New Phytologist

Research

The evolutionary dynamics of ancient and recent polyploidy in the African semiaquatic species of the legume genus *Aeschynomene*

Clémence Chaintreuil[1], Djamel Gully[1], Catherine Hervouet[2], Panlada Tittabutr[3], Herizo Randriambanona[4], Spencer C. Brown[5], Gwilym P. Lewis[6], Mickaël Bourge[5], Fabienne Cartieaux[1], Marc Boursot[1], Heriniaina Ramanankierana[4], Angélique D'Hont[2], Neung Teaumroong[3], Eric Giraud[1] and Jean-François Arrighi[1]

[1]Laboratoire des Symbioses Tropicales et Méditerranéennes, IRD, UMR LSTM, Campus International de Baillarguet, 34398 Montpellier, France; [2]CIRAD, UMR AGAP, Plateau de Cytogénétique Moléculaire, 34398 Montpellier, France; [3]School of Biotechnology, Institute of Agricultural Technology, Suranaree University of Technology, Nakhon Ratchasima, Thailand; [4]Laboratoire de Microbiologie de l'Environnement/Centre National de Recherche sur l'Environnement, Antananarivo 101, Madagascar; [5]Institute of Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91 198 Gif-sur-Yvette, France; [6]Comparative Plant and Fungal Biology Department, Royal Botanic Gardens Kew, Richmond, Surrey, TW9 3AB, UK

Author for correspondence:
Jean-François Arrighi
Tel: +33 (0) 4 67 59 38 82
Email: jean-francois.arrighi@ird.fr

Received: *15 January 2016*
Accepted: *4 March 2016*

New Phytologist (2016) **211:** 1077–1091
doi: 10.1111/nph.13956

Key words: *Aeschynomene*, dysploidy, genome downsizing, polyploidy, stem nodulation, symbiosis.

Summary

- The legume genus *Aeschynomene* is notable in the ability of certain semiaquatic species to develop nitrogen-fixing stem nodules. These species are distributed in two clades. In the first clade, all the species are characterized by the use of a unique Nod-independent symbiotic process. In the second clade, the species use a Nod-dependent symbiotic process and some of them display a profuse stem nodulation as exemplified in the African *Aeschynomene afraspera*.
- To facilitate the molecular analysis of the symbiotic characteristics of such legumes, we took an integrated molecular and cytogenetic approach to track occurrences of polyploidy events and to analyze their impact on the evolution of the African species of *Aeschynomene*.
- Our results revealed two rounds of polyploidy: a paleopolyploid event predating the African group and two neopolyploid speciations, along with significant chromosomal variations. Hence, we found that *A. afraspera* (8x) has inherited the contrasted genomic properties and the stem-nodulation habit of its parental lineages (4x).
- This study reveals a comprehensive picture of African *Aeschynomene* diversification. It notably evidences a history that is distinct from the diploid Nod-independent clade, providing clues for the identification of the specific determinants of the Nod-dependent and Nod-independent symbiotic processes, and for comparative analysis of stem nodulation.

Introduction

Rhizobia interact symbiotically with leguminous plants by inducing nitrogen-fixing nodules, generally exclusively formed on the roots. However, in very few tropical legumes that are hydrophytic, nodulation by rhizobia can also occur on stems. This unusual behavior is shared by species belonging to three papilionoid genera, *Aeschynomene*, *Discolobium* and *Sesbania*, and to the mimosoid genus *Neptunia* (Boivin *et al.*, 1997; James *et al.*, 2001). While stem nodulation is uncommon, >15 *Aeschynomene* species have been reported to form stem nodules, making this character widespread in this genus (Alazard, 1985; Boivin *et al.*, 1997; Chaintreuil *et al.*, 2013). The hydrophytic legumes can differ in the environmental conditions by which they form stem nodules. In *Discolobium pulchellum*, *Discolobium leptophyllum* and *Aeschynomene fluminensis*, the development of stem nodules requires permanent submergence in water (Loureiro *et al.*, 1994, 1995; James *et al.*, 2001). Other species such as *Aeschynomene afraspera*, *Aeschynomene aspera*, *Aeschynomene nilotica* and *Sesbania rostrata* are distinctive in their ability to develop stem nodules even in unflooded conditions (Boivin *et al.*, 1997). These aerial stem-nodulating species are among the most active nitrogen-fixing legumes (Alazard & Becker, 1987; Alazard & Duhoux, 1987; Devi, 2013a,b). As a consequence, they have been proposed as green manure crops, notably to enhance rice production.

An original feature that is found only in some stem-nodulating *Aeschynomene* species is their capacity to symbiotically interact with photosynthetic *Bradyrhizobium* strains (Evans *et al.*, 1990; Giraud & Fleischman, 2004). The photosynthetic activity of such strains has been shown in *Aeschynomene sensitiva* to facilitate *ex planta* survival and infectivity, and to directly supply energy to the bacterium that can be used for nitrogen fixation during stem nodulation (Giraud *et al.*, 2000). However, the most unexpected feature

encountered in certain *Aeschynomene* species is that they are nodulated by *Bradyrhizobium* strains lacking the *nodABC* genes required for the synthesis of the key signal molecules called Nod factors (Giraud *et al.*, 2007). This revealed the existence of a Nod-independent process in these species. A phylogenetic study of the semiaquatic *Aeschynomene* revealed that the Nod-independent species form a monophyletic clade referred as the Nod-independent group, whereas stem nodulation in aerial condition is present in two clades (Fig. 1) (Chaintreuil *et al.*, 2013). The first one corresponds to the Nod-independent group that is of American origin and contains only stem-nodulating species. The second one comprises species using only a Nod-dependent symbiotic process, some of them, such as *A. afraspera*, being able to develop aerial stem nodules. Since, this clade contains mostly African species, it is here named the African Nod-dependent clade (Fig. 1).

Recent advances on the Nod-independent *Aeschynomene* clade revealed it contained species that are mainly diploid ($2n = 20$) but that multiple hybridization and polyploidization events significantly contributed to its radiation (Arrighi *et al.*, 2014). From this knowledge, the diploid *Aeschynomene evenia* has been proposed as a new model species to decipher the molecular mechanisms of the Nod-independent process (Arrighi *et al.*, 2012). It is thus expected to provide important insights into the evolution of nodulation in this group (Sprent & James, 2008). But to progress in this issue, it would also be interesting to consider other *Aeschynomene* species using a classical Nod-dependent symbiotic process. Indeed, comparative analysis between *Aeschynomene* species that use a distinct symbiotic process could facilitate the identification of the specific determinant of each process. In addition, besides differing in the signal triggering nodule organogenesis, striking differences are also observed in bacteroid differentiation. Nod-independent species contain elongated bacteroids, while those of the Nod-dependent species contain spherical ones (Czernic *et al.*, 2015). Therefore, comparison between these two groups could also provide clues to help us better understand the mechanisms of bacterial adaptation during symbiosis.

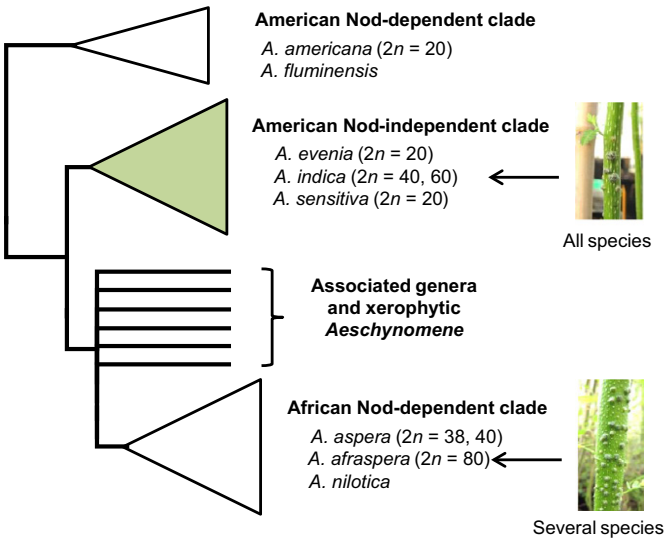

Fig. 1 Tree representation of the phylogeny of the legume genus *Aeschynomene*. Triangles represent the different hydrophytic *Aeschynomene* clades with the naming used in this study and some notable species indicated. The Nod-independent clade is colored in green and the occurrence of aerial stem nodulation in the different clades is indicated on the right. The figure is a simplified and stylized phylogeny based on a tree in Chaintreuil *et al.* (2013). Chromosome numbers are indicated when known (referenced in Index to Plant Chromosome Numbers (IPCN) and from Arrighi *et al.* (2014)).

Despite the scientific and agronomic interest in the *Aeschynomene* species of the African clade, knowledge of their genetics and genomics is limited. Chromosome counts indicated that *A. afraspera* has $2n = 80$ and *A. aspera* has $2n = 38$ or 40 (Index to Plant Chromosome Numbers (IPCN)), suggesting polyploid origins. This raises questions about the evolution of ploidy status within the African *Aeschynomene* clade compared with the Nod-independent group, which is of diploid origin ($2n = 20$). Polyploidy is recognized as an important process in plant evolution and is often invoked as a driver of diversification (Soltis *et al.*, 2009). It can act alone, resulting in autopolyploidy, or in concert with hybridization, producing allopolyploids, and both modes lead to plant speciation. In addition to polyploidy, chromosomal rearrangements can play an important role in the evolution of genome size and karyotype, so they are contributors to genome diversification. Therefore, to unravel the evolutionary dynamics of the African *Aeschynomene* clade and compare it with the history of the Nod-independent group, we undertook a molecular and cytogenetic approach. We determined the ploidy levels and the phylogenetic relationships for the different species of the group, studied their genome structures and rearrangements and analyzed the potential correlation of ploidy changes with the evolution of the breeding system and stem nodulation.

Materials and Methods

Plant material and microscopic observation

All the accessions of *Aeschynomene* used in this study, their geographic origin and their sources are listed in Table 1 and Supporting Information Table S1. Seeds were scarified with sulfuric acid for germination and plants were grown in pots filled with atapulgite under glasshouse conditions (temperature, 26–36°C; relative humidity, 70–80%; insect-proof screens) as detailed in Arrighi *et al.* (2012). Phenotypic traits such as flower morphology and stem nodulation were directly analyzed in the glasshouse. Microscopic observations were performed on whole and 80-µm-thick vibratome (Leica VT1000S, Leica Biosystems, Nussloch, Germany) sections of fresh nodule samples using a stereomicroscope (Nikon AZ100; Campigny-sur-Marne, France). Pictures of nodule cross-sections were assembled using Adobe Photoshop.

