

Supplementary Materials for **Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis**

Maximilian Griesmann, Yue Chang, Xin Liu, Yue Song, Georg Haberer, Matthew B. Crook, Benjamin Billault-Penneteau, Dominique Lauressergues, Jean Keller, Leandro Imanishi, Yuda Purwana Roswanjaya, Wouter Kohlen, Petar Pujic, Kai Battenberg, Nicole Alloisio, Yuhu Liang, Henk Hilhorst, Marco G. Salgado, Valerie Hocher, Hassen Gherbi, Sergio Svistoonoff, Jeff J. Doyle, Shixu He, Yan Xu, Shanyun Xu, Jing Qu, Qiang Gao, Xiaodong Fang, Yuan Fu, Philippe Normand, Alison M. Berry, Luis G. Wall, Jean-Michel Ané, Katharina Pawlowski, Xun Xu, Huanming Yang, Manuel Spannagl, Klaus F. X. Mayer, Gane Ka-Shu Wong, Martin Parniske*, Pierre-Marc Delaux*, Shifeng Cheng*

*Corresponding author. Email: chengshf@genomics.cn (S.C.); pierre-marc.delaux@lrsv.ups-tlse.fr (P.-M.D.); parniske@lmu.de (M.P.)

Published 24 May 2018 on *Science* First Release
DOI: 10.1126/science.aat1743

This PDF file includes:

Materials and Methods
Figs. S1 to S27
Tables S1 to S37
Captions for Data S1 and S2
References

Other Supporting Online Material for this manuscript includes the following: (available at www.sciencemag.org/cgi/content/full/science.aat1743/DC1)

Data S1 (FASTA)
Data S2 (Excel)

Materials and Methods

Seed germination and plant growth conditions

Seeds of *Casuarina glauca* were grown for one month in a soil:vermiculite substrate (4:1, v:v) in a greenhouse under natural light at temperatures between 25 and 30 °C. After one month, seedlings were transferred to a hydroponic growth system for three weeks in a modified Broughton and Dilworth (BD) medium (26) supplemented with nitrogen (KNO_3 5mM) and cultivated in a growth chamber (25 °C, 16 h/8 h day/night, ~45% humidity, $\text{PAR} = 150 \mu\text{mol m}^{-2} \text{s}^{-1}$). Seeds of *Mimosa pudica* were germinated by incubating in 60 °C sterile water for 15 minutes at 60–70 rpm at room temperature. Then they were incubated on a water plate with 1.5% agar and 1 µM gibberellic acid, in the dark at room temperature for 2–3 days. For samples used for whole transcriptome (RNA) and genome (DNA) sequencing, 3-day-old seedlings of *M. pudica* were transferred to 8 × 8 cm pots filled with Sungro® Metro-Mix® 360 potting soil (Sungro). The pots were watered with tap water and placed in a growth chamber (23 °C with constant light). Seeds of *Nissolia schotii* were scarified in H_2SO_4 for 5 minutes and rinsed five times in a large volume of water. After two hours of imbibition in water, seeds were plated on 1% agar plates and germinated in the dark at room temperature. Seedlings were moved to a mix of clay and sand (50:50, v:v) and grown in a growth chamber (23 °C, 16 h/8 h day/night cycles). Plants were watered thrice a week with BNM medium containing 10 mM KNO_3 . Seeds of *Chaetocalyx sp.*, *Chaetocalyx scandens*, *Chaetocalyx brasiliensis*, *Lathyrus sativus*, *Lathyrus oleraceus*, *Lathyrus oleraceus var arvense*, *Ornithopus perpusillus*, *Vicia sativa*, *Vigna mungo*, *Vigna unguilurata ssp sesquipedalis*, and *Adesmia muricata* were scarified for five minutes in H_2SO_4 and rinsed five times with a large volume of water. After 3 hours of imbibition, seeds were transferred to pots with regular potting soil. Plants were grown for twelve days in a growth

chamber (16h/8h day/night cycles, 23°C constant temperature). *Sophora flavescens*, *Chamaecrista mimosoides* and *Chamaecrista stricta* were treated with H₂SO₄ for ten minutes and subsequently washed six times with water. Seeds were kept at 4°C for 24h and pre-germinated for 48h at room temperature. *Castanospermum australe* seeds were stripped of their seed coats and germinated on moist filter paper in a closed pot system under controlled conditions (26°C/22°C, 16h/8h day/night cycles and 80% relative humidity) for 5-10 days. *Sophora flavescens*, *Chamaecrista mimosoides*, *Chamaecrista stricta* and *Castanospermum australe* were transferred in potting soil and grown for an additional 4 weeks in a conditioned greenhouse (22°C/18°C, 16h/8h day/night cycles and 65% relative humidity for *Sophora flavescens* and 26°C/22°C, 16h/8h day/night cycles and 80% relative humidity for *Chamaecrista mimosoides*, *Chamaecrista stricta* and *Castanospermum australe*). *Begonia fuchsioides* cuttings were grown at the LMU Munich in a growth cabinet at low light (2 fluorescent lights removed from the cabinet) at 16 h light (18.5 °C) / 8 h dark (12 °C). They were grown in poor soil 1:1 mixture of sand (grain size 1-1.2 mm): Stender Vermehrungssubstrat A 210; Stender AG, Schermbeck, Germany. Plants were regularly watered with deionized water and once a week with ¼ Hoagland's (97). *Datisca glomerata* was cultivated at Stockholm University (Sweden) and at the LMU Munich (Germany). For surface sterilization, *D. glomerata* seeds were incubated for 30 sec in 25% H₂SO₄, followed by two washes with sterile deionized water. Then they were incubated for 5 min in 2.5 % NaOCl and washed six times with sterile deionized water. Seeds were transferred to pots with sand, or to Petri dishes with ¼ Hoagland's with 10 mM KNO₃ and 1% agar. For leaf production, *D. glomerata* was grown in the greenhouse at LMU Munich (18 °C/12 °C, 16 h/8 h day/night cycles, light intensity of 150 µmol.m⁻².s⁻¹). Growth substrate was A210 (perlite, with 0.5 kg/m³ of 14-10-18 NPK, with Sphagnum peat H3-H5, pH 6.2) from

Stender AG (Schermbeck, Germany). Seeds of *D. drummondii* were surface sterilized with 30% of H₂O₂ for 10 minutes and 15 minutes, respectively. Seeds were then rinsed with at least three washes of deionized water and incubated in the dark at 22 °C on water plate with 0.8% agar. The 12-days-old seedlings of *D. drummondii* were transferred on 1-4 mm diameter particles of expensed clay in sterile jars or pots (7 × 7 × 8 cm). The jars were placed in a growth chamber (20 °C; 16 h/9 h day/night cycles; light intensity around 125 µmol.m⁻².s⁻¹) and the pots were placed in a greenhouse (night temperature of 12°C ±1; day temperature of 18°C ±5; mercury-vapor lamps were used to ensure a minimum 14 h of light per day). 35mL of ¼ Hoagland solution¹⁷ (with 1mM of KNO₃) were added in the jars; the plants in the greenhouse were watered once a week with the same solution and twice a week with tap water. Seeds of *Discaria trinervis* were surface sterilized and germinated as previously described by Valverde *et al.* (98) The sterilized seeds were placed on petri dishes with 1% agar and then incubated in a culture chamber in the dark at 25 ° C for 2 weeks. *D. trinervis* plants were grown in a greenhouse at Quilmes National University, Bernal, Argentina, in a hydroponic system with BD nutrient solution⁵ with KNO₃ (16 h/8 h day/night cycle, temperature between 19 and 25 °C; relativity humidity of 50–90%). Seeds of *Alnus glutinosa* were sown, left to germinate and grown for six weeks in a soil/vermiculite substrate (1:1, vol/vol) in a greenhouse under natrium lighting (10 to 12 klx) with a 16 h/8 h day/night regime at a temperature regime of 21 °C (day) and 16 °C (night). Seedlings were transferred to 500 ml Fåhraeus' mineral solution (99) with 5 mM KNO₃ in opaque plastic pots (8 seedlings/pot) and grown for four weeks.

DNA extraction

For *A. glutinosa* DNA extraction, 2 g of young leaves were ground to a fine powder using liquid nitrogen in a porcelain mortar in the presence of 2 g of Polyclar AT (Serva, Paris France). The

resulting powder was then transferred to 50 ml centrifuge tubes, containing 15 ml of EB buffer (100 mM Tris/HCl pH 8.0, 40 mM EDTA, 500 mM NaCl, and 15 µl of β-mercaptoethanol (Sigma)) and mixed by vortexing. One ml of 20% w/vol SDS was added and mixed gently by pipetting several times and incubated for 10 min at 65°C. 5 ml of 5M K-acetate (pH not adjusted) was added, mixed several times by inverting the tube and left on ice for 15 min. Cells debris were removed by centrifugation at 4°C, 2700 × g for 12 min. The supernatant was filtered through 100 µm sterile filter (Easystrainer, Dutscher, Brumath France) into a new 50 ml tubes containing 1 g of Polyclar AT, mixed carefully and incubated at room temperature for 15 min. Polyclar AT and bound polysaccharides and polyphenols were removed by centrifugation, 2700 × g at RT, for 12 min. Supernatants were filtered through 100 µm sterile filters into new 50 ml tubes. DNA was precipitated by adding 0.66 volumes of isopropanol, mixed and collected by centrifugation at 4450 × g at RT for 20 min. The supernatant was removed, and the DNA pellet was washed with 1 ml of 75% ethanol and transferred together into a 2 ml Eppendorf tube. Sample was centrifuged in a micro centrifuge Eppendorf at RT for 5 min at 10000 × g. Supernatant was removed quantitatively and pellet washed in 1 ml of 3M NaCl for 15 min, supernatant was removed after centrifugation. Pellet was washed in 1 ml of 75% ethanol and supernatant removed quantitatively after centrifugation. DNA pellet was air dried and solubilized in 600 µl of sterile deionised water and extracted with 600 µl of phenol: chlorophorm:isoamyl alcohol 25: 24: 1 (v/v/v) pH 7.8. Aqueous phase was transferred into a new 2 ml Eppendorf tube. DNA was precipitated by adding one tenth of 3M sodium acetate pH 5.2 and two volumes of ethanol. DNA was collected by centrifugation in Eppendorf centrifuge at 13 000 × g, RT for 10 min, supernatant was discarded and pellet washed in 500 µL of 75 % ethanol, centrifuged for 1 minute at 13 000 x g and air dried after ethanol removal. The

DNA sample was solubilized in 600 µl of TE buffer pH8 and the RNA digested by adding 50 µg of RNase A (Macherey Nagel) during one hour at 37 °C. DNA sample was purified by phenol chlorophorm isoamyl alcohol extraction and precipitated as before and then solubilised in 200 µl of TE buffer pH 8 at + 4 °C for 24 hours before quantification. Modified from Dellaporta *et al.* (100). For *Casuarina glauca* high quality genomic DNA was extracted from young leaves sampled on one month old seedlings by the company ADNiD (www.adnid.fr) using a confidential CTAB based method. DNA quality was assessed through gel electrophoresis, endonuclease degradation test and *HinfI* digestion test. Quantification was done using PicoGreen DNAassay (Invitrogen, USA) following manufacturer's instructions. For *Cercis canadensis* DNA isolation was conducted on leaves collected from a tree growing on the Iowa State University campus. All samples were ground in liquid nitrogen and subjected to a CTAB extraction, partially derived from Healey *et al.* (101). Briefly, 1 g of ground material was added to 10 ml of CTAB buffer (2% CTAB; 100 mM Tris, pH 8.0; 25 mM EDTA, pH 8.0; 1.5 M NaCl; 1% PVP 40; in RNase-free water, adjust with HCl to a final pH of 5.0) and 30 µl β-mercaptoethanol, vortexed for 30 seconds, and incubated at 65 °C for 1 hour. Proteins were removed by extraction with 24:1 chloroform:isoamyl alcohol followed by centrifugation. RNA was removed by treatment with RNase A at 37 °C for 15 minutes followed by another 24:1 chloroform:isoamyl alcohol extraction and centrifugation. DNA was precipitated with chilled (4 °C) 70% ethanol, incubation at –20 °C for 30 – 60 minutes. After centrifugation, the DNA pellet was washed three times with 70% ethanol, and resuspended in TE. Leaves from 2-month-old *Chamaecrista fasciculata* plants grown in a growth chamber were collected, ground in liquid nitrogen and subjected to a CTAB extraction, partially derived from Healey *et al.* (101). Briefly, 1 g of ground material was added to 10 ml of CTAB buffer (2% CTAB; 100 mM Tris, pH 8.0; 25 mM EDTA,

pH 8.0; 1.5 M NaCl; 1% PVP 40; in double distilled water) and 30 µl β-mercaptoethanol, vortexed for 30 seconds, and incubated at 65 °C for 1 hour. Proteins were extracted with 24:1 chloroform:isoamyl alcohol followed by centrifugation. RNA was removed by treatment with RNase A at 37 °C for 15 minutes followed by another 24:1 chloroform:isoamyl alcohol extraction and centrifugation. DNA was precipitated by adding chilled (-20°C) ethanol to a final concentration of 70% and incubation at -20 °C for 60 minutes. After centrifugation, the pellet was washed three times with 70% ethanol, and resuspended in TE. For *Mimosa pudica*, DNA isolation was conducted on leaves from 2-month-old plants grown in a growth chamber. All samples were ground in liquid nitrogen and subjected to a CTAB extraction, partially derived from Healey *et al.* Briefly, 1 g of ground material was added to 10 ml of CTAB buffer (2% CTAB; 100 mM Tris, pH 8.0; 25 mM EDTA, pH 8.0; 1.5 M NaCl; 1% PVP 40; in RNase-free water, adjust with HCl to a final pH of 5.0) and 100 µl β-mercaptoethanol, vortexed for 20 seconds, and incubated at 65 °C for 1 hour. Proteins were removed extracted with 24:1 chloroform:isoamyl alcohol followed by centrifugation. RNA was removed by treatment with RNase A at 37 °C for 15 minutes followed by another 24:1 chloroform:isoamyl alcohol extraction and centrifugation. DNA was precipitated with chilled (4 °C) ethanol added to a final concentration of 70% and incubation at -20°C for 30–60 minutes. After centrifugation, the pellet was washed three times with 70% ethanol, and resuspended in TE. For *D. glomerata* DNA isolation, leaves from greenhouse-grown plants were used. For *B. fuchsoides* DNA isolation, leaves from the shrub in the Nymphenburg Botanical garden were used. *Dryas drummondii* used for DNA extractions came from seeds bought in a specialized seed production company: Jelitto®. All seeds came from the same seed bags. DNA was isolated using the lysis method developed by Dellaporta *et al.* (100) with some modifications, followed by a CsCl gradient

purification according to Ribeiro *et al.* (102). DNA of *D. trinervis* was extracted from leaves of 2 to 4-month-old plants grown in a greenhouse. DNA extractions were performed by Productos Bio-Lógicos S.A. (www.pb-l.com.ar, Quilmes, Buenos Aires, Argentina) based on Healey *et al.* (101). Briefly, 1 g of liquid nitrogen ground material was added to 10 ml of CTAB buffer (2% CTAB; 100 mM Tris, pH 8.0; 25 mM EDTA, pH 8.0; 1.5 M NaCl; 1% PVP 40; in RNase-free water, final pH 5.0) and 100 µl β-mercaptoethanol, vortexed for 20 seconds, and incubated at 65 °C for 1 hour. Proteins were removed extracted with 24:1 chloroform:isoamyl alcohol followed by centrifugation. RNA was removed by treatment with RNase A at 37 °C for 15 minutes followed by another 24:1 chloroform:isoamyl alcohol extraction and centrifugation. DNA was precipitated with chilled (4°C) 70% ethanol, incubated at –20°C for 30–60 minutes, washed three times with 70% ethanol, and resuspended in TE. For the experiment presented in Figure S5, DNA was extracted using the Qiagen DNA plant mini kit from young leaves of 12-days old plants for *Chaetocalyx sp.*, *Chaetocalyx scandens*, *Chaetocalyx brasiliensis*, *Lathyrus sativus*, *Lathyrus oleraceus*, *Lathyrus oleraceus var arvense*, *Ornithopus perpusillus*, *Vicia sativa*, *Vigna mungo*, *Vigna unguiculata ssp sesquipedalis*, *Mimosa pudica*, *Lupinus angustifolius* and *Adesmia muricata*. For *Sophora flavescens*, *Chamaecrista mimosoides*, *Chamaecrista stricta* and *Castanospermum australe* DNA was extracted from young shoot tissue (~0.5 cm shoot tips) using a CTAB based method.

RNA extraction

Leaves and roots of *Casuarina glauca* were sampled. Total RNA were purified by ultracentrifugation (103). Residual DNA was removed from RNA samples using the Turbo DNA free kit (Ambion, USA), quantified using a NanoDrop spectrophotometer (ThermoFisher Scientific, Courtaboeuf, France) and qualitatively assessed using a Bioanalyser 2100 (Agilent p/n

G2938A) according to the manufacturer's instructions (Agilent, USA). For each sample, 4 µg of total RNA was pooled in a single sample. For *Cercis canadensis* all samples were ground in liquid nitrogen and subjected to a CTAB extraction, partially derived from Healey *et al.* (101). Briefly, 250 of ground material was added to 1.5 ml of CTAB buffer (2% CTAB; 100 mM Tris, pH 8.0; 25 mM EDTA, pH 8.0; 1.5 M NaCl; 1% PVP 40; in RNase-free water, adjust with HCl to a final pH of 5.0) and 15 µl β-mercaptoethanol, vortexed for 1 minute, and incubated at 65°C for 20 minutes. Proteins were removed by two 24:1 chloroform:isoamyl alcohol extractions followed by centrifugation. Nucleic acids were precipitated from the aqueous phase with 100% isopropanol and then resuspended in RNase-free water. RNA extractions were then performed with a PureLink® RNA Mini Kit (<https://www.thermofisher.com/us/en/home/life-science/dna-rna-purification-analysis/rna-extraction/rna-types/total-rna-extraction/purelink-rna-mini-kit.html>). Contaminating DNA was removed using an Ambion TURBO DNA-free™ Kit (ThremoFisher). For *Chamaecrista fasciculata* all samples were ground in liquid nitrogen and subjected to a CTAB extraction, partially derived from Healey *et al.* Briefly, 250 of ground material was added to 1.5 ml of CTAB buffer (2% CTAB; 100 mM Tris, pH 8.0; 25 mM EDTA, pH 8.0; 1.5 M NaCl; 1% Polyvinylpyrrolidone (PVP) 40; in RNase-free water, final pH 5.0) and 15 µl β-mercaptoethanol, vortexed for 45 seconds, and incubated on ice for 2 minutes. Nucleic acids were precipitated from the aqueous phase with 70% ethanol and then resuspended in RNase-free water. RNA extractions were then performed with a PureLink® RNA Mini Kit (Thermofisher). Contaminating DNA was removed using an Ambion TURBO DNA-free™ Kit (Thermofisher). For *Mimosa pudica* RNA extractions were performed on 2-month-old plants grown in a growth chamber. All samples were ground in liquid nitrogen and subjected to a CTAB extraction, partially derived from Healey *et al.* (101) Briefly, 250 of ground material was

added to 1.5 ml of CTAB buffer (2% CTAB; 100 mM Tris, pH 8.0; 25 mM EDTA, pH 8.0; 1.5 M NaCl; 1% PVP 40; in RNase-free water, final pH 5.0) and 15 µl β-mercaptoethanol, vortexed for 1 minute, and incubated at 65 °C for 20 minutes. Proteins were removed by two 24:1 chloroform:isoamyl alcohol extractions followed by centrifugation. Nucleic acids were precipitated with 100% isopropanol and then resuspended in RNase-free water. RNA extractions were then performed with a PureLink® RNA Mini Kit ([Thermofisher](#)). Contaminating DNA was removed using an Ambion TURBO DNA-*free*™ Kit ([Thermofisher](#)). Leaf and root samples of *Nissolia schotii* were collected, soil particles quickly removed and the samples were frozen in liquid nitrogen. RNA was extracted using the Qiagen RNA plant mini kit following manufacturer's instructions. Roots, nodulated roots and leaves of *Alnus glutinosa* were sampled. Total RNA were purified with the RNeasy plant mini kit and the RNase-free Dnase set (Qiagen, Courtaboeuf, France). Residual DNA was removed from RNA samples using the RQ1 RNase-free DNase (Promega, Charbonnières-les-Bains, France) and the Turbo DNA free kit (Ambion, USA). After DNase treatments, RNA samples were purified using the RNeasy mini cleanup protocol (Qiagen). For *Begonia fuchsioides* material from cuttings was harvested after *ca.* three months of growth, and shock-frozen in liquid nitrogen. Roots and leaves were frozen separately. The entire root systems were harvested except for the top part of *ca.* 1 cm which showed secondary growth and lignification. All RNA extractions were performed using the Sigma Spectrum™ Plant Total RNA extraction kit; the only modification of the kit protocol was that Polyclar AT (Serva, Heidelberg, Germany) was added to the extraction buffer (2% w/v). For *Dryas drummondii*, RNA extractions were done on 12-day-old seedlings and leaves harvested in the botanical garden of Munich. All RNA extractions were performed with the Sigma Spectrum™ Plant Total RNA extraction kit). For extraction of RNA from leaves PolyclarAT

(Serva, Heidelberg, Germany) was added to the extraction buffer (2% w/v) prior to extraction. After 7 weeks of growth, *D. glomerata* root systems were harvested and shock-frozen in liquid nitrogen. Ca. 100 mg of plant material were used per root RNA isolation. Nodules were harvested 7 weeks after infection. Young ca. 4 cm long leaves were harvested from a mature plant which had not yet started to form flowers (*ca.* 3 months after nodulation). RNA of *Discaria trinervis* was extracted from roots and leaves of 2-month-old plants grown in a greenhouse. All samples were ground in liquid nitrogen before extraction with a PURO® Plant RNA Kit (Productos Bio-lógicos).

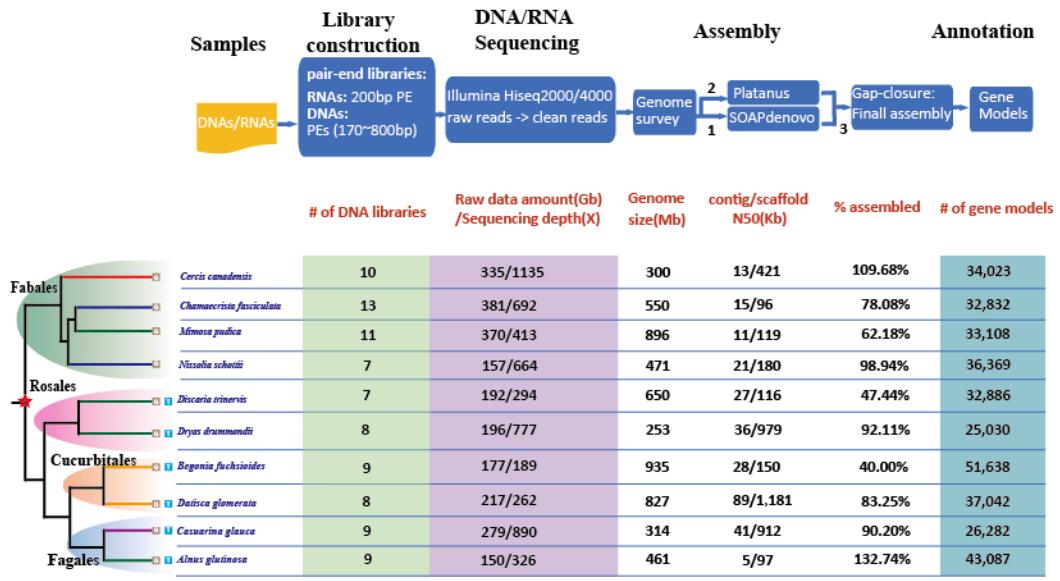


Fig. S1. Overview of the genomics strategy and dataset. The general genome sequencing strategy is presented in the top panel. The bottom panel summarizes the genome assembly characteristics for the ten newly sequenced species.

