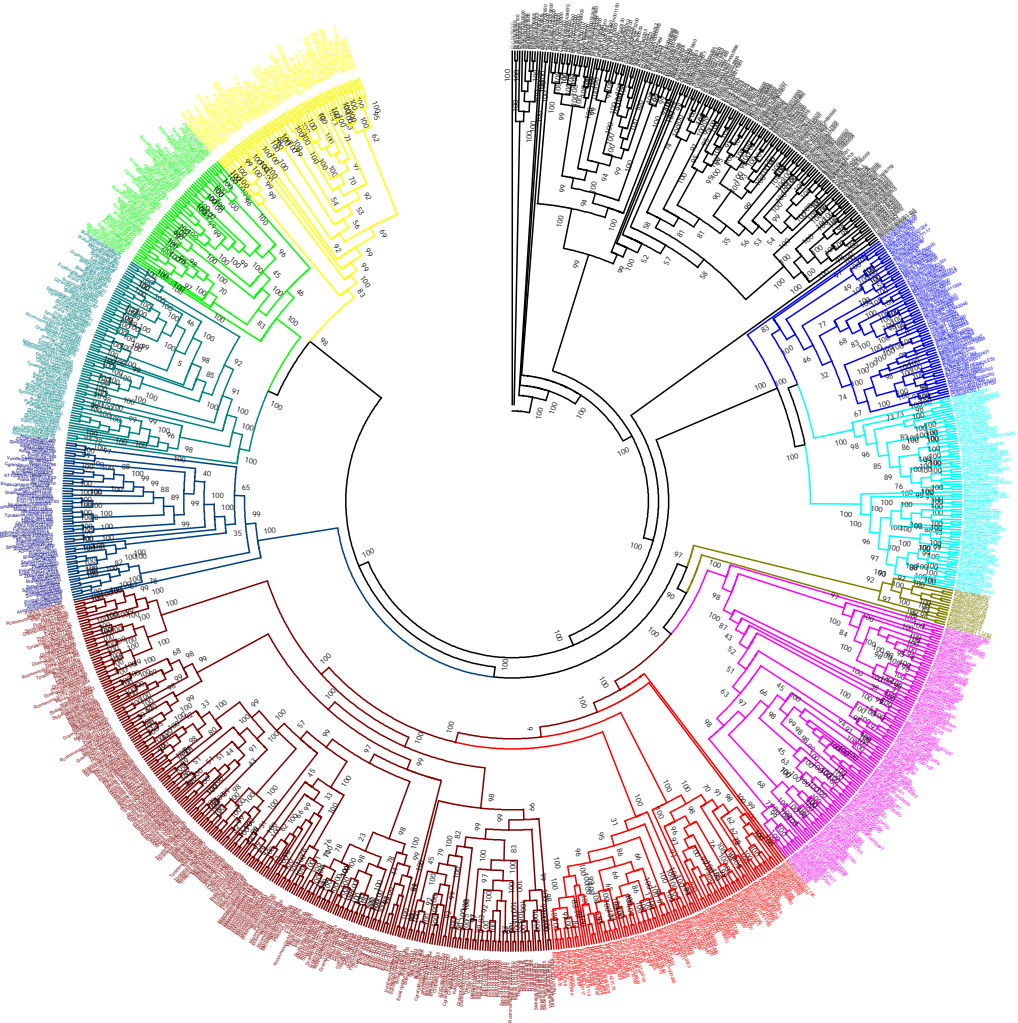# Genome-wide identification of CrRLK1L subfamily genes in diverse species and comparative analysis of their roles in the legume–rhizobia symbiosis