Molecular methods

Genomic DNA was isolated from fresh leaves using the CTAB extraction method. The nuclear ribosomal internal transcribed spacer region (ITS: ITS1-5.8S rDNA gene-ITS2) and the chloroplast *matK* gene were amplified with the primers listed in Table S2. Single sequence repeat (SSR) motifs present in the *A. afraspera* transcriptomes developed by Czernic *et al.* (2015)

Table 1 Characteristics of the *Aeschynomene* species and samples used in this study

Species	LSTM code	2C DNA content (pg)	Chromosome number (2*n*)	Ploidy level (x)	Country of origin	Voucher	Seed bank
A. afraspera	1	5.28 ± 0.03	76	8	Senegal	LSTM1	LSTM
	2				Senegal	LSTM2	LSTM
	225	5.28 ± 0.05			Chad	CIAT 22491	CIAT
	227				Gambia	IRRI 14142	IRRI
	228				Guinea	IRRI 14143	IRRI
A. aspera	47	4.26 ± 0.04	38	4	Sri Lanka	IRRI 013020	IRRI
	264	3.91 ± 0.03			Cambodia	LSTM264	LSTM
	266	3.92 ± 0.06			Thailand	LSTM266	LSTM
	294	4.22 ± 0.02			India	LSTM294	LSTM
A. crassicaulis	13	3.10 ± 0.03	40	4	Senegal	LSTM13	LSTM
A. cristata	169	1.92 ± 0.02	38	4	Central African Republic	ILRI 16880	ILRI
A. elaphroxylon	112	3.20 ± 0.02	40	4	Senegal	LSTM112	LSTM
	148				Burundi	IRRI 12148	IRRI
	159				Ethiopia	ILRI 010590	ILRI
A. evenia	76	0.85 ± 0.02	20	2	Malawi	CIAT22838	CIAT
A. fluitans	118	4.16 ± 0.04	38	4	Zambia	CPI 052338	AusPGRIS
	119				Zambia	CPI 073047	AusPGRIS
A. nilotica	53	4.11 ± 0.03	38	4	Senegal	IRRI 014040	IRRI
	247				Mali	IRRI 14152	IRRI
A. pfundii	54	3.64 ± 0.07	40	4	Zimbabwe	LSTM54	LSTM
A. schimperi	123	3.93 ± 0.04	28	4	Tanzania	ILRI 000659	ILRI
'grandiflora'	124	3.89 ± 0.03			Tanzania	ILRI 001170	ILRI
	138				Tanzania	CIAT 22523	CIAT
A. schimperi	49*	4.82 ± 0.12	28	4	Democratic Republic of Congo	IRRI 012146	IRRI
	59	8.54 ± 0.02			Senegal	IRRI 012156	IRRI
	125	9.12 ± 0.25			Ethiopia	ILRI 002056	ILRI
	126	8.49 ± 0.11	56	8	Ethiopia	ILRI 009675	ILRI
	128	8.49 ± 0.07			Ethiopia	ILRI 013490	ILRI
	129	9.20 ± 0.12	56	8	Ethiopia	ILRI 013814	ILRI
	171	8.50 ± 0.01			Madagascar	IRRI 012059	IRRI
	309	8.49 ± 0.04			Madagascar	LSTM309	LSTM
A. uniflora	61	2.62 ± 0.02	40	4	DR Congo	IRRI 013158	IRRI
	114	2.72 ± 0.02			Zambia	CPI 052333	AusPGRIS
	137	2.57 ± 0.02			Uganda	CPI 060164	AusPGRIS
	213	2.68 ± 0.01			Tanzania	ILRI 700	ILRI
	318	2.69 ± 0.02			Madagascar	LSTM318	LSTM
	319	2.70 ± 0.04			Madagascar	LSTM319	LSTM
	320	2.68 ± 0.02			Madagascar	LSTM320	LSTM

*Identified as *A. cristata* in the IRRI Aquatic Legume Collection.

and E. Giraud (unpublished) were identified and analyzed as performed in Arrighi *et al.* (2013). Single nuclear genes were identified in the *A. evenia* (CIAT22838) transcriptome (J. F. Arrighi, unpublished) and used to identify orthologs in the *A. afraspera* transcriptomes by BLAST search (Table S2). Primers were designed using the *A. afraspera* sequences in order to amplify all homeologs (Table S2). PCR amplifications, cloning and sequencing of PCR products were performed as described in Arrighi *et al.* (2013, 2014). The DNA sequences generated in this study were deposited in GenBank under accession numbers KT821096–KT821213 (Table S2).

Sequence data analysis

For the phylogenetic analyses, the gene sequences were aligned in CLUSTALX, v.1.81b (Thompson *et al.*, 1997), using the default parameters and the alignments were checked in GENEDOC v.2.7 (Nicholas & Nicholas, 1997). Phylogenetic reconstructions were performed with MEGA v.5 (Tamura *et al.*, 2011) using the maximum likelihood approach and the Tamura 3-parameter model with a 1000× bootstrap (BS). The data are presented as rooted trees using *A. evenia* as outgroup.

For an estimate of divergence times in *Aeschynomene*, a simple molecular clock assumption was employed. Coding sequences of six single nuclear genes from *A. evenia* and homeologous copies in *A. afraspera* were obtained from the transcriptome databases where they are available (Table S2). The clock calibration was carried out as already performed by Moretzsohn *et al.* (2013), based on well-established evolutionary divergences within legumes: separation of the Dalbergioid clade (containing *Aeschynomene*) *c.* 55 million yr ago (Ma), separation of the Millettioid (containing *Glycine*) and Galegoid (containing *Lotus* and *Medicago*) clades *c.* 50 Ma (Lavin *et al.*, 2005), and separation of the two components of the soybean genome *c.* 13 Ma (Schmutz *et al.*, 2010). For this, sequences orthologous to those identified in *Aeschynomene* spp. were searched in the sequenced genomes of

Glycine max, *Lotus japonicus* and *Medicago truncatula*. Coding sequences were cropped, aligned and used to calculate the number of silent nucleotide substitutions per synonymous site (= K_s values) with the DnaSP 5.10 software (Librado & Rozas, 2009). Substitution rates (SRs) were calculated for the four species divergences with previously known estimates and, assuming the substitution rate per time unit (SR Ma^{-1}) for the *Aeschynomene* genome divergences is the same as for the other species divergences, an average SR was obtained. Divergence times (T) in *Aeschynomene* were deduced using the equation $T = K_s/2 \times SR$.

Genome size estimation

Genome sizes were measured by flow cytometry using leaf material as described in Arrighi *et al.* (2012). Genome size estimations resulted from measurements for three plants per accession, and *Lycopersicum esculentum* 'Roma' (2C = 1.99 pg) was used as the internal standard.

Cytogenetics

Chromosomes were prepared from root tips as explained in Arrighi *et al.* (2012). For genomic *in situ* hybridization (GISH) on 8x *Aeschynomene* species, genomic DNA of putative genome donors were labeled with digoxigenin-11-dUTP and detected with Texas Red (red) or labeled with biotin-14-dUTP and detected with fluorescein isothiocyanate (green). DNA labeling, chromosome hybridization and fluorescent image capture were performed as described in Arrighi *et al.* (2014).

Results

Assessing the African *Aeschynomene* group

To investigate the characteristics of the whole group, we increased the taxonomic sampling performed by Chaintreuil *et al.* (2013) and used for comparison the diploid *A. evenia* as a representative of the Nod-independent group that was previously analyzed in Arrighi *et al.* (2014) (Fig. 1). A total of 38 accessions were sampled, procured from either different seed banks or direct field collections (Table 1). All come from the African continent, with the exception of *A. aspera* which is exclusively Asian. Species identity of the accessions was checked based on morphological features. Among the species of the African group, *Aeschynomene schimperi* is described as displaying a noticeable variability in flower sizes, and Verdcourt (1971) suggested that larger-flowered specimens may be polyploid. We observed in our collection of accessions two flower morphotypes with significantly different sizes, being either *c.* 16–20 mm or *c.* 24 mm long. To distinguish them, accessions displaying the largest flowers were qualified as 'grandiflora', as previously proposed by Peter (1928) (Table 1).

DNA contents were estimated by flow cytometry for most accessions in order to provide a first view of the characteristics of the genomes encountered in the African group (Table 1). When comparing the species, genome size was found to present an extended range, varying from 1.92 pg/2C for *A. cristata* to 5.28 pg/2C for *A. afraspera* and up to 9.20 pg/2C for some *A. schimperi* accessions. These values are higher than those observed in the Nod-independent clade, with *c.* 1 pg/2C for the diploid species including *A. evenia* (Arrighi *et al.*, 2014). These genome sizes may be indicative of different genome structures underlined by ploidy changes and generally larger genomes for the whole African group. It is also worth noting that for *A. schimperi*, the 'grandiflora' morphotype displays a 3.9 pg/2C genome size while other accessions are endowed with a 4.8, 8.5 or 9.2 pg/2C genome (Table 1).

To link the genome sizes to ploidy levels, some accessions for each species were selected for chromosome counting (Table 1; Fig. S1). Five chromosome numbers were obtained: $2n = 28$ for *A. schimperi* 'grandiflora' and one *A. schimperi* accession; $2n = 38$ for *A. aspera*, *A. fluitans*, *A. nilotica* and *A. cristata*; $2n = 40$ for *A. crassicaulis*, *A. elaphroxylon*, *A. pfundii* and *A. uniflora*; $2n = 56$ for other *A. schimperi* accessions; and $2n = 76$ for *A. afraspera*. When compared with *A. evenia* ($2n = 20$) (Fig. 1), these data suggest that the African *Aeschynomene* group may have a general tetraploid status. But the occurrence of chromosome number variations also points to possible dysploidy/aneuploidy events and higher ploidy levels.

Genetic relationships in the African *Aeschynomene* group

To test these hypotheses regarding the evolution of chromosome numbers in the African *Aeschynomene*, the genetic relationships between the different species was investigated. For this, phylogenetic reconstruction of the whole group was undertaken using the nuclear ribosomal *ITS* region and the chloroplast *matK* gene (Tables S1, S2). The maximum likelihood analysis of these two sequences produced very similar topologies (Fig. 2a,b) resolving the accessions in five well-supported clades (BS > 74% for *ITS* and BS > 98% for *matK*) with coherent chromosome numbers. Clade 1 groups *A. aspera* ($2n = 38$), *A. fluitans* ($2n = 38$), *A. nilotica* ($2n = 38$), *A. cristata* ($2n = 38$) and *A. afraspera* ($2n = 76$) together. Clade 2 clusters *A. schimperi* 'grandiflora' ($2n = 28$) and *A. schimperi* ($2n = 28,56$). Clades 3 and 4 are monospecific and correspond to *A. uniflora* ($2n = 40$) and *A. elaphroxylon* ($2n = 40$), respectively, while clade 5 gathers *A. pfundii* ($2n = 40$) with *A. crassicaulis* ($2n = 40$). The observed tree topologies support the hypothesis of a basically tetraploid group ($2n = 4x = 40$) that would be characterized by two dysploid/aneuploid events occurring in clade 1 ($2n = 4x = 38$) and clade 2 ($2n = 4x = 28$) and by the presence of two putative octoploid species, *A. afraspera* ($2n = 8x = 76$) and *A. schimperi* ($2n = 8x = 56$), that possess twice the chromosome number of the related species or accessions within their relative clades.

