DNA Research, 2016, 23(4), 365–376
doi: 10.1093/dnares/dsw020
Advance Access Publication Date: 13 June 2016
Full Paper

Full Paper

A gene-based map of the Nod factor-independent *Aeschynomene evenia* sheds new light on the evolution of nodulation and legume genomes

Clémence Chaintreuil[1], Ronan Rivallan[2], David J. Bertioli[3], Christophe Klopp[4], Jérôme Gouzy[5], Brigitte Courtois[2], Philippe Leleux[1,4], Guillaume Martin[2], Jean-François Rami[2], Djamel Gully[1], Hugues Parrinello[6], Dany Séverac[6], Delphine Patrel[1,7], Joël Fardoux[1], William Ribière[7], Marc Boursot[1], Fabienne Cartieaux[1], Pierre Czernic[1], Pascal Ratet[8,9], Pierre Mournet[2], Eric Giraud[1], and Jean-François Arrighi[1,*]

[1]IRD, UMR LSTM, Campus International de Baillarguet, F-34398 Montpellier, France, [2]CIRAD, UMR AGAP, Campus de Lavalette, F-34398 Montpellier, France, [3]University of Brasília, Institute of Biological Sciences, Campus Darcy Ribeiro, 70910-900 Brasília, DF, Brazil, [4]INRA, Plateforme GenoToul Bioinfo, UR 875, INRA Auzeville, F-31326 Castanet-Tolosan, France, [5]INRA, UMR441 LIPM, INRA Auzeville, F-31326 Castanet-Tolosan, France, [6]MGX-Montpellier GenomiX, Institut de Génomique Fonctionnelle, F-34094 Montpellier, France, [7]IRD, Centre IRD de Montpellier France Sud, F-34394 Montpellier, France, [8]Institute of Plant Sciences Paris Saclay IPS2, CNRS, INRA, Université Paris-Sud, Université Evry, Université Paris-Saclay, 91405 Orsay, France and [9]Institute of Plant Sciences Paris-Saclay IPS2, Paris Diderot, Sorbonne Paris-Cité, 91405 Orsay, France

*To whom correspondence should be addressed. Tel. +33 04 67 59 38 82. Fax. +33 04 67 59 38 02.
Email: jean-francois.arrighi@ird.fr

Edited by Dr Sachiko Isobe

Received 11 March 2016; Accepted 2 May 2016

Abstract

Aeschynomene evenia has emerged as a new model legume for the deciphering of the molecular mechanisms of an alternative symbiotic process that is independent of the Nod factors. Whereas most of the research on nitrogen-fixing symbiosis, legume genetics and genomics has so far focused on Galegoid and Phaseolid legumes, *A. evenia* falls in the more basal and understudied Dalbergioid clade along with peanut (*Arachis hypogaea*). To provide insights into the symbiotic genes content and the structure of the *A. evenia* genome, we established a gene-based genetic map for this species. Firstly, an RNAseq analysis was performed on the two parental lines selected to generate a F_2 mapping population. The transcriptomic data were used to develop molecular markers and they allowed the identification of most symbiotic genes. The resulting map comprised 364 markers arranged in 10 linkage groups ($2n = 20$). A comparative analysis with the sequenced genomes of *Arachis duranensis* and *A. ipaensis*, the diploid ancestors of peanut, indicated blocks of conserved macrosynteny. Altogether, these results provided important clues regarding the evolution of symbiotic genes in a Nod factor-independent context. They provide a

basis for a genome sequencing project and pave the way for forward genetic analysis of symbiosis in *A. evenia*.

Key words: *Aeschynomene evenia*; symbiosis; legume; genetic map; comparative genomics

1. Introduction

Fabaceae, or legumes, account for ~27% of the world's primary crop production. The capacity of most legumes to form nitrogen fixing nodules with rhizobia underlies their importance as a source of protein for human and animal diet and of nitrogen in both natural and agricultural ecosystems.[1] Among the 20,000 legume species, around 14,000 fall into the Papilionoideae subfamily that is divided into four main clades Genistoids, Dalbergioids, Phaseoloids and Galegoids (Fig. 1).[7]

To advance in the understanding of how legumes have evolved to nitrogen-fixing symbiosis with rhizobia, two model species belonging to the Galegoid clade and with favorable genetic attributes, *Medicago truncatula* and *Lotus japonicus*, have been extensively studied. This study has revealed a sophisticated symbiotic process that is triggered by the recognition of rhizobial signal molecules called Nod factors (NF) and that involves both the formation of an infection thread guiding the bacteria inside the root and a distant induction of a nodule primordium. Furthermore, the genetic dissection of the nodulation process has allowed the identification and elucidation of the role of many genes that are essential for the different steps of nodule development, some of them being also involved in mycorrhizal symbiosis.[8] Such mechanisms are likely to be highly conserved in the Phaseolid and Galegoid clades that share similar symbiotic infection and nodule organogenesis processes (Fig. 1). In addition, since most of the legumes used as crops are members of the Phaseoloids and Galegoids, genetic and genomic studies, including genome sequencing and comparative genomics, have greatly advanced in these two clades.[9–17]

Conversely, the Genistoid and the Dalbergioid clades, which are more basal in their divergence within the Papilionoideae, have lagged far behind, even though they contain important crop legumes such as the genistoid lupine (*Lupinus angustifolius*) and the dalbergioid peanut (*Arachis hypogaea*). This is partly due to the large and complex nature of their genomes that makes their study challenging. Hence, in the case of peanut, which is a recent allotetraploid species, the choice was made to sequence first the genomes of its two diploid progenitors, *Arachis duranensis* and *A. ipaensis*.[18] Similarly, genistoid and dalbergioid legumes have been understudied for their symbiotic properties although they display a distinct infection process that is initiated in an intercellular fashion and by the formation of nodules that originate directly from dividing infected cells (Fig. 1).[2] But most surprisingly, in some *Aeschynomene* species, which are phylogenetically related to *Arachis*, an unconventional symbiotic process has also been described where some *Bradyrhizobium* strains are able to form nodules in the absence of NF synthesis (Fig. 1).[19] Unravelling the molecular mechanisms of this NF-independent process would bring important insights on the evolution of rhizobium-legume symbiosis.[20] For this purpose, *Aeschynomene evenia* was proposed as a new model legume because of its advantageous genetic and developmental characteristics for molecular genetics.[21,22]

To provide insights into the symbiotic gene content and on the structure of the NF-independent *A. evenia* genome, we undertook the development of a gene-based genetic map. RNAseq data obtained for each parental line were mined for symbiotic gene discovery and molecular marker development. These markers were used to

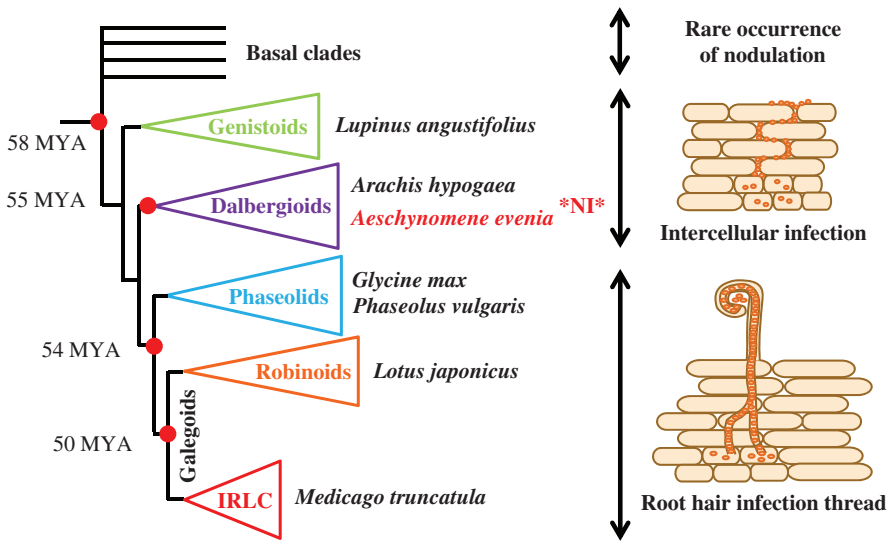

Figure 1. Phylogeny of the Papilionoid lineage and evolution of the symbiotic infection process. The tree is a simplified representation with triangle representing the major clades and the two subclades of the Galegoids: the Robinoids and the plastid DNA inverted repeat-lacking clade (IRLC). Known divergence times are indicated in MYA (million years ago) for some nodes (circles) and notable species are listed for each clade. Symbiotic infection processes encountered in the different clades are reported, excepted for the basal clades for which occurrence of nodulation is restricted to few understudied species. Note that *Aeschynomene* is the only legume genus where a Nod-independent symbiotic infection process (*NI*) has been reported for several species, including *A. evenia*. Phylogeny and drawings are adaptations, date estimates and nodulation properties come from previous publications.[2-6]

genotype a F_2 mapping population and construct a high-density genetic map that was subjected to comparative analysis with the *Arachis duranensis* and *A. ipaensis* genomes.