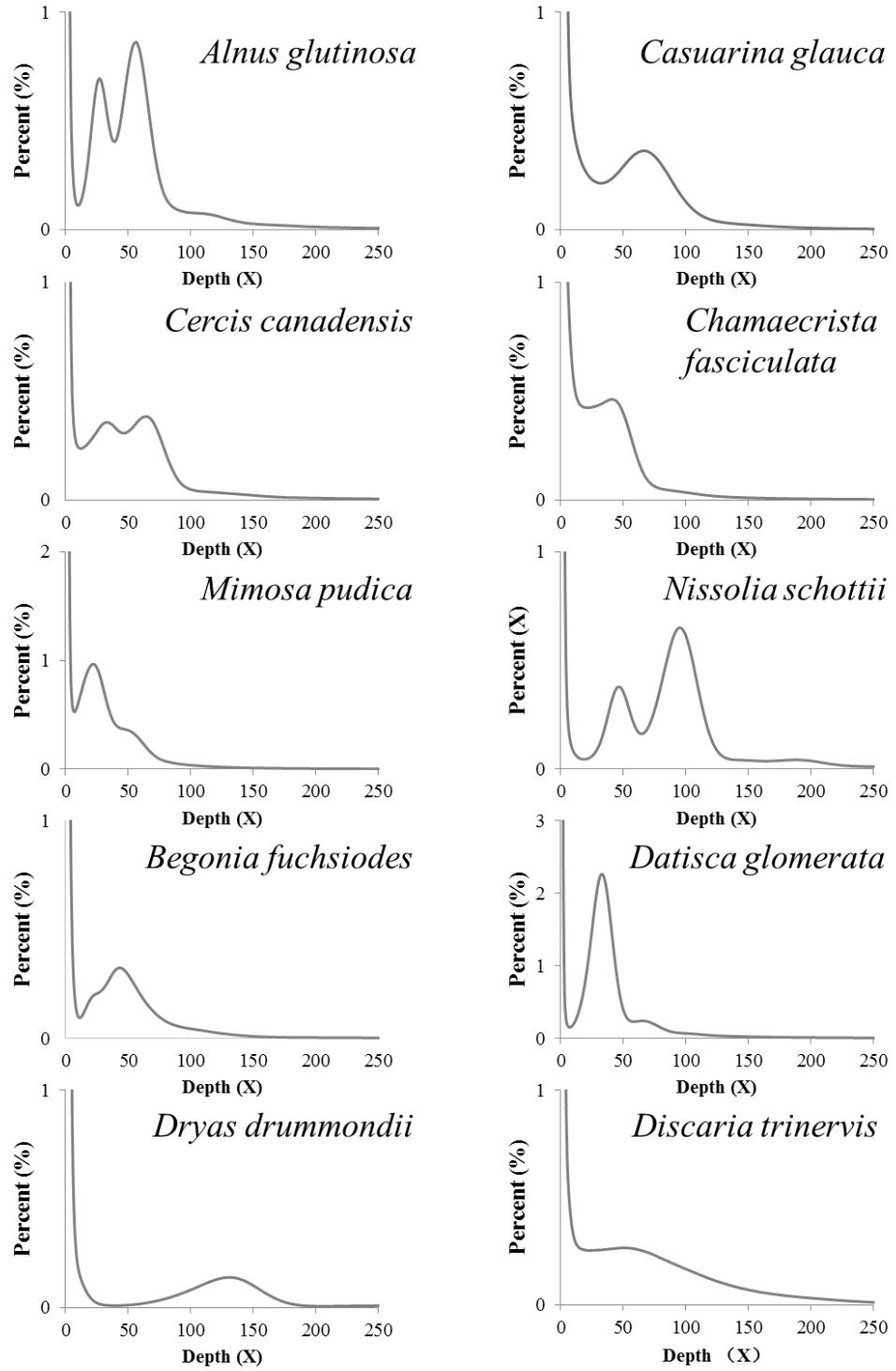


Fig. S2. 17-kmer frequency distribution of the 10 newly sequenced genomes. The main (1C) peak indicates the target diploid genome. A 2C peak (2 fold 1C peak on x-axis) or 0.5C peak indicate high heterozygosity rate.

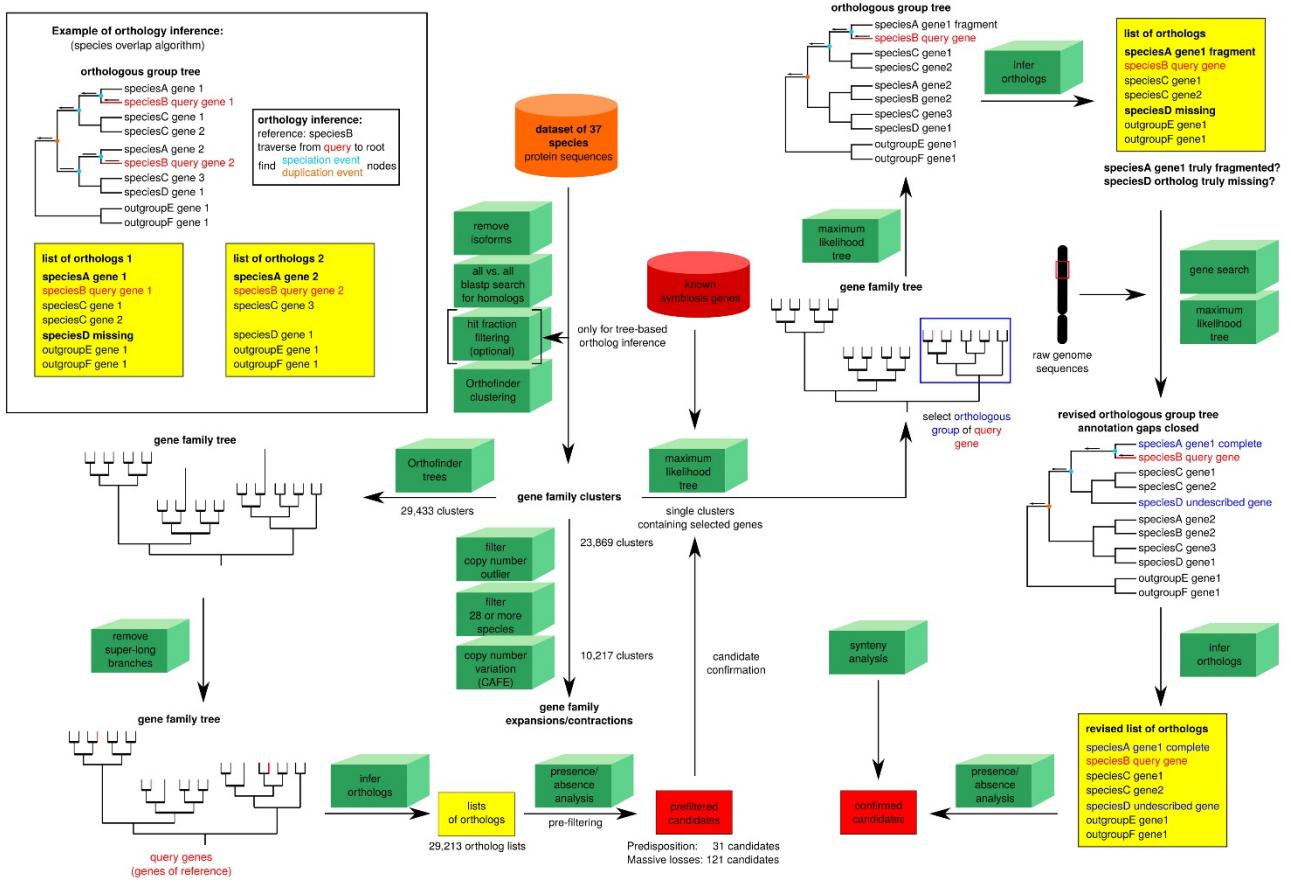
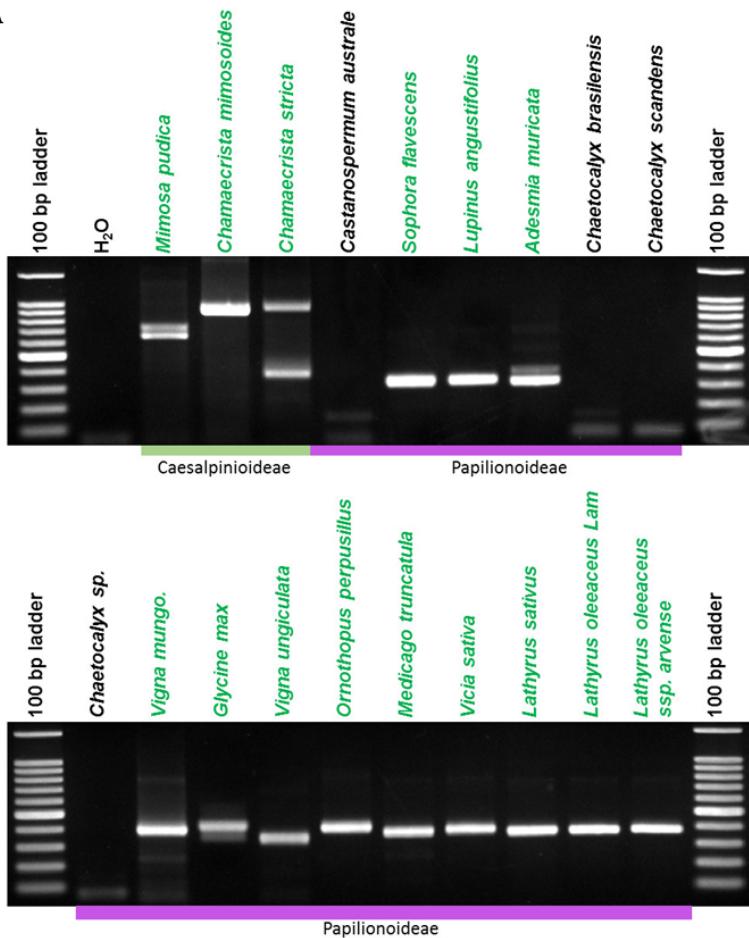


Fig. S3. Overview of pipeline for the genome-wide comparative phylogenomic analysis.



Fig. S4. Maximum-likelihood phylogenetic tree of NIN.

A



B

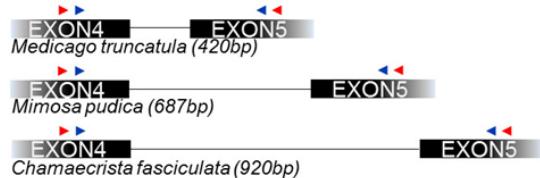


Fig. S5. Validation of the absence of NIN in non-nodulating legumes by PCR. A) Using a degenerated-PCR followed by a nested PCR, NIN from 15 nodulating legumes were amplified. *Mimosa pudica*, *Chamaecrista mimosoides* and *Chamaecrista stricta* belong to be Caesalpinoideae subfamily and harbor unusually large introns as found in the genome assemblies of *Mimosa pudica* and *Chamaecrista fasciculata*. *Adesmia muricata* belongs to the Dalbergieae tribe and is one of the closest nodulating relative to the non-nodulating genera *Nissolia* and *Chaetocalyx*. *Castanospermum australe* is one of the earliest diverging Papilionoideae and is non-nodulating. No band corresponding at least to the minimal 290bp of a putative intron-less fragment was amplified for the four tested non-nodulating species. B) Schematic representation of the intron size in three representative species with available genomes. Red arrowheads: position of the primers used for the first PCR, Blue arrowheads: position of the primers used for the second PCR.

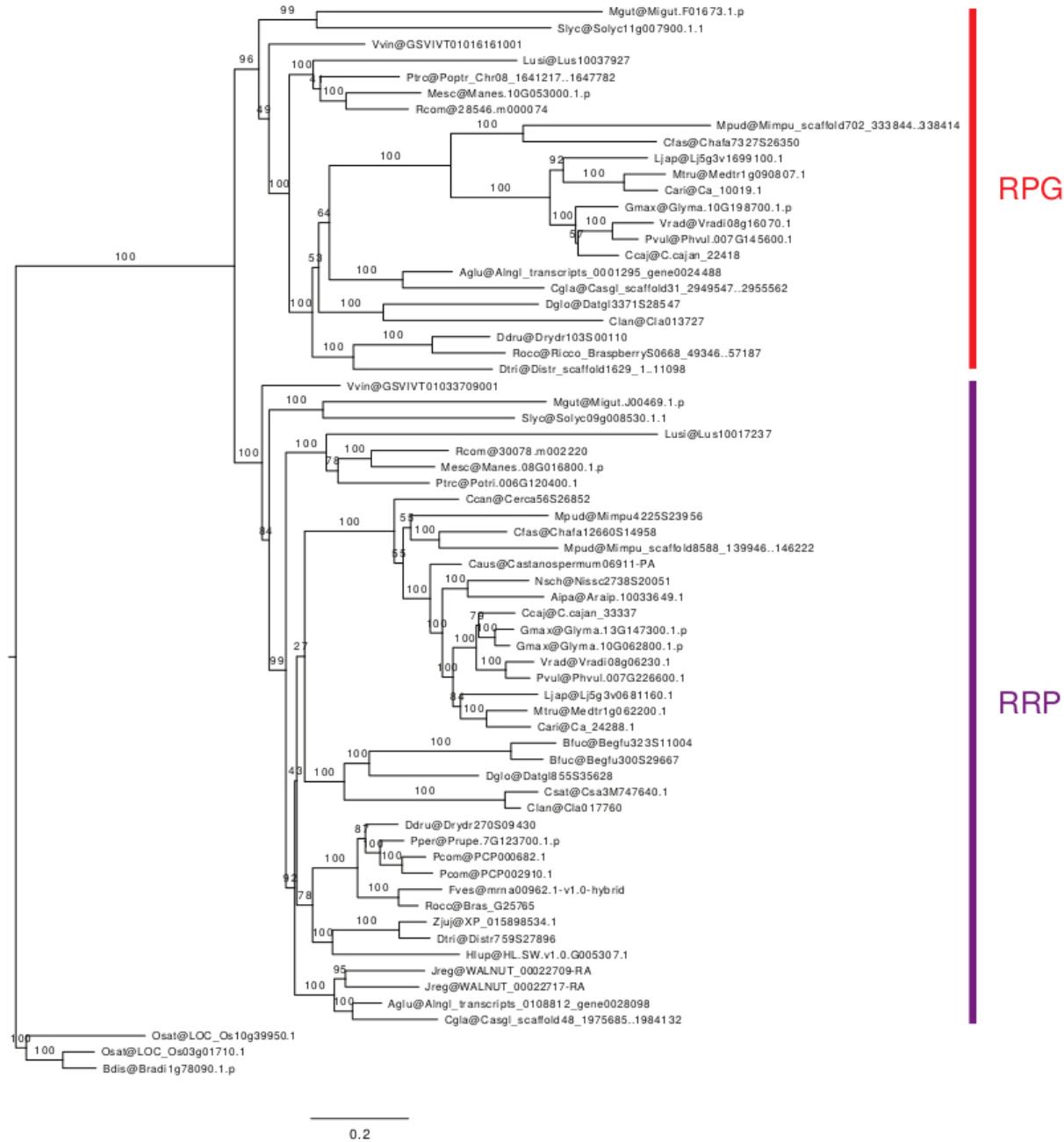


Fig. S6. Maximum-likelihood phylogenetic tree of RPG.

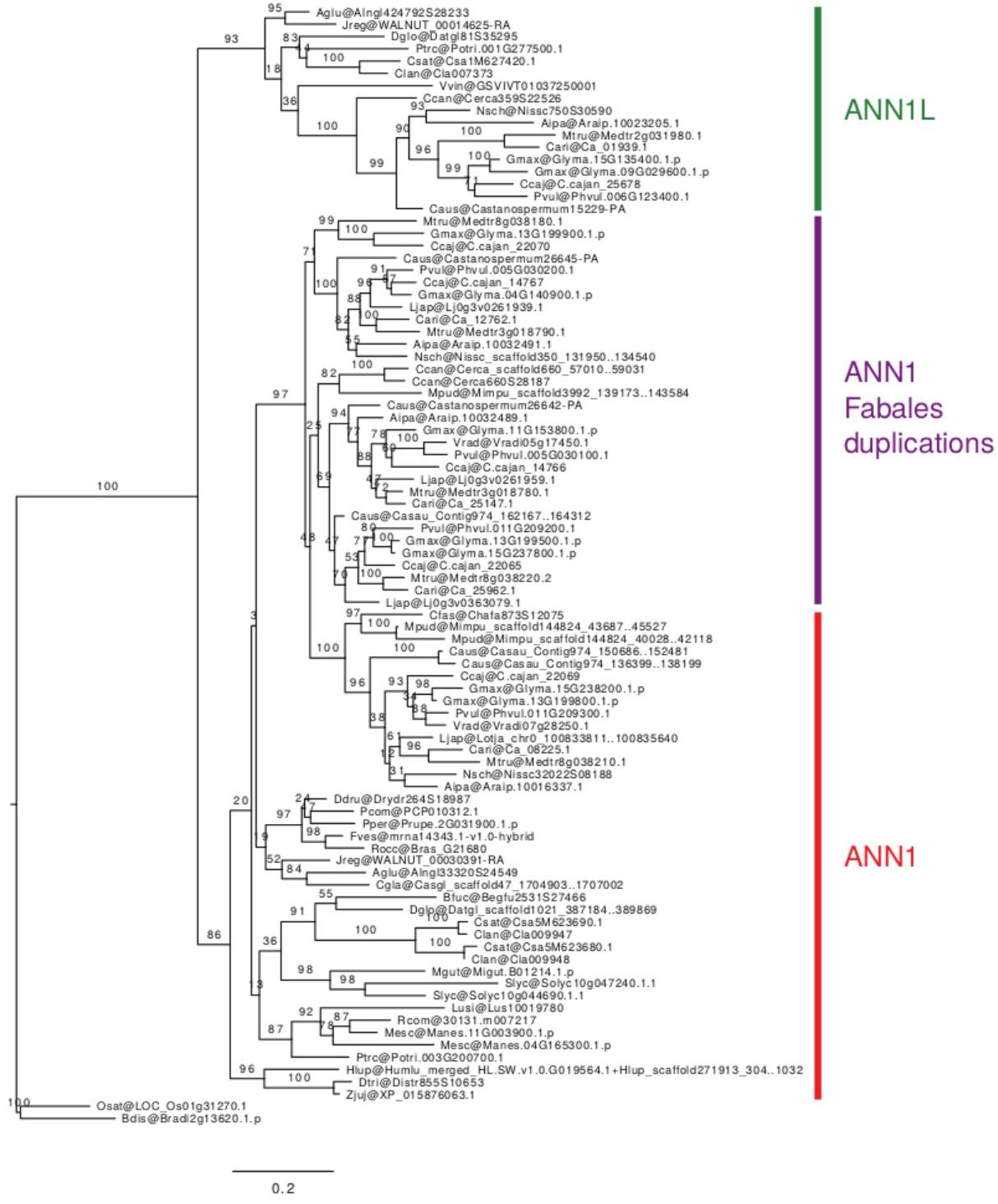


Fig. S7. Maximum-likelihood phylogenetic tree of ANN1.

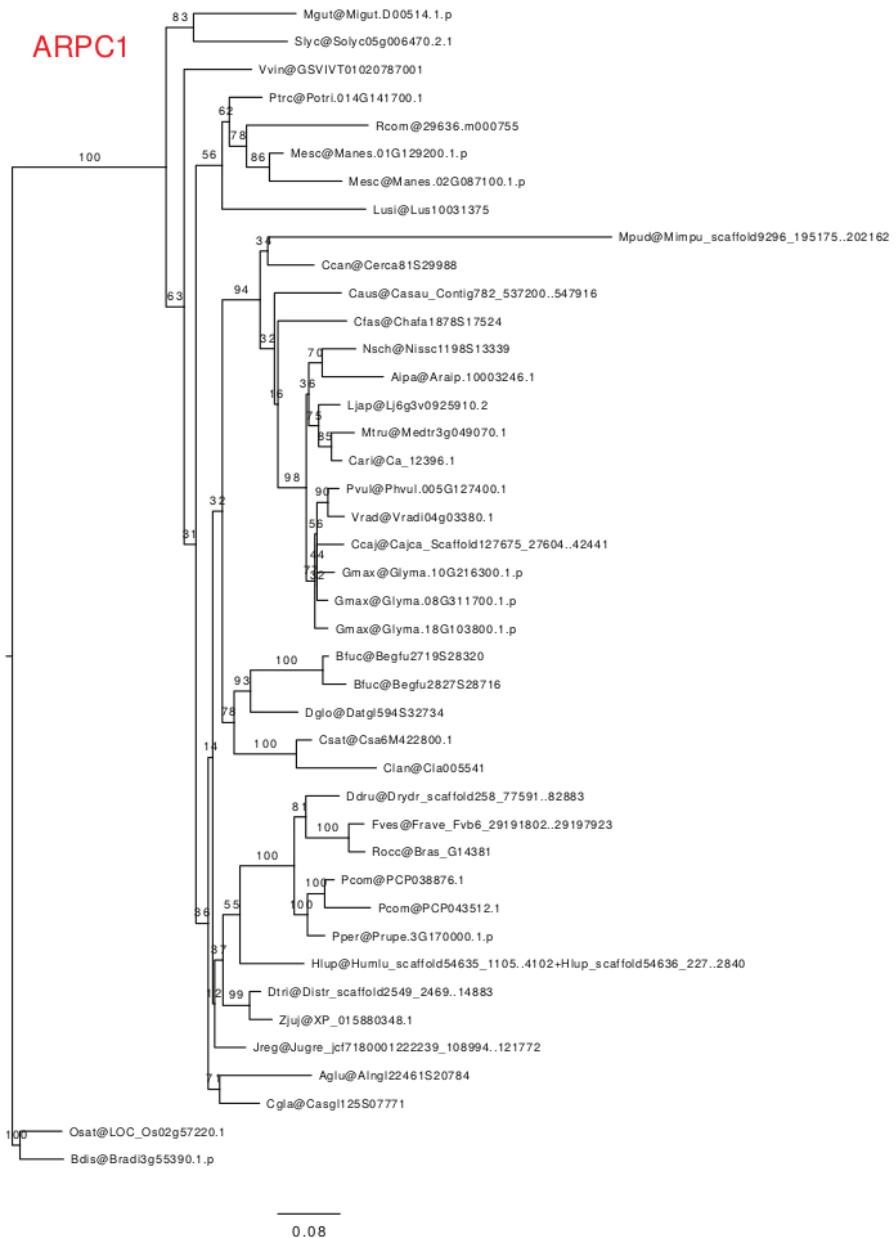


Fig. S8. Maximum-likelihood phylogenetic tree of ARPC1.

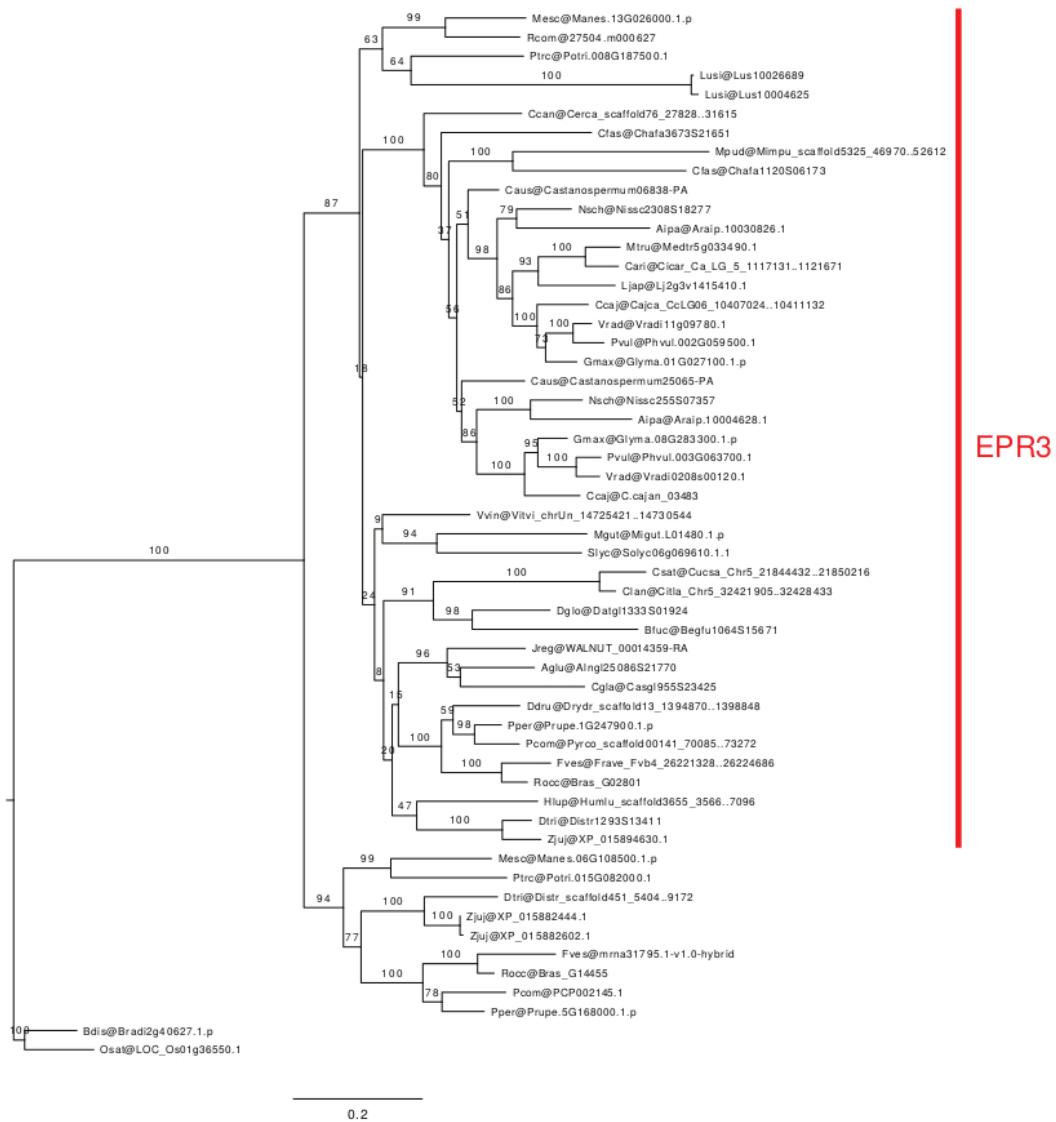


Fig. S9. Maximum-likelihood phylogenetic tree of EPR3.

