The LINNEAN SOCIETY of London BOTANICAL Journal of the Linnean Society

Botanical Journal of the Linnean Society, 2015, **177**, 15–26. With 3 figures

Interstitial telomere-like repeats in the monocot family Araceae

ARETUZA SOUSA* and SUSANNE S. RENNER

Department of Biology, University of Munich (LMU), 80638 Munich, Germany

Received 6 May 2014; revised 26 September 2014; accepted for publication 4 October 2014

Combining molecular cytogenetics and phylogenetic modelling of chromosome number change can shed light on the types of evolutionary changes that may explain the haploid numbers observed today. Applied to the monocot family Araceae, with chromosome numbers of $2n = 8$ to $2n = 160$, this type of approach has suggested that descending dysploidy has played a larger role than polyploidy in the evolution of the current chromosome numbers. To test this, we carried out molecular cytogenetic analyses in 14 species from 11 genera, using probes for telomere repeats, 5S rDNA and 45S rDNA and a plastid phylogenetic tree covering the 118 genera of the family, many with multiple species. We obtained new chromosome counts for six species, modelled chromosome number evolution using all available counts for the family and carried out fluorescence *in situ* hybridization with three probes (5S rDNA, 45S rDNA and *Arabidopsis*-like telomeres) on 14 species with $2n = 14$ to $2n = 60$. The ancestral state reconstruction provides support for a large role of descending dysploidy in Araceae, and interstitial telomere repeats (ITRs) were detected in *Anthurium leuconerum*, *A. wendlingeri* and *Spathyphyllum tenerum*, all with $2n = 30$. The number of ITR signals in *Anthurium* (up to 12) is the highest so far reported in angiosperms, and the large repeats located in the pericentromeric regions of *A. wendlingeri* are of a type previously reported only from the gymnosperms *Cycas* and *Pinus*. © 2014 The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2015, **177**, 15–26.

ADDITIONAL KEYWORDS: 5S and 45S rDNA – ancestral trait reconstruction – Araceae, gymnosperms – Bayesian and maximum-likelihood inference – dysploidy – FISH – interstitial telomeric signals.

INTRODUCTION

A phylogenetic analysis establishes the direction of evolution and allows reconstruction of the likely time-frame and sequence of events that may have led to the character states seen in the included species. With the availability of DNA-based phylogenetic trees, cytogeneticists have increasingly turned to 'trait reconstruction' to infer the direction of change in chromosome numbers. Among the insights coming from these efforts is that over the course of evolution descending chromosome numbers (dysploidy) may be a more common phenomenon than traditionally thought, for example in Brassicaceae (Yogeeswaran *et al*., 2005; Lysak *et al*., 2006; Mandakova & Lysak, 2008; Cheng *et al*., 2013), Rosaceae (Vilanova *et al*., 2008; Illa *et al*., 2011; Jung *et al*., 2012), Poaceae (Luo *et al*., 2009) and Melanthiaceae (Pellicer *et al*., 2014). Probably the best studied case of chromosome rearrangements leading to descending dysploidy is *Arabidopsis* (DC.) Heynh., in which $n = 8$ is ancestral to $n = 5$ through inversions, fusions and translocations (Lysak *et al*., 2006). Descending dysploidy in *Arabidopsis* could be inferred only by combining phylogenetic trees for the relevant species with fluorescence *in situ* hybridization (FISH). Work on chromosome evolution in the large monocot family Araceae, with 3790 species in 118 genera (Boyce & Croat, 2011), revealed that in this family dysploidy may also have played a greater role than polyploidy (Cusimano, Sousa & Renner, 2012: table S1 lists all counts for Araceae; Sousa, Cusimano & Renner, 2014). This inference, however, was based on a relatively sparse sample of species representing the family (Cusimano *et al*., 2012) and a follow-up study on one derived tribe, Areae (Sousa *et al*., 2014). The hypothesis of frequent chromosome losses due to descending dysploidy in Araceae is therefore in need of further cytogenetic testing.

A cytogenetic test for a possible reduction in chromosome number by chromosome fusion is the presence of interstitial telomere repeats (ITRs), which can be visualized using standard probes for plant

*Corresponding author. E-mail: aretuzasousa@gmail.com

telomere repeats (Ijdo *et al*., 1991; Fuchs, Brandes & Schubert, 1995; Weiss-Schneeweiss *et al*., 2004). Thus, ITRs have been used as indicators of chromosome fusion in *Vicia faba* L., *Sideritis montana* L. and *Typhonium laoticum* Gagnep. (Schubert *et al*., 1992; Raskina *et al*., 2008; Sousa *et al*., 2014), and telomere signals near a centromere may indicate the fusion of two telocentric chromosomes (Schubert *et al*., 1992). The absence of ITRs, however, does not exclude the possibility of chromosome number reduction via chromosome rearrangements (see Lysak *et al*., 2006). Thus far, *Pinus* L. is the genus with the most conspicuous interstitial telomere FISH signals, with often up to four signals near the centromere and in interstitial positions (Fuchs *et al*., 1995; Lubaretz *et al*., 1996; Schmidt *et al*., 2000; Hizume *et al*., 2002; Islam-Faridi, Nelson & Kubisiak, 2007). Based on the inferred large role of descending dysploidy in Araceae (Cusimano *et al*., 2012; Sousa *et al*., 2014), we decided to carry out cytogenetic analyses of the distribution of telomere repeats in 14 species from 11 genera, selected to represent lineages of Araceae not previously well studied. The enlarged Araceae phylogenetic analysis and new cytogenetic data on which we report here reveal an unexpected frequency of conspicuous ITRs in this family.

MATERIAL AND METHODS

Plant material and DNA sequencing

We augmented the DNA data matrix of Nauheimer, Metzler & Renner (2012) by adding sequences for 29 further species from GenBank and by sequencing 14 additional species (on which cytogenetic studies were performed) for the same gene loci used by Nauheimer *et al*. (the plastid *trnL* intron and *trnL-F* spacer, the *matK* gene and partial *trnK* intron and the *rbcL* gene). We used standard primers (Cabrera *et al*., 2008), except for *matK* for which we used the primers listed in Cusimano *et al*. (2010). Total DNA from silica-dried leaves was extracted with the NucleoSpin plant II kit according to the manufacturer's protocol (Macherey-Nagel). Polymerase chain reactions (PCRs) were performed using 1.25 units of *Taq* DNA polymerase (New England Biolabs). Each PCR was composed of 17.55 μL H_2O (Sigma), 2.5 μL 10× PCR buffer, 1 μL 10 mM dNTPs, 0.75 μL 25 mM M_gCl_2, 0.2 μL *Taq* polymerase, 1 μL forward primer and 1 μL reverse primer (Taberlet *et al*., 1991). The PCR products were purified with Exo I and FastAP (Fermentas). Sequencing was done on an ABI 3130 four-capillary sequencer and sequences were assembled and edited with Sequencher 4.2 (Gene Codes Corp.). The newly studied and sequenced species, with their taxonomic authorities, herbarium vouchers and GenBank accession numbers, are listed in Supporting Information Table S1. For voucher information on the previously sequenced Araceae, see Nauheimer *et al*. (2012; table S1). The final alignment included 160 of the 3790 species from each of the 118 genera of Araceae and 11 outgroups representing the remaining families of Alismatales.