2. Materials and methods

2.1. Plant material, culture, crossing and observation

Seed germination, plant culture and hybridizations were performed as indicated.[21] To develop the mapping population, individual seeds of the *A. evenia* accessions CIAT8232 (Brazil) and CIAT22838 (Malawi) obtained from CIAT (Colombia) were selfed three times to generate inbred lines and then crossed manually. The hybrid nature of the F_1 plants was confirmed with molecular markers as described.[22] Two F_1 plants obtained by bi-directional hybridizations were selfed to develop the F_2 mapping population in greenhouse. For pollen viability analysis, buds were fixed in Carnoy's fixative prior to anthesis, when pollen was mature but anthers non-dehiscent, and then stained using a simplified method of the Alexander's stain as detailed.[3] Pollen viability was scored under the light microscope for at least three flowers per plant by counting aborted pollen grains, which stained pale turquoise blue and nonaborted pollen grains, which stained dark blue or purple.

2.2 DNA and RNA isolation

Genomic DNA was extracted from young leaves using the CTAB method with the addition of β-mercaptoethanol 2% and PVPP 2% to the CTAB solution in order to limit polysaccharides and polyphenols co-extraction.[21] DNA quality and quantity were evaluated in 1% agarose gel electrophoresis and by spectrophotometer before DNA normalization to a concentration of 10 ng/µl.

For RNA extractions, tissue samples were collected from *in vitro* cultured plants: roots and leaves on 7-days old un-inoculated plants, nodules at 4, 7 and 14-days after inoculation with the *Bradyrhizobium* strain ORS278. Total RNA was extracted using the SV total RNA Isolation System (Promega) excepted for leaves for which RNA was extracted using a CTAB protocol.[23] RNA was quantified using a NanoDrop ND-1000 spectrophotometer and its quality verified using a 2100 Bioanalyzer RNA Nanochip (Agilent, Santa Clara, CA, USA). For each parental line, a total of 12 µg of RNA was pooled equally from the five tissues for Illumina library construction.

2.3 Development of Illumina transcriptomes

Two mRNA libraries were built and sequenced in one lane as described.[24] The Illumina paired-end sequencing technology generated 2× 100 bp independent reads from the 200 bp insert libraries. RNA-seq data were processed for the *de novo* transcriptome assemblies using the Velvet 1.2.07 and CAP3 08/06/13 softwares and orthologous relationships between contigs of the two transcriptome assemblies were identified by best reciprocal hit search. To evaluate the quality of the assembly, all usable reads were realigned to the contigs using BWA-MEM 0.7.12-r1039 and SAMtools idxstats 1.1. To assess the depth of gene coverage through comparative genome analysis, ORFs were searched in *A. evenia* transcripts with TransDcoder_r20140704 and used for BLASTX alignment (E value $< 10^{-20}$) with Blastall 2.2.26 onto the protein database of *A. duranensis* (Araip. V14167.a1.M1.peptide.fa file downloaded from www.peanut.org). Raw Illumina sequence data have been deposited in the NCBI Sequence Read Archive (SRA) under the accession numbers SRR3276128 and SRR3285082, for the CIAT8232 and CIAT22838 lines respectively.

2.4 Sequence search and comparison

Symbiotic gene sequences were obtained from *M. truncatula* and *L. japonicus* by searching in the NCBI database. These sequences were used to find orthologous unigenes in an *A. evenia* local database developed with a Galaxy work bench (INRA, Toulouse, France) (https://bbric-pipelines.toulouse.inra.fr/galaxy/) and containing the CIAT8232 and CIAT22838 transcriptome data (Supplementary Tables S1, S3; Files S1, S2).

For estimation of divergence time, single-copy genes identified in the CIAT8232 and CIAT22838 transcriptomes were used to obtain orthologous sequences for *A. evenia* IRLFL6945 from its transcriptome database (http://esttik.cirad.fr) and for *A. duranensis* and *A. ipaensis* from their genome sequence database (http://peanutbase.org/) (Supplementary Table S5). Coding sequences were cropped, aligned and used to calculate Ks values with the DnaSP 5.10 software (Supplementary Fig. S1). Divergence times were obtained based on a simple molecular clock assumption and by using the previously known ~49 MYA divergence time estimate between *Aeschynomene* and *Arachis*.[6]

2.5 SSR and INDEL genotyping

Detection of microsatellites in the transcriptome data were performed using the SPUTNIK software by searching for di- to hexanucleotide SSRs with a minimum of three to six repetitions depending on the motifs. Orthologous sequences in the two transcriptome assemblies comprising different lengths of the same SSR motif, with a contig length over 500 nt and with at least 50 nt surrounding the SSR motifs were selected as *in silico* polymorphic SSRs (Supplementary Table S1). For specific gene mapping, SSR and INDELs were searched by sequence comparison from the two transcriptomes. Primer pairs were designed using Primer3 with tree parameters: (i) a primer length of 20 bp with an annealing temperature of 55 °C, (ii) three ranges of PCR product size of 90–120, 150–180 and 220–250 bp, (iii) the addition of a 5′-end M13 tail (5′-CACGACGTTGTAAAACGAC-3′) on the forward primer (Supplementary Table S2). SSR markers were checked for PCR-amplification and co-dominant behavior, organized in 12-SSR multiplexes by combining both the three ranges of fragment sizes and the four fluorochromes (FAM, VIC, NED and PET), and then used to genotype the F_2 mapping population as described.[25]

2.6 SNP genotyping

To detect single nucleotide polymorphisms for a set of target genes, sequences of orthologous contigs identified in the two transcriptome datasets were aligned with the MUSCLE program (Supplementary Table S3). Positions of the SNPs were determined by examining the intron/exon structure of orthologous genes in *Arachis* (http://peanutbase.org/). Whenever necessary, additional SNPs were searched by intron sequencing. Up to three SNPs were selected for each target gene and flanking sequences 50 bp both upstream and downstream each target SNP were sent to KBioscience for primer design (Supplementary Table S4). Genotyping of the F_2 mapping population was performed with the KASP genotyping chemistry (KBioscience Ltd., Hoddesdon, UK) on Fluidigm 96.96 Dynamic Arrays using the BioMark HD system (Fluidigm Corp, San Francisco, CA). On the basis of obtained fluorescence, allele call data were viewed

graphically as a scatter plot for each marker assayed by using Fluidigm SNP Genotyping Analysis software and checked manually.

2.7 Genetic map construction

Linkage analysis was performed on segregated genotypic data from the F_2 mapping population using JoinMap® version 4 (Plant Research International BV, Wageningen, Netherlands). The marker loci were attributed to linkage groups using the JoinMap® grouping module with a logarithm of odds (LOD) scores of 6.0. Marker order and genetic distance (in cM) were calculated using a regression mapping algorithm with the following parameters: Kosambi's mapping function, recombination frequency ≤ 0.40, and LOD threshold of 1.0 (Supplementary Table S6). The segregation data were tested for deviation from the expected Mendelian ratio using a Chi-square test. LG numbering and orientation were attributed to best fit the *A. duranensis and A. ipaensis* chromosomes.

2.8 Comparative genomics

Gene sequences underlying the molecular marker loci assigned to the linkage map were blasted against the *A. duranensis* and *A. ipaensis* genome sequences (http://peanutbase.org/) to detect significant matches at minimum threshold E values $< 10^{-60}$ (Supplementary Table S6). For the purpose of graphic preparation, cM distance on the *A. evenia* linkage groups were scaled by factors calculated on the basis of the genome size of *A. ipaensis*, to provide matching chromosomal lengths in base pairs. The graphical comparative maps were drawn using the Circos program (http://circos.ca/) and the SpiderMap software (JF Rami, unpublished).

3. Results and discussion

3.1 Selection of parental genotypes to develop a mapping population

3.1.1 Choice of the reference line

The *A. evenia* species was proposed as a new model legume to investigate the evolution of the nitrogen-fixing symbiosis because it possesses a number of interesting characteristics for both classical and molecular genetics.[21,22] Key attributes of *A. evenia* include diploidy ($2n = 2\times = 20$), self-pollination, a small genome for a legume, a short growth cycle and a high level of diversity. In addition it is efficiently nodulated by the well-characterized strains *Bradyrhizobium* ORS278 and ORS285 and it can be transformed using the hairy root system. First nodulation studies were performed on the cultivar IRFL6945 belonging to the *A. evenia* ssp. *serrulata*.[21] But subsequent analysis of the intra-specific diversity revealed that *A. evenia* ssp. *evenia* is more appropriate due to a smaller genome size (415 vs 465 Mb) and the presence of genotypes with a more rapid reproductive cycle and a non-branching habit that facilitate plant management (Fig. 2A,B).[22] We made use of this knowledge to select from within the subspecies *evenia* var. *evenia* the genotype Mbao CIAT22838, which displays these advantageous characteristics, as the reference genotype. Its nodulation properties were previously shown to be representative of the species with ~11nodules/plant, and its ability to be transformed by *Agrobacterium rhizogenes* was found to be similar to what was described for the line IRLF6945 with a 60% root transformation rate and subsequent 90% nodulation rate upon inoculation with *Bradyrhizobium* ORS278.[21] Even though *A. evenia* is a self-pollinated plant and was previously shown to be predominantly homozygous, a single seed from the reference genotype was selfed three times to ensure high homozygosity.

