

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

A phylogenetic framework of the legume genus *Aeschynomene* for comparative genetic analysis of the Nod-dependent and Nod-independent symbioses

Laurent Brottier^{*1}, Clémence Chaintreuil^{*1}, Paul Simion², Céline Scornavacca², Ronan Rivallan^{3,4}, Pierre Mournet^{3,4}, Lionel Moulin⁵, Gwilym P. Lewis⁶, Joël Fardoux¹, Spencer C. Brown⁷, Mario Gomez-Pacheco⁷, Mickaël Bourges⁷, Catherine Hervouet^{3,4}, Mathieu Gueye⁸, Robin Duponnois¹, Heriniaina Ramanankierana⁹, Herizo Randriambanona⁹, Hervé Vandrot¹⁰, Maria Zabaleta¹¹, Maitrayee DasGupta¹², Angélique D'Hont^{3,4}, Eric Giraud¹ and Jean-François Arrighi^{1}**

¹IRD, Laboratoire des Symbioses Tropicales et Méditerranéennes, UMR LSTM, Campus International de Baillarguet, 34398 Montpellier, France, ²Institut des Sciences de l'Evolution (ISE-M), Université de Montpellier, CNRS, IRD, EPHE, 34095 Montpellier Cedex 5, France, ³CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement), UMR AGAP, F-34398 Montpellier, France, ⁴AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, 34060 Montpellier, France ⁵IRD, Interactions Plantes Microorganismes Environnement, UMR IPME, 34394 Montpellier, France, ⁶Comparative Plant and Fungal Biology Department, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, United Kingdom, ⁷Institute of Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette, France, ⁸Laboratoire de Botanique, Institut Fondamental d'Afrique Noire, Ch. A. Diop, BP 206 Dakar, Sénégal, ⁹Laboratoire de Microbiologie de l'Environnement/Centre National de Recherche sur l'Environnement, Antananarivo 101, Madagascar, ¹⁰IAC, Laboratoire de Botanique et d'Ecologie Végétale Appliquée, UMR AMAP, 98825 Pouembout, Nouvelle-Calédonie, ¹¹Department of Biochemistry and Microbial Genomics. IIBCE. Montevideo 11600. Uruguay, ¹²Department of Biochemistry, University of Calcutta, Kolkata, 700019, India

*: contributed equally to this work

**: Author for correspondence

Jean-François Arrighi

35 Tel: +33 (0) 4 67 59 38 82

36 Fax: +33 (0) 4 67 59 38 02

37 E-mail: jean-francois.arrighi@ird.fr

38

39

40 Total word count: 6053

41 Introduction: 823

42 Materials and Methods: 1083

43 Results: 2601

44 Discussion: 1470

45 Acknowledgements: 76

46 Number of figures: 5 figures in color

47 Number of tables: 0

48 Supporting information: 12 figures and 4 tables

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69 **SUMMARY**

70

71 • Some *Aeschynomene* legume species have the property of being nodulated by
72 photosynthetic *Bradyrhizobium* lacking the *nodABC* genes. Knowledge of this unique Nod
73 (factor)-independent symbiosis has been gained from the model *A. evenia* but our
74 understanding remains limited due to the lack of comparative genetics with related taxa using
75 a Nod-dependent process.

76 • To fill this gap, this study significantly broadened previous taxon sampling, including in
77 allied genera, to construct a comprehensive phylogeny. This backbone tree was matched with
78 data on chromosome number, genome size, low-copy nuclear genes and strengthened by
79 nodulation tests and a comparison of the diploid species.

80 • The phylogeny delineated five main lineages that all contained diploid species while
81 polyploid groups were clustered in a polytomy and were found to originate from a single
82 paleo-allopolyploid event. In addition, new nodulation behaviours were revealed and Nod-
83 dependent diploid species were shown to be tractable.

84 • The extended knowledge of the genetics and biology of the different lineages in the legume
85 genus *Aeschynomene* provides a solid research framework. Notably, it enabled the
86 identification of *A. americana* and *A. patula* as the most suitable species to undertake a
87 comparative genetic study of the Nod-independent and Nod-dependent symbioses.

88

89

90

91

92

93

94

95

96

97

98

99 **Keywords**

100 *Aeschynomene*, genetics, legumes, nodulation, phylogenetics, polyploidy, symbiosis

101

102

103 **INTRODUCTION**

104

105 In the field of nitrogen-fixing symbiosis, scientists have a long-standing interest in the tropical
106 papilionoid legume genus *Aeschynomene* since the discovery of the ability of the species *A.*
107 *afraspera* to develop abundant nitrogen-fixing stem nodules (Hagerup, 1928). This nodulation
108 behavior is uncommon in legumes, being shared by very few hydrophytic species of the
109 genera *Discolobium*, *Neptunia* and *Sesbania*, but it is exceptionally widespread among the
110 semi-aquatic *Aeschynomene* species (Alazard, 1985; Boivin *et al.*, 1997; Chaintreuil *et al.*,
111 2013). These stem-nodulating *Aeschynomene* species are able to interact with *Bradyrhizobium*
112 strains that display the unusual property of being photosynthetic (Giraud *et al.*, 2000; Miché
113 *et al.*, 2010). Most outstanding is the evidence that some of these photosynthetic
114 *Bradyrhizobium* strains lack both the *nodABC* genes required for the synthesis of the key
115 “Nod factors” symbiotic signal molecules and a type III secretion system (T3SS) that is
116 known in other rhizobia to activate or modulate nodulation (Giraud *et al.*, 2007; Okazaki *et*
117 *al.*, 2013, 2015). These traits revealed the existence of an alternative symbiotic process
118 between rhizobia and legumes that is independent of the Nod factors.

119 As in the legume genus *Arachis* (peanut), *Aeschynomene* uses an intercellular symbiotic
120 infection process instead of infection thread formation that can be found in other legume
121 groups (Sprent *et al.*, 2017). This led to the suggestion that the Nod-independent process
122 might correspond to the ground state of the rhizobial symbiosis, although we cannot exclude
123 that it represents an alternative symbiotic process compared to the one described in other
124 legumes (Sprent & James, 2008; Madsen *et al.*, 2010, Okubo *et al.*, 2012). It is noteworthy
125 that all the Nod-independent species form a monophyletic clade within the *Aeschynomene*
126 phylogeny and jointly they also display striking differences in the bacteroid differentiation
127 process compared to other *Aeschynomene* species (Chaintreuil *et al.*, 2013; Czernic *et al.*,
128 2015). To decipher the molecular mechanisms of this distinct symbiosis, the Nod-independent
129 *A. evenia* has been used as a new model legume, because its genetic and developmental
130 characteristics (diploid with a reasonable genome size $2n=20$, 415 Mb/1C-, short perennial
131 and autogamous, can be hybridized and transformed) make this species tractable for
132 molecular genetics (Arrighi *et al.*, 2012, 2013, 2015). Functional analyses revealed that some
133 symbiotic determinants identified in other legumes (*SYMRK*, *CCaMK*, *HK1* and *DNF1*) are
134 recruited, but several key genes involved in bacterial recognition (e.g. *LYK3*), symbiotic
135 infection (e.g. *EPR3* and *RPG*), and nodule functioning (e.g. *DNF2* and *FEN1*) were found
136 not to be expressed in *A. evenia* roots and nodules, based on RNAseq data (Czernic *et al.*,

137 2015; Fabre *et al.*, 2015; Chaintreuil *et al.*, 2016a; Nouwen *et al.*, 2017). This suggested that
138 the Nod-independent symbiosis is distinct from the Nod-dependent one.

139 Forward genetics are now expected to allow the identification of the specific molecular
140 determinants of the Nod-independent process in *A. evenia* (Arrighi *et al.*, 2012; Chaintreuil *et*
141 *al.*, 2016a). In addition, comparing *A. evenia* with closely related Nod-dependent
142 *Aeschynomene* species will promote our understanding how symbiosis evolved in legumes.
143 The genus *Aeschynomene* (restricted now to the section *Aeschynomene* as discussed in
144 Chaintreuil *et al.* (2013)) is traditionally composed of three infrageneric taxa, subgenus
145 *Aeschynomene* (which includes all the hydrophytic species) and subgenera *Bakerophyton* and
146 *Rueppellia* (Rudd, 1955; Gillet *et al.*, 1971). The genus has also been shown to be
147 paraphyletic, with a number of related genera being nested within it, but together they form a
148 distinct clade in the tribe Dalbergieae (Chaintreuil *et al.*, 2013; Rudd, 1981; Lavin *et al.*,
149 2001; Klitgaard *et al.*, 2005; LPWG, 2017). Within this broad clade, two groups of semi-
150 aquatic *Aeschynomene* have been well-studied from a genetic and genomic standpoint: the *A.*
151 *evenia* group, which contains all the Nod-independent species (most of them being 2x), and
152 the *A. afraspera* group (all species being Nod-dependent) that appears to have a 4x origin
153 (Arrighi *et al.*, 2014, Chaintreuil *et al.* 2016b, 2018). For comparative analyses, the use of
154 Nod-dependent species with a diploid structure would be more appropriate, but such
155 *Aeschynomene* species are poorly documented.

156 To overcome these limitations, our aim was to produce a species-comprehensive phylogenetic
157 tree supplemented with genetic and nodulation data. For this, we made use of an extensive
158 taxon sampling in both the genus *Aeschynomene* and in closely related genera to capture the
159 full species diversity of the genus and to clarify phylogenetic relationships between taxa. For
160 most species, we also documented chromosome number, genome size and molecular data for
161 low-copy nuclear genes, thus allowing the identification of diploid species as well as
162 untangling the genome structure of polyploid taxa. In addition, these species were
163 characterized for their ability to nodulate with various *Bradyrhizobium* strains containing or
164 lacking *nod* genes and, finally, the diploid species were submitted to a comparative analysis
165 of their properties. In light of the data obtained in this study, we discuss the interest of two
166 *Aeschynomene* species, *A. americana* and *A. patula*, to set up a comparative genetic system to
167 complement the *A. evenia* model.

168

169

170 **MATERIALS & METHODS**

171

172 **Plant material**

173 All the accessions of *Aeschynomene* used in this study, including their geographic origin and
174 collection data are listed in Tables S1 and S4. Seed germination and plant cultivation in the
175 greenhouse were performed as indicated in Arrighi *et al.* (2012). Phenotypic traits such as the
176 presence of adventitious root primordia and nodules on the stem were directly observed in the
177 glasshouse.

178

179 **Nodulation tests**

180 Nodulation tests were carried out using *Bradyrhizobium* strains ORS278 (originally isolated
181 from *A. sensitiva* nodules), ORS285 (originally isolated from *A. afraspera* nodules),
182 ORS285 Δ *nod* and DOA9 (originally isolated from *A. americana* nodules) (Giraud *et al.*,
183 2007; Teamtisong *et al.*, 2014; Bonaldi *et al.*, 2011). *Bradyrhizobium* strains were cultivated
184 at 34°C for seven days in Yeast Mannitol (YM) liquid medium supplemented with an
185 antibiotic when necessary (Howieson *et al.*, 2016). Plant *in vitro* culture was performed in
186 tubes filled with buffered nodulation medium (BNM) as described in Arrighi *et al.* (2012).
187 Five-day-old plants were inoculated with 1 mL of bacterial culture with an adjusted OD at
188 600nm to 1. Twenty one days after inoculation, six plants were analysed for the presence of
189 root nodules. Nitrogen-fixing activity was estimated on the entire plant by measurement of
190 acetylene reducing activity (ARA) and microscopic observations were performed using a
191 stereo-microscope (Nikon AZ100, Champigny-sur-Marne, France) as published in Bonaldi *et*
192 *al.* (2011).