Phylogenetic analyses

Alignments were generated in MAFFT (Katoh & Standley, 2013; http://mafft.cbrc.jp/alignment/server/) and checked visually using MEGA5 (Tamura *et al*., 2011). We removed 249 poorly aligned positions and the combined plastid matrix was then used for maximum-likelihood (ML) tree searches in RAxML (Stamatakis, 2006; Stamatakis, Hoover & Rougemont, 2008), under the GTR + G substitution model with four rate categories because this model fits the data best, as assessed with Modeltest (Posada & Crandall, 1998).

Bootstrapping under ML used 1000 replicates. We also generated ultrametric trees in BEAST v. 1.7.5 via the portal CIPRES science gateway (Drummond & Rambaut, 2007; Miller, Pfeiffer & Schwartz, 2010), using the same substitution model for the entire concatenated alignment and a pure-birth Yule model as the tree prior. The analysis was run for 100 million generations, sampling every 1000th step. The burn-in fraction, i.e. the number of trees to be discarded before reconstructing a consensus tree (the maximum clade credibility tree) from the remaining trees, was assessed using Tracer v. 1.4.1, which is part of the BEAST package.

Inference of chromosome number change

For ML and Bayesian phylogenetic inferences of ancestral haploid chromosome numbers we used ChromEvol v. 1.4 with eight models (Mayrose, Barker & Otto, 2010; http://www.tau.ac.il/~itaymay/cp/chromEvol/index.html), the fit of which was assessed via likelihood ratio testing, using the Akaike information criterion (AIC). ChromEvol models have the following parameters: polyploidization (chromosome number duplication) with constant rate ρ, demi-duplication (fusion of gametes of different ploidy) with constant rate μ and dysploidization with either constant or linearly changing rates (ascending: chromosome gain rates λ or λ_1; descending: chromosome loss rates δ or δ_1). We fitted all models to a phylogram (in which branch lengths are proportional to numbers of substitutions) and an ultrametric depiction of the phylogenetic tree (in which branch lengths are proportional to time). The phylogram was the RAxML tree, and the ultrametric tree was the BEAST

maximum clade credibility tree. For each model, we ran 10 000 simulated repetitions to compute the expected number of changes along each branch of the phylogenetic tree and the ancestral haploid chromosome numbers at nodes. The maximum possible ancestral number of chromosomes was set to ten times the highest number found in the empirical data, and the minimum number was set to 1. Haploid chromosome numbers of species were obtained from Cusimano *et al*. (2012; table S1) and from the *Index to Plant Chromosome Numbers* (http://www.tropicos.org/Project/IPCN); species without known numbers were coded as 'unknown' (X) and changes among character states (chromosome numbers) were assigned equal likelihood. Results were plotted in R using the ChromEvol functions version 1 of N. Cusimano (http://www.sysbot.biologie.uni-muenchen.de/en/people/cusimano/).

Chromosome preparation, FISH analyses and DNA probes

Root tips were collected from potted plants cultivated in the greenhouses of the Munich Botanical Garden, one individual per species. Species authority names and voucher material for each species are given in Table S1. Root tips were pretreated in 2 mM 8-hydroxyquinoline for 20 h at 4 °C, fixed in freshly prepared 3:1 (v/v) ethanol/glacial acetic acid at room temperature overnight and kept at –20 °C. For chromosome preparations, fixed root tips were washed three times for 5 min in distilled water, digested with 1% cellulase (w/v; Onozuka RS, Serva), 0.4% pectolyase (w/v; Sigma) and 0.4% cytohelicase (w/v; Sigma) in citric buffer, pH 4.8, for 30 min at 37 °C in a humid chamber, dissected in a drop of 45% acetic acid and squashed. Coverslips were removed after freezing in dry ice and preparations were air-dried at room temperature. The quality of spreads was checked using phase-contrast microscopy and only preparations with at least ten well-spread metaphases were used for FISH.

We performed FISH with probes for telomere repeats, 5S rDNA and 45S rDNA. For some species, we had little material and could only use one or two of the three probes. To locate the rDNAs, we used the 18S–5.8S–25S rDNA repeat unit of *Arabidopsis thaliana* (L.) Heynh. in the pBSK+ plasmid, labelled with digoxigenin-11-dUTP (Roche) by nick translation and a 349-bp fragment of the 5S rRNA gene repeat unit from *Beta vulgaris* L. cloned into pBSK+ (Schmidt, Schwarzacher & Heslop-Harrison, 1994) and labelled with biotin-16-dUTP (Roche) by PCR. Telomere repeats were visualized with the *Arabidopsis*-like telomere probe of Ijdo *et al*. (1991) using the oligomer primers $(5'\text{-TTTAGGG-}3')_5$ and $(5'\text{-CCCTAAA-}3')_5$, labelled with digoxigenin-11-dUTP by nick translation. Hybridization mixes consisted of 50% formamide (w/v), 2× SSC, 10% (w/v) dextran sulfate and 100–200 ng of labelled probe. The hybridization mix was denatured at 75 °C for 10 min and immediately cooled on ice for 10 min; 10–15 μL of the mix was then added to each slide and the slides plus the hybridization mix were denatured at 75 ºC for 5 min. Hybridization was carried out in a humid chamber at 37 °C for 20 h. Post-hybridization washes were performed with a stringency of 80%. The 5S rDNA was detected with streptavidin–Cy3 conjugate (Sigma) and the 45S rDNA and the *Arabidopsis*-like telomere with anti-DIG–FITC conjugate (Roche) at 37 °C for 1 h. The chromosomes were counterstained with DAPI (2 μg mL^{-1}) and mounted in Vectashield (Vector).

Slides were first analysed with the probes for telomeres and 5S rDNA. They were then de-stained and a second hybridization was performed on the same slide with 45S rDNA. For some species with multiple 45S rDNA sites or with ITRs, further single-probe experiments were carried out to confirm the number of signals. More than 20 metaphases per species were analysed after FISH. Images were taken with a Leica DMR microscope equipped with a KAPPA-CCD camera and the KAPPA software. They were optimized for best contrast and brightness using Adobe Photoshop CS3 version 10.0.

RESULTS

Chromosomal evolution in Araceae

The plastid DNA matrix of 4928 aligned nucleotides for 171 species yielded a well-supported ML phylogenetic tree that we used to infer the evolution of chromosome numbers in Araceae (Figs 1, S1, S2, S4). The changes inferred on the ultrametric tree for Araceae are shown in Figures 1 and S4, and a reconstruction on the phylogram in Figures S1 and S2. The statistical support for both trees is shown in Figures S3 and S5 and the inferred rates of change and numbers of events are summarized in Table 1. On the ultrametric tree, the four-parameter-constant-rate model, which assumes constant chromosome gain and loss rates and a polyploidization rate that differs from the demi-polyploidization rate, best explained the data (AIC = 732.6 compared with 736.6 for the next best model), whereas on the phylogram, the best model was the four-parameter-linear-rate model, which includes rates of gain and loss that depend linearly on the current chromosome number (AIC = 844.4 compared with 982.8 for the next best). In both trees, chromosomes loss was the most common event. On the ultrametric tree, the next most common events were duplication of the entire chromosome complement and demi-duplications (Figs 1 and S4 insets in the lower left, Table 1); on the