Both phylogenies confirmed botanical species delineation. Intriguingly, the positions of *A. schimperi* 'grandiflora' (4x) and *A. schimperi* (8x) accessions were found to be intermingled, whereas *A. schimperi* (4x) was more distinct, suggesting complex genetic relationships in clade 2. But the most noticeable feature concerns how *A. afraspera* (8x) is positioned compared with the 4x species of clade 1. Indeed, in the *ITS* tree, this species strongly

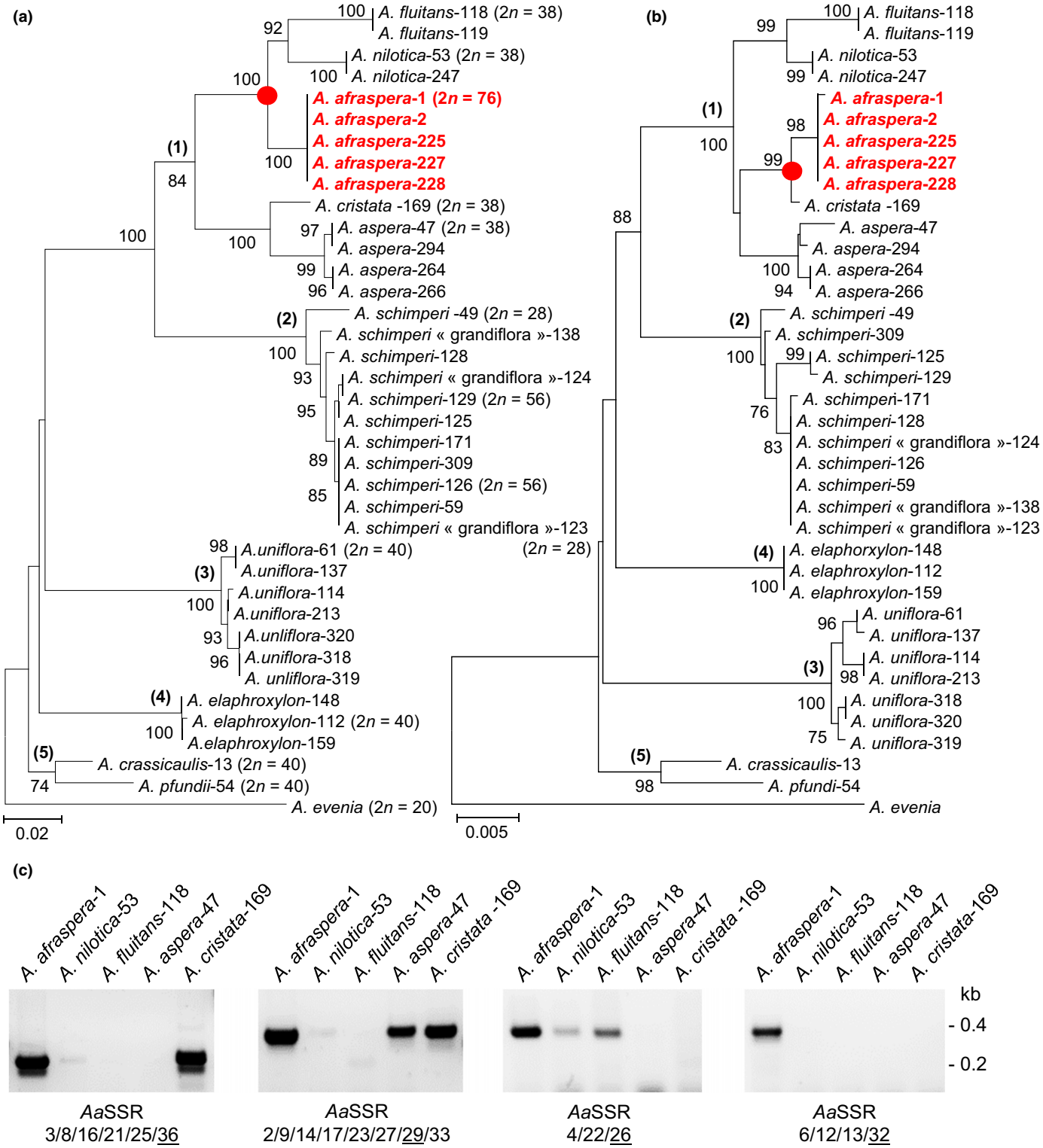

Fig. 2 Phylogenetic and genetic relationships. (a, b) Maximum likelihood phylogenies were obtained with the nuclear internal transcribed spacer (ITS) sequence (a) and the cytoplasmic *matK* sequence (b). *Aeschynomene evenia* is used as an outgroup. Numbers at nodes represent bootstrap values (% of 1000 replicates). The different clades are identified by numbers into brackets and the main disagreement regarding *A. afraspera* between the two trees is highlighted by a red circle. Chromosome numbers are indicated when determined for the accessions. (c) Cross-amplification tests using *Aa*SSR markers. Single sequence repeat (SSR) markers designed from *Aeschynomene afraspera* transcriptome sequences were tested for PCR amplification in the different species of the *Afraspera* lineage. AaSSR markers producing the same amplification profile are listed below the picture and the one used for illustration is underlined. Numbers following the species names refer to the accession code shown in Table 1.

clusters with *A. nilotica* and *A. fluitans* (BS of 100%), whereas in the *matK* tree it appears to be closely related to *A. cristata* (BS of 99%). In conjunction with the chromosome numbers, this discordance between the *ITS* and the *matK* trees suggests an allopolyploid origin for *A. afraspera*.

To further substantiate the hypothesis of a hybrid origin, we took advantage of a transcriptome database available for *A. afraspera* (Czernic *et al.*, 2015) to develop a set of 36 *Aa*SSR markers (Table S1) and to perform cross-amplification tests on the species related to *A. afraspera*, using one accession per species. Fifteen *Aa*SSR markers amplified in the five accessions tested, whereas 21 *Aa*SSR markers displayed restricted amplification profiles. In the latter case, they could be sorted in three categories (Fig. 2c). The first one corresponded to markers amplifying specifically in *A. cristata* (*Aa*SSR3, 8, 16, 21, 25, 36) or also in *A. aspera* (*Aa*SSR2, 9, 14, 17, 23, 27, 29, 33). The second gathered markers (*Aa*SSR4, 22, 26) that amplified in *A. nilotica* and *A. fluitans*, while the remaining markers (*Aa*SSR6, 12, 13, 32) were not amplified in any of the species related to *A. afraspera*. These differential amplification profiles indicate that *A. afraspera* presents dual genome affinities with, on the one hand, *A. cristata* (and to a lesser extent with *A. aspera*) and, on the other hand, both *A. nilotica* and *A. fluitans*. The specific amplification of some *Aa*SSR markers may be explained by the differentiation of the *Aeschynomene* genomes. But it is also possible that another species, which remains to be identified, could amplify these specific markers.

Genome structure of *A. afraspera* deduced from transcriptome data

In order to get further insights into the genome structure of *A. afraspera*, we used the transcriptome databases developed for this species (Czernic *et al.*, 2015; E. Giraud, unpublished) and for *A. evenia* (J. F. Arrighi, unpublished), and we performed gene comparisons. Interestingly, *in silico* analysis revealed three to four distinct sequences for many genes for *A. afraspera* when only one sequence was encountered for *A. evenia* that is known to be diploid. Six genes were selected for detailed comparisons: ascorbate peroxidase 2 (*APX2*), asparagine synthetase 2 (*AS2*), eukaryotic translation initiation factor α (*eiF1α*), plasma membrane intrinsic protein 2;7 (*PIP2;7*), sucrose synthase (*SUS*) and tonoplast intrinsic protein 1;1 (*TIP1;1*) genes (Table S1). Phylogenetic analyses of the coding sequences of the six single-copy nuclear genes by the maximum likelihood method always resulted in a topology that clustered the distinct sequences into two highly divergent clades (BS of 87–100%). These clades were arbitrarily named 'A' and 'B' and the gene copies contained in these clades as '1' and '2' (Fig. 3a). These phylogenetic data are consistent with *A. afraspera* being endowed with a polyploid genome of a complex structure corresponding to the association of four subgenomes: A1A2B1B2. These gene copies would correspond to homeologs, each one putatively representing a different parental lineage. Therefore, it is likely that *A. afraspera* has undergone two rounds of polyploidization events, as it diverged from other diploid *Aeschynomene* such as *A. evenia*.

For a rough estimate of the divergence time of the four subgenomes of *A. afraspera* along with the one of *A. evenia*, a simple molecular clock assumption was employed. The clock calibration was carried out based on previously known estimates of the divergence of main lineages in the legume phylogeny: 55 Ma, *Aeschynomene–Lotus*/*Medicago*/*Glycine*; 50 Ma, *Glycine–Lotus*/*Medicago* (Lavin *et al.*, 2005); and 10–13 Ma, *Glycine–Glycine* (Egan & Doyle, 2010; Schmutz *et al.*, 2010). Using this calibration together with the observed rates of silent substitutions per synonymous site between the different species for the six genes detailed earlier, it was possible to infer for *Aeschynomene* a mean divergence rate of 1.13×10^{-8} mutations per site yr^{-1} and thus to make age estimations for the major *Aeschynomene* genome divergences. This gives an approximate date of 20.8 ± 3.6 Ma for the *A. evenia–A. afraspera* divergence, 18.6 ± 3.1 Ma for the A and B genome divergence and 5.5 ± 2 Ma for the *A. afraspera* parental lineage divergence (Table 2; Fig. S2). The two latter divergence times constitute the maximum age estimates for the two polyploidization events evidenced for *A. afraspera*.

Genome structure of the African clade inferred from molecular phylogenies

For the reconstruction of the polyploidization history of *A. afraspera* in line with the other members of the African clade, we selected two genes, *eif1α* and *SUS*, being either intronless or presenting a long exon, and sequenced them in selected accessions representing the different species of the group. To capture all possible copies of these genes, several primers were developed based on the transcriptome sequences of *A. afraspera* (Table S1). In most cases, we isolated a pair of putative homeologs, and in several other instances three or four putative homeologs. The two genes treated separately gave similar results when subjected to maximum likelihood analysis (Fig. 3b,c; Table S2) with the distinct sequence types clustering in two major clades (one of the two clades supported by a BS of *c.* 80%) and the diploid *A. evenia* being used as an outgroup. Importantly, for each species, the homeologous copies nested into each of the two major clades. One clade contained the 'A' versions for the *eif1α* and *SUS* genes identified in *A. afraspera*, while the other one gathered the 'B' versions. The homeologous copies were accordingly referred to as 'A' and 'B' for the other species. These data support the idea that the whole African group is fundamentally tetraploid (4x) with a common AB genome structure, indicating that a single ancient polyploid event occurred and preceded the diversification of the African group.

