G3, 2024, **14(3)**, jkad284
https://doi.org/10.1093/g3journal/jkad284
Advance Access Publication Date: 14 December 2023
Genome Report

Chromosome-level genome of the transformable northern wattle, *Acacia crassicarpa*

Isabelle Massaro,[1] Richard Scott Poethig ,[2] Neelima R. Sinha ,[3] Aaron R. Leichty [1,3,4,5,*]

[1]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA 94720, USA
[2]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA
[3]Department of Plant Biology, University of California Davis, Davis, CA 95616, USA
[4]USDA Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
[5]Present address: 800 Buchanan Street, Albany, CA 94710, USA

*Corresponding author: USDA Plant Gene Expression Center, 800 Buchanan Street, Albany, CA, USA 94710. Email: aaron.leichty@usda.gov

The genus *Acacia* is a large group of woody legumes containing an enormous amount of morphological diversity in leaf shape. This diversity is at least in part the result of an innovation in leaf development where many *Acacia* species are capable of developing leaves of both bifacial and unifacial morphologies. While not unique in the plant kingdom, unifaciality is most commonly associated with monocots, and its developmental genetic mechanisms have yet to be explored beyond this group. In this study, we identify an accession of *Acacia crassicarpa* with high regeneration rates and isolate a clone for genome sequencing. We generate a chromosome-level assembly of this readily transformable clone, and using comparative analyses, confirm a whole-genome duplication unique to Caesalpinoid legumes. This resource will be important for future work examining genome evolution in legumes and the unique developmental genetic mechanisms underlying unifacial morphogenesis in *Acacia*.

Keywords: *Acacia crassicarpa*; Acra3RX; Caesalpinioideae; northern wattle; transformation; unifacial morphogenesis; whole-genome duplication

Introduction

The genus *Acacia* is an immense group of woody legumes from Australia and South East Asia that arose nearly ~23 MYA and diversified into over 1,000 species (Renner *et al.* 2020; McLay *et al.* 2022). Many of these species are of economic, cultural, and agricultural importance, and several are noxious invasives in other regions of the world (Miller *et al.* 2011). One of the defining features of the genus is the morphology of their adult leaves. All *Acacia* species begin vegetative development producing bifacial compound leaves that are horizontally positioned relative to the meristem and have dorsal–ventral (or abaxial and adaxial) polarity (Gardner *et al.* 2008). In the vast majority of species, the shoot transitions to producing simple leaves with unifacial morphology (i.e. all leaf surfaces are of identical morphology; Fig. 1a). In some species, the leaves are radially symmetric, but in most, the leaf blade has been flattened to produce a photosynthetic organ homologous to a bifacial leaf blade (Boke 1940; Kaplan 1980).

The evolution of unifacial development in *Acacia* has resulted in an astounding array of leaf morphologies (Fig. 1c). Leaves range extensively in size and shape, from the small triangular phyllodes of *Acacia cultriformis* to the long narrowly elliptic phyllodes of *A. macradenia* (Fig. 1c). A great amount of this variation is associated with adaptation to some of the most extreme climatic variations in Australia, and a number of hypotheses have been put forward for the functional significance of the unifacial phyllode (Pasquet-Kok *et al.* 2010; Renner *et al.* 2021). In at least some of these species, it is likely that the vertical orientation of the unifacial blade plays a role in thermoregulation and light inception (Liu *et al.* 2003) similar to what has been reported for another major Australian genus of trees, the *Eucalyptus* (King 1997).

Despite the fact that unifacial leaves are often superficially quite similar to bifacial leaves, the mechanisms for their development remain a mystery. In bifacial leaves, blade growth occurs in the medio-lateral direction as a consequence of adaxial and abaxial polarity. When genes that specify ad- or abaxial polarity are lost, the leaf becomes radially symmetric, losing the production of a flattened lamina (e.g. Waites and Hudson 1995; Kerstetter *et al.* 2001). However, in some unifacial monocots, and almost all *Acacia* species, growth of the leaf blade occurs predominately in the ad-abaxial plane, suggesting the evolution of novel mechanisms for expansion independent of ad-abaxial polarity (Yamaguchi *et al.* 2010; Nakayama *et al.* 2022; Golenberg *et al.* 2023). To adequately investigate these physiologically important developmental phenomena, improvements in genetic and genomic resources are needed in these nonmodel systems.

In an effort to establish *Acacia* as a model for the study of variation in leaf development and developmental timing, we have cloned a line of *A. crassicarpa*, one of the few *Acacia* species where genetic transformation is readily available. Using this highly amenable clone, Acra3RX (Fig. 1b), we generate a haploid assembly to the chromosomal level and present evidence that at least some Caesalpinioid legumes have undergone a whole-genome duplication (WGD) event independent of other Papilionoid legumes.

Received on 13 September 2023; accepted on 01 December 2023
Published by Oxford University Press on behalf of The Genetics Society of America 2023.

Materials and methods

Plant material

For the genome assembly of *A. crassicarpa*, leaf tissue from a single individual of accession 19724 from the Australian Tree Seed Center was used. This individual, Acra3RX, was clonally propagated in sterile culture and clones were also grown in the greenhouse for generating large amounts of tissue for HMW DNA extraction.

For RNA-seq, seeds were germinated by clipping the seed coat and placed in a petri dish with a moist paper towel at room temperature for 3–4 days. The seeds were transferred to 2″×2″ pots containing Fafard-2 and grown in a growth chamber maintained at 25°C, with 16 h light/8 h dark, and 190–220 μmol m^{-2} s^{-1} using white fluorescent lights. Apices, including leaf primordia up to 3 mm and associated stipules for 2 developmental stages, were sampled. The first sample included nodes 1–3, where leaves are only pinnate or bipinnately compound. The second sample included nodes only after each individual plant had switched to producing phyllodes (on average node 6). The samples were pooled and total RNA was extracted using the Spectrum Plant Total RNA Kit (Sigma). Additionally, a second RNA sample was sequenced from whole ramets grown in culture.