193

194 **Molecular methods**

195 Plant genomic DNA was isolated from fresh material using the classical CTAB (Cetyl
196 Trimethyl Ammonium Bromide) extraction method. For herbarium material, the method was
197 adapted by increasing the length of the incubation (90 min), centrifugation (20 min) and
198 precipitation (15 min) steps. The nuclear ribosomal internal transcribed spacer region (ITS),
199 the chloroplast *matK* gene and four low-copy nuclear genes (*CYP1*, *eiF1 α* , *SuSy*, and *TIP1;1*)
200 previously identified in the *A. evenia* and *A. afraspera* transcriptomes were used for
201 phylogenetic analyses (Arrighi *et al.*, 2014; Chaintreuil *et al.*, 2016). The genes were PCR-
202 amplified, cloned and sequenced as described in Arrighi *et al.* (2014) (Table S2). For genomic
203 DNA extracted from herbarium specimens, a battery of primers was developed to amplify the

204 different genes in overlapping fragments as short as 250 bp (Table S2). The DNA sequences
205 generated in this study were deposited in GenBank (Table S3).

206

207 **Phylogenetic analyses and traits mapping**

208 Sequences were aligned using MAFFT (*--localpair --maxiterate 1000*; Katoh & Standley,
209 2013). Phylogenetic reconstructions were performed for each gene as well as for concatenated
210 datasets under a Bayesian approach using Phylobayes 4.1b (Lartillot & Philippe, 2004) and
211 the site-heterogeneous CAT+F81+F4 evolution model. For each analysis, two independent
212 chains were run for 10,000 Phylobayes cycles with a 50% burn-in. Ancestral states
213 reconstruction was done through stochastic character mapping using the Phytools R package
214 (Revell, 2012) running 10 simulations for each character.

215

216 **Species networks and hybridizations**

217 To test if the phylogeny obtained by concatenating the four low-copy nuclear genes (*CYP1*,
218 *eiFla*, *SuSy*, and *TIP1;1*) was most likely obtained by gene duplications followed by
219 differential losses or by a combination of duplications, losses coupled with one or several
220 allopolyploidy events involving *A. patula* and *Soemmeringia semperflorens*, the method
221 presented in To & Scornavacca (2015) was used. In short, this method computes a
222 reconciliation score by comparing a phylogenetic network and one or several gene trees. The
223 method allows allopolyploidy events at hybridization nodes while all other nodes of the
224 network are associated to speciation events; meanwhile, duplication and loss events are
225 allowed at a cost (here, arbitrarily fixed to 1) on all nodes of the gene tree.

226 Thus, the set of 4 nuclear gene trees was used to score different phylogenetic networks
227 corresponding to four different potential evolutionary histories. Two alternative networks with
228 no reticulation corresponding to the two topologies obtained either with the group A (T1) or
229 group B (T2) served to evaluate a no-allopolyploidisation hypothesis. The topology yielding
230 the best score (T2) served to generate and compare all phylogenetic networks with one or two
231 hybridization nodes, involving *A. patula* and/or *S. semperflorens*, to test successively a one-
232 allopolyploidisation scenario (N1-best) and a two-allopolyploidisation evolutionary scenario
233 (N2-best).

234

235 **GBS analysis**

236 A GBS library was constructed based on a described protocol (Oueslati *et al.*, 2017). For each
237 sample, a total of 150 ng of genomic DNA was digested using the two-enzyme system, PstI

238 (rare cutter) and Mse (common cutter) (New England Biolabs, Hitchin, UK), by incubating at
239 37°C for 2 h. The ligation reaction was performed using the T4 DNA ligase enzyme (New
240 England Biolabs, Hitchin, UK) at 22°C for 30 min and the ligase was inactivated at 65°C for
241 30 min. Ligated samples were pooled and PCR-amplified using the Illumina Primer 1
242 (barcoded adapter with PstI overhang) and Illumina Primer 2 (common Y-adapter). The
243 library was sequenced on an Illumina HiSeq 3000 (1x 150 pb) (at the Get-PlaGe platform in
244 Toulouse, France).

245 The raw sequence data were processed in the same way as in the study described in Garsmeur
246 *et al.* (2018). SNP calling from the raw Illumina reads was performed using the custom
247 python pipeline VcfHunter (available at <https://github.com/SouthGreenPlatform/VcfHunter/>)
248 (Guillaume Martin, CIRAD, France). For all samples, these sequence tags were aligned to the
249 *A. evenia* 1.0 reference genome (JF Arrighi, unpublished data). The SNP results from all the
250 samples were converted into one large file in VCF format and the polymorphism data were
251 subsequently analysed using the web-based application SNIPlay3 (Dereeper *et al.*, 2015).
252 First, the SNP data were treated separately for each species and filtered out to remove SNP
253 with more than 10% missing data as well as those with a minor allele frequency (MAF) of
254 0.01 using integrated VCFtools. Second, an overall representation of the species diversity
255 structures was obtained by making use of the PLINK software as implemented in SNIPlay3.
256 This software is based on the multidimensional-scaling (MSD) method to produce two-
257 dimensional plots. The Illumina HiSeq 3000 sequencing raw data are available in the NCBI
258 SRA (Sequence Read Archive) under the study accession number: SRP149516.

259

260 **Genome size estimation and chromosome counting**

261 Genome sizes were measured by flow cytometry using leaf material as described in Arrighi *et*
262 *al.* (2012). Genome size estimations resulted from measurements of three plants per accession
263 and *Lycopersicon esculentum* (Solanaceae) cv “Roma” (2C = 1.99 pg) was used as the
264 internal standard. The 1C value was calculated and the conversion factor 1 pg DNA = 978 Mb
265 was used to express it in Mb/1C. To count chromosome number, metaphasic chromosomes
266 were prepared from root-tips, spread on slides, stained with 4',6-diamidino-2-phenylindole
267 (DAPI) and their image captured with a fluorescent microscope as detailed in Arrighi *et al.*
268 (2012) .

269

270

271 **RESULTS**

272

273 **A comprehensive phylogeny of the genus *Aeschynomene* and allied genera**

274 To obtain an in-depth view of the phylogenetic relationships within the genus *Aeschynomene*
275 subgenus *Aeschynomene*, which contains the hydrophytic species, we significantly increased
276 previous sampling levels by the addition of new germplasm accessions and, if these were not
277 available, we used herbarium specimens. DNA was isolated for 40 out of the 41 species
278 (compared to the 27 species used in Chaintreuil *et al.* (2013)) included in this group in
279 taxonomic and genetic studies (Table S1) (Arrighi *et al.*, 2014; Chaintreuil *et al.*, 2012,
280 2016b, 2018; Rudd, 1955). In addition, to determine the phylogenetic relationship of this
281 subgenus with respect to *Aeschynomene* subgenera *Bakerophyton* and *Rueppellia*, unclassified
282 *Aeschynomene* species, as well as with the allied genera *Bryaspis*, *Cyclocarpa*, *Geissaspis*,
283 *Humularia*, *Kotschya*, *Smithia* and *Soemmeringia*, representatives of these 10 taxa were also
284 sampled (compared to the 5 taxa present in Chaintreuil *et al.* (2013)) (Rudd, 1981; Lewis,
285 2005). This added 21 species to our total samples (Table S1). The dalbergioid species *Pictetia*
286 *angustifolia* was used as outgroup (Chaintreuil *et al.* 2013; LPWG, 2017).

287 Phylogenetic reconstruction of all the taxa sampled was undertaken using Bayesian analysis
288 of the chloroplast *matK* gene and the nuclear ribosomal *ITS* region (Table S2 and S3). The
289 *matK* and *ITS* gene trees distinguished almost all the different *Aeschynomene* groups and
290 related genera (Fig. S1 and S2). The two phylogenetic trees have a very similar topology
291 although some branches of both trees can be lowly supported. Incongruences were also
292 observed for *A. deamii* and the genus *Bryaspis*, but the conflicting placements are poorly
293 supported and were thus interpreted as a lack of resolution typical of single-marker trees,
294 rather than strong incongruence. To improve the phylogenetic resolution among the major
295 lineages, the *matK* gene and the *ITS* sequence datasets were combined into a single
296 phylogenetic analysis where only well-supported nodes were considered (posterior probability
297 (PP) ≥ 0.5) (Fig. 1). Our analysis recovered a grade of five main lineages with a branching
298 order that received robust support (PP ≥ 0.92): (1) a basally branching lineage represented by
299 *A. americana*, (2) an *A. montevidensis* lineage, (3) an *A. evenia* lineage corresponding to the
300 Nod-independent clade (Arrighi *et al.*, 2012, 2014), (4) a newly-identified lineage containing
301 *A. patula* and (5) a lineage represented by an unresolved polytomy clustering the *A. afraspera*
302 clade (Chaintreuil *et al.*, 2016b) with all the remaining taxa.

303 In large part, our work also provided good species-level resolution and demonstrated that
304 *Aeschynomene* subgenus *Aeschynomene* (as currently circumscribed) is interspersed on the
305 phylogenetic tree with the lineage containing *A. patula*, the two other subgenera of

306 *Aeschynomene* and a number of other genera related to *Aeschynomene* (Fig. 1) (Chaintreuil *et*
307 *al.*, 2013; Lavin *et al.*, 2001; LPWG, 2017; Du Puy *et al.*, 2002). The combined analysis also
308 grouped the genus *Bryaspis* with the species related to *A. afraspera* in a highly supported
309 clade but its exact position with respect to other taxa remained inconclusive, as previously
310 observed (Fig. 1) (Chaintreuil *et al.*, 2013). Most noticeably, several intergeneric relationships
311 are consistently recovered, notably sister-clade relationships between *Cyclocarpa* and *Smithia*
312 as well as between *Aeschynomene* subgenera *Bakerophyton* and *Rueppellia* together with the
313 genus *Humularia* (referred to as the BRH clade herein after) (Fig. 1). This clade supports
314 previous observations of a morphological continuum between *Aeschynomene* subgenus
315 *Rueppellia* and the genus *Humularia* and brings into question their taxonomic separation
316 (Gillett *et al.*, 1971).