However, transcriptome data had indicated that *A. afraspera* contains three or four homeologous sequences for these two genes. The phylogenetic trees assigned one A or B homeologous version to the sister of *A. cristata* and the other (related) one to A. *fluitans* and *A. nilotica*, but with poor to moderate BS support (Fig. 3b,c). Despite this, the obtained topologies are congruent with those observed in the *ITS* and *matK* phylogenies that resolved strong relationships for *A. afraspera* with *A. fluitans* and *A. nilotica* in one case and with *A. cristata* in the other (Fig. 2a,b). Hence, these data further support the possibility that *A. afraspera*

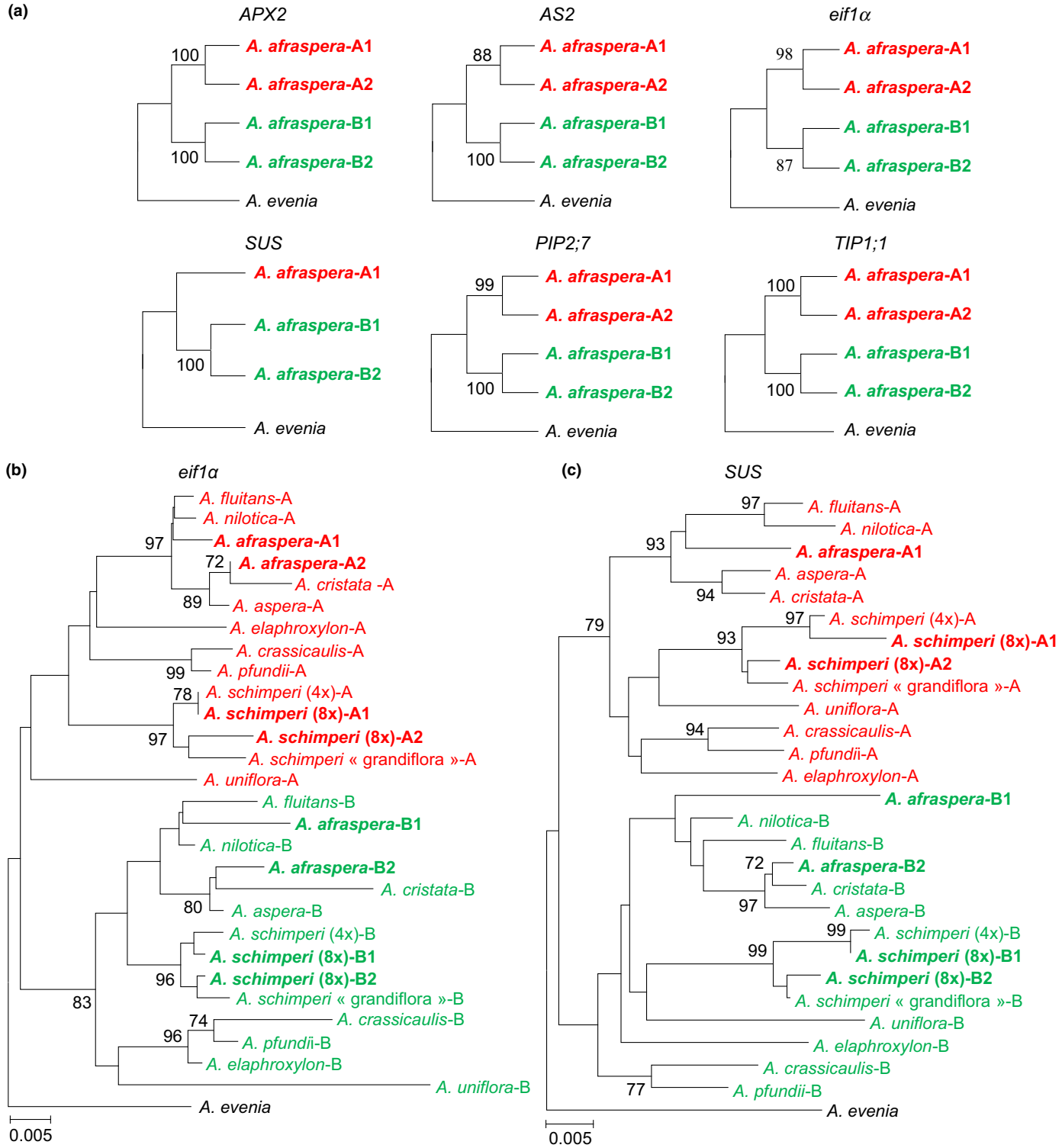

Fig. 3 Phylogenetic trees of duplicated copies of nuclear genes. Maximum likelihood phylogenies obtained with transcriptome data of *Aeschynomene afraspera* and *Aeschynomene evenia*, and after sequencing of PCR products for the other *Aeschynomene* species. (a) Ascorbate peroxidase 2 (*APX2*), asparagine synthetase 2 (*AS2*), eukaryotic translation initiation factor α (*Eif1α*), plasma membrane intrinsic protein 2;7 (*PIP2;7*), sucrose synthase (*SUS*) and tonoplast intrinsic protein 1;1 (*TIP1,1*) genes. (b) *eif1α* gene. (c) *SUS* gene. The diploid *Aeschynomene evenia* is used as an outgroup. References of the accessions used are given in Supporting Information Table S2. -A, -A1,-A2, -B, B1 and -B2 indicate the different copies found and are colored in red for the A versions and green for the B versions. Species inferred to be 8x are indicated in bold. Numbers at nodes represent bootstrap values (% of 1000 replicates).

Table 2 Divergence time for different splits within the *Aeschynomene*

Evolutionary divergence	Substitution per site (Ks)							Divergence	
	APX2 (750 pb)	*AS2* (1629 pb)	*eiF1α* (435 pb)	*PIP2;7* (827 pb)	*SUS* (1128 pb)	*TIP1;1* (750 pb)	Mean (SR)	(SR Ma^{-1})	(Ma ± SD)
Medicago–Glycine	0.6298	0.4500	–	0.6084	0.6915	0.6890	0.6137	0.0123	50
Lotus–Glycine	0.5402	0.3935	–	–	0.4904	0.5004	0.4811	0.0096	50
Medicago–A. evenia	0.5922	0.6997	–	0.8147	0.6611	0.6096	0.6755	0.0123	55
Lotus–A. evenia	0.5430	0.6458	–	–	0.7126	0.4426	0.5860	0.0107	55
A. evenia–Glycine	0.5554	0.6149	0.6124	0.6606	0.7068	0.4710	0.6035	0.0110	55
A. afraspera–Glycine	0.5008	0.5703	0.7249	0.6652	0.6558	0.5900	0.6178	0.0112	55
Glycine A–Glycine B	0.1469	0.0853	0.2971	0.0777	0.1439	0.1885	0.1566	0.0120	13
A. afraspera X1–X2	0.03755	0.0678	0.0426	0.0476	0.0995	0.0589	0.0590	0.0113*	**5.2 ± 2**
A. afraspera A–B	0.1800	0.2418	0.2099	0.1563	0.2373	0.2370	0.2104		**18.6 ± 3.1**
A. afraspera–A. evenia	0.1965	0.2623	0.1878	0.2258	0.2937	0.2447	0.2351		**20.8 ± 3.6**

Substitution rates (SR) for synonymous substitutions in the coding sequence of nuclear genes for the four different legume group divergences with previously known time estimates and for unknown divergences within *Aeschynomene*. Numbers in bold are inferred values. Ma, million years ago.
*Mean of the known divergences.
–, data that could not be estimated.

is an octoploid (8x) species of allopolyploid origin. Although *A. cristata* was identified as a probable genome donor, the variable and poorly supported relationships with *A. fluitans* and *A. nilotica* suggest that the other parent has not been sampled. Whether this species is not yet identified, is extinct or corresponds to the ancestor of *A. fluitans* and *A. nilotica* remains to be determined. Whatever the case, all these data corroborate the observations made with the *ITS*/*matK* discordance and the *Aa*SSR cross-amplifications. These results also confirm that *A. afraspera* and *A. aspera*, which have long been confused as a result of morphological similarity, are different species (Léonard, 1954), and further indicate that *A. aspera* is not a genome donor of *A. afraspera*, but that the sister species *A. cristata* is. Similarly, *A. schimperi* (8x) was characterized by four gene copies, supporting the fact that it, too, is an octoploid species (Fig. 3b,c). Well supported sister-group relationships were recovered in three cases out of four with *A. schimperi* (4x) and weaker supported relationships were observed with *A. schimperi* 'grandiflora'. However, in that latter case, the tree topologies corroborate the highly supported relationships observed in the *ITS* and *matK* phylogenies (Fig. 2a,b). All taken together, these results also point to an allopolyploid origin for *A. schimperi* (8x) that would derive from *A. schimperi* (4x) and *A. schimperi* 'grandiflora' (4x) or closely related taxa and, thus, they clarify the genetic relationships between the different *A. schimperi* accessions.

Cytogenetic analysis of the hybrid origin of *A. afraspera* and *A. schimperi* (8x)

Aeschynomene afraspera is hypothesized to result from the hybridization of *A. cristata* and another genome donor related to *A. fluitans* and *A. nilotica*, based on phylogenetic analyses. In clade 1, *A. cristata* is notable in that it displays a relatively small genome (1.92 pg/2C) that corresponds to only 46% of the genome size of three other 4*x* species, *A. aspera*, *A. fluitans* and *A. nilotica* (Table 1). Thus, *A. cristata* shows a drastic and specific genome downsizing. In order to understand its impact on the genome constitution of the derived hybrid *A. afraspera*, metaphase chromosomes of related species were compared (Fig. 4). *A. aspera* and *A. nilotica* both displayed well-shaped chromosomes, while those of *A. cristata* sharply differed in their obviously smaller sizes. Interestingly, *A. afraspera* displayed an extended range of chromosome sizes (Fig. 4). Therefore, one might speculate that *A. afraspera* has inherited its smaller chromosomes from *A. cristata* and the longest chromosomes from the other genome donor. To examine this hypothesis, GISH was carried out on chromosome preparations of *A. afraspera* using biotin-labeled genomic DNA from *A. cristata* and digoxigenin-labeled genomic DNA from *A. nilotica*. The hybridized chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) and analyzed for the hybridization signals (Fig. 5a). All the chromosomes were labeled in red (biotin) but with a more marked signal on small chromosomes (Fig. 5b). Conversely, the green labeling (digoxigenin) was restricted to the longest chromosomes and the hybridization was uniform along the chromosome arms (Fig. 5c). Superposed exposures of the metaphase clearly showed the presence of two differentiated groups of chromosomes based on both their size and the hybridization signals, with the set of small chromosomes presenting a high affinity for the *A. cristata* genome probe and the set of the longest chromosomes for the *A. nilotica* genome probe (Fig. 5d).