In vitro culture

Ten, single and multiparent seed accessions (Acra1–Acra11) were screened for their ability to propagate and regenerate in culture (Supplementary Table 1). To do this, individual plants were cloned for each accession and propagated by shoot culture in magenta boxes. Specifically, to start sterile cultures of single plants, seed coats were clipped and then boiled for 30 s, followed by 3 washes with sterile water. Seeds were grown in magenta boxes on germination media consisting of ½ MS basal media, 1% sucrose, 1.3 mM MES, 0.7% agar, and a pH of 5.8. Plants surviving sterilization and free of contamination were further propagated on rooting media as previously outlined (Yang *et al.* 2006). After the establishment of 3–6 ramets per individual plant, phyllode leaf primordia were sampled for regeneration tests. Briefly, leaf primordia ranging from 1 to 4 cm were excised from shoots, placed on a sterile paper towel, and cut transversely into 1 cm pieces. The resulting explants were cultured for 45 days on regeneration media as previously described (Yang *et al.* 2006), except with the addition of 0.75 mg/L meta-topolin and 0.125 mg/L of trans-zeatin. For each accession, 7–9 individual plants were individually evaluated for their ability to regenerate with a minimum of 5 explants per cloned plant. For an analysis of root regeneration, ramets were scored as either "good rooting" if primary roots developed lateral roots, "poor rooting" if only primary roots were present, or "no root" if ramets formed no roots. Data were collected at 2 months with 14–18 ramets evaluated per accession. Shoot height was also measured at this time. Figures for in vitro analyses were generated using R, v3.6.3 (R Core Team 2020).

Estimation of genome size and ploidy

Genomic DNA was extracted from leaves of an in vitro grown clone of Acra3RX using the Qiagen Plant DNeasy Kit (Redwood City, CA, USA). A TruSeq library was prepared and sequenced by Admera Health (South Plainfield, NJ, USA) following the manufacturer's protocol (Illumina, San Diego, CA, USA). Genome size was estimated using a kmer-based method as implemented in GenomeScope2 (Ranallo-Benavidez *et al.* 2020). Raw reads from an Illumina library (ACRA3RX_WGS2; Supplementary Table 2) were used with KMC for kmer counting (Kokot *et al.* 2017). These reads were also used with a smudgeplot to confirm the ploidy of Acra3RX (Ranallo-Benavidez *et al.* 2020).

PacBio library and sequencing

HMW DNA was extracted from young leaf primordia and shoot apices of an adult clone of Acra3RX using an in-house CTAB protocol at Cantata Bio (Scotts Valley, CA, USA). DNA samples were quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20 kb) for PacBio Sequel was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) and the manufacturer-recommended protocol. The library was bound to polymerase using Sequel II Binding Kit 2.0 (PacBio) and loaded onto PacBio Sequel II. Sequencing was performed on PacBio Sequel II 8M SMRT cells.

Dovetail Omni-C library preparation and sequencing

A Dovetail Omni-C library was prepared by Cantata Bio following the manufacturer's protocol. In brief, chromatin was fixed in place with formaldehyde in the nucleus. Fixed chromatin was digested with DNase I and then extracted chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed and the DNA purified. The purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. Raw reads (ACRA3RX_OmniC1; Supplementary Table 2) were processed using the ArimaGenomics mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline).

mRNA-seq of whole plant and mixed apices

RNA from whole, rooted in vitro-grown clones of Acra3RX was extracted using the Sigma Spectrum Plant Total RNA Kit and on-column DNase digestion treatment method (St. Louis, MO, USA). The NEBNext Ultra II Directional RNA Library Prep Kit was used to construct a Poly A library according to the manufacturer's protocol (New England Biolabs, Ipswich, MA, USA). The library was sequenced with PE150 format by Admera Health.

RNA from shoot apices with leaf primordia <3 mm was extracted using the Sigma Spectrum Plant Total RNA Kit and on-column DNase digestion (St. Louis, MO, USA). Each extraction consisted of a pool of 5–10 apices from plants grown from the seed accession 19724 from the Australian Tree Seed Centre. For each extraction, apices were either from nodes 2 to 3, with leaf primordia having juvenile morphology, or from nodes >7, having adult morphology. RNA concentration was measured using a Qubit Fluorometer (Life Technologies), and a single pool of RNA was created using equal amounts of 4 juvenile extractions and 4 adult extractions. The NEBNext Ultra II Directional RNA Library Prep Kit was used to construct a Poly A library according to the manufacturer's protocol (New England Biolabs). The library was sequenced with PE150 format by Admera Health.

Genome assembly

PacBio HiFi reads were assembled using hifiasm and Omni-C reads to generate haplotype-resolved assemblies (Cheng *et al.* 2021, 2022). Each haplotype was then used for scaffolding using SALSA with the same Omni-C reads (Ghurye *et al.* 2017, 2019). The larger of the 2 hifiasm assemblies (1,151 contigs) was used as the primary assembly, unless otherwise noted.

Gene prediction and annotation

De novo identification and classification of repeats was done using RepeatModeler v.2.0.1 (Flynn *et al.* 2020). RepeatMasker v.4.0.7 (Smit *et al.* 2013) was used with the *A. crassicarpa* repeat library to soft-mask the genome using the following parameters: -nolow -norna -xsmall.