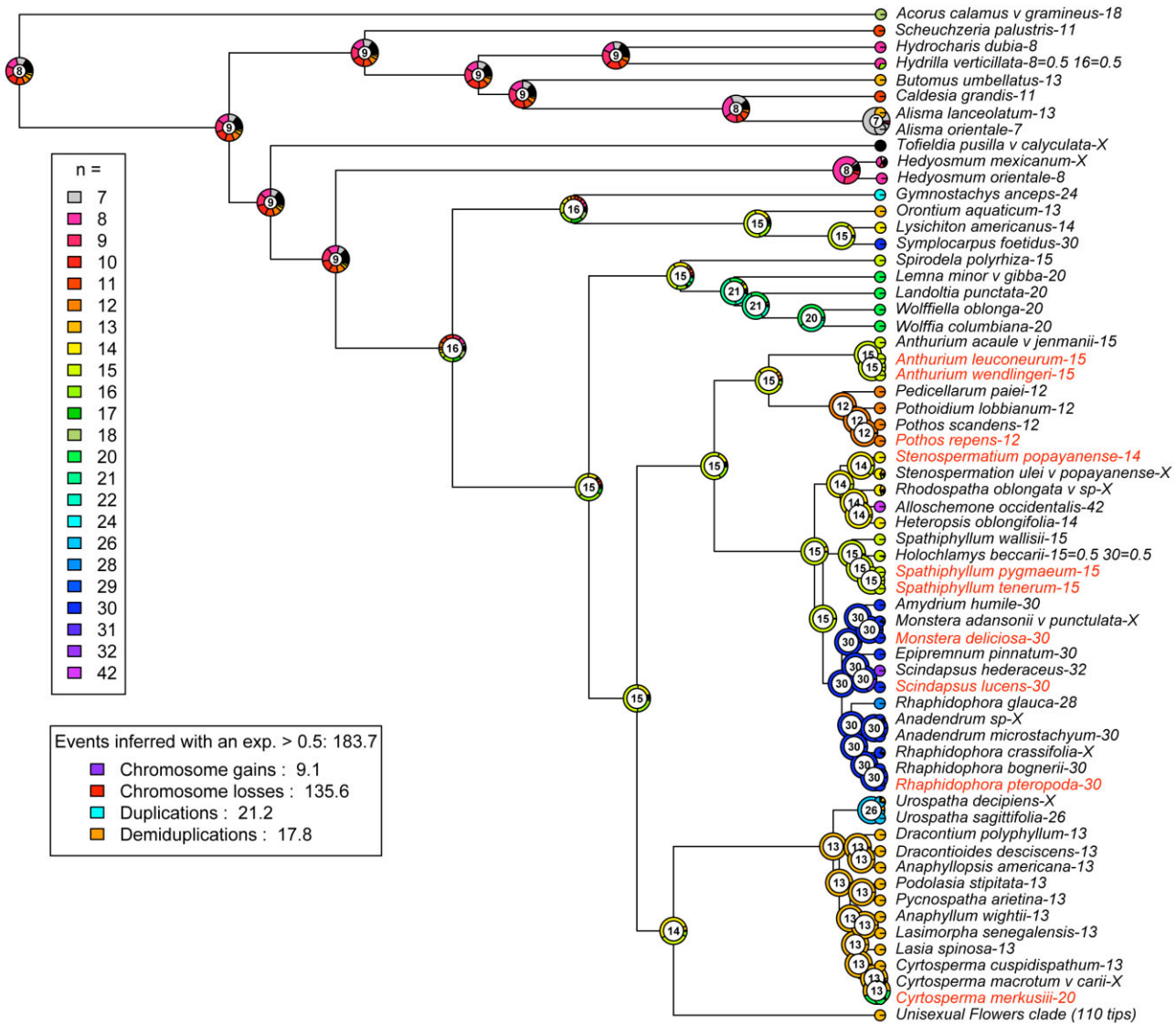

Figure 1. Chromosome number reconstruction for Araceae on an ultrametric tree, rooted on *Acorus calamus*. Pie charts represent the probabilities of inferred chromosome numbers, with the number inside a pie having the highest probability. The rectangular inset shows the frequency with which each event type (gains, losses, duplications, demiduplication) occurs along the branches of the tree. The colour coding of chromosome numbers is explained in the inset on the left. Species investigated by FISH are labelled in red.

phylogram, the next most common events were single chromosome gains, duplication of the entire set and demi-duplications (Figs S1 and S2 insets in the lower left, Table 1). The inferred ancestral haploid numbers in the Araceae decrease from $a = 16$ to 15 and 14 on the ultrametric tree and from $a = 16$ to 14 and 13 (and then back to $a = 14$) on the phylogram.

Results of the FISH experiments

FISH was performed for 14 species from 11 genera representing early and derived lineages of Araceae (Table 2). New chromosome numbers were obtained for *Cyrtosperma merkusii*, *Pothos repens*, *Spathiphyllum pygmaeum*, *S. tenerum*, *Stylochaeton puberulus* and *Ulearum sagittatum*. Their somatic numbers varied from $2n = 14$ to $2n = 60$ (Table 2). The chromosome variation found within each genus is presented in Table S2. *Ulearum sagittatum*, with $2n = 14$, has especially large chromosomes (Fig. 2A, D). The remaining species with relatively high chromosome numbers ($2n = 24$, 26, 28 and 30) have large or medium-sized chromosomes (Figs 2, 3 and S6); species with $2n = 60$ all have medium to small chromosomes (Figs 3 and S6).

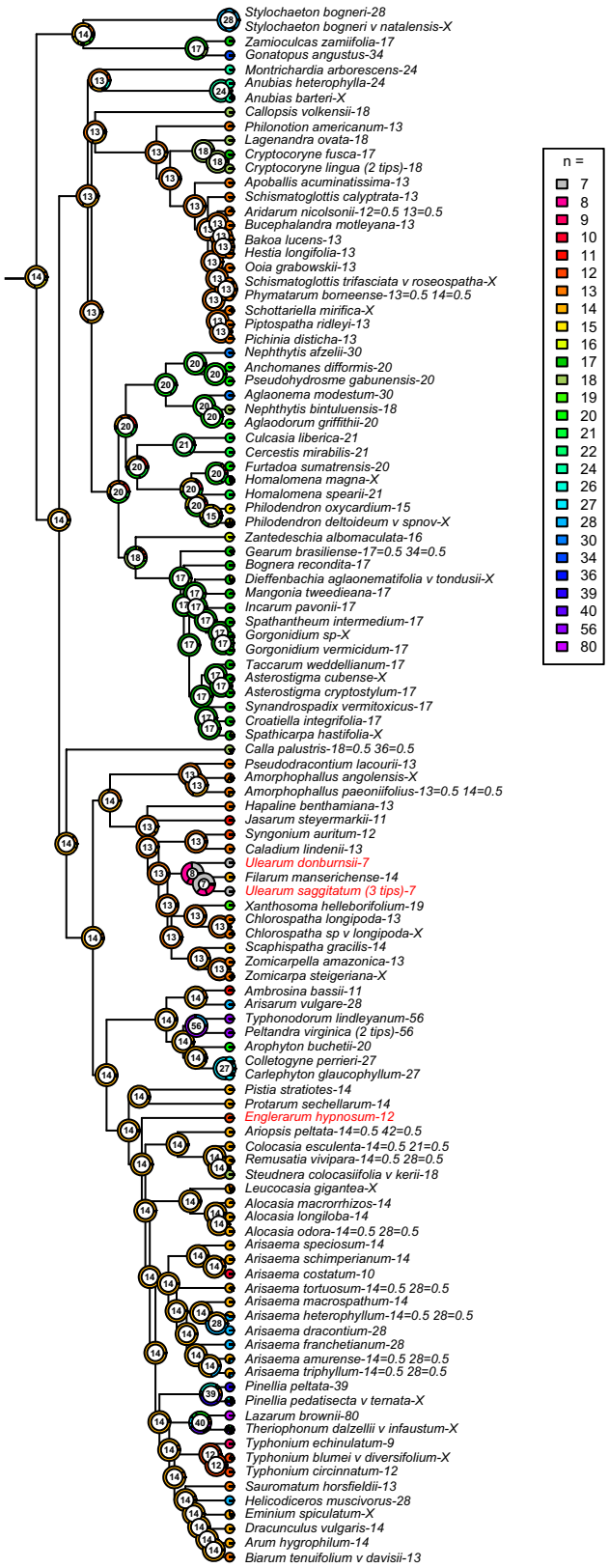

Figure 1. *Continued*

Table 1. Inferred chromosome number evolution in Araceae and their immediate outgroups under the best-fitting model