PIR1

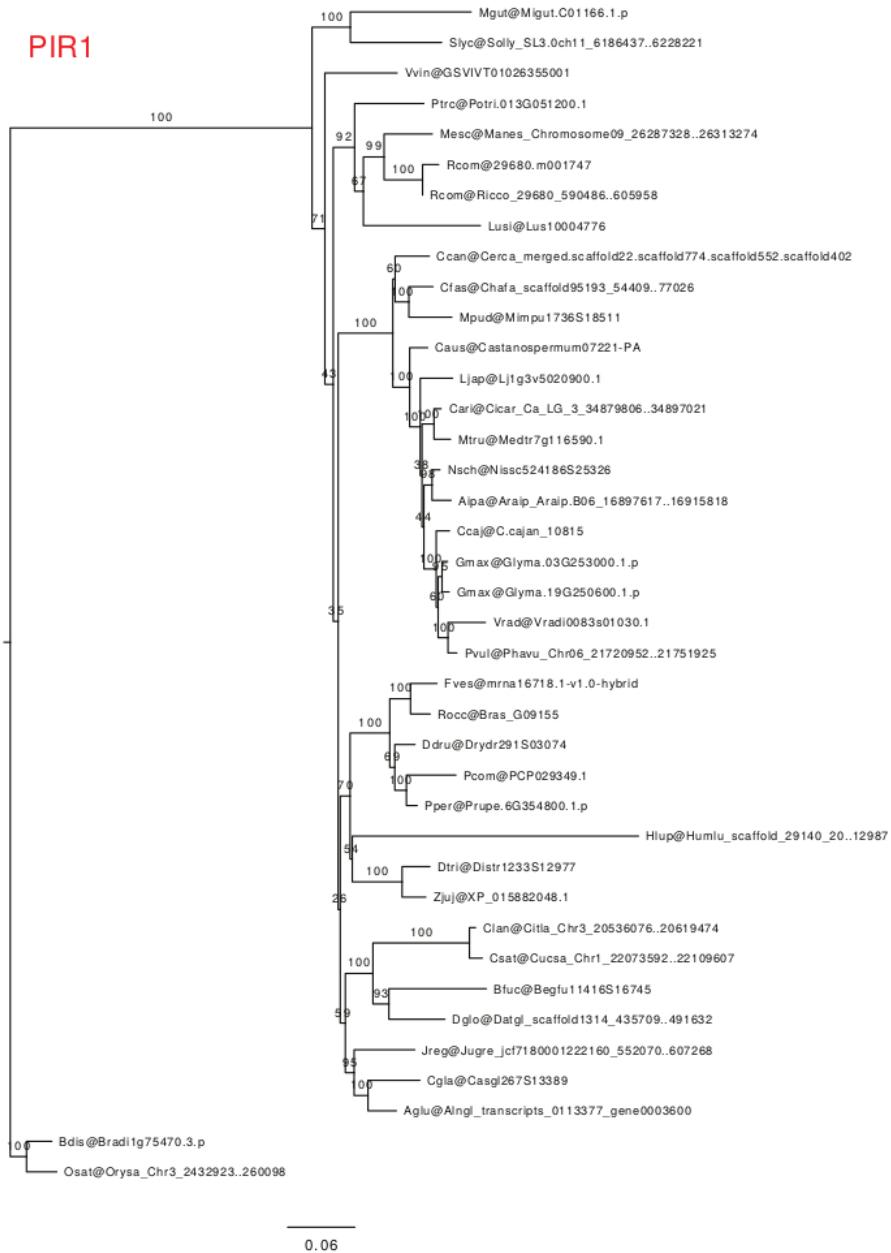


Fig. S10. Maximum-likelihood phylogenetic tree of PIR1.

PUB1

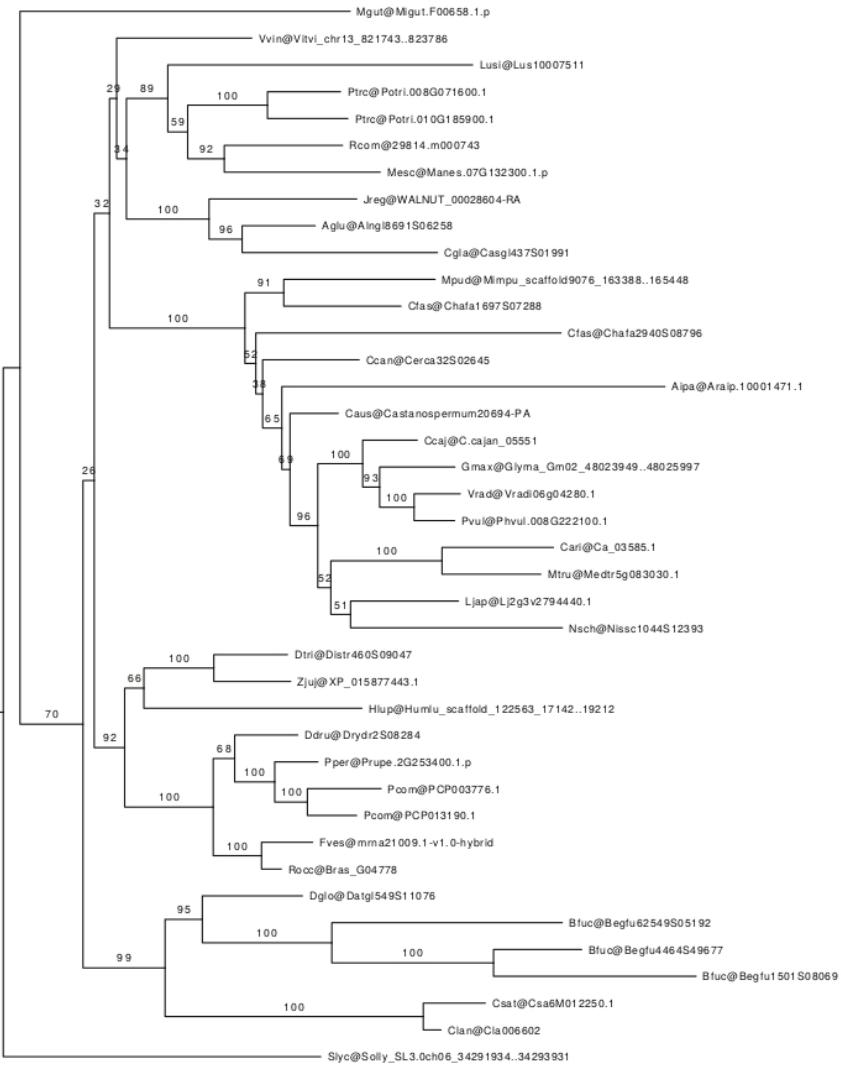


Fig. S11. Maximum-likelihood phylogenetic tree of PUB1.

NAP1

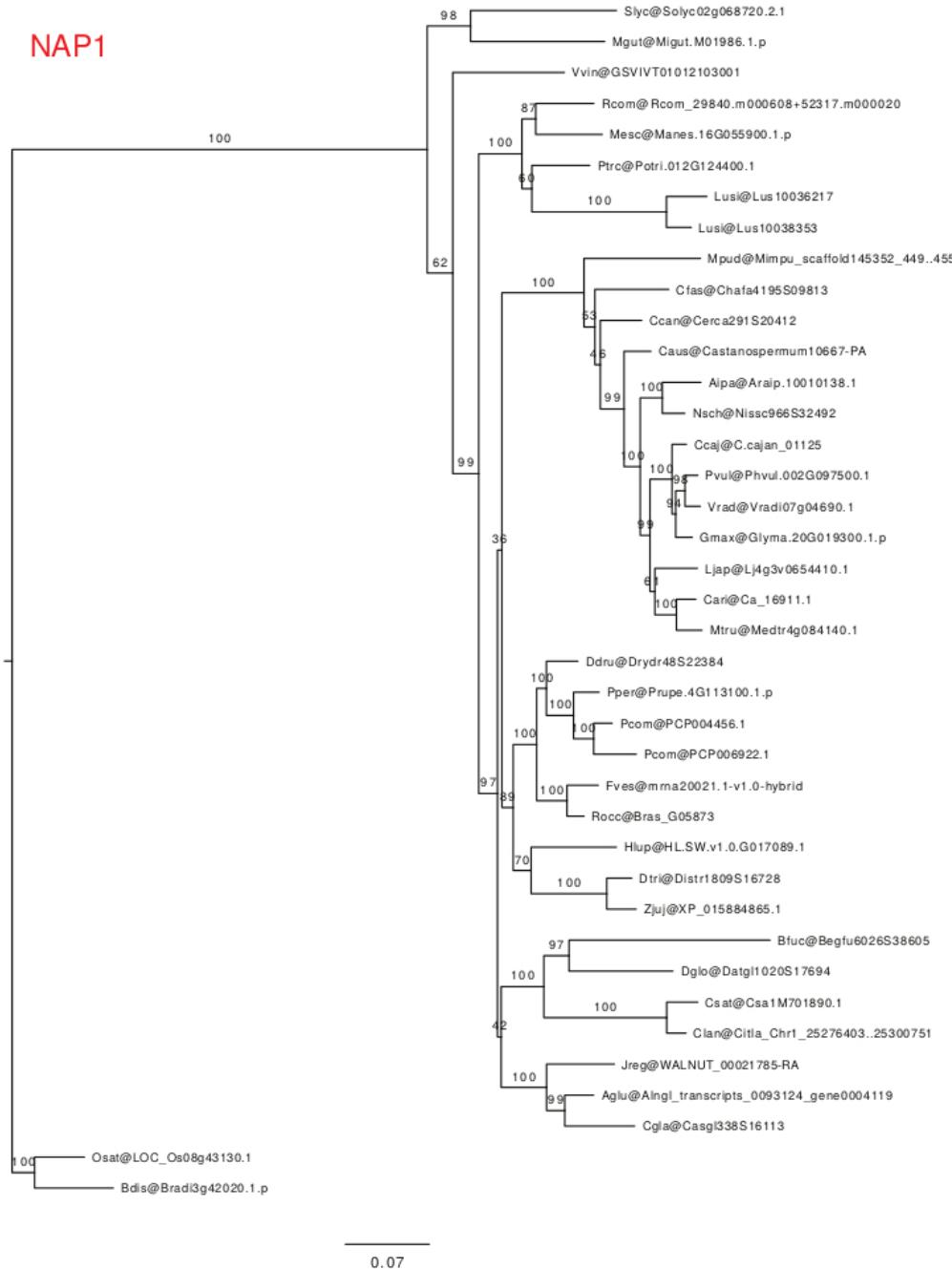


Fig. S12. Maximum-likelihood phylogenetic tree of NAP1.

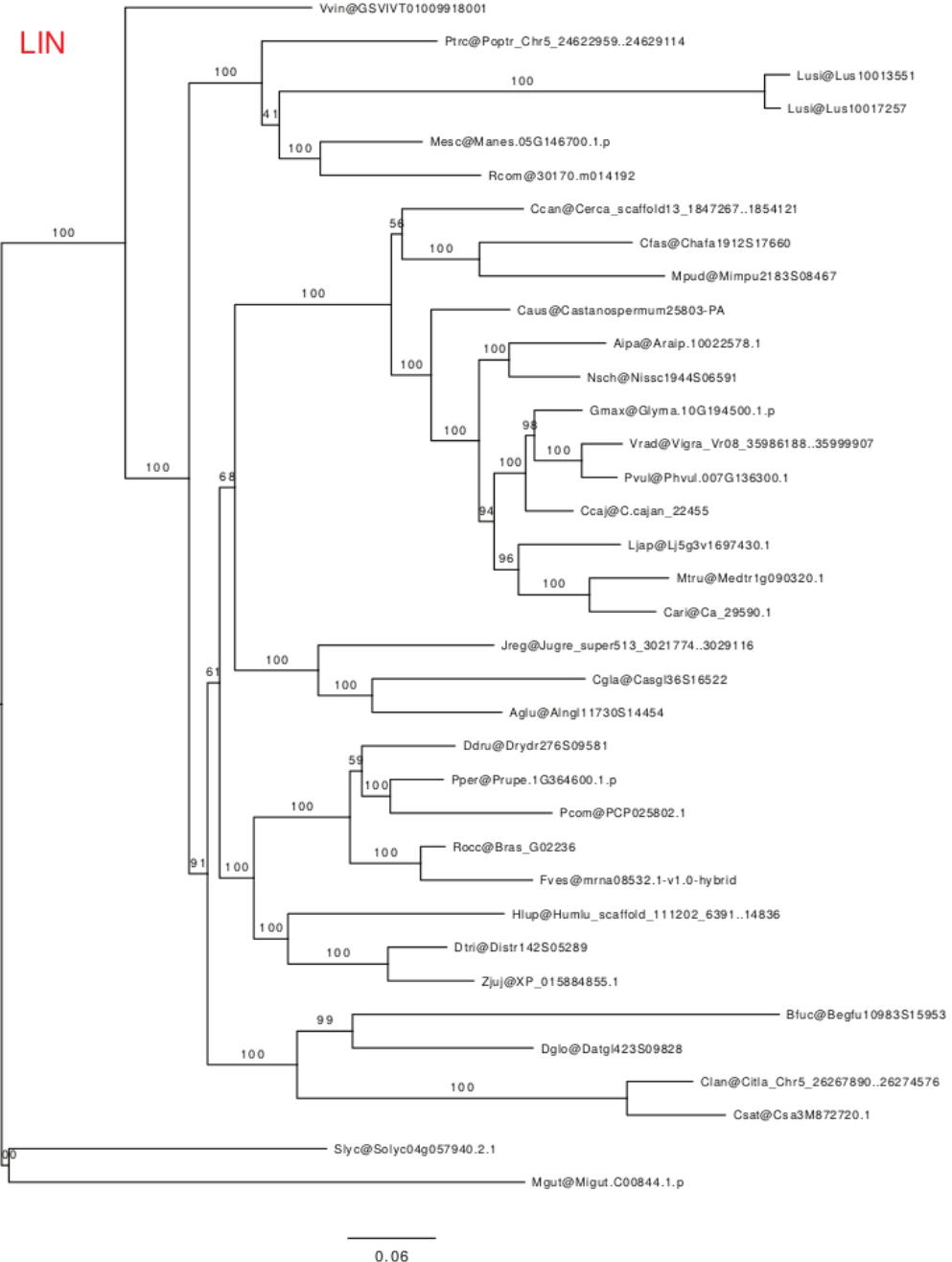


Fig. S13. Maximum-likelihood phylogenetic tree of LIN.

ASTRAY

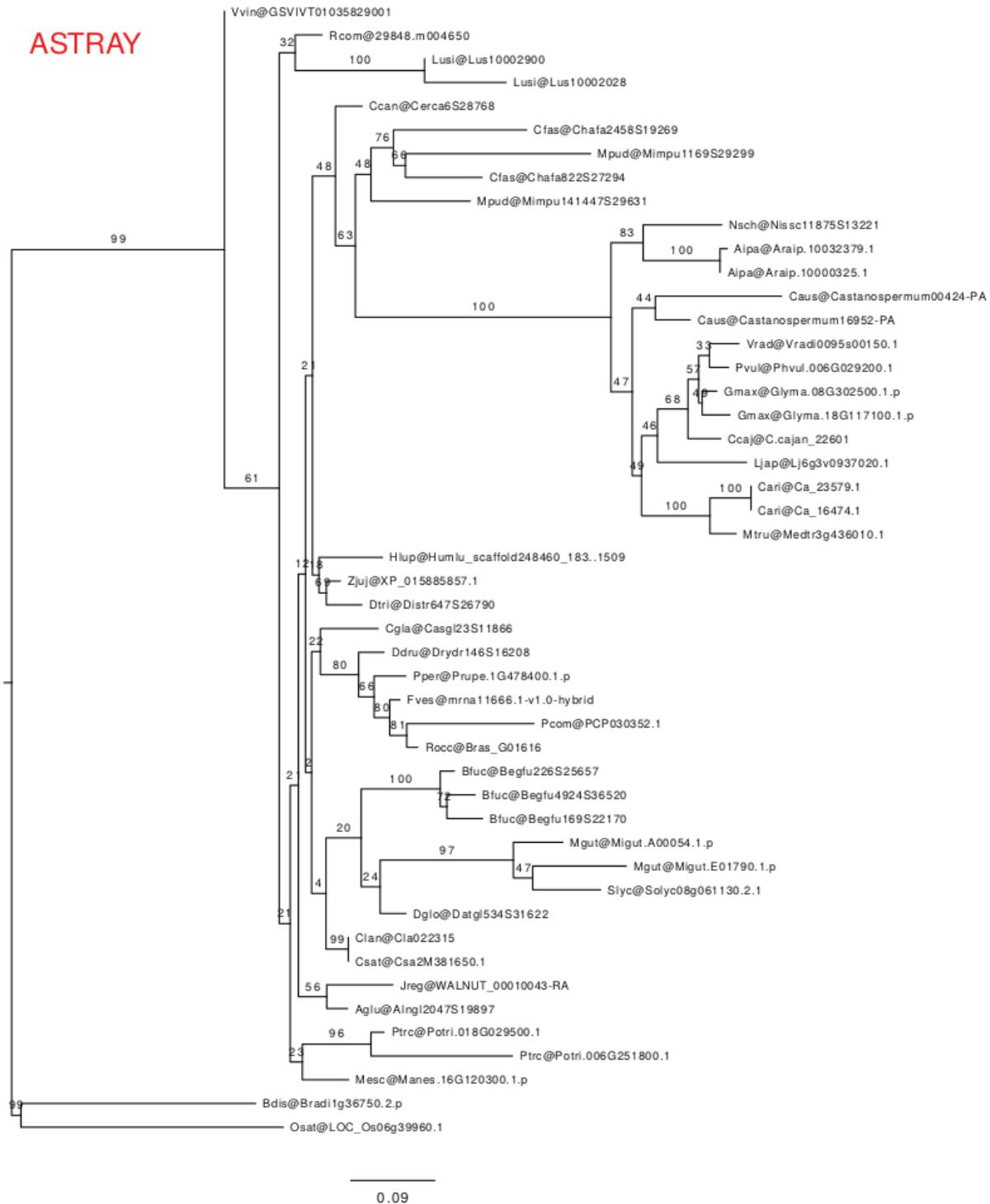


Fig. S14. Maximum-likelihood phylogenetic tree of ASTRAY.

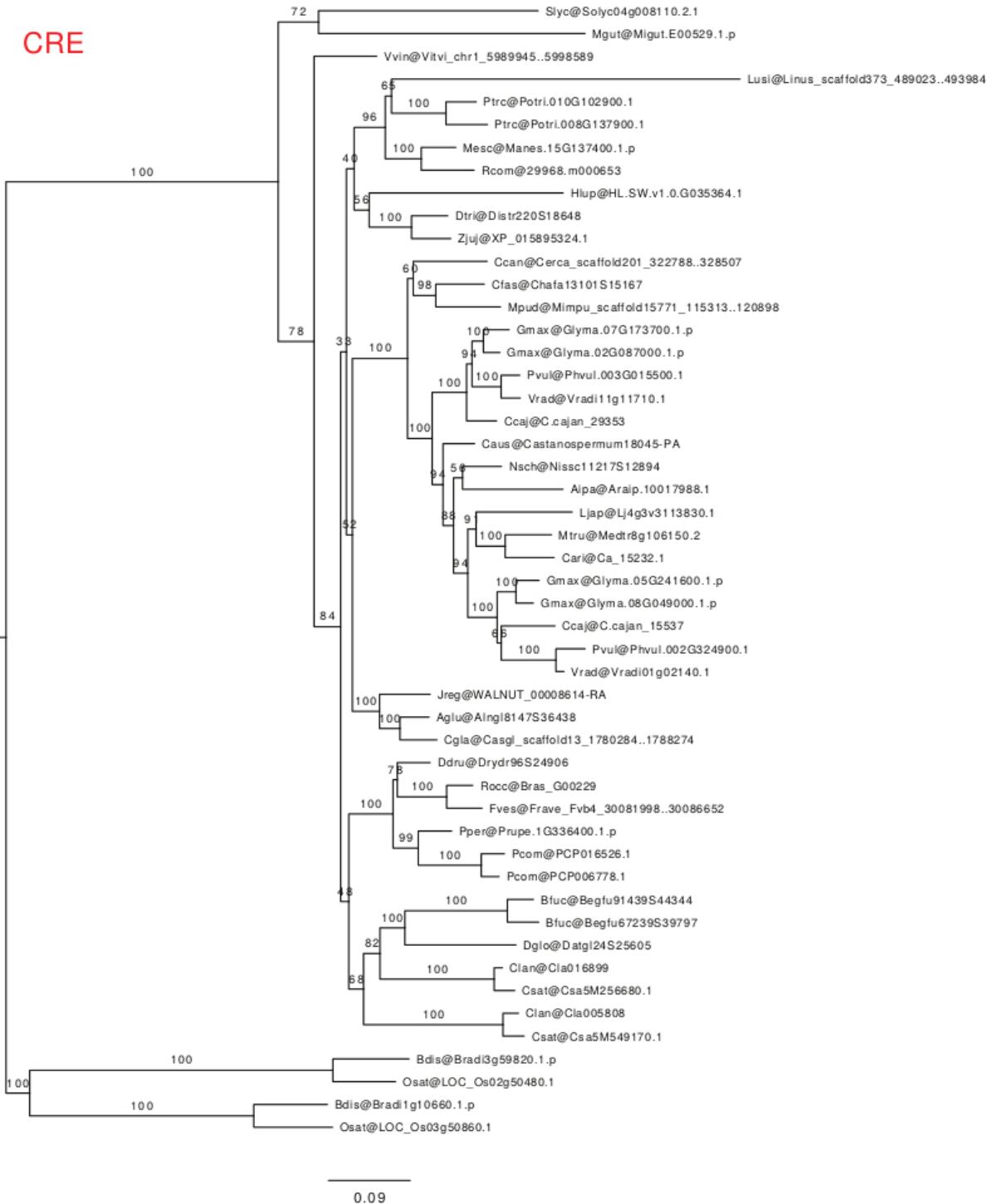


Fig. S15. Maximum-likelihood phylogenetic tree of CRE.

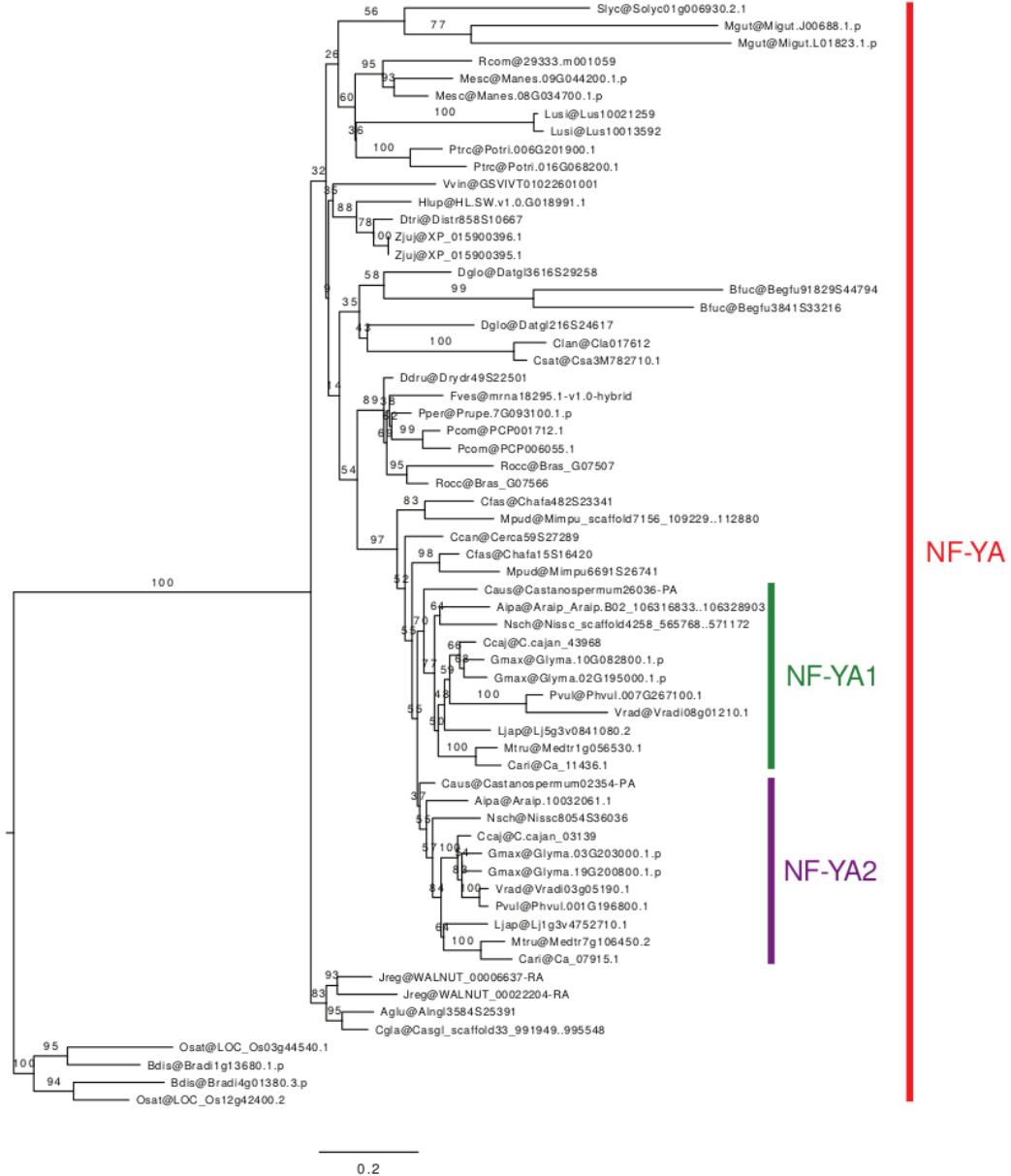


Fig. S16. Maximum-likelihood phylogenetic tree of NF-YA.