BRAKER3 was used to annotate genes on the soft-masked genome (Hoff *et al.* 2016, 2019; Brůna *et al.* 2021; Gabriel *et al.* 2023). First, fastp was used for trimming and quality filtering RNA-seq reads (Chen *et al.* 2018). Hisat2 v2.2.1 (Kim *et al.* 2019) was used to map the RNA-seq libraries and the resulting mapping files were supplied to the BRAKER pipeline along with UniProt proteins from the Viridiplantae dataset (The UniProt Consortium 2021). This pipeline uses StringTie (Pertea *et al.* 2015) to assemble the RNA-seq reads followed by rounds of GeneMark and AUGUSTUS training and gene prediction (Stanke and Waack 2003). Gene sets were combined with TSEBRA (Gabriel *et al.* 2021).

Genome assessment and comparative genomics

Assessment of genome completeness was conducted with BUSCO v5 using the Embryophyta *odb10* dataset (Manni *et al.* 2021). Alignment of DNA-seq reads was done with bowtie2 v2.4.4 using the following options: –end-to-end –sensitive (Langmead and Salzberg 2012). Alignment of RNA-seq reads was done with Hisat2 v2.2.1 (Kim *et al.* 2019). Feature counting of RNA-seq reads was done using htseq-count (Anders *et al.* 2015).

Alignment of the hard-masked *A. crassicarpa* genome with the *A. melanoxylon* (Jones *et al.* 2021, 2023) and *A. pycnantha* (McLay *et al.* 2022) genomes was done using MashMap (Jain *et al.* 2018) and the following parameters: -f none –pi 98. Micro and macrosynteny analyses were done using the MCscan pipeline (Tang *et al.* 2008). In particular, comparisons with *Medicago truncatula* were done using Mt4.0v1 (Tang *et al.* 2014), Lj1.0v1 was used for comparisons with *Lotus japonicus* (Li *et al.* 2020), ISU-01 v2.1 was used for *Glycine max* (https://phytozome-next.jgi.doe.gov/info/GmaxWm82ISU_01_v2_1), and MpudA1P6v1 was used for *Mimosa pudica* (Libourel *et al.* 2023). Details of this analysis can be found at the DRYAD submission associated with this paper.

Results and discussion

Isolation of an amenable *A. crassicarpa* line

The massive amount of within-generation (Fig. 1a) and between-species (Fig. 1c) variation in unifacial leaf morphology makes the genus *Acacia* one of the most tantalizing groups for the study of this alternative form of leaf development. To date, stable transformation of *Acacia* has been reported only in *A. mangium* (Xie and Hong 2002) and *A. crassicarpa* (Yang *et al.* 2008). In both of these systems, explant material is usually derived from seedlings or surface-sterilized adult phyllodes. Given the extensive genetic diversity in these woody outcrossing species, we decided to screen a set of 10 single and multiparent seed accessions for their ability to propagate and regenerate in culture (Supplementary Table 1). Across all accessions (Acra1–Acra11), shoots were capable of forming de novo roots in culture, with the best accessions, Acra5, Acra9, and Acra10 forming at least a primary root at 2 months across all ramets tested (Supplementary Fig. 1a). However, rooting performance showed a low correlation with the ability of phyllode explants to regenerate by 45 days, with Acra5, Acra9, and Acra10 regenerating at rates of 3%, 0%, and 7%, respectively (Supplementary Fig. 1b). The highest shoot regenerating accession was Acra3 at a rate of 67% (Supplementary Fig. 1b), which was an 11% improvement over the previously reported rate for *A. crassicarpa* at 2 months of culture (Yang *et al.* 2006). While Acra3 tended to have one of the lower growth rates (Supplementary Fig. 1c), its rooting performance was comparable to other accessions, forming at least a primary root on 94% of ramets (Supplementary Fig. 1a). From this accession, we cloned a single seedling, Acra3RX, which has shown great promise for rapid transformation protocols (Fig. 1b) and was used for genome sequencing and assembly.

Genome sequencing and assembly of Acra3RX

To sequence the genome of Acra3RX, we generated 2.58M PacBio HiFi reads with a median length of 12,069 bp (Supplementary Table 2). Given the previous estimate of 745 Mb for the size of the *A. crassicarpa* genome (Gallagher *et al.* 2011), this represents 42× coverage. We additionally generated 83.5M Dovetail OmniC chromosome conformation capture sequencing reads (Supplementary Table 2). Together, these data were used to generate haplotype-resolved assemblies. The primary haplotype assembly consisted of 1,151 contigs totalling 792 Mb, and an NG50 and LG50 of 17.7 Mb and 15 contigs, respectively. The second haplotype assembly consisted of 660 contigs totalling 782 Mb, and NG50 and LG50 of 26.5 and 10 contigs, respectively. Using the OmniC reads, we scaffolded the primary assembly, which resulted in a final assembly of 1,106 scaffolds and an NG50 and LG50 of 56 Mb and 6 scaffolds, respectively (Table 1). In total, there were 15 scaffolds larger than 10 Mb in the final assembly, approximating the haploid chromosome number for *Acacia* ($n = 13$). Additionally, after scaffolding, the second haplotype assembly had 13 scaffolds larger than 10 Mb, which constituted 98% of the predicted genome size.

As a measure of genome completeness, 94% of reads from an Illumina library (Supplementary Table 2) mapped to the 24 scaffolds that were longer than 1 Mb, and 99% mapped to the total 1,106 scaffolds. A kmer based estimate of genome size using this same library found the Acra3RX genome to be 647 Mb, and confirmed its status as a diploid (Supplementary Fig. 2). Given the fact that previous studies reported underestimates of genome size by kmer methods, in this study, we used an expected genome size of 745 Mb (Gallagher *et al.* 2011).