Tree	Factor	Total tree length	Root-tip length	Best model	Log likelihood	AIC	Rates				No. of events			Total events
							λ	δ	ρ	μ	Gains	Losses	Duplications	Demi.
Ultrametric	0.3	3.6	0.75	crde	-362.3	732.6	6.96	43.67	8.28	7.83	9.1	135.6	21.2	17.8	183.7
Phylogram	0.2	1.5	0.9	lrde	-416.2	844.4	0	75.11	18.67	15.78	35.4	133.1	18.4	15.8	202.7

Column 2 refers to the factor used to multiply branch lengths to obtain a suitable root-to-tip length for the tree (see Sousa *et al.*, 2014); columns 3 and 4 give the lengths obtained after adjusting branch lengths by the multiplication factor; column 5 gives the logarithmic likelihood, and column 6 the AIC scores to the likelihood ratio tests; the symbols for the rates inferred for all events in the tree are: λ, chromosome gain rate; δ, chromosome loss rate; ρ, duplication rate; μ, demiduplication rate; and the number of events refers to the four event types with an expectation > 0.5 (demi.: demiduplication). The last column shows the total number of events inferred on the respective tree.

Table 2. Species of Araceae investigated with their chromosome number, presence of interstitial telomere repeats (ITRs) and the number and distribution of 5S and 45S rDNA signals

Species	2*n*	ITRs	No. of 5S rDNA loci†	Distribution	No. of 45S rDNA loci†	Distribution
Anthurium leuconerum	30	12	1	Subterminal	2	Pericentromeric
Anthurium wendlingeri	30	∞	X	**?**	X	**?**
*Cyrtosperma merkusii**	39	–	1(3)	Subterminal	1(3)	Terminal
Englerarum hypnosum	24	–	1(3)	Interstitial	NA	NA
Monstera deliciosa	60	–	1	Interstitial	1	Terminal
Rhaphidophora pteropoda	60	–	X	**?**	2	Terminal
Scindapsus lucens	60	–	1	Subterminal	2	Terminal
*Spathiphyllum pygmaeum**	30	–	1	Subterminal	3	Interstitial
*Spathiphyllum tenerum**	30	4	1	Subterminal	8	Interstitial
Stenospermatium papayanense	28	–	1	Interstitial	1	Terminal
*Stylochaeton puberulus**	26	–	1	Interstitial	1	Terminal
Ulearum donburnsii	14	–	1	Interstitial	2	Pericentromeric
*Ulearum sagittatum**	14	–	1	Interstitial	2	Pericentromeric
*Pothos repens**	24	–	NA	NA	NA	NA

Species authority names and voucher information are given in Table S1. Asterisks mark species for which chromosome counts were newly obtained. An 'X' indicates species where the hybridization did not work or the pattern was not clear, hence the question mark. The symbol ∞ indicates the presence of ITRs in all chromosomes.
†Unusual number of 5S and 45S rDNA signals is shown in parentheses.
NA, not applicable.

Of the 12 species tested with the 5S rDNA probe, ten had one 5S rDNA locus in one chromosome pair, six of them interstitial and five subterminal (Figs 2B, E, H, K, N; 3B, E, H, K, S6B, E; Table 2). Only *Anthurium wendlingeri* lacked any 5S rDNA signal. In *Rhaphidophora pteropoda*, one 5S rDNA signal was detected on a single chromosome in some cells but its homologue was never seen (data not shown). In *Cyrtosperma merkusii*, with $2n = 39$, and in *Englerarum hypnosum* (the former *Alocasia hypnosa*; Nauheimer & Boyce, 2014), with $2n = 24$, three instead of two 5S rDNA signals were detected and we also found an extra chromosome (Figs 3H and S6E, respectively).

Of the 11 species tested with the 45S rDNA probe, some had one locus, and others up to eight loci on eight chromosome pairs. Both *Ulearum* spp. (with $2n = 14$) had two loci in two chromosome pairs (Fig. 2C, F). For the two *Spathiphyllum* spp. (with $2n = 30$), one had three loci in three chromosome pairs, and the other eight loci in eight pairs, mostly close to, or inside, the pericentromeric regions (Fig. 3C, F). The localization of the rDNA signals is summarized in Table 2 (owing to a lack of material, no experiments could be performed on *Englerarum hypnosum* and no 45S signal was found in *Anthurium wendlingeri*). In *Cyrtosperma merkusii*, three 45S rDNA signals were detected (Fig. 3I), the same number of 5S rDNA signals in that species (Fig. 3H). The centromeric regions of *Ulearum* and *Anthurium* were not DAPI-positive (Fig. 2C, F, O; Table 2).

Of the 13 species tested with the telomere probe, all had telomeric signals at both chromosome ends (Figs 2A, D, J, M, 3A, D, G, J, S6A, D, F, G, H) and three had additional interstitial telomere-like signals, namely *Anthurium leuconerum*, *A. wendlingeri* and *Spathiphyllum tenerum*, with 12, 30 or four interstitial signals localized in pericentromeric regions, respectively (Figs 2M, S6G and 3D).

DISCUSSION

Modelling of chromosome number change in Araceae

Our new chromosome counts (Table 2) and previously published numbers reveal an overall variation in Araceae from $2n = 8$ (*Typhonium* spp.) to $2n = 160$ (*Lazarum* spp.). However, in contrast to what might be expected from such numbers, polyploidy does not appear to have played a large role. Instead, our model-based ML inference of the probable direction in chromosome number change points to dysploidy as the predominant event in karyotype evolution in the family. Caveats applying to our study are the sparse species sampling in Araceae, which comprise 3790 species (26% of them with chromosome counts), and that few chromosome counts are available for the