Jorge Solis-Miranda, Citlali Fonseca-Garcia, Nava Noreide, Ronal Pacheco, and Carmen Quinto,*
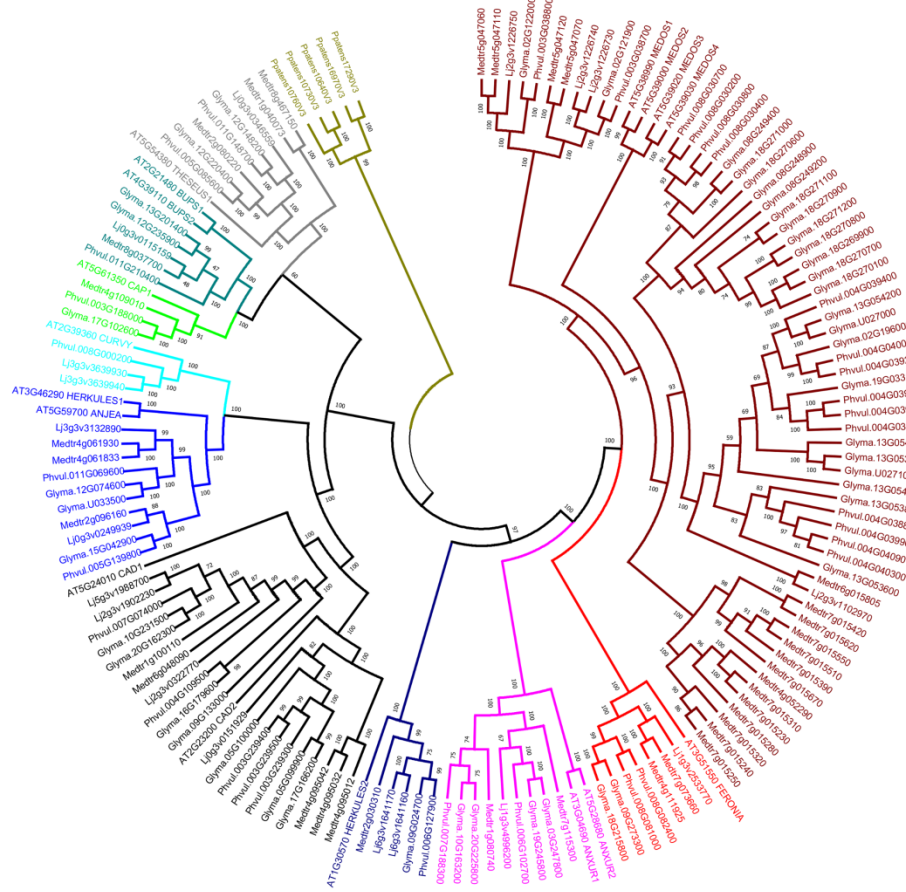
Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Colonia Chamilpa, Cuernavaca, Morelos 62210, Mexico; jsolis@ibt.unam.mx (J.S.-M), fonsecac@ibt.unam.mx (C.F.-G), noreide@ibt.unam.mx (N.N), ronal.pacheco@mail.ibt.unam.mx (R.P.),

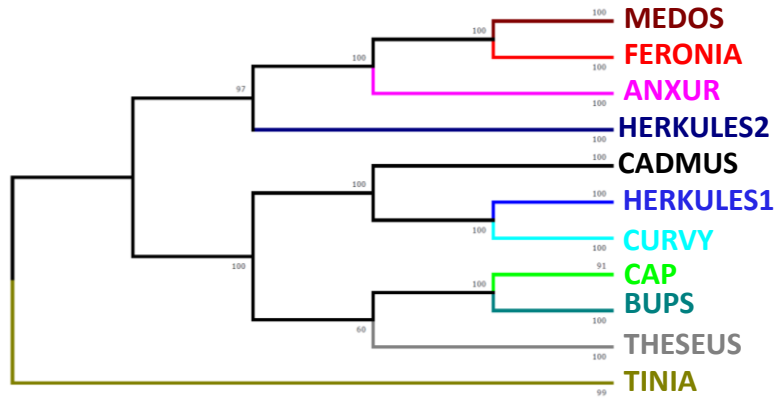*Correspondence: quinto@ibt.unam.mx (C.Q.)