Comparison with other Mimosoid genomes

Until recently, most of the available genomes for Mimosoid species were constructed using short-read libraries [e.g. *Faidherbia albida*, (Chang *et al.* 2019); *M. pudica*, (Griesmann *et al.* 2018); *Vachellia collinsii*, (Leichty and Poethig 2019)]. However, there is now a growing number of long-read assemblies that have far exceeded the assembly metrics of these earlier assemblies [e.g. *Prosopis alba*, University of Georgia NCBI direct submission; *M. pudica* (Libourel *et al.* 2023); *A. pycnantha*, (McLay *et al.* 2022)]. Of these, only the *M. pudica* genome (MpudA1P6v1) approaches a chromosome-level assembly (Table 1) with 45 scaffolds over 10 Mb in size for this tetraploid species ($n = 26$). The genome of *A. melanoxylon*, which is of comparable size and is phylogenetically close to *A. crassicarpa* (Miller *et al.* 2011), is also near-chromosomal, with 20 scaffolds larger than 10 Mb (Table 1). Comparatively, our assembly of *A. crassicarpa* is similar to, or exceeds, these assemblies, with 15 scaffolds >10 Mb.

A comparison of the *A. crassicarpa* assembly with *A. melanoxylon* demonstrates the high contiguity of the assembly. In general, many *A. melanoxylon* contigs align to single *A. crassicarpa* scaffolds, and there are only a few examples where an *A. crassicarpa* scaffold does not span multiple *A. melanoxylon* contigs (Supplementary Fig. 3b).

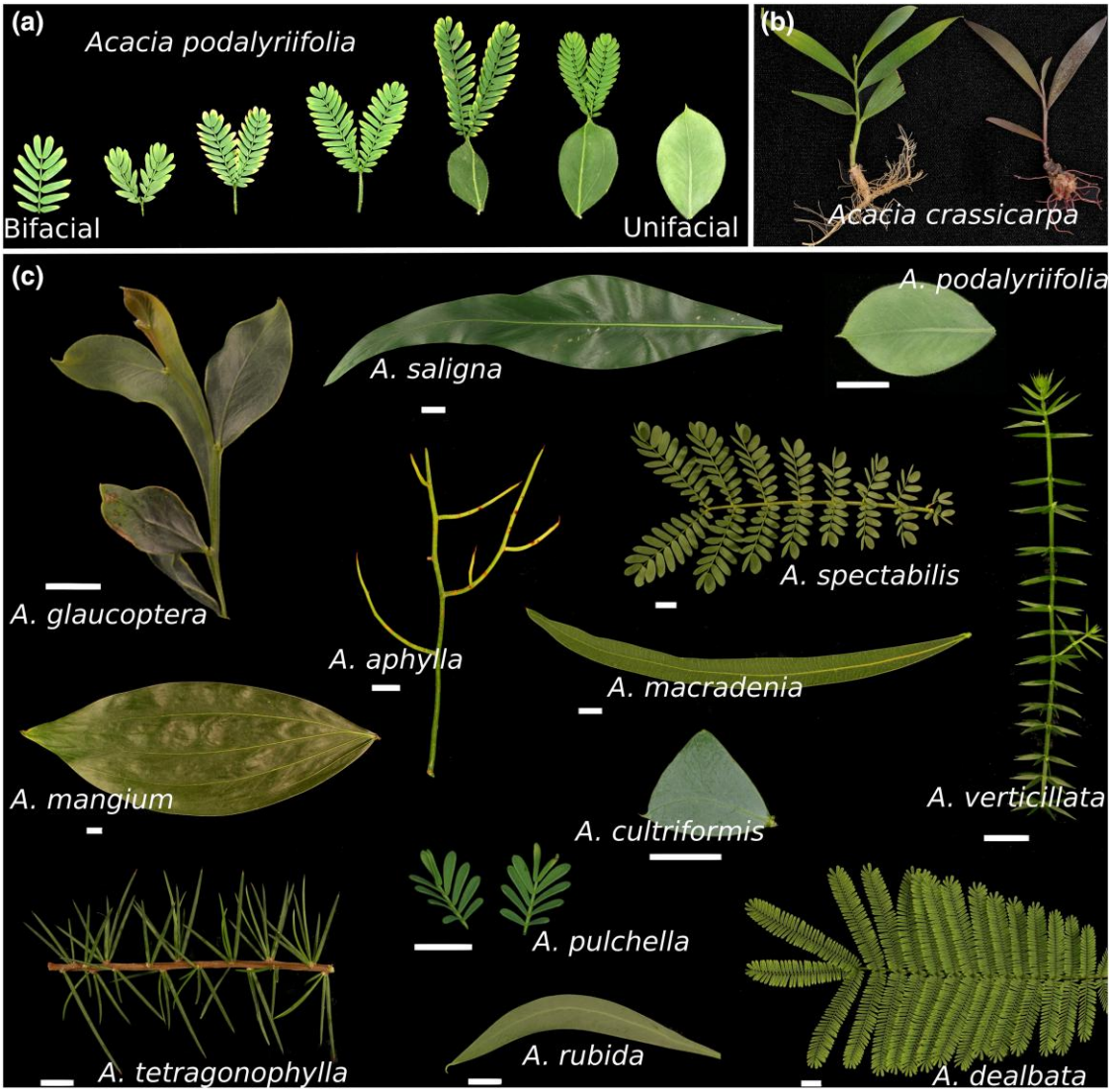

Fig. 1. Leaf morphological diversity in the genus *Acacia*. a) An example of the age-dependent transition from bifacial to unifacial leaf development in *Acacia podalyriifolia*. b) In vitro-grown ramets of clonal *A. crassicarpa* line, Acra3RX (left), and version transformed with a construct expressing the RUBY marker system (right; He *et al.* 2020). c) Examples of adult leaf diversity. *Acacia glaucoptera*, *A. teragonophylla*, *A. aphylla*, and *Asclepias verticillata* images are of whole shoots and not of single leaves. Scale bars have a length of 1 cm.

Table 1. A comparison of the *A. crassicarpa* assembly with other recent Mimosoid genomes.