317

318 **Ploidy level of the species and genomic structure of the different lineages**

319 The revised *Aeschynomene* phylogeny was used as a backbone tree to investigate the
320 evolution of ploidy levels. Previous studies had demonstrated that the *A. evenia* clade is
321 mostly diploid ($2n=2x=20$) even if some species such as *A. indica* ($2n=4x=40$, $2n=6x=60$)
322 appear to be of recent allopolyploid origin (Arrighi *et al.*, 2014; Chaintreuil *et al.*, 2018).
323 Conversely, all the species of the *A. afraspera* group were found to be polyploid
324 ($2n=4x=28,38,40$, $2n=8x=56,76$) and to have a common AB genome structure but the origin
325 of the polyploidy event remained undetermined (Chaintreuil *et al.*, 2016b). To assess the
326 ploidy levels in *Aeschynomene* species and related genera, chromosome numbers and nuclear
327 DNA content were determined (appended to labels in Fig. 2a, Table S1, Fig. S3 and S4). We
328 provide evidence that the lineages containing *A. americana*, *A. montevidensis*, *A. evenia* and
329 *A. patula*, as well as *Soemmeringia semperflorens*, are diploid with $2n=20$, with the smallest
330 $2x$ genome found in *A. patula* (0.58 pg/2C) and the largest $2x$ genome in *A. deamii* (1.93
331 pg/2C). With the exception of *S. semperflorens*, all the groups that are part of the polytomy
332 were characterized by higher chromosome numbers $2n=28,36,38,40$ (up to 76). These
333 chromosome numbers equate to approximately twice that of the diploid species (except for
334 $2n=28$), suggesting that the corresponding groups are most probably polyploid. Species with
335 chromosome numbers departing from $2n=40$ are likely to be of dispolyploid origin as already
336 described in the *A. afraspera* clade (Chaintreuil *et al.*, 2016b). Here again, important genome
337 size variations ranging from 0.71 pg/2C for the *Geissaspis* species to 4.82 pg/2C for the $4x$ *A.*
338 *schimperii* highlight the genomic differentiation of the various taxa (Fig. 2a, Table S1).

339 To firmly link chromosome numbers to ploidy levels and to clarify genetic relationships
340 between the different lineages, we cloned and sequenced four nuclear-encoded low-copy
341 genes in selected species: *CYPI* (Cyclophilin 1), *eiFl α* (eukaryotic translation initiation factor
342 α), *SuSy* (Sucrose Synthase) and *TIP1;1* (tonoplast intrinsic protein 1;1) (Table S2). For all
343 diploid species, only one gene sequence was obtained, while for all the polyploid species, in
344 almost all cases, a pair of putative homeologues was isolated, thus confirming their genetic
345 status inferred from the karyotypic data (Table S3). In general, the duplicated copies were
346 highly divergent and nested in two different major clades in the resulting Bayesian phylogenic
347 trees generated for each gene (Fig. S5). One clade contained all the A copies (except for one
348 anomalous sequence for *Bryaspis lupulina* in the *eiFl α* tree) and the other clade gathered all
349 the B copies previously identified in *A. afraspera* (Chaintreuil *et al.*, 2016b). These two
350 clades A and B did not always receive high support, however it is notable that the A copies
351 formed a monophyletic group with, or sister to, the *A. patula* sequence and similarly the B
352 copies with, or sister to, the *Soemmerignia semperflorens* sequence, in all gene trees (Fig. S5).
353 In an attempt to improve phylogenetic resolution, the four gene data sets were concatenated.
354 This combination resulted in a highly supported Bayesian tree that places the A copy clade as
355 the sister to the diploid *A. patula* (PP =1), and the B copy clade as sister to the diploid *S.*
356 *semperflorens* (PP =1) (Fig. 2b). As a result, these phylogenetic analyses combined to
357 karyotypic data show that all the five main lineages contain diploid species. They also reveal
358 that all the polyploid groups share the same AB genome structure, with the diploid *A. patula*
359 and *S. semperflorens* species being the closest modern representatives of the ancestral donors
360 of the A and B genomes.

361 In addition, an ancestral state reconstruction analysis performed on the *ITS+matK* phylogeny
362 indicates that diploidy is the ancestral condition in the whole revised group and that
363 tetraploidy most likely evolved once in the polytomy (Fig. S6). To provide support on a
364 probable single origin of the allopolyploidy event, separate and concatenated nuclear gene
365 trees were further used for a phylogenetic network analysis. In this analysis, the two non-
366 allopolyploidisation hypotheses (T1 and T2) were found to be more costly (scores of 207 and
367 196) than the two hypotheses allowing for hybridization (N1-best and N2-best with scores of
368 172 and 169, respectively) (Fig. S7a-d). The one-allopolyploidisation hypothesis (N1-best)
369 strongly indicates that a hybridization involving the lineages that contain *A. patula* and *S.*
370 *semperflorens* gave rise to all the polyploid groups (Fig. S7c). Although the two-
371 allopolyploidisation hypothesis (N2-best) yielded the absolute best score, the score
372 improvement was very low (169 vs 172) and the resulting network included the hybridization

373 inferred with the one-allopolyploidisation hypothesis making this latter hypothesis most
374 probably the correct one (Fig. S7d).

375

376 **Nodulation properties of the different *Aeschynomene* lineages**

377 Species of *Aeschynomene* subgenus *Aeschynomene* are known to be predominantly
378 amphibious and more than 15 of these hydrophytic species (found in the *A. evenia* and *A.*
379 *afraspera* clades, as well as *A. fluminensis*) have been described as having the ability to
380 develop stem nodules (Boivin *et al.*, 1997; Chaintreuil *et al.*, 2016b; Lock, 1989; Rudd,
381 1955). In *A. fluminensis*, these nodules are observed only on submerged stems (as also seen in
382 the legume *Discolobium pulchellum*), while they occur on aerial stems within the *A. evenia*
383 and *A. afraspera* clades (Fig. 3a) (Alazard & Duhoux, 1987; Chaintreuil *et al.*, 2013; Loureiro
384 *et al.*, 1994, 1995). Phenotypic analysis of representatives of the different lineages under
385 study revealed that they all display adventitious root primordia along the stem (Fig. 3a,b).
386 Adventitious roots are considered to be an adaptation to temporary flooding and they also
387 correspond to nodulation sites in stem-nodulating *Aeschynomene* species (Fig. 3b) (Alazard &
388 Duhoux, 1987). Given that the *A. evenia* and *A. afraspera* clades are now demonstrated not to
389 share the same genomic components provides a genetic argument for independent
390 developments of stem nodulation by photosynthetic bradyrhizobia. Reconstruction of
391 ancestral characters based on the *ITS+matK* phylogeny confirmed that the whole group was
392 ancestrally a wet ecology taxon endowed with adventitious root primordia but that the stem
393 nodulation ability evolved several times, as previously inferred (Fig. S8, S9; and S10)
394 Chaintreuil *et al.*, 2013, 2016b).

395 To investigate whether the newly studied species could be nodulated by photosynthetic
396 bradyrhizobia, we extended the results obtained by Chaintreuil *et al.* (2013) by testing the
397 nodulation abilities of 22 species available (listed in Fig. 4a) for which adequate seed supply
398 was available. Three different strains of *Bradyrhizobium* equating to the three cross-
399 inoculation (CI) groups defined by Alazard (1985) were used: DOA9 (non-photosynthetic
400 *Bradyrhizobium* of CI-group I), ORS285 (photosynthetic *Bradyrhizobium* with *nod* genes of
401 CI-group II) and ORS278 (photosynthetic *Bradyrhizobium* lacking *nod* genes of CI-group
402 III). These strains were used to inoculate the 22 species and their ability to nodulate them was
403 analysed at 21 dpi. For this, we recorded nodule formation and compared nitrogen fixation
404 efficiency by an acetylene reduction assay (ARA) and observation of plant vigor. Nodulation
405 was observed on all species tested except for *Smithia sensitiva* that had a problem with root

406 development, *A. montevidensis* and *S. semperflorens*. For these three species, either the
407 culture conditions or the *Bradyrhizobium* strains used were not appropriate (Fig. 4a).
408 The non-photosynthetic strain DOA9 displayed a wide host spectrum but was unable to
409 nodulate the Nod-independent species, *A. deamii*, *A. evenia* and *A. tambacoundensis*. The
410 photosynthetic strain ORS285 efficiently nodulated *A. afraspera* and the Nod-independent
411 *Aeschynomene* species (Fig 4a), as previously reported (Chaintreuil *et al.*, 2013).
412 Interestingly, the ORS285 strain was also able to induce nitrogen-fixing nodules in *A. patula*
413 and ineffective nodules were observed on *A. fluminensis* and the genera *Bryaspis*, *Cyclocarpa*
414 and *Smithia* (Fig. 4a). To examine if in these species the nodulation process relies on a Nod-
415 dependent or Nod-independent symbiotic process, we took advantage of the availability of a
416 Δnod mutant of the strain ORS285. None of them were found to be nodulated by
417 ORS285 Δnod , suggesting that the nodule formation depended on a Nod signaling in these
418 species (Fig. 4a). In fact, the ORS285 Δnod mutated strain was able to nodulate only species
419 of the *A. evenia* clade, similarly as to the photosynthetic strain ORS278 naturally lacking *nod*-
420 genes (Fig. 4a). Analysis of the evolution of these nodulation abilities by performing an
421 ancestral state reconstruction on the revisited phylogeny indicated several emergences of the
422 ability to interact with photosynthetic bradyrhizobia and a unique emergence of the ability to
423 be nodulated by the *nod* gene-lacking strain as observed earlier (Fig. S11 and Fig. S12)
424 (Chaintreuil *et al.*, 2013). From these nodulation tests, different nodulation patterns emerged
425 for the diploid *Aeschynomene* species (as detailed in Fig. 4b-d) with the DOA9 and ORS278
426 strains being specific to the Nod-dependent and Nod-independent groups respectively and
427 ORS285 showing a gradation of compatibility between both.

428

429 **Diversity of the diploid species outside the Nod-independent clade**

430 To further characterize the diploid species that fall outside of the Nod-independent clade, in
431 which *A. evenia* lies, they were analysed for their developmental properties and genetic
432 diversity (Fig. 5a). All species are described as annuals or short-lived perennials (Du Puy *et*
433 *al.*, 2002; Lewis, 2005; Rudd, 1955). As for *A. evenia*, *A. americana*, *A. villosa*, *A.*
434 *fluminensis*, *A. parviflora* and *A. montevidensis* are robust and erect, reaching up to 2 m high
435 when mature, whilst *A. patula* and *S. semperflorens* are creeping or decumbent herbs. These
436 differences in plant habit are reflected by the important variation in seed size between these
437 two groups (Fig. 5a). This has an impact on plant manipulation, because for *A. patula* and *S.*
438 *sempreflorens* seed scarification needs to be adapted (25 min with concentrated sulfuric acid
439 instead of 40 min for the other species) and *in vitro* plant growth takes slightly more time to

440 get a root system sufficiently developed for inoculation with *Bradyrhizobium* strains (10 days-
441 post-germination instead of the 5-7 dpi for other species) (Arrighi *et al.*, 2012). Consistent
442 flowering and seed production was observed for *A. americana*, *A. villosa*, *A. patula* and *S.*
443 *semperflorens* when grown under full ambient light in the tropical greenhouse in short days
444 conditions as previously described for *A. evenia*, making it possible to develop inbred lines by
445 successive selfing (Fig. 5a) (Arrighi *et al.*, 2012). For *A. fluminensis*, *A. parviflora* and *A.*
446 *montevidensis*, flowering was sparse or not observed, indicating that favorable conditions for
447 controlled seed set were not met (Fig. 5a).