3.1.2 Choice of a mapping parent

The establishment of a mapping population required the identification of a second parent that has favorable polymorphism while the resulting F_2 population should not have developmental and genetic abnormalities. Characterization of the diversity in *A. evenia* had previously revealed that the genetic polymorphism was the highest between the subspecies *evenia* and *serrulata* but that inter-subspecies crosses led to sterile F_1 hybrids. Conversely, intra-subspecific hybridizations were shown to generate fully fertile F_1 progenies.[22] Therefore, to maximize genetic polymorphism, the genotype Bahia CIAT8232 that belongs to the subspecies *evenia* var. *pauciciliata* and that presents a distinct geographic origin (African for Mbao versus American for Bahia) was developed as an inbred line and used as the second mapping parent (Fig. 2A). The Mbao and Bahia genotypes present a similar DNA content (415 and 405 Mb, respectively) but they were shown to display a 32% SSR polymorphism.[22] This situation is similar to the one observed in *Medicago truncatula* with the two mapping parents Jemalong and DZA315.16 and between three ecotypes of *Arabidopsis thaliana*, indicating it is convenient for generating genetic maps.[26] In addition, the two genotypes display phenotypic differences making them readily distinguishable such as the branching habit, the timing of flower openings and the presence of a dark throat on the upper petal (Fig. 2B).

3.1.3 Development of a F2 mapping population

Artificial hybridizations were performed using the genotype Mbao CIAT22838 as female and male alternatively and producing F_{1A} and F_{1B} progenies respectively. These F_1 plants showed hybrid vigor, intermediate flower characteristics (opening time and shape) and were self-fertile (Fig. 2B). In particular, they displayed equivalent pod set (Bahia: 8.6 ± 0.9, Mbao: 8.2 ± 1, F_1: 8.8 ± 1.1 seeds/pod). The resulting F_{2A} and F_{2B} offsprings had no albino or dwarf individuals. Altogether, this suggested that Bahia CIAT8232 was a suitable crossing partner to Mbao CIAT22838 and that the obtained F_2 offsprings could be used as mapping population to develop a genetic map.

3.2 Transcriptome sequencing for identification of symbiotic genes and genetic markers

3.2.1 Illumina paired-end sequencing and *de novo* transcriptome assembly

High-throughput transcriptomic sequencing allows generating a large transcript sequence dataset for gene discovery and development of molecular markers that serve for gene-based map construction. To be able to achieve a broad survey of genes associated with symbiosis, RNA was extracted and pooled from various organs: roots, leaves and nodules at 4, 7 and 14-dpi. In addition, to increase the efficiency of molecular marker development by prior *in silico* polymorphic analysis, a RNA library was developed for each parental genotype.

Using Illumina paired-end sequencing and after stringent quality check of the raw data, ~72 million reads were kept for the Mbao genotype and ~66 million for the Bahia genotype. Based on these high-quality reads, a total of 51,763 and 52,648 contigs were assembled respectively, with a common N50 of 1,800 bp and a consistent contig size distribution (Table 1). This indicated that a similar transcript content was obtained for both mapping parents. Furthermore,

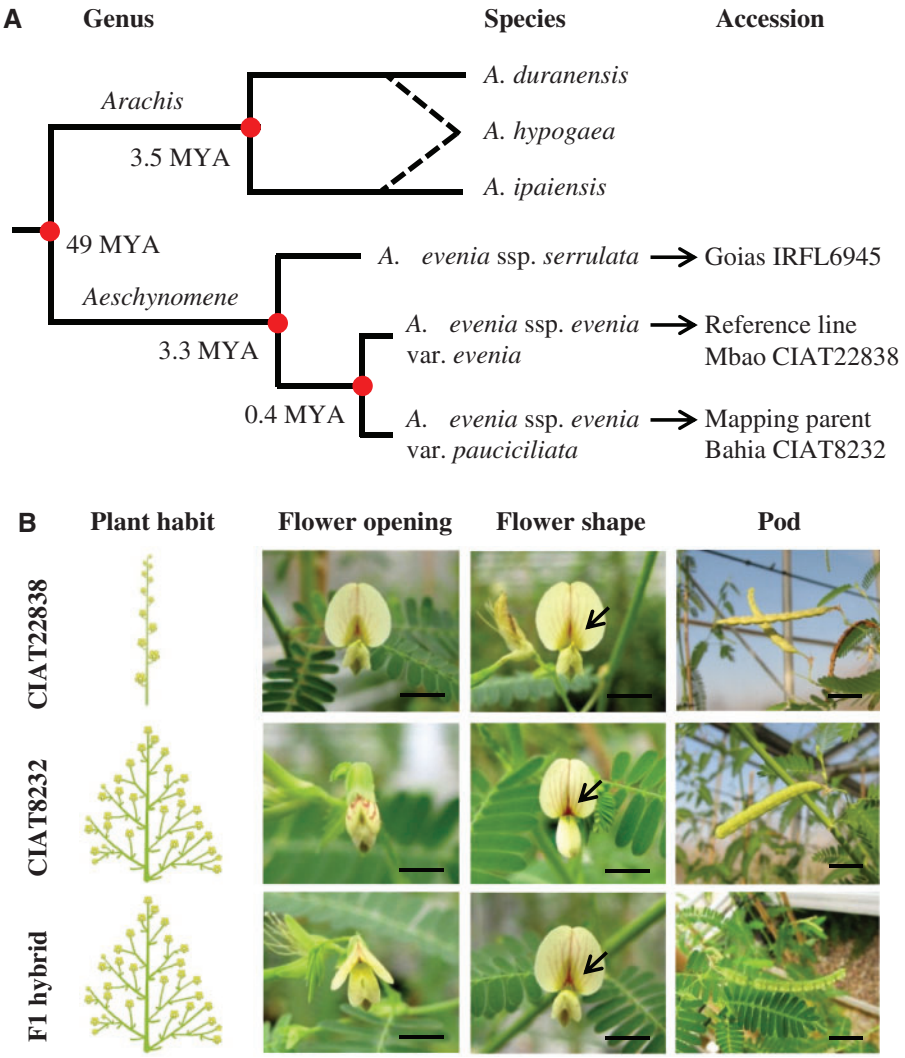

Figure 2. Differentiation within the *A. evenia* species and its use for genomic and genetic studies. (A) Schematic phylogeny depicting the genetic differentiation within the *A. evenia* species and the relationships with *Arachis* sp.[18,22] Dash lines connect the tetraploid *A. hypogaea* to its most probable diploid genome donors *A. duranensis* and *A. ipaensis*. Divergence times at certain nodes (circles) were estimated using a relaxed molecular clock (Table 3). (B) Phenotypic patterns which distinguish CIAT22838 and CIAT8232 lines and impact on the F1 hybrid obtained by cross-pollination: plant habit (ramified or not), time of flower opening (early and late), flower shapes (standard shape and dark throat indicated by an arrow), pod production. Bars = 5 mm.

realignment of all the usable reads to the contigs revealed that ~35,000 contigs were remapped by more than 100 reads and, among them, ~24,500 contigs were realigned with more than 500 reads, suggesting the contigs were well overlapped by the sequencing reads (Fig. 3A). *A. evenia* being a member of the Dalbergioid clade along with the recently sequenced *A. duranensis* and *A. ipaensis*, their genomic data could be also used to assess the quality and coverage of the assembled transcriptomes.[22] Since *A. duranensis* has not experienced the local gene duplications evidenced for *A. ipaensis*, the former species was better suited for this purpose. Thus, comparative analysis revealed that 60% of total contigs (~31,850) and 83% of contigs over 500 bp in length (~26,378) had BLAST matches (*E*-value $< 10^{-5}$) onto the ~36,700 predicted genes in the *A. duranensis* genome (Fig. 3B). In addition, plotting the contigs onto top-hit homologous coding sequences of *A. duranensis* (*E*-value $< 10^{-30}$) indicated a high proportion of genes with complete coding sequence (Fig. 3C). Altogether, these data suggested that a comprehensive gene capture was achieved for *A. evenia*.