Aeschynomene schimperi (8x) is also supposed to derive from hybridization between *A. schimperi* (4x) and *A. schimperi* 'grandiflora' (4x). As these putative progenitors present a 20% difference in their genome size (4.82 and 3.93 pg/2C, respectively) (Table 1), a similar cytogenetic approach was performed. Metaphase chromosomes of *A. schimperi* (4x) appeared relatively longer than those of *A. schimperi* 'grandiflora', but most strikingly they presented an atypical plump morphotype and well-developed telomeres (Fig. 4). Metaphase spreads were observed for two *A. schimperi* accessions that present distinct genome sizes. Both *A. schimperi*-126 (8x) (8.49 ± 0.11 pg/2C) and

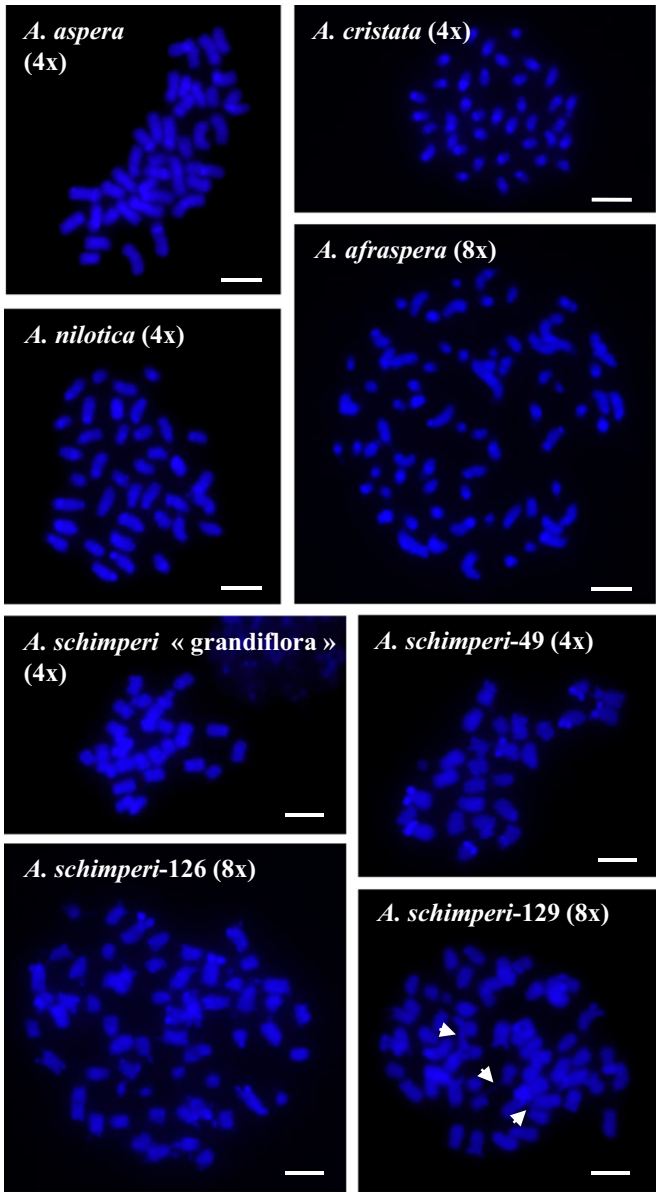

Fig. 4 Karyotypes features of the 8x *Aeschynomene* and related 4x species. Root tip metaphase chromosomes stained with 4′,6-diamidino-2-phenylindole (blue) of *A. aspera*-47, *A. nilotica*-53, *A. cristata*-169, *A. afraspera*-1, *A. schimperi* 'grandiflora'-123, *A. schimperi*-49, *A. schimperi*-126 and *A. schimperi*-129. Arrowheads indicate interchromosomal connections. Numbers following the species names refer to the accession code figuring in Table 1. Bars, 5 μm.

A. schimperi-129 (8x) (9.20 ± 0.12 pg/2C) revealed the two types of chromosomes (Fig. 4). An additional phenotype characterized many chromosomes of *A. schimperi*-129 by the presence of interchromosomal connectives (Fig. 4). To determine the origin of the two chromosome types in the 8x cytotype, GISH was performed on mitotic chromosomes of *A. schimperi*-126 (Fig. 5e) using digoxigenin labeling for the DNA from *A. schimperi* 'grandiflora' (4x) (detected in red) and biotin labeling for the DNA from *A. schimperi* (4x) (detected in green). Analysis of the hybridization signals on DAPI-counterstained chromosomes showed uniform labeling patterns on all the chromosomes but

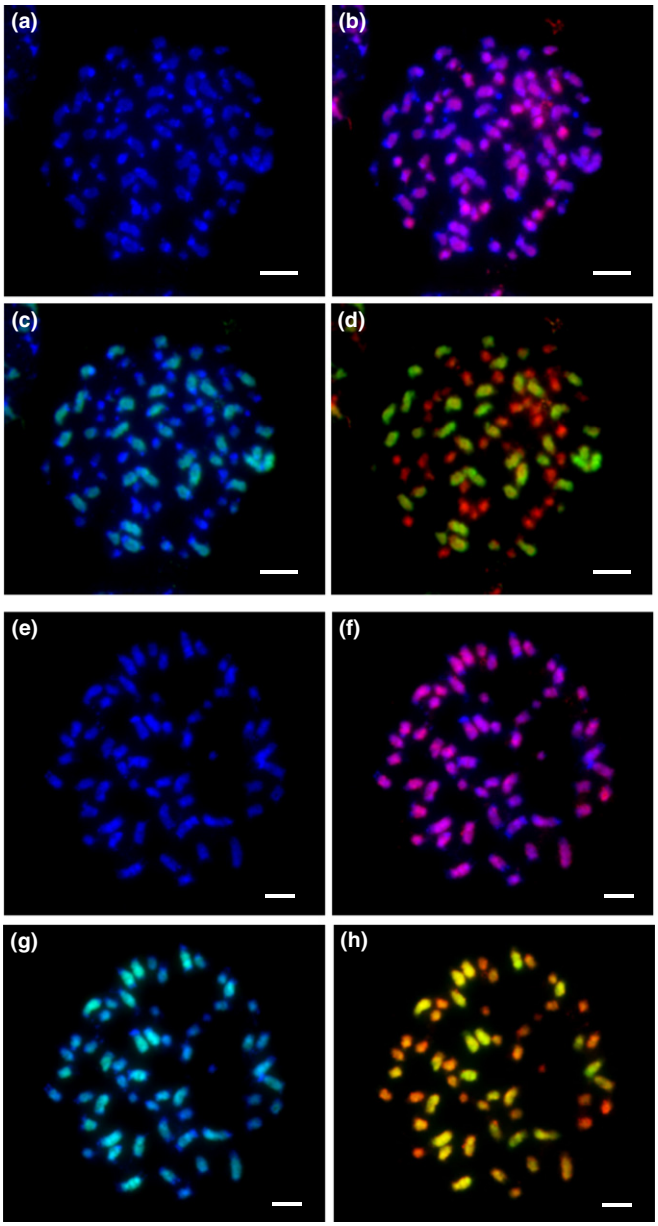

Fig. 5 Genomic *in situ* hybridization (GISH) analysis of the hybrid origin of *Aeschynomene afraspera* and *Aeschynomene schimperi* 8x. (a–d) GISH on mitotic chromosomes of *A. afraspera*-1 using *A. cristata*-169 and *A. nilotica*-47 as putative genome donors: (a) chromosomes stained with 4′,6-diamidino-2-phenylindole (DAPI) (blue); (b) GISH signal of digoxigenin-labeled genomic DNA from *A. cristata*-169 (detected in red) on DAPI-stained chromosomes; (c) GISH signal of biotin-labeled genomic DNA from *A. nilotica* (detected in green) on DAPI-stained chromosomes; (d) merged red and green GISH signals from both putative progenitors' genomic DNAs. (e–h) GISH on mitotic chromosomes of *A. schimperi*-126 (8x) using *A. schimperi*-49 (4x) and *A. schimperi* 'grandiflora'-123 (4x) as putative genome donors: (e) chromosomes stained with DAPI (blue); (f) GISH signal of digoxigenin-labeled genomic DNA from *A. schimperi* 'grandiflora'-123 (detected in red) on DAPI-stained chromosomes; (g) GISH signal of biotin-labeled genomic DNA from *A. schimperi*-49 (detected in green) on DAPI-stained chromosomes; (h) merged red and green GISH signals from both putative progenitors' genomic DNAs. Numbers following the species names refer to the accession code shown in Table 1. Bars, 5 μm.

with variation in red and green fluorescence intensities. Overlay of the two hybridization signals allowed us to distinguish two sets of chromosomes, with the smaller ones preferentially labeled in red and the longer ones –with a plump shape and developed telomeres – preferentially labeled in green, hence showing differential affinity for the *A. schimperi* (4x) and *A. schimperi* 'grandiflora' genome probes (Fig. 5f–h).

As a whole, these cytogenetic analyses confirm that both *A. afraspera* and *A. schimperi* (8x) are allopolyploid species and provide further evidence that they have conserved the contrasting chromosomal characteristics of their respective putative progenitors.

Characterization of the breeding systems and the stem nodulation in the African group

We investigated the different breeding systems encountered in the African group. First sorting of the species came from their ability to produce seeds in our pollinator-free glasshouse. During the 2 yr of observation, all species, except for *A. elaphroxylon*, readily bloomed. Most of them were able to develop pods abundantly, indicating that they correspond to selfing species. This class contains both 4x species – *A. aspera, A. crassicaulis*, *A. nilotica*, *A. uniflora* – and the two 8x species – *A. afraspera* and *A. schimperi* (Fig. 6 and data not shown). The second class corresponds to species that flowered but without seed production. It includes only 4x species: *A. cristata*, *A. fluitans*, *A. pfundii*, *A. schimperi* (4x) and *A. schimperi* 'grandiflora' (Fig. 6). These recalcitrant species displayed flowers that are composed of a long stigma emerging from the corolla and short anthers, and some species also developed flowers with a very short stigma and long anthers emerging outside the corolla Fig. 6). In both cases, self-pollination is mechanically impossible. Therefore, such species can be considered as outcrossing and so their reproduction probably implies an insect-mediated pollination. Among these outcrossing species, *A. cristata* and *A. pfundii* also have in common the fact that the upper parts of stems are conspicuously covered with glandular trichomes, forming a sticky coating that may provide protection against pollen theft by undesired nonflying insects (Fig. 6).

Previous studies recorded the *Aeschynomene* species that are endowed with stem nodulation (Alazard & Duhoux, 1987, 1988; Chaintreuil *et al.*, 2013; Devi, 2013a). Here, we took advantage of the fact that some *Aeschynomene* species spontaneously form stem nodules under our glasshouse conditions because of the presence of a highly compatible bradyrhizobial population in the atmosphere to observe stem nodule formation in the African species. Five of them readily and repeatedly developed aerial stem nodules in a profuse fashion: *A. aspera*, *A. afraspera*, *A. nilotica*, *A. cristata* and *A. uniflora* (Fig. 7a). In all cases, the stem nodules appear to develop around the adventitious root as evidenced by the presence of the root tip at the top of the nodules (insets in Fig. 7a). Two other species were also conspicuous for their floating stems with nodules: *A. crassicaulis* and *A. fluitans* (Fig. 7b). In these latter cases, root primordia could develop into adventitious roots. The well-developed stem nodules of *A. fluitans* were used to analyze their structure (Fig. 7c). Cross-sections revealed a green cortical parenchyma surrounding an infected tissue. This latter feature was composed of two distinct areas on both sides of the adventitious root vascular traces to which the nodular vasculature was connected.