NSP1

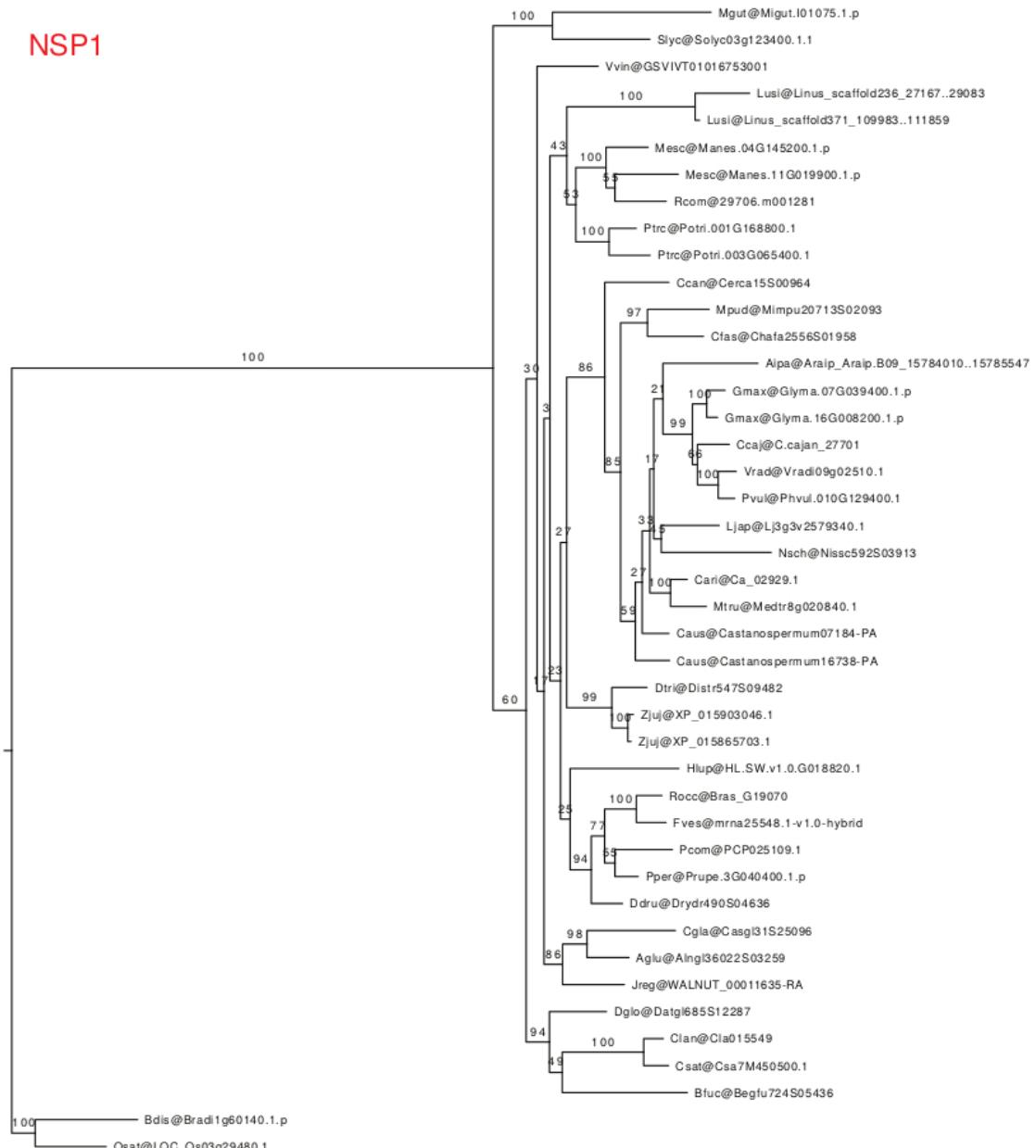


Fig. S17. Maximum-likelihood phylogenetic tree of NSP1.

NSP2

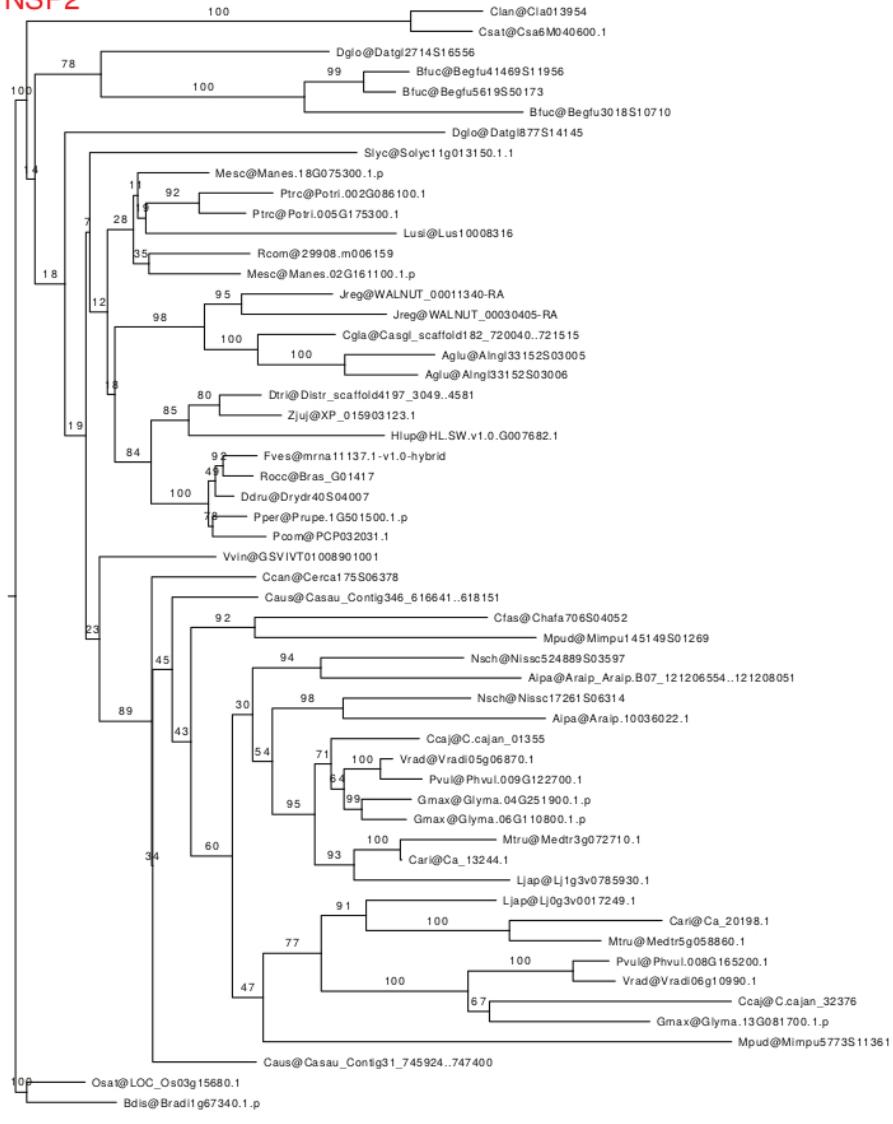


Fig. S18. Maximum-likelihood phylogenetic tree of NSP2.



Fig. S19. Maximum-likelihood phylogenetic tree of RDN1.

SUNER1

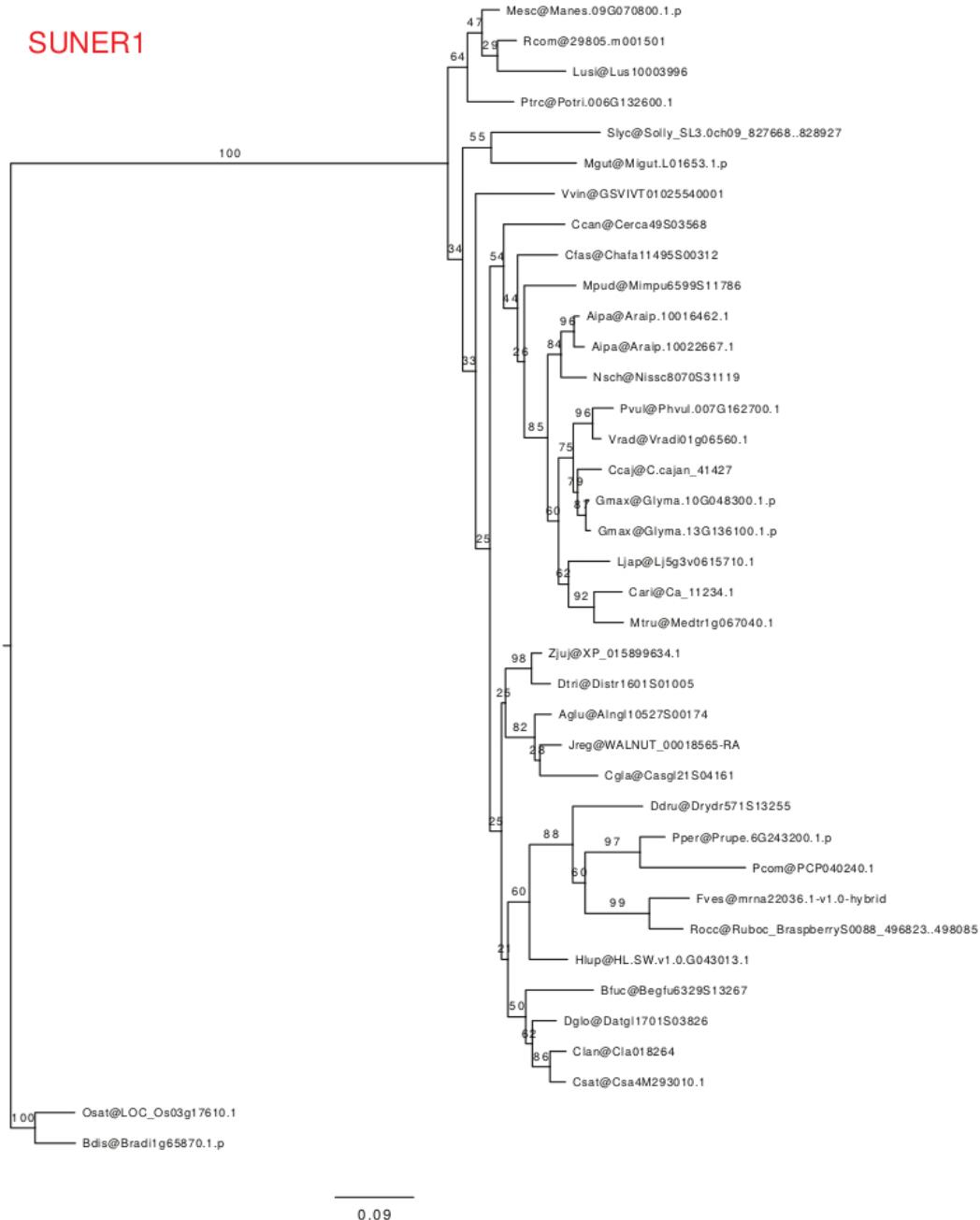


Fig. S20. Maximum-likelihood phylogenetic tree of SUNER1.

SUNN

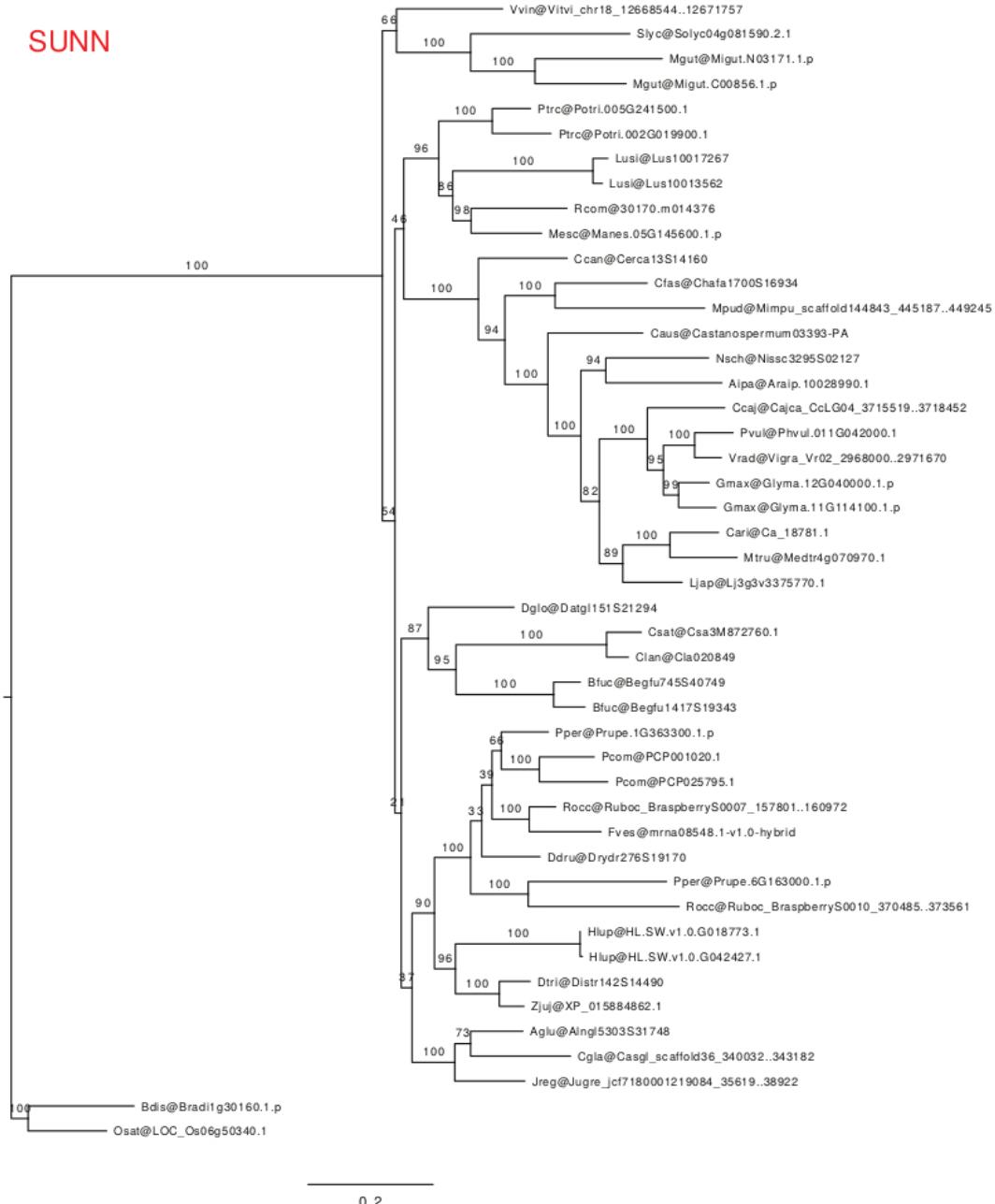


Fig. S21. Maximum-likelihood phylogenetic tree of SUNN.

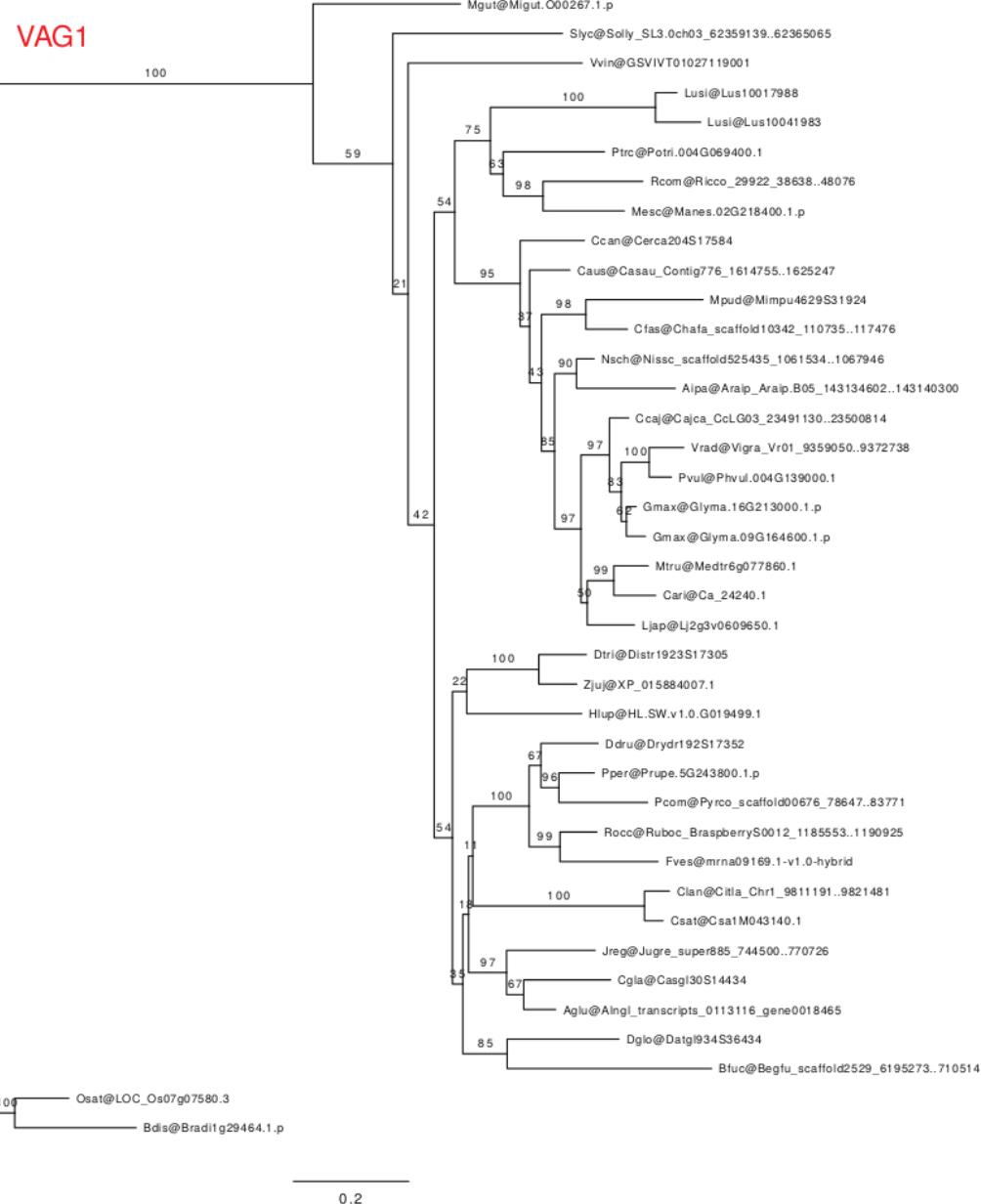


Fig. S22. Maximum-likelihood phylogenetic tree of VAG1.

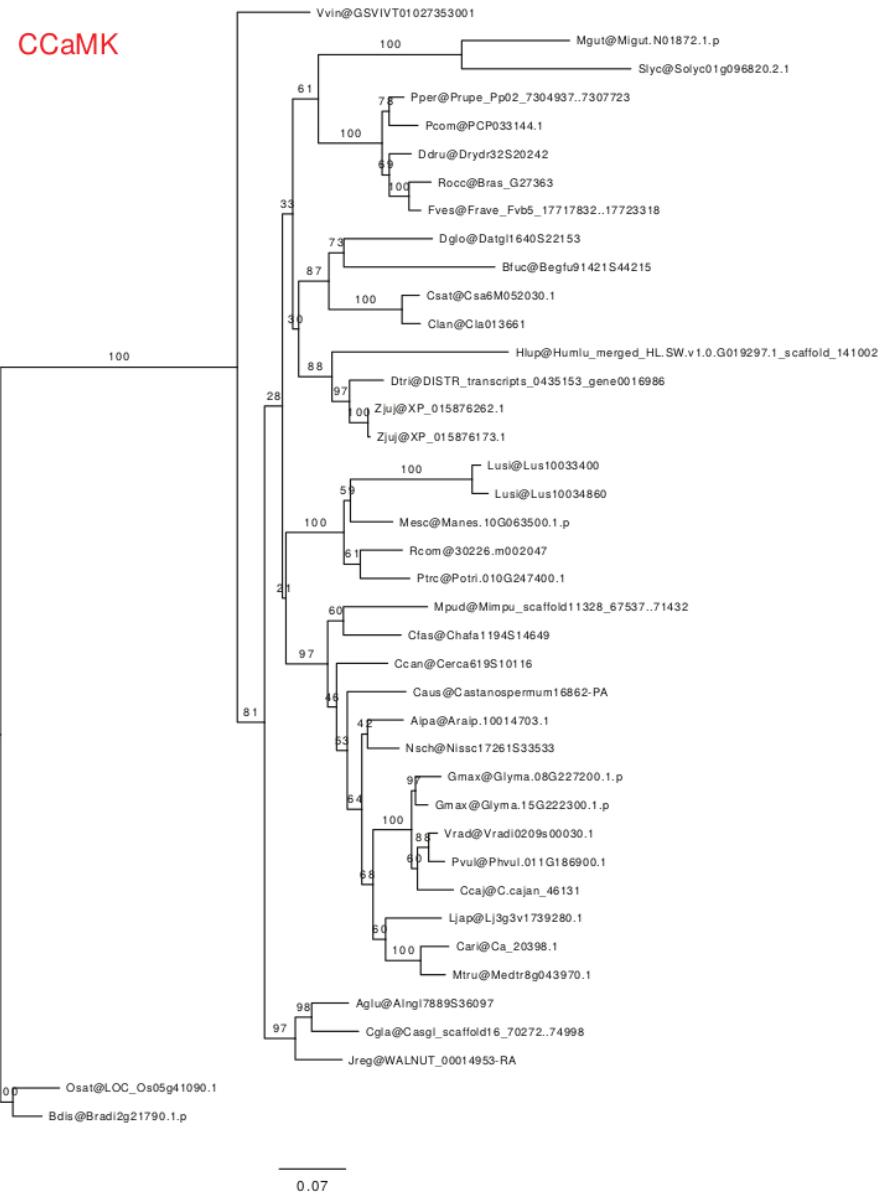


Fig. S23. Maximum-likelihood phylogenetic tree of CCaMK.

CYCLOPS

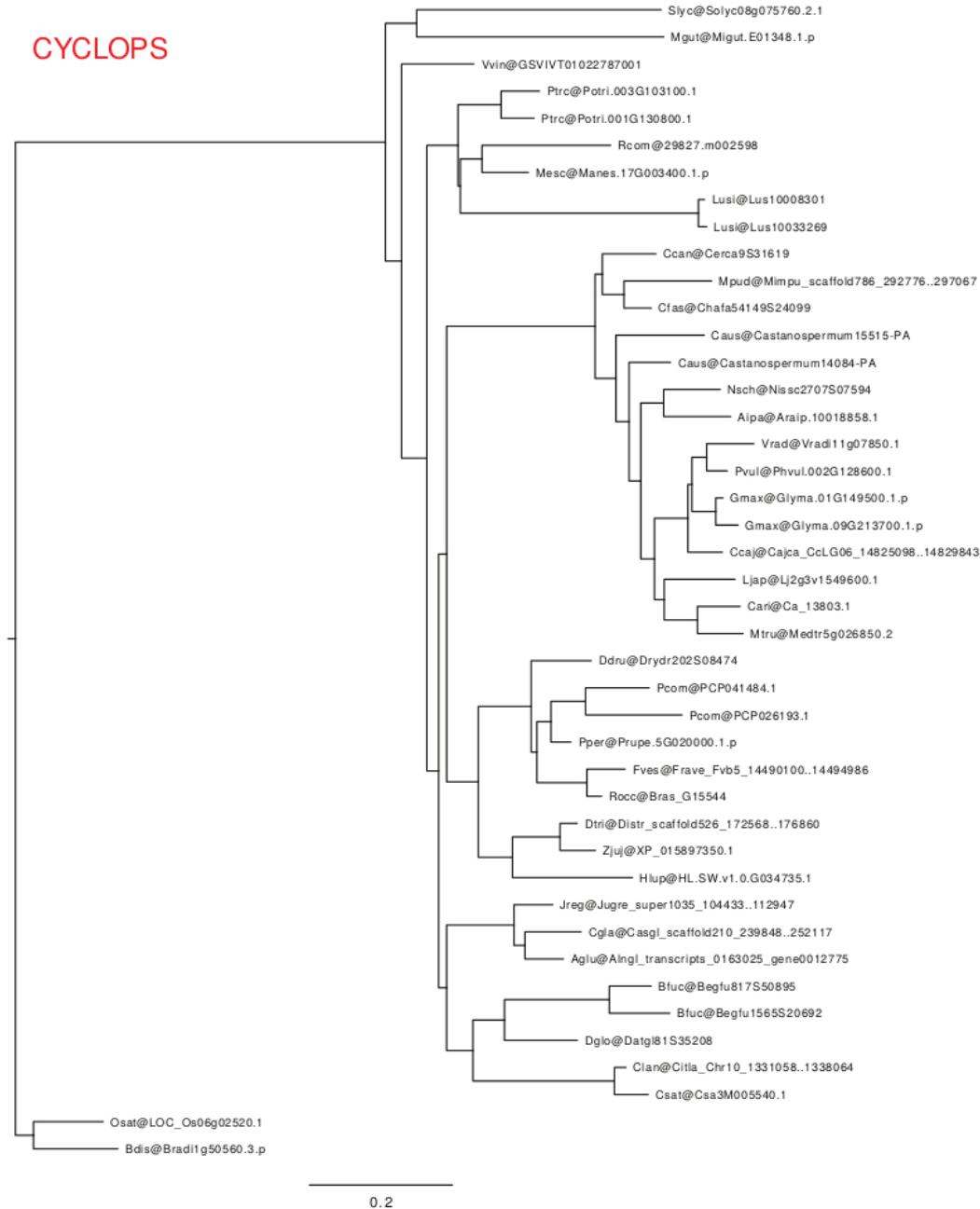


Fig. S24. Maximum-likelihood phylogenetic tree of CYCLOPS.