3.2.2 Identification of putative orthologs of symbiotic genes

To determine which symbiotic genes can be identified in *A. evenia*, a BLAST search was performed in the Bahia and Mbao transcriptomes using the sequence of symbiotic genes characterized in the two model legumes *M. truncatula* and *L. japonicus*.[8,27–33] This allowed the identification of a single sequence in *A. evenia* for 36 out of 47 symbiotic genes analysed, for which the putative orthology was checked by reciprocal BLAST and sequence alignment (Fig. 4; Files S1, S2). These genes completely covered the different steps of the nodulation process: Nod signaling, activation and regulation of transcription, nodule organogenesis and regulation of nodulation (Fig. 4). This coverage suggested that these steps were conserved in *A. evenia*, but it could not be excluded that some of these genes were only involved in the arbuscular mycorrhizal symbiosis or in other biological processes. So far, only the *A. evenia SYMRK*, *CCaMK*, *CRE1* and *DNF1* orthologs were effectively demonstrated to play a role in nodulation.[34,35]

Table 1. Overview of the sequencing and assembly

Nucleotide length (pb)	*A. evenia* Illumina libraries	
	Bahia CIAT8232	Mbao CIAT22838
150–500	21,020	20,269
501–1,000	10,109	9,990
1,001–2,000	13,127	13,216
2,001–3,000	5,673	5,561
3,001–4,000	1,767	1,769
4,001–5,000	570	578
>5,000	382	379
Raw reads	68,945,277	75,745,315
Filtered reads	65,865,857	72,387,963
Number of contigs	51,763	52,648
min length (bp)	150	150
max length (bp)	16,895	16,886
N50 (bp)	1,799	1,803
Average length (bp)	1,092	1,082
Total nucleotide length (bp)	56,526,002	57,007,824

Interestingly, for some other symbiotic genes, transcripts were not detected despite the use of root and nodule materials for the transcriptomes. Missing genes are involved in three symbiotic processes: bacterial recognition (i.e. *LYK3* and *LYR3*), rhizobial infection (i.e. *ANN1*, *EPR3*, *FLOT* and *RPG*) as well as nodule functioning and immunity (i.e. *DNF2*, *FEN1*, *SUNERGOS1*, *SymCRK* and *VAG1*) (Fig. 4). In the absence of genomic data for *A. evenia*, it was not possible to determine whether these genes were absent, not expressed or missed when mining the transcriptome data. But, it is worth noting that the concerned genes are involved in symbiotic processes that present sharp differences between *A. evenia*, *M. truncatula* and *L. japonicus* (Fig. 1).[2,19] As a consequence, these first observations represented important cues to further investigate the evolution of rhizobial symbiosis.

3.2.3 Development of EST-based molecular markers

To allow the development of molecular markers, the Bahia and Mbao transcriptomes were searched for microsatellites. These were subsequently analysed for *in silico* polymorphism by comparing the orthologous SSR-containing contigs of the two parental genotypes and filtered to allow primer design on flanking sequences. 1,500 microsatellites complying with these criteria were identified, the most abundant type of repeat motif being tri-nucleotides (46.4%), followed by di-nucleotides (35.9%) (Table 2). 500 SSR sites were selected for primer design and used for assessment of the polymorphism between the two parental genotypes and the F_1 progeny by capillary sequencing. Of the 412 successfully amplified SSRs, 335 (81%) were co-dominant, 50 (12%) were dominant and 27 (7%) were not polymorphic. Hence, thanks to the *in silico* analysis, the efficiency of polymorphic SSR development was increased by 2.5-fold compared to previous work with no prior filtering of the genic SSR tested on the same Bahia and Mbao genotypes.[22] Finally, a set of 318 co-dominant SSR markers was retained for the genotyping of the F_2 mapping population (Supplementary Tables S1, S2).

In order to map the symbiotic genes, polymorphisms were also searched for by comparing the orthologous sequences from the two parental genotypes and were converted into molecular markers. When no polymorphic SSR and no INDEL could be detected, Single Nucleotide Polymorphism (SNP) was used instead for KASP assay design (Supplementary Tables S1–S4). Subsequently, at least one

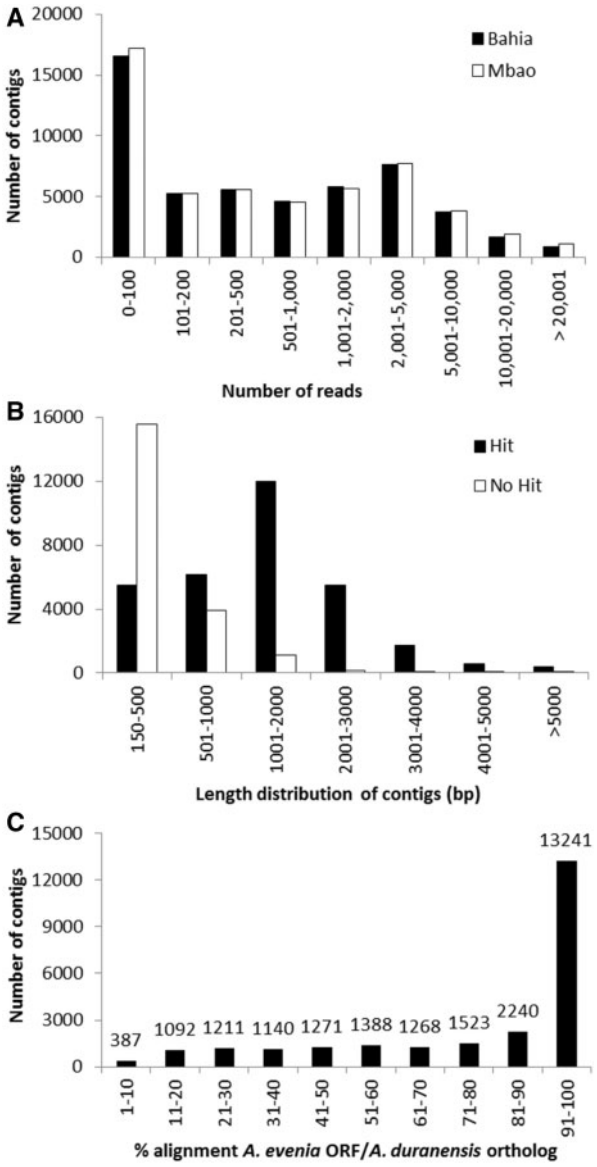

Figure 3. Assessment of assembly quality and gene coverage for the Illumina transcriptomes. (A) Distribution of mapped reads within the assembled contigs for the Illumina transcriptomes of the two parental genotypes Bahia and Mbao. B, Comparison of contig length between hit and no-hit contigs from the Mbao transcriptome assembly after tBLASTX analysis with the protein database of *A. ipaensis*. C, Comparison of Mbao contigs to putative orthologous *A. ipaensis* sequences by evaluating the % of alignment of corresponding coding sequences.

molecular marker could be satisfactorily developed for each identified symbiotic gene, except for two genes, *SEN1* and *RSD*, that lack of polymorphism. A few additional markers were also used, notably to locate the 5S and 45S loci that were previously identified on *A. evenia* chromosomes by cytogenetics (Supplementary Tables S1, S2).[21]

3.3 Construction of the genetic map and mapping of symbiotic genes

3.3.1 Genotyping of the 220 F_2 mapping population

110 F_{2A} and the 110 F_{2B} individuals, which were obtained by bi-directional crossings of the Bahia and Mbao genotypes, formed the

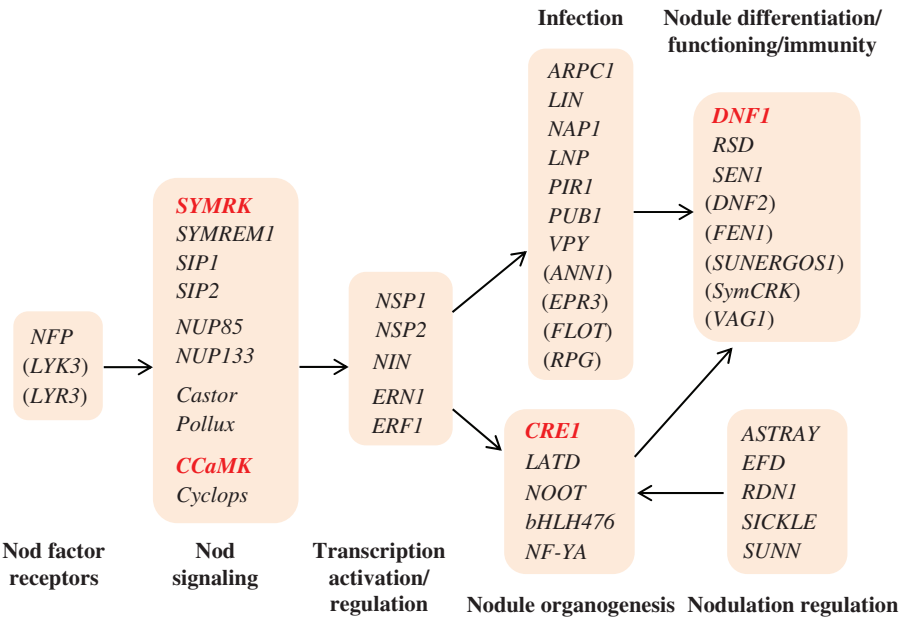

Figure 4. Simplified model for the symbiotic signaling pathway in legumes. Symbiotic genes identified in *M. truncatula* and *L. japonicus* are listed and classified in main symbiotic functions.[8,27-33] Putative orthologs were searched in the Illumina transcriptome of the two mapping parents. Those that could not be identified are between parentheses. Genes highlighted in bold characters are those which were demonstrated to be involved in the Nod-independent nodulation process in *A. evenia*.[33,34]

Table 2. Distribution of repeat types and number of repeats within the Mbao library

Repeat type	Number of repeat units									Total	%
	3	4	5	6	7	8	9	10	>10		
Di	–	–	–	64	105	95	107	68	120	559	35.9
Tri	–	68	141	181	132	93	49	29	31	724	46.4
Tetra	12	58	53	29	11	2	2	1	–	168	10.8
Penta	25	51	27	1	1	1	–	–	–	106	6.8
Hexa	–	2	–	–	–	–	–	–	–	2	0.1

mapping population. DNAs were extracted from the 220 F_2 individuals and were genotyped for the selected 318 SSR markers by capillary sequencing. Multiplexes of 12 SSR markers were chosen so as to combine both three ranges of fragment sizes and four fluorochromes (FAM, VIC, NED and PET). 46 additional markers comprising SSRs, INDELs and SNPs, which target notable putative symbiotic genes, were also used for genotyping by capillary sequencing or KASPAR technology. This genotyping generated a matrix with <0.6% of missing data for the 364 co-dominant markers used.