Species	*A. crassicarpa*[a]	*A. pycnantha* McLay *et al.* (2022)	*A. melanoxylon* Jones *et al.* (2023; 2021)	*A. acuminata* (NCBI ASM1902265v1)	*F. albida*[c] Chang *et al.* (2019)	*M. pudica* Libourel *et al.* (2023)
Chromosome number (*n*)	13	13	13	13	13	26
Predicted genome size (Mb)	745	850	780	907	661	900
Assembly size (Mb)	751	814	749	880	640	797
Number of scaffolds	62	1272	188	633	9058	74
Scaffolds > 0.1 M	62	629	166	520	1113	74
Scaffolds > 1 M	24	226	125	287	141	60
Scaffolds > 10 M	15	8	20	3	0	45
Number of predicted protein-coding genes	29,488	47,624	NA	NA	28,979	73,541
Busco (embryophyta)[b]	C:99.4% S:94.1% D:5.3%	C:99.0% S:86.4% D:12.6%	C:99.3% S:91.1% D:8.2%	C:99.2% S:87.7%, D:11.5%	C:99.4% S:97.3% D:2.1%	C:99.3% S:3.8% D:95.5%

[a] The assembly with scaffolds larger than 1 Mb was used for annotating protein-coding genes, and scaffolds larger than 100 kb were used for other summary statistics in this table.
[b] Percent of embryophyta gene models that are complete (C), and a portion that exists as single copy (S) or duplicates (D) in the assembly.
[c] Assembly was filtered for scaffolds larger than 1 kb.

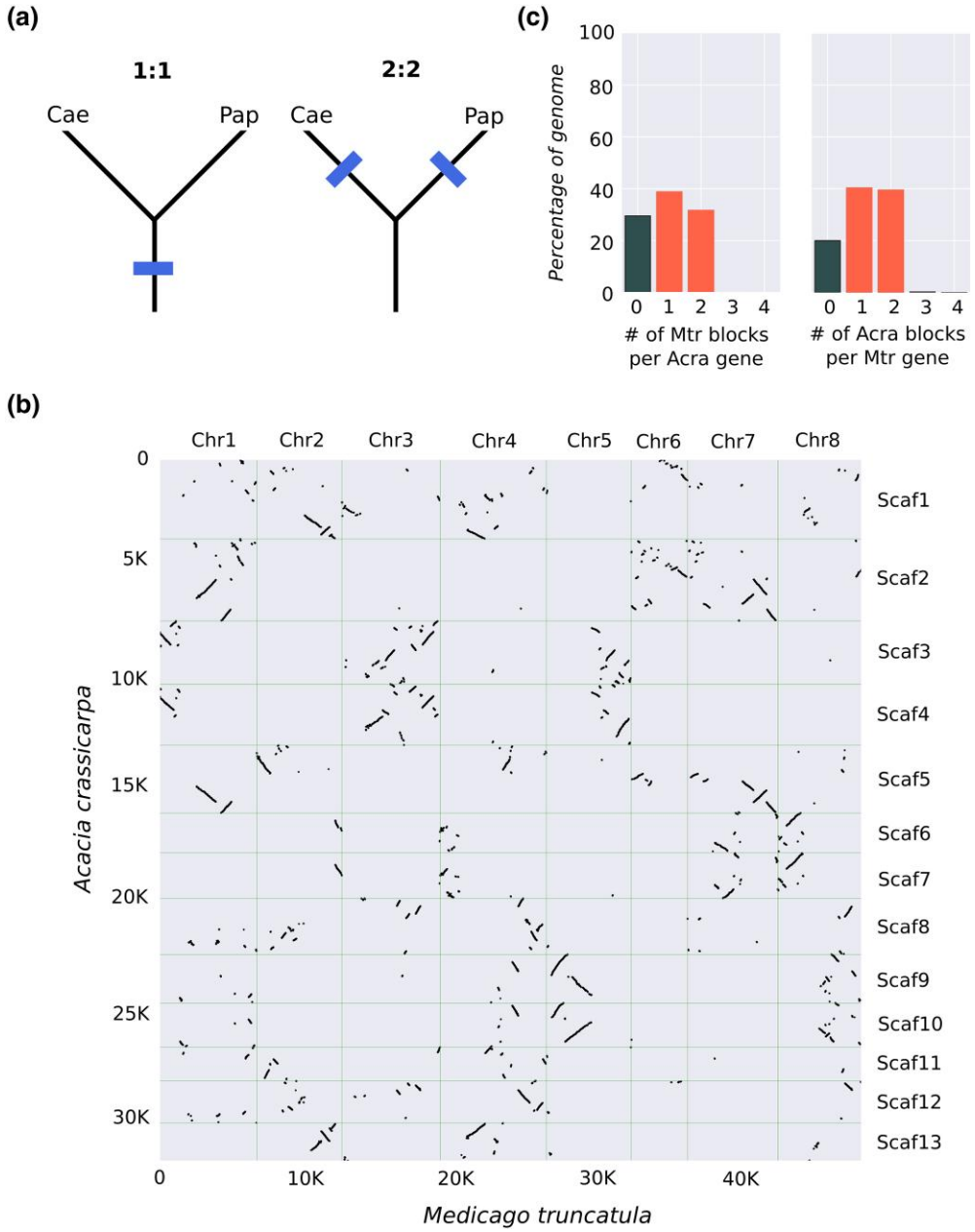

Fig. 2. A comparison of the *A. crassicarpa* genome with a *M. truncatula*. a) Predicted syntenic depth between a Caesalpinoid (Cae) and a Papilionoid (Pap) genome under the 2 leading hypothesis for WGD within legumes. Either the WGD event previously confirmed for Papilionoid species was shared with Caesalpinoid species (left tree) or the event was independent and a second event unique to the Caesalpinoid lineage happened after subfamily divergence (right tree). The former scenario results in a 1:1 ratio between orthologous blocks, while the latter would produce a 2:2 ratio (i.e. 2 orthologous blocks per gene). b) A comparison of the *A. crassicarpa* genome with *M. truncatula*. The cumulative number of gene orthologs is plotted on the y- and x-axes as a function of position along the *M. truncatula* chromosomes (horizontal axes) and *A. crassicarpa* scaffolds (vertical axes). c) Syntenic depth for orthologous genomic blocks from either the *A. crassicarpa* or the *M. truncatula* genome (i.e. the number of orthologous blocks for a gene from the comparison species).