448 Five species (*A. villosa*, *A. fluminensis*, *A. parviflora*, *A. montevidensis* and *S. semperflorens*)
449 are strictly American while *A. americana* is a pantropical species and *A. patula* is endemic to
450 Madagascar (Du Puy *et al.*, 2002; Lock, 1989; Rudd, 1955). Several species have a narrow
451 geographic distribution or seem to be infrequent, explaining the very limited accession
452 availability in seedbanks (Fig. 5a). This is in sharp contrast with both *A. americana* and *A.*
453 *villosa* that are well-collected, being widely found as weedy plants and sometimes used as
454 component of pasture for cattle (Fig. 5a) (Cook *et al.*, 2005). To assess the genetic diversity of
455 these two species, a germplasm collection containing 79 accessions for *A. americana* and 16
456 accessions for *A. villosa*, and spanning their known distribution was used (Table S4). A
457 Genotyping-By-Sequencing (GBS) approach resulted in 6370 and 1488 high quality
458 polymorphic SNP markers for *A. americana* and *A. villosa* accessions, respectively. These
459 two SNP datasets subsequently served for a clustering analysis based on the
460 multidimensional-scaling (MSD) method. The MSD analysis distinguished three major
461 groups of accessions for both *A. americana* and *A. villosa* along coordinate axes 1 and 2 (Fig.
462 5b). When mapping the accessions globally, the three groups identified for *A. villosa* were
463 observed together in Mexico and only group (3) extended to the northern part of South
464 America (Fig. 5c, Table S4). Conversely, a clear geographical division was observed for *A.*
465 *americana* with group (1) occupying the central part of South America, group (2) being found
466 in the Caribbean area while group (3) was present in distinct regions from Mexico to Brazil
467 and in across the Palaeotropics (Fig. 5c, Table S4). *A. americana* is hypothesized to be native
468 in America and naturalized elsewhere (Cook *et al.*, 2005). The observed distributions in
469 combination with the MSD analysis, accessions being tightly clustered in group (3) compared
470 to groups (1) and (2), support this idea and indicate that group (3) recently spread worldwide.

471

472

473 **DISCUSSION**

474

475 **A well-documented phylogenetic framework for the legume genus *Aeschynomene***

476 We produced a new and comprehensive phylogeny of the genus *Aeschynomene* and its closely
477 related genera complemented by gene data sets, genome sizes, karyotypes and nodulation
478 assays. For plant genera, there are few for which documentation of taxonomic diversity is so
479 extensive and supported by a well-resolved, robustly supported phylogeny which reveals the
480 evolutionary history of the group (Govindarajulu *et al.*, 2011). Here, the whole group, which
481 includes the genus *Aeschynomene* with its 3 subgenera and its 7 allied genera, is shown to
482 have experienced cladogenesis leading to five main lineages, including the Nod-independent
483 clade, with diploid species found in all these lineages. The multigene data analysis provided
484 robust evidence that two of them, represented by the two diploid species *A. patula* and *S.*
485 *semperflorens*, are involved in an ancient allotetraploidization process that gave rise to the
486 different polyploid lineages clustering in a polytomy. Separate allopolyploidization events
487 from the same diploid parents or a single allopolyploid origin are plausible explanations for
488 the formation of these lineages. However, the consistent resolution of the phylogenetic tree
489 obtained with the combined gene data, where *A. patula* and *S. semperflorens* are sisters to the
490 A and B subgenomic sequences, favours the hypothesis of a single allopolyploid origin, as
491 also argued for other ancient plant allopolyploid events in *Asimitellaria* (Saxifragaceae) and
492 *Leucaena* (Leguminosae) (Govindarajulu *et al.*, 2011; Okuyama *et al.*, 2012). The
493 phylogenetic network analysis also supports the one-allopolyploidisation hypothesis.
494 However, additional nuclear genes will be needed to conclusively confirm that no additional
495 hybridization event occurred. Although not the focus of the present study, it is worth noting
496 that most diploid species are found in the Neotropics, the two modern representatives of the A
497 and B genome donors that gave rise to the 4x lineages are located on different continents (*S.*
498 *semperflorens* in South America and *A. patula* in Madagascar) and that all the 4x lineages are
499 located in the Palaeotropics (Lewis *et al.*, 2005). This raises questions about the evolution of
500 the whole group and the origin of the 4x lineages. In addition, the presence of a polytomy
501 suggests that this allopolyploid event preceded a rapid and major diversification of 4x groups
502 that have been ascribed to different *Aeschynomene* subgenera or totally distinct genera that
503 altogether represent more than 80% of the total species of the whole group (LPWG, 2017;
504 Whitefiel & Lockhart, 2007). Diversification by allopolyploidy occurred repeatedly in the
505 genus *Aeschynomene* since several neopolyploid species are found in both the *A. evenia* clade
506 and the *A. afraspera* clade as exemplified by *A. indica* (4x, 6x) and *A. afraspera* (8x) (Arrighi
507 *et al.*, 2014; Chaintreuil *et al.*, 2016b). Dense sampling for several *Aeschynomene* taxa or

508 clades also a more precise delimitation of species boundaries (for morphologically similar
509 taxa but which are genetically differentiated or correspond to different cytotypes) and
510 revealed intraspecific genetic diversity that is often geographically-based as showed for the
511 pantropical species *A. americana* (this study), *A. evenia*, *A. indica* and *A. sensitiva*
512 (Chaintreuil *et al.*, 2018). All these *Aeschynomene* species share the presence of adventitious
513 root primordia on the stem that correspond to the infection sites for nodulation. The consistent
514 presence of adventitious root primordia in all taxa of the whole group, together with an
515 ancestral state reconstruction, substantiates the two-step model proposed earlier for the
516 evolution of stem nodulation in *Aeschynomene*, with a common genetic predisposition at the
517 base of the whole group to produce adventitious root primordia on the stem, as an adaptation
518 to flooding, and subsequent mutations occurring independently in various clades to enable
519 stem nodulation (Chaintreuil *et al.*, 2013). The ability to interact with photosynthetic
520 bradyrhizobia that are present in aquatic environments also appears to have evolved several
521 times. This photosynthetic activity is important for the bacterial symbiotic lifestyle as it
522 provides energy usable for infection and subsequently for nitrogenase activity inside the stem
523 nodules (Giraud *et al.*, 2000). To date, natural occurrence of nodulation by photosynthetic
524 bradyrhizobia has been reported only for the *A. evenia* and *A. afraspera* clades, and for *A.*
525 *fluminensis* (Loureiro *et al.*, 1995; Miché *et al.*, 2010; Molouba *et al.*, 1997). Nevertheless, we
526 could not test the photosynthetic strains isolated from *A. fluminensis* nodules and the nature of
527 the strains present in those of the newly studied species *A. patula* has not been investigated
528 yet. They would allow the comparison of their nodulation efficiency with the reference
529 photosynthetic *Bradyrhizobium* ORS278 and ORS285 strains. In addition, we can ask if the
530 semi-aquatic lifestyle and/or nodulation with photosynthetic bradyrhizobia may have
531 facilitated the emergence of the Nod-independent symbiosis in the *A. evenia* clade.

532

533 ***Aeschynomene* species for a comparative analysis of nodulation with *A. evenia***

534 To uncover whether the absence of detection for several key symbiotic genes in the root and
535 nodule transcriptomic data of *A. evenia* are due to gene loss or extinction, and to identify the
536 specific symbiotic determinants of the Nod-independent symbiosis, a genome sequencing
537 combined with a mutagenesis approach is presently being undertaken for *A. evenia* in our
538 laboratory. A comparative analysis with Nod-dependent *Aeschynomene* species is expected to
539 consolidate this genomic and genetic analysis performed in *A. evenia* by contributing to
540 elucidate the genetic changes that enabled the emergence of the Nod-independent process.
541 Phylogenomics and comparative transcriptomics, coupled with functional analysis, are

542 undergoing increased development in the study of symbiosis. They enabled unravelling gene
543 loss linked to the lack of developing a symbiosis in certain plant lineages but also to identify
544 new symbiosis genes (for arbuscular mycorrhizal symbiosis (Delaux *et al.*, 2014; Bravo *et al.*,
545 2016); for the nodulating symbiosis (Delaux *et al.*, 2015; Griesmann *et al.*, 2018).
546 Comparative work on symbiotic plants is often hindered, however, either by the absence of
547 closely related species which display gain or loss of symbiotic function or, when these are
548 present, by the lack of a well-understood genetic framework, as outlined in Behm *et al.*
549 (2014), Delaux *et al.* (2015), Geurts *et al.* (2016), Sprent (2017). Nevertheless, in the case of
550 the nodulating *Parasponia*/non-nodulating *Trema* system, a fine comparative analysis was
551 very powerful to demonstrate a parallel loss of the key symbiotic genes *NFR5*, *NIN* and *RGP*,
552 in the non-nodulating species, challenging the long-standing assumption that *Parasponia*
553 specifically acquired the potential to nodulate (Behm *et al.*, 2014; Geurts *et al.*, 2016; van
554 Velzen *et al.*, 2018). In this respect, the uncovering of the genetic evolution of the genus
555 *Aeschynomene* and related genera along with the identification of diploid species outside of
556 the Nod-independent clade, provided a robust phylogenetic framework that can now be
557 exploited to guide the choice of Nod-dependent diploid species for comparative genetic
558 research. Among these, some species are discarded because of the lack of nodulation with
559 reference *Bradyrhizobium* strains or the inability to produce seeds under greenhouse
560 conditions. Based on efficient nodulation, short flowering time and ease of seed production,
561 *A. americana* (2n=20, 600 Mb) and *A. patula* (2n=20, 270 Mb) appear to be the most
562 promising Nod-dependent diploid species to develop a comparative genetic system with *A.*
563 *evenia* (2n=20, 400 Mb). In contrast to *A. evenia*, *A. americana* is nodulated only by non-
564 photosynthetic bradyrhizobia, and in this respect, it behaves in a similar way to other legumes.
565 This species is widespread in the tropics, so that adequate quantities of germplasm are
566 available, and it has already been subject to detailed studies, notably to isolate its nodulating
567 *Bradyrhizobium* strains, among which is the DOA9 strain (Noisangiam *et al.*, 2012;
568 Teamtisong *et al.*, 2014). As *A. americana* belongs to the most basal lineage in the
569 *Aeschynomene* phylogeny, it may be representative of the ancestral symbiotic mechanisms
570 found in the genus. On the other hand, *A. patula* has a restricted Madagascan distribution with
571 only one accession being available, but it is of interest due to its relatively smaller plant size
572 and genome size (actually the smallest diploid genome in the group) making this species the
573 “Arabidopsis” of the genus *Aeschynomene*. As for *A. americana*, this species is efficiently
574 nodulated by non-photosynthetic bradyrhizobia, but it is also compatible with the
575 photosynthetic *nod* gene-containing ORS285 strain. This property makes *A. patula*

576 particularly interesting as it allows direct comparisons of mechanisms and pathways between
577 it and *A. evenia* without the problem of potential strain effects on symbiotic responses. In
578 addition, when considering the *Aeschynomene* phylogeny, *A. patula* is more closely related to
579 *A. evenia* than is *A. americana*, and so it may be more suitable to demonstrate the changes
580 necessary to switch a Nod-dependent to a Nod-independent process or vice-versa. Developing
581 sequence resources and functional tools for *A. americana* and/or *A. patula* is now necessary to
582 set up a fully workable comparative *Aeschynomene* system. In the long run, handling such a
583 genetic system will be instrumental in understanding how photosynthetic *Bradyrhizobium* and
584 some *Aeschynomene* species co-evolved and in unravelling the molecular mechanisms of the
585 Nod-independent symbiosis.