SYMRK

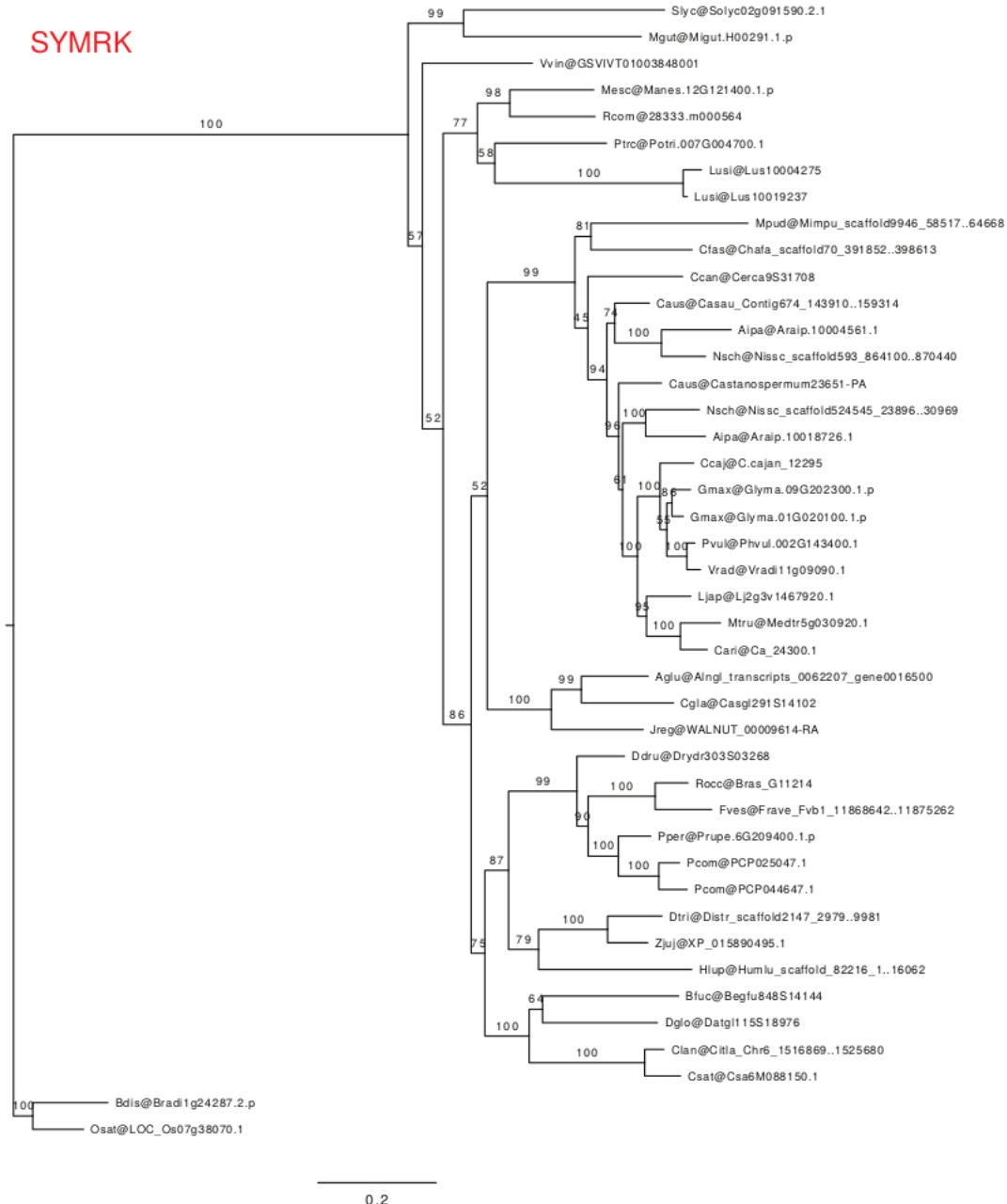


Fig. S25. Maximum-likelihood phylogenetic tree of SYMRK.

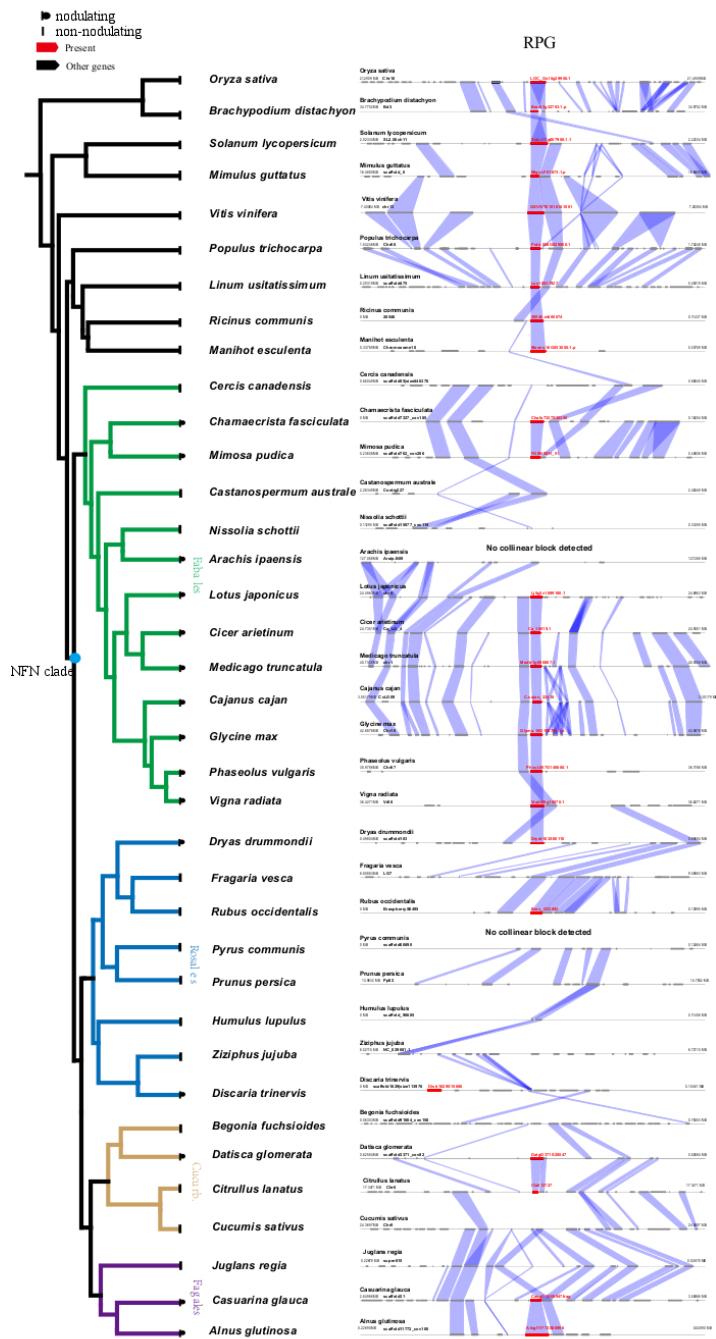


Fig. S26. Syntenic relationships of the *RPG* region. *RPG* genes are colored in red. *RPG* gene IDs are shown above the gene symbol. The synteny analysis upholds orthologous relationships drawn from the phylogenetic analysis and supports the absence of *RPG* in several species by verifying the existence of contiguous *RPG* regions without *RPG* genes.

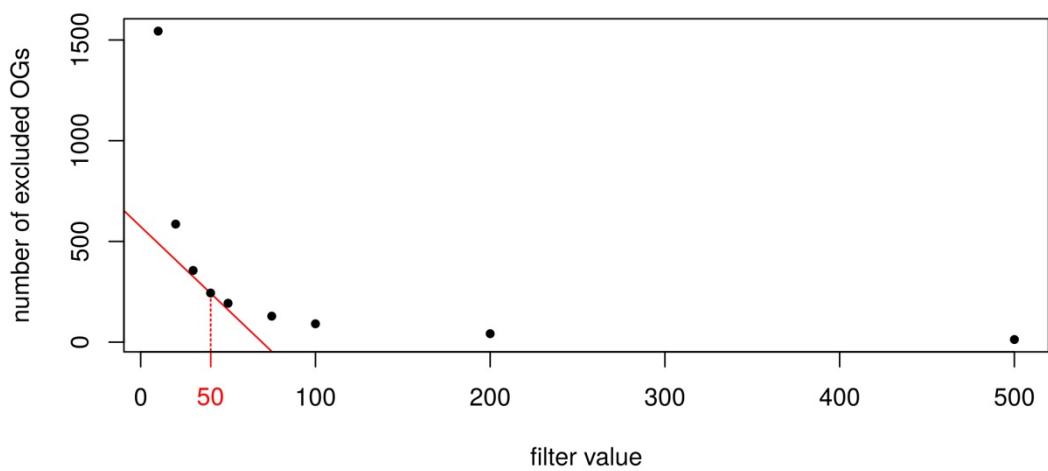


Fig. S27. Rationale for the exclusion of clusters due to strong outliers in gene copy number.
 To identify strong outliers in gene copy number for each cluster we counted the number of genes per species. We subtracted the maximum in gene copy number from the median gene copy number of all species. If the difference was larger than 50 we excluded the Orthofinder cluster. The figure shows that the value 50 meets the elbow criterion to identify the optimal number of clusters in a clustering problem. The value 50 leads to the exclusion of 193 gene clusters.

Species abbreviation	Species tax name	Genome release version	Genome source
Osat	<i>Oryza sativa</i>	7	Phytozome
Bdis	<i>Brachypodium distachyon</i>	3,1	Phytozome
Mgut	<i>Mimulus guttatus</i>	2	Phytozome
Slyc	<i>Solanum lycopersicum</i>	2,4	SGN
Vvin	<i>Vitis vinifera</i>	12x	Phytozome
Mesc	<i>Manihot esculenta</i>	6,1	Phytozome
Ptrc	<i>Populus trichocarpa</i>	3	Phytozome
Lusi	<i>Linum usitatissimum</i>	1	Phytozome
Rcom	<i>Ricinus communis</i>	0,1	Phytozome
Ccan	<i>Cercis canadensis</i>	1	this study (GigaDB)
Mpud	<i>Mimosa pudica</i>	1	this study (GigaDB)
Cfas	<i>Chamaecrista fasciculata</i>	1	this study (GigaDB)
Caus	<i>Castanospermum australe</i>	1	Henk Hillhorst
Nsch	<i>Nissolia schottii</i>	1	this study (GigaDB)
Aipa	<i>Arachis ipaensis</i>	K30076.a1.G1	PeanutBase
Ljap	<i>Lotus japonicus</i>	3	Kazusa
Cari	<i>Cicer arietinum</i>	2	CSFL
Mtru	<i>Medicago truncatula</i>	4	JCVI
Ccaj	<i>Cajanus cajan</i>	5	GigaDB
Gmax	<i>Glycine max</i>	Wm82.a2.v1	Phytozome
Pvul	<i>Phaseolus vulgaris</i>	1	Phytozome
Vrad	<i>Vigna radiata</i>	6	LIS
Ddru	<i>Dryas drummondii</i>	1_b2	this study (GigaDB)
Fves	<i>Fragaria vesca</i>	1,1	Phytozome
Rocc	<i>Rubus occidentalis</i>	1.0.a1	GDR
Pcom	<i>Pyrus communis</i>	1	GDR
Pper	<i>Prunus persica</i>	2,1	Phytozome
Hlup	<i>Humulus lupulus</i>	1.1 shinsuwase	HopBase
Zjuj	<i>Ziziphus jujuba</i>	1,1	NCBI
Dtri	<i>Discaria trinervis</i>	1	this study (GigaDB)
Bfuc	<i>Begonia fuchsoides</i>	1	this study (GigaDB)
Dglo	<i>Datisca glomerata</i>	1_b2	this study (GigaDB)
Clan	<i>Citrullus lanatus</i>	97103 1.0	ICuGI
Csat	<i>Cucumis sativus</i>	Chinese Long 2.0	ICuGI
Jreg	<i>Juglans regia</i>	1	HWG
Cgla	<i>Casuarina glauca</i>	1,1	this study (GigaDB)
Aglu	<i>Alnus glutinosa</i>	1	this study (GigaDB)

Table S1. Genome and annotation resources used in this study

Species	Raw data (Gb)	clean data (Gb)	clean data X	raw read (#Million)	clean read (#Million)
<i>Alnus glutinosa</i>	150.1	79.11	171.57	1 501	852
<i>Casuarina glauca</i>	278.65	106.95	341.06	2 786	1 120
<i>Cercis canadensis</i>	335.07	134.99	220.74	3 253	1 365
<i>Chamaecrista fasciculata</i>	381.07	270.76	492.29	3 704	2 750
<i>Mimosa pudica</i>	370.07	167.28	287.65	3 582	1 750
<i>Nissolia schottii</i>	144.06	70.41	149.48	1 440	749
<i>Begonia fuchsioides</i>	177	118.58	126.96	1 769	1 260
<i>Datisca glomerata</i>	216.55	91.39	110.51	2 165	977
<i>Dryas drummondii</i>	196.49	75.46	298.42	1 964	804
<i>Discaria trinervis</i>	191.53	97.55	120	1 915	1 060

Table S2. Detailed statistics of data production

Species	Predict size	Main Assemblers	Gapfilled scaffold			Gapfilled scaftig		
			Total length (bp)	N50 (bp)	Assembly Percentage (%)	Total length (bp)	N50 (bp)	Assembly Percentage (%)
<i>Alnus glutinosa</i>	461Mb	Platanus	612Mb	96 508	132.74%	552Mb	5 330	119.73%
<i>Casuarina glauca</i>	314Mb	Soapdenovo	283MB	912 668	90.2%	276Mb	41 115	88.02%
<i>Cercis canadensis</i>	301Mb	Celera Assembler	330Mb	421 030	109.68%	318Mb	12 883	105.61%
<i>Chamaecrista fasciculata</i>	550Mb	Platanus	429Mb	96 643	78.08%	413Mb	14 934	75.21%
<i>Mimosa pudica</i>	896Mb	Platanus	557Mb	119 676	62.18%	530Mb	11 069	59.16%
<i>Nissolia schottii</i>	471Mb	Platanus	466Mb	179 654	98.94%	456Mb	20 655	96.88%
<i>Begonia fuchsoides</i>	935Mb	Platanus	374Mb	154 265	40.00%	368Mb	29 236	39.33%
<i>Datisca glomerata</i>	827Mb	Platanus	688Mb	1 186 304	83.25%	687Mb	89 992	83.04%
<i>Dryas drummondii</i>	253Mb	Soapdenovo	233Mb	979 416	92.11%	229Mb	35 876	90.55%
<i>Discaria trinervis</i>	650Mb	Celera Assembler	310Mb	115 599	47.71%	308Mb	27 329	47.44%

Table S3. The statistics of genome assemblies.

Species	BUSCO result				
	Complete	Single	Duplicate	Fragment	Missing
<i>Alnus glutinosa</i>	88.2%	79.3%	8.9%	4.9%	6.9%
<i>Casuarina glauca</i>	91.2%	66%	25.2%	2.8%	6%
<i>Cercis canadensis</i>	95.7%	92.6%	3.1%	1.3%	3%
<i>Chamaecrista fasciculata</i>	93.2%	86.1%	7.1%	1.7%	5.1%
<i>Mimosa pudica</i>	94.3%	92.5%	1.8%	1.7%	4%
<i>Nissolia schottii</i>	95%	93.1%	1.9%	1.5%	3.5%
<i>Begonia fuchsoides</i>	96.7%	93.7%	3%	0.8%	2.5%
<i>Datisca glomerata</i>	96.2%	91.6%	4.6%	1.5%	2.3%
<i>Dryas drummondii</i>	90.8%	86.4%	4.4%	3.9%	5.3%
<i>Discaria trinervis</i>	92.5%	85.6%	6.9%	2.2%	5.3%

Table S4. Evaluation of genome assembly completeness using BUSCO.

Species	Gene number	BUSCO result				
		Complete	Single	Duplicate	Fragment	Missing
<i>Alnus glutinosa</i>	43 087	81.4%	70.6%	10.8%	10.6%	8%
<i>Begonia fuchsioides</i>	51 638	92.5%	66.5%	26%	2.5%	5%
<i>Casuarina glauca</i>	26 282	95.3%	92.4%	2.9%	2.3%	2.4%
<i>Cercis canadensis</i>	34 023	96.2%	93.3%	2.9%	1.9%	1.9%
<i>Chamaecrista fasciculata</i>	32 832	94.3%	86.4%	7.9%	2.6%	3.1%
<i>Datisca glomerata</i>	37 042	90.6%	88.8%	1.8%	3.2%	6.2%
<i>Discaria trinervis</i>	32 886	91.7%	89.2%	2.5%	1.9%	6.4%
<i>Dryas drummondii</i>	25 030	90.3%	87.9%	2.4%	1.9%	7.8%
<i>Mimosa pudica</i>	33 108	78%	62.8%	15.2%	3.1%	18.9%
<i>Nissolia schottii</i>	36 369	92.8%	86.3%	6.5%	2.6%	4.6%

Table S5. Evaluation of gene models using BUSCO.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	16.14	35.01	94_94	14.15	30.69
170	100_100	16.42	35.62	90_90	13.67	29.65
250	100_100	10.21	22.15	94_94	8.74	18.96
250	100_100	11.46	24.86	92_92	9.6	20.82
350	100_100	9.14	19.83	94_94	7.97	17.29
2000	100_100	13.03	28.26	93_93	8.34	18.09
6000	100_100	19.11	41.45	94_94	6.11	13.25
10000	100_100	27.66	60	92_92	6.41	13.9
20000	100_100	26.94	58.44	94_94	4.11	8.92
Total		150.1	325.62		79.11	171.57

Table S6. Statistics of DNA sequencing data of *Alnus glutinosa*.

	Scaffolds		Scaftrigs	
	Length (bp)	Number	Length (bp)	Number
Maximal length	799 831		79 120	
N90	1 154	38 099	830	129 463
N80	4 763	11 701	1 643	83 220
N70	17 664	4 288	2 631	56 458
N60	55 573	2 384	3 913	39 359
N50	96 508	1 553	5 330	27 265
N40	138 007	1 025	7 038	18 237
N30	185 871	641	9 286	11 390
N20	245 612	352	12 356	6 216
N10	347 897	141	17 356	2 392
Total length	611 918 954		551922622	
Number >=100bp		167 455		262 559
Number >=2kb		24 871		70 759

Table S7. Statistics of the assembled sequence length of *Alnus glutinosa*.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (x)
170	100_100	23.23	74.22	96_96	1.67	53.44
250	100_100	22.66	72.39	96_96	1.53	48.82
350	100_100	11.04	35.29	96_96	6.32	20.18
500	100_100	18.03	57.63	95_95	1.61	51.28
800	100_100	9.89	31.59	95_95	8.61	27.51
2000	100_100	4.46	142.41	95_95	1.14	36.29
6000	100_100	2.92	93.23	95_95	8.02	25.64
10000	100_100	6.8	217.29	96_96	1.13	36.13
20000	100_100	5.2	166.21	95_95	1.33	42.39
Total		278.65	890.25		106.95	341.69

Table S8. Statistics of DNA sequencing data of *Casuarina glauca*.

	Scaffold		ScafTig	
	Length(bp)	Number	Length(bp)	Number
Max length	6 487 367		310 342	
N90	32 449	600	3 940	8 972
N80	207 270	276	12 865	5 351
N70	387 466	178	21 600	3 714
N60	629 167	121	30 995	2 650
N50	912 668	84	41 115	1 879
N40	1 146 642	57	53 276	1 290
N30	1 531 651	35	67 357	829
N20	2 084 067	19	85 676	462
N10	3 014 060	7	117 303	181
Total length	282 815 317		275 989 893	
num>=100bp		39 801		52 963
num>=2kb		2 057		11 362

Table S9. Statistics of the assembled sequence length of *Casuarina glauca*.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	25.9	86.08	94_91	14.64	48.65
170	100_100	53.03	176.28	94_92	30.6	101.99
250	100_100	25.19	83.73	94_91	12.45	41.4
250	100_100	52.17	173.41	94_92	26.18	87.66
350	100_100	14.48	48.13	94_91	4.96	16.47
350	100_100	40.48	134.56	94_92	20.23	67.44
500	150_150	15.24	50.67	144_141	11.59	38.53
800	150_150	13.89	46.18	140_140	10.74	35.7
2000	100_100	43.64	145.07	94_91	0.15	0.51
10000	100_100	51.06	169.73	91_91	3.45	11.47
Total		335.07	1113.85		134.99	362.16

Table S10. Statistics of DNA sequencing data of *Cercis canadensis*.

	Scaffold		ScafTig	
	Length(bp)	Number	Length(bp)	Number
Max length	3 438 730		199 472	
N90	50 393	964	3 320	25 260
N80	139 655	588	5 606	18 011
N70	235 376	406	7 834	13 240
N60	313 771	285	10 180	9 690
N50	421 030	193	12 883	6 917
N40	570 332	127	16 088	4 706
N30	740 456	76	20 389	2 948
N20	1 048 042	38	26 687	1 575
N10	1 926 456	14	39 179	568
Total length	329 952 350		317707279	
num>=100bp		8 831		52 306
num>=2kb		3 359		31 384

Table S11. Statistics of the assembled sequence length of *Cercis canadensis*.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	Clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	26.69	37.82	95_95	20.8	48.54
170	100_100	52	75.27	95_95	41.4	94.55
250	100_100	24.3	35.17	95_95	19.35	44.18
350	100_100	22.37	35.6	95_95	19.58	40.67
170	100_100	63.68	98.26	95_95	54.04	115.78
500	150_150	18.53	28.24	145_145	15.53	33.7
500	100_100	23.31	36.64	95_95	20.15	42.39
800	150_150	13.42	20.19	145_145	11.11	24.39
800	100_100	12.87	20.93	95_95	11.51	23.41
2000	100_100	31.58	38.17	95_95	20.99	57.42
2000	100_100	27.03	20.54	95_95	11.29	49.15
6000	100_100	21.75	24.05	95_95	13.23	39.55
10000	100_100	43.5	21.41	95_95	11.78	79.1
Total		381.07	692.85		270.76	492.29

Table S12. Statistics of DNA sequencing data of *Chamaecrista fasciculata*.

Scaffold		ScafTig	
		Length(bp)	Number
Max length	723 940		138 716
N90	11 197	5 583	2 435
N80	31 933	3 452	5 322
N70	51 642	2 403	8 274
N60	72 058	1 701	11 461
N50	96 643	1 188	14 934
N40	128 519	803	18 976
N30	162 222	505	23 954
N20	215 809	273	30 733
N10	316 121	107	42 393
Total length	429 272 528		413 494 121
num>=100bp		56 700	94 086
num>=2kb		9 483	35 073

Table S13. Statistics of the assembled sequence length of *Chamaecrista fasciculata*.

Insert size(bp)	Read length (bp)	Raw data (Gb)	Raw data depth (X)	clean length (bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	27.67	30.88	94_91	16.03	17.89
170	100_100	51	56.91	90_90	32.82	36.62
250	100_100	26.76	29.86	94_91	13.03	14.54
250	100_100	52.65	58.75	90_90	27.78	31
350	100_100	15.85	17.68	94_91	5.76	6.43
350	100_100	50.47	56.32	90_90	26.08	29.1
500	100_100	16.58	18.5	142_129	13.24	14.77
800	100_100	11.12	12.4	139_139	8.41	9.38
2000	100_100	40.49	45.18	94_91	12.34	13.77
6000	100_100	30.63	34.18	91_91	8.58	9.58
10000	100_100	46.86	52.29	90_90	3.22	3.59
Total		370.07	412.96		167.28	186.67

Table S14. Statistics of DNA sequencing data of *Mimosa pudica*.

	Scaffold		Scafseq	
	Length(bp)	Number	Length(bp)	Number
Max length	839 160		97 653	
N90	4 438	8 048	1 606	57 106
N80	30 591	3 921	4 008	37 083
N70	59 165	2 616	6 196	26 508
N60	88 032	1 844	8 531	19 252
N50	119 676	1 302	11 069	13 786
N40	153 844	890	14 011	9 528
N30	198 373	572	17 447	6 127
N20	252 136	321	22 306	3 430
N10	344 353	128	30 347	1 362
Total length	557 207 566		530 112 146	
num>=100bp		97 900		154 619
num>=2kb		12 044		52 304

Table S15. Statistics of the assembled sequence length of *Mimosa pudica*.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	Clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	28.01	59.46	94_94	16.98	36.05
250	100_100	24.52	62.65	94_94	12.97	27.53
350	100_100	17.41	38.98	94_92	6.89	14.63
500	100_100	25.87	122.29	94_92	21.69	14.25
800	100_100	14.02	172.44	94_91	7.63	17.55
2000	100_100	23.14	105.56	94_92	6.99	8.13
6000	100_100	23.72	102.8	94_92	3.97	8.43
Total		156.69	664.2		77.12	163.74

Table S16. Statistics of DNA sequencing data of *Nissolia schottii*.

	Scaffold		ScafTig	
	Length(bp)	Number	Length(bp)	Number
Max length	2 079 550		308 931	
N90	1 427	18 385	1 147	43 107
N80	11 670	3 642	3 852	20 188
N70	42 996	1 562	8 699	12 401
N60	94 157	837	14 356	8 341
N50	179 654	483	20 655	5 690
N40	313 492	284	27 649	3 779
N30	497 247	166	36 661	2 341
N20	703 731	87	49 585	1 268
N10	1 134 304	32	70 349	486
Total length	466 120 566		456 386 186	
num>=100bp		116 257		146 370
num>=2kb		12 569		29 227

Table S17. Statistics of the assembled sequence length of *Nissolia schottii*.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	Clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	49.34	52.77	94_94	30.62	32.75
250	100_100	20.93	22.39	94_92	17.04	18.22
250	100_100	13.73	14.68	94_92	11.46	12.26
350	100_100	19.87	21.25	94_92	16.66	17.82
350	100_100	12.48	13.35	94_92	10.64	11.38
500	100_100	14.56	15.57	94_94	10.26	10.97
800	100_100	11.7	12.51	94_94	10.21	10.92
2000	100_100	12.32	13.18	94_92	7.61	8.14
6000	100_100	22.06	23.6	94_94	4.08	4.36
Total		177	189.3		118.58	126.83

Table S18. Statistics of DNA sequencing data of *Begonia fuchsioides*.