3.3.2 Building of the genetic map

The *A. evenia* genetic map was built using the JOINMAP software with a minimum LOD score value of 6. The markers were assembled into 10 linkage groups, corresponding to the haploid chromosome set of *A. evenia* ($2n = 20$). Interestingly, no marker remained unlinked and no linkage group could be split by increasing the LOD score value up to a LOD of 15. In addition, manual checking of the marker positions using color-mapping revealed no discrepancy in the genotyping data, thus corroborating the robustness of the 10 linkage groups obtained. They were thus numbered AeLG1 to AeLG10 and orientated as explained below (Fig. 5).

The resulting genetic map spans 1,036 Kosambi cM, with an average of 400 kb/cM. The 10 linkage groups have different genetic sizes ranging from 82 to 130.8 cM and contained 13–49 markers (Table 3). This leads to average distances between two adjacent markers varying from 2.6 to 7.5 cM but with maximum interval sizes per linkage group ranging from 10.7 to 25.1 cM (Table 3). This situation is reminiscent of what has been already observed in *Vigna unguiculata* and *Lens culinaris* for which genetic maps were also developed with gene-based markers.[36,37] The gaps have been proposed to correspond to gene-poor regions or to recombination hotspots. In addition, mapping of the 5S and 45S rDNA loci further enabled distinguishing two linkage groups. Thus the 5S rDNA locus was found to locate on linkage group AeLG8 with a central position (at 55.6 cM on total 96.2 cM length) that is in accordance with the proximal localization previously observed on metaphase chromosomes by GISH.[21] Conversely, two 45S rDNA loci were shown to locate to in secondary constrictions of satellite chromosomes.[21] Sequence polymorphism for the ITS region allowed mapping only one of the two loci on the upper part of the linkage group AeLG10 (at 19.4 cM on a total 82 cM length) (Fig. 5).[22] Finally, gene-specific markers corresponding to symbiotic genes were found to be distributed in all but one (AeLG2) linkage groups, with AeLG5, AeLG6, AeLG7 and AeLG9 containing the majority of them (Fig. 5).

3.3.3 Regions of segregating distortion (RSD)

Although almost all markers complied with the 1:2:1 expected Mendelian ratio for co-dominant genetic markers, two regions of segregation distortion were observed (Table 3, Fig. 6A). The AeLG2 RSD comprised up to 10 markers covering a map distance of up 70 cM and the AeLG9 RSD contained 11 markers (among which the symbiotic *NSP1* and *EFD* genes) distributed over 45 cM, altogether representing 10% of the genetic map. In both cases, there was a strong overrepresentation of the Bahia allele and a near complete disappearance of the Mbao allele where the distortion culminated,

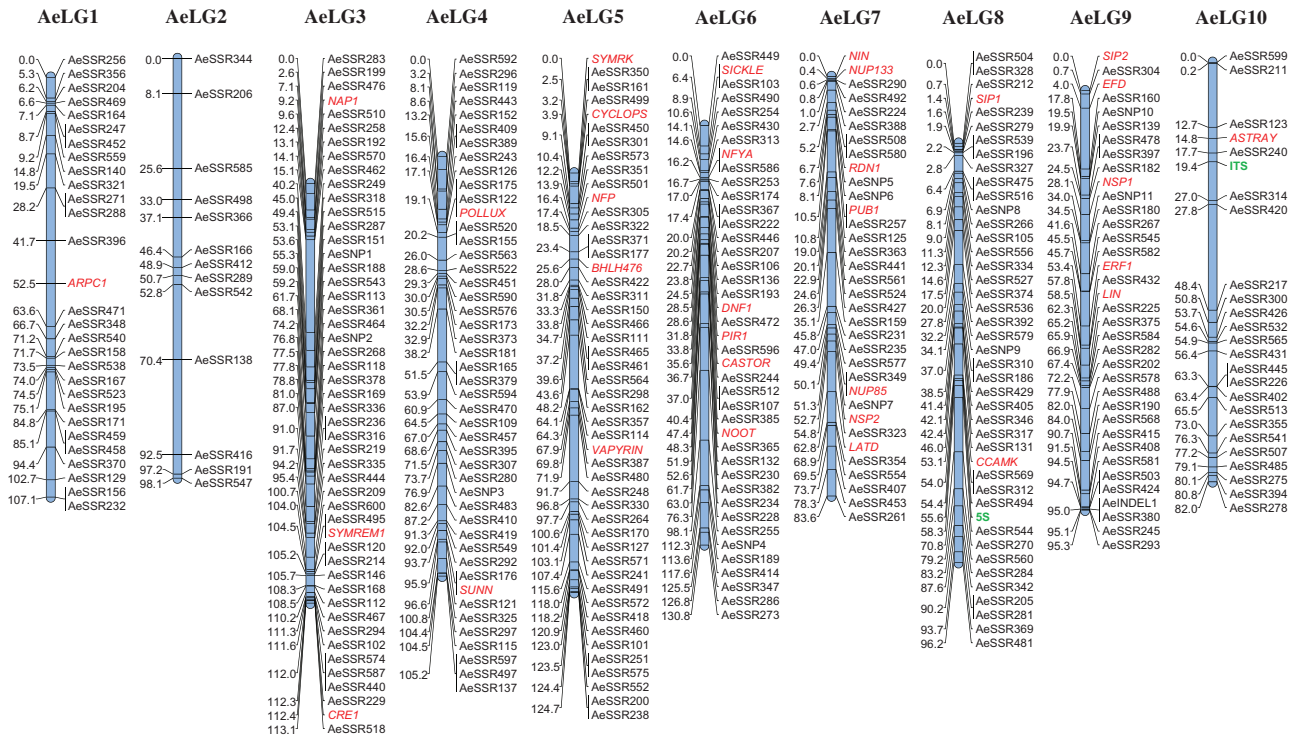

Figure 5. Genetic map of *Aeschynomene evenia*. The genetic map based on F2 mapping population CIAT22838 × CIAT8232 is comprised of 364 gene based markers including SSR, INDEL and SNP (KASP) markers. The ten linkage groups are designated as AeLG1-AeLG10. The code to the right of the linkage groups refers to the marker or symbiotic gene name. The numbers to the left of the linkage groups refers to the genetic distances (Kosambi cM) from the top.

Table 3. Characteristics of the *Aeschynomene evenia* genetic map

LG	Total length (cM)	Number of loci	Number of markers	Average distance (cM)	Min–max distances (cM)	Distorted markers ($P < 0.001$)
AeLG1	107.1	25	29	4.3	0–13.5	0
AeLG2	98.1	13	13	7.5	0–22.1	10
AeLG3	113.1	44	49	2.6	0–25.1	0
AeLG4	105.2	38	46	2.8	0–13.3	0
AeLG5	124.7	42	48	3.0	0–19.8	0
AeLG6	130.8	37	41	3.5	0–21.8	0
AeLG7	83.6	31	34	2.7	0–10.7	0
AeLG8	96.2	37	43	2.6	0–12.6	0
AeLG9	95.3	33	36	2.9	0–13.9	15
AeLG10	82	24	25	3.4	0–20.6	0

whereas the proportion of heterozygotes remained unaltered. Judged from these skewed genotypic frequencies, a strong selection in favor of the Bahia alleles occurs at these RSD.

Distortions can be caused by chromosomal structural rearrangements, differences in DNA content, gene incompatibilities or deleterious recessive alleles. Inversion or translocation events have been shown to be the cause of distortions in genetic maps of other legume such as *M. truncatula*, *L. japonicus* and *L. culinaris*.[13,26,38–41] Such events are more likely to occur when performing interspecific crosses than when using related genotypes, but one cannot exclude that a chromosomal rearrangement occurred within the subspecies *evenia* to which both the Bahia and Mbao genotypes belong. In addition, even if they display a similar overall DNA content, this DNA content may not be identical in the two RSD regions. Comparison of the genotypic data for the F_{2A} and F_{2B} mapping populations, revealed similar allele skewing in the two RSD regions, indicating that the observed distortions were independent of maternal and paternal effects. However, pollen viability analysis showed that the F_1 progeny contained ~38% aborted pollen in contrast to the Bahia and Mbao parental genotypes that showed more than 99% pollen viability (Fig. 6B,C). This F_1 semisterility may account for the observed skewed genetic frequencies, as it has been reported for some *M. truncatula* crosses.[38] Although, this RSD can alter genetic distances, it does not compromise marker ordering in the SDR regions.