586

587

588 **ACKNOWLEDGMENTS**

589 We thank the different seed banks and herbaria for provision of seeds and herbarium vouchers
590 that were used in this study. The present work has benefited from the facilities and expertise
591 of the cytometry facilities of Imagerie-Gif ([http://www.i2bc.paris-](http://www.i2bc.paris-saclay.fr/spip.php?article279)
592 [saclay.fr/spip.php?article279](http://www.i2bc.paris-saclay.fr/spip.php?article279)) and of the molecular cytogenetic facilities of the AGAP
593 laboratory (<http://umr-agap.cirad.fr/en/plateformes/plateau-de-cytogenetique-moleculaire>).
594 This work was supported by a grant from the French National Research Agency (ANR-
595 AeschyNod-14-CE19-0005-01) that financed the design of the study, experimentation and
596 analysis of the data.

597

598

599 **AUTHOR CONTRIBUTIONS**

600 J.F.A. designed the experiments. L.B., C.C., R.R., J.F., M.G.M, S.C.B., C.H., M.D. and J.F.A.
601 performed the experiments and obtained the data. P.S., C.S., L.M. undertook the phylogenetic
602 analyses. P.M., J.Q., G.P.L., X.P., A.D'H., E.G. and J.F.A. analysed the data. M.G., R.D., H.
603 Randriambanona, H. Ramanankierna, H.V. and M.Z. contributed to the acquisition and
604 analysis of accessions. J.F.A. wrote the paper. L.B. and C.C contributed equally. All authors
605 read and approved the final manuscript.

606

607

608 **REFERENCES**

609

- 610 **Alazard D. 1985.** Stem and root nodulation in *Aeschynomene* spp. *Applied and*
611 *Environmental Microbiology* **50**: 732-734.
- 612 **Alazard D, Duhoux E. 1987.** Nitrogen-fixing stem nodules on *Aeschynomene afraspera*.
613 *Biological fertility Soils*. **4**: 61-66.
- 614 **Arrighi JF, Cartieaux F, Brown SC, Rodier-Goud M, Boursot M, Fardoux J, Patrel D,**
615 **Gully D, Fabre S, Chaintreuil C et al. 2012.** *Aeschynomene evenia*, a Model Plant for
616 Studying the Molecular Genetics of the Nod-Independent Rhizobium-Legume Symbiosis.
617 *Molecular Plant-Microbe Interactions* **25**: 851-861.
- 618 **Arrighi JF, Cartieaux F, Chaintreuil C, Brown SC, Boursot M, Giraud E. 2013.**
619 Genotype delimitation in the Nod-independent model legume *Aeschynomene evenia*. *PloS*
620 *One* **8**:e63836.
- 621 **Arrighi JF, Chaintreuil C, Cartieaux F, Cardi C, Rodier-Goud M, Brown SC, Boursot**
622 **M, d’Hont A, Dreyfus B, Giraud E. 2014.** Radiation of the Nod-independent *Aeschynomene*
623 relies on multiple allopolyploid speciation events. *New Phytologist* **201**: 1457-68.
- 624 **Arrighi JF, Cartieaux F. 2015.** Out of water of a new model legume: the Nod-independent
625 *Aeschynomene evenia*. In *Biological Nitrogen Fixation*, Vol 2 Ed. de Bruijn. Wiley-Blackwell
626 publishers. Chapter **45**, 447-54.
- 627 **Behm JE, Geurts R, Kiers ET. 2014.** *Parasponia*: a novel system for studying mutualism
628 stability. *Trends Plant Sci* **19**: 757-63.
- 629 **Boivin C, Ndoye I, Molouba F, De Lajudie P, Dupuy N, Dreyfus B. 1997.** Stem nodulation
630 in legumes: diversity, mechanisms, and unusual characteristics. *Critical Reviews in Plant*
631 *Sciences* **16**: 1-30.
- 632 **Bonaldi K, Gargani D, Prin Y, Fardoux J, Gully D, Nouwen N, Goormachtig S, Giraud**
633 **E. 2011.** Nodulation of *Aeschynomene afraspera* and *A. indica* by photosynthetic
634 *Bradyrhizobium* sp. strain ORS285: The Nod-dependent versus the Nod-independent
635 symbiotic interaction. *M.P.M.I.* **24**: 1359-1371.
- 636 **Bravo A, York T, Pumplun N, Mueller LA, Harrison MJ. 2016.** Genes conserved for
637 arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nat Plants* **18**: 15208.
- 638 **Chaintreuil C, Arrighi JF, Giraud E, Miché L, Moulin L, Dreyfus B, Munive-Hernandez**
639 **J, Villegas-Hernandez M, Béna G. 2013.** Evolution of symbiosis in the legume genus
640 *Aeschynomene*. *New Phytologist* **200**:1247-59.
- 641 **Chaintreuil C, Rivallan R, Bertoli DJ, Klopp C, Gouzy J, Courtois B, Leleux P, Martin**
642 **G, Rami JF, Gully D et al. 2016a.** A gene-based map of the Nod factor-independent

- 643 *Aeschynomene evenia* sheds new light on the evolution of nodulation and legume genomes.
644 *DNA Res* **23**:365-76.
- 645 **Chaintreuil C, Gully J, Hervouet C, Tittabutr P, Randriambanona H, Brown SC, Lewis**
646 **GP, Bourge M, Cartieaux F, Boursot M et al. 2016b.** The evolutionary dynamics of
647 ancient and recent polyploidy in the African semi-aquatic species of the legume genus
648 *Aeschynomene*. *New Phytologist* **211**:1077-91.
- 649 **Chaintreuil C, Perrier X, Martin G, Fardoux J, Lewis GP, Brottier L, Rivallan R,**
650 **Gomez-Pacheco M, Bourges M, Lamy L et al. 2018.** Naturally occurring variations in the
651 nod-independent model legume *Aeschynomene evenia* and relatives: a resource for
652 nodulation genetics. *BMC Plant Biol.* **18**:54.
- 653 **Czernic P, Gully D, Cartieaux F, Moulin L, Guefrachi I, Patrel D, Pierre O, Fardoux J,**
654 **Chaintreuil C, Nguyen P et al. 2015.** Convergent Evolution of Endosymbiont Differentiation
655 in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific
656 Cysteine-Rich Peptides. *Plant Physiol* **169**: 1254-65.
- 657 **Cook, B.G., Pengelly, B.C., Brown, S.D., Donnelly, J.L., Eagles, D.A., Franco, M.A.,**
658 **Hanson, J., Mullen, B.F., Partridge, I.J., Peters, M. and Schultze-Kraft, R. 2005.** Tropical
659 Forages: an interactive selection tool. CSIRO, DPI&F(Qld), CIAT and ILRI, Brisbane,
660 Australia. <http://www.tropicalforages.info>
- 661 **Dereeper A, Homa F, Andres G, Sempere G, Sarah G, Hueber Y, Dufayard JF, Ruiz M.**
662 **2015.** SNIPlay3: a web-based application for exploration and large scale analyses of genomic
663 variations. *Nucleic Acids Res* **43**(W1):W295-300.
- 664 **Delaux PM, Varala K, Edger PP, Coruzzi GM, Pires JC, Ané JM. 2014.** Comparative
665 phylogenomics uncovers the impact of symbiotic associations on host genome evolution.
666 *PLoS Genet* **10**: e1004487.
- 667 **Delaux PM, Radhakrishnan G, Oldroyd G. 2015.** Tracing the evolutionary path to
668 nitrogen-fixing crops. *Curr Opin Plant Biol* **26**: 95-9.
- 669 **Du Puy DJ, Labat J-N, Rabevohitra R, Villiers J-F, Bosser J, Moat J. 2002.** The
670 *Leguminosae* of Madagascar. *Royal Botanic Gardens, Kew.*
- 671 **Fabre S, Gully D, Poitout A, Patrel D, Arrighi JF, Giraud E, Czernic P, Cartieaux F.**
672 **2015.** Nod Factor-Independent Nodulation in *Aeschynomene evenia* Required the Common
673 Plant-Microbe Symbiotic Toolkit. *Plant Physiol* **169**: 2654-64.
- 674 **Geurts R, Xiao TT, Reinhold-Hurek B. 2016.** What Does It Take to Evolve A Nitrogen-
675 Fixing Endosymbiosis? *Trends Plant Sci* **21**:199-208.

- 676 **Gillett JB, Polhill RM, Verdcourt B. 1971.** Leguminosae (part 3): subfamily Papilionoideae
677 (part 1). In Milne-Redhead E & Polhill RM (eds). Flora of Tropical East Africa. *Royal*
678 *Botanic Gardens, Kew.*
- 679 **Giraud E, Hannibal L, Fardoux J, Vermeglio A, Dreyfus B. 2000.** Effect of
680 *Bradyrhizobium* photosynthesis on stem nodulation of *Aeschynomene sensitiva*. *P.N.A.S.* **97**:
681 14795-14800.
- 682 **Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D,**
683 **Cartieaux F, Prin Y, et al. 2007.** Legumes symbioses: absence of Nod genes in
684 photosynthetic bradyrhizobia. *Science* **316**: 1307-1312.
- 685 **Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin**
686 **G, Charron C, Hervouet C et al. 2018.** A mosaic monoploid reference sequence for the
687 highly complex genome of sugarcane. *Nature Communications* **9**:2638.
- 688 **Govindarajulu R, Hughes CE, Alexander PJ, Bailey CD. 2011.** The complex evolutionary
689 dynamics of ancient and recent polyploidy in *Leucaena* (Leguminosae; Mimosoideae). *Am J*
690 *Bot* **98**:2064-76.
- 691 **Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B,**
692 **Lauressergues D, Keller J, Imanishi L et al. 2018.** Phylogenomics reveals multiple losses
693 of nitrogen-fixing root nodule symbiosis. *Science* **361**(6398). pii: eaat1743.
- 694 **Hagerup O. 1928.** En hygrofil baelplante (*Aeschynomene aspera* L.) med bakterieknolde
695 paa staenglen. *Dansk Bot. Arkiv.* **14**: 1-9.
- 696 **Howieson J.G. and Dilworth M.J. (Eds.). 2016.** Working with rhizobia. Australian Centre
697 for International Agricultural Research: Canberra.
- 698 **Katoh K, Standley DM. 2013.** MAFFT multiple sequence alignment software version 7:
699 improvements in performance and usability. *Mol Biol Evol* **30**:772-80.
- 700 **Klitgaard, B.B. & Lavin, M. 2005.** Tribe Dalbergieae *sens. lat.* In G. Lewis *et al.* (eds.)
701 Legumes of the World. *Royal Botanic Gardens, Kew.* 307-335.
- 702 **Lartillot N, Philippe H. 2004.** A Bayesian Mixture Model for Across-Site Heterogeneities in
703 the Amino-Acid Replacement Process. *Mol Biol Evol* **21**:1095-1109.
- 704 **Lavin M, Pennington RT, Klitgaard BB, Sprent JI, de Lima HC, Gasson PE. 2001.** The
705 dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *Am J Bot*
706 **88**: 503-533.
- 707 **Lewis G, Schrire B, Mackinder B, Lock M. 2005.** Legumes of the world. *Royal Botanic*
708 *Gardens, Kew.*
- 709 **Lock JM. 1989.** Legumes of Africa. A check-list. *Royal Botanic Gardens, Kew, England.*