	Scaffold		Scafseq	
	Length(bp)	Number	Length(bp)	Number
Max length	1 899 432		507 112	
N90	4 356	6 188	2 383	20 114
N80	23 940	2 328	6 461	10 860
N70	57 770	1 352	12 686	6 843
N60	101 506	869	20 198	4 556
N50	154 265	569	29 236	3 040
N40	219 486	363	40 313	1 967
N30	292 035	215	54 435	1 178
N20	419 274	109	77 946	610
N10	683 419	38	116 876	219
Total length	373 930 519		367 616 434	
num>=100bp		55 040		74 847
num>=2kb		10 679		22 358

Table S19. Statistics of the assembled sequence length of *Begonia fuchsoides*.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	22.42	27.11	94_94	19.5	23.58
250	100_100	28.06	33.93	93_91	8.44	10.21
350	100_100	9.92	11.99	94_94	8.29	10.02
500	100_100	36.88	44.6	94_91	29.91	36.17
800	100_100	8.55	10.34	93_91	6.77	8.18
2000	100_100	30.94	37.42	94_91	8.13	9.83
6000	100_100	21.13	25.55	91_91	3.66	4.42
10000	100_100	58.66	70.94	91_91	6.69	8.09
Total		216.55	261.87		91.39	110.51

Table S20. Statistics of DNA sequencing data of *Datisca glomerata*.

	Scaffold		ScafTig	
	Length(bp)	Number	Length(bp)	Number
Max length	6 997 071		460 556	
N90	265 133	608	23 472	7 932
N80	517 372	426	40 289	5 734
N70	739 139	317	55 663	4 294
N60	949 652	236	72 175	3 213
N50	1 186 304	171	89 992	2 363
N40	1 545 777	120	109 917	1 670
N30	1 785 324	80	131 396	1 098
N20	2 386 593	46	164 067	628
N10	2 880 265	19	215 251	257
Total length	688 405 811		686 655 549	
num>=100bp		13 868		27 262
num>=2kb		2 125		13 952

Table S21. Statistics of the assembled sequence length of *Datisca glomerata*.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	23.8	94.12	94_92	13.12	51.89
250	100_100	34.26	135.51	94_92	15.52	61.39
350	100_100	16.33	64.59	94_91	5.62	22.22
500	100_100	27.82	110.04	94_91	21.88	86.55
800	100_100	6.77	26.78	94_91	5.17	20.45
2000	100_100	32.99	130.47	94_91	7.43	29.37
6000	100_100	18.6	73.57	94_91	3.59	14.18
10000	100_100	35.91	142.02	91_91	3.13	12.37
Total		196.49	777.1		75.46	298.42

Table S22. Statistics of DNA sequencing data of *Dryas drummondii*.

	Scaffold		ScafTig	
	Length(bp)	Number	Length(bp)	Number
Max length	5 217 279		210 940	
N90	166 916	274	7 631	6 894
N80	351 279	181	14 767	4 788
N70	515 990	127	21 283	3 501
N60	749 113	89	28 601	2 575
N50	979 416	62	35 876	1 860
N40	1 272 740	41	44 922	1 290
N30	1 614 940	25	55 933	834
N20	2 373 954	13	71 067	470
N10	3 758 852	6	95 044	191
Total length	232 912 488		228 963 732	
num>=100bp		13 426		25 960
num>=2kb		1 292		10 336

Table S23. Statistics of the assembled sequence length of *Dryas drummondii*.

Insert size(bp)	Read length (read1_read2. bp)	Raw data (Gb)	Raw data depth (X)	clean length (read1_read2. bp)	Clean data (Gb)	Clean data depth (X)
170	100_100	23.51	36.2	94_94	13.79	21.23
170	100_100	51.99	80.04	94_92	31.64	48.71
250	100_100	27.21	41.89	94_94	13.4	20.63
350	100_100	16.53	25.45	94_94	5.96	9.18
500	100_100	24.27	37.36	90_90	19.07	29.37
800	100_100	9.85	15.16	91_91	7.89	12.15
2000	100_100	38.17	58.77	92_92	5.79	8.91
Total		191.53	294.88		97.55	150.19

Table S24. Statistics of DNA sequencing data of *Discaria trinervis*.

Scaffold		ScafTig		
	Length(bp)	Number	Length(bp)	Number
Max length	698 860		292 169	
N90	31 817	2 743	6 955	11 646
N80	52 048	1 990	11 790	8 290
N70	72 153	1 488	16 468	6 101
N60	93 452	1 113	21 617	4 471
N50	115 599	814	27 329	3 204
N40	140 733	571	34 258	2 193
N30	173 322	374	43 225	1 391
N20	216 079	213	55 867	760
N10	279 342	88	79 061	288
Total length	309 860 975		308 140 234	
num>=100bp		8 269		26 610
num>=2kb		4 704		17 257

Table S25. Statistics of the assembled sequence length of *Discaria trinervis*.

Species	Kmer	Kmer number	Peak depth	Estimated genome size	Used bases	# of Used reads	Used data X
<i>Alnus glutinosa</i>	17	2,63E+10	57	461Mb	3,18E+10	3,40E+08	68.79
<i>Casuarina glauca</i>	17	2,26E+10	72	314Mb	2,71E+10	2,83E+08	86.42
<i>Cercis canadensis</i>	17	1,95E+10	65	300Mb	2,20E+10	1,54E+08	73.19
<i>Chamaecrista fasciculate</i>	17	2,26E+10	41	550Mb	2.54E+10	1,76E+08	46.12
<i>Mimosa pudica</i>	17	1,97E+10	22	896Mb	2,21E+10	1,54E+08	24.75
<i>Nissolia schottii</i>	17	4,48E+10	95	471Mb	5,38E+10	5,67E+08	114.2
<i>Begonia fuchsioides</i>	27	4,02E+10	43	935Mb	5,58E+10	6,00E+08	59.69
<i>Datisca glomerata</i>	17	2,72E+10	33	827Mb	3,29E+10	3,50E+08	39.76
<i>Dryas drummondii</i>	17	3,31E+10	131	253Mb	4,00E+10	4,33E+08	158.39
<i>Discaria trinervis</i>	17	3,18E+10	49	649Mb	3,82E+10	4,02E+08	58.92

Table S26. Result of K-mer analysis.

	<i>Ailnus glutinosa</i>	<i>Casuarina glauca</i>	<i>Cercis canadensis</i>	<i>Chamaecrista fasciculata</i>	<i>Mimosa pudica</i>	<i>Nissolia schottii</i>	<i>Begonia fuchsioides</i>	<i>Datisca glomerata</i>	<i>Dryas drummondii</i>	<i>Discaria trinervis</i>	
Total	gene number	43 087	26 282	34 023	32 832	33 108	36 369	51 638	37 042	25 030	32 886
	exon number	190 680	145 573	173 547	162 383	166 964	184 570	260 820	169 486	130 647	174 760
	intron number	147 593	119 291	139 524	129 551	133 856	148 201	209 182	132 444	105 617	141 874
	intron length(bp)	94 817 110	71 442 241	73 542 998	67 402 144	75 937 949	74 165 233	68 470 827	77 462 991	46 502 539	56 789 932
average	mRNA length(bp)	3329.11	4035.57	3408.1	3370.15	3661.78	3251.14	2467.97	3104.33	3208.54	2934.62
	cds length(bp)	1128.52	1317.28	1246.54	1317.21	1368.13	1211.9	1141.99	1013.11	1350.66	1207.75
	exon number	4.43	5.54	5.1	4.95	5.04	5.07	5.05	4.58	5.22	5.31
	exon length(bp)	255.01	237.82	244.38	266.33	271.29	238.8	226.09	221.42	258.77	227.27
	intron length(bp)	642.42	598.89	527.1	520.27	567.31	500.44	327.32	584.87	440.29	400.28

Table S27. Transposable elements in assemblies.

Table S28. Statistic of annotated genes.

	<i>Alnus glutinosa</i>	<i>Casuarina glauca</i>	<i>Cercis canadensis</i>	<i>Chamaecrista fasciculata</i>	<i>Mimosa pudica</i>	<i>Nissolia schottii</i>	<i>Begonia fuchsioides</i>	<i>Datiscia glomerata</i>	<i>Dryas drummondii</i>	<i>Discaria trinervis</i>										
	Number	Percent(%)	Number	Percent(%)	Number	Percent(%)	Number	Percent(%)	Number	Percent(%)										
Total	43 087		26 282		34 023		32 832		33 108		36 369		51 638		37 042		25 030		32 886	
InterPro	30 010	69.65	19 127	72.78	23 557	69.24	24 073	73.32	25 517	77.07	24 923	68.53	35 244	68.25	21 913	59.16	19 216	76.77	19 042	57.9
GO	17 488	40.59	12 765	48.57	14 169	41.658	14 320	43.62	15 978	48.26	14 674	40.35	22 286	43.16	12 126	32.74	12 188	48.69	10 496	31.92
KEGG	32 390	75.17	19 382	73.75	25 474	74.87	24 730	75.32	25 881	78.17	26 175	71.97	37 264	72.16	27 382	73.92	19 258	76.94	23 874	72.6
Swissprot	30 330	70.39	20 051	76.29	23 514	69.11	23 873	72.71	25 149	75.96	24 987	68.7	38 355	74.28	23 895	64.51	18 610	74.35	22 232	67.6
Annotated	36 173	83.95	21 872	83.22	28 356	83.34	27 661	84.25	28 832	87.08	29 565	81.29	42 272	81.86	30 272	81.72	21 687	86.64	26 458	80.45
Unannotated	6 914	16.05	4 410	16.78	5 667	16.66	5 171	15.75	4 276	12.92	6 804	18.71	9 366	18.14	6 770	18.28	3 343	13.36	6 428	19.55

Table S29. Summary of genes with function description.

Medicagoid	Medicago description	Confirmation
Medtr1g018480.1	calcium-binding EF hand-like protein	not specific for NFN clade
Medtr1g063940.3	myb transcription factor	not specific for NFN clade
Medtr1g067750.4	DNA-binding protein, putative	not specific for NFN clade
Medtr1g076840.1	RmIC-like cupins superfamily protein	not specific for NFN clade
Medtr2g039340.1	cytokinin oxidase/dehydrogenase-like protein	not specific for NFN clade
Medtr2g076650.1	epidermis-specific secreted EP1-like glycoprotein	not specific for NFN clade
Medtr2g078310.1	PHD finger and bromo-adjacent-like domain protein	not specific for NFN clade
Medtr2g099155.2	Myb-like DNA-binding domain protein	not specific for NFN clade
Medtr3g073100.1	phosphatidylinositol-4-phosphate 5-kinase family protein	not specific for NFN clade
Medtr3g094260.1	nucleotide/sugar transporter family protein	not specific for NFN clade
Medtr3g094270.1	nucleotide/sugar transporter family protein	not specific for NFN clade
Medtr3g112060.1	mannan endo-1,4-beta-mannosidase-like protein	low phylogenetic signal
Medtr3g114310.1	serpin-like protein	low phylogenetic signal
Medtr4g093080.1	receptor lectin kinase	not specific for NFN clade
Medtr4g093140.1	concanavalin A-like lectin kinase family protein, putative	not specific for NFN clade
Medtr4g093950.1	hypothetical protein	not specific for NFN clade
Medtr4g094472.1	transmembrane protein, putative	not specific for NFN clade
Medtr4g109550.1	presenilin plant-like protein	not specific for NFN clade
Medtr4g126160.1	cytokinin oxidase/dehydrogenase-like protein	not specific for NFN clade
Medtr5g044170.1	fasciclin-like arabinogalactan protein	not specific for NFN clade
Medtr5g083330.1	dehydration-responsive element-binding protein	not specific for NFN clade
Medtr5g459450.1	phosphatidylinositol-4-phosphate 5-kinase family protein	not specific for NFN clade
Medtr6g027720.1	LRR receptor-like kinase	not specific for NFN clade
Medtr6g090500.2	DNA-directed RNA polymerase I, II	not specific for NFN clade
Medtr7g015770.1	disease resistance protein (CC-NBS-LRR class) family protein	not specific for NFN clade
Medtr7g016060.1	NADH-ubiquinone reductase complex 1 MLRQ subunit	low phylogenetic signal
Medtr7g086420.1	receptor-like kinase	not specific for NFN clade
Medtr8g013950.1	GATA transcription factor-like protein	not specific for NFN clade
Medtr8g102250.1	dual specificity phosphatase domain protein	not specific for NFN clade
Medtr8g445730.1	transmembrane protein, putative	not specific for NFN clade
Medtr8g466030.1	quinone oxidoreductase	not specific for NFN clade

Table S30. Candidate genes identified by the Orthofinder pipeline for the predisposition before curation. Genes identified by the Orthofinder pipeline using relaxed selection criteria are presented. The ones highlighted in red have been excluded because they are not specific to the NFN clade. Genes highlighted in yellow do not allow reconstructing phylogenies with robust support.

cluster name	SWISSPROT human readable description	number of hypothetical gain nodes with expansions	number of nodes with expansions inside NFN clade	number of nodes with expansions outside NFN clade	number of nodes with expansions
OG0001284	BON1-associated protein 2	1	16	6	22
OG000245	Protein PELPK2	3	17	10	27
OG000283	NADH dehydrogenase [ubiquinone] iron-sulfur protein 1	3	16	10	26
OG000334	Probable alpha,alpha-trehalose-phosphate acyltransferase	3	20	8	28
OG000752	Cullin-1	3	21	8	29
OG001084	Plastid division protein PDV2, Leghemoglobin	3	16	6	22
OG001398	3-hydroxy-3-methylglutaryl-coenzyme A 1	3	13	5	18
OG002006	Repetitive proline-rich cell wall protein 2	3	18	4	22
OG002759	Senescence-associated carboxylesterase	3	16	4	20
OG000064	Probable glutathione S-transferase	4	28	10	38
OG000120	Cytochrome P450 81E8	4	27	11	38
OG000131	Protein SRG1	4	24	10	34
OG000176	Probable pectate lyase 16	4	27	12	39
OG000209	WAT1-related protein	4	25	9	34
OG000268	(+)-neomenthol dehydrogenase SDRI	4	25	11	36
OG000277	Calmodulin-binding protein 60 B	4	24	7	31
OG000376	Heavy metal-associated isoprenylated protein	4	23	7	30
OG000501	Dirigent Protein 19? (only 3 hits)	4	19	7	26
OG000793	Beta-glucosidase 11	4	17	8	25
OG001126	Protein DOWNY MILDEW RESISTANCE	4	23	4	27
OG000067	Beta-glucosidase 12	5	28	9	37
OG000077	Thaumatin-like protein 1	5	29	13	42
OG000135	Cytochrome P450 CYP72A219	5	24	8	32
OG000148	Pathogenesis-related protein STH-2	5	24	10	34
OG000159	Cellulose synthase A catalytic subunit 3	5	25	10	35
OG000223	Fasciclin-like arabinogalactan protein 11	5	25	10	35
OG000243	UDP-glycosyltransferase 91A1	5	21	10	31
OG000280	Phospholipid-transporting ATPase 10	5	20	7	27
OG000300	Chlorophyll a-b binding protein 40, chlorophyll a/b binding protein 40	5	21	9	30
OG000354	Protein EXORDIUM	5	22	8	30
OG000408	E3 ubiquitin-protein ligase RGLG2	5	19	6	25
OG000540	Leucine-rich repeat receptor-like serine/threonine kinase 1	5	22	8	30
OG000041	Laccase-17	6	29	10	39
OG000066	1-amino cyclopropane-1-carboxylate oxidase	6	32	9	41
OG000089	MLP-like protein 43	6	30	11	41
OG000126	Geraniol 8-hydroxylase	6	24	10	34
OG000164	Dirigent protein 19	6	25	10	35
OG000221	Ethylene-responsive transcription factor 1	6	24	10	34
OG000225	Cytokinin riboside 5'-monophosphate phosphotransferase	6	25	10	35
OG000247	Basic 7S globulin	6	21	11	32
OG000261	Serine/threonine-protein kinase RIPK	6	27	11	38
OG000271	Alcohol dehydrogenase-like 6	6	22	5	27
OG000297	Chalcone synthase	6	24	7	31
OG000423	Arabinogalactan peptide 23	6	29	11	40
OG000655	Protein NRT1/ PTR FAMILY 1.2	6	20	7	27
OG000061	Expansin-A4	7	35	12	47
OG000110	Isoliquiritigenin 2'-O-methyltransferase	7	27	8	35
OG000216	21 kDa protein, Pectinesterase inhibitor	7	23	12	35
OG000218	3,9-dihydroxypterocarpan 6A-monoxygenase	7	25	8	33
OG000235	Tubulin beta-1 chain	7	22	9	31
OG000548	Extensin-3	7	22	7	29
OG000174	Potassium transporter 6	8	27	8	35
OG000075	Beta-galactosidase 3	9	30	11	41

Table S31. Occurrence of clusters enriched with nodule-regulated *Medicago truncatula* genes in nodes among the dataset. The table shows at how many nodes each enriched cluster expanded. Distinguished are nodes outside the NFN clade and inside the NFN clade. In addition, it is shown at how many of such nodes each cluster expanded, at which an independent gain of the root nodule symbiosis was postulated (Figure 2, blue boxes).

MedicagoID	Medicago description	Confirmation
Medtr00480130.1	hypothetical protein	low phylogenetic signal
Medtr04958020.1	cytochrome C oxidase subunit 3	low phylogenetic signal
Medtr02989001.1	trichome brefingheme-like protein, putative	low phylogenetic signal
Medtr10569001.1	hypothetical protein	low phylogenetic signal
Medtr10706240.1	app-1P processing enzyme family protein	all non-nodulators present
Medtr10711880.1	cytochrome b-c1 complex subunit 7	all non-nodulators present
Medtr10728220.1	topless-like protein	nodulators missing
Medtr10768990.1	hypothetical protein	all non-nodulators present
Medtr10769440.1	alpha amylose domain protein	nodulators missing
Medtr10782110.1	VAC14-like protein	all non-nodulators present
Medtr10782730.1	trehalose-6-phosphate synthase domain protein	all non-nodulators present
Medtr10785000.1	transmembrane amino acid transporter family protein	all non-nodulators present
Medtr10785340.1	ubiquitin tñmer protein	nodulators missing
Medtr10789150.2	HIT zinc finger protein	nodulators missing
Medtr10892235.1	photosystem I reaction center subunit IV A	all non-nodulators present
Medtr10774460.1	plant/F9H34-like protein	more than 50% non-nodulators present (9 of 13)
Medtr1078400.1	eukaryotic release factor 1 (eRF1) family protein	all non-nodulators present
Medtr1086070.1	leucogalactin N epimerase-like protein	nodulators missing
Medtr1090807.1	myosin heavy chain-like protein, putative (RPG)	nodulators missing
Medtr1090817.1	O-sialoglycoprotein endopeptidase	all non-nodulators present
Medtr1090825.1	hypothetical protein	nodulators missing
Medtr11122020.1	transportin-1 protein	all non-nodulators present
Medtr10971120.1	ZOC-F6-like oxygenase family oxidoreductase	nodulators missing
Medtr1098770.1	pumilio family RNA-binding repeat protein	nodulators missing
Medtr10971930.1	ankyrin repeat protein	low phylogenetic signal
Medtr10972000.1	methionine sulfoxide reductase B	low phylogenetic signal
Medtr10972090.3	methionine sulfoxide reductase B 2	all non-nodulators present
Medtr109736460.1	receptor-like kinase	nodulators missing
Medtr109738230.1	hypothetical protein	more than 50% non-nodulators present (11 of 13)
Medtr109740500.1	protein phosphatase 2C AB1-like protein	more than 50% non-nodulators present (10 of 13)
Medtr109759490.1	eukaryotic aspartyl protease family protein	low phylogenetic signal
Medtr109762220.1	oxidoreductase/transition metal ion-binding protein	more than 50% non-nodulators present (9 of 13)
Medtr109767650.1	epidemiologically specific secreted EP1-like glycoprotein	all non-nodulators present
Medtr109778310.1	PHD finger and bromo-adjacent-like domain protein	all non-nodulators present
Medtr109798430.1	UDP-glucosyltransferase family protein	more than 50% non-nodulators present (10 of 13)
Medtr109799790.3	calmodulin-binding family protein	nodulators missing
Medtr109838150.1	potassium transporter 12	all non-nodulators present
Medtr109835130.1	long-chain acyl-CoA synthetase	low phylogenetic signal
Medtr109837570.1	peptidylprolyl cis/trans isomerase, NIMA-interacting protein	all non-nodulators present
Medtr109854220.1	co-chaperone GrpE family protein	all non-nodulators present
Medtr109869000.1	zinc-finger metalloprotease family protein	all non-nodulators present
Medtr10986230.1	photosystem II core complex family psbY protein	low phylogenetic signal
Medtr109868640.1	hypothetical protein	low phylogenetic signal
Medtr1096310.1	GDP-mannose transporter GONST1	all non-nodulators present
Medtr10912290.1	BEL1-like homeodomain protein	all non-nodulators present
Medtr109116080.1	papain family cysteine protease	low phylogenetic signal
Medtr10935430.1	expansin A10	all non-nodulators present
Medtr109409410.1	hypothetical protein	nodulators missing
Medtr109414630.1	gamma interferon inducible lysosomal thiol reductase	nodulators missing
Medtr109430910.1	ARM repeat kinase family protein	all non-nodulators present
Medtr10935955.1	glycosyl hydrolase family 9 protein	nodulators missing
Medtr10945903.1	indole-3-acetyl acid-amido synthetase	all non-nodulators present
Medtr10957270.1	transcription factor	all non-nodulators present
Medtr10961360.3	PRR response regulator	all non-nodulators present
Medtr10972060.1	NADP-dependent D-orbitol-6-phosphate dehydrogenase	all non-nodulators present
Medtr10975590.1	major intrinsic protein (MIP) family transporter	all non-nodulators present
Medtr10982330.1	transmembrane protein, putative	nodulators missing
Medtr1098245.1	AP2/ERF domain transcription factor	nodulators missing
Medtr10984550.1	wuschel-related homeobox protein	nodulators missing
Medtr10984710.1	serine carboxypeptidase-like protein	all non-nodulators present
Medtr10993080.1	receptor lectin kinase	low phylogenetic signal
Medtr10993140.1	concanavalin A-like lectin kinase family protein, putative	low phylogenetic signal
Medtr10993950.1	hypothetical protein	low phylogenetic signal
Medtr1094202.1	ICE-like protease (caspase) p20 domain protein	nodulators missing
Medtr109494908.1	WRKY family transcription factor	all non-nodulators present
Medtr10912220.1	DUF1442 family protein	low phylogenetic signal
Medtr109176940.1	hypothetical protein	low phylogenetic signal
Medtr10912290.1	plant/T25P12-18 protein	nodulators missing
Medtr109127530.1	DUF4228 domain protein	all non-nodulators present
Medtr109485940.1	UDP-glucosyltransferase 73B2	nodulators missing
Medtr10906970.1	alba DNA/RNA-binding protein	low phylogenetic signal
Medtr10908150.1	eukaryotic aspartyl protease family protein	all non-nodulators present
Medtr10909930.1	SGT1-like protein	all non-nodulators present
Medtr10911980.2	Lipid transfer protein	low phylogenetic signal
Medtr10926380.1	hypothetical protein	nodulators missing
Medtr10927210.1	UDP-glucosyltransferase family protein	low phylogenetic signal
Medtr10932880.1	1-aminocyclopropane-1-carboxylate oxidase	nodulators missing
Medtr10937230.1	short hypocotyl in white light1 protein	nodulators missing
Medtr109474700.1	cationic peroxidase	all non-nodulators present
Medtr109574860.1	peroxisome family protein	all non-nodulators present
Medtr109575260.1	nucleodiytransferase	all non-nodulators present
Medtr109580750.1	hypothetical protein	nodulators missing
Medtr10981230.1	serine/threonine-protein phosphatase PP1	all non-nodulators present
Medtr10987550.1	3-hydroxy-3-methylglutaryl-coenzyme A reductase-like protein	low phylogenetic signal
Medtr10995990.1	meiotic recombination SPO11-like protein	all non-nodulators present
Medtr10996020.1	ribosomal RNA small subunit methyltransferase B	more than 50% non-nodulators present (11 of 13)
Medtr10971780.1	serine carboxypeptidase-like protein	all non-nodulators present
Medtr1099360.1	reduce inspection protein (NIN)	accepted
Medtr10945257.1	heavy metal-associated domain protein, putative	all non-nodulators present
Medtr10964590.1	hypothetical protein	low phylogenetic signal
Medtr109606260.1	phosphoenopyruvate carboxylase-like protein	all non-nodulators present
Medtr109715770.1	disease resistance protein (CC-NBS-LRR class) family protein	nodulators missing
Medtr109716060.1	NADH-ubiquinone reductase complex 1 MLRQ subunit	all non-nodulators present
Medtr109768190.1	receptor-like kinase	all non-nodulators present
Medtr109777870.1	hypothetical protein	all non-nodulators present
Medtr109822660.1	plastid movement impaired-like protein	all non-nodulators present
Medtr109822730.1	hypothetical protein	all non-nodulators present
Medtr109822760.1	hypothetical protein	all non-nodulators present
Medtr109822760.1	C2 domain protein	nodulators missing
Medtr10983010.1	strictosidine synthase family protein	more than 50% non-nodulators present (11 of 13)
Medtr10983370.1	peroxidase family protein	more than 50% non-nodulators present (9 of 13)
Medtr10984500.1	receptor Serine/Threonine kinase	all non-nodulators present
Medtr109842450.1	UDP-glucosyltransferase family protein	low phylogenetic signal
Medtr109810330.1	F-box protein PP2-A13	all non-nodulators present
Medtr1098113670.1	hypothetical protein	nodulators missing
Medtr109809830.1	NADP-dependent alkenal double bond reductase	low phylogenetic signal
Medtr109810330.1	NADP-dependent alkenal double bond reductase	low phylogenetic signal
Medtr109827285.1	ORM1 family protein	all non-nodulators present
Medtr109827765.1	riboflavose-5'-phosphate phosphatase	nodulators missing
Medtr109831010.1	hypothetical protein	all non-nodulators present
Medtr109832040.1	NADP-dependent malic enzyme	all non-nodulators present
Medtr10984590.2	nucleic acid-binding, CB-fold-like protein	low phylogenetic signal
Medtr109848440.1	protein disulfide isomerase (PDI)-like protein	more than 50% non-nodulators present (10 of 13)
Medtr109877000.1	RNA transcription initiation factor III, putative	all non-nodulators present
Medtr109887750.1	DUF793 family protein	nodulators missing
Medtr109898757.1	enoyl (acyl carrier) reductase	nodulators missing
Medtr109898445.1	WRKY family transcription factor	low phylogenetic signal
Medtr109899000.1	RNA nucleotidylyltransferase/poly(A) polymerase	all non-nodulators present
Medtr109866030.1	quinone oxidoreductase	low phylogenetic signal
Medtr10986240.1	UDP-glucosyltransferase family protein	nodulators missing
Medtr109871160.1	3-hydroxyacyl-CoA dehydratase	all non-nodulators present