3.4 Comparative analysis with *Arachis*

3.4.1 Estimate of time divergences

With the aim of fostering genomic and symbiotic comparative analysis, we wanted to further our knowledge of the genetic differentiation in *A. evenia* and of the evolution within the Dalbergioid clade that contains *Aeschynomene* and *Arachis*. A ~49 MYA divergence of *Aeschynomene* and *Arachis* was previously estimated from chloroplast gene phylogenies (Fig 2A).[6] Using this date for calibration and assuming a simple molecular clock, we made a rough estimate for the divergence times between different *A. evenia* genotypes and between the two *Arachis* ancestral species of peanut. The coding reading frame of seven single-copy nuclear genes was obtained from the transcript databases available for *A. evenia* genotypes Bahia (CIAT8232), Goias (IRFL6945) and Mbao (CIAT22838) and their *Arachis* orthologs in the genome sequences of *A. duranensis* and *A. ipaensis* (Fig 2A, Table 4, Supplementary Table S5). Synonymous substitution rates (Ks) were calculated for binned values and an

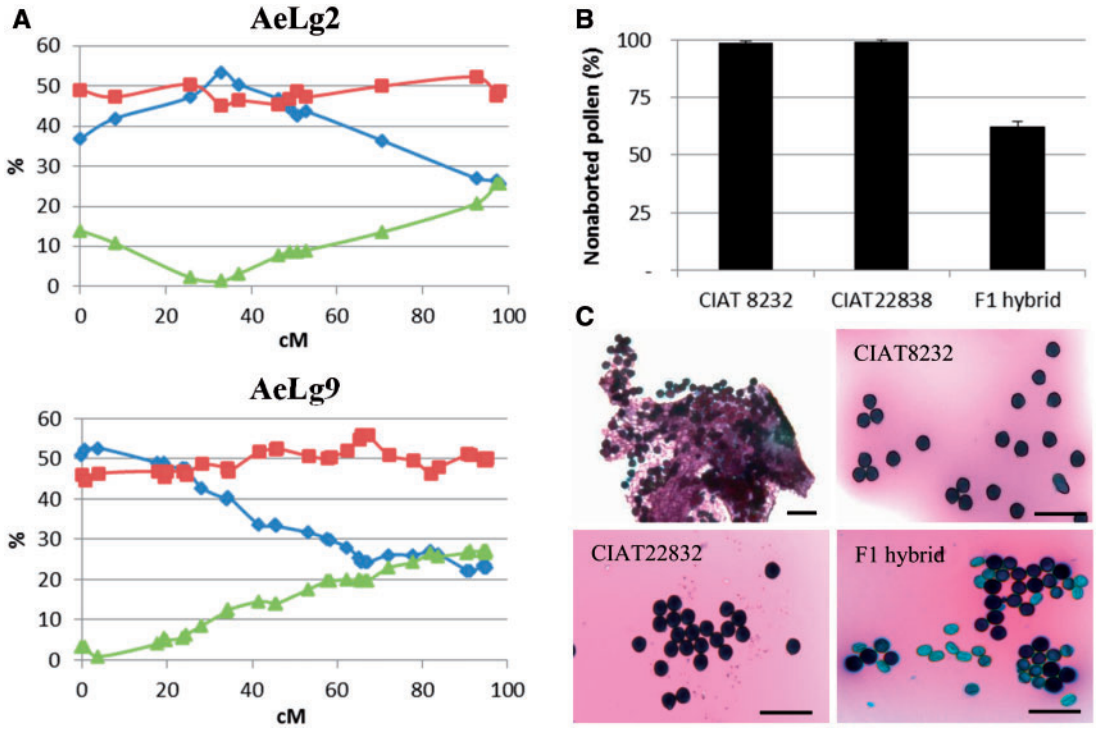

Figure 6. Segregation distortion and pollen viability of F1 hybrids. (A) Segregation distortion of the co-dominant markers along linkage groups *AeLG*2 and *AeLG*9 in the F2 mapping population. Triangles refer to the CIAT22838 genotype, lozenges to the CIAT8232 genotype and squares to the heterozygote. *X*-axis: genetic distance from the top of the linkage group in Kosambi cM. *Y*-axis: frequency (%) of segregation of the different genotypes. If no distortion occurs, the segregation values should be 25%/50%/25%. B, Proportion of non-aborted pollen grains in the mapping parents and the F1 hybrid. C, Anther from the CIAT22838 line (upper left panel) and pollen grains of the mapping parents and the F1 hybrid. Aborted pollen stains pale and nonaborted pollen stains dark. Bars = 50 μm.

Table 4. Estimation of divergence times between *Aeschynomene* and *Arachis* genomes

Evolutionary divergence	Substitution per site (Ks)							Average (SR)	Divergence	
	APX2 (738 pb)	AS2 (1,629 pb)	β-Fructofuranosidase (1,536 pb)	PIP2;7 (864 pb)	Proline-tRNA ligase (1,098 pb)	SPAG1 (861 pb)	TIP1;1 (750 pb)		(SR/Mya)	(Mya)
A. duranensis–*A. evenia* Mbao	0.581	0.556	0.504	0.797	0.380	0.431	0.442	0.527	0.00534*	49*
A. duranensis–*A. evenia Mbao*	0.569	0.576	0.531	0.758	0.367	0.421	0.419	0.520		
A. duranensis–*A. ipaensis*	0.029	0.039	0.037	0.053	0.025	0.049	0.026	0.037	0.00534	**3.5**
A. evenia Goias-Bahia	0.035	0.056	0.039	0.024	0.029	0.038	0.026	0.035	0.00534	**3.3**
A. evenia Goias-Mbao	0.035	0.050	0.045	0.024	0.038	0.032	0.026	0.036	0.00534	**3.3**
A. evenia Bahia-Mbao	0.000	0.011	0.008	0.000	0.008	0.005	0.000	0.005	0.00534	**0.4**

*Average divergence rate obtained by using the known 49 MYA divergence time between *Aeschynomene* and *Arachis*.[6] Numbers in bold are inferred values.

average mutation rate of 5.34×10^{-3} substitutions/synonymous site/Mya was obtained by using the 49 MYA value for the *Aeschynomene*–*Arachis* divergence.[6]

This enabled an estimate of divergence time between the two mapping parents that belong to *A. evenia* ssp. *evenia* var. *evenia* and var. *pauciciliata* at ~0.4 Mya, reflecting both the low difference observed in polymorphism and the cross-compatibility between the two varieties.[22] In contrast, a much older divergence time of ~3.3 Mya was deduced for *A. evenia* ssp. *evenia*–*A. evenia* ssp. *serrulata*. This is in agreement with the pronounced genetic differentiation of the two *A. evenia* subspecies evidenced by the difficulty of making successful crosses and the almost completely sterile hybrids.[22] This divergence time is similar to the ~3.5 Mya obtained for *A. duranensis*–*A. ipaensis*, which are genetically isolated, thus supporting the idea that the two *A. evenia* subspecies are at the final stage of the speciation process. Interestingly, the value calculated for the *Arachis* species couple is in accordance with other estimates (2.1–3.5 Mya).[18,42] Therefore, our approach seems to be valid to provide a timeframe for *A. evenia* evolution.

3.4.2 Genome conservation and evolution with *Arachis ipaensis*

Several studies have undertaken comparative genomics within the Papilionoids revealing that the degree of synteny is correlated with