**Figure S1: Phylogenetic relationship among 1050 CrRLK1L proteins.** High-resolution unrooted approximately maximum-likelihood phylogenetic tree inferred from the 1,050 CrRLK1L proteins present in 57 plant species. The clades, indicated in different colors, are named based on the *A. thaliana* CrRLK1L names. The CADMUS clade contains uncharacterized CrRLK1Ls from *A. thaliana*, and the TINIA clade corresponds to a clade formed only with the CrRLK1L proteins of *S. moellendorffii*, *S. fallax*, *P. patens*, and *M. polymorpha*. The phylogenetic tree was constructed using IQ-TREE software with the JTT+F+R10 substitution model with 1,000 bootstrap iterations. Gene names are highlighted in colors as follows: *ANX* in purple, *BUPS* in green, *CAD* in black, *CAP* in lime, *CRV* in blue, *FER* in red, , *HERK1* (*ANJ*) in blue, *HERK2* in dark blue, *MEDOS* in brown, and *THE* in gray.

**Figure S2. Unrooted approximately maximum-likelihood phylogenetic tree of the CrRLK1L subfamily proteins in various species. A)** Inferred phylogenetic tree of 155 CrRLK1L proteins from *P. vulgaris*, *L. japonicus*, *M. truncatula*, *G. max*, *A. thaliana*, and *P. patens*. **B)** Simplified dendrogram representing the 11 clades of the inferred phylogenetic tree in A. Local support values correspond to those in A. Gene names are highlighted in colors as follows: *ANX* in purple, *BUPS* in green, *CAD* in black, *CAP* in lime, *CRV* in blue, *FER* in red, , *HERK1* (*ANJ*) in blue, *HERK2* in dark blue, *MEDOS* in brown, and *THE* in gray.

**A** *P. vulgaris*
**B** *L. japonicus*
**C** *G. max*
**D** *M. truncatula*
**E** *A. thaliana*
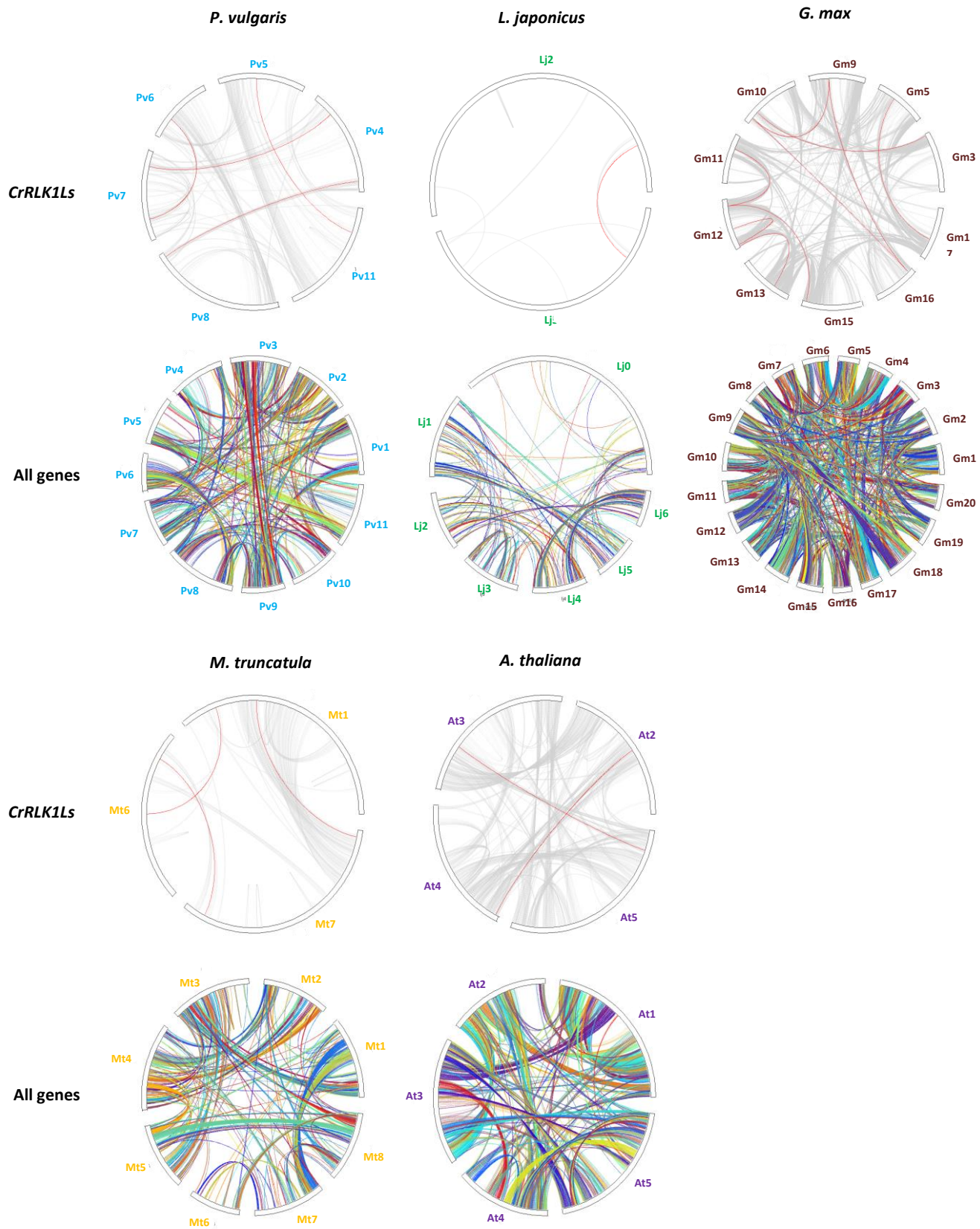**F** *S. bicolor*