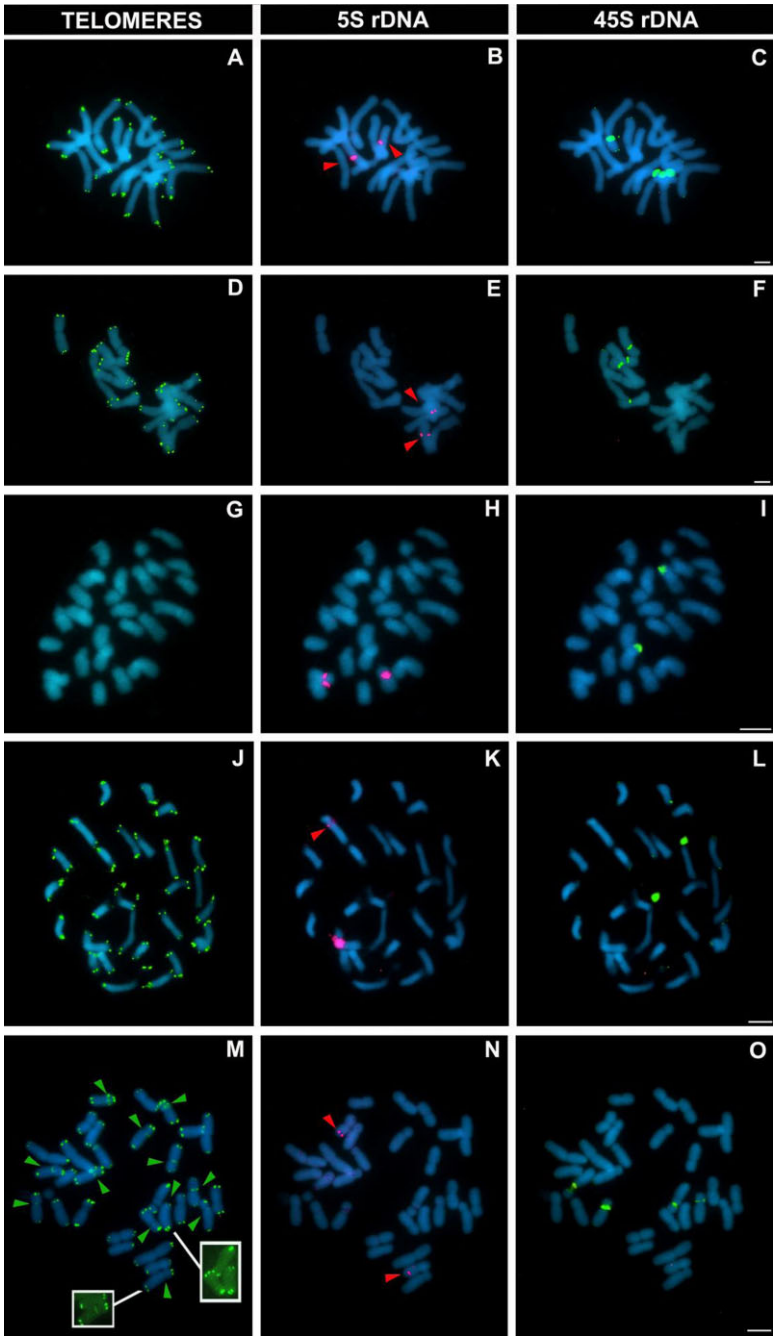

Figure 2. Detection of telomeric, 5S and 45S rDNA signals by FISH in chromosomes of: A–C, *Ulearum donburnsii* ($2n = 14$); D–F, *Ulearum sagittatum* ($2n = 14$); G–I, *Stylochaeton puberulus* ($2n = 26$); J–L, *Stenospermatium papayanense* ($2n = 28$); and M–O, *Anthurium leuconerum* ($2n = 30$). Detection of the telomeres was not performed in *Stylochaeton puberulus*. Red arrowheads indicate the position of weak 5S rDNA signals in some cells, while green ones in M indicate ITRs. Insets in M show chromosomes, without being overlapped with DAPI, with weak ITRs treated with a differential brightness/contrast. Scale bars = 5 μm, and are valid for plates in each row.

outgroup families (Figs 1, S1, S2 and S4), which are phylogenetically distant from Araceae. To infer the most likely events, the ChromEvol approach (Mayrose *et al*., 2010) uses the frequencies of tip states (i.e. chromosome counts) with branch lengths in gene trees (as a proxy for time). In combination, the long phylogenetic branches, especially near the base of the phylogenetic tree, and few species counts result in

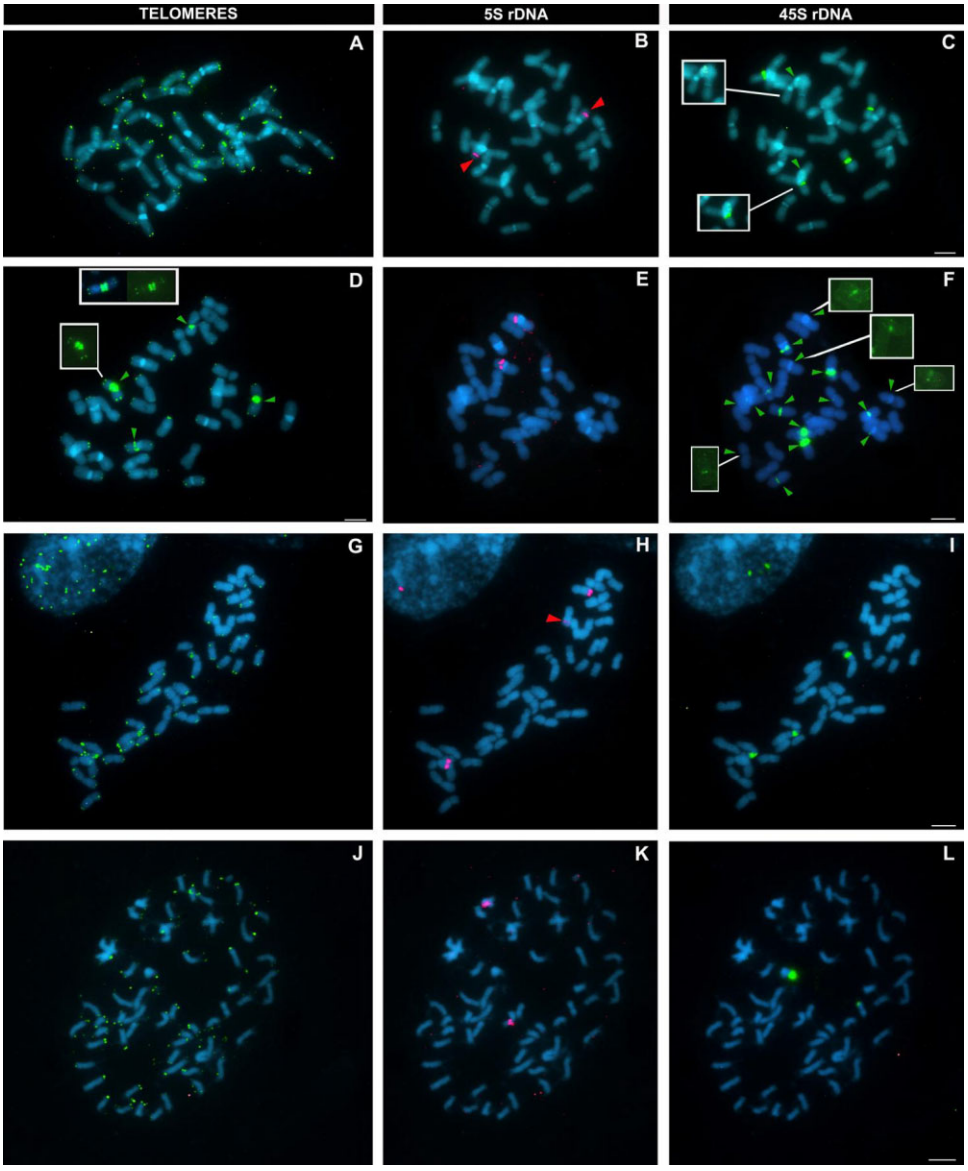

Figure 3. Detection of telomeric, 5S and 45S rDNA signals by FISH in chromosomes of: A–C, *Spathiphyllum pygmaeum* ($2n = 30$); D–F, *Spathiphyllum tenerum* ($2n = 30$); G–I, *Cyrtosperma merkusii* ($2n = 39$); and J–L, *Monstera deliciosa* ($2n = 60$). Red arrowheads indicate the position of weak 5S rDNA signals in some cells, while green ones in D indicate the position of ITRs and in F of 45S rDNA signals. Insets in D display chromosome with telomeric probe without the overlapping with DAPI, and a chromosome from another cell (top) presenting similar telomeric distribution, and in F show chromosomes, without being overlapped with DAPI, with weak 45S rDNA signals treated with a differential brightness/contrast. Scale bars = 5 μm, and are valid for plates in each row.

