Evolutionary dynamics of LysM and LRR type symbiotic receptor-like kinases and species-specificity in root-nodule symbiosis

with the supervision of Prof. Laura Rose

Mohammad Tanbir Habib

Heinrich Heine University Düsseldorf
September, 2019

Evolutionary dynamics of LysM and LRR type symbiotic receptor-like kinases and species-specificity in root-nodule symbiosis

---

Inaugural dissertation

for the attainment of the title of doctor

at the Heinrich Heine University Düsseldorf

---

presented by−

Mohammad Tanbir Habib

from Rajshahi, Bangladesh

Düsseldorf, September 2019

from the institute for population genetics
at the Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. Laura Rose
Co-supervisor: Prof. Dr. Petra Bauer

Date of the oral examination:

Declaration

I hereby declare that I have produced the presented thesis by myself and without unpermitted help, that all sources and aids used are indicated, and that this dissertation has not been presented elsewhere as an examination paper.

Düsseldorf, _____ _____

Signature

Zusammenfassung

Die Wurzeln von Leguminosen interagieren mit Bodenbakterien und entwickeln dabei in manchen Fällen charakteristische Organe genannt Wurzelköllchen. Diese Wurzelknöllchen sind Organe des Wirts und des Bakteriums der Familie der Rhizobien und dienen der Fixierung von elementarem Stickstoff (N2). Diese Interaktion ist die Hauptquelle des biologisch verfügbaren Stickstoffs in terrestrischen Ökosystemen und ist fast Ausschließlich in Leguminosen zu finden.

Ein früher und essentieller Schritt in der Initiation der Knöllchenbildung ist die Erkennung bakteriell sekretierter Nod-Faktoren (NFs) durch Oberflächenrezeptoren der Pflanzenzelle. Bei den Nod-Faktor Rezeptoren NFR1 und NFR5 der Leguminosen handelt es sich um LysM-Typ "Receptor like Kinases" (LysM-RLKs), welche von entscheidender Bedeutung bei der Erkennung des Symbiosepartners sind, so dass in diesen Genen mutierte Pflanzen keine Erkennungsreaktion mehr zeigen. Modifikationen von spezifischen Aminosäuren der extrazellulären LysM Domäne von NFR5 führen zu einem Wechsel des Symbionten, was ein weiterer Hinweis auf die Rolle von NFR5 bei der Erkennung des Symbionten ist. NFR1, der Exopolysccharidrezeptor 3 (EPR3) und andere LysM-RLKs sind zudem Kandidatengene für die artspezifische Erkennung des Rhizobium Partners. Darüber, wie Leguminosen in natürlichen Ökosystemen zwischen eng verwandten Mikroben unterscheiden können, ist immer noch nicht im Detail verstanden.

In dieser Arbeit rekonstruieren wir die Evolution dieser großen Genfamilie des Leguminosen Genus Lotus. Wir quantifizieren das Verhältnis von nicht synonymen Substitutionen (Austausche mit Änderung der Aminosäure) zu synonymen Substitutionen (Austausche ohne Änderung der Aminosäure) für jedes Codon dieser Gene als ein Maß für evolutionäre Beschränkungen. Die Sequenzvariabilität in NFR1 und NFR5 zeigte einen erhöhten Anteil an adaptiv evolvierenden Aminosäurepositionen ($Ka/Ks > 1$) in den Extrazellulären LysM Domänen.

Um die Evolution der Spezifität zu verstehen, charakterisierten wir zusätzlich die Rhizobienpräferenzen der Lotusarten. Wir konnten zeigen, dass die Variation der adaptiv evolvierenden Positionen in NFR1 und NFR5 mit der Kompatibilität mit Mesorhizobium loti zusammenhängt. Zusätzlich untersuchten wir die evolutionäre Geschichte von SymRK, einer, für die Symbiose essentiellen, "receptor-like" Kinase.

SymRK codiert eine “rezeptor-like” Kinase des LRR Typs (LRR-RLK), die sowohl für die Mycorrhizza Symbiose als auch für die Symbiose mit Rhizobien notwendig ist. Auch wenn die LRR-RLKs, die an der Wurzelknöllchen Symbiose beteiligt sind, keine klaren Anzeichen von adaptiver Evolution zeigten (Ka/Ks <= 1), so korrelierte die genetische Variation von SymRK klar mit der Spezifität mit *M. loti*.

Zusammenfassend kann man sagen, dass unsere Analyse Positionen der NFR1, NFR5 und der SymRK Gene aufzeigte, die an der spezifischen Arterkennung während der Ausbildung der Wurzelköllchen Symbiose beteiligt sein könnten. Des Weiteren zeigen wir neue evolutionäre Aspekte zweier kürzlich beschriebener *L. japonicus* Immunrezeptoren, Lys6 und Lys12, auf.

Abstract

The roots of legume plants interact with soil bacteria and, in some cases, develop distinctive organs called nodules. The nodules are host-Rhizobium dual organs, which house the bacteria and are where nitrogen fixation takes place. This distinct interaction is the main source of biologically fixed nitrogen available in our terrestrial ecosystems and can be found almost exclusively in legumes. An early and essential event for the initiation of nodule is the recognition of bacterial secreted nod factors (NFs) by plant cell-surface receptors. The Nod Factor Receptors, NFR1 and NFR5, of legumes are LysM-type receptor like kinases (LysM-RLKs) and are crucial for the recognition of symbiotic partner. Mutants in these genes are unresponsive to their symbiotic microbes. Modification of a specific amino acid in the extracellular LysM domain of NFR5 causes a symbiont shift, highlighting the role of NFR5 in legume-Rhizobium specificity. The NFR1, NFRe, Exopolysaccharide Receptor 3 (EPR3) and other LysM-RLKs are also candidate genes for species-specific discrimination of Rhizobium partners. However, despite progress, how legumes discriminate between closely related microbial species in natural ecosystems is still poorly understood.

Here we reconstruct the molecular evolution of this large gene family within the legume genus *Lotus*. We quantify the ratio of non-synonymous substitutions (changes resulting in a different amino acid) to synonymous substitutions (changes that do not alter the amino acid) for each codon of these genes as a proxy for evolutionary constraint. Sequence variability within NFR1 and NFR5 revealed an elevated proportion of adaptively evolving amino acid sites ($Ka/Ks > 1$) in the extracellular LysM domains. To understand the evolution of specificity, we also characterized the Rhizobium preferences of the same focal individuals. We discovered that variation at adaptively evolving sites in NRF1 and NFR5 was associated with compatibility with *Mesorhizobium loti*. Additionally, we studied the evolutionary history of SymRK, an essential symbiosis receptor-like kinase. SymRK encodes an LRR type receptor-like kinase (LRR-RLK) required for both mycorrhizal and rhizobial symbiosis. Although the LRR-RLKs involved in root-nodule symbiosis did not show clear evidence of adaptive evolution ($Ka/Ks <= 1$), genetic variation at SymRK was correlated with *M. loti* specificity. In summary, our analysis uncovered specific sites of the NFR1, NFR5 and SymRK genes that might be associated with species-specific recognition during root-nodule symbiosis. Furthermore, we present novel evolutionary aspects of two recently identified *L. japonicus* immune receptors, Lys6 and Lys12.

Dedication

To the Rohingyas.

Acknowledgements

I express my deepest gratitude to my supervisor Prof. Laura Rose for accepting me as a graduate student and for providing all necessary support to finish my research project. This project would not have been possible without her powerful scientific vision, critical advice and enormous support during my study period.

I am also grateful to Prof. Sheng Yang He of the Michigan State University for accepting me as a visiting student. It was a great pleasure for me to work for six months on a pant-microbe interaction project in his laboratory.

I am deeply indebted to the iGRAD-plant graduate school for providing the funding and guideline for doing the Dr.rer.nat. at the Heinrich Heine University of Düsseldorf. I am grateful to the i-GRAD plant co-ordinators Dr. Petra Fackendahl and Dr. Sigrun Wegener-Feldbrügge for their inspirations, critical suggestions for research and for providing all general administrative support.

I would like to thank Dr. Takaki Maekawa, MPI Cologne for providing the *Mesorhizobium loti* strain labeled with DsRed. I am also grateful to Dr. Dario I. Ojeda Alayon (`http://darioi.weebly.com`) for proving some rare genotypes of the genus Lotus collected from the Canary islands.

I would like to express my sincere gratitude to Mr. Predrag Marinoski for his help during the inoculation experiments.

The master's student Andreas Kukuk helped me during the inoculation experiment with Bradyrhizobium and in the cloning and sequencing process of the Clavata2 and Klavier genes. The bachelor's student Anna Spillner helped during the cloning and sequencing of Lys15 and Lys16 genes. I express my sincere thanks to both of them for their kind support.

This Project is a part of the DFG funded Graduiertenkolleg "GRK 1525: The dynamic response of plants to a changing environment".

Table of Contents

List of Tables

List of Figures

Abbreviations

BEB	Bayes Empirical Bayes
bp	base pair
cDNA	Complementary DNA
CEBiP	Chitin Elicitor Binding Protein
CERK1	Chitin Elicitor Receptor Kinase 1
DNA	Deoxyribonucleic acid
DsRed	Red Fluorescent Protein
EPR3	Exo-Polysaccharide Receptor 3
IT	Infection Threads
ITS	Internal Transcribed Spacer
Kappa	Transition to transversion rate ratio
Ka	The number of nonsynonymous substitutions per non-synonymous site
Ks	The number of synonymous substitutions per synonymous site
LHK	Lotus Histidine Kinase
LRR	Leucine Rich Repeat
LRT	Likelihood Ratio Test
LysM	Lysin Motif
MAMPs	Microbe-Associated Molecular Patterns
matk3	maturase K3
MPI	Max Planck Institute
N1LB	NFR1 Left-Border
N1RB	NFR1 Right-Border
NFR1	Nod Factor Receptor 1
NFR5	Nod Factor Receptor 5
NFRs	Nod Factor Receptors
Omega	Ka to Ks ratio
PAML	Phylogenetic Analysis by Maximum Likelihood
PCR	Polymerase Chain Reaction
rbcl	Ribulose-1,5-Bisphosphate Carboxylase/oxygenase Large subunit
RLKs	Receptor-Like Kinases
rpm	Rounds per minute
SDS	Sodium Dodecyl Sulphate
SNF	Symbiotic Nitrogen Fixation
SymRK	Symbiotic Receptor-Like Kinase
UK	United Kingdom
USA	United States of America

Chapter 1

Introduction

Symbiotic nitrogen fixation (SNF) is an evolutionary innovation with immense agricultural and ecological importance. The nitrogen-fixing Rhizobia are hosted and nourished inside the distinctive organs of legume roots called nodules. In exchange for plant fixed carbon, Rhizobia convert atmospheric nitrogen into biological accessible form and increase host fitness in nitrogen limiting conditions. Interestingly, this interaction is often highly specific, in which only particular species or strains of Rhizobia can establish symbiosis with certain legume species and vice versa (Andrews and Andrews, 2017). For example, the Rhizobia associated with the roots of African and Asian soybean cultivars are different from those that associate with American soybean roots(Abaidoo *et al.*, 2000 ; Annapurna and Krishnan, 2003).

The legume-Rhizobium specificity has been extensively investigated in the model legume species *L. japonicus*, which preferentially nodulates with *Mesorhizobium loti* over *Bradyrhizobium* (Gossmann *et al.*, 2012). However, the genetic basis of species-specificity in the legume-Rhizobium symbiosis is still poorly understood. In this study, I investigated dozens of species from the legume genus *Lotus* for their ability to produce nodules with *Mesorhizobium loti* labeled with DsRed. Our studies indicate that different species of the genus *Lotus* exhibit clade-specific symbiotic preferences.

One of the earliest steps of a successful symbiosis is the recognition of bacterial nod factors by plant cell-surface receptors, known as the nod factor receptors (NFRs). Nod factors are lipo-chitooligosaccharide effectors secreted by Rhizobia. Closely related species produce structurally similar, but distinctive nod factors (Bek *et al.*, 2010 ; Via *et al.*, 2016). NFR1 and NFR5 are the corresponding plant receptor-like kinases that recognize the nod factors of different Rhizobium origin by direct binding (Broghammer

et al., 2012). The earliest noticeable indication of a successful association is the initiation of infection threads, a process in which the Rhizobia enter the cortical cell layer via root hairs. In *Lotus*, NFR1 and NFR5 are absolutely required for the initiation of infection threads and mutants for these genes are dysfunctional at this earliest step of nodule formation (Madsen *et al.*, 2003 ; Radutoiu *et al.*, 2003).

The general structure NFR1, NFR5 and other LysM-RLKs comprise as many as three extracellular LysM domains and an internal cytoplasmic kinase domain connected through a short transmembrane domain. The LysM (lysin motif) domain is a widely distributed protein domain (PF01476) of around 40 amino acids, found mostly in prokaryotes at the substrate-binding sites of several peptidoglycan hydrolases (Buist *et al.*, 2008). Some plant pathogenic fungi use LysM as secreted virulence effectors to invade plants (Jonge *et al.*, 2010 ; Mentlak *et al.*, 2012 ; Lee *et al.*, 2014 ; Kombrink *et al.*, 2017 ; Dölfors *et al.*, 2019). Interestingly, the combination of extracellular LysM domains with a cytoplasmic kinase domain is found exclusively in plants.

A specific mutation at the site L118 of the putative ligand-binding LysM2 extracellular domain of NFR5 results in partial symbiotic alteration, implying that additional sites of NFR5 might also contribute in species-specificity (Radutoiu *et al.*, 2007). Moreover, the results of a taxonomic assay revealed significant changes in root associated microbiota in the *nfr5* knockout mutants, in which several bacterial orders were almost entirely depleted compared to the wild type roots (Zgadzaj *et al.*, 2016).

NFR1 dimerizes with NFR5 and co-operatively recognizes the nod factors secreted by the microbial partner (Madsen *et al.*, 2011). NFR1, NFR5 and other LysM-RLKs encode multiple extracellular LysM domains and a cytoplasmic kinase domain for signal transduction. Like NFR5, NFR1 is essential for nodulation and mutants fail to develop infection threads (Radutoiu *et al.*, 2003). The NFR5 kinase domain has consistently been demonstrated to be catalytically inactive, untill the recent discovery of NiCK4 (NFR5-interacting cytoplasmic kinase 4), which binds and phosphorylates the cytoplasmic domains of NFR5 (Wong *et al.*, 2019). Interestingly, the closest homolog of NFR1 in the model plant species *Arabidopsis thaliana* is the *At* CERK1, a major chitin elicitor receptor kinase for fungal and bacterial pathogens (Liu *et al.*, 2012 ; Petutschnig *et al.*, 2010 ; Miya *et al.*, 2007). Although the ligand binding sites of *At* CERK1 are known, how NFR1 distinguishes between symbiotic and non-symbiotic signals is still unknown.

NFR5 and NFR1 are not the only LysM-RLKs involved in root-nodule symbiosis. The Lys1 or NFRe is an epidermal LysM-RLK participating in the NFR1-NFR5 mediated signalling pathway in a spatio-temporal manner (Murakami *et al.*, 2018). Another example is the Exopolysaccharide Receptor 3 or EPR3 (previously known as Lys3), a LysM-RLK that recognizes Rhizobium-secreted exopolysaccharides with structural similarities to nod factors (Kawaharada *et al.*, 2015). However, unlike *nfr1* or *nfr5*, *epr3* mutants are not completely defective in infection thread initiation, but nevertheless display a reduction in nodulation events (Kawaharada *et al.*, 2015). Recently, two novel immune receptors, Lys6 and Lys12, were identified in *Lotus japonicus* (Bozsoki *et al.*, 2017 ; Fuechtbauer *et al.*, 2017). In total, the genome of *L. japonicus* encodes 17 LysM-type RLKs, most of which are still uncharacterized (Lohmann *et al.*, 2010).

Another molecular signaling partner is the SymRK protein, an LRR-type RLK with malectin-like carbohydrate binding domain outside the plasma membrane. This protein is the fundamental symbiotic receptor-like kinase required for multiple plant-microbe symbioses including the ancient arbuscular mycorrhizal symbiosis (Stracke *et al.*, 2002 ; Markmann *et al.*, 2008 ; Gherbi *et al.*, 2008). SymRK co-operates with the NFRs and SymRK mutants are unable to produce any nodules (Stracke *et al.*, 2002). Overexpression of NFR1, NFR5 or SymRK initiates spontaneous nodule organogenesis, even in absence of the bacterium (Ried *et al.*, 2014). SymRK physically interacts with the NFRs and the extracellular region of SymRK is cleaved after perceiving symbiotic signals (Antolıén-Llovera *et al.*, 2014). However, considering its essential role for the arbuscular mycorrhizal symbiosis, sequence variation in the SymRK protein within the genus *Lotus* is not anticipated to be exceptionally high.

Here I investigated the evolutionary history and evaluated the evidence for natural selection in multiple of genes involved in plant-microbe discrimination from multiple *Lotus* species. For these genes (NFR5, NFR1, Lys3 or EPR3, NFRe or Lys1, Lys2, Lys4, Lys6, Lys11-16, Lys21, SymRK, HAR1 and Klavier) I estimated the ratio of non-synonymous (*Ka*) to synonymous (*Ks*) substitutions (or ω) to gauge the evolutionary constraint across the gene, using the well-developed phylogenetic program PAML (Yang, 2007). Values of $\omega > 1$ indicate positive selection, while an $\omega = 0$ is found under extreme purifying section. Values of ω around 1 are typical in the absence of selection, such as for pseudogenes. Our evolutionary and phenotypic analyses identified new candidate sites in NFR5 that may be associated with microbe discrimination.

Our data highlighted that RLKs receiving symbiotic effectors have experienced dispro-

portional evolutionary constraints on ligand binding and signal transduction domains. It also revealed that the ligand binding domain of LysM-type symbiotic RLKs are adaptively evolving, while those from the LRR-type symbiotic RLKs are evolving neutrally. The signaling efficient kinase domains of symbiotic receptor like kinases are largely maintained by purifying selection. We also showed that the section Lotus shows species-specific symbiotic interactions and this is decided early during the interaction. The phylogenetic position of *L. conimbricensis* indicates that *M. loti* compatibility within section Lotus has possibly re-emerged from a *Bradyrhizobium* preferring ancestor. Remarkably, the NFR5 and the NFR1 emerged as the best candidate genes among the sixteen RLKs analyzed to contribute in species-specific discrimination of Rhizobia.

Chapter 2

Materials and Methods

2.1 Biological Materials

A total of 50 different *Lotus* species originating from mainly Europe were retrieved from seed banks in Germany (`http://www.ipk-gatersleben.de`), the U.S.A. (`http://www.ars-grin.gov`), the UK (`http://www.kew.org`), Czech Republic (www.vupt.cz) and through personal communication with Dr. Dario I. Ojeda Alayon (`http://darioi.weebly.com`). The *M. loti* DsRed strain MAFF303099 was kindly provided by Dr. Takaki Maekawa, MPI Cologne, Germany (Maekawa *et al.*, 2009). The *Bradyrhizobium* sp. (strain NZP2309) was purchased from the Belgian coordinated collection of microorganisms (`http://bccm.belspo.be`). *E. coli* Top10 cells were used for regular cloning purposes.

2.2 Analysis of species relationships

DNA was extracted from leaf tissues using DNeasy Plant Mini extraction kit (Qiagen, Hilden, Germany). I used three standard phylogenetic markers: rbcL, matK3, and ITS to infer the relationships among the 50 *Lotus* species. The rbcL primer set amplified a 524 bp region of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene. The matK3 primers amplified a 867 bp region of the maturase K gene. Both rbcL and matK3 are chloroplast encoded genes and represent mainly the coding region. The ITS primers amplified an approximately 649 bp region consisting of the non-coding internal transcribed spacer 1, internal transcribed spacer 2 and the coding region of 5.8S ribosomal RNA gene. PCRs were performed using the phusion high-

fidelity polymerase enzyme (New England Biolabs) and the lengths of amplicons were confirmed on a 1% agarose gel. The PCR products were purified with peqGOLD Cycle-Pure Kit (Peqlab, Erlangen, Germany) and sequenced with Sanger sequencing facilities of Eurofins (Ebersberg, Germany). Sequences were aligned with MAFFT (Katoh and Standley, 2013) and the rbcL, matK3 and ITS alignments were concatenated using FASconCAT-G_v1.02 resulting in a 2040 bp sequence alignment (Kück and Longo, 2014). The maximum likelihood tree was inferred using the phylogenetic software PhyML and 1000 bootstrap replicates were computed (Guindon *et al.*, 2010).

2.3 Inoculation of *Lotus* plants

Seeds were scarified using sandpaper and surface sterilized for 10 minutes with 10% (w/v) bleach containing 0.1% (w/v) SDS solution. The surface sterilized seeds were rinsed four times with sterile dH2O and incubated at 4°C overnight. The seeds were then transferred to plates containing minimum nitrogen medium Fahraeus Plant (FP) and sealed with parafilm (Fahraeus, 1957). The plates were then covered with aluminum foil to exclude light and kept in the dark for three days in a nearly upright position. 500 ml glass jars (Weck, Wehr, Germany) containing 300 ml of seramis (Mars) and 50 ml of FP buffer were autoclaved. Five plantlets (three-day old) were placed in each glass jar and inoculated with bacteria in 50 ml of FP buffer to achieve a final optical density (OD600) of 0.05. For the mock inoculation, we used sterile FP buffer. The glass jars were closed with cut-open rubber gaskets and the cuts were sealed with permeable 2M micropore tape to allow for airflow. The nodulation phenotypes of the plants were assessed at four weeks post inoculation. Growth conditions were 24°C/8°C at 16-h-light/8- h-dark cycles. The nodulation phenotypes of the plants were assessed categorically at four weeks post inoculation as developed nodules or developing primordia. General developmental parameters including shoot-height and shoot-weight were also documented. For the inoculation experiments, the *M. loti* strain MAFF303099 labeled with DsRed was grown in liquid YEM media with 10 mg/mL of Gentamicin for 4-6 days at 200 rpm to reach the desired optical density of 1 to 1.2 (Vincent, 1970). The *Bradyrhizobium* sp. (strain NZP2309) was also grown in liquid YEM media with 2.5 µg/mL of Tetracycline for 4-6 days at 200 rpm to reach the desired optical density.

2.4 Phylogenetic analysis of LysM and LRR-type symbiotic receptors

Gene specific primers were designed based on the *L. japonicus* MG20 genome version 1.0 (`http://www.kazusa.or.jp/lotus/`) at approximately 200 bp upstream of the start codon and 200 bp downstream of the stop codon (Table 6). A Blast search of the *L. japonicus* genome was used to ensure that the primer was specific to the targeted gene. The following 19 genes were analyzed in this study: NFR1, NFR5, Lys3 or EPR3, Lys1 or NFRe, SymRK, HAR1, Klavier, Clavata2, LHK1, Lys2(Lys2 left-border only - L2LB), Lys4, Lys6 (AA #34 - stop codon), Lys11, Lys12, Lys13, Lys14, Lys15, Lys16 and Lys21. PCRs were performed using the phusion high-fidelity polymerase enzyme (New England Biolabs) and lengths of amplicons were confirmed on 1% agarose gel before subsequent cloning into the TOPO-blunt cloning vector (Life Technologies). The clones were selected on LB plates with 50 µg/mL Kanamycin (Bertani, 1951). For longer genes (i.e. NFR1 and Lys2), additional primers were designed at the middle of the gene and the left side and right side were cloned separately with approximately 50 bp overlap. Sequencing was done with Sanger sequencing approach by Europhins (Ebersberg, Germany) using the M13Forward and M13Reverse universal primers and additional sequencing primers where necessary (Table 7). Sequences of LysM-RLK genes were assembled using the Sequencher software (`http://genecodes.com/`). All insertion/deletions (indels) and substitutions were examined by eye. The cDNA coding sequences were extracted based on the *L. japonicus* MG20 genome version 1.0 gene models. The resultant cDNA and protein sequences of individual genes were aligned with MAFFT and the cDNA alignments were converted to codon alignments based on the protein alignment using PAL2NAL (Suyama *et al.*, 2006).The phylogenetic trees used for PAML analysis are presented in Figure 4, Figure 5, Figure 9 and Figure 21-37. The phylogenetic trees were inferred from nucleotide sequences for each individual gene using maximum likelihood and the branch labels were removed using a custom pearl script (available on request). The edited tree was used in the maximum likelihood based phylogenetic program PAML for evolutionary analyses of individual genes and identification of adaptively evolving sites (Yang, 2007).

I used the random-sites codon substitution models implemented in the CodeML program of PAML to detect evidence of positive selection (Yang *et al.*, 2000). Two pairs of site-specific models were used, M1a (nearly neutral) versus M2a (selection) and M7 (neutral, beta) versus M8 (selection, beta & ω). In these comparisons, M1a

and M7 neutral models (null hypothesis) do not include positive selection, while M2a and M8 alternative models allow for positive selection. Additionally, I also used the model M0 (one ratio) that assumes the same ω ratio for all sites. The branch-sites models were used to detect adaptive evolution of individual codons along specific lineages (Zhang *et al.*, 2005).

The domain structures were predicted on-line (`http://smart.embl-heidelberg.de/`) by the SMART domain prediction service (Schultz *et al.*, 1998). The fixed-sites models were used to detect differences in selection pressure between the extracellular and intracellular domains (Yang and Swanson, 2002). The codon frequency model F3x4 was fitted to all the alignments.

2.5 Genotype-phenotype association analysis

The R package Genphen was used to evaluate the association between the amino acid sequences and the average number of *M. loti* nodules (Kitanovski, 2016). I used a phenotypic metadata composed of three independent experiments done with the *M. loti* - DsRed following the exact same protocol (please see the inoculation experiment 1, 2 and 3 presented in Figure 19 & 20). I calculated the average number of developed nodules for each species from the data presented in Figure 19 & 20 and disregarded the developing primordia data. I also disregarded the nodulation data presented in Figure 3. I calculated Cohen's d for the 596 amino acid sites of NFR5 based on the average number of *M. loti* nodules of 24 *Lotus* species. For other genes I used variable number of genotypes common between the Figure 19 & 20 and individual gene tree. The *Lotus* genotypes used for association analysis are highlighted on individual gene tree presented in Figure 4, 5, 9 & 21-37. Sites showing significant association with the average number of nodules (p-value < 0.05) are listed in Table 5.

2.6 Data visualization

I used the R package ggplot2 for visualization of the phenotypic data. The phylogenetic trees were visualized and annotated with the R package ggtree (Yu *et al.*, 2017). Genotype-phenotype association between individual sites and the average number of nodules were visualized using the "plotSpecificGenotype" function of the R package Genphen. The I-TASSER web server was used to determine the homology modeling of protein domains. The resulted models were visualized with PyMol.

Chapter 3

Results

3.1 Two well differentiated subclades within the genus Lotus

We assembled a collection of 50 different *Lotus* species supplied from four national seed-bank collections and our colleagues (Table 16). Our collection captures a range of genetic, ecological and life-history diversity, ideal for examining the evolutionary history of host-symbiont specificity (Figure 2).

To understand the evolution of symbiont specificity, I first determined the evolutionary relationships among the sampled host species. I amplified and sequenced three standard DNA regions used in phylogenetic analysis: ITS, matk3, and rbcL from these plants. Our phylogenetic analysis, based on nucleotide sequence, is largely congruent with previous analyses of this group (Degtjareva *et al.*, 2006 ; Allan *et al.*, 2004). Two major clades can be distinguished (Figure 1). Clade A includes 23 species, dominated by the species-rich taxonomic section *Lotus* (17 species). Clade A species were mainly sampled from Europe, but also included a small number of species from Asia. Clade A includes two well-known European representatives, *L. corniculatus* and *L. pedunculatus*, previously characterized by differences in their symbiotic preferences (Gossmann *et al.*, 2012). Clade B includes 25 species collected from locations in Europe, Africa and Australia. Clade B is more diverse as it includes species from numerous taxonomic sections including *Pedrosia* (12 species), *Loteae* (5 species), *Heinekenia* (2 species), and *Krokeria* (1 species). Specialized island endemics including *L. campylocladus*, *L. dumetorum* and *L. spartioides* and a critically endangered species,

L. kunkelii, also belong to Clade B. The tree was rooted with two species of the section *Anisolotus* collected from South America. The sampling of a similar number of taxa from clade A and B provides an excellent opportunity to investigate the evolution of species-specificity in *Lotus-Rhizobium* symbiosis within a comparative phylogenetic framework.

3.2 The genus *Lotus* exhibits clade-specific symbiotic preferences

I evaluated the compatibility of 24 *Lotus* species with *Mesorhizobium loti* constitutively expressing DsRed. Most *Lotus* species developed characteristic nodules with *M. loti* (Figure 19 & 20). However, a few species, such as *L. castellanus*, *L. conimbricensis*, *L. halophilus*, and *L. pedunculatus*, failed to do so. These species instead showed compatibility with *Bradyrhizobium* sp., initially isolated from a *L. pedunculatus* nodule (Figure 3).

The species compatible with *M. loti* were: *L. alpinus*, *L. angustissimus*, *L. arenarius*, *L. corniculatus*, *L. drepanocarpus*, *L. gebelia*, *L. glinoides*, *L. krylovii*, *L. mearnsii*, *L. preslii*, *L. weilleri*, among others. These *M. loti* compatible hosts were from both Clade A and Clade B (Figure 1). The *Bradyrhizobium*-compatible taxa included: *L. castellanus*, *L. conimbricensis*, *L. halophilus*, and *L. pedunculatus*. However, a range of nodulation phenotypes was observed across species, from exclusive symbiont specificity, in which host plants formed nodules with only one symbiont, to less discriminating hosts, in which host plants formed nodules with both symbionts. For example, *L. pedunculatus* and *L. castellanus* formed white primordia with *M. loti* and functional pink nodules with *Bradyrhizobium* (Figure 3). The species *L. arabicus*, *L. cytisoides*, *L. ornithopodioides*, and *L. tenuis* also responded modestly to the *M. loti* inoculation (Figure 19 & 20). In contrast, *L. conjugatus* and *L. conimbricensis* did not respond to inoculation with *M. loti* and only few or no infection threads (ITs) were observed at 7dpi (Figure 39). *L. conimbricensis* and *L. halophilus* preferentially formed distinctive nodules with *Bradyrhizobium* and failed to induce nodulation with *M. loti*. The phylogenetic position of the monophyletic *Mesorhizobium* compatible clade (or Meso-clade) embedded within the paraphyletic *Bradyrhizobium* compatible clade (or Brady-clade) of section *Lotus* may indicate that the *Mesorhizobium* preference has re-emerged from a *Bradyrhizobium* preferring ancestor (Figure 1).

3.3 The evolutionary pattern at NFR5 correlates with rhizobial host specificity

The *NFR5* gene has been shown to play an important role in rhizobial preference (Radutoiu *et al.*, 2007). Therefore, I evaluated the evolutionary history of this gene across the Lotus genus. The *NFR5* gene could be amplified and sequenced from 50 individuals. The gene genealogy of *NFR5* roughly recapitulates the species phylogeny (Figure 4). The same major clades (Clade A, Clade B, Meso-clade, and Brady-clade) are well supported in the *NFR5* gene tree.

To take a closer look at the role of natural selection in shaping the evolution of *NFR5*, the phylogenetic program PAML was used, which can distinguish between different models of evolution. Using the random-sites models, we could identify individual positions in NFR5 under positive selection (Nielsen and Yang, 1998 ; Yang *et al.*, 2000). The random-sites selection model (M8) provided a statistically significant better fit to the NFR5 data compared to the null model without positive selection (M7), implying that some sites within the *NFR5* gene have experienced positive selection (Table 2; n=50, LRT= 39.86, p-value < 0.001).

The BEB statistics implemented in PAML revealed six codon sites of the *NFR5* gene as positively selected (probability $> 95\%$). Four of these sites are found in the extracellular LysM2 domain of NFR5 : sites L118, D123, E153 and V155 (Figure 6a). The codon site, L118, had the highest ω value (ω= 2.801, probability $> 99\%$) and this site has previously been identified to be involved in distinguishing between symbiotic rhizobial species (Radutoiu *et al.*, 2007). This codon site, together with the D123 (ω= 2.709), controls the rhizobial specificity by putative interaction with the chitin backbone of nod factors (Figure 45).

Furthermore, I identified sites V155 (ω= 2.797, probability $> 99\%$) and E153 (ω= 2.782) as another hot spot of protein evolution. Variation at position V155 was correlated with rhizobial specificity within the genus Lotus (Figure 40). In our study, all alleles from the Meso-clade encode a Valine at site 155, while alleles from the Brady-clade encode either a Leucine or Isoleucine. This results in a significant association between protein variation and microbe preference (Table 5; Cohen's d = 3.53). Sites E153 and V155 are located in close proximity to the putative ligand-binding region of the extracellular domain of NFR5 and possibly represents another specificity determining region. Specifically, these amino acids of NFR5 putatively interact with the nod factors secreted by the *Rhizobium* (Figure 45). The two other

adaptively evolving sites of NFR5 extracellular domain are sites K30 and T51 (ω= 2.712 and 2.716, respectively). Among those, variation at site T51 was correlated with *M. loti* specificity (Table 5; Cohen's d = -0.88).

In contrast to the extracellular domain, ω was greater than one at only a few sites in the NFR5 cytoplasmic kinase domain, and none of these had a likelihood greater than 95%. Furthermore, sites classified as neutrally evolving were more abundant in the extracellular regions compared to the kinase domain, indicating that evolutionary constraints may have differed on either side of the transmembrane region. Nevertheless, site Y371 and site P399 of the NFR5 kinase domain showed strong association with *Mesorhizobium* compatibility (Table 5; Cohen's d = -2.01 and 2.14, respectively).

3.4 Evidence of adaptive evolution in the extracellular domain of NFR1

NFR1 is another gene involved in symbiont recognition in *Lotus* (Radutoiu *et al.*, 2003). I evaluated the evolution of this gene from these plant species as well. However, this gene is much longer due to the presence of eleven introns in *L. japonicus*. Therefore, I amplified and analyzed the 5' and 3' parts of the *NFR1* gene separately. The NFR1 left-border (N1LB) corresponds to the extracellular LysM domains and the NFR1 right-border (N1RB) corresponds to the cytoplasmic kinase domain.

The random-sites model in PAML revealed that the NFR1 kinase domain (N1RB) is under strong purifying selection and functionally well conserved across different *Lotus* species (Table 2; n=37, LRT= 2.79, p-value = 0.249). In contrast, the NFR1 extracellular domains (or N1LB) showed signatures of positive selection (Table 2; n=36, LRT= 23.08, p-value < 0.001). Three codon positions in the NFR1 extracellular region have an ω-value significantly greater than one (Figure 6b). However, no strong association between variation at these putatively adaptive sites (I45, N61, and S107) and nodulation phenotype were detected (Table 5).

A radically different amino acid residue at site I45 is present in the alleles of NFR1 from both Clade A and Clade B that can be correlated with few or no nodule after *M. loti* inoculation (Figure 5 & 41). This makes this residue an interesting candidate for species specificity. Furthermore, the allele of NFR1 from *L. conimbricensis* carried two amino acid insertions at position Y67 and K68, which are absent in all other 35 *Lotus* species evaluated. However, the main feature of NFR1 is that it shows a more

conserved kinase domain compared to that of NFR5, underscoring the importance of NFR1 kinase domain for the downstream signal transduction and initiation of the symbiotic process.

3.5 The form and strength of natural selection differs between the extracellular and intracellular regions of NFR1 and NFR5

I used the fixed-sites models to determine whether the form and strength of natural selection differed between a *priori*-defined regions of NFR5 and NFR1 (Yang and Swanson, 2002). In this model the following parameters were estimated from the data: ω *(Ka/Ks)* ratio and κ (transition/transversion) ratio. The two regions I compared were the 5' gene region that encodes the signal peptide plus the extracellular receptor domain ($\omega_{\text{Extracellular}}$) and the 3' gene region encoding the transmembrane domain plus the intracellular kinase domain ($\omega_{\text{Intracellular}}$). The null model, model C, assumes that the ω values and their distribution are the same for both regions (i.e, $\omega_{\text{Extracellular}} = \omega_{\text{Intracellular}}$). In contrast, Model E allows for the ω values and their distribution to differ depending on domain. Model E fits the NFR5 sequence data significantly better than Model C in which the intracellular and extracellular domains were constrained to have the same distribution of ω values (Figure 13, Table 3, $\omega_{\text{Extracellular}} > \omega_{\text{Intracellular}}$, LRT = 61.01, p-value < 0.00001). The estimated ω value for the extracellular region of NFR5 was higher than the ω of the intracellular region (Figure 13; Table 3; $\omega_{\text{Extracellular}} = 0.54$, $\omega_{\text{Intracellular}} = 0.23$), revealing a higher rate of amino acid evolution in the extracellular region. In an analogous manner, I analyzed the sequence variation for the two NFR1 regions. Here as well, the model allowing the intracellular and extracellular regions to have different ω values fit significantly better than the model constraining the two regions to have the same ω values (Table 3; $\omega_{\text{Extracellular}} > \omega_{\text{Intracellular}}$, LRT = 100.79, p-value < 0.00001).

Despite its lower ω (Ka/Ks ratio), the synonymous rate of substitution was higher at the NFR1 intracellular region (or N1RB) compared to the NFR1 extracellular region (or N1LB) (Table 1). This indicates that mutations continue to accumulate both sides of the plasma membrane. However, natural selection is effectively distinguishing between nonsynonymous and synonymous mutations, favoring nonsynonymous mutations in the N-terminal extracellular region and removing nonsynonymous mutations in the

C-terminal intracellular region via purifying selection.

3.6 Specific lineages within Lotus show adaptive evolution at NFR1 and NFR5 associated with divergence in symbiont-specificity

The PAML branch-sites models were used to evaluate whether a lineage specific change in symbiont preference was correlated with a shift in the evolutionary rate parameter, ω, in the *NFR1* and *NFR5* genes. The branch-sites model examines whether positive selection may have acted on an a *priori* defined lineage, or foreground branch, assuming a common $\omega <= 1$ for all background branches (Yang and Nielsen, 2002 ; Zhang *et al.*, 2005). For each gene, I evaluated two possible pre-defined foreground branches (or lineages) corresponding to the following clades of alleles: Clade A and within that, the Meso-clade. Clade A contains species that nodulate with either *Bradyrhizobium* and *Mesorhizobium* and the Meso-clade is a smaller clade of 14 taxa and includes species that preferentially nodulate with *Mesorhizobium*.

The Likelihood ratio tests of the branch-site model on the NFR5 gene tree revealed that the ω-values for the foreground branches (Clade A and Meso-clade) are significantly elevated over the background branches (Figure 4, Table 4). This is consistent with an evolutionary shift specifically associated with these two lineages. The inferred ω-value for the Meso-clade is $\omega_{\text{Meso}} = 11.24$, while the inferred ω-value for the Clade A is $\omega_{\text{Clade-A}} = 6.07$. The higher ratio for the Meso-clade may indicate that positive selection has been especially strong on this lineage, perhaps linked to greater symbiont discrimination and specificity.

The lineage specific analysis of the extracellular domain of NFR1 fell just short of statistical significance (Table 4, $\omega_{\text{Clade-A}} = 4.18$, p-value $= 0.069$). However, in this case the signature of positive selection was lower along the lineage leading to the Meso-clade ($\omega_{\text{Meso}} = 2.27$, p-value $= 0.67$) as compared to Clade A ($\omega_{\text{Clade-A}} > \omega_{\text{Meso}}$). In summary, these analyses allowed us to not only pinpoint specific residues within the NFR1 and NFR5 genes that have been likely targets of adaptive evolution, but to also identify whether specific evolutionary lineages (such as those sharing a common symbiont preference) experienced an accelerated rate of protein evolution.

3.7 No strong evidence of adaptive evolution in the Exopolysaccharide Receptor 3 (EPR3)

I evaluated the evolutionary history of ten LysM-RLKs, in addition to NFR1 and NFR5 in our collection of *Lotus* species. According to the random-sites analyses of these genes, the model that allows for some sites to have experienced positive selection fits better than a model excluding positive selection (Table 2, Figure 12). Two exceptional genes were: the recently characterized Exopolysaccharide Receptor 3 (EPR3, previously called Lys3) and the putative pseudogene Lys21. For these two genes, the model with and without positively selected sites fit the data equally well (Table 2). In EPR3, only two sites in the extracellular domain (sites P14 and M22) showed weak signals of adaptive evolution (Figure 7a).

Overall, purifying selection appears to be the predominant form of natural selection at EPR3 (Figure 7a). However, according to the branch-sites analysis, the ω value was higher for Meso lineage ($\omega_{\text{Meso}} = 7.14$) compared to Clade A lineage ($\omega_{\text{Clade-A}} = 1$), indicating a difference in evolutionary constraints along these two lineages (Table 4). As observed for NFR1 and NFR5, selective constraints are stronger in the intracellular domain of EPR3, resulting in a statistically significant difference between intracellular and extracellular values of ω ($\omega_{\text{intracellular}} = 0.09534$, $\omega_{\text{extracellular}} = 0.27257$, Table 3, Figure 13) . A moderate correlation between sequence variation at EPR3 and *Mesorhizobium loti* compatibility data may point to a role of EPR3 in *Rhizobium*-specificity (Table 5, Figure 42).

In contrast to EPR3, the gene wide mean ω for Lys21 is higher and more similar to the values observed for Lys1, Lys4, Lys 6, Lys11, Lys12, Lys15 and Lys16 (Figure 12). None of the Lys21 sites were identified as candidates for positive selection (Table 2, Figure 15b). There is also no evidence that selective constraints differed significantly on either side of the plasma membrane ($\omega_{\text{intracellular}} = 0.24983$, $\omega_{\text{extracellular}} = 0.37004$, Table 3, Figure 13). Taken together, this gene stands out compared to other LysM genes, by not showing any evidence for adaptive evolution in the extracellular domain.

3.8 The kinase domain of the immune receptor Lys12 shows signatures of adaptive evolution

Since I observed that selective constraints in the extracellular and cytoplasmic domains of NFR1, NFR5 and EPR3 differed, I evaluated whether this pattern also applied to other members of the LysM-RLKs gene family. I observed that the ω ratio was significantly higher in the extracellular regions compared to the cytoplasmic regions for three additional genes: Lys6, Lys 13 and Lys14 (Table 3, Figure 13). In contrast, the ω ratios for the intracellular and extracellular regions did not differ significantly for Lys4, Lys11 and Lys12 (Table 3, Figure 13). Lys6 and Lys12 have been characterized as immune receptors, rather than required for rhizobial discrimination (Bozsoki *et al.*, 2017 ; Fuechtbauer *et al.*, 2017). Adaptively evolving residues were identified in both proteins. In Lys6, the four putative adaptive sites were in the extracellular domain (Table 2, Figure 8a), while in Lys12 one site falls in the signal peptide (C3) and one in the kinase domain (L513)(Table 2, Figure 8b). Unlike the NFR1 and EPR3, sites in the kinase domains of Lys4, Lys11, Lys12, Lys13 and Lys15 displayed evidence of positive selection (Table 2, Figure 8b, Figure 14a, Figure 15a, Figure 16a, Figure 17a). Furthermore a statistically different ω ratio was observed for the Clade A lineage for the following genes: Lys11, Lys13, Lys14 and Lys16 (Table 4).

3.9 Less evidence of positive selection in the extracellular LRR domains of SymRK, Klavier and HAR1

We examined the evolutionary history of three LRR-type RLKs (SymRK, HAR1 and Klavier) involved in *Lotus-Rhizobium* symbiosis. A model allowing for positive selection did not fit the data significantly better than one lacking positive selection for SymRK, Klavier or HAR1 (Table 2). No putative adaptively evolving residues were identified for SymRK or Klavier (Table 2, Figure 10a & 11b). The rate of amino acid evolution is strongly constrained in all three genes, especially in the kinase region (Table 3, Figure 10a & 11). The ω values differed significantly between the intracellular and extracellular regions for all three genes (Table 3, Figure 13). However, despite lack of evidence for adaptive evolution, amino acid variation in SymRK is

correlated with *Mesorhizobium* compatibility (Figure 9, Table 5). The sites, P475 and S374, (Cohen's d = 2.2 and 1.92, respectively) exhibited strong association with *Mesorhizobium* compatibility. The HAR1 gene showed a weak signal of adaptive evolution, but also shows the highest rate of synonymous evolution out of all the genes analyzed (Table 1 & 2, Figure 11a and Figure 38). Clavata2 is a receptor-like protein without any kinase domain and differed from the three LRR-RLKs analyzed. It showed strong evidence of adaptive evolution, with five sites identified as putative targets of positive selection (S6, P8, I81, L126 and N127)(Table 2, Figure 18). The branch-sites analysis indicates that there has been a shift in evolutionary rates in Clavata2 along the branch subtending the Meso-clade (Table 4, Figure 25).

3.10 The two transmembrane domains of LHK1 (cytokinin receptor) show signatures of positive selection

Given that almost all the LysM and LRR-type RLKs with known contribution to the legume-*Rhizobium* symbiosis displayed elevated *Ka*/*Ks* ratios in the extracellular regions, I evaluated the pattern in the well-known *Lotus* cytokinin receptor LHK1. The LHK1 is a histidine kinase with two transmembrane domains flanking an extracellular CHASE domain, responsible for binding cytokinin ligand. A model postulating adaptive evolution at LHK1 fit the data significantly better than one without (Table 2). Most of the putatively adaptively evolving residues (sites P40, F46, A54, H58, and Y335) were localized to the transmembrane regions (Table 2, Figure 10b). The ω ratio did not statistically differ between the intracellular and extracellular regions (Table 4, Figure 13). No lineage specific rate change for the branch leading to Clade A was detected for the LHK1 gene (Table 4, Figure 26) and protein variation at LHK1 did not correlate with a *Mesorhizobium* preference (Table 5, Figure 44).

Table 1: Basic evolutionary measurements by CodeML model 0

Gene or Region	**n**	**Fragment**	**Codons**	Ka	Ks	ω	κ	**Tree length**	**lnl**
NFR5	50	1869	596	0.27	0.76	0.36	2.12	1.20	-6852.85
NFR1 left border (N1LB)	36	2507	320	0.23	0.56	0.41	1.64	0.91	-2970.80
NFR1 right border (N1RB)	37	3434	328	0.04	0.81	0.05	2.49	0.65	-2542.85
Lys3 or EPR3	31	5150	620	0.12	0.75	0.16	1.92	0.82	-5639.54
Lys1 or NFRe	20	4660	600	0.08	0.24	0.33	2.58	0.35	-3866.43
Lys2 left border (L2LB)	11	2625	352	0.29	0.42	0.69	1.67	0.95	-3232.80
Lys4	28	4345	642	0.18	0.58	0.31	2.33	0.84	-5917.66
Lys6	15	6795	590	0.12	0.45	0.27	2.15	0.62	-4362.39
Lys11	27	1804	596	0.21	0.50	0.41	2.43	0.84	-5594.08
Lys12	22	1989	667	0.10	0.30	0.33	2.59	0.46	-4441.36
Lys13	35	2201	673	0.42	1.47	0.29	1.83	2.06	-10755.35
Lys14	29	2260	672	0.34	1.31	0.26	1.79	1.77	-9165.00
Lys15	21	1849	603	0.03	0.14	0.23	2.45	0.18	-3186.99
Lys16	25	2165	683	0.14	0.47	0.31	1.86	0.67	-5562.44
Lys21	21	2705	497	0.13	0.42	0.32	2.01	0.62	-3945.15
Clavata2	26	2433	725	0.10	0.46	0.22	2.59	0.57	-5658.84
HAR1	25	4851	987	0.12	1.48	0.08	2.16	1.33	-11004.20
Klavier	15	3511	1152	0.06	0.29	0.19	2.16	0.33	-6961.15
SymRK	42	5689	923	0.16	0.52	0.30	2.39	0.75	-8144.22
LHK1	13	6483	1003	0.07	0.27	0.27	1.94	0.36	-6241.62
NFR1 (Concatinated)	32	NA	624	0.15	0.76	0.20	1.83	0.85	-5488.44

Note:

n = Number of species

Fragment = Cloned fragment size (bp) based on the *Lotus japonicus* gemone

Codons = Number of codon triplets in an alignment

Ka = Non-synonymous substituton rate

Ks = Synonymous substitution rate

$\omega = Ka/Ks$

κ = Transition-Transversion Ratio

Tree length = Tree length calculated by CodeML model 0

lnl = log-likelihood of CodeML model 0

Table 2: Identifying positive selection via CodeML M2a and M8

	M1a(neutral) & M2a(selection)				**M7(beta) & M8(beta with ω)**				
Gene	**M1a**	**M2a**	$2\Delta lnl$	$p-$**value**	**M7**	**M8**	$2\Delta lnl$	$p-$**value**	**M8 sites**
NFR5	-6775.07	-6760.73	28.68	5.92e-07	-6782.43	-6762.50	39.86	2.21e-09	K30*, T51*, L118**, D123*, E153*, V155**
NFR1 left border (N1LB)	-2932.41	-2921.02	22.77	1.14e-05	-2932.51	-2920.97	23.08	9.74e-06	I45**, N61**, S107*
NFR1 right border (N1RB)	-2533.08	-2533.08	0.00	1.00e+00	-2534.51	-2533.12	2.78	2.49e-01	
Lys1 or NFRe	-3837.41	-3815.55	43.73	3.19e-10	-3838.36	-3815.76	45.19	1.54e-10	F19*, Q37*, Q53**, R69**, E97**, E103**, S265*
Lys3 or EPR3	-5560.52	-5558.86	3.32	1.90e-01	-5563.46	-5558.86	9.21	1.00e-02	P14*, M22*
Lys2 left border (L2LB)	-3190.14	-3169.89	40.50	1.61e-09	-3190.61	-3170.11	40.99	1.26e-09	N53**, V82**, H83**, T84**, E110*, T120**, T140*
Lys4	-5846.55	-5837.94	17.21	1.83e-04	-5848.93	-5837.67	22.51	1.29e-05	M112* , T130*, Q456*, T492*, L587*
Lys6	-4310.68	-4303.32	14.71	6.39e-04	-4311.33	-4303.15	16.37	2.80e-04	S34*, Q37**, N150*, A197*
Lys11	-5542.73	-5536.60	12.27	2.17e-03	-5544.76	-5536.91	15.70	3.91e-04	S213*, H214*, N306*, E526*
Lys12	-4412.60	-4403.62	17.96	1.26e-04	-4412.61	-4403.67	17.89	1.31e-04	C3**, N513*
Lys13	-10403.18	-10374.84	56.67	4.95e-13	-10413.88	-10376.16	75.45	2.20e-16	T105*, K167*, F170*, A171*, P207**, D108**, A269*, V291*, P312**, A313*, S318**, Q605**
Lys14	-8949.27	-8941.35	15.85	3.62e-04	-8953.94	-8938.33	31.23	1.65e-07	T16*, M17*, P31*, S32**, K43*
Lys15	-3166.38	-3160.44	11.88	2.63e-03	-3166.85	-3160.44	12.81	1.65e-03	T16*, Q29**, S463**, K527*
Lys16	-5504.15	-5493.01	22.28	1.45e-05	-5506.55	-5494.40	24.31	5.27e-06	L53**, S133**, L219*, S267*
Lys21	-3919.62	-3918.57	2.11	3.48e-01	-3919.43	-3918.28	2.31	3.15e-01	
SymRK	-8096.70	-8096.79	0.18	9.15e-01	-8097.92	-8096.19	3.46	1.77e-01	
HAR1	-10875.85	-10875.85	0.00	1.00e+00	-10871.37	-10866.72	9.30	9.55e-03	R4*, K474*, A531*
Klavier	-6943.17	-6941.78	2.79	2.48e-01	-6943.63	-6941.61	4.04	1.33e-01	
Clavata2	-5591.72	-5572.08	39.28	2.96e-09	-5593.17	-5572.33	41.68	8.89e-10	S6**, P8**, I81*, L126**, N127**
LHK1	-6192.79	-6181.05	23.48	7.97e-06	-6193.37	-6181.20	24.35	5.15e-06	F5*, P40*, F46*, A54**, H58**, Y335*

Note:
$2\Delta lnl$ = Likelyhood Ratio Test
df=1
M1a = Likelihood-ratio of Model 1a
M2a = Likelihood-ratio of Model 2a
M7 = Likelihood-ratio of Model 7
M8 = Likelihood-ratio of Model 8
Sites = Significant sites identified by Model 8 (* indicates significant at 5% and ** indicate sighificant at 1%)

Table 3: PAML fixed-sites test of Intra- and Extra-cellular parts

RLKs	**Gene**	**Models**	#	ω_{extra}	ω_{intra}	κ_{extra}	κ_{intra}	**lnl**	$2\Delta lnl$	$p-value$
LysM	NFR5	Model C	100	0.3581	0.3581	2.1051	2.10510	-6839.260	33.5594	1.00e-07
		Model E	102	0.5406	0.2320	1.9263	2.31934	-6822.480		
	NFR1 (Concatinated)	Model C	64	0.2003	0.2003	1.7485	1.74845	-5617.991	100.7887	0.00e+00
		Model E	66	0.4701	0.0785	1.4950	2.37463	-5567.597		
	Lys3 or EPR3	Model C	62	0.1603	0.1603	1.9116	1.91156	-5627.970	29.4076	4.00e-07
		Model E	64	0.2726	0.0953	1.8444	1.92824	-5613.266		
	Lys1 or NFRe	Model C	40	0.3356	0.3356	2.5398	2.53984	-3836.820	1.7254	4.22e-01
		Model E	42	0.2874	0.4173	2.3060	2.85660	-3835.957		
	Lys12	Model C	44	0.3300	0.3300	2.5912	2.59120	-4385.525	1.1099	5.74e-01
		Model E	46	0.3106	0.3581	2.8258	2.32867	-4384.970		
	Lys6	Model C	30	0.2682	0.2682	2.1373	2.13730	-4299.009	24.4250	5.00e-06
		Model E	32	0.3805	0.1404	1.8351	2.89060	-4286.796		
	Lys11	Model C	54	0.4109	0.4109	2.4254	2.42537	-5573.048	0.7146	7.00e-01
		Model E	56	0.4269	0.3962	2.2755	2.57399	-5572.691		
	Lys13	Model C	70	0.2604	0.2604	1.9300	1.93001	-10723.305	25.9338	2.30e-06
		Model E	72	0.3551	0.1996	1.9308	1.91163	-10710.338		
	Lys14	Model C	58	0.2422	0.2422	1.9053	1.90525	-9118.788	41.4106	0.00e+00
		Model E	60	0.3447	0.1661	1.7218	2.10930	-9098.083		
	Lys15	Model C	42	0.2341	0.2341	2.4584	2.45840	-3178.098	0.7063	7.03e-01
		Model E	44	0.2808	0.2023	2.4038	2.48725	-3177.745		
	Lys16	Model C	50	0.3047	0.3047	1.8645	1.86453	-5549.494	10.2784	5.86e-03
		Model E	52	0.4253	0.2284	1.8095	1.87062	-5544.355		
	Lys21	Model C	42	0.3374	0.3374	2.2901	2.29005	-3944.133	4.8911	8.67e-02
		Model E	44	0.3700	0.2498	1.8734	2.25730	-3941.688		
	Lys4	Model C	56	0.3119	0.3119	2.3243	2.32433	-5915.602	0.6034	7.40e-01
		Model E	58	0.3328	0.2887	2.3763	2.26806	-5915.301		
LRR	SymRK	Model C	84	0.2971	0.2971	2.3567	2.35669	-8114.948	68.2547	0.00e+00
		Model E	86	0.4599	0.1005	2.5722	2.10459	-8080.820		
	Klavier	Model C	30	0.1879	0.1879	2.1189	2.11888	-6945.478	22.1350	1.56e-05
		Model E	32	0.2367	0.0548	2.0846	2.43707	-6934.410		
	HAR1	Model C	50	0.0743	0.0743	2.2161	2.21608	-10969.034	61.0105	0.00e+00
		Model E	52	0.0994	0.0242	2.1435	2.45778	-10938.529		
HK	LHK1	Model C	26	0.2644	0.2644	1.9403	1.94029	-6234.878	4.9927	8.24e-02
		Model E	28	0.1697	0.3032	1.9623	1.93002	-6232.382		

Note:

Model = Fixed-sites model used by PAML (E or C)

= Number of parameters used by PAML

$\omega_{extra} = Ka/Ks$ or ω of Extracellular part plus the Signal Peptide Region

$\omega_{intra} = Ka/Ks$ or ω of Cytoplasmic part plus the Transmembrane Region

κ_{extra} = Transition-Transversion Ratio (κ) of Extracellular part plus the Signal Peptide Region

κ_{intra} = Transition-Transversion Ratio (κ) of Cytoplasmic part plus the Transmembrane Region

lnl=Log-Likelyhood

$2\Delta lnl$ = Likelyhood Ratio Test; df=2

Table 4: PAML branch-site models comparison of Lotus clades

Gene or Region	**Foreground**	**Foreground** ω	**Neutral** lnl	**Selection** lnl	$2\Delta lnl$	$p-$**value**
NFR5	Clade-A	6.06920	-6697.881	-6688.216	19.329	1.10e-05
	Meso	11.24000	-6699.052	-6696.116	5.872	1.54e-02
NFR1 left Border (N1LB)	Clade-A	4.17690	-2860.805	-2859.151	3.307	6.90e-02
	Meso	2.26860	-2860.440	-2860.351	0.177	6.74e-01
Lys3 or EPR3	Clade-A	1.00000	-5483.824	-5483.824	0.000	1.00e+00
	Meso	7.14610	-5484.914	-5484.137	1.555	2.12e-01
Lys1 or NFRe	Clade-A	NA	NA	NA	NA	NA
	Meso	16.42691	-3655.638	-3653.257	4.761	2.91e-02
Lys2 left border (L2LB)	Clade-A	NA	NA	NA	NA	NA
	Meso	14.07210	-3166.509	-3161.922	9.176	2.45e-03
Lys4	Clade-A	4.33770	-5703.781	-5702.698	2.167	1.41e-01
	Meso	1.00000	-5702.406	-5702.406	0.000	1.00e+00
Lys6	Clade-A	1.00000	-4296.719	-4296.719	0.000	1.00e+00
	Meso	1.00000	-4296.719	-4296.719	0.000	1.00e+00
Lys11	Clade-A	3.81240	-5447.290	-5438.895	16.790	4.18e-05
	Meso	7.37010	-5448.107	-5446.978	2.258	1.33e-01
Lys12	Clade-A	NA	NA	NA	NA	NA
	Meso	12.85320	-4241.581	-4224.256	34.650	3.95e-06
Lys13	Clade-A	2.66270	-10114.695	-10093.915	41.560	1.14e-07
	Meso	5.76190	-10121.853	-10121.131	1.444	2.30e-01
Lys14	Clade-A	4.76660	-7924.206	-7920.475	7.461	6.30e-03
	Meso	9.90360	-7924.114	-7921.630	4.968	2.58e-02
Lys15	Clade-A	NA	NA	NA	NA	NA
	Meso	7.00390	-3166.235	-3160.861	10.748	1.04e-03
Lys16	Clade-A	16.13480	-5323.423	-5311.564	23.718	1.11e-06
	Meso	2.30370	-5322.467	-5322.396	0.141	7.07e-01
Lys21	Clade-A	NA	NA	NA	NA	NA
	Meso	24.56240	-3820.493	-3812.511	15.963	6.46e-05
SymRK	Clade-A	1.00000	-7904.608	-7904.608	0.000	1.00e+00
	Meso	1.00000	-7905.325	-7905.325	0.000	1.00e+00
HAR1	Clade-A	NA	NA	NA	NA	NA
	Meso	1.51700	-10833.492	-10833.138	0.709	4.00e-01
Klavier	Clade-A	NA	NA	NA	NA	1.10e-01
	Meso	12.91360	-6717.167	-6715.889	2.558	1.10e-01
Clavata2	Clade-A	NA	NA	NA	NA	NA
	Meso	19.52310	-5587.570	-5582.542	10.057	1.52e-03
LHK1	Clade-A	1.00000	-6148.806	-6148.806	0.000	1.00e+00
	Meso	1.00000	-6148.755	-6148.755	0.000	1.00e+00

Note:

Neutral lnl = Likelihood-ratio of PAML branch-site neutral model

Selection lnl = Likelihood-ratio of PAML branch-site selection model

Foreground ω = Please see figure 4, 5, 21-37

$2\Delta lnl$ = Likelyhood Ratio Test

df = 1

The NA rows express lack of individuals for analysis

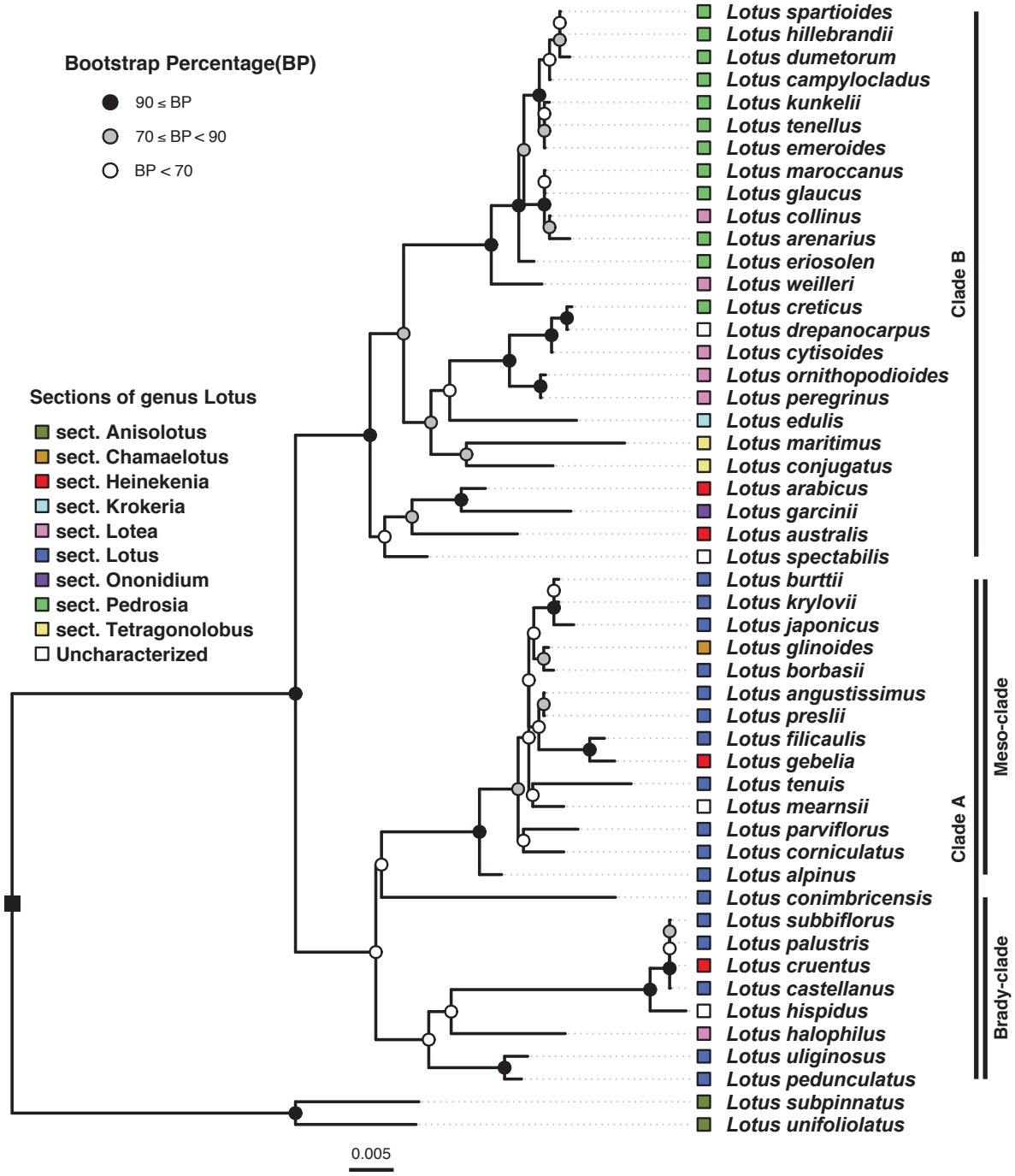

Figure 1. Phylogenetic relationship within the legume genus Lotus. Maximum likelihood phylogeny based on ITS, rbcL and matk3 sequences from 50 *Lotus* species. Bootstrap support from 1000 replicates are indicated by shaded circles on the nodes. The black box represents the placement of the root inferred using *Lotus subpinnatus* and *Lotus unifoliatus* as outgroups.

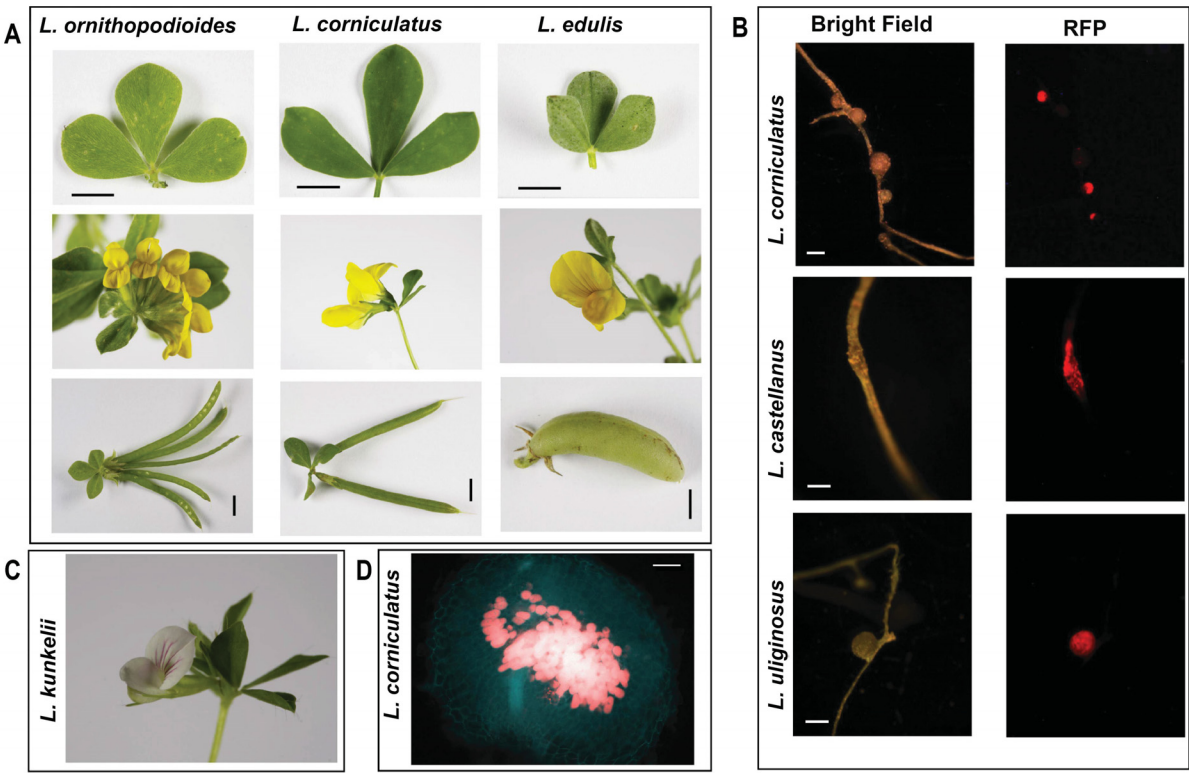

Figure 2. Morphological variation within the genus Lotus. (A) Variation in leaf, flower and fruit shape between *L. ornithopodioides*, *L. corniculatus* and *L. edulis* (scale = 5 mm). (B) Compatibility of *M. loti* DsRed with *L. corniculatus* (developed nodules), *L. castellanus* (developing primordia) and *L. uliginosus* (occasional white nodule) (scale = 0.5 mm) (C) A white flower of the endangered species *L. kunkelii*. (D) Cross-section of a *L. corniculatus* nodule colonized by *M. loti* constitutively expressing DsRed (scale = 100 um). The 75 um vibratome section was stained with 10 mg/L Calcofluor White stain (Sigma).

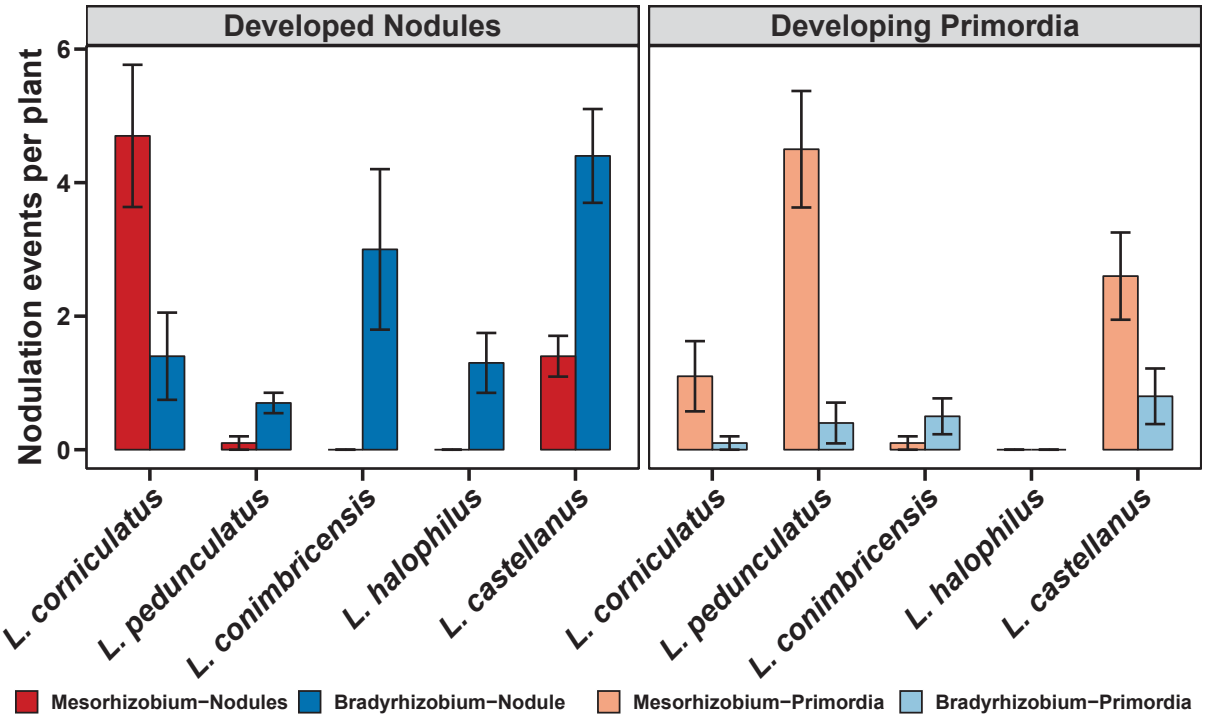

Figure 3: Compatibility of five *Lotus* species with two species of Rhizobia. Plants were grown in glass canning jars (Weck) on seramis and minimum nitrogen media. Nodulation events were scored at 4 weeks after infection (n=10, bars show SE) with *M. loti* (strain DsRed MAFF303099) or *Bradyrhizobium* sp. (strain NZP2309). The raw data and t-test results are presented in Table 14 and Table 15, respectively.

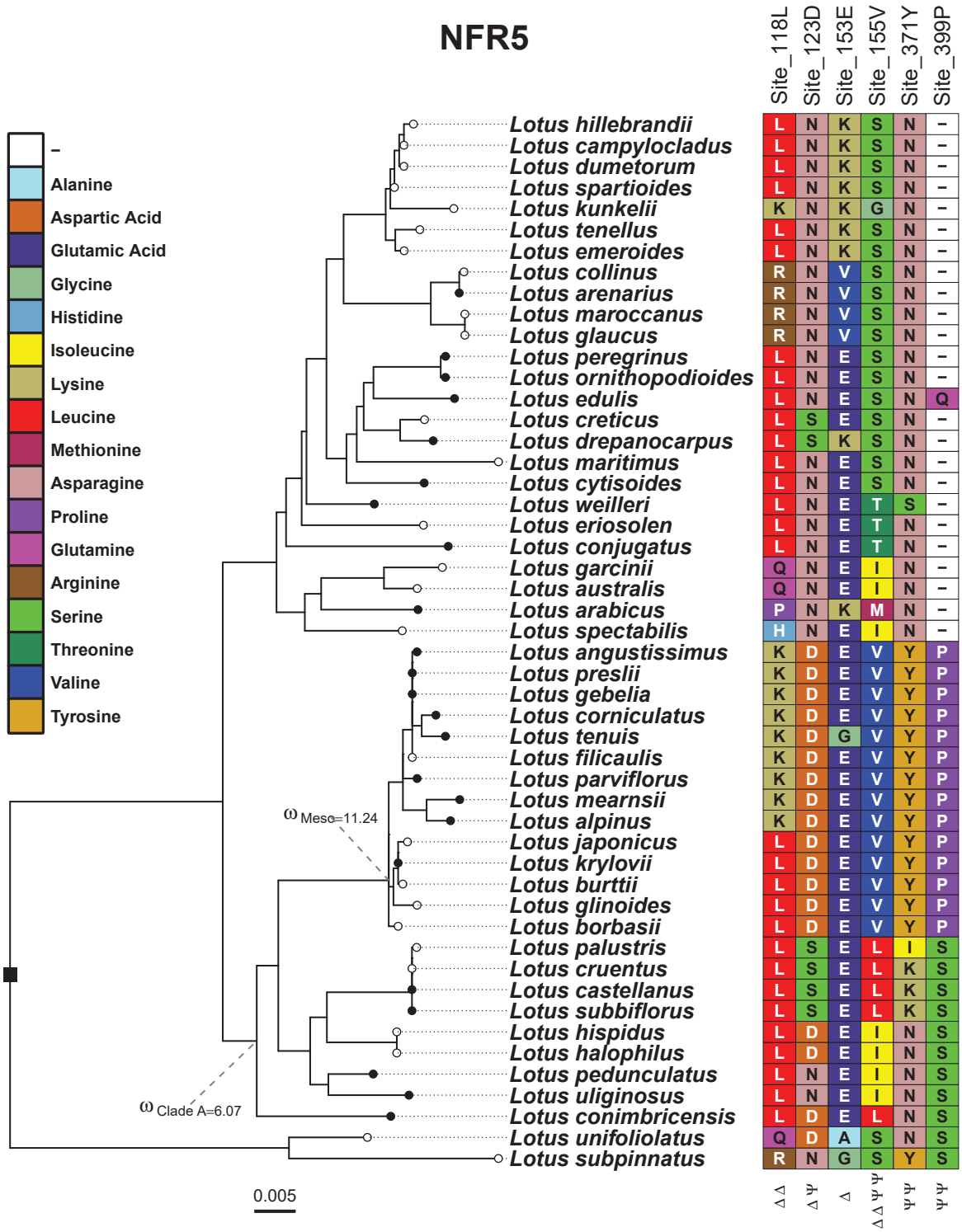

Figure 4. Phylogenetic reconstruction of alleles of NFR5 from 50 *Lotus* species. Maximum likelihood analysis of NFR5 nucleotide sequences. Mid-point rooting was applied. The black branch tips indicate species used in genetic association study with *M. loti* compatibility. The compatibility data are reported in figure 19

& 20. The columns on the right display the amino acid residue located at six sites identified either based on their strong association according to the Genphen analysis or for positive selection in the CodeML analysis. Ψ denotes significant sites in Genphen association analysis (Ψ = p-value < 0.05; $\Psi\Psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The foreground branches used in the CodeML branch-sites model are indicated with the dashed lines.

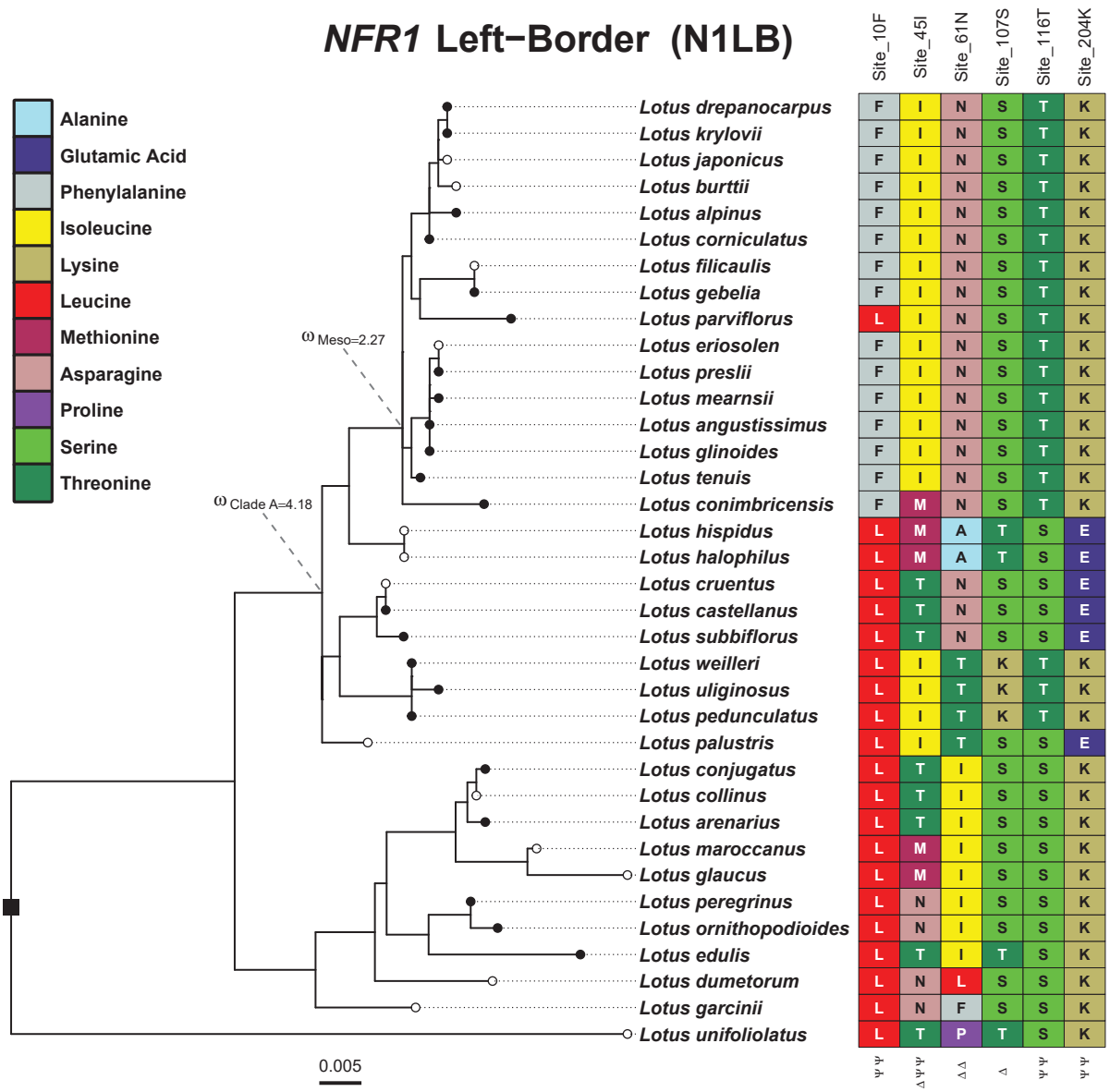

Figure 5. Maximum likelihood phylogeny based on NFR1 left-border (N1LB) sequences from 36 *Lotus* species. The black tips indicate species used in association study with *M. loti* compatibility data. The compatibility data are found in the figures 19 & 20. The tree was rooted using mid-point rooting. Ψ denotes significant sites in Genphen association analysis (Ψ = p-value < 0.05; ΨΨ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; ΔΔ = p-value < 0.01). The foreground branches used in the CodeML branch-sites model are indicated with the dashed lines.

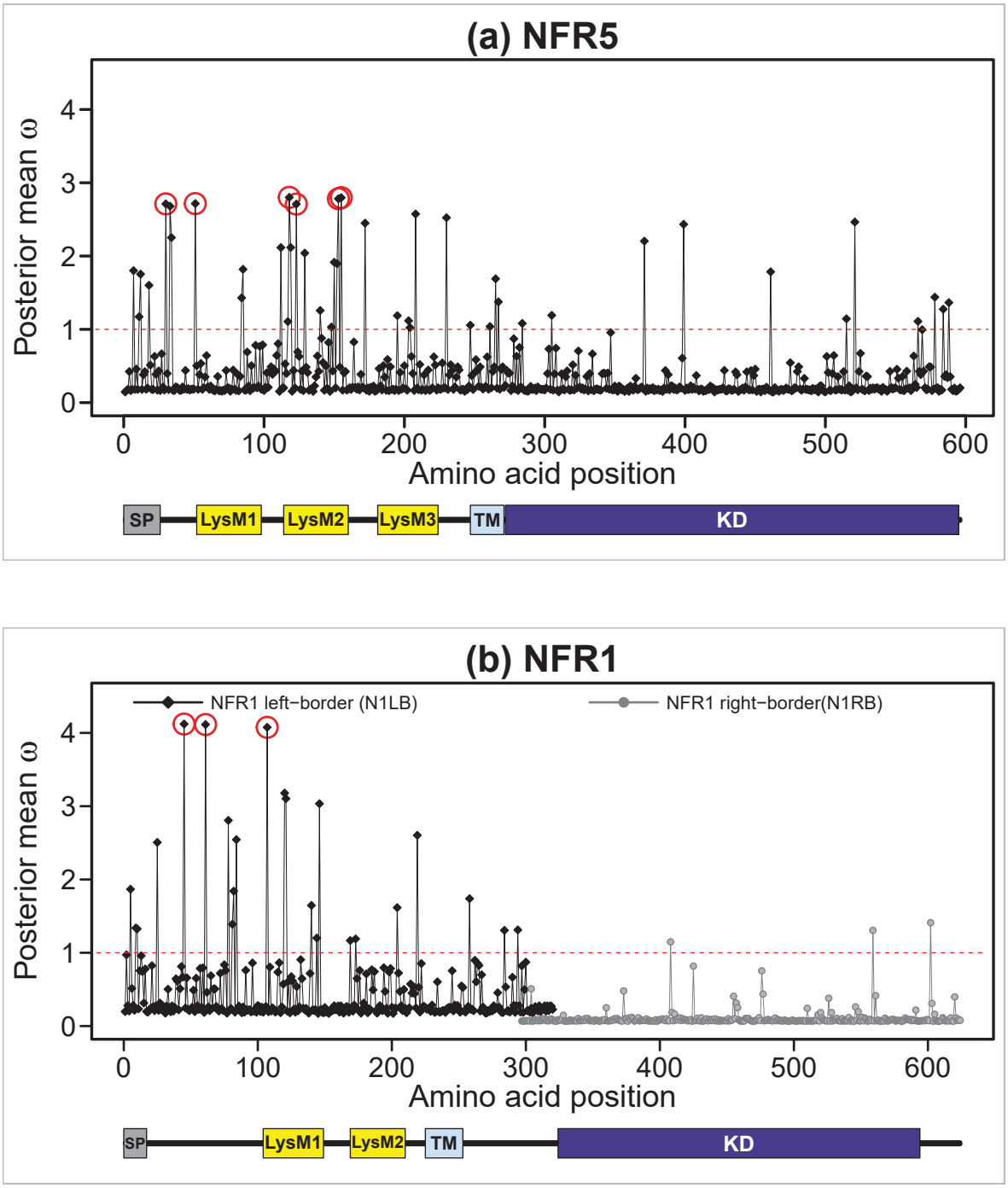

Figure 6. Distribution of posterior mean ω across NFR5 and NFR1. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to $\omega = 1$, expected under neutrality, and the predicted domain positions are plotted below individual figures. The NFR1 left and right borders were cloned, sequenced and analyzed separately.

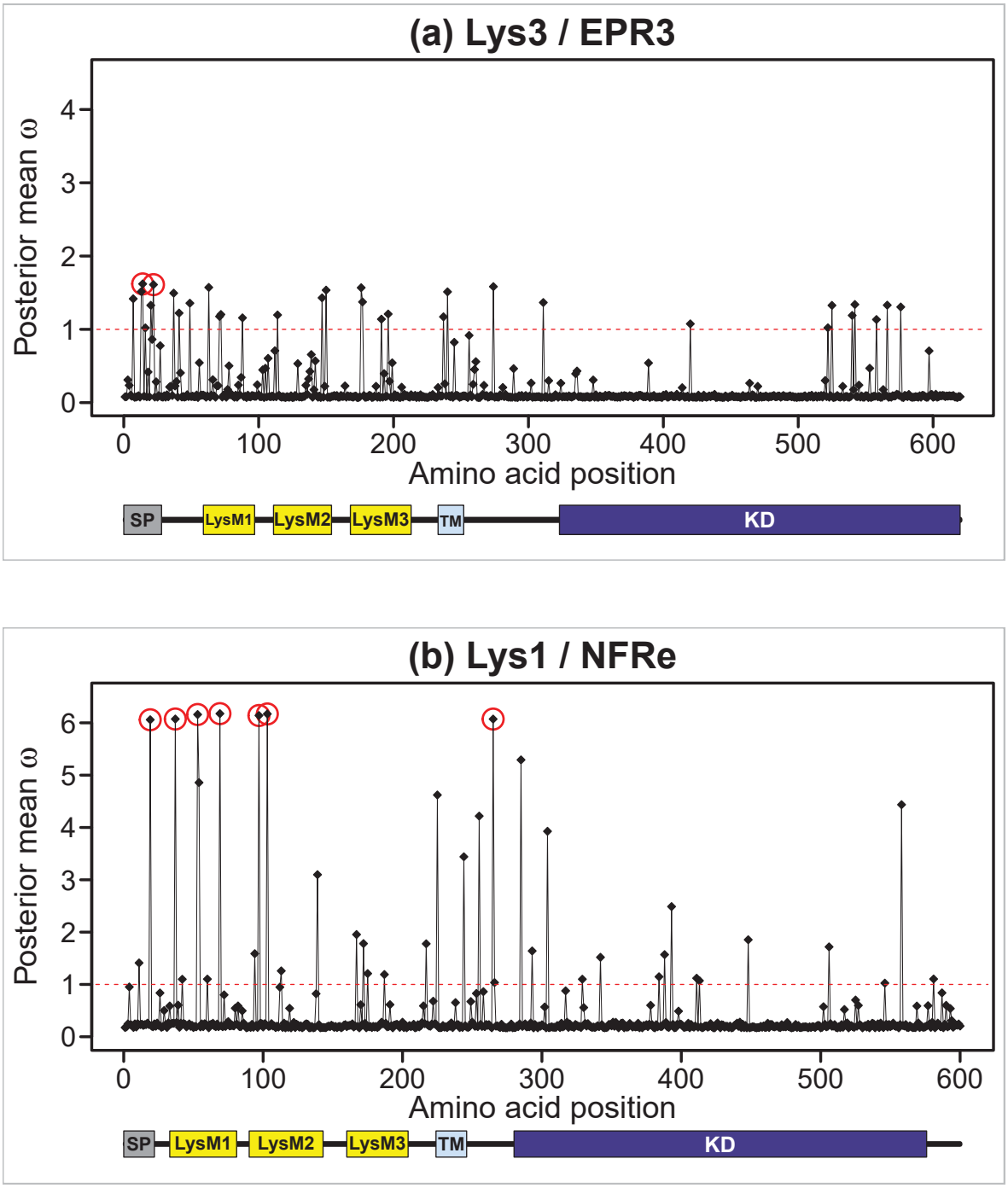

Figure 7. Distribution of posterior mean ω across Lys3/EPR3 and Lys1/NFRe. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to $\omega = 1$, expected under neutrality, and the predicted domain positions are plotted below individual figures.

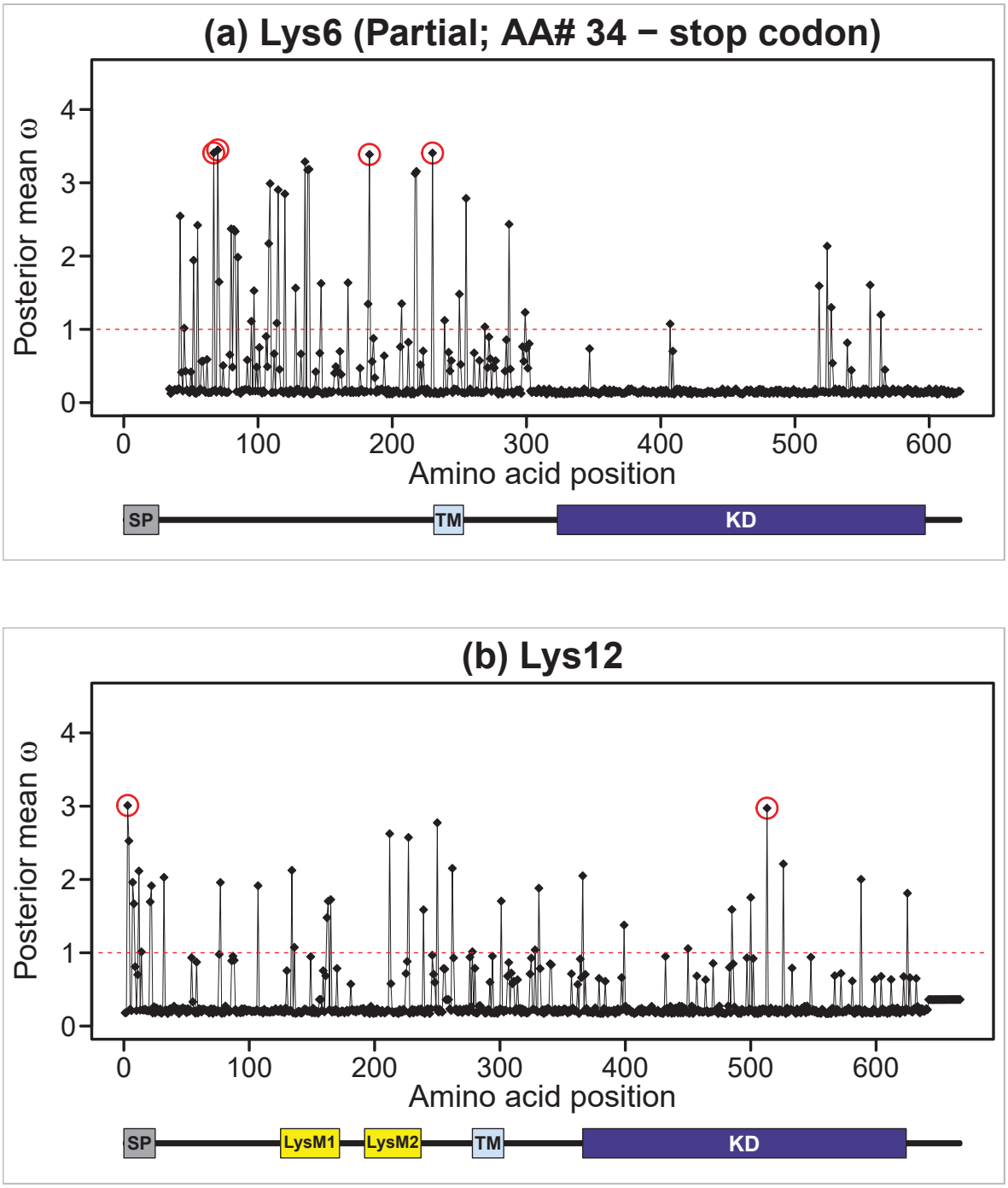

Figure 8. Distribution of posterior mean ω across Lys6 (partial, AA# 34 - stop codon) and Lys12. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to $\omega = 1$, expected under neutrality, and the predicted domain positions are plotted below individual figures.

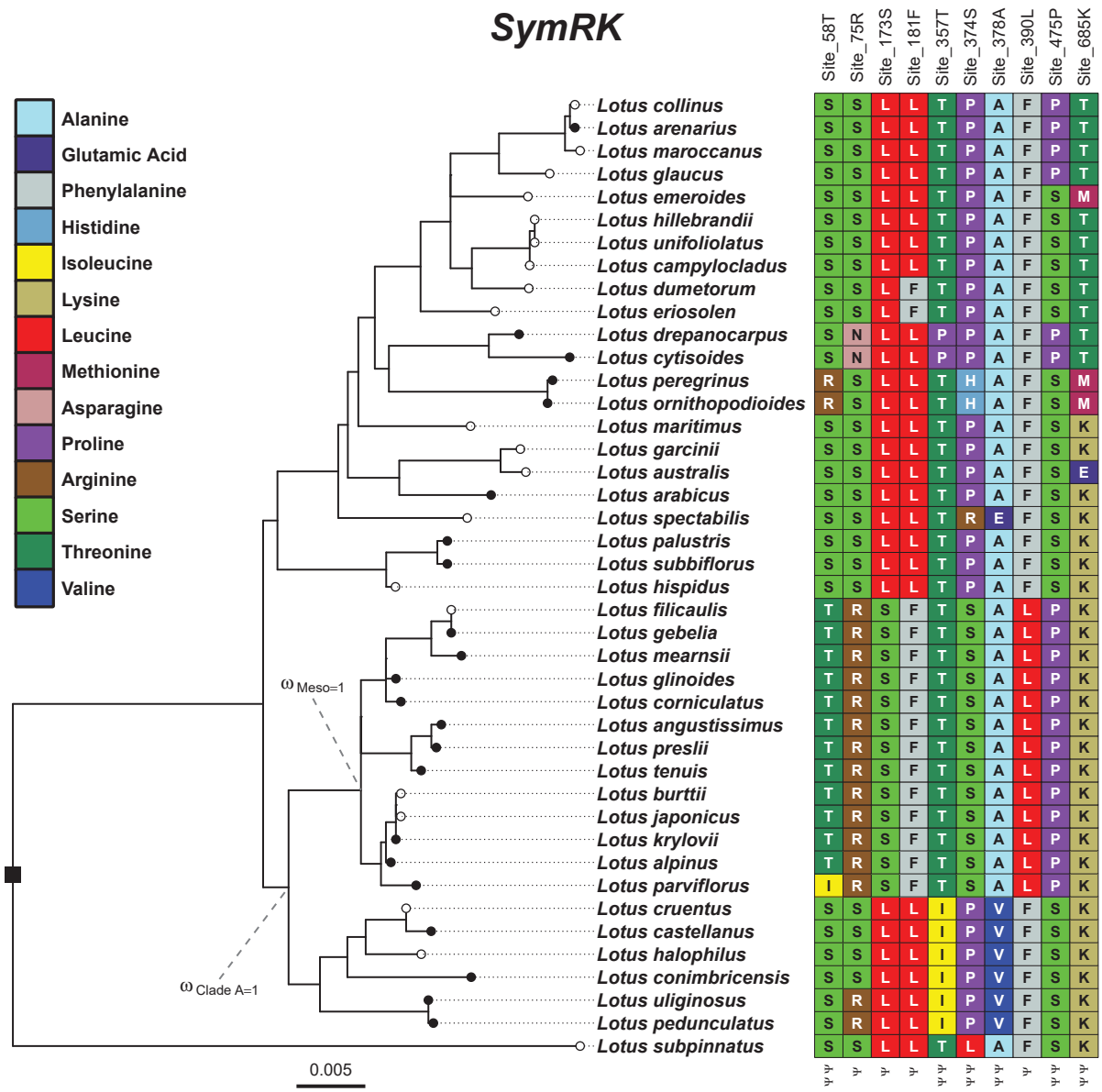

Figure 9. Maximum likelihood phylogeny based on SymRK sequences from 42 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figure 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

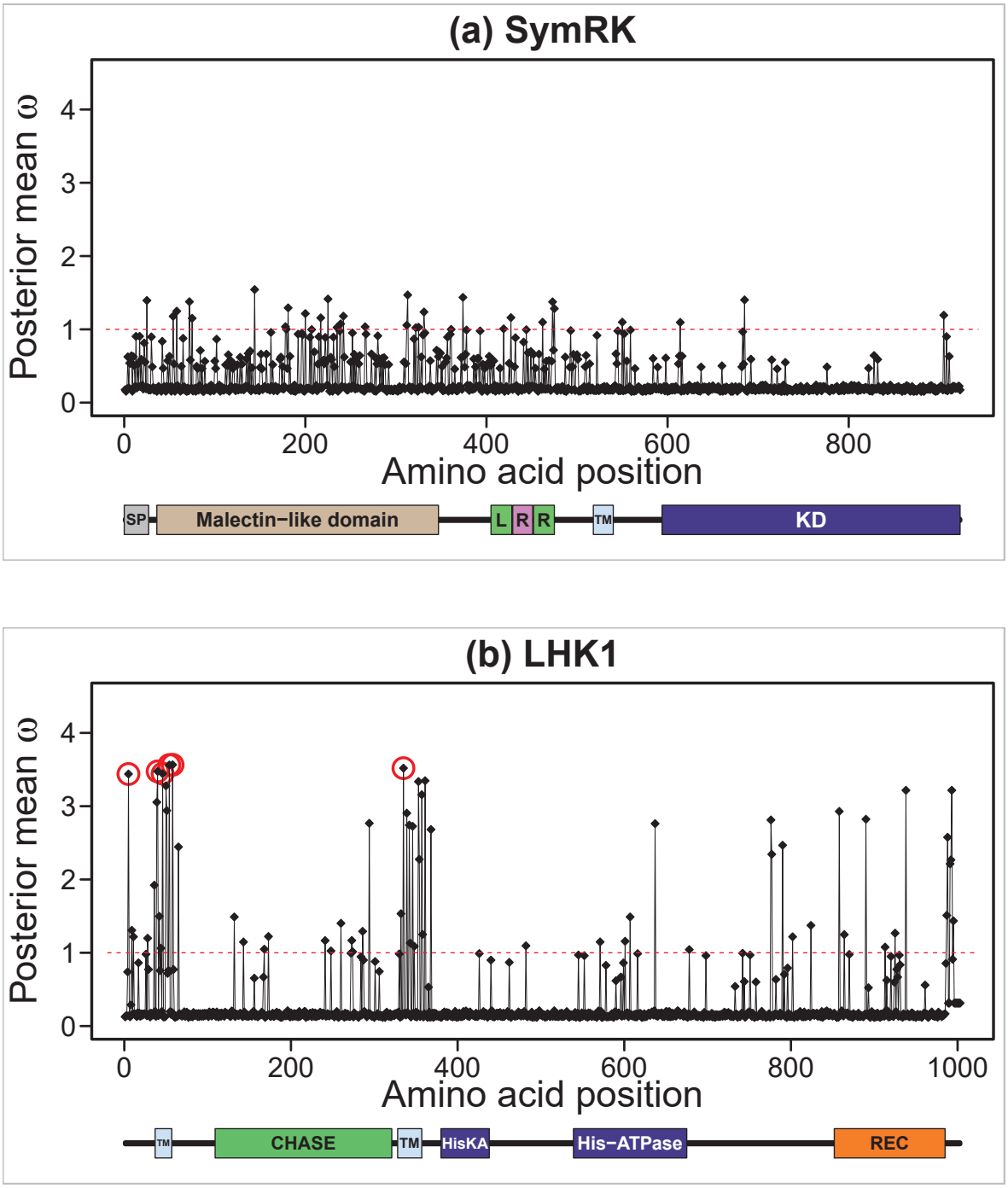

Figure 10. Distribution of posterior mean ω across the LRR-type symbiotic receptors-like kinase the SymRK and the cytokinin receptor LHK1. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to $\omega = 1$, expected under neutrality, and the predicted domain positions are plotted below individual figures.

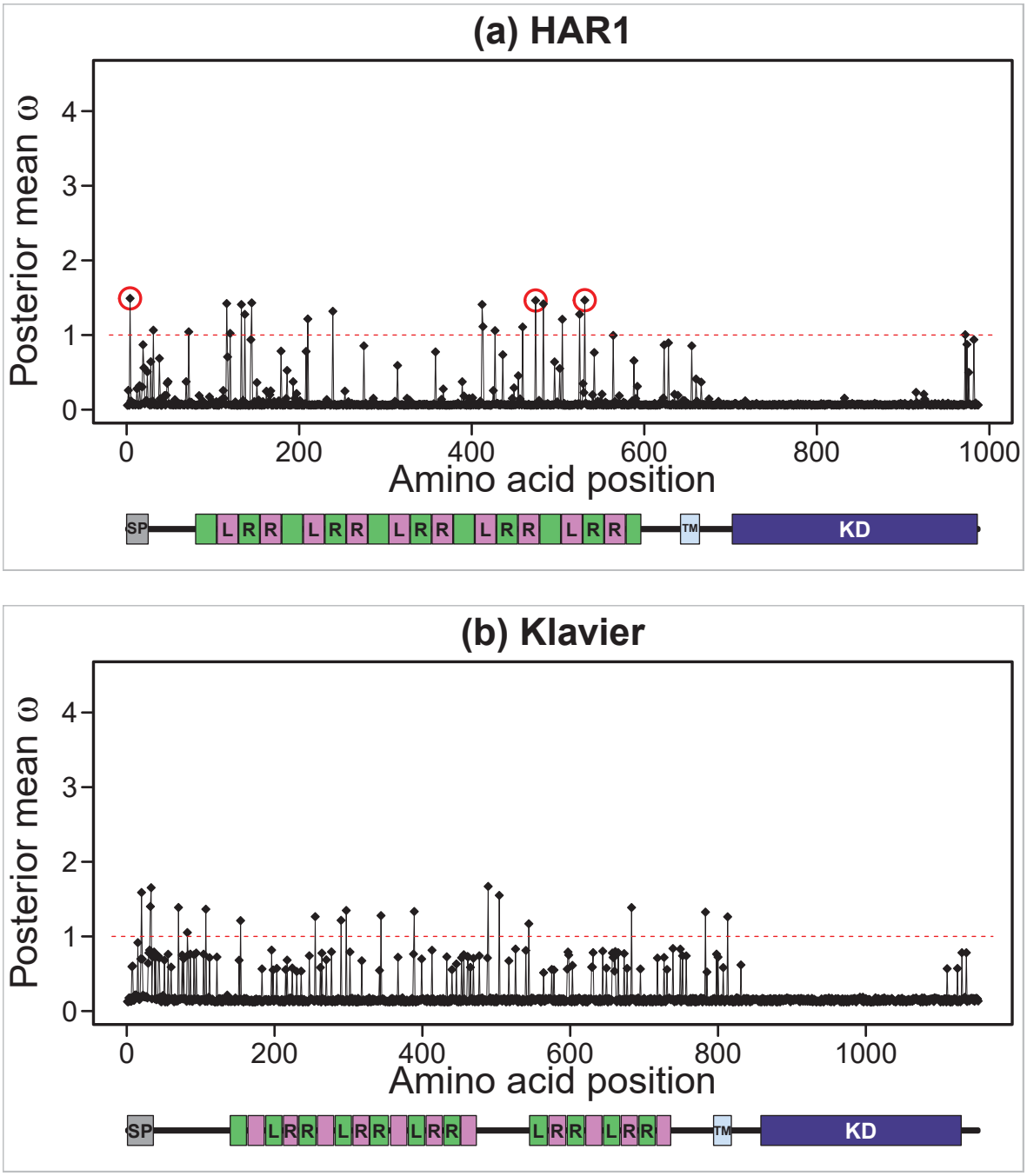

Figure 11. Distribution of posterior mean ω across the LRR-type symbiotic receptors-like kinases the HAR1 and Klavier. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to $\omega = 1$, expected under neutrality, and the predicted domain positions are plotted below individual figures.

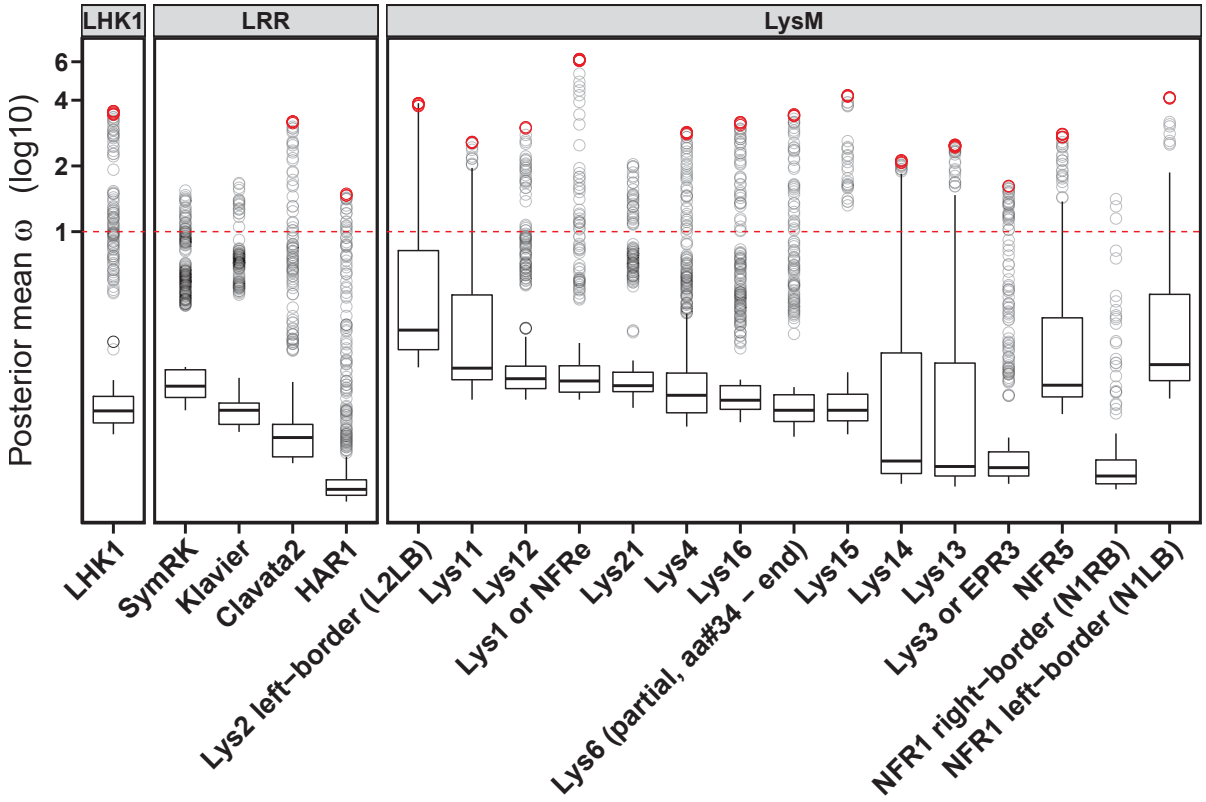

Figure 12. Distribution of posterior mean ω per codon across 20 symbiotic receptor-like kinases. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to ω=1, expected under neutrality.

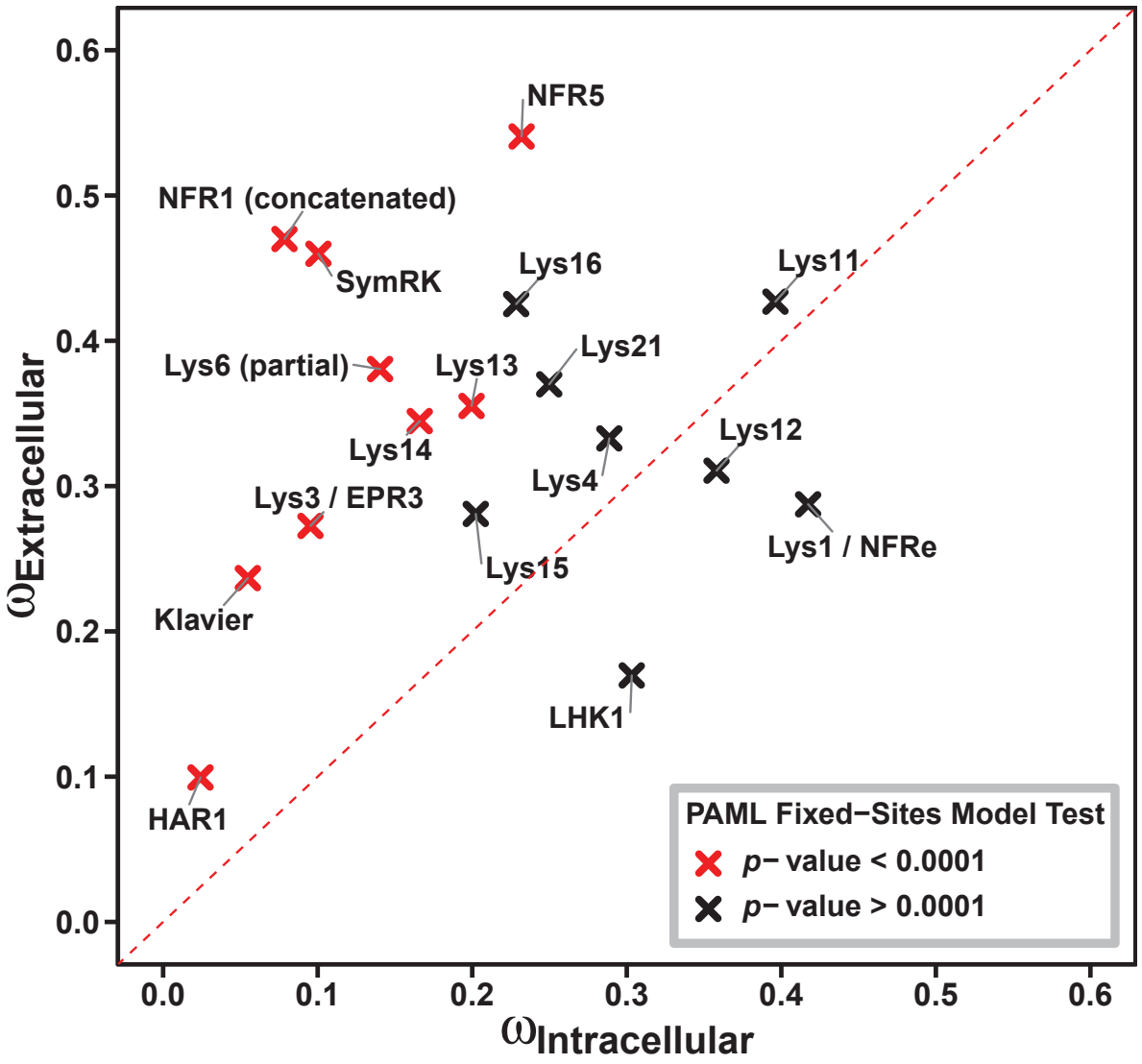

Figure 13. Comparison of ratio of substitution rates (ω) between ectodomain ($\omega_{\text{Extracellular}}$) versus transmembrane and cytoplasmic domains ($\omega_{\text{Intracellular}}$). The ratio of substitution rates (ω) was calculated using the Fixed-site model-E. Genes with significant likelihood ratio tests for Model C vs Model E are highlighted in red. These genes show a significantly elevated ω in the ectodomain compared to the transmembrane and cytoplasmic domains. The dotted red line indicate where $\omega_{\text{Extracellular}} = \omega_{\text{Intracellular}}$, indicating no difference in omega values depending on domain. See Table 3 for further details.

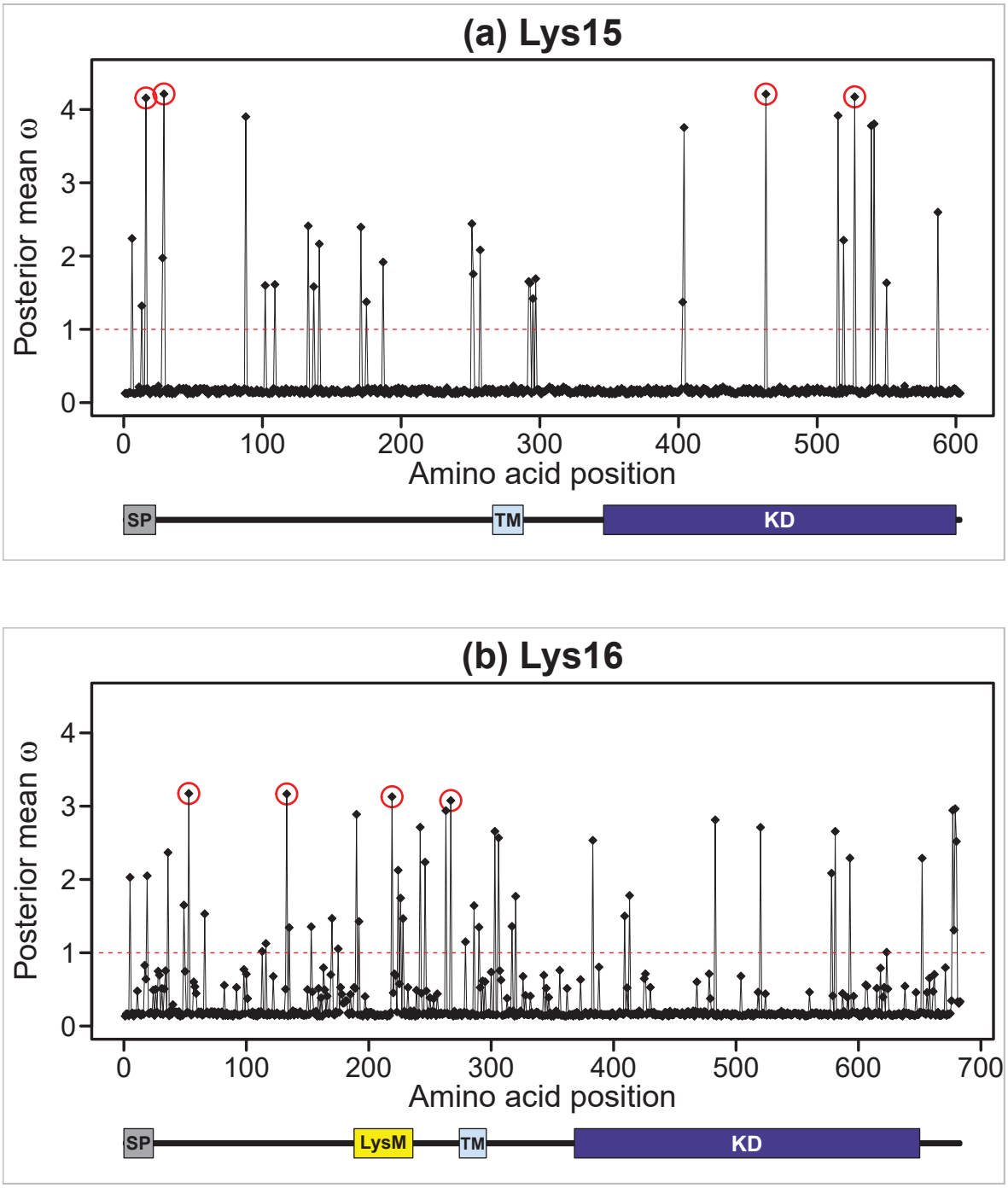

Figure 14. Distribution of posterior mean ω across Lys15 and Lys16. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to ω=1, expected under neutrality, and the predicted domain positions are plotted below the figure.