Table S32. Candidate genes identified by the Orthofinder pipeline for the multiple losses before curation. Genes identified by the Orthofinder pipeline using relaxed selection criteria are presented.

gene_name	organism	ID	source	publication_DOI	function
ANN1	<i>M. truncatula</i>	CAA75308	Genbank	10.1094/MPMI.1998.11.6.504	infection
ARPC1	<i>L. japonicus</i>	JX446368	Genbank	10.1104/pp.112.202572	infection
ASTRAY	<i>L. japonicus</i>	BAC20318	Genbank	10.1073/pnas.222302699	nodulation regulation
CCaMK	<i>L. japonicus</i>	CAJ76699	Genbank	10.1038/nature04862	common symbiosis gene
CRE	<i>L. japonicus</i>	ABI48271	Genbank	10.1126/science.1132514	nodule organogenesis
CYCLOPS	<i>L. japonicus</i>	ABU63668	Genbank	10.1073/pnas.0806858105	common symbiosis gene
EPR3	<i>L. japonicus</i>	BAI79284	Genbank	10.1038/nature14611	infection control
LIN	<i>M. truncatula</i>	EU926660	Genbank	10.1104/pp.109.143933	infection
NAP1	<i>M. truncatula</i>	ADM22319	Genbank	10.1094/MPMI-06-10-0144	infection
NF-YA1	<i>M. truncatula</i>	EF488826	Genbank	10.1101/gad.461808	nodule organogenesis
NF-YA2	<i>M. truncatula</i>	KEH24257	Genbank	10.1111/tpj.12587	nodule organogenesis
NIN	<i>L. japonicus</i>	CAB61338	Genbank	10.1038/46058	infection, nodule organogenesis, nodulation regulation
NSP1	<i>M. truncatula</i>	CAJ00010	Genbank	10.1126/science.1111025	nodulation regulation
NSP2	<i>M. truncatula</i>	CAH55768	Genbank	10.1126/science.1110951	nodulation regulation
PIR1	<i>L. japonicus</i>	CAQ17049	Genbank	10.1105/tpc.108.063693	infection
PUB1	<i>M. truncatula</i>	DAA33939	Genbank	10.1105/tpc.110.075861	infection
RDN1	<i>M. truncatula</i>	ADV35716	Genbank	10.1104/pp.111.178756	nodulation regulation
RPG	<i>M. truncatula</i>	ABI51615	Genbank	10.1073/pnas.0710273105	infection
SUNERGOS1	<i>L. japonicus</i>	AID62216	Genbank	10.1111/tpj.12520	nodule regulation
SUNN	<i>M. truncatula</i>	AAW71475	Genbank	10.1007/s11103-005-8102-y	nodulation regulation
SYMRK	<i>L. japonicus</i>	AAM67418	Genbank	10.1038/nature00841	common symbiosis gene
VAG1	<i>L. japonicus</i>	BAO79526	Genbank	10.1242/dev.107946	nodule regulation

Table S33. List of candidate genes explored in targeted phylogenetic approach.

		<i>NIN</i>			<i>RPG</i>		
Model A		lnL	LRT	p-value (df = 1)	lnL	LRT	p-value (df = 1)
Null		-11179.401054	1.463358	0.226397	-8172.069458	3.529036	0.060303
Alternative		-11178.669375			-8170.304940		

Table S34. Selection pressure on *NIN* and *RPG*. Results of the model A “null” and “alternative” hypotheses confronted. *P-values* were calculated by likelihood ratio test.

Accession code	SRA code	Library ID	Insert size (bp)	sample name	species
SAMN06393017	SRA 538085	NITgmnDAEADAAPEI-87	170	Aln_g1_1	<i>Alnus glutinosa</i>
SAMN06393018		NITgmnDAHDAAPEI-95	170	Aln_g1_2	<i>Alnus glutinosa</i>
SAMN06393019		NITgmnDAFDAAPEI-88	250	Aln_g1_3	<i>Alnus glutinosa</i>
SAMN06393020		NITgmnDAIDAAPEI-96	250	Aln_g1_4	<i>Alnus glutinosa</i>
SAMN06393021		NITgmnDAAFDAAPEI-89	350	Aln_g1_5	<i>Alnus glutinosa</i>
SAMN06393022		NITgmnDABDWAAPEI-91	2000	Aln_g1_6	<i>Alnus glutinosa</i>
SAMN06393023		NITgmnDABDLAAPEI-92	6000	Aln_g1_7	<i>Alnus glutinosa</i>
SAMN06393024		NITgmnDABDUAAPEI-93	10000	Aln_g1_8	<i>Alnus glutinosa</i>
SAMN06393025		NITgmnDABDUAAPEI-94	20000	Aln_g1_9	<i>Alnus glutinosa</i>
SAMN06393026		RSZABPI007591-140	170	Beg_fu_1	<i>Begonia fuchsoides</i>
SAMN06393027		RSZAXPI007685-29	250	Beg_fu_2	<i>Begonia fuchsoides</i>
SAMN06393028		NITgmnDAAACDAAPEI-85	250	Beg_fu_3	<i>Begonia fuchsoides</i>
SAMN06393029		RSZAXPI007684-30	350	Beg_fu_4	<i>Begonia fuchsoides</i>
SAMN06393030		NITgmnDAADDAAAPEI-86	350	Beg_fu_5	<i>Begonia fuchsoides</i>
SAMN06393031		RSZABPI007455-37	500	Beg_fu_6	<i>Begonia fuchsoides</i>
SAMN06393032		RSZAMP007456-51	800	Beg_fu_7	<i>Begonia fuchsoides</i>
SAMN06393033		NITgmnDAADWAAPEI-90	2000	Beg_fu_8	<i>Begonia fuchsoides</i>
SAMN06393034		NITgmnDAJDLAAPEI-149	6000	Beg_fu_9	<i>Begonia fuchsoides</i>
SAMN06393035	SRA 540436	RSZABP006199-23	170	Cas_g1_1	<i>Casuarina glauca</i>
SAMN06393036		RSZAXPI006202-25	250	Cas_g1_2	<i>Casuarina glauca</i>
SAMN06393037		RSZAXPI006205-24	350	Cas_g1_3	<i>Casuarina glauca</i>
SAMN06393038		RSZAXPI006207-35	500	Cas_g1_4	<i>Casuarina glauca</i>
SAMN06393039		RSZAMP006210-39	800	Cas_g1_5	<i>Casuarina glauca</i>
SAMN06393040		NITgmnDAGDWAAPEI-43	2000	Cas_g1_6	<i>Casuarina glauca</i>
SAMN06393041		NITgmnDADCLAAPEI-122	6000	Cas_g1_7	<i>Casuarina glauca</i>
SAMN06393042		NITgmnDACTDAAPEI-45	10000	Cas_g1_8	<i>Casuarina glauca</i>
SAMN06393043		NITgmnDAGDUAAPEI-84	20000	Cas_g1_9	<i>Casuarina glauca</i>
SAMN06393044		RSZABP0066907-98	170	Cer_ca_1	<i>Cercis canadensis</i>
SAMN06393045		RSZABP007590-130	170	Cer_ca_2	<i>Cercis canadensis</i>
SAMN06393046		RSZAXPI006909-104	250	Cer_ca_3	<i>Cercis canadensis</i>
SAMN06393047		RSZAXPI007589-131	250	Cer_ca_4	<i>Cercis canadensis</i>
SAMN06393048		RSZAXPI006908-108	350	Cer_ca_5	<i>Cercis canadensis</i>
SAMN06393049		RSZAXPI007588-132	350	Cer_ca_6	<i>Cercis canadensis</i>
SAMN06393050		RSZAXPI006799-66	500	Cer_ca_7	<i>Cercis canadensis</i>
SAMN06393051		RSZAMP006795-59	800	Cer_ca_8	<i>Cercis canadensis</i>
SAMN06393052		NITgmnDABDWAAPEI-114	2000	Cer_ca_9	<i>Cercis canadensis</i>
SAMN06393053		NITgmnDAKDTAAPEI-160	10000	Cer_ca_10	<i>Cercis canadensis</i>
SAMN06393054	SRA 540439	RSZABP006911-101	170	Cha_fa_1	<i>Chamaecrista fasciata</i>
SAMN06393055		RSZABP007595-142	170	Cha_fa_2	<i>Chamaecrista fasciata</i>
SAMN06393056		RSZAXPI006910-106	250	Cha_fa_3	<i>Chamaecrista fasciata</i>
SAMN06393057		RSZAXPI006912-109	350	Cha_fa_4	<i>Chamaecrista fasciata</i>
SAMN06393058		RSZAXPI007596-143	350	Cha_fa_5	<i>Chamaecrista fasciata</i>
SAMN06393059		RSZAXPI006796-67	500	Cha_fa_6	<i>Chamaecrista fasciata</i>
SAMN06393060		RSZAXPI007482-15	500	Cha_fa_7	<i>Chamaecrista fasciata</i>
SAMN06393061		RSZAMP006796-64	800	Cha_fa_8	<i>Chamaecrista fasciata</i>
SAMN06393062		RSZAMP007481-17	800	Cha_fa_9	<i>Chamaecrista fasciata</i>
SAMN06393063		NITgmnDADFWAAPEI-141	2000	Cha_fa_10	<i>Chamaecrista fasciata</i>
SAMN06393064		NITgmnDALDWAAPEI-34	2000	Cha_fa_11	<i>Chamaecrista fasciata</i>
SAMN06393065		NITgmnDADFLAAPEI-148	6000	Cha_fa_12	<i>Chamaecrista fasciata</i>
SAMN06393066		NITgmnDADDTAAPEI-13	10000	Cha_fa_13	<i>Chamaecrista fasciata</i>
SAMN06393067		RSZABP0067089-39	170	Dat_gl_1	<i>Datiscia glomerata</i>
SAMN06393068		RSZAXPI007090-40	250	Dat_gl_2	<i>Datiscia glomerata</i>
SAMN06393069	SRA 540443	RSZAXPI007086-37	350	Dat_gl_3	<i>Datiscia glomerata</i>
SAMN06393070		RSZAXPI007079-17	500	Dat_gl_4	<i>Datiscia glomerata</i>
SAMN06393071		RSZAMP007082-18	800	Dat_gl_5	<i>Datiscia glomerata</i>
SAMN06393072		NITgmnDABDWAAPEI-142	2000	Dat_gl_6	<i>Datiscia glomerata</i>
SAMN06393073		NITgmnDABDLAAPEI-150	6000	Dat_gl_7	<i>Datiscia glomerata</i>
SAMN06393074		NITgmnDABDTAAPEI-161	10000	Dat_gl_8	<i>Datiscia glomerata</i>
SAMN06393075		RSZABP006200-27	170	Dis_tr_1	<i>Discaria trinervis</i>
SAMN06393076		RSZABP007597-147	170	Dis_tr_2	<i>Discaria trinervis</i>
SAMN06393077		RSZAXPI006203-31	250	Dis_tr_3	<i>Discaria trinervis</i>
SAMN06393078		RSZAXPI006206-32	350	Dis_tr_4	<i>Discaria trinervis</i>
SAMN06393079		RSZAXPI006208-40	500	Dis_tr_5	<i>Discaria trinervis</i>
SAMN06393080		RSZAMP006211-41	800	Dis_tr_6	<i>Discaria trinervis</i>
SAMN06393081		NITgmnDADCWAAPEI-34	2000	Dis_tr_7	<i>Discaria trinervis</i>
SAMN06393082	SRA 540444	RSZABP007087-43	170	Dry_dr_1	<i>Dryas drummondii</i>
SAMN06393083		RSZAXPI007083-41	250	Dry_dr_2	<i>Dryas drummondii</i>
SAMN06393084		RSZAXPI007084-42	350	Dry_dr_3	<i>Dryas drummondii</i>
SAMN06393085		RSZAXPI007093-20	500	Dry_dr_4	<i>Dryas drummondii</i>
SAMN06393086		RSZAMP00707801-21	800	Dry_dr_5	<i>Dryas drummondii</i>
SAMN06393087		NITgmnDAGDWAAPEI-143	2000	Dry_dr_6	<i>Dryas drummondii</i>
SAMN06393088		NITgmnDAGDLAAPEI-152	6000	Dry_dr_7	<i>Dryas drummondii</i>
SAMN06393089		NITgmnDACDTAAPEI-13	10000	Dry_dr_8	<i>Dryas drummondii</i>
SAMN06393109		RSZABP006913-102	170	Mim_pu_1	<i>Mimosa pudica</i>
SAMN06393110		RSZABP007594-133	170	Mim_pu_2	<i>Mimosa pudica</i>
SAMN06393111		RSZAXPI006915-107	250	Mim_pu_3	<i>Mimosa pudica</i>
SAMN06393112		RSZAXPI007593-135	250	Mim_pu_4	<i>Mimosa pudica</i>
SAMN06393113		RSZAXPI006914-113	350	Mim_pu_5	<i>Mimosa pudica</i>
SAMN06393114		RSZAXPI007592-139	350	Mim_pu_6	<i>Mimosa pudica</i>
SAMN06393115		RSZAXPI006797-68	500	Mim_pu_7	<i>Mimosa pudica</i>
SAMN06393116	SRA 540445	RSZAMP006794-65	800	Mim_pu_8	<i>Mimosa pudica</i>
SAMN06393117		NITgmnDADWAAPEI-146	2000	Mim_pu_9	<i>Mimosa pudica</i>
SAMN06393118		NITgmnDAAFLAAPEI-147	6000	Mim_pu_10	<i>Mimosa pudica</i>
SAMN06393119		NITgmnDAEDETAAPEI-14	10000	Mim_pu_11	<i>Mimosa pudica</i>
SAMN06393120		RSZABP006198-21	170	Nis_sc_1	<i>Nissolia schottii</i>
SAMN06393121		RSZAXPI006201-20	250	Nis_sc_2	<i>Nissolia schottii</i>
SAMN06393122		RSZAXPI006204-22	350	Nis_sc_3	<i>Nissolia schottii</i>
SAMN06393123		RSZAXPI006209-33	500	Nis_sc_4	<i>Nissolia schottii</i>
SAMN06393124		RSZAMP006212-34	800	Nis_sc_5	<i>Nissolia schottii</i>
SAMN06393125		NITgmnDADDWAAPEI-43	2000	Nis_sc_6	<i>Nissolia schottii</i>
SAMN06393126		NITgmnDAEFLAAPEI-67	6000	Nis_sc_7	<i>Nissolia schottii</i>
SAMN06458495	SRA 540452	RALDgEAAAARAAPEI-72	170	Alnus glutinosa	<i>Alnus glutinosa</i>
SAMN06458496	SRA 540453	RCEAkmEADAANRAAPEI-28	200	Casuarina glauca	<i>Casuarina glauca</i>
SAMN06458497	SRA 540454	RNTYkvTBAACRAAPEI-115	200	Nissolia schottii	<i>Nissolia schottii</i>
SAMN06458498	SRA 540455	RBEGilTAAAFRAAPEI-141	170	Begonia fuchsoides_1	<i>Begonia fuchsoides</i>
SAMN06458499	SRA 540456	RBEGilTAAAHRAAPEI-146	170	Begonia fuchsoides_2	<i>Begonia fuchsoides</i>
SAMN06458500	SRA 540456	RDTAgdEAAAARAAPEI-77	170	Datiscia glomerata_1	<i>Datiscia glomerata</i>
SAMN06458501	SRA 540456	RDTAgdEAAAARFAAPEI-90	170	Datiscia glomerata_2	<i>Datiscia glomerata</i>
SAMN06458502	SRA 540457	RDTAgdEAAAARAAPEI-95	170	Datiscia glomerata_3	<i>Datiscia glomerata</i>
SAMN06458503	SRA 540457	DRDYacnEAAAARAAPEI-12	170	Dryas drummondii_1	<i>Dryas drummondii</i>
SAMN06458504	SRA 540457	DRDYacnEAAAACRAAPEI-14	170	Dryas drummondii_2	<i>Dryas drummondii</i>
SAMN06458505	SRA 540460	RCASnzaEAAJRAAPEI-69	170	Discaria trinervis	<i>Discaria trinervis</i>

Species	Library									
	170bp	250bp	350bp	500bp	800bp	2kb	6kb	10kb	20kb	
<i>Alnus glutinosa</i>	2	2	1			1	1	1	1	
<i>Casuarina glauca</i>	1	1	1	1	1	1	1	1	1	
<i>Cercis canadensis</i>	2	2	2	1	1	1				
<i>Chamaecrista fasciculata</i>	2	1	2	2	2	2	1	1		
<i>Mimosa pudica</i>	2	2	2	1	1	1	1	1		
<i>Nissolia schottii</i>	1	1	1	1	1	1	1			
<i>Begonia fuchsoides</i>	1	2	2	1	1	1	1			
<i>Datisca glomerata</i>	1	1	1	1	1	1	1	1		
<i>Dryas drummondii</i>	1	1	1	1	1	1	1	1		
<i>Discaria trinervis</i>	2	1	1	1	1	1				

Table S36. List of prepared libraries for DNA sequencing

Species Name	Tissues	# Sample
<i>Alnus glutinosa</i>	Nodule, Root, Leaf	3
<i>Casuarina glauca</i>	Leaf	1
<i>Nissolia schottii</i>	Root, leaves	2
<i>Begonia fuchsioides</i>	Root, Leaf	2
<i>Datisca glomerata</i>	Nodule, Root, Leaf	3
<i>Dryas drummondii</i>	Seedling, Leaf	2
<i>Discaria trinervis</i>	Leaf	1

Table S37. List of RNA samples sequenced in this study.

Data S1. Sequences used for the phylogenies presented in Fig. S4, S6 – S25.

Data S2. List of the plant species used for DNA and RNA extraction in this study. Repository and accession numbers are indicated.

References and Notes

1. P. C. Dos Santos, Z. Fang, S. W. Mason, J. C. Setubal, R. Dixon, Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* **13**, 162 (2012). [doi:10.1186/1471-2164-13-162](https://doi.org/10.1186/1471-2164-13-162) [Medline](#)
2. P. H. Beatty, A. G. Good, Future prospects for cereals that fix nitrogen. *Science* **333**, 416–417 (2011). [doi:10.1126/science.1209467](https://doi.org/10.1126/science.1209467) [Medline](#)
3. L. Hiltner, Über die Bedeutung der Wurzelknöllchen von *Alnus glutinosa* für die Stickstoffernährung dieser Pflanze. *Landw. Versuchsstat* **46**, 153–161 (1895).
4. H. Hellriegel, H. Wilfarth, *Untersuchungen über die Stickstoffnahrung der Gramineen und Leguminosen* (Buchdruckerei der Post Kayssler, Berlin, 1888).
5. D. E. Soltis, P. S. Soltis, D. R. Morgan, S. M. Swensen, B. C. Mullin, J. M. Dowd, P. G. Martin, Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2647–2651 (1995). [doi:10.1073/pnas.92.7.2647](https://doi.org/10.1073/pnas.92.7.2647) [Medline](#)
6. C. Kistner, M. Parniske, Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* **7**, 511–518 (2002). [doi:10.1016/S1360-1385\(02\)02356-7](https://doi.org/10.1016/S1360-1385(02)02356-7) [Medline](#)
7. J. J. Doyle, Phylogenetic perspectives on the origins of nodulation. *Mol. Plant Microbe Interact.* **24**, 1289–1295 (2011). [doi:10.1094/MPMI-05-11-0114](https://doi.org/10.1094/MPMI-05-11-0114) [Medline](#)
8. J. I. Sprent, J. Ardley, E. K. James, Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytol.* **215**, 40–56 (2017). [doi:10.1111/nph.14474](https://doi.org/10.1111/nph.14474) [Medline](#)
9. K. Pawłowski, K. N. Demchenko, The diversity of actinorhizal symbiosis. *Protoplasma* **249**, 967–979 (2012). [doi:10.1007/s00709-012-0388-4](https://doi.org/10.1007/s00709-012-0388-4) [Medline](#)
10. S. M. Swensen, The evolution of actinorhizal symbioses: Evidence for multiple origins of the symbiotic association. *Am. J. Bot.* **83**, 1503–1512 (1996). [doi:10.1002/j.1537-2197.1996.tb13943.x](https://doi.org/10.1002/j.1537-2197.1996.tb13943.x)
11. G. D. A. Werner, W. K. Cornwell, J. I. Sprent, J. Kattge, E. T. Kiers, A single evolutionary innovation drives the deep evolution of symbiotic N₂-fixation in angiosperms. *Nat. Commun.* **5**, 4087 (2014). [doi:10.1038/ncomms5087](https://doi.org/10.1038/ncomms5087) [Medline](#)
12. S. M. Swensen, D. R. Benson, in *Nitrogen-fixing Actinorhizal Symbioses*, K. Pawłowski, W. E. Newton, Eds. (Springer Netherlands, 2008), pp. 73–104.
13. H. L. Li, W. Wang, P. E. Mortimer, R.-Q. Li, D.-Z. Li, K. D. Hyde, J.-C. Xu, D. E. Soltis, Z.-D. Chen, Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Sci. Rep.* **5**, 14023 (2015). [doi:10.1038/srep14023](https://doi.org/10.1038/srep14023) [Medline](#)
14. J. A. Eisen, C. M. Fraser, Phylogenomics: Intersection of evolution and genomics. *Science* **300**, 1706–1707 (2003). [doi:10.1126/science.1086292](https://doi.org/10.1126/science.1086292) [Medline](#)
15. P. M. Delaux, G. Radhakrishnan, G. Oldroyd, Tracing the evolutionary path to nitrogen-fixing crops. *Curr. Opin. Plant Biol.* **26**, 95–99 (2015). [doi:10.1016/j.pbi.2015.06.003](https://doi.org/10.1016/j.pbi.2015.06.003) [Medline](#)

16. Materials and methods are available as supplementary materials.
17. N. D. Young, F. Debelle, G. E. D. Oldroyd, R. Geurts, S. B. Cannon, M. K. Udvardi, V. A. Benedito, K. F. X. Mayer, J. Gouzy, H. Schoof, Y. Van de Peer, S. Proost, D. R. Cook, B. C. Meyers, M. Spannagl, F. Cheung, S. De Mita, V. Krishnakumar, H. Gundlach, S. Zhou, J. Mudge, A. K. Bharti, J. D. Murray, M. A. Naoumkina, B. Rosen, K. A. T. Silverstein, H. Tang, S. Rombauts, P. X. Zhao, P. Zhou, V. Barbe, P. Bardou, M. Bechner, A. Belloc, A. Berger, H. Bergès, S. Bidwell, T. Bisseling, N. Choisne, A. Couloux, R. Denny, S. Deshpande, X. Dai, J. J. Doyle, A.-M. Dukez, A. D. Farmer, S. Fouteau, C. Franken, C. Gibelin, J. Gish, S. Goldstein, A. J. González, P. J. Green, A. Hallab, M. Hartog, A. Hua, S. J. Humphray, D.-H. Jeong, Y. Jing, A. Jöcker, S. M. Kenton, D.-J. Kim, K. Klee, H. Lai, C. Lang, S. Lin, S. L. Macmil, G. Magdelenat, L. Matthews, J. McCrosson, E. L. Monaghan, J.-H. Mun, F. Z. Najar, C. Nicholson, C. Noirot, M. O'Bleness, C. R. Paule, J. Poulain, F. Prion, B. Qin, C. Qu, E. F. Retzel, C. Riddle, E. Sallet, S. Samain, N. Samson, I. Sanders, O. Saurat, C. Scarpelli, T. Schiex, B. Segurens, A. J. Severin, D. J. Sherrier, R. Shi, S. Sims, S. R. Singer, S. Sinharoy, L. Sterck, A. Viollet, B.-B. Wang, K. Wang, M. Wang, X. Wang, J. Warfsmann, J. Weissenbach, D. D. White, J. D. White, G. B. Wiley, P. Wincker, Y. Xing, L. Yang, Z. Yao, F. Ying, J. Zhai, L. Zhou, A. Zuber, J. Dénarié, R. A. Dixon, G. D. May, D. C. Schwartz, J. Rogers, F. Quétier, C. D. Town, B. A. Roe, The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520–524 (2011).
[doi:10.1038/nature10625](https://doi.org/10.1038/nature10625) [Medline](#)
18. J. Schmutz, P. E. McClean, S. Mamidi, G. A. Wu, S. B. Cannon, J. Grimwood, J. Jenkins, S. Shu, Q. Song, C. Chavarro, M. Torres-Torres, V. Geffroy, S. M. Moghaddam, D. Gao, B. Abernathy, K. Barry, M. Blair, M. A. Brick, M. Chovatia, P. Gepts, D. M. Goodstein, M. Gonzales, U. Hellsten, D. L. Hyten, G. Jia, J. D. Kelly, D. Kudrna, R. Lee, M. M. S. Richard, P. N. Miklas, J. M. Osorno, J. Rodrigues, V. Thareau, C. A. Urrea, M. Wang, Y. Yu, M. Zhang, R. A. Wing, P. B. Cregan, D. S. Rokhsar, S. A. Jackson, A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **46**, 707–713 (2014). [doi:10.1038/ng.3008](https://doi.org/10.1038/ng.3008) [Medline](#)
19. J. Schmutz, S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, Q. Song, J. J. Thelen, J. Cheng, D. Xu, U. Hellsten, G. D. May, Y. Yu, T. Sakurai, T. Umezawa, M. K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.-C. Zhang, K. Shinozaki, H. T. Nguyen, R. A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R. C. Shoemaker, S. A. Jackson, Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183 (2010).
[doi:10.1038/nature08670](https://doi.org/10.1038/nature08670) [Medline](#)
20. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015). [doi:10.1186/s13059-015-0721-2](https://doi.org/10.1186/s13059-015-0721-2) [Medline](#)
21. D. L. Stern, The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013). [doi:10.1038/nrg3483](https://doi.org/10.1038/nrg3483) [Medline](#)
22. N. Shubin, C. Tabin, S. Carroll, Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009). [doi:10.1038/nature07891](https://doi.org/10.1038/nature07891) [Medline](#)