- 710 **Loureiro MF, De Faria SM, James EK, Pott A, Franco AA. 1994.** Nitrogen-Fixing Stem
711 Nodules of the Legume, *Discolobium Pulchellum* Benth. *New Phytologist* **128**: 283-295.
- 712 **Loureiro MF, James EK, Sprent JI, Franco AA. 1995.** Stem and root nodules on the
713 tropical wetland legume *Aeschynomene fluminensis*. *New Phytologist* **130**: 531-544.
- 714 **LPWG (The Legume Phylogeny Working Group). 2017.** A new subfamily classification of
715 the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* **66**: 44-77.
- 716 **Madsen LH, Tirichine L, Jurkiewicz A, Sullivan JT, Heckmann AB, Bek AS, Ronson
717 CW, James EK, Stougaard J. 2010.** The molecular network governing nodule
718 organogenesis and infection in the model legume *Lotus japonicus*. *Nature Communication* **1**:
719 10.
- 720 **Miché L, Moulin L, Chaintreuil C, Contreras-Jimenez JL, Munive-Hernandez JA, Del
721 Carmen Villegas-Hernandez M, Crozier F, Béna G. 2010.** Diversity analyses of
722 *Aeschynomene* symbionts in Tropical Africa and Central America reveal that nod-independent
723 stem nodulation is not restricted to photosynthetic bradyrhizobia. *Environ Microbiol* **12**:
724 2152-2164.
- 725 **Molouba, F, Lorquin J, Willems A, Hoste B, Giraud E, Dreyfus B, Gillis M, DeLajudie P
726 and Masson-Boivin C. 1999.** "Photosynthetic bradyrhizobia from *Aeschynomene* spp. are
727 specific to stem-nodulated species and form a separate 16S ribosomal DNA restriction
728 fragment length polymorphism group." *Applied and Environmental Microbiology* **65**: 3084-
729 3094.
- 730 **Noisangiam R, Teamtisong K, Tittabutr P, Boonkerd N, Toshiki U, Minamisawa K,
731 Teaumroong N. 2012.** Genetic diversity, symbiotic evolution, and proposed infection process
732 of *Bradyrhizobium* strains isolated from root nodules of *Aeschynomene americana* L. in
733 Thailand. *Appl Environ Microbiol* **78**:6236-50.
- 734 **Nouwen N, Arrighi JF, Cartieaux F, Gully D, Klopp C, Giraud E. 2017.** The role of
735 rhizobial (NifV) and plant (FEN1) homocitrate synthesis in *Aeschynomene* - photosynthetic
736 *Bradyrhizobium* symbiosis. *Scientific Reports*, **7**: 448.
- 737 **Okazaki S, Kaneko T, Sato S, Saeki K. 2013.** Hijacking of leguminous nodulation signaling
738 by the rhizobial type III secretion system. *P.N.A.S.* **110**:17131-6.
- 739 **Okazaki S, Tittabutr P, Teulet A, Thouin J, Fardoux J, Chaintreuil C, Gully D, Arrighi
740 JF, Furuta N, Miwa H et al. 2015.** Rhizobium-legume symbiosis in the absence of Nod
741 factors: two possible scenarios with or without the T3SS. *ISME J* **10**: 64-74.

- 742 **Okubo T, Fukushima S, Minamisawa K. 2012.** Evolution of *Bradyrhizobium-*
743 *Aeschynomene* mutualism: living testimony of the ancient world or highly evolved state?
744 *Plant & Cell Physiology* **53**:2000-2007.
- 745 **Okuyama Y, Tanabe AS, Kato M. 2012.** Distinguishing ancient allotetraploidization in Asian
746 *Mitella*: an integrated approach for multilocus combinations. *Mol Biol Evol* **29**: 429-39.
- 747 **Revell LJ. 2012.** Phytools: An R package for phylogenetic comparative biology (and other
748 things). *Methods Ecol. Evol* **3**: 217-223.
- 749 **Oueslati A, Salhi-Hannachi A, Luro F, Vignes H, Mournet P, Ollitrault P. 2017.**
750 Genotyping by sequencing reveals the interspecific *C. maxima* / *C. reticulata* admixture along
751 the genomes of modern citrus varieties of mandarins, tangors, tangelos, orangelos and
752 grapefruits. *PLoS One* **12**:e0185618.
- 753 **Rudd VE. 1955.** The American species of *Aeschynomene*. *Contributions of the United States*
754 *National Herbarium* **32**: 1-172.
- 755 **Rudd VE 1981.** *Aeschynomeneae*. In: Polhill RM, Raven PH eds. *Advances in legume*
756 *systematics*. Kew: Royal Botanic Gardens. **Part1**: 347-354.
- 757 **Sprent JI, James EK. 2008.** Legume-rhizobial symbiosis: an anorexic model? *New*
758 *Phytologist* **179**: 3-5.
- 759 **Sprent JI, Ardley J, James EK. 2017.** Biogeography of nodulated legumes and their
760 nitrogen-fixing symbionts. *New Phytologist* doi: 10.1111/nph.14474.
- 761 **Teamtisong K, Songwattana P, Noisangiam R, Piromyou P, Boonkerd N, Tittabutr P,**
762 **Minamisawa K, Nantagij A, Okazaki S, Abe M, Uchiumi T, Teaumroong N. 2014.**
763 Divergent nod-containing *Bradyrhizobium* sp. DOA9 with a megaplasmid and its host range.
764 *Microbes Environ.* **29**: 370-6.
- 765 **To TH, Scornavacca C. 2015.** Efficient algorithms for reconciling gene trees and species
766 networks via duplication and loss events. *BMC Genomics* **16** Suppl 10:S6.
- 767 **van Velzen R, Holmer R, Bu F, Rutten L, van Zeijl A, Liu W, Santuari L, Cao Q,**
768 **Sharma T, Shen D et al. 2018.** Comparative genomics of the nonlegume Parasponia reveals
769 insights into evolution of nitrogen-fixing rhizobium symbioses. *PNAS* **115**:E4700-E4709.
- 770 **Whitfield JB, Lockhart PJ. 2007.** Deciphering ancient rapid radiations. *Trends in Ecology*
771 *and Evolution* **22**: 258-265.

772

773

774 **FIGURE LEGENDS**

775

776 **Figure 1: Phylogeny of the genus *Aeschynomene* and allied genera.**

777 The Bayesian phylogenetic reconstruction was obtained using the concatenated *ITS* (Internal
778 Transcribed Spacer) + *matK* sequences. Numbers at branches indicate posterior probability
779 above 0.5. The five main lineages are identified with a circled number and the two previously
780 studied *Aeschynomene* groups are framed in a red box bordered with a dashed line. On the
781 right are listed *Aeschynomene* subgenus *Aeschynomene* (in green), other *Aeschynomene*
782 subgenera or species groups (in blue) and related genera (in orange).

783

784 **Figure 2: Genomic characteristics and phylogenetic relationships.**

785 (a) Simplified Bayesian *ITS+matK* phylogeny with representative species of different lineages
786 and groups. The *A. evenia*, *A. afraspera* and BRH (*Bakerophyton-Rueppelia-Humularia*)
787 clades are represented by black triangles and the polytomy is depicted in bold. Chromosome
788 numbers are indicated in brackets. (b) Phylogenetic relationships based on the combination of
789 4 concatenated nuclear low-copy genes (*CYP1*, *eif1a*, *SuSy* and *TIP1;1* genes detailed in
790 Figure S5). Diploid species ($2n=20$) are in blue, polyploid species ($2n\geq 28$) in black. The A
791 and B subgenomes of the polyploid taxa are delineated by red and green boxes in dashed
792 lines, respectively. Nodes with a posterior probability inferior to 0.5 were collapsed into
793 polytomies. Posterior probability above 0.5 are indicated at every node. (c) The one-
794 allopolyploidisation hypothesis (N1-best) obtained with the phylogenetic network analysis
795 based on the T2 tree with reticulations in blue (detailed in Fig. S7).

796

797 **Figure 3: Occurrence of adventitious root primordia and of stem nodulation.**

798 (a) Simplified Bayesian *ITS+matK* phylogeny of the whole group with the *A. evenia*, *A.*
799 *afraspera* and BRH (*Bakerophyton-Rueppelia-Humularia*) clades represented by black
800 triangles. The polytomy is depicted in bold. The shared presence of adventitious root
801 primordia is depicted on the stem by a blue circle. Dashed red boxes indicate groups
802 comprising aerial stem-nodulating species. Asterisks refer to illustrated species in (b) for
803 aerial stem-nodulation. (b) Stems of representatives for the different lineages and groups.
804 Small spots on the stem correspond to dormant adventitious root primordia and stem nodules
805 are visible on the species marked by an asterisk. Bars: 1cm.

806

807 **Figure 4: Comparison of the root nodulation properties.**

808 (a) Species of different lineages and groups that were tested for nodulation are listed in the
809 simplified Bayesian phylogeny on the left. Root nodulation tests were performed using the

810 DOA9, ORS285, ORS285 Δ *nod* and ORS278 strains. E, effective nodulation; e, partially
811 effective nodulation; i, ineffective nodulation, -, no nodulation; blank, not tested. (b) Number
812 of nodules per plant, (c) relative acetylene-reducing activity (ARA) and (d) aspect of the
813 inoculated roots developing nodules or not (some nodules were cut to observe the
814 leghemoglobin color inside) after inoculation with *Bradyrhizobium* DOA9, ORS285 and
815 ORS278 on *A. americana*, *A. patula*, *A. afraspera* and *A. evenia*. Error bars in (b) and (c)
816 represent s.d. (n=6). Scale bar in (d): 1 mm.

817

818 **Figure 5: Characteristics of diploid species.**

819 (a) Development and germplasm data for species that are listed in the simplified phylogeny on
820 the left. *A. evenia* from the Nod-independent clade (NI) is also included for comparison.
821 Germplasm numbers correspond to the sum of accessions available at CIAT, USDA, Royal
822 Botanic Gardens (Kew), AusPGRIS, IRRI and at LSTM. (b) Multi-dimensional scaling
823 (MSD) plots of the genetic diversity among *A. americana* (left) and *A. villosa* (right)
824 accessions according to coordinates 1 and 2 (C1, C2). Identified groups are delimited by
825 circles and labeled with numbers. (c) Geographical distribution of the of *A. americana* and
826 *A. villosa* accessions. Taxon colours and group numbers are the same as in (b). Details of the
827 accessions are provided in Table S4.

828

829 **Figure S1: *matK* phylogeny of the genus *Aeschynomene* and allied genera.**

830 Bayesian phylogenetic reconstruction obtained using the chloroplast *matK* gene. Numbers at
831 branches are posterior probability.

832

833 **Figure S2: *ITS* phylogeny of the genus *Aeschynomene* and allied genera.**

834 Bayesian phylogenetic reconstruction obtained using the Internal Transcribed Spacer (*ITS*)
835 sequence. Numbers at branches are posterior probability.