The two phenotypic traits were mapped on a simplified tree representing the phylogenetic relationships between 4x species and the inferred patterns of hybridization for the 8x species (Fig. 8). For the modes of reproduction, selfing and outcrossing are evenly distributed among 4x species, suggesting that the breeding systems are versatile and that parallel evolutions have occurred. Strikingly, the two 8x species, *A. afraspera* and *A. schimperi*, are autogamous, while their putative genome donors or related 4x species appear to be outcrossing (Fig. 8). This strongly suggests that allogamy has promoted interspecific hybridizations, while transition to selfing in the 8x progenies may have favored allopolyploid survival. Regarding stem nodulation, nodules developing on aerial stems or on floating stems were observed in two unrelated lineages in each case (Fig. 8). This suggested two independent emergences for these nodulation traits. Furthermore, this works also reports *A. cristata* as a new stem-nodulating *Aeschynomene* species, while superimposition of the stem-nodulation trait upon the phylogenetic and genomic relationships indicates that the hybrid *A. afraspera* has inherited its stem-nodulation ability from its two parental lineages.

Discussion

By using an integrated molecular and cytogenetic approach, the origin and evolution of the African *Aeschynomene* have been unraveled. We discuss the main features of the dynamics of this group and how this knowledge can be exploited to better understand the evolution of the nitrogen-fixing symbiosis in *Aeschynomene* species.

Genome evolution through paleopolyploidy and neopolyploidy

Chromosome number counting and single-copy nuclear gene phylogenies revealed that the African *Aeschynomene* group has encountered two rounds of polyploidization.

The ancient polyploid event, referred as paleopolyploidy, occurred before the group diversified as both A and B homeologs of the investigated genes were found in all the African *Aeschynomene* species. These findings support a single tetraploid origin with a common AB genome for the whole group. The use of a simple molecular clock assumption provided a rough time frame that places the A and B genome divergence at 18.6 Ma. Ancient polyploidization events are involved early in the formation of different papilionoid genera, notably *Lupinus* and *Glycine* (Shoemaker *et al.*, 2006; Krock *et al.*, 2014). Molecular dating in soybean indicated the maximum age of the *Glycine* paleopolyploid event to be 10–13 million yr (Egan & Doyle, 2010; Schmutz *et al.*, 2010). But the nature of this ancient polyploid event remains unclear as allopolyploidy was inferred from cytogenetic

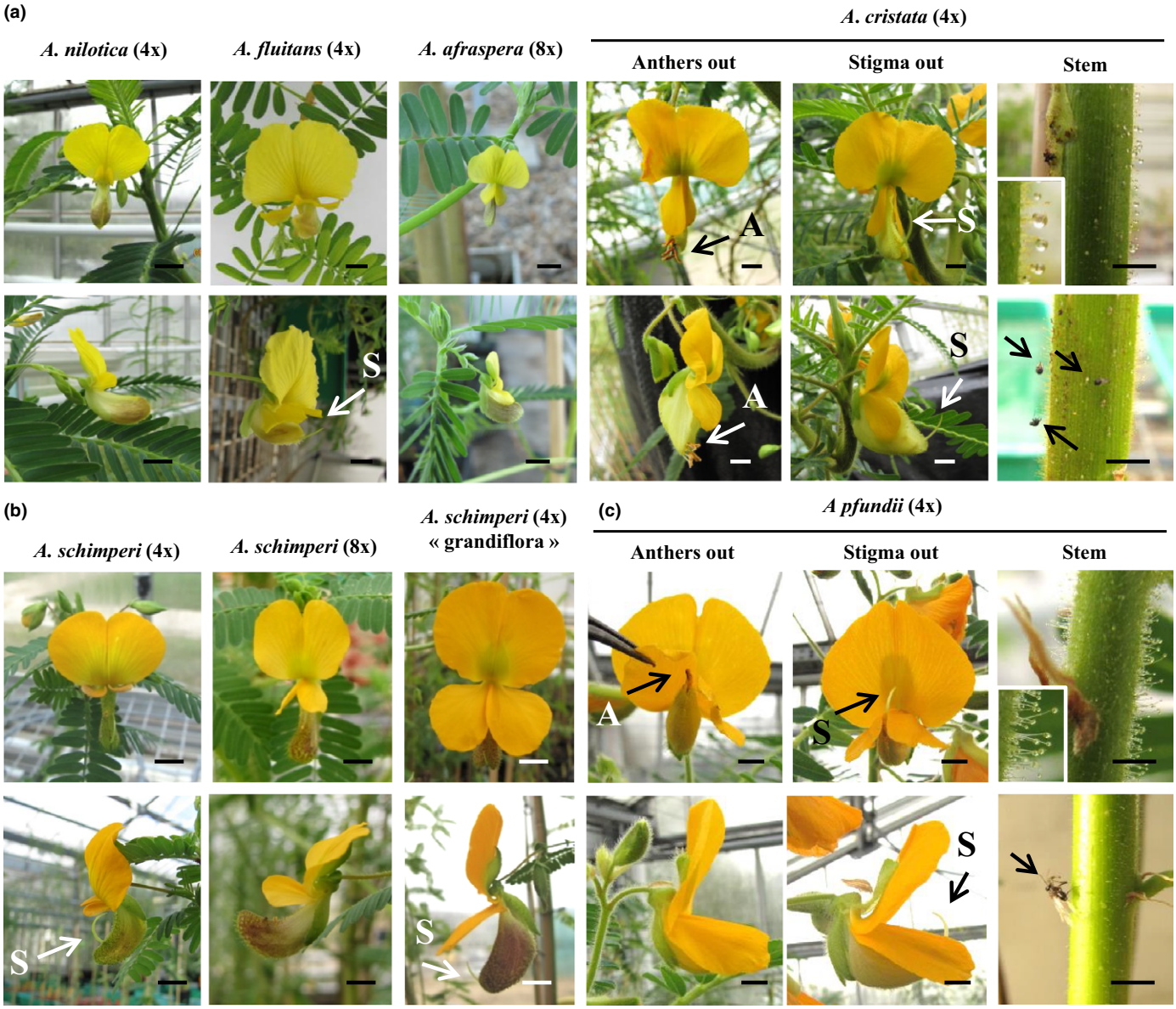

Fig. 6 Floral morphology and flower protection in *Aeschynomene* species. (a) *Aeschynomene* species of clade 1 containing *A. afraspera*; (b) *Aeschynomene* species of clade 2 containing *A. schimperi*; (c) *A. pfundii*. Upper and lower panels show front and lateral views of the flowers, respectively, with the sexual organs, style (S) and anthers (A) indicated when emerging from the corolla. Right-hand side panels show stems covered with glandular trichomes (zoom in the inset) and trapped insects (arrows). Bars, 5 mm.

analyses, while patterns of genome fractionation suggested autopoloidy (Gill *et al.*, 2009; Garsmeur *et al.*, 2013). Phylogenetic analysis of duplicated genes was also inconclusive in resolving its origin because of the absence of the diploid genome donors that are likely to be extinct (Straub *et al.*, 2005). For the African *Aeschynomene* group, the same problem is encountered with the present phylogenetic approach, which does not enable us to discriminate between autopolyploidy and allopolyploidy regarding the paleopolyploid event. Interestingly, the genus *Aeschynomene* was recently shown to be paraphyletic with the genera *Soemmeringia*, *Cyclocarpa*, *Smithia*, *Kotschya*, *Humularia*, *Geissaspis*, *Bryaspis* and also xerophytic *Aeschynomene* species being embedded between the Nod-independent *Aeschynomene* clade (2x) and the African *Aeschynomene* clade (4x) (Fig. 1) (Lavin *et al.*, 2001; Chaintreuil *et al.*, 2013). Therefore, investigating whether these associated genera share the same AB genome or contain the potential A and B genome donors should shed light on the origin of the paleopolyploid event.

Phylogenetic analyses further pointed to two more recent polyploid events, corresponding to neopolyploidy, for *A. schimperi* (8x) and *A. afraspera* (8x), the latter occurring < 5.5 Ma, and they also allowed the identification of potential tetraploid progenitors. In addition, for *A. afraspera*, both cross-amplification tests with *Aa*SSR markers and the *ITS*/matK discordance corroborated *A. cristata* as one genome donor and indicated that the other genome donor, although not identified, was closely related to both *A. nilotica* and *A. fluitans*. Whether this second genome donor corresponds to their ancestor or to a different species,

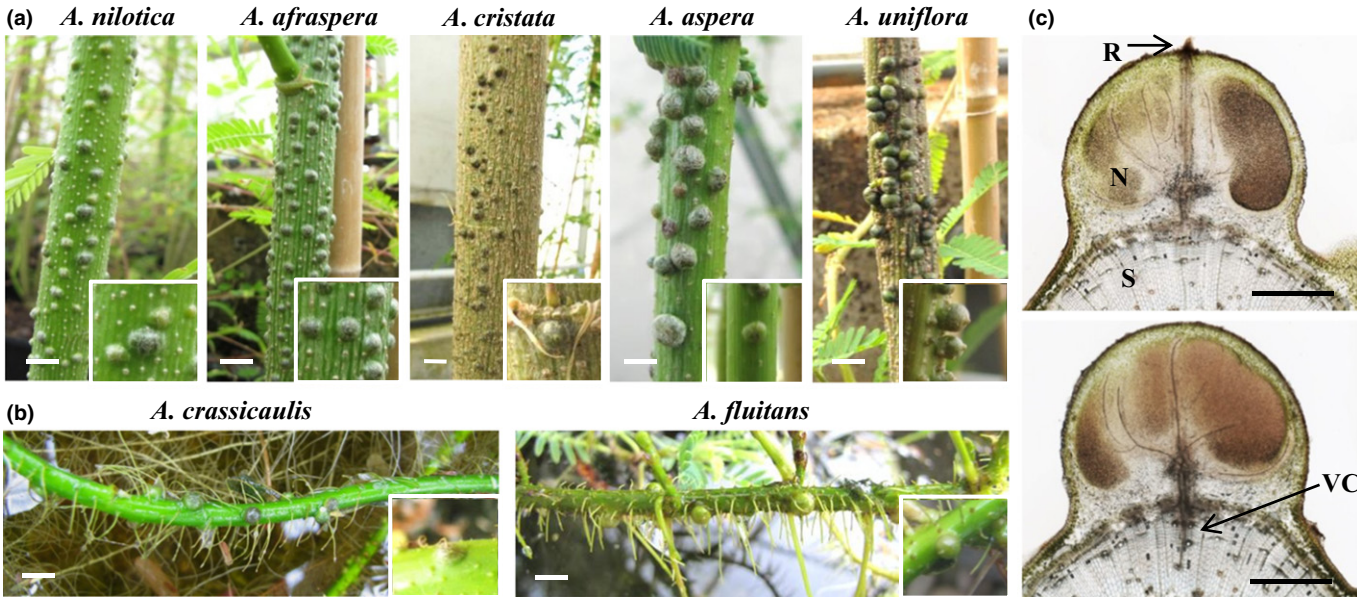

Fig. 7 Stem nodulation in *Aeschynomene* species. (a) Nodules developed on aerial stems of 2-month-old *Aeschynomene* spp. Small spots on the stem correspond to dormant root primordia. (b) Nodules developed on floating stems of 2-month-old *Aeschynomene* spp. Note that in this case root primordia have developed into adventitious roots. Insets show a close-up of stem nodules centered on an adventitious root emergence. Bars, 5 mm. (c) Serial cross-sections of a stem nodule of *Aeschynomene fluitans*. Root (R), nodule (N), stem (S) and vascular connection (VC) are indicated. Note the nodular vasculature connecting to the root vascular bundles (upper panel), themselves entering the stem tissues (lower panel). Bars, 100 μm.