great uncertainty for the inferred events. The ancestral chromosome number in the family may be $a = 16$ (our Figs 1 and S1, also Cusimano *et al*., 2012, when inferred on a phylogram; $a = 18$ on their ultrametric tree), but this has no statistical support. In contrast, the more recent evolutionary downward trend in chromosome numbers is strongly supported, going from $a = 16$ to 15 to 14 on the ultrametric tree (Fig. 1) and from $a = 16$ to 14 to 13 and back to 14 on the phylogram (Fig. S1).

Previously proposed basic chromosome numbers for Araceae were $x = 7$ (Larsen, 1969; Marchant, 1973) or $x = 14$ (Petersen, 1993), but this was before the inclusion in the family of the five genera of Lemnoideae (in the past treated as Lemnaceae), all of which have relatively high chromosome numbers ($2n = 20$ to $2n = 126$; Cao, 2013) and before it was known that other early lineages of Araceae also have high numbers ($n = 13$, 14, 15, 20, 24 and 30; Figs 1 and S1: tips). They were also arrived at in an idiosyncratic

manner that did not incorporate phylogenetic relationships, even if these relationships had been known (Sousa dos Santos, 2014). Model-based inference of chromosome number, as used here, has the advantage of being reproducible and explicitly incorporating relationships among taxa. Nevertheless and as stressed above, our inferences of numbers near the root have no statistical support and might change with the inclusion of more outgroup chromosome numbers and more early lineages of Araceae.

No evidence for polyploidy from standard FISH probes

We performed FISH in 14 Araceae (telomere repeats: 13 species, 5S rDNA: 12 species, 45S rDNA: 11 species) of which 11 belong to early lineages of the family and three to derived lineages. Although chromosome numbers are now known for 26% of Araceae (Cusimano *et al*., 2012: table S1), FISH studies are scarce and have focused on a derived genus (Sousa *et al*., 2014). In the present study, we sampled early diverging lineages, namely *Anthurium*, *Pothos*, *Stenospermatium*, *Spathiphyllum*, *Monstera*, *Scindapsus* and *Rhaphidophora* (Figs 1 and S1–S5). The results for these genera showed a conserved number of 5S rDNA loci (one per chromosome pair) but variable numbers of 45S rDNA loci (one to eight distributed on one to eight chromosome pairs; see Table 2). Atypical rDNA signals (three instead of four; one on an extra chromosome) were observed in *Cyrtosperma merkusii* ($2n = 39$; Fig. 3H, I) and *Englerarum hypnosum* ($2n = 24$; Fig. S6E). The evolutionary events that led to the reduction of rDNA signals in these species, either loss of an entire chromosome or loss of just the 45S rDNA sequences from its homologues, remain unclear. Also unclear is the evolutionary significance of odd chromosome numbers (Fig. 3H, I), such as found here in *Cyrtosperma merkusii* and earlier in *Amorphophallus*, *Anthurium*, *Apoballis*, *Arisaema*, *Caladium*, *Cryptocoryne*, *Piptospatha*, *Schismatoglottis*, *Typhonium* and *Xanthosoma* (Cusimano *et al*., 2012: table S1; Sousa *et al*., 2014). Although rDNA is the most conserved fraction of eukaryotic genomes, this region in *Anthurium wendlingeri* does not seem homologous to the standard probe from *Arabidopsis thaliana* (Fig. S6G), judging from several attempts of hybridization that produced no detectable FISH signals. This resembles the situation in *Rhynchospora* (Cyperaceae), in which cloning and sequencing of 5S rDNA units from a species of this genus were necessary to detect FISH signals (Sousa *et al*., 2011).

There was no correlation between the number of rDNA loci and ploidy. Polyploids may have at least twice the number of rDNA loci (depending on ploidy) than their parental species (additive polyploidy; see Adams *et al*., 2000; Vanzela, Cuadrado & Guerra, 2003; Ansari, Ellison & Williams, 2008; Robert *et al*., 2008; Weiss-Schneeweiss *et al*., 2008), but they can also have lower numbers of sites (Kovarik *et al*., 2005) or exceed those expected under additive inheritance (De Melo & Guerra, 2003). In some genera, such as *Aloe* L. (Xanthorrhoeaceae) and *Medicago* L. (Fabaceae), the number of rDNA loci is stable regardless of ploidy (Adams *et al*., 2000; Falistocco, 2000). In the genera of Araceae investigated here (*Anthurium*, *Spathiphyllum*, *Ulearum*), we found no evidence of polyploidy (see Table 2) and even species with high chromosome numbers (e.g. $2n = 60$) had just one or two 45S rDNA loci (*Monstera deliciosa* and *Scindapsus lucens*, Figs 3L and S6C). The multiple rDNA loci in *Spathiphyllum* were mainly located in the pericentromeric region, close to or within heterochromatic DAPI-positive bands (Fig. 3C, F). Pericentromeric regions are prone to the insertion of mobile elements, which can mediate the amplification of rDNA in a genome (reviewed by Raskina *et al*., 2008).

Large ITRs in Araceae

The normal distribution of telomere motif repeats at the ends of each chromosome makes it possible to reliably identify chromosomes even in cells with numerous hereditary units. In all species studied here (Figs 2A, D, J, M; 3A, D, G, J; S6A, D, F, G, H), the telomere repeat sequences were detected at both ends of each chromosome, but three species (*Anthurium leuconerum*, *A. wendlingeri* and *Spathiphyllum tenerum*) had additional ITRs (Figs 2M, 3D and S6G). Unexpectedly, we found no ITRs in the two *Ulearum* spp. with the largest chromosomes, whereas *Anthurium leuconerum* and *A. wendlingeri* (Figs 2M, S6G), with medium-sized chromosomes, had ITRs in most or all chromosomes. These signals were located close to the centromere or in subterminal regions (Fig. 2M, S6G) and their number (12 and multiple signals) is the highest so far reported for any angiosperm. That they were discovered in *Anthurium* was unexpected because 80% of the 171 *Anthurium* spp. for which chromosomes have been counted (out of 835 species in the genus) have counts of $2n = 30$ (Cusimano *et al*., 2012: table S1). This consistent chromosome number makes the discovery of ITRs, which could be a sign of chromosome restructuring, surprising.