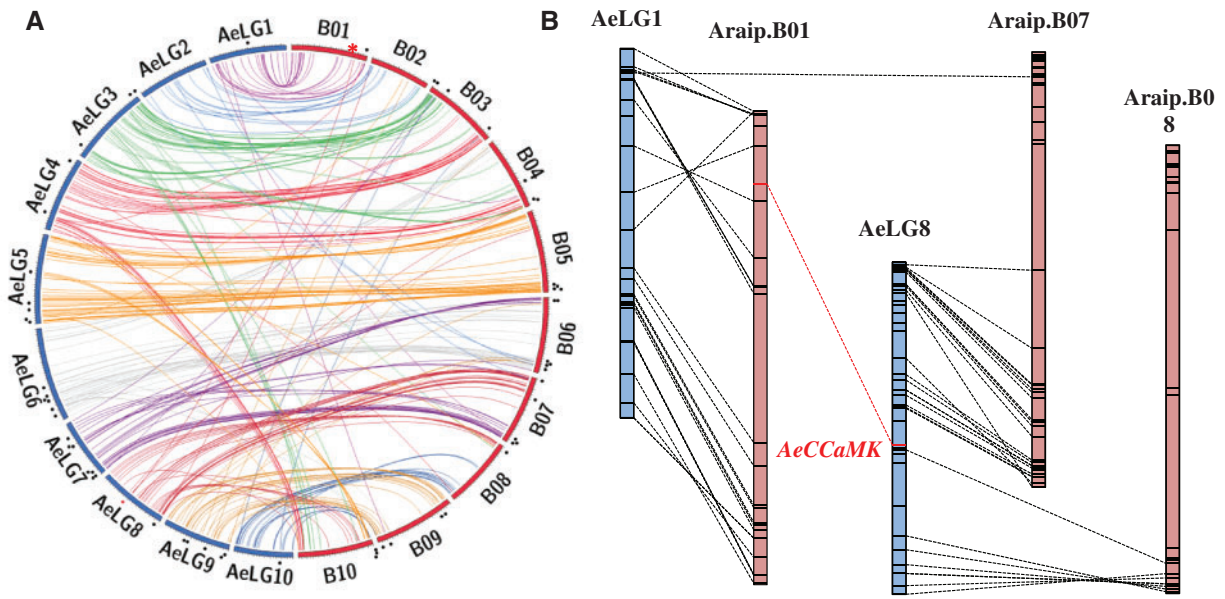

Figure 7. Syntenic relationships of *A. evenia* with the sequenced Dalbergioid legume *Arachis ipaiensis*. (A) Comparison of the genetic map of *A. evenia* with the genome sequence of *A. ipaensis*. Circled bars on the left side correspond to the linkage groups of *A. evenia* while those on the right side represent the chromosomes of *A. ipaensis*. Homologous loci are connected by lines and spots correspond to symbiotic genes. Spots indicate conserved macrosyntenic locations and asterisks non-conserved ones. (B) Detailed example of macrosynteny between *A. evenia* and *A. ipaensis*. Alignment of linkage groups AeLG1 and AeLG8 from the developed genetic map of *A. evenia* with the corresponding A1 and A8 chromosomes of *A. ipaensis*. Dot lines connect orthologous loci between the two species. The dot line link the non-macrosyntenic location of the *AeCcaMK* gene.

phylogenetic distance of the legume species.[9–17] The first glimpses of genome evolution within the Dalbergioid clade were only recently obtained by the genome sequencing of the two putative genome donors of peanut. This revealed a mostly one-to-one correspondence between the two species, with the genome of *A. duranensis*, but not of *A. ipaensis*, having undergone several major rearrangements.[22] To perform the *Aeschynomene-Arachis* genome comparison, we made use of the gene-based nature of the markers mapped on the genetic map of *A. evenia* ($2n = 20$) to locate orthologous sequences on the chromosomes of *A. duranensis* and *A. ipaensis* ($2n = 20$) (Supplementary Table S6). Although all the transcripts used to develop the markers did not necessarily contain predicted coding reading frames, ~97% of them (353 out of 364) showed significant matches to *Arachis* sequences. To facilitate comparison, the *Aeschynomene* linkage groups have been numbered and oriented to best match the corresponding *Arachis* chromosomes (Supplementary Fig. S2A,B).

When the *A. evenia* linkage groups and the *A. ipaensis* chromosomes were aligned, extensive stretches of shared colinearity between them were evident (Fig. 7A, Supplementary Fig. S2A). Some linkage groups of *A. evenia* showed hits mainly with one chromosome of *A. ipaensis*, notably AeLG1 that matches Araip.A01, AeLG2 with Araip.A02, AeLG5 with Araip.05, AeLG6 with Araip.06 and AeLG9 with Araip.09 (Fig. 7A). However, detailed analysis of macrosynteny between homologous chromosomes evidence internal rearrangements such as translocations and inversions (Fig. 7B). In addition, even though *A. evenia* and *A. ipaensis* share the same base chromosome number of 10, one-to-one relationships do not hold true for the remaining *A. evenia* linkage groups. Instead, they appear to be composed of segmental syntenic blocks matching different chromosomal positions in *A. ipaensis* as exemplified by AeLG4 that contains three conserved blocks found in the chromosomes Araip. A03, A04 and A10. It is worth noting that some gaps observed within *Aeschynomene* linkage groups are coincident with syntenic block junctions and thus may correspond to more variable regions prone to genome restructuring as suggested earlier.[4] In addition, *Aeschynomene* loci mostly match and cover the euchromatic arms of *Arachis* chromosomes while only few loci map pericentromeric regions (Fig. 7A).[18,22] This observation is consistent with all the euchromatic regions of the *A. evenia* genome being captured by the gene-based markers composing the genetic map.

3.4.3 Macrosynteny analysis for the symbiotic genes

Considering the divergence time between *Aeschynomene* and *Arachis* (~49 Mya) relative to the isolation time of the Dalbergioid clade from the Phaseolids and Galegoids (~55 Mya), along with the 3 to 3.5-fold larger genome of *A. duranensis* and *A. ipaensis* (1,250 and 1,560 Mb, respectively) compared to *A. evenia* (415 Mb), the high level of macrosynteny documented here between *Arachis* and *Aeschynomene* is remarkable. Much of the expansion in *Arachis* genomes may be due to retroelements, but it has done little to disrupt macrosynteny that is dominated by large chromosome arms-size rearrangements. Such macrosynteny was found to be maintained for the symbiotic genes mapped onto the *A. evenia* linkage groups, suggesting that true orthologous relationships were identified (Fig. 7A). One noticeable exception concerns the *CCaMK* gene that is mapped onto AeLG8 and is positioned between two syntenic blocks, while its *A. ipaensis* ortholog locates on Araip.A01 chromosome within a block that is syntenic to AeLG1 (Fig. 7A,B). Although it is not possible to determine the extent to which macrosyntenic relationships identified by genetic mapping are indicative of conserved microsynteny, it is interesting to note that a retrotransposon insertion site was previously identified 564-bp upstream of the *AeCCamK* gene.[34] Since retrotransposons are known to be involved in genome restructuring and also to induce gene mobility, further studies may help

determining their potential impact on the genomic environment of this key symbiotic gene in *A. evenia*.

4. Conclusions and perspectives

To advance in our understanding of the NF-independent symbiosis in the model legume *A. evenia*, we developed a genetic map incorporating expressed genes that are likely to be involved in nodulation. Previous characterization of different genotypes in *A. evenia* allowed us to select the Mbao genotype as a reference for molecular genetics and the Bahia genotype as a suitable crossing parent to develop a F_2 mapping population. Transcriptome *in silico* analysis for the two parental genotypes enabled an efficient selection of polymorphic molecular markers that were used to genotype the mapping population. The choice of genotyping methods, i.e. multiplexing and capillary sequencing for SSR and INDEL markers along with the KASP assays for SNP markers resulted in 99.4% complete genotyping data. The genetic map was resolved in 10 robust linkage groups that most likely represented the 10 chromosome pairs in *A. evenia*. The availability of a high-density map containing easy-to-use and cost-effective markers will facilitate gene and QTL mapping work in *A. evenia*.

Given the gene-based nature of the markers used, this genetic map provided the first insights into the structure of the genome gene-space and, in particular, uncovered the distribution of expressed orthologs of known symbiotic genes. Comparative genomic analysis with *Arachis* further evidenced that the *Aeschynomene* genome constituted of macrosyntenic blocks. Therefore, this genetic map of *Aeschynomene* could form the basis for future efforts in the scaffold anchoring and assembly in an *Aeschynomene* genome sequencing initiative. The availability of a whole genome sequence is now feasible thanks to the next-generation sequencing technologies. This will allow much better comparison with *Arachis* to characterize genome evolution within the Dalbergioid clade. In particular, microsynteny analysis will allow investigating if chromosome rearrangements and transposable elements had an impact on certain symbiotic genes as suspected for the *AeCCaMK* gene and those that could not be found in the transcriptome datasets. Understanding how these symbiotic genes evolved (by changes of gene expression or by gene loss) in a NF-independent context would be an important advance in understanding the evolution of nodulation.

In addition, an *a priori* identification of symbiotic genes in *A. evenia* using a mutagenesis approach should be a powerful tool. In this perspective, a strategy combining gene mapping with sequencing at gene or whole genome-level will allow an efficient identification of new alleles and new symbiotic loci among mutant lines, as has been successfully performed in *M. truncatula* and *L. japonicus*.[31,43–46] New methods, such as SHOREmap and MutMap, now integrate simultaneous mapping and mutation identification in a single step by deep whole-genome resequencing of pooled DNA from a F_2 mutant population.[47,48] This population can be obtained either by crossing the mutant to the mapping parent or to the original WT parental line used for mutagenesis. In the latter case, mapping relies on the detection of mutagenesis-induced polymorphisms. The similar genetic backgrounds present the advantage of both simplifying segregation patterns and suppressing the occurrence of distortions. As a result, applying this strategy to *A. evenia*, the identification of new symbiotic genes should be straightforward and should facilitate the deciphering of the molecular mechanisms of the NF-independent process.

Acknowledgements

We thank A.M. Risterucci and F. de Bellis (AGAP, CIRAD, France) for valuable suggestions on genetic map construction and L. Legrand (LIPM, INRA/CNRS, France) for informatics assistance.

Accession numbers

SRR3276128, SRR3285082

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported by Agropolis Fondation through the « Investissements d'avenir » programme (ANR-10-LABX-0001-01) under the reference ID "AeschyMap" AA1202-009, and by a grant from the French National Research Agency (ANR-AeschyNod-14-CE19-0005-01). Funding to pay the Open Access publication charges for this article was provided by the French National Research Agency.