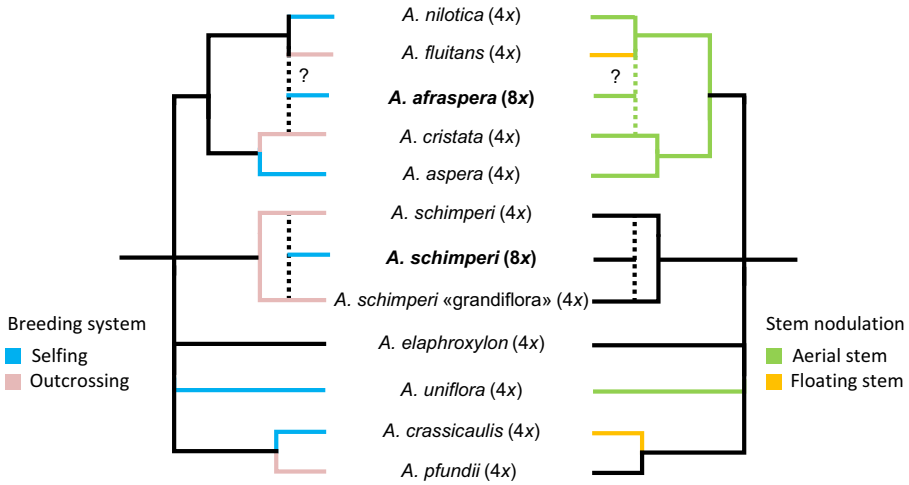

Fig. 8 Evolution of selected characters in the African semiaquatic *Aeschynomene*. Breeding system (left) and stem nodulation (right) mapped on the simplified tree representing the phylogenetic relationships between the 4x species and the inferred patterns of allopolyploid origins for the 8x species. Octoploid species are highlighted in bold and are connected by dashed lines to their putative genome donors or to the most closely related tetraploid species when not conclusively identified (?).

remains to be determined. In the case of *A. schimperi*, this study uncovered an important diversity at both the genetic and genomic levels, suggesting that it actually corresponds to a large allopolyploid species complex. The phylogeny indicated that *A. schimperi* (4x) and *A. schimperi* 'grandiflora' were the potential progenitors of *A. schimperi* (8x). However, octoploid *A. schimperi* accessions show variations in their *ITS* and *matK* sequences and in their genome size. Therefore, two plausible explanations can be proposed: the different accessions arose either via separate allopolyploid events involving the same parents or related genotypes or from a single allopolyploidization event followed by divergent evolution. In addition, as some *A. schimperi* (8x) accessions and *A. schimperi* 'grandiflora' share the same *matK* sequence, we cannot exclude the possibility that an introgression has occurred. Such problems are recurrent in polyploid formations and have also been described in the legume genera *Arachis* (Moretzsohn *et al.*, 2004), *Glycine* (Doyle *et al.*, 2004) and *Leucaena* (Govindarajulu *et al.*, 2011). Further analyses are necessary to clarify the relationships and differences betweeen the different accessions referred to as *A. schimperi*.

Genome reshaping by descendent dysploidy and genome downsizing

Chromosome number counting confirmed or re-evaluated some previous reports for *A. afraspera*, *A. aspera*, *A. elaphroxylon*

(referenced in the IPCN), *A. schimperi* and *A. uniflora* (Bielig, 1997). The basic chromosome number is $2n = 40$, confirming the tetraploid status of this group compared with the Nod-independent clade which is mainly $2n = 20$ (Arrighi *et al.*, 2014). Deviations from the $2n = 40$ were observed in the two lineages containing *A. afraspera* and *A. schimperi* with $2n = 38$ or 76 and $2n = 28$ or 56, respectively. The mechanisms responsible for the diminution of chromosome number could be aneuploidy (chromosome loss) or dysploidy (chromosome rearrangements). As the genome sizes do not seem to be impacted, dysploidy events are more likely. The decrease in chromosome number may be a consequence of chromosome fusions and rearrangements (Schubert & Lysak, 2011). It should be noted that, among the dysploid species, at least *A. schimperi*-129 (8x) is distinguished by numerous interchromosomal connectives linking different mitotic chromosomes together. Similar phenomena have been observed in other organisms, notably in mitotic and meiotic cells of *Ornithogalum virens* (Godin & Stack, 1976; Ashley, 1979). Although end-to-end connections are supposed to facilitate pairing of homologous chromosomes during meiosis, their significance during mitosis remains poorly understood.

The African clade had genome sizes that were more than twice the mean size of diploid genomes observed in the Nod-independent *Aeschynomene* clade (Arrighi *et al.*, 2014). This can be explained partly by the changes of ploidy level, but it also indicates a general larger genome that must be linked with a differential evolution of transposable elements, a phenomenon now accepted as the main factor of genome size increase and decrease besides polyploidy in plants (*Oxalis*, Vaio *et al.*, 2013; *Oryza*, Piegu *et al.*, 2006). Although the transposable elements have not yet been characterized in the *Aeschynomene* species of the African group, their abundant presence is suggested by the GISH experiments that revealed a uniform chromosome labeling contrarily to the Nod-independent *A. indica* for which labeling was previously shown to be restricted to the centromeres (Arrighi *et al.*, 2014). Considering that GISH is essentially based on the detection of repeated sequences (D'Hont, 2005), the labeling pattern observed is indicative of the different genome composition of these two *Aeschynomene* groups. In the African group, *A. cristata* presents a discordant genome size corresponding to only 46% of the most related genomes. This suggests that a dramatic and recent genome downsizing has occurred in this species and may correspond to a massive removal of transposons. Such a process is often interpreted as an adaption to sharp environmental changes and constraints. Interestingly, it was recently shown in *Primulina* that genome size was strongly positively correlated to plant N content, leading to the proposition that N limitation may be a driver of genome size variation (Kang *et al.*, 2015). The specific mechanism of genome downsizing in *A. cristata* remains unknown, although it is endowed with nitrogen-fixing nodulation. At a cytological level, the very visible consequence is a twofold size decrease of the mitotic chromosomes compared with related species. This also accounts for the GISH patterns observed for the derived hybrid *A. afraspera*, with the smaller chromosomes inherited from *A. cristata* and the longest ones from the second genome donor. This situation is reminiscent of the bimodal karyotype of *Milium montianum*, which contains two distinct size classes of chromosomes donated by each parent (Bennett *et al.*, 1992).

Comparative evolution of the African and Nod-independent *Aeschynomene* groups

The Nod-independent and the African *Aeschynomene* groups are characterized by distinct origins, American vs African, and by different genome structures, diploid vs tetraploid genomes, which are indicative of independent evolutionary histories. This idea is supported by the recent evidence that these two groups are separated by several embedded genera in the paraphyletic genus *Aeschynomene* (Fig. 1) (Chaintreuil *et al.*, 2013). In addition, the Nod-independent species differ from other *Aeschynomene* species, including those of the African group, in their symbiotic program in two main ways: at the early stage of the interaction, the nodulation process is triggered in the absence of Nod factor recognition, and in mature nodules, the differentiation in bacteroids is distinct (Giraud *et al.*, 2007; Czernic *et al.*, 2015). For comparative analysis of the symbiotic processes, *A. afraspera* is the Nod-dependent species that has been the most studied (Bonaldi *et al.*, 2011; Okazaki *et al.*, 2016). But the polyploid past of the African *Aeschynomene* species will make their genetic analysis difficult. Instead, it would be advisable to use diploid species. Characterizing the genome of American Nod-dependent *Aeschynomene* species, such as *A. americana* which is known to be 2x, along with those of the associated genera will allow the identification of the species that have a simple genome structure (Fig. 1). Determination of such diploid species will be an important prerequisite to performing genetic and comparative phylogenomic analyses with the aim of identifying the determinants of the Nod-dependent and Nod-independent processes.

Despite their independent evolution histories, the Nod-independent and the African *Aeschynomene* groups have in common the same semiaquatic lifestyle. As a probable adaptation to temporary flooding, stem nodulation occurs in both clades. Aerial stem nodules were first described in *A. afraspera* (but initially referred to as *A. aspera*) (Hagerup, 1928; Alazard & Duhoux, 1987) and subsequent functional analyses of the photosynthetic activity of Bradyhrizobia in the development of the stem nodules were performed in *A. sensitiva* (Giraud *et al.*, 2000). Like most of the other Nod-independent *Aeschynomene*, stem nodulation in *A. sensitiva* is sparse, whereas it is profuse in some African *Aeschynomene* species. This difference was explained by Alazard & Duhoux (1988) who observed that in the Nod-independent group dormant root primordia usually just underlie the stem epidermis and so are less prone to be infected by *Bradyrhizobium*. By contrast, those of the African stem-nodulating *Aeschynomene* break the stem epidermis, facilitating the access of the *Bradyrhizobium* to the base of the root primordium. This corresponds to the infection site and it is also important for the subsequent nodule organogenesis in that it provides vascular connection. Genetic analysis in both *Aeschynomene* groups is now necessary to uncover the genetic determinism of stem nodulation. The diploid nature of the new model species

A. evenia should facilitate such a genetic analysis (Arrighi & Cartieaux, 2015), but the African semiaquatic *Aeschynomene* will provide the opportunity to investigate whether paleopolyploidy might have provided genetic material leading or contributing to the evolution of stem nodulation, which is profuse in some species of this group.

Acknowledgements

We thank the different gene banks, AusPGRIS (Australia), CIAT (Columbia), ILRI (Ethiopia) and IRRI (Philipines), for providing seeds; R. Razakamalala (Missouri Botanical Garden, Madagascar), B. R. Harinisainana and A. T. E. Razakatiana (CNRE, Madagascar) for assistance with field work; and N. Nouwen (LSTM, IRD, France), P. Czernic (LSTM, UM, France) and B. Dreyfus (IRD, France) for their critical reading of the manuscript. The present work has benefited from the facilities and expertise of the cytometry facilities of Imagerie-Gif (http://www.i2bc.paris-saclay.fr/spip.php?article279) and of the molecular cytogenetic facilities of the AGAP laboratory (http://umr-agap.cirad.fr/en/plateformes/plateau-de-cytogenetique-moleculaire). This work was supported by grants from the French National Research Agency (ANR-BugsInaCell-13-BSV7-0013) and the Franco-Thai PHC Siam Program (project SIAM no. 29589XA).