836

837 **Figure S3: Chromosome numbers in *Aeschynomene* species.**

838 Root tip metaphase chromosomes stained in blue with DAPI (4',6-diamidino-2-phenylindole).
839 Chromosome numbers are indicated in brackets. Scale bars: 5 μ m.

840

841 **Figure S4: Chromosome numbers in species of genera related to *Aeschynomene***

842 Root tip metaphase chromosomes stained in blue with DAPI (4',6-diamidino-2-phenylindole).
843 Chromosome counts are indicated in brackets. Scale bars: 5 μ m.

844

845 **Figure S5: Phylogenetic trees based on nuclear low-copy genes.**

846 Bayesian phylogenetic reconstructions obtained for the *CYP1*, *eif1a*, *SuSy* and *TIP1;1* genes.
847 Diploid species ($2n=20$) are in blue, polyploid species ($2n\geq 28$) in black excepted *A. afraspera*
848 for which the A and B gene copies are distinguished in red and green respectively. -A, -A1, -
849 A2, -B, -B1 and -B2 indicated the different copies found. Putative A and B subgenomes of the
850 polyploid taxa are delineated by red and green boxes in dashed lines, respectively. Numbers at
851 branches represent posterior probability.

852

853 **Figure S6: Ancestral state reconstruction of ploidy levels in the genus *Aeschynomene***
854 **and allied genera.**

855 Ancestral state reconstruction was estimated in SIMMAP software using the 50% majority-
856 rule topology obtained by Bayesian analysis of the combined *ITS+matK* sequences. Ploidy
857 levels are indicated by different colors. Unknown ploidy levels are denoted by a dash.

858

859 **Figure S7: Phylogenetic networks based on the four nuclear *CYP1*, *eif1a*, *SuSy* and**
860 ***TIP1;1* genes.**

861 (a) No-allopolyploidisation hypothesis (T1) based on the concatenated gene tree obtained
862 taking into account the group A (Fig. 2b). (b) No-allopolyploidisation hypothesis (T2) based
863 on the concatenated gene tree obtained taking into account the group B (Fig. 2b). (c) One-
864 allopolyploidisation hypothesis (N1-best). (d) Two-allopolyploidisation hypothesis (N2-best).
865 Blue lines indicate reticulations while other nodes of the network are associated to speciation
866 events. Reconciliation scores obtained for each phylogenetic network are indicated.

867

868 **Figure S8: Ancestral state reconstruction of adventive root primordia in the genus**
869 ***Aeschynomene* and allied genera.**

870 Ancestral state reconstruction was estimated in SIMMAP software using the 50% majority-
871 rule topology obtained by Bayesian analysis of the combined *ITS+matK* sequences. Data on
872 the adventitious root primordia come from the present analysis and pertinent previously
873 published data. Presence or not of adventitious root primordia is indicated by different colors.

874

875 **Figure S9: Ancestral state reconstruction of ecological habit in the genus *Aeschynomene***
876 **and allied genera.**

877 Ancestral state reconstruction was estimated in SIMMAP software using the 50% majority-
878 rule topology obtained by Bayesian analysis of the combined *ITS+matK* sequences. Data on
879 the species ecology come from pertinent previously published data. Ecological habits are
880 indicated by different colors.

881

882 **Figure S10: Ancestral state reconstruction of the aerial stem nodulation ability in the**
883 **genus *Aeschynomene* and allied genera.**

884 Ancestral state reconstruction was estimated in SIMMAP software using the 50% majority-
885 rule topology obtained by Bayesian analysis of the combined *ITS+matK* sequences. Data on
886 the occurrence of stem nodulation come from pertinent previously published data. Occurrence
887 or not of stem nodulation is indicated by different colors.

888

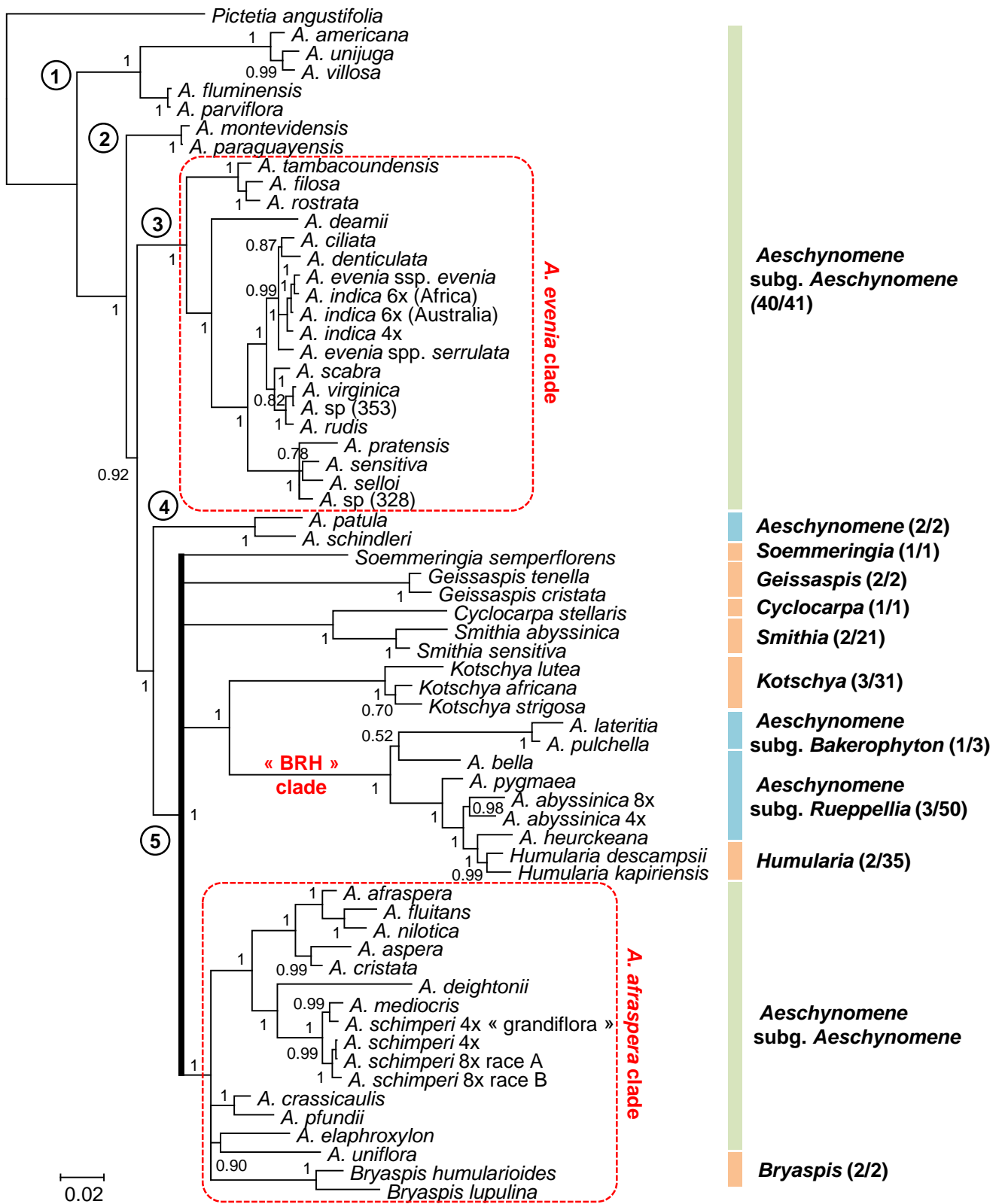
889 **Figure S11: Ancestral state reconstruction of the ability to nodulate with the**
890 **photosynthetic *Bradyrhizobium* strains in the genus *Aeschynomene* and allied genera.**

891 Ancestral state reconstruction was estimated in SIMMAP software using the 50% majority-
892 rule topology obtained by Bayesian analysis of the combined *ITS+matK* sequences. Data on
893 nodulation with photosynthetic *Bradyrhizobium* strains come from the present analysis and
894 pertinent previously published data. Nodulation with photosynthetic *Bradyrhizobium* strains is
895 considered positive only if reported as occurring naturally or being efficient *in vitro*.

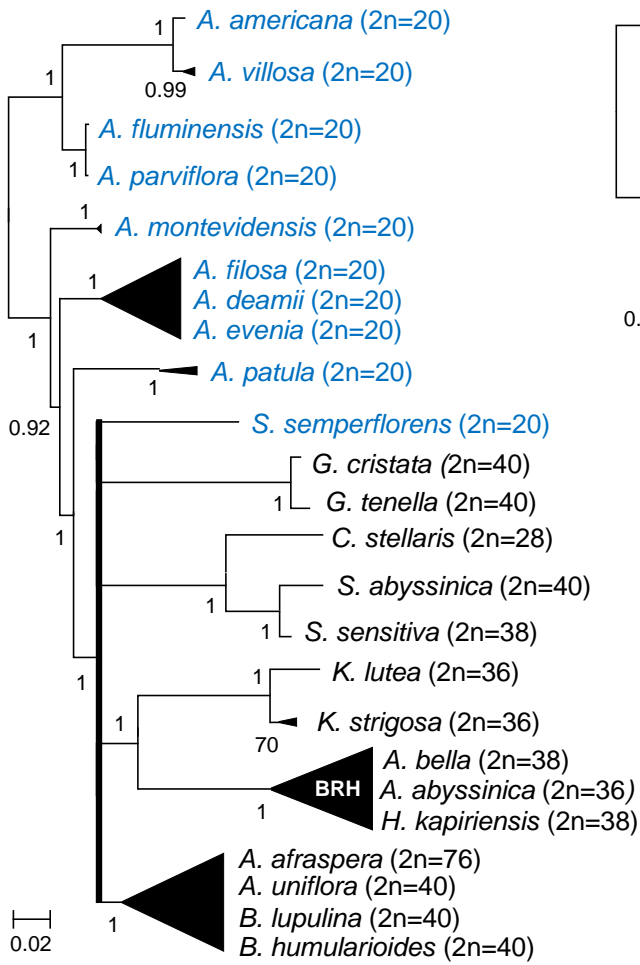
896

897 **Figure S12: Ancestral state reconstruction of the ability to nodulate with the**
898 **photosynthetic *Bradyrhizobium* strain ORS278 in the genus *Aeschynomene* and allied**
899 **genera.**