References

1. Graham, P.H. and Vance, C.P. 2003, Legumes: importance and constraints to greater use, *Plant Physiol.*, **131**, 872–7.
2. Sprent, J.I. 2007, Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytol.*, **174**, 11–25.
3. Peterson, R., Slovin, J.P., Chen, C. 2010, A simplified method for differential staining of aborted and non-aborted pollen grains. *Int. J. Plant Biol.*, **1**:e13.
4. Bertioli, D.J., Moretzsohn, M.C., Madsen, L.H. et al. 2009, An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. *BMC Genomics*, **10**, 45–55.
5. Held, M., Hossain, M.S., Yokota, K., et al. 2010, Common and not so common entry. *Trends Plant Sci.*, **15**, 540–5.
6. Lavin, M., Herendeen, P.S., Wojciechowski, M.F. 2005, Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst. Biol.*, **54**, 575–594.
7. Doyle, J.J. and Luckow, M.A. 2003, The rest of the iceberg. Legume diversity and evolution in a phylogenetic context, *Plant Physiol.*, **131**, 900–10.
8. Oldroyd, G.E.D., Murray J.D., Poole P.S., et al. 2011, The rules of engagement in the legume-rhizobial symbiosis, *Annu. Rev. Genet.*, **45**, 119–44.
9. Choi, H.K., Mun, J.H., Kim, D.J., et al. 2004, Estimating genome conservation between crop and model legume species, *Proc. Natl. Acad. Sci.*, **101**, 15289–94.
10. Isobe, S.N., Hisano, H., Sato S., et al. 2012, Comparative genetic mapping and discovery of linkage disequilibrium across linkage groups in white clover (*Trifolium repens* L.). *G3*, **2**, 607–17.
11. Sato, S., Nakamura, Y., Kaneko, T., et al. 2008, Genome structure of the legume, *Lotus japonicus*. *DNA Res.*, **15**, 227–39.
12. Schmutz, J., Cannon, S.B., Schlueter, J., et al. 2010, Genome sequence of the palaeopolyploid soybean, *Nature*, **463**, 178–83.

13. Young, N.D., Debellé, F., Oldroyd, G.E., et al. 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature*, **480**, 520–4.
14. Varshney, R.K., Chen, W., Li, Y., et al. 2012, Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol*., **30**, 83–9.
15. Zhu, H., Choi, H.K., Cook, D.R., Shoemaker, R.C. 2005, Bridging model and crop legumes through comparative genomics. *Plant Physiol*., **137**, 1189–96.
16. Varshney, R.K., Song, C., Saxena, R.K. et al. 2012, Draft genome sequence of Chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol*., **31**, 240–6.
17. Schmutz, J., Mc Clean, P.E., Mamido, S. et al. 2014, A reference genome for common bean and genome wide analysis of dual domestications. *Nature Genet*., **46**, 707–13.
18. Bertioli, D.J., Cannon, S.B., Froenicke, L. et al. 2016, The genome sequences of *Arachis duranensis* and *Arachis ipaensis* the diploid ancestors of cultivated peanut. *Nature Genet*. doi:10.1038/ng.3517.
19. Giraud, E., Moulin, L., Vallenet, D. et al. 2007, Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science*, **316**, 1307–1312.
20. Sprent, J.I. and James, E.K. 2008, Legume-rhizobial symbiosis: an anorexic model? *New Phytol*., **179**, 3–5.
21. Arrighi, J.F., Cartieaux, F., Brown, S.C., et al. 2012, *Aeschynomene evenia*, a model plant for studying the molecular genetics of the Nod-independent rhizobium-legume symbiosis. *Mol. Plant-Microbe Interact*., **25**, 851–861.
22. Arrighi, J.F., Cartieaux, F., Chaintreuil, C., et al. 2013, Genotype delimitation in the Nod-independent model legume *Aeschynomene evenia*. *PloS One*, **8**, e63836.
23. Rubio-Piña, J.A. and Zapata-Pérez, O. 2011, Isolation of total RNA from tissues rich in polyphenols and polysaccharides of mangrove plants. *Plant Biotech*., doi:10.2225/Vol02304-issue5-fulltext-8.
24. Combes, M.C., Dereeper, A., Severac, D., et al. 2013, Contribution of subgenomes to the transcriptomes and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytol*., **200**, 251–60.
25. Allegre, M., Argoult, X., Boccara, M., et al. 2012, Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. *DNA Res*., **19**, 23–35.
26. Thoquet, P., Ghérardi, M., Journet, E.P., et al. 2002, The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. *BMC Plant Biology*, **2**, 1.
27. Gourion, B., Berrabah, F., Ratet., P., et al. 2014, Rhizobium-legume symbioses: the crucial role of plant immunity. *Trends Plant Sci*., **20**, 186–94.
28. Reid, D.E., Ferguson, B.J., Hayashi, S., et al. 2011, Molecular mechanisms controlling legume autoregulation of nodulation. *Ann. Bot*., **108**, 789–95.
29. Couzigou, J.M., Zhukov, V., Mondy, S., et al. 2012, NODULE ROOT and COCHLEATA maintain nodule development and are legume orthologs of *Arabidopsis* BLADE-ON-PETIOLE genes. *Plant Cell*., **24**, 4498–510.
30. Suzaki, T., Ito, M., Yoro, E., et al. 2014, Endoreduplication-mediated initiation of symbiotic organ development in *Lotus japonicus*. *Development*, **141**, 2441–5.
31. Yoon, H.J., Hossain, M.S., Held, M., et al. 2014, *Lotus japonicus* SUNERGOS1 encodes a predicted subunit A of a DNA topoisomerase VI that is required for nodule differentiation and accommodation of rhizobial infection. *Plant J*., **78**, 811–21.
32. Kawaharada, Y., Kelly, S., Nielsen, M.W., et al. 2015, Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature*, **523**, 308–12.
33. Fliegmann, J., Canova, S., Lachaud, C., et al. 2013. Lipo-chitooligosaccharidic symbiotic signals are recognized by LysM receptor-like kinase LYR3 in the legume *Medicago truncatula*. *ACS Chem. Biol*., **8**, 1900–6.
34. Fabre, S., Gully, D., Poitout, A., et al. 2015, The Nod factor-independent nodulation in *Aeschynomene evenia* required the common plant-microbe symbiotic "toolkit". *Plant Physiol*., **169**, 2654–64.
35. Czernic, P., Gully, D., Cartieaux, F., et al. 2015, Convergent evolution of endosymbiont diffenrentiation in Dalbergioid and IRLC legumes mediated by nodule-specific cysteine-rich peptides. *Plant Physiol*., **169**, 1254–65.
36. Xu, P., Wu, X., Wang, B., et al. 2011, A SNP and SSR based genetic map of Asparagus Bean (*Vigna unguiculata* ssp. *sesquipedialis*) and comparison with the broader species. *PLoS ONE*, **6**, e15952.
37. Sharpe, A.G., Ramsay, L, Sanderson, L.A., et al, Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. *BMC Genomics*, **14**, 192.
38. Kamphuis, L.G., Williams, A.H., D'Souza, N.K., et al. 2007, The *Medicago truncatula* reference accession A17 has an aberrant chromosomal configuration. *New Phytol*., **174**, 299–303.
39. Hayashi, M., Miyahara, A., Sato, S., et al. 2001, Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F2 population. *DNA Res*., **8**, 301–10.
40. Sandal, N., Krusell. L., Radutoiu, S., et al. 2002, A genetic linkage map of the model legume *Lotus japonicus* and strategies for fast mapping of new loci. *Genetics*, **161**, 1673–83.
41. Gujaria-Verma, N., Vail, S.L., Carrasquilla-Garcia, N., et al. 2014, Genetic mapping of legume orthologs reveals high conservation of synteny between lentil species and the sequenced genomes of *Medicago* and chickpea. *Front. Plant. Sci*., **5**, 676.
42. Nielen, S., Vigidal, B.S., Leal-Bertioli, S.C.M., et al. 2012, *Matita*, a new retroelement from peanut: characterization and evolutionary context in the light of the *Arachis* A-B genome divergence. *Mol. Genet. Genomics*, **287**, 21–38.
43. Sandal, N., Petersen, T.R., Murray, J., et al. 2006, Genetics of symbiosis in *Lotus japonicus*: recombinant inbred lines, comparative genomic maps, and map position of 35 symbiotic loci. *MPMI*, **16**, 80–91.
44. Sandal, N., Jin, H., Rodriguez-Navarro, D.N., et al. 2012, A set of *Lotus japonicus* Gifu x *Lotus burttii* recombinant inbred lines facilitates map-based cloning and QTL mapping. *DNA Res*., **19**, 317–23.
45. Murray, J., Karas, B., Ross, L., et al. 2006. Genetic suppressors of the *Lotus japonicus har1-1* hypernodulation phenotype. *MPMI*, **19**, 1082–91.
46. Kim, M., Chen, Y., Xi, J., et al. An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis. *Proc. Natl. Acad. Sci. USA*., **112**, 15238–43.
47. Schneeberger, K., Ossowski, S., Lanz, C., et al. 2009, SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods*, **6**, 550–1.
48. Aba, A., Kosugi, S., Yoshida, K., et al. 2012, Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotech*. **30**, 174–9.