–218.

Baidouri ME, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O. 2014. Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Research* **24**: 831–838.

Baidouri ME, Panaud O. 2013. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biology and Evolution* **5**: 954–965.

Barrón MG, Fiston-Lavier AS, Petrov DA, González J. 2014. Population genomics of transposable elements in *Drosophila*. *Annual Review of Genetics* **48**: 561–581.

Bennett MD, Leitch IJ. 2012. *Angiosperm DNA C-values database (release 8.0, Dec. 2012).* [WWW document] URL http://data.kew.org/cvalues/ [accessed 7 March 2017].

Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annual Review of Plant Biology* **65**: 505–530.

Bilinski P, Albert PS, Berg JJ, Birchler J, Grote M, Lorant A, Quezada J, Swarts K, Yang J, Ross-Ibarra J. 2017. Parallel altitudinal clines reveal adaptive evolution of genome size in *Zea mays*. *bioRxiv*. doi: 10.1101/134528.

Blumenstiel JP. 2011. Evolutionary dynamics of transposable elements in a small RNA world. *Trends in Genetics* **27**: 23–31.

Brookfield JFY. 2005. The ecology of the genome – mobile DNA elements and their hosts. *Nature Reviews Genetics* **6**: 128–136.

Caswell H. 2001. *Matrix population models: construction, analysis and interpretation*, 2nd edn. Sunderland, MA, USA: Sinauer Associates.

Chamala S, Chanderbali AS, Der JP, Lan T, Walts B, Albert V, DePamphilis CW, Leebens-Mack J, Rounsley S, Schuster SC *et al.* 2013. Assembly and validation of the genome of the nonmodel basal angiosperm *Amborella*. *Science* **342**: 1516–1517.

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G *et al.* 2010. Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology* **28**: 951–956.

Chan JE, Kolodner RD. 2011. A genetic and structural study of genome rearrangements mediated by high copy repeat *Ty1* elements. *PLoS Genetics* **7**: e1002089.

Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Goicoechea JL, Liang C *et al.* 2013. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nature Communications* **4**: 1595.

Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu JK, Bohnert HJ *et al.* 2011. The genome of the extremophile crucifer *Thellungiella parvula*. *Nature Genetics* **43**: 913–918.

Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G *et al.* 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **345**: 1181–1184.

Díez CM, Gaut BS, Meca E, Scheinvar E, Montes-Hernandez S, Eguiarte LE, Tenaillon MI. 2013. Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist* **199**: 264–276.

Dobler S, Dalla S, Wagschal V, Agrawal AA. 2012. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na, K-ATPase. *Proceedings of the National Academy of Sciences, USA* **109**: 13040–13045.

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* **2**: 2233–2244.

Doležel J, Sgorbati S, Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum* **85**: 625–631.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics* **9**: 18.

Fawcett JA, Rouze P, Van de Peer Y. 2012. Higher intron loss rate in *Arabidopsis thaliana* than *A. lyrata* is consistent with stronger selection for a smaller genome. *Molecular Biology and Evolution* **29**: 849–859.

Fedoroff N. 2012. Transposable elements, epigenetics, and genome evolution. *Science* **338**: 758–767.

Feschotte C, Jiang N, Wessler SR. 2002. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* **3**: 329–341.

Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V *et al.* 2015. Convergent evolution of the genomes of marine mammals. *Nature Genetics* **47**: 272–275.

Giri C, Ochieng E, Tieszen LL, Zhu Z, Singh A, Loveland T, Masek J, Duke N. 2011. Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography* **20**: 154–159.

Grandbastien MA. 1998. Activation of plant retrotransposons under stress conditions. *Trends in Plant Science* **3**: 181–187.

Hawkins JS, Proulx SR, Rapp RA, Wendel JF. 2009. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proceedings of the National Academy of Sciences, USA* **106**: 17811–17816.

Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledgebase of divergence times among organisms. *Bioinformatics* **22**: 2971–2972.

Hellsten U, Wright KM, Jenkins J, Shu S, Yuan Y, Wessler SR, Schmutz J, Willis JH, Rokhsar DS. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences, USA* **110**: 19478–19482.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* **19**: 1419–1428.

Hou Y, Lin S. 2009. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. *PLoS ONE* **4**: e6978.

Houle D, Nuzhdin S. 2004. Mutation accumulation and the effect of copia insertions in *Drosophila melanogaster*. *Genetical Research* **83**: 7–18.

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett J, Grimwood J, Gundlach H *et al.* 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics* **43**: 476–481.

Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, Zhang L, Niu X, Zhang X, Meng M *et al.* 2013. Draft genome of the kiwifruit *Actinidia chinensis*. *Nature Communications* **4**: 2640.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P *et al.* 2009. The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics* **41**: 1275–1281.

Ibarra-Laclette E, Albert VA, Herrera-Estrella A, Herrera-Estrella L. 2011. Is GC bias in the nuclear genome of the carnivorous plant *Utricularia* driven by ROS-based mutation and biased gene conversion? *Plant Signaling & Behavior* **6**: 1631–1634.

Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang TH, Lan T, Welch AJ, Juárez MJA, Simpson J *et al.* 2013. Architecture and evolution of a minute plant genome. *Nature* **498**: 94–98.

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C *et al.* 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463–467.

Jian S, Tang T, Zhong Y, Shi S. 2004. Variation in inter-simple sequence repeat (ISSR) in mangrove and non-mangrove populations of *Heritiera littoralis* (Sterculiaceae) from China and Australia. *Aquatic Botany* **79**: 75–86.

Jurka J. 2000. Repbase update: a database and an electronic journal of repetitive elements. *Trends in Genetics* **16**: 418–420.

Kelley JL, Peyton JT, Fiston-Lavier A-S, Teets NM, Yee M-C, Johnston JS, Bustamante CD, Lee RE, Denlinger DL. 2014. Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. *Nature Communications* **5**: 4611.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111–120.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**: 204–220.

Levin HL, Moran J. 2011. Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics* **12**: 615–627.

Liang S, Fang L, Zhou R, Tang T, Deng S, Dong S, Huang Y, Zhong C, Shi S. 2012. Transcriptional homeostasis of a mangrove species, *Ceriops tagal*, in saline environments, as revealed by microarray analysis. *PLoS ONE* **7**: e36499.

Lisch D. 2013. How important are transposons for plant evolution? *Nature Reviews Genetics* **14**: 49–61.

Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP *et al.* 2011. The gypsy database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Research* **39**: 70–74.

Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* **65**: 1827–1840.

Ma J, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences, USA* **101**: 12404–12410.

Ma J, Devos KM, Bennetzen JL. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Research* **14**: 860–869.

McDonald JF, Matyunina LV, Wilson S, Jordan IK, Bowen NJ, Miller WJ. 1997. LTR retrotransposons and the evolution of eukaryotic enhancers. *Genetica* **100**: 3–13.

Mewes HW, Dietmann S, Frishman D, Gregory R, Mannhaupt G, Mayer KFX, Münsterkötter M, Ruepp A, Spannagl M, Stümpflen V *et al.* 2008. MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Research* **36**: 196–201.

Michael TP. 2014. Plant genome size variation: bloating and purging DNA. *Briefings in Functional Genomics* **13**: 308–317.

Miele V, Penel S, Duret L. 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12**: 116.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly B, Lewis KLT *et al.* 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **452**: 991–996.

Morris SC. 2003. *Life's solution: inevitable humans in a lonely universe.* Cambridge, UK: Cambridge University Press.

Myburg A, Grattapaglia D, Tuskan G, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* **509**: 356–362.

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR. 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **461**: 1130–1134.

Orel LE, Crick FHC. 1980. Selfish DNA: the ultimate parasite. *Nature* **284**: 604–607.

Ouyang S, Buell CR. 2004. The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Research* **32**: D360–D363.

Papaceit M, Ávila V, Aguadé M, García-Dorado A. 2007. The dynamics of the *roo* transposable element in mutation-accumulation lines and segregating populations of *Drosophila melanogaster*. *Genetics* **177**: 511–522.

Parida AK, Jha B. 2010. Salt tolerance mechanisms in mangroves: a review. *Trees* **24**: 199–217.

Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**: 228–231.

Pasyukova EG, Nuzhdin SV, Morozova TV, Mackay TFC. 2004. Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. *Journal of Heredity* **95**: 284–290.

Pereira V. 2004. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biology* **5**: R79.

Petrov DA. 2001. Evolution of genome size: new approaches to an old problem. *Trends in Genetics* **17**: 23–28.

Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Molecular Biology and Evolution* **20**: 880–892.

Petrov DA, Fiston-Lavier AS, Lipatov M, Lenkov K, Gonzalez J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Molecular Biology and Evolution* **28**: 1633–1644.

Pfeiffer P, Goedecke W, Obe G. 2000. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis* **15**: 289–302.

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T *et al.* 2012. The cassava genome: current progress, future directions. *Tropical Plant Biology* **5**: 88–94.

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Robberecht C, Voet T, Esteki MZ, Nowakowska BA, Vermeesch JR. 2013. Nonallelic homologous recombination between retrotransposable elements is a driver of *de novo* unbalanced translocations. *Genome Research* **23**: 411–418.

Röeser M, Johnson MAT, Hanson L. 1997. Nuclear DNA amounts in palms (Arecaceae). *Botanica Acta* **110**: 79–89.

Rothschild LJ, Mancinelli RL. 2001. Life in extreme environments. *Nature* **409**: 1092–1101.

Sanmiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. 1998. The paleontology of intergene retrotransposons of maize. *Nature Genetics* **20**: 43–45.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C *et al.* 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics* **46**: 707–713.

Schuermann D, Molinier J, Fritsch O, Hohn B. 2005. The dual nature of homologous recombination in plants. *Trends in Genetics* **21**: 172–181.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP *et al.* 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* **43**: 109–116.

Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009. Fine-grained annotation and classification of *de novo* predicted LTR retrotransposons. *Nucleic Acid Research* **37**: 7002–7013.

Teeling EC. 2009. Hear, hear: the convergent evolution of echolocation in bats? *Trends in Ecology and Evolution* **24**: 351–354.

Tenaillon MI, Hollister JD, Gaut BS. 2010. A triptych of the evolution of plant transposable elements. *Trends in Plant Science* **15**: 471–478.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.

The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**: 635–641.

Thomas GWC, Hahn MW. 2015. Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. *Molecular Biology and Evolution* **32**: 1232–1236.

Tomlinson PB. 1986. *The botany of mangroves*. Cambridge, UK: Cambridge University Press.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D *et al.* 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* **42**: 833–839.

Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F *et al.* 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics* **45**: 487–494.

Vinogradov AE. 2003. Selfish DNA is maladaptive: evidence from the plant Red List. *Trends in Genetics* **19**: 609–614.

Vinogradov AE. 2004a. Genome size and extinction risk in vertebrates. *Proceedings of the Royal Society of London Series B: Biological Sciences* **271**: 1701–1705.

Vinogradov AE. 2004b. Evolution of genome size: multilevel selection, mutation bias or dynamical chaos? *Current Opinion in Genetics & Development* **14**: 620–626.

Vitte C, Bennetzen JL. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proceedings of the National Academy of Sciences, USA* **103**: 17638–17643.

Vitte C, Panaud O. 2003. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Molecular Biology and Evolution* **20**: 528–540.

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S *et al.* 2012b. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature genetics* **44**: 1098–1103.

Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C, Zhang Y, Zhang X, Wang Y, Hua W *et al.* 2014. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biology* **15**: R39.

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F *et al.* 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43**: 1035–1039.

Wang X, Weigel D, Smith LM. 2013. Transposon variants and their effects on gene expression in *Arabidopsis*. *PLoS Genetics* **9**: e1003255.

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R *et al.* 2012a. The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. *Plant Journal* **72**: 461–473.

Wu HJ, Zhang Z, Wang JY, Oh DH, Dassanayake M, Liu B, Huang Q, Sun HX, Xia R, Wu Y *et al.* 2012. Insights into salt tolerance from the genome of *Thellungiella salsuginea*. *Proceedings of the National Academy of Sciences, USA* **109**: 12219–12224.

Xu Q, Chen LL, Ruan X, Chen D, Zhu A, Chen C, Bertrand D, Jiao WB, Hao BH, Lyon MP *et al.* 2013. The draft genome of sweet orange (*Citrus sinensis*). *Nature Genetics* **45**: 59–66.

Xu S, He Z, Guo Z, Zhang Z, Wyckoff GJ, Greenberg A, Wu CI, Shi S. 2017a. Genome-wide convergence during evolution of mangroves from woody plants. *Molecular Biology and Evolution* **34**: 1008–1015.

Xu S, He Z, Zhang Z, Guo Z, Guo W, Lyu H, Li J, Yang M, Du Z, Huang Y *et al.* 2017b. The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole genome sequencing. *National Science Review*. doi: 10.1093/nsr/nwx065.

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J *et al.* 2011. Genome sequence and analysis of the tuber crop potato. *Nature* **475**: 189–195.

Yang G, Zhou R, Tang T, Chen X, Ouyang J, He L. 2011. Gene expression profiles in response to salt stress in *Hibiscus tiliaceus*. *Plant Molecular Biology Reporter* **29**: 609–617.

Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H *et al.* 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW *et al.* 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**: 1311–1320.

Zhao M, Ma J. 2013. Co-evolution of plant LTR-retrotransposons and their host genomes. *Protein and Cell* **4**: 493–501.

Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel molecular evolution in an herbivore community. *Science* **337**: 1634–1637.

Zou Z, Zhang J. 2015. No genome-wide protein sequence convergence for echolocation. *Molecular Biology and Evolution* **32**: 1237–1241.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Work flow for prediction of intact LTR-RTs, TE load, and truncated LTR-RTs and Solo-LTR.

Fig. S2 Phylogenetic comparison of TE loads with ancestral state reconstruction.

Fig. S3 Genome size comparisons between true mangroves and nonmangrove relatives.

Fig. S4 Comparison of genome size between mangrove associates and nonmangrove relatives.

Fig. S5 Insertion time distributions of LTR-RT genomes of 17 nonmangroves.

Fig. S6 Cumulative distributions of LTR-RT frequencies in mangroves and nonmangroves.

Table S1 Version and source of all whole genome sequences in this study

Table S2 Repetitive sequence and TE proportions from collected genomic information

Table S3 Genome size and TE contents within each genome

Table S4 Genome size data of mangroves used in this study

Table S5 Genome size of nonmangroves in each family

Table S6 The number of original and filtered intact LTR-RTs, solo-LTRs and truncated LTRs

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.

See also the Commentary on this article by Gaut, **217:** 5–7.