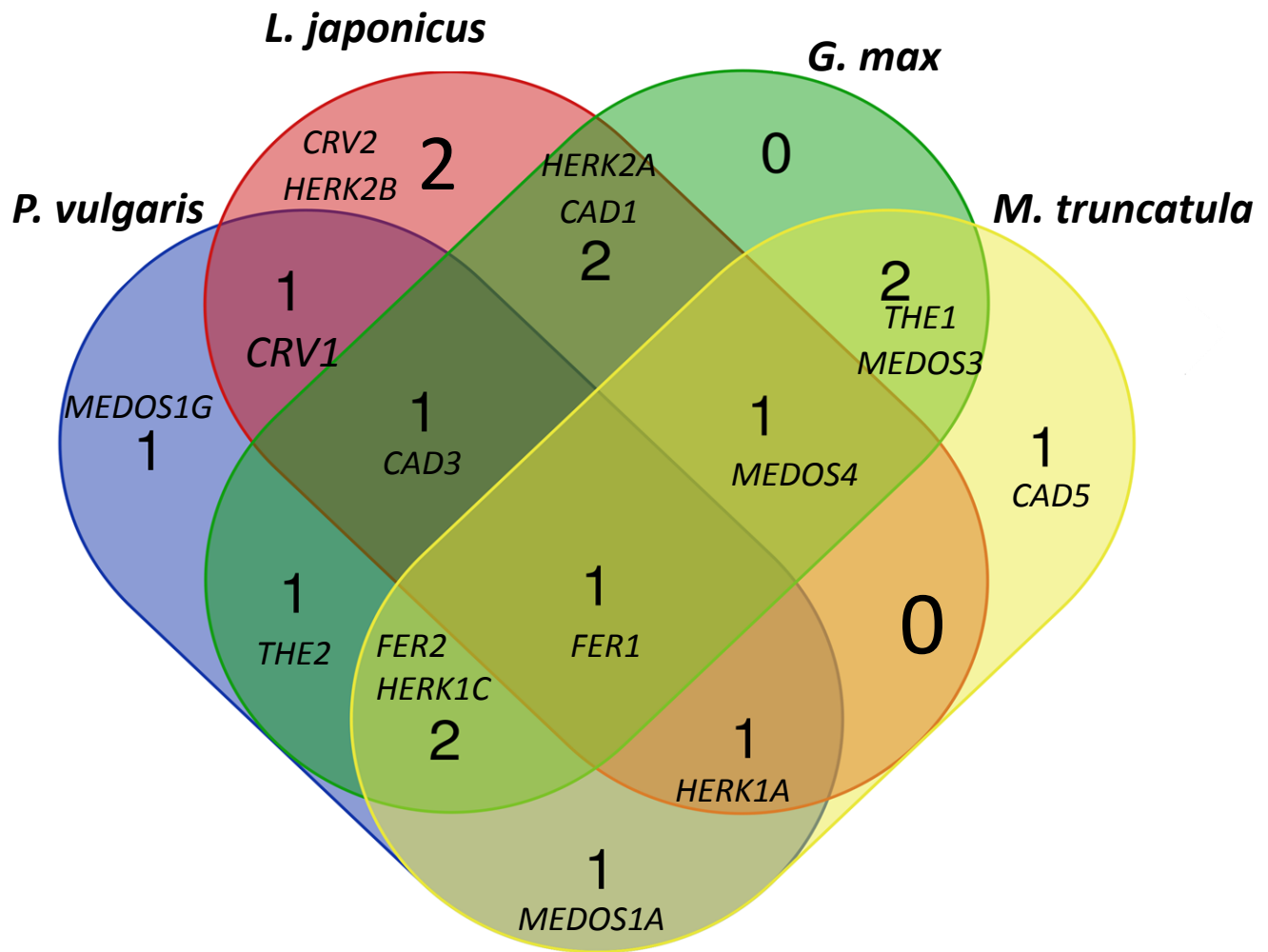● CLADO FER/ANX   ● CLADO BUPS/CAP   ● CLADO THE/HERK   ● CLADO CAD   ● CLADO MEDOS