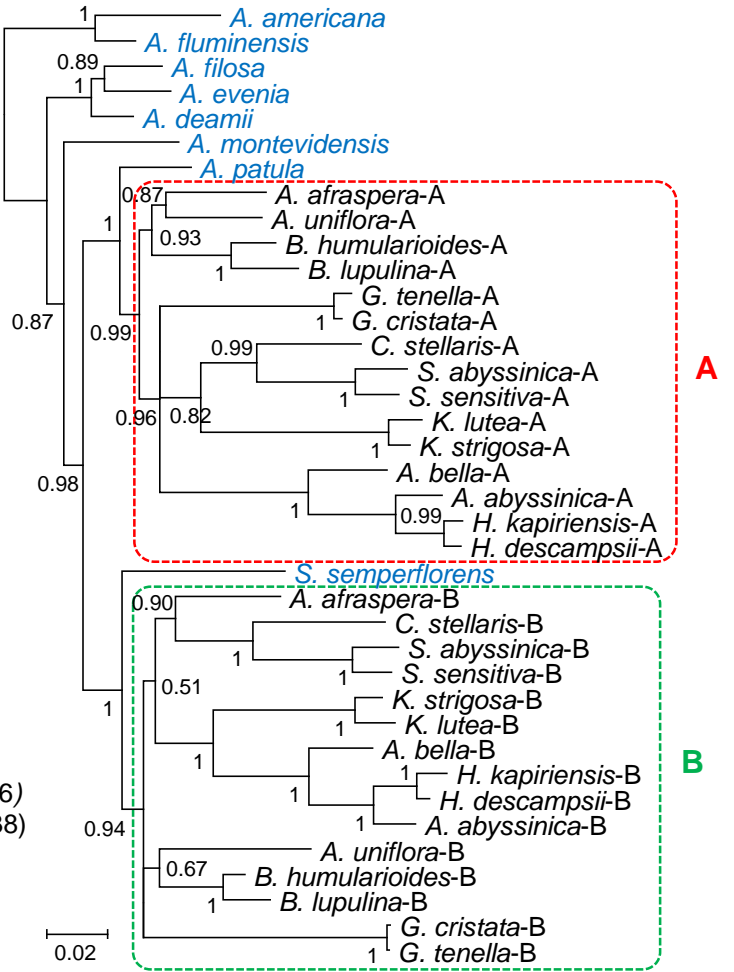
900 Ancestral state reconstruction was estimated in SIMMAP software using the 50% majority-
901 rule topology obtained by Bayesian analysis of the combined *ITS+matK* sequences. Data on
902 nodulation with ORS278 come from the present analysis and pertinent previously published
903 data. Ability or not to nodulate with ORS278 is indicated by different colors.



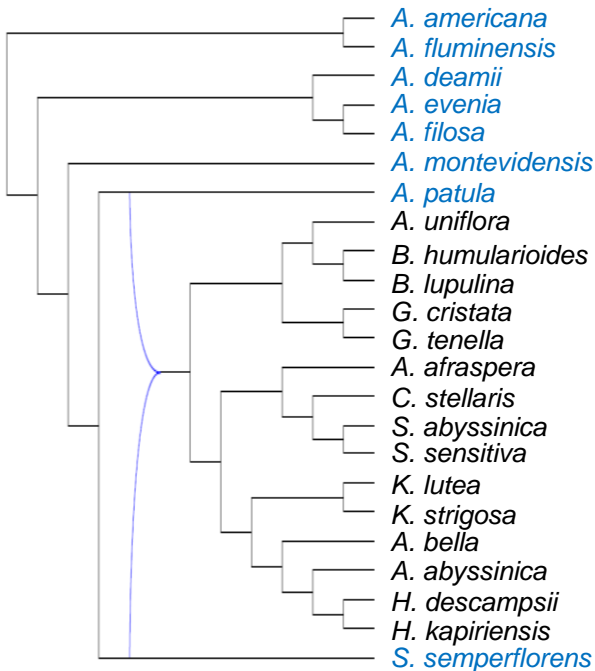
a ITS + matK

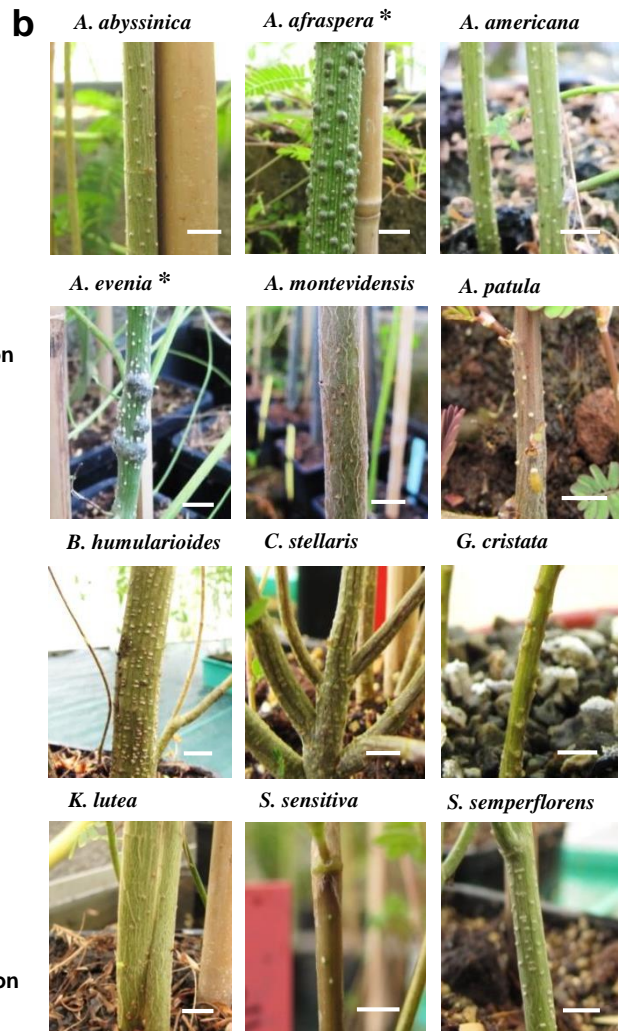
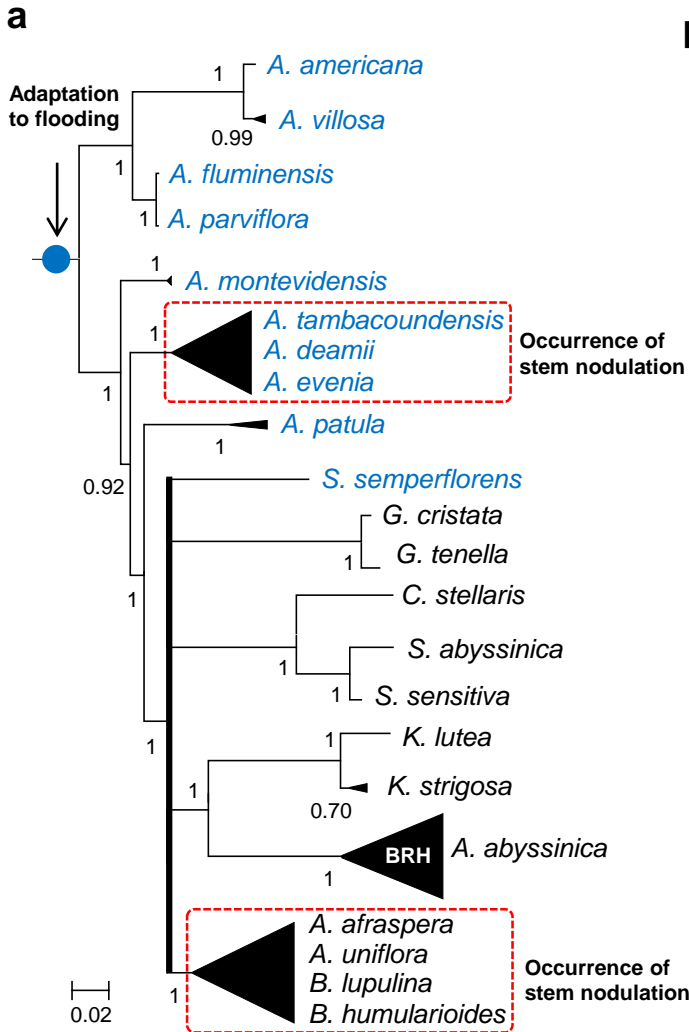


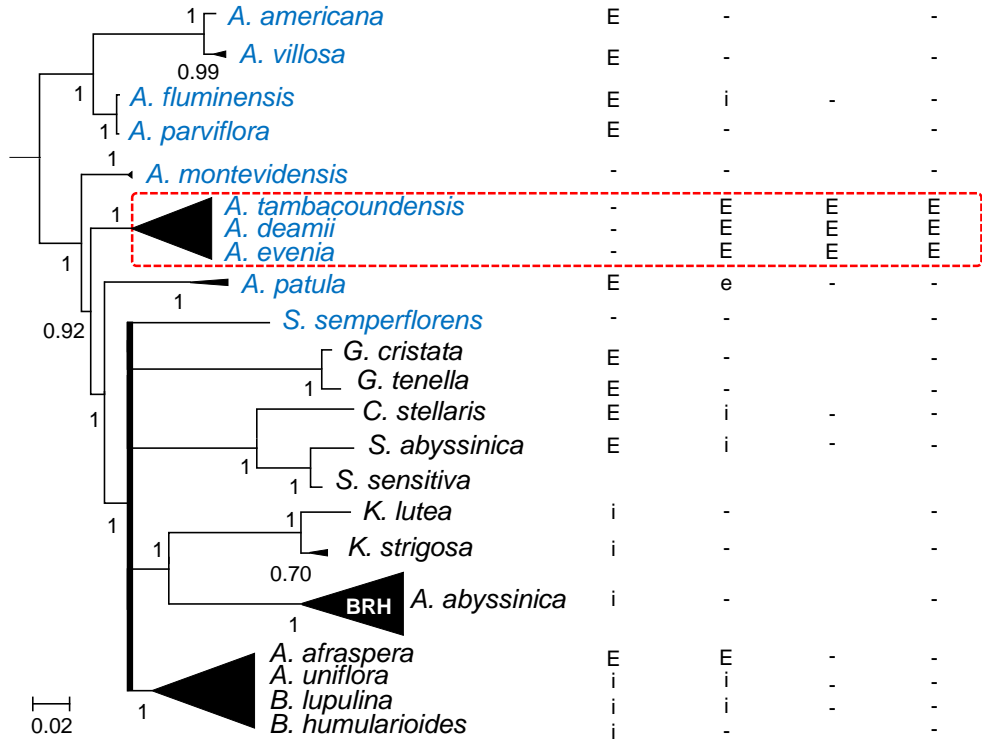
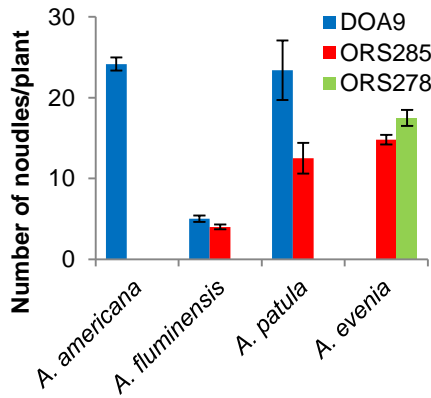
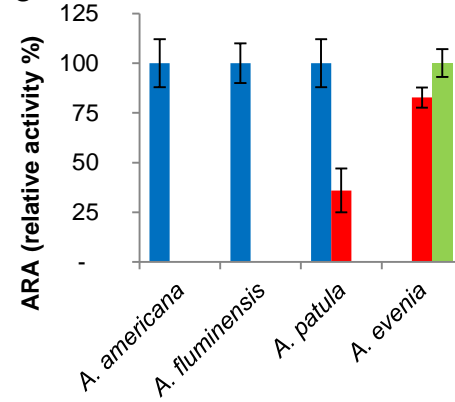
b Concatenated nuclear genes



C Phylogenetic network





a**b****c****d**