Interstitial telomeric sites are rare, but are known from *Vicia faba* (Schubert, Rieger & Fuchs, 1995; Fuchs *et al*., 1995: fig. 1), *Eleocharis subarticulata* Boeckeler (Da Silva, González-Elizondo & Vanzela, 2005), *Cephalanthera damasonium* Druce (Moscone *et al*., 2007), *Sideritis montana* (Raskina *et al*., 2008) and two species of *Typhonium* Schott (Sousa *et al*., 2014). In *Vicia faba*, the presence of ITRs was related

to the existence of fusion–fission cycles and in *Typhonium* to Robertsonian-fusion-like rearrangements. The latter mechanism differs from the former in involving the formation of a chromosome with a single centromere after a reciprocal translocation involving two acro- or telocentric chromosomes (Sousa *et al*., 2014). The four ITRs in *Spathiphyllum tenerum* (Fig. 3D) could also have originated via Robertsonian-fusion-like rearrangements, implying a chromosome number reduction, whereas the large ITR bands in *A. wendlingeri* (Fig. S6G) could result from massive repeat amplification of telomere-like sequences. Such large ITR bands have so far been reported only from the gymnosperms *Cycas revoluta* Thunb., *Pinus elliottii* Engelm. var. *elliottii*, *P. densiflora* Siebold & Zucc., *P. taeda* L. and *P. sylvestris* L. (Fuchs *et al*., 1995; Hizume *et al*., 1998; Schmidt *et al*., 2000; Shibata, Matsusaki & Hizume, 2005; Islam-Faridi *et al*., 2007), with each chromosome usually displaying more than one signal (up to six). The *Arabidopsis* telomere motif TTTAGGG or its degenerate variants have previously been reported within tandem repeats, but could not be detected in interstitial positions by FISH using telomere-specific probes (Mlinarec *et al*., 2009).

In *Pinus elliottii* var. *elliottii* and *P. densiflora*, some of the ITRs co-localize with positive DAPI bands, but the regular terminal telomere signals could be visualized only after differential brightness/contrast treatment or not at all (Schmidt *et al*., 2000; Shibata *et al*., 2005; similarly in *P. taeda*, Fuchs *et al*., 1995). In Araceae studied here, we also found co-localization of ITRs and positive DAPI bands in *A. wendlingeri* and *S. tenerum* (Fig. 3D: inserts and Fig. S6G), but in contrast to *Pinus*, their terminals were normally labelled with the standard plant telomeric probe. Whatever their origin, the ITR bands as reported here suggest that nuclear genome assembly in the Araceae may be challenging.

ACKNOWLEDGEMENTS

We are grateful to Josef Bogner for providing material of Araceae, Itay Mayrose for help with modifications to his ChromEvol program, Natalie Cusimano for help with R, Martina Silber for support in the laboratory and to the associate editor and two anonymous reviewers for helpful suggestions. This work was supported by the DFG RE 603/7-1.

REFERENCES

Adams SP, Leitch IJ, Bennett MD, Chase MW, Leitch AR. 2000. Ribosomal DNA evolution and phylogeny in *Aloe* (Asphodelaceae). *American Journal of Botany* **87:** 1578–1583.

Ansari HA, Ellison NW, Williams WM. 2008. Molecular and cytogenetic evidence for an allotetraploid origin of *Trifolium dubium* (Leguminosae). *Chromosoma* **117:** 159–167.

Boyce PB, Croat TB. 2011. *The Überlist of Araceae. Totals for published and estimated numbers of species in aroid genera*. Available at: http://www.aroid.org/genera/130307uberlist.pdf

Cabrera LI, Salazar GA, Chase MW, Mayo SJ, Bogner F, Davila P. 2008. Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. *American Journal of Botany* **95:** 1153–1165.

Cao H. 2013. Chromosomal integration of the *Spirodela polyrhiza* reference genome. 2nd International Conference on Duckweed Research & Applications, New Brunswick, 21–24 August. Available at: http://duckweed2013.rutgers.edu/presentations/10_cao_hieu.pdf

Cheng F, Mandakova T, Wu J, Xie Q, Lysak MA, Wang X. 2013. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *The Plant Cell* **25:** 1541–1554.

Cusimano N, Barrett M, Hetterscheid WLA, Renner SS. 2010. A phylogeny of the Areae (Araceae) implies that *Typhonium*, *Sauromatum* and the Australian species of *Typhonium* are distinct clades. *Taxon* **59:** 439–447.

Cusimano N, Sousa A, Renner SS. 2012. Maximum likelihood inference implies a high, not a low, ancestral haploid chromosome number in the Araceae, with a critique of the bias introduced by '*x*'. *Annals of Botany* **109:** 681–692.

Da Silva CRM, González-Elizondo MS, Vanzela ALL. 2005. Reduction of chromosome number in *Eleocharis subarticulata* (Cyperaceae) by multiple translocations. *Botanical Journal of the Linnean Society* **149:** 457–464.

De Melo NF, Guerra M. 2003. Variability of 5S and 45S rDNA sites in *Passiflora* L. species with distinct base chromosome numbers. *Annals of Botany* **92:** 309–316.

Drummond AJ, Rambaut A. 2007. Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7:** 214.

Falistocco E. 2000. Physical mapping of rRNA genes in *Medicago sativa* and *M. glomerata* by fluorescent *in situ* hybridization. *Journal of Heredity* **91:** 256–260.

Fuchs J, Brandes A, Schubert I. 1995. Telomere sequence localization and karyotype evolution in higher plants. *Plant Systematics and Evolution* **196:** 227–241.

Hizume M, Kurose N, Shibata F, Kondo K. 1998. Molecular cytogenetic studies on sex chromosomes and proximal heterochromatin containing telomere-like sequences in *Cycas revoluta*. *Chromosome Science* **2:** 63–72.

Hizume M, Shibata F, Matsusaki Y, Garajova Z. 2002. Chromosome identification and comparative karyotypic analyses of *Pinus* species. *Theoretical Applied Genetics* **105:** 491–497.

Ijdo JW, Wells RA, Baldini A, Reeders ST. 1991. Improved telomere detection using a telomere repeat probe $(TTAGGG)_n$ generated by PCR. *Nucleic Acids Research* **19:** 4780.

Illa E, Sargent DJ, Girona EL, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, van der Knaap E, Iezzoni A, Gardiner S, Velasco R, Arus P, Chagne D, Troggio M. 2011. Comparative analysis of Rosaceous

genomes and the reconstruction of a putative ancestral genome for the family. *BMC Evolutionary Biology* **11:** 9.

Islam-Faridi MN, Nelson CD, Kubisiak TL. 2007. Reference karyotype and cytomolecular map for loblolly pine (*Pinus taeda* L.). *Genome* **50:** 241–251.

Jung S, Cestaro A, Troggio M, Main D, Zheng P, Cho I, Folta KM, Sosinski B, Abbott A, Celton JM, Arus P, Shulaev V, Verde I, Morgante M, Rokhsar D, Velasco R, Sargent DJ. 2012. Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceous subfamilies. *BMC Genomics* **13:** 129.

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30:** 772–780.

Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R, Rocca J, Soltis DE, Soltis PS. 2005. Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics* **169:** 931–944.

Larsen K. 1969. Cytology of vascular plants: III. A study of aroids. *Dansk Botanisk Arkiv* **27:** 39–59.

Lubaretz O, Fuchs J, Ahne R, Meister A, Schubert I. 1996. Karyotyping of three Pinaceae species via fluorescent *in situ* hybridization and computer-aided chromosome analysis. *Theoretical Applied Genetics* **92:** 411– 416.

Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Heo HY, Huo N, Lazo G, Ma Y, Matthews DE, McGuire PE, Morrell PL, Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W, Dvorak J. 2009. Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 15780–15785.

Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 5224–5229.

Mandakova T, Lysak MA. 2008. Chromosomal phylogeny and karyotype evolution in $x = 7$ crucifer species (Brassicaceae). *The Plant Cell* **20:** 2559–2570.