These observations are even more clear when compared with the recently assembled *A. pycnantha* genome where all *A. crassicarpa* scaffolds span more than one *A. pycnantha* scaffold. The chromosome scale of the assembly was further demonstrated by a comparison of *A. crassicarpa* with the newly assembled *M. pudica* genome. Alignment of these genomes revealed that *A. crassicarpa* scaffolds are entire chromosomes, while chromosome arms remain unlinked in the current *M. pudica* assembly (Supplementary Fig. 4). In general, these genome-wide comparisons demonstrate a high degree of consistent genomic content, with very little evidence for large-scale duplication events (either real or artifactual) in the assemblies (Supplementary Figs. 3a, 3b, and 4).

The completeness of the *A. crassicarpa* genome was measured by the presence of embryophyta single-copy orthologs, with 99.4% present as complete gene copies, of which 94.1% were present as single copies and 5.3% as duplicates (Table 1). These values were nearly identical across all of the available Mimosoid genomes, although the *A. pycnantha*, *A. melanoxylon*, and *A. acuminata* genomes had duplication rates ranging from 8.2% to 12.6% compared with 5.3% in *A. crassicarpa* and 2.1% in *F. albida* (as a tetraploid *M. pudica* had a predictably high level of 95.5%; Table 1).

These latter 2 species also had a similar number of predicted protein-coding genes, with *A. crassicarpa* having 29,488 and *F. albida* having 28,979, compared with 47,624 in *A. pycnantha*. It will be interesting to see how these gene counts change with improvements on each genome. However, we do not anticipate significant improvements for *A. crassicarpa*, given that 95% of the reads from our RNA-seq of Acra3RX mapped to the 24 largest scaffolds in the *A. crassicarpa* genome, of which 90% mapped to annotated protein genes. This suggests that the current scaffolds contain most of the expressed genomes.

Evidence for a Caesalpinioideae WGD event

Genome duplication events are known to have played a major role in the evolution of legumes (Young *et al.* 2011). However, it remains unclear whether the WGD event detected in the subfamily Papilionoideae was shared or independent of a WGD event detected in Caesalpinioideae (which contains Mimosoid legumes; Cannon *et al.* 2015; Koenen *et al.* 2021; Zhao *et al.* 2021; McLay *et al.* 2022). Most recently, using synonymous site analysis in one-to-one orthologs, McLay *et al.* (2022) found evidence for independent Papilionoid and Caesalpinioid WGD events. If this were true, then it would be expected that when the *A. crassicarpa* genome is compared with a Papilionoid species such as *M. truncatula*, synthenic blocks would show a 2:2 ratio (i.e. there would be 2 syntenic *M. truncatula* blocks for each orthologous *A. crassicarpa* gene, and vice versa; Tang *et al.* 2011). Conversely, if the WGD is shared between the 2 subfamilies, then syntenic blocks would be expected to have a 1:1 pattern (Fig. 2a).

Alignment of *A. crassicarpa* to *M. truncatula* revealed that 2 independent WGD events appear to have occurred (i.e. one in each subfamily; Fig. 2). In *A. crassicarpa*, there are 3 sets of scaffolds (scaffold 3 and 4, 6 and 7, 9 and 10) that show highly correlated alignment patterns to the *M. truncatula* genome. Scaffold 1 and 13, and 2 and 5, also show a high degree of concordance, suggesting their shared evolutionary origin (Fig. 2b). Importantly, a comparison of *M. truncatula* and *A. crassicarpa* shows a 2:2 pattern, with 2 orthologous syntenic blocks per orthologous gene in 40% and 32% of the *M. truncatula* and *A. crassicarpa* genomes, respectively (Fig. 2c). Comparisons with other Papilionoid species, *L. japonicus* and *G. max*, were consistent with the observations from *M. truncatula*, further supporting that WGD events occurred independently in Caesalpinioid and Papilionoid legumes (Supplementary Table 3). Additionally, comparisons within subfamilies consistently show 1:1 syntenic depths between species (or 1:2, in the case of *G. max* and *M. pudica*), which is expected if WGD events are shared between these species (Supplementary Table 3).

An analysis of the putative chromosomal duplications within *Acacia* revealed a high degree of conservation. For example, an analysis of scaffolds 9 and 10, which contain duplicate copies of the single-copy gene first identified in *M. truncatula*, *PALMATE-LIKE PENTAFOLIATA1* (*PALM1*), reveals high levels of macro-scale (Supplementary Fig. 5a) and micro-scale synteny (Supplementary Fig. 5b and c). This synteny is highly conserved within *Acacia*, but seems to largely be lost between the putative homologous regions of chromosomes 5 and 8 in *M. truncatula* (where chromosome 5 shows high syntenic conservation, but chromosome 8 shows a fractionated pattern for the same region, Fig. 2b). Given the large adaptive radiations associated with the genus *Acacia*, and many lineages within the Caesalpinioid subfamily (Azani *et al.* 2017), it will be interesting to see how these duplication events have been conserved or lost in association with novel adaptations such as the unifacial leaves of *Acacia*.