**Figure S3. Chromosomal location of *CrRLK1L* genes in various species.** Chromosomal location of *CrRLK1L* genes in **A)** *P. vulgaris*, **B)** *L. japonicus*, **C)** *G. max*, **D)** *M. truncatula*, **E)** *A. thaliana*, and **F)** *S. bicolor*. The colored dots and lines indicate the chromosomal location of the different genes.

*P. vulgaris*

Pv5
Pv6
Pv4
Pv7
*CrRLK1Ls*
Pv11
Pv8

*L. japonicus*

Lj2
*CrRLK1Ls*
Lj

*G. max*

Gm9
Gm10
Gm5
Gm11
Gm3
*CrRLK1Ls*
Gm12
Gm17
Gm13
Gm16
Gm15

All genes

Pv3
Pv4
Pv2
Pv5
Pv1
Pv6
Pv7
Pv11
Pv8
Pv10
Pv9

Lj0
Lj1
Lj6
Lj2
Lj5
Lj3
Lj4

Gm6 Gm5
Gm7 Gm4
Gm8 Gm3
Gm9 Gm2
Gm10 Gm1
Gm11 Gm20
Gm12 Gm19
Gm13 Gm18
Gm14 Gm17
Gm15 Gm16

*M. truncatula*

Mt1
Mt6
Mt7

*A. thaliana*

At3
At2
At4
At5

*CrRLK1Ls*

All genes

Mt3
Mt2
Mt4
Mt1
Mt5
Mt8
Mt6
Mt7

At2
At1
At3
At5
At4

**Figure S4. Synteny of all *CrRLK1L* genes of *P. vulgaris*, *L. japonicus*, *G. max*, *M. truncatula*, and *A. thaliana* and of the complete genome of each species.** First and third rows, the *CrRLK1L* syntenic genes are represented as connected red lines. Only chromosomes containing syntenic genes are outlined. Second and fourth rows, all the syntenic genes present in the genome of each species are represented as connected colored lines.

**Figure S5. Venn diagram of *CrRLK1Ls* genes expressed in nodules of *P. vulgaris*, *L. japonicus*, *G. max*, and *M. truncatula*.** Venn diagram representing the number and names of the genes expressed in nodules. Each of the four ovals represents a set of nodule expressed genes on each legume. Names and numbers in the interconnections represent the nodule-expressed genes shared between the four legumes according to data from PvGEA, Lotus Base, MtGEA, and BAR.