Marchant CJ. 1973. Chromosome variation in Araceae IV: from Acoreae to Lasieae. *Kew Bulletin* **28:** 199–210.

Mayrose I, Barker MS, Otto SP. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology* **59:** 132–144.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, New Orleans, 1–8.

Mlinarec J, Chester M, Siljak-Yakovlev S, Papeš D, Leitch AR, Besendorfer V. 2009. Molecular structure and chromosome distribution of three repetitive DNA families in *Anemone hortensis* L. (Ranunculaceae). *Chromosome Research* **17:** 331–346.

Moscone EA, Samuel R, Schwarzacher T, Schweizer D, Pedrosa-Harand A. 2007. Complex rearrangements are involved in *Cephalanthera* (Orchidaceae) chromosome evolution. *Chromosome Research* **15:** 931–943.

Nauheimer L, Boyce PC. 2014. *Englerarum* (Araceae, Aroideae): a new genus supported by plastid and nuclear phylogenies. *Plant Systematic and Evolution* **300:** 709–715.

Nauheimer L, Metzler D, Renner SS. 2012. Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytologist* **195:** 938–950.

Pellicer J, Kelly LJ, Leitch IJ, Zomlefer WB, Fay MF. 2014. A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. *New Phytologist* **201:** 1484–1497.

Petersen G. 1993. Chromosome numbers of the genera of Araceae. *Aroideana* **16:** 37–46.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Raskina O, Barber JC, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* **120:** 351–357.

Robert ML, Lim KY, Hanson L, Sanchez-Teyer F, Bennett MD, Leitch AR, Leitch IJ. 2008. Wild and agronomically important *Agave* species (Asparagaceae) show proportional increases in chromosome number, genome size, and genetic markers with increasing ploidy. *Botanical Journal of the Linnean Society* **158:** 215–222.

Schmidt A, Doudrick RL, Heslop-Harrison JS, Schmidt T. 2000. The contribution of short repeats of low sequence complexity to large conifer genomes. *Theoretical and Applied Genetics* **101:** 7–14.

Schmidt T, Schwarzacher T, Heslop-Harrison JS. 1994. Physical mapping of rRNA genes by fluorescent *in situ* hybridization and structural analysis of 5S rRNA genes and intergenic spacer sequences in sugar beet (*Beta vulgaris*). *Theoretical and Applied Genetics* **88:** 629–636.

Schubert I, Rieger R, Fuchs J. 1995. Alteration of basic chromosome number by fusion–fission cycles. *Genome* **38:** 1289–1292.

Schubert I, Schriever-Schwemmer G, Werner T, Adler I-D. 1992. Telomeric signals in Robertsonian fusion and fission chromosomes: implications for the origin of pseudoaneuploidy. *Cytogenetics and Cell Genetics* **59:** 6–9.

Shibata F, Matsusaki Y, Hizume M. 2005. AT-rich sequences containing *Arabidopsis*-type telomere sequence and their chromosomal distribution in *Pinus densiflora*. *Theoretical and Applied Genetics* **110:** 1253–1258.

Sousa A, Barros e Silva AE, Cuadrado A, Loarce Y, Alves MV, Guerra M. 2011. Distribution of 5S and 45S rDNA in plants with holokinetic chromosomes and the 'chromosome field' hypothesis. *Micron* **42:** 625–631.

Sousa A, Cusimano N, Renner SS. 2014. Combining FISH and model-based predictions to understand chromosome evolution in *Typhonium* (Araceae). *Annals of Botany* **113:** 669–680.

Sousa dos Santos A. 2014. *Molecular cytogenetics and phylogenetic modeling to study chromosome evolution in the Araceae and sex chromosomes in the Cucurbitaceae*. Doctoral dissertation, Faculty of Biology, Ludwig-Maximilians University (LMU), Munich.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology* **57:** 758–771.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17:** 1105–1110.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731–2739.

Vanzela ALL, Cuadrado A, Guerra M. 2003. Localization of 45S rDNA and telomeric sites on holocentric chromosomes of *Rhynchospora tenuis* Link (Cyperaceae). *Genetics and Molecular Biology* **26:** 199–201.

Vilanova S, Sargent DJ, Arus P, Monfort A. 2008. Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). *BMC Plant Biology* **8:** 67.

Weiss-Schneeweiss H, Riha K, Jang CG, Puizina J, Scherthan H, Schweizer D. 2004. Chromosome termini of the monocot plant *Othocallis siberica* are maintained by telomerase, which specifically synthesises vertebrate-type telomere sequences. *Plant Journal* **37:** 484–493.

Weiss-Schneeweiss H, Tremetsberger K, Schneeweiss GM, Parker JS, Stuessy TF. 2008. Karyotype diversification and evolution in diploid and polyploid South American *Hypochaeris (*Asteraceae) inferred from rDNA localization and genetic fingerprint data. *Annals of Botany* **101:** 909–918.

Yogeeswaran K, Frary A, York TL, Amenta A, Lesser AH, Nasrallah JB, Tanksley SD, Nasrallah ME. 2005. Comparative genome analyses of *Arabidopsis* spp.: inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Research* **15:** 505–515.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Chromosome number reconstruction for Araceae on a phylogram, rooted on *Acorus calamus*. Pie charts represent the probabilities of inferred chromosome numbers, with the number inside the pie having the highest probability. The rectangular inset shows the frequency with which each event type (gains, losses, duplications, demiduplication) occurs along the branches of the tree. The colour coding of chromosome numbers is explained in the elongate inset on the left. Species investigated by FISH are labelled in red.

Figure S2. Chromosome number reconstruction for Araceae on a phylogram tree rooted on *Acorus calamus*. The inferred frequency of the four possible events (gains, losses, duplications, demiduplications) is shown above branches. The colour coding of event types is explained in the inset. Species investigated by FISH are labelled in red.

Figure S3. Phylogeny obtained from ML analysis of a plastid DNA matrix of 4928 aligned nucleotides for 171 species. Bootstrap support is indicated at nodes.

Figure S4. Chromosome number reconstruction for Araceae on an ultrametric tree rooted on *Acorus calamus*. The inferred frequency of the four possible events (gains, losses, duplications, demiduplications) is shown above branches. The colour coding of event types is explained in the inset. Species investigated by FISH are labelled in red.

Figure S5. Maximum clade credibility tree of a molecular data set of 171 species. Posterior probabilities are indicated at nodes.

Figure S6. Detection by FISH of: telomeric, 5S and 45S rDNA signals in chromosomes of (A–C) *Scindapsus lucens* ($2n = 60$); telomeric and 5S rDNA signals in chromosomes of (D, E) *Englerarum hypnosum* ($2n = 24$); telomeric signals in chromosomes of (F) *Pothos repens* ($2n = 24$) and (G) *Anthurium wendlingeri* ($2n = 30$); and telomeric and 45S rDNA signals in chromosomes of (H, I) *Rhaphidophora pteropoda* ($2n = 60$). Red arrowheads indicate the position of weak 5S rDNA signals, while green arrows in C and I indicate the position of weak 45S rDNA signals. Empty plates labelled 'NO' indicate that experiments using these probes were not made in these species while 'YES' means that experiments were performed but failed or yielded unsatisfactory results. Bars correspond to 5 μm and apply to all plates of a row.

Table S1. Sequence source information.

Table S2. Information on the genera newly studied here.