Data availability

The genome assembly and raw RNA and DNA-seq data have been submitted to Genbank under the bioproject: PRJNA975180. A breakdown of the associated SRA accessions can be found in Supplementary Table 2. Genome annotations and additional assembly data can be found on DRYAD (https://doi.org/10.5061/dryad.573n5tbdr).

Supplemental material available at G3 online.

Acknowledgments

We are thankful to the staff at Cantata Bio and Admera Health for their help in obtaining the sequencing data.

Funding

Funding for this work was provided by the United States Department of Agriculture (CRIS 2030-21000-054-000D). IM was supported by the University of California, Berkeley. NRS was supported by National Science Foundation IOS-211980. RSP was supported by National Institutes of Health GM51893. ARL was supported by a National Science Foundation Postdoctoral Research Fellowship in Biology (IOS-1812043) and the University of California Davis Katherine Esau Postdoctoral Fellowship, and is currently supported by the USDA-ARS.

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 31(2):166–169. doi:10.1093/bioinformatics/btu638.

Azani N, Babineau M, Bailey CD, Banks H, Barbosa AR, Pinto RB, Boatwright JS, Borges LM, Brown GK, Bruneau A, *et al.* 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny—the Legume Phylogeny Working Group (LPWG). Taxon. 66(1):44–77. doi:10.12705/661.3.

Boke NH. 1940. Histogenesis and morphology of the phyllode in certain species of Acacia. Am J Bot. 27(2):73–90. doi:10.2307/2436690.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 3(1):lqaa108. doi:10.1093/nargab/lqaa108.

Cannon SB, McKain MR, Harkess A, Nelson MN, Dash S, Deyholos MK, Peng Y, Joyce B, Stewart CN Jr, Rolf M, *et al.* 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Mol Biol Evol. 32(1):193–210. doi:10.1093/molbev/msu296.

Chang Y, Liu H, Liu M, Liao X, Sahu SK, Fu Y, Song B, Cheng S, Kariba R, Muthemba S, *et al.* 2019. The draft genomes of five agriculturally important African orphan crops. GigaScience. 8(3):giy152. doi:10.1093/gigascience/giy152.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34(17):i884–i890. doi:10.1093/bioinformatics/bty560.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly

graphs with hifiasm. Nat Methods. 18(2):170–175. doi:10.1038/s41592-020-01056-5.

Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, Li H. 2022. Haplotype-resolved assembly of diploid genomes without parental data. Nat Biotechnol. 40(9):1332–1335. doi:10.1038/s41587-022-01261-x.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 117(17):9451–9457. doi:10.1073/pnas.1921046117.

Gabriel L, Bruna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. 2023. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. bioRxiv. doi:10.1101/2023.06.10.544449.

Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. 2021. TSEBRA: transcript selector for BRAKER. BMC Bioinformatics. 22(1):566. doi:10.1186/s12859-021-04482-0.

Gallagher RV, Leishman MR, Miller JT, Hui C, Richardson DM, Suda J, Trávníček P. 2011. Invasiveness in introduced Australian acacias: the role of species traits and genome size. Divers Distrib. 17(5):884–897. doi:10.1111/j.1472-4642.2011.00805.x.

Gardner S, Drinnan A, Newbigin E, Ladiges P. 2008. Leaf ontogeny and morphology in Acacia Mill. (Mimosaceae). Muelleria. 26(1):43–50. doi:10.5962/p.292492.

Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. 2017. Scaffolding of long read assemblies using long range contact information. BMC Genomics. 18(1):527. doi:10.1186/s12864-017-3879-z.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 15(8):e1007273. doi:10.1371/journal.pcbi.1007273.

Golenberg EM, Popadić A, Hao W. 2023. Transcriptome analyses of leaf architecture in Sansevieria support a common genetic toolkit in the parallel evolution of unifacial leaves in monocots. Plant Direct. 7(8):e511. doi:10.1002/pld3.511.

Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, *et al.* 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science. 361(6398):eaat1743. doi:10.1126/science.aat1743.

He Y, Zhang T, Sun H, Zhan H, Zhao Y. 2020. A reporter for noninvasively monitoring gene expression and plant transformation. Hortic Res. 7(1):152. doi:10.1038/s41438-020-00390-1.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 32(5):767–769. doi:10.1093/bioinformatics/btv661.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. Methods Mol Biol. 1962:65–95. doi:10.1007/978-1-4939-9173-0_5.

Jain C, Koren S, Dilthey A, Phillippy AM, Aluru S. 2018. A fast adaptive algorithm for computing whole-genome homology maps. Bioinformatics. 34(17):i748–i756. doi:10.1093/bioinformatics/bty597.

Jones A, Stanley D, Ferguson S, Schwessinger B, Borevitz J, Warthmann N. 2023. Cost-conscious generation of multiplexed short-read DNA libraries for whole-genome sequencing. PLoS One. 18(1):e0280004. doi:10.1371/journal.pone.0280004.

Jones A, Torkel C, Stanley D, Nasim J, Borevitz J, Schwessinger B. 2021. High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing. PLoS One. 16(7):e0253830. doi:10.1371/journal.pone.0253830.

Kaplan DR. 1980. Heteroblastic leaf development in Acacia. morphological and morphogenetic implications. Cellule. 73:137–203.

Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS. 2001. KANADI regulates organ polarity in Arabidopsis. Nature. 411(6838):706–709. doi:10.1038/35079629.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 37(8):907–915. doi:10.1038/s41587-019-0201-4.

King DA. 1997. The functional significance of leaf angle in Eucalyptus. Aust J Bot. 45(4):619–639. doi:10.1071/bt96063.

Koenen EJM, Ojeda DI, Bakker FT, Wieringa JJ, Kidner C, Hardy OJ, Toby Pennington R, Herendeen PS, Bruneau A, Hughes CE. 2021. The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the cretaceous–paleogene (K–Pg) mass extinction event. Syst Biol. 70(3):508–526. doi:10.1093/sysbio/syaa041.

Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. Bioinformatics. 33(17):2759–2761. doi:10.1093/bioinformatics/btx304.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9(4):357–359. doi:10.1038/nmeth.1923.

Leichty AR, Poethig RS. 2019. Development and evolution of age-dependent defenses in ant-acacias. Proc Natl Acad Sci U S A. 116(31):15596–15601. doi:10.1073/pnas.1900644116.

Li H, Jiang F, Wu P, Wang K, Cao Y. 2020. A high-quality genome sequence of model legume *Lotus japonicus* (MG-20) provides insights into the evolution of root nodule symbiosis. Genes (Basel). 11(5):483. doi:10.3390/genes11050483.

Libourel C, Keller J, Brichet L, Cazalé AC, Carrère S, Vernié T, Couzigou JM, Callot C, Dufau I, Cauet S, *et al.* 2023. Comparative phylotranscriptomics reveals ancestral and derived root nodule symbiosis programmes. Nat Plants. 9(7):1067–1080. doi:10.1038/s41477-023-01441-w.

Liu L-X, Xu S-M, Woo KC. 2003. Influence of leaf angle on photosynthesis and the xanthophyll cycle in the tropical tree species *Acacia crassicarpa*. Tree Physiol. 23(18):1255–1261. doi:10.1093/treephys/23.18.1255.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 38(10):4647–4654. doi:10.1093/molbev/msab199.

McLay TGB, Murphy DJ, Holmes GD, Mathews S, Brown GK, Cantrill DJ, Udovicic F, Allnutt TR, Jackson CJ. 2022. A genome resource for Acacia, Australia's largest plant genus. PLoS One. 17(10):e0274267. doi:10.1371/journal.pone.0274267.

Miller JT, Murphy DJ, Brown GK, Richardson DM, González-Orozco CE. 2011. The evolution and phylogenetic placement of invasive Australian Acacia species. Divers Distrib. 17(5):848–860. doi:10.1111/j.1472-4642.2011.00780.x.

Nakayama H, Leichty AR, Sinha NR. 2022. Molecular mechanisms underlying leaf development, morphological diversification, and beyond. Plant Cell. 34(7):2534–2548. doi:10.1093/plcell/koac118.

Pasquet-Kok J, Creese C, Sack L. 2010. Turning over a new 'leaf': multiple functional significances of leaves versus phyllodes in Hawaiian Acacia koa. Plant Cell Environ. 33(12):2084–2100. doi:10.1111/j.1365-3040.2010.02207.x.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 33(3):290–295. doi:10.1038/nbt.3122.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 11(1):1432. doi:10.1038/s41467-020-14998-3.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna (Austria): R Foundation for Statistical Computing.

Renner MAM, Foster CSP, Miller JT, Murphy DJ. 2020. Increased diversification rates are coupled with higher rates of climate space exploration in Australian Acacia (Caesalpinioideae). New Phytol. 226(2):609–622. doi:10.1111/nph.16349.

Renner MAM, Foster C, Miller J, Murphy D. 2021. Phyllodes and bipinnate leaves of Acacia exhibit contemporary continental-scale environmental correlation and evolutionary transition-rate heterogeneity. Aust Syst Bot. 34(6):595–608. doi:10.1071/SB21009.

Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0. 2015. http://www.repeatmasker.org.

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and new intron submodel. Bioinformatics. 19(suppl_2): ii215–ii225. doi:10.1093/bioinformatics/btg1080.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. Science. 320(5875): 486–488. doi:10.1126/science.1153917.

Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, Gentzbittel L, Childs KL, Yandell M, Gundlach H, *et al.* 2014. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. BMC Genomics. 15(1):312. doi:10.1186/1471-2164-15-312.

Tang H, Lyons E, Pedersen B, Schnable JC, Paterson AH, Freeling M. 2011. Screening synteny blocks in pairwise genome comparisons through integer programming. BMC Bioinformatics. 12(1):102. doi:10.1186/1471-2105-12-102.

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49(D1):D480–D489. doi:10.1093/nar/gkaa1100.

Waites R, Hudson A. 1995. Phantastica: a gene required for dorsoventrality of leaves in *Antirrhinum* majus. Development. 121(7): 2143–2154. doi:10.1242/dev.121.7.2143.

Xie D, Hong Y. 2002. Agrobacterium-mediated genetic transformation of *Acacia mangium*. Plant Cell Rep. 20(10):917–922. doi:10.1007/s00299-001-0397-9.

Yamaguchi T, Yano S, Tsukaya H. 2010. Genetic framework for flattened leaf blade formation in unifacial leaves of Juncus prismatocarpus. Plant Cell. 22(7):2141–2155. doi:10.1105/tpc.110.076927.

Yang M, Xie X, He X, Zhang F. 2006. Plant regeneration from phyllode explants of *Acacia crassicarpa* via organogenesis. Plant Cell Tissue Organ Cult. 85(2):241–245. doi:10.1007/s11240-006-9082-6.

Yang M, Xie X, Zheng C, Zhang F, He X, Li Z. 2008. Agrobacterium tumefaciens-mediated genetic transformation of *Acacia crassicarpa* via organogenesis. Plant Cell Tissue Organ Cult. 95(2):141–147. doi:10.1007/s11240-008-9424-7.

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, *et al.* 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature. 480(7378):520–524. doi:10.1038/nature10625.

Zhao Y, Zhang R, Jiang KW, Qi J, Hu Y, Guo J, Zhu R, Zhang T, Egan AN, Yi TS, *et al.* 2021. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. Mol Plant. 14(5):748–773. doi:10.1016/j.molp.2021.02.006.

Editor: P. Ingvarsson