Author contributions

J.F.A. designed the experiments. C.C., D.G., C.H., S.C.B., M. Bourge, F.C., M. Boursot and J.F.A. performed the experiments. J.F.A., A.D'H. and E.G. analyzed the data. G.P.L. carried out critical species identification. P.T., H. Randriambanona, H. Ramanankierna and N.T. carried out data collection. J.F.A. and E.G. wrote the paper.

References

Alazard D. 1985. Stem and root nodulation in *Aeschynomene* spp. *Applied and Environmental Microbiology* **50**: 732–734.

Alazard D, Becker M. 1987. *Aeschynomene* as green manure for rice. *Plant and Soil* **101**: 141–143.

Alazard D, Duhoux E. 1987. Nitrogen-fixing stem nodules on *Aeschynomene afraspera*. *Biological Fertility Soils* **4**: 61–66.

Alazard D, Duhoux E. 1988. Diversity of stem nodulation sites in *Aeschynomene* spp. *Journal of Plant Physiology* **132**: 123–125.

Arrighi JF, Cartieaux F. 2015. Out of water of a new model legume: the Nod-independent *Aeschynomene evenia*. In: de Bruijn F, ed. *Biological nitrogen fixation, vol. 2, chap. 45*. Oxford, UK: Wiley-Blackwell, 447–454.

Arrighi JF, Cartieaux F, Brown SC, Rodier-Goud M, Boursot M, Fardoux J, Patrel D, Gully D, Fabre S, Chaintreuil C *et al.* 2012. *Aeschynomene evenia*, a model plant for studying the molecular genetics of the Nod-independent rhizobium–legume symbiosis. *Molecular Plant–Microbe Interactions* **25**: 851–861.

Arrighi JF, Cartieaux F, Chaintreuil C, Brown SC, Boursot M, Giraud E. 2013. Genotype delimitation in the Nod-independent model legume *Aeschynomene evenia*. *PLoS ONE* **8**: e63836.

Arrighi JF, Chaintreuil C, Cartieaux F, Cardi C, Rodier-Goud M, Brown SC, Boursot M, d'Hont A, Dreyfus B, Giraud E. 2014. Radiation of the Nod-independent *Aeschynomene* relies on multiple allopolyploid speciation events. *New Phytologist* **201**: 1457–1468.

Ashley T. 1979. Specific end-to-end attachment of chromosomes in *Ornithogalum virens*. *Journal of Cell Science* **38**: 357–367.

Bennett ST, Kenton AY, Bennett MD. 1992. Genomic *in situ* hybridization reveals the allopolyplloid nature of *Milium montianum* (Gramineae). *Chromosoma* **101**: 420–424.

Bielig LM. 1997. Chromosome numbers in the forage legume genus *Aeschynomene* L. *SABRAO Journal* **29**: 33–39.

Boivin C, Ndoye I, Molouba F, De Lajudie P, Dupuy N, Dreyfus B. 1997. Stem nodulation in legumes: diversity, mechanisms, and unusual characteristics. *Critical Reviews in Plant Sciences* **16**: 1–30.

Bonaldi K, Gargani D, Prin Y, Fardoux J, Gully D, Nouwen N, Goormachtig S, Giraud E. 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by photosynthetic *Bradyrhizobium* sp. strain ORS285: the Nod-dependent versus the Nod-independent symbiotic interaction. *Molecular Plant–Microbe Interactions* **24**: 1359–1371.

Chaintreuil C, Arrighi JF, Giraud E, Miche L, Moulin L, Dreyfus B, Munive-Hernadez J, Villegas-Hernadez M, Béna G. 2013. Evolution of symbiosis in the legume genus *Aeschynomene*. *New Phytologist* **200**: 1247–1259.

Czernic P, Gully D, Cartieaux F, Guefrachi I, Patrel D, Pierre O, Fardoux J, Chaintreuil C, Nguyen P, Gressent F *et al.* 2015. Convergent evolution of endosymbiont differentiation in Dalbergioid and IRLC legumes mediated by nodule-specific cysteine-rich peptides. *Plant Physiology* **169**: 1254–1265.

Devi WR. 2013a. *Aeschynomene aspera* L., a nitrogen fixing stem nodulation plant from Manipur. *International Journal of Scientific & Technology Research* **2**: 232–235.

Devi WR. 2013b. *Aeschynomene aspera* L., a potential stem nodulated plant as green manure for rice cultivation in Manipur. *Journal of Medicinal Plants* **1**: 28–31.

D'Hont A. 2005. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenetic and Genome Research* **109**: 27–33.

Doyle JJ, Doyle JL, Rausher JT, Brown AHD. 2004. Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): a study of contrasts. *Biological Journal of the Linnean Society* **82**: 583–597.

Egan A, Doyle JJ. 2010. A comparison of global, gene-specific, and relaxed clock methods in a comparative genomics framework: dating the polyploid history of soybean (*Glycine max*). *Systematic Biology* **59**: 534–547.

Evans WR, Fleischman DE, Calvert HE, Pyati PV, Alter GM, Rao NS. 1990. Bacteriochlorophyll and photosynthetic reaction centers in *Rhizobium* strain BTAi 1. *Applied and Environment Microbiology* **56**: 3445–3449.

Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. 2013. Two evolutionarily distinct classes of paleopolyploidy. *Molecular Biology and Evolution* **31**: 448–454.

Gill N, Findley S, Wallig JG, Hans C, Ma J, Doyle J, Stacey G, Jackson SA. 2009. Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiology* **151**: 1167–1174.

Giraud E, Fleischman D. 2004. Nitrogen-fixing symbiosis between photosynthetic bacteria and legumes. *Photosynthesis Research* **82**: 115–130.

Giraud E, Hannibal L, Fardoux J, Vermeglio A, Dreyfus B. 2000. Effect of *Bradyrhizobium* photosynthesis on stem nodulation of *Aeschynomene sensitiva*. *Proceedings of the National Academy of Sciences, USA* **97**: 14795–14800.

Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y *et al.* 2007. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science* **316**: 1307–1312.

Godin DE, Stack SM. 1976. Homologous and non-homologous chromosome associations by interchromosomal chromatin connectives in *Ornothogalum virens*. *Chromosoma* **57**: 309–318.

Govindarajulu R, Hughes CE, Alexander PJ, Bailey CD. 2011. The complex evolutionary dynamics of ancient and recent polyploidy in *Leucaena* (Leguminosae; Mimosoideae). *American Journal of Botany* **98**: 2064–2076.

Hagerup O. 1928. En hygrofil baelgplante (*Aeschynomene aspera* L.) med bakterieknolde paa staenglen. *Dansk Botanisk Arkiv* **14**: 1–9.

James EK, Loureiro FM, Pott A, Pott VJ, Martins CM, Franco AA, Sprent JI. 2001. Flooding-tolerant legume symbioses from the Brazilian Pantanal. *New Phytologist* **150**: 723–738.

Kang M, Wang J, Huang H. 2015. Nitrogen limitation as a driver of genome size evolution in a group of karst plants. *Scientific Reports* **5**: 11636.

Krock M, Koczyk G, Swiecicki W, Kilian A, Nelson MN. 2014. New evidence of ancestral polyploidy in the Genistoid legume *Lupinus angustifolius* L.

(narrow-leafed lupin). *TAG. Theoretical and Applied Genetics* **127**: 1237–1249.

Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* **54**: 575–594.

Lavin M, Pennington RT, Klitgaard BB, Sprent JI, de Lima HC, Gasson PE. 2001. The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *American Journal of Botany* **88**: 503–533.

Léonard J. 1954. Notulae Systematicae XV. Papilionaceae: Hedysareae Africanae (*Aeschynomene, Alysicarpus, Ormocarpum*). *Bulletin du Jardin Botanique de l'Etat à Bruxelles* **24**: 63–106.

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

Loureiro MF, Defaria SM, James EK, Pott A, Franco AA. 1994. Nitrogen-fixing stem nodules of the legume, *Discolobium pulchellum* Benth. *New Phytologist* **128**: 283–295.

Loureiro MF, James EK, Sprent JI, Franco AA. 1995. Stem and root nodules on the tropical wetland legume *Aeschynomene fluminensis*. *New Phytologist* **130**: 531–544.

Moretzsohn MC, Gouvea EG, Inlglis PW, Leal-Bertioli SCM, Valls JFM, Bertioli DJ. 2013. A study of relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellites markers. *Annals of Botany* **111**: 113–126.

Moretzsohn C, Hopkins MS, Mitchell SE, Kresovich S, Valls JF, Ferreira ME. 2004. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biology* **4**: 11.

Nicholas KB, Nicholas HB Jr 1997. *GeneDoc: a tool for editing and annotating multiple sequence alignments*. Distributed by the authors. URL http://www.psc.edu/biomed/genedoc [accessed 3 September 2015].

Okazaki S, Tittabur P, Teulet A, Thouin J, Farcdoux J, Chaintreuil C, Gully D, Arrighi JF, Furuta N, Miwa H *et al.* 2016. Rhizobium-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *ISME Journal* **10**: 64–74.

Peter. 1928. Abhandl. Ges. Wiss. Göttingen, Math.-Phys. *Klasse, Neue Folge* **13**: 82.

Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar D, Jackson S, Wing RA *et al.* 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Research* **16**: 1262–1269.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J *et al.* 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **463**: 178–183.

Schubert I, Lysak MA. 2011. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends in Genetics* **27**: 207–216.

Shoemaker RC, Schlueter J, Doyle JJ. 2006. Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinion in Plant Biology* **9**: 104–109.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, de Pamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**: 336–348.

Sprent JI, James EK. 2008. Legume–rhizobial symbiosis: an anorexic model? *New Phytologist* **179**: 3–5.

Straub SCK, Pfeil BE, Doyle JJ. 2005. Testing the polyploid past of soybean using a low-copy nuclear gene – is *Glycine* (Fabaceae: Papilionoideae) an auto- or allopolyploid? *Molecular Phylogenetics and Evolution* **39**: 580–584.

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* **10**: 2731–2739.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876–4882.

Vaio M, Gardner A, Emshwiller E, Guerra M. 2013. Molecular phylogeny and chromosome evolution among the creeping herbaceous *Oxalis* species of sections *Corniculatae* and *Ripariae* (Oxalidaceae). *Molecular Phylogenetics and Evolution* **68**: 199–211.

Verdcourt B. 1971. *Aeschynomene*. In: Gillet JB, Polhill RM, Verdcourt B, eds. *Flora of Tropical East Africa, Leguminosae, Papilionoideae*. Kew, UK: Royal Botanic Gardens.

Supporting Information

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

Fig. S1 Somatic metaphases in different *Aeschynomene* species.

Fig. S2 Alignment of the coding sequence (partial or complete) of single-copy nuclear genes.

Table S1 Genes and SSR markers used in this study

Table S2 GenBank numbers for the sequences used in the phylogenetic analyses

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.