23. J. J. Doyle, Chasing unicorns: Nodulation origins and the paradox of novelty. *Am. J. Bot.* **103**, 1865–1868 (2016). [doi:10.3732/ajb.1600260](https://doi.org/10.3732/ajb.1600260) [Medline](#)
24. J. F. Marsh, A. Rakocevic, R. M. Mitra, L. Brocard, J. Sun, A. Eschstruth, S. R. Long, M. Schultze, P. Ratet, G. E. D. Oldroyd, *Medicago truncatula NIN* is essential for rhizobia-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol.* **144**, 324–335 (2007). [doi:10.1104/pp.106.093021](https://doi.org/10.1104/pp.106.093021) [Medline](#)
25. F. Clavijo, I. Diedhiou, V. Vaissayre, L. Brottier, J. Acolatse, D. Moukouanga, A. Crabos, F. Auguy, C. Franche, H. Gherbi, A. Champion, V. Hocher, D. Barker, D. Bogusz, L. S. Tisa, S. Svistoonoff, The *Casuarina NIN* gene is transcriptionally activated throughout *Frankia* root infection as well as in response to bacterial diffusible signals. *New Phytol.* **208**, 887–903 (2015). [doi:10.1111/nph.13506](https://doi.org/10.1111/nph.13506) [Medline](#)
26. H. Gherbi, K. Markmann, S. Svistoonoff, J. Estevan, D. Autran, G. Giczey, F. Auguy, B. Péret, L. Laplaze, C. Franche, M. Parniske, D. Bogusz, SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and *Frankiabacteria*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4928–4932 (2008). [doi:10.1073/pnas.0710618105](https://doi.org/10.1073/pnas.0710618105) [Medline](#)
27. L. Schauer, A. Roussis, J. Stiller, J. Stougaard, A plant regulator controlling development of symbiotic root nodules. *Nature* **402**, 191–195 (1999). [doi:10.1038/46058](https://doi.org/10.1038/46058) [Medline](#)
28. A. van Zeijl, T. A. K. Wardhani, M. Seifi Kalhor, L. Rutten, F. Bu, M. Hartog, S. Linders, E. E. Fedorova, T. Bisseling, W. Kohlen, R. Geurts, CRISPR/Cas9-mediated mutagenesis of four putative symbiosis genes of the tropical tree *Parasponia andersonii* reveals novel phenotypes. *Front. Plant Sci.* **9**, 284 (2018). [doi:10.3389/fpls.2018.00284](https://doi.org/10.3389/fpls.2018.00284) [Medline](#)
29. S. Svistoonoff, F. M. Benabdoun, M. Nambiar-Veetil, L. Imanishi, V. Vaissayre, S. Cesari, N. Diagne, V. Hocher, F. de Billy, J. Bonneau, L. Wall, N. Ykhlef, C. Rosenberg, D. Bogusz, C. Franche, H. Gherbi, The independent acquisition of plant root nitrogen-fixing symbiosis in Fabids recruited the same genetic pathway for nodule organogenesis. *PLOS ONE* **8**, e64515 (2013). [doi:10.1371/journal.pone.0064515](https://doi.org/10.1371/journal.pone.0064515) [Medline](#)
30. K. Markmann, G. Giczey, M. Parniske, Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLOS Biol.* **6**, e68 (2008). [doi:10.1371/journal.pbio.0060068](https://doi.org/10.1371/journal.pbio.0060068) [Medline](#)
31. L. G. Nagy, R. A. Ohm, G. M. Kovács, D. Floudas, R. Riley, A. Gácser, M. Sipiczki, J. M. Davis, S. L. Doty, G. S. de Hoog, B. F. Lang, J. W. Spatafora, F. M. Martin, I. V. Grigoriev, D. S. Hibbett, Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat. Commun.* **5**, 4471 (2014). [doi:10.1038/ncomms5471](https://doi.org/10.1038/ncomms5471) [Medline](#)
32. C. E. Conn, R. Bythell-Douglas, D. Neumann, S. Yoshida, B. Whittington, J. H. Westwood, K. Shirasu, C. S. Bond, K. A. Dyer, D. C. Nelson, Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* **349**, 540–543 (2015). [doi:10.1126/science.aab1140](https://doi.org/10.1126/science.aab1140) [Medline](#)
33. J. He, V. A. Benedito, M. Wang, J. D. Murray, P. X. Zhao, Y. Tang, M. K. Udvardi, The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics* **10**, 441 (2009). [doi:10.1186/1471-2105-10-441](https://doi.org/10.1186/1471-2105-10-441) [Medline](#)

34. G. Criscuolo, V. T. Valkov, A. Parlati, L. M. Alves, M. Chiurazzi, Molecular characterization of the *Lotus japonicus* NRT1(PTR) and NRT2 families. *Plant Cell Environ.* **35**, 1567–1581 (2012). [doi:10.1111/j.1365-3040.2012.02510.x](https://doi.org/10.1111/j.1365-3040.2012.02510.x) [Medline](#)
35. L. Aravind, H. Watanabe, D. J. Lipman, E. V. Koonin, Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11319–11324 (2000). [doi:10.1073/pnas.200346997](https://doi.org/10.1073/pnas.200346997) [Medline](#)
36. R. Albalat, C. Cañestro, Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016). [doi:10.1038/nrg.2016.39](https://doi.org/10.1038/nrg.2016.39) [Medline](#)
37. P. M. Delaux, K. Varala, P. P. Edger, G. M. Coruzzi, J. C. Pires, J.-M. Ané, Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLOS Genet.* **10**, e1004487 (2014). [doi:10.1371/journal.pgen.1004487](https://doi.org/10.1371/journal.pgen.1004487) [Medline](#)
38. A. Y. Borisov, L. H. Madsen, V. E. Tsyanov, Y. Umehara, V. A. Voroshilova, A. O. Batagov, N. Sandal, A. Mortensen, L. Schausler, N. Ellis, I. A. Tikhonovich, J. Stougaard, The *Sym35* gene required for root nodule development in pea is an ortholog of *Nin* from *Lotus japonicus*. *Plant Physiol.* **131**, 1009–1017 (2003). [doi:10.1104/pp.102.016071](https://doi.org/10.1104/pp.102.016071) [Medline](#)
39. N. Azani, M. Babineau, C. D. Bailey, H. Banks, A. R. Barbosa, R. B. Pinto, J. S. Boatwright, L. M. Borges, G. K. Brown, A. Bruneau, E. Candido, D. Cardoso, K.-F. Chung, R. P. Clark, A. S. Conceição, M. Crisp, P. Cubas, A. Delgado-Salinas, K. G. Dexter, J. J. Doyle, J. Duminil, A. N. Egan, M. De La Estrella, M. J. Falcão, D. A. Filatov, A. P. Fortuna-Perez, R. H. Fortunato, E. Gagnon, P. Gasson, J. G. Rando, A. M. G. Azevedo Tozzi, B. Gunn, D. Harris, E. Haston, J. A. Hawkins, P. S. Herendeen, C. E. Hughes, J. R. V. Iganci, F. Javadi, S. A. Kanu, S. Kazempour-Osaloo, G. C. Kite, B. B. Klitgaard, F. J. Kochanowski, E. J. M. Koenen, L. Kovar, M. Lavin, M. Roux, G. P. Lewis, H. C. de Lima, M. C. López-Roberts, B. Mackinder, V. H. Maia, V. Malécot, V. F. Mansano, B. Marazzi, S. Mattapha, J. T. Miller, C. Mitsuyuki, T. Moura, D. J. Murphy, M. Nageswara-Rao, B. Nevado, D. Neves, D. I. Ojeda, R. T. Pennington, D. E. Prado, G. Prenner, L. P. de Queiroz, G. Ramos, F. L. Ranzato Filardi, P. G. Ribeiro, M. L. Rico-Arce, M. J. Sanderson, J. Santos-Silva, W. M. B. São-Mateus, M. J. S. Silva, M. F. Simon, C. Sinou, C. Snak, É. R. de Souza, J. Sprent, K. P. Steele, J. E. Steier, R. Steeves, C. H. Stirton, S. Tagane, B. M. Torke, H. Toyama, D. T. Cruz, M. Vatanparast, J. J. Wieringa, M. Wink, M. F. Wojciechowski, T. Yahara, T. Yi, E. Zimmerman, A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny – The Legume Phylogeny Working Group (LPWG). *Taxon* **66**, 44–77 (2017). [doi:10.12705/661.3](https://doi.org/10.12705/661.3)
40. J. I. Sprent, Evolving ideas of legume evolution and diversity: A taxonomic perspective on the occurrence of nodulation. *New Phytol.* **174**, 11–25 (2007). [doi:10.1111/j.1469-8137.2007.02015.x](https://doi.org/10.1111/j.1469-8137.2007.02015.x) [Medline](#)
41. J.-F. Arrighi, O. Godfroy, F. de Billy, O. Saurat, A. Jauneau, C. Gough, The *RPG* gene of *Medicago truncatula* controls *Rhizobium*-directed polar growth during infection. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9817–9822 (2008). [doi:10.1073/pnas.0710273105](https://doi.org/10.1073/pnas.0710273105) [Medline](#)
42. M. R. Chandler, Some observations on infection of *Arachis hypogaea* L. by *Rhizobium*. *J. Exp. Bot.* **29**, 749–755 (1978). [doi:10.1093/jxb/29.3.749](https://doi.org/10.1093/jxb/29.3.749)

43. Z. Peng, F. Liu, L. Wang, H. Zhou, D. Paudel, L. Tan, J. Maku, M. Gallo, J. Wang, Transcriptome profiles reveal gene regulation of peanut (*Arachis hypogaea* L.) nodulation. *Sci. Rep.* **7**, 40066 (2017). [Medline](#)
44. D. W. Gorbet, J. C. Burton, A non-nodulating peanut. *Crop Sci.* **19**, 727–728 (1979). [doi:10.2135/cropsci1979.0011183X001900050045x](https://doi.org/10.2135/cropsci1979.0011183X001900050045x)
45. S. De Mita, A. Streng, T. Bisseling, R. Geurts, Evolution of a symbiotic receptor through gene duplications in the legume-rhizobium mutualism. *New Phytol.* **201**, 961–972 (2014). [doi:10.1111/nph.12549](https://doi.org/10.1111/nph.12549) [Medline](#)
46. M. Parniske, Arbuscular mycorrhiza: The mother of plant root endosymbioses. *Nat. Rev. Microbiol.* **6**, 763–775 (2008). [doi:10.1038/nrmicro1987](https://doi.org/10.1038/nrmicro1987) [Medline](#)
47. P. Favre, L. Bapaume, E. Bossolini, M. Delorenzi, L. Falquet, D. Reinhardt, A novel bioinformatics pipeline to discover genes related to arbuscular mycorrhizal symbiosis based on their evolutionary conservation pattern among higher plants. *BMC Plant Biol.* **14**, 333 (2014). [doi:10.1186/s12870-014-0333-0](https://doi.org/10.1186/s12870-014-0333-0) [Medline](#)
48. A. Bravo, T. York, N. Pumplin, L. A. Mueller, M. J. Harrison, Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nat. Plants* **2**, 15208 (2016). [doi:10.1038/nplants.2015.208](https://doi.org/10.1038/nplants.2015.208) [Medline](#)
49. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007). [doi:10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088) [Medline](#)
50. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008). [doi:10.1016/j.cell.2008.06.030](https://doi.org/10.1016/j.cell.2008.06.030) [Medline](#)
51. R. van Velzen, R. Holmer, F. Bu, L. Rutten, A. van Zeijl, W. Liu, L. Santuari, Q. Cao, T. Sharma, D. Shen, Y. Roswanjaya, T. A. K. Wardhani, M. S. Kalhor, J. Jansen, J. van den Hoogen, B. Güngör, M. Hartog, J. Hontelez, J. Verver, W. C. Yang, E. Schijlen, R. Repin, M. Schilthuizen, M. E. Schranz, R. Heidstra, K. Miyata, E. Fedorova, W. Kohlen, T. Bisseling, S. Smit, R. Geurts, Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4700–E4709 (2018). [Medline](#)
52. L. G. Nagy, Evolution: Complex multicellular life with 5,500 genes. *Curr. Biol.* **27**, R609–R612 (2017). [doi:10.1016/j.cub.2017.04.032](https://doi.org/10.1016/j.cub.2017.04.032) [Medline](#)
53. M. A. O’Malley, J. G. Wideman, I. Ruiz-Trillo, Losing complexity: The role of simplification in macroevolution. *Trends Ecol. Evol.* **31**, 608–621 (2016). [doi:10.1016/j.tree.2016.04.004](https://doi.org/10.1016/j.tree.2016.04.004) [Medline](#)
54. D. S. LeBauer, K. K. Treseder, Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology* **89**, 371–379 (2008). [doi:10.1890/06-2057.1](https://doi.org/10.1890/06-2057.1) [Medline](#)
55. J. Streeter, P. P. Wong, Inhibition of legume nodule formation and N₂-fixation by nitrate. *Crit. Rev. Plant Sci.* **7**, 1–23 (1988). [doi:10.1080/07352688809382257](https://doi.org/10.1080/07352688809382257)
56. E. T. Kiers, R. A. Rousseau, S. A. West, R. F. Denison, Host sanctions and the legume-rhizobium mutualism. *Nature* **425**, 78–81 (2003). [doi:10.1038/nature01931](https://doi.org/10.1038/nature01931) [Medline](#)

57. H. Fujita, S. Aoki, M. Kawaguchi, Evolutionary dynamics of nitrogen fixation in the legume-rhizobia symbiosis. *PLOS ONE* **9**, e93670 (2014). [doi:10.1371/journal.pone.0093670](https://doi.org/10.1371/journal.pone.0093670) [Medline](#)
58. L. Carro, P. Pujic, M. E. Trujillo, P. Normand, *Micromonospora* is a normal occupant of actinorhizal nodules. *J. Biosci.* **38**, 685–693 (2013). [doi:10.1007/s12038-013-9359-y](https://doi.org/10.1007/s12038-013-9359-y) [Medline](#)
59. J. I. Sprent, *Legume Nodulation: A Global Perspective* (Wiley, 2009).
60. R. Zgadzaj, R. Garrido-Oter, D. B. Jensen, A. Koprivova, P. Schulze-Lefert, S. Radutoiu, Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7996–E8005 (2016). [doi:10.1073/pnas.1616564113](https://doi.org/10.1073/pnas.1616564113) [Medline](#)
61. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011). [doi:10.1093/bioinformatics/btr011](https://doi.org/10.1093/bioinformatics/btr011) [Medline](#)
62. E. Veeckman, T. Ruttink, K. Vandepoele, Are we there yet? Reliably estimating the completeness of plant genome sequences. *Plant Cell* **28**, 1759–1768 (2016). [doi:10.1105/tpc.16.00349](https://doi.org/10.1105/tpc.16.00349) [Medline](#)
63. R. Arratia, D. Martin, G. Reinert, M. S. Waterman, Poisson process approximation for sequence repeats, and sequencing by hybridization. *J. Comput. Biol.* **3**, 425–463 (1996). [doi:10.1089/cmb.1996.3.425](https://doi.org/10.1089/cmb.1996.3.425) [Medline](#)
64. R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S.-M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T.-W. Lam, J. Wang, SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012). [doi:10.1186/2047-217X-1-18](https://doi.org/10.1186/2047-217X-1-18) [Medline](#)
65. R. Kajitani, K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura, M. Okuno, M. Yabana, M. Harada, E. Nagayasu, H. Maruyama, Y. Kohara, A. Fujiyama, T. Hayashi, T. Itoh, Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014). [doi:10.1101/gr.170720.113](https://doi.org/10.1101/gr.170720.113) [Medline](#)
66. E. W. Myers, G. G. Sutton, A. L. Delcher, I. M. Dew, D. P. Fasulo, M. J. Flanigan, S. A. Kravitz, C. M. Mobarry, K. H. Reinert, K. A. Remington, E. L. Anson, R. A. Bolanos, H. H. Chou, C. M. Jordan, A. L. Halpern, S. Lonardi, E. M. Beasley, R. C. Brandon, L. Chen, P. J. Dunn, Z. Lai, Y. Liang, D. R. Nusskern, M. Zhan, Q. Zhang, X. Zheng, G. M. Rubin, M. D. Adams, J. C. Venter, A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204 (2000). [doi:10.1126/science.287.5461.2196](https://doi.org/10.1126/science.287.5461.2196) [Medline](#)
67. M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011). [doi:10.1093/bioinformatics/btq683](https://doi.org/10.1093/bioinformatics/btq683) [Medline](#)
68. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015). [doi:10.1093/bioinformatics/btv351](https://doi.org/10.1093/bioinformatics/btv351) [Medline](#)

69. A. Smit, R. Hubley, P. Green, RepeatMasker Open–4.0, 2013–2015; www.repeatmasker.org.
70. Y. Han, S. R. Wessler, MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199 (2010). [doi:10.1093/nar/gkq862](https://doi.org/10.1093/nar/gkq862) [Medline](#)
71. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008). [doi:10.1186/1471-2105-9-18](https://doi.org/10.1186/1471-2105-9-18) [Medline](#)
72. G. Gremme, S. Steinbiss, S. Kurtz, GenomeTools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **10**, 645–656 (2013). [doi:10.1109/TCBB.2013.68](https://doi.org/10.1109/TCBB.2013.68) [Medline](#)
73. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). [doi:10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) [Medline](#)
74. M. S. Campbell, M. Law, C. Holt, J. C. Stein, G. D. Moghe, D. E. Hufnagel, J. Lei, R. Achawanantakun, D. Jiao, C. J. Lawrence, D. Ware, S.-H. Shiu, K. L. Childs, Y. Sun, N. Jiang, M. Yandell, MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* **164**, 513–524 (2014). [doi:10.1104/pp.113.230144](https://doi.org/10.1104/pp.113.230144) [Medline](#)
75. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr., L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003). [doi:10.1093/nar/gkg770](https://doi.org/10.1093/nar/gkg770) [Medline](#)
76. M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004). [doi:10.1093/nar/gkh379](https://doi.org/10.1093/nar/gkh379) [Medline](#)
77. A. V. Lukashin, M. Borodovsky, GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115 (1998). [doi:10.1093/nar/26.4.1107](https://doi.org/10.1093/nar/26.4.1107) [Medline](#)
78. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004). [doi:10.1186/1471-2105-5-59](https://doi.org/10.1186/1471-2105-5-59) [Medline](#)
79. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011). [doi:10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883) [Medline](#)
80. E. M. Zdobnov, R. Apweiler, InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001). [doi:10.1093/bioinformatics/17.9.847](https://doi.org/10.1093/bioinformatics/17.9.847) [Medline](#)
81. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006). [doi:10.1093/bioinformatics/btl158](https://doi.org/10.1093/bioinformatics/btl158) [Medline](#)
82. Y. Yang, S. A. Smith, Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics.

Mol. Biol. Evol. **31**, 3081–3092 (2014). [doi:10.1093/molbev/msu245](https://doi.org/10.1093/molbev/msu245) [Medline](#)

83. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015). [doi:10.1093/molbev/msv150](https://doi.org/10.1093/molbev/msv150) [Medline](#)
84. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016). [doi:10.1093/molbev/msw046](https://doi.org/10.1093/molbev/msw046) [Medline](#)
85. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). [doi:10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010) [Medline](#)
86. G. Tan, M. Muffato, C. Ledergerber, J. Herrero, N. Goldman, M. Gil, C. Dessimoz, Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* **64**, 778–791 (2015). [doi:10.1093/sysbio/syv033](https://doi.org/10.1093/sysbio/syv033) [Medline](#)
87. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). [doi:10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033) [Medline](#)
88. V. Solovyev, in *Handbook of Statistical Genetics*, D. J. Balding, M. Bishop, C. Cannings, Eds. (Wiley, 2008), pp. 97–159.
89. C. D. Bell, D. E. Soltis, P. S. Soltis, The age and diversification of the angiosperms revisited. *Am. J. Bot.* **97**, 1296–1303 (2010). [doi:10.3732/ajb.0900346](https://doi.org/10.3732/ajb.0900346) [Medline](#)
90. Z. Xi, B. R. Ruhfel, H. Schaefer, A. M. Amorim, M. Sugumaran, K. J. Wurdack, P. K. Endress, M. L. Matthews, P. F. Stevens, S. Mathews, C. C. Davis, Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17519–17524 (2012). [doi:10.1073/pnas.1205818109](https://doi.org/10.1073/pnas.1205818109) [Medline](#)
91. M. T. J. Johnson, E. J. Carpenter, Z. Tian, R. Bruskiewich, J. N. Burris, C. T. Carrigan, M. W. Chase, N. D. Clarke, S. Covshoff, C. W. Depamphilis, P. P. Edger, F. Goh, S. Graham, S. Greiner, J. M. Hibberd, I. Jordon-Thaden, T. M. Kutchan, J. Leebens-Mack, M. Melkonian, N. Miles, H. Myburg, J. Patterson, J. C. Pires, P. Ralph, M. Rolf, R. F. Sage, D. Soltis, P. Soltis, D. Stevenson, C. N. Stewart Jr., B. Surek, C. J. M. Thomsen, J. C. Villarreal, X. Wu, Y. Zhang, M. K. Deyholos, G. K.-S. Wong, Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLOS ONE* **7**, e50226 (2012). [doi:10.1371/journal.pone.0050226](https://doi.org/10.1371/journal.pone.0050226) [Medline](#)
92. M. V. Han, G. W. C. Thomas, J. Lugo-Martinez, M. W. Hahn, Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013). [doi:10.1093/molbev/mst100](https://doi.org/10.1093/molbev/mst100) [Medline](#)
93. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). [doi:10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300) [Medline](#)

94. D. T. Hoang, L. S. Vinh, T. Flouri, A. Stamatakis, A. von Haeseler, B. Q. Minh, MPBoot: Fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evol. Biol.* **18**, 11 (2018). [doi:10.1186/s12862-018-1131-3](https://doi.org/10.1186/s12862-018-1131-3) [Medline](#)
95. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017). [doi:10.1038/nmeth.4285](https://doi.org/10.1038/nmeth.4285) [Medline](#)
96. M. Griesmann, Y. Chang, X. Liu, Y. Song, G. Haberer, M. B. Crook, B. Billault-Penneteau, D. Laressergues, L. Imanishi, Y. P. Roswanjaya, W. Kohlen, P. Pujic, K. Battenberg, N. Alloisio, W. Sun, H. Hilhorst, M. G. Salgado, V. Hocher, H. Gherbi, S. Svistoonoff, J. J. Doyle, S. He, Y. Xu, W. Xian, Y. Fu, P. Normand, A. M. Berry, L. G. Wall, J. M. Ané, K. Pawłowski, X. Xu, H. Yang, M. Spannagl, K. F. X. Mayer, G. K. Wong, M. Parniske, P. M. Delaux, S. Cheng, Data supporting multiple independent losses of nitrogen-fixing root nodule symbiosis. GigaDB (2018); [doi:10.5524/100300](https://doi.org/10.5524/100300)
97. V. Tolieng, B. Prasirtsak, J. Sitdhipol, N. Thongchul, S. Tanasupawat, Identification and lactic acid production of bacteria isolated from soils and tree barks. *Malays. J. Microbiol.* **13**, 100–108 (2017).
98. C. Valverde, L. G. Wall, Time course of nodule development in the *Discaria trinervis* (Rhamnaceae) – *Frankia* symbiosis. *New Phytol.* **141**, 345–354 (1999). [doi:10.1046/j.1469-8137.1999.00345.x](https://doi.org/10.1046/j.1469-8137.1999.00345.x)
99. G. Fahraeus, The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. *J. Gen. Microbiol.* **16**, 374–381 (1957). [Medline](#)
100. S. L. Dellaporta, J. Wood, J. B. Hicks, A plant DNA minipreparation: Version II. *Plant Mol. Biol. Report.* **1**, 19–21 (1983). [doi:10.1007/BF02712670](https://doi.org/10.1007/BF02712670)
101. A. Healey, A. Furtado, T. Cooper, R. J. Henry, Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **10**, 21 (2014). [doi:10.1186/1746-4811-10-21](https://doi.org/10.1186/1746-4811-10-21) [Medline](#)
102. A. Ribeiro, A. D. L. Akkermans, A. van Kammen, T. Bisseling, K. Pawłowski, A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development. *Plant Cell* **7**, 785–794 (1995). [doi:10.1105/tpc.7.6.785](https://doi.org/10.1105/tpc.7.6.785) [Medline](#)
103. J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald, W. J. Rutter, Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299 (1979). [doi:10.1021/bi00591a005](https://doi.org/10.1021/bi00591a005) [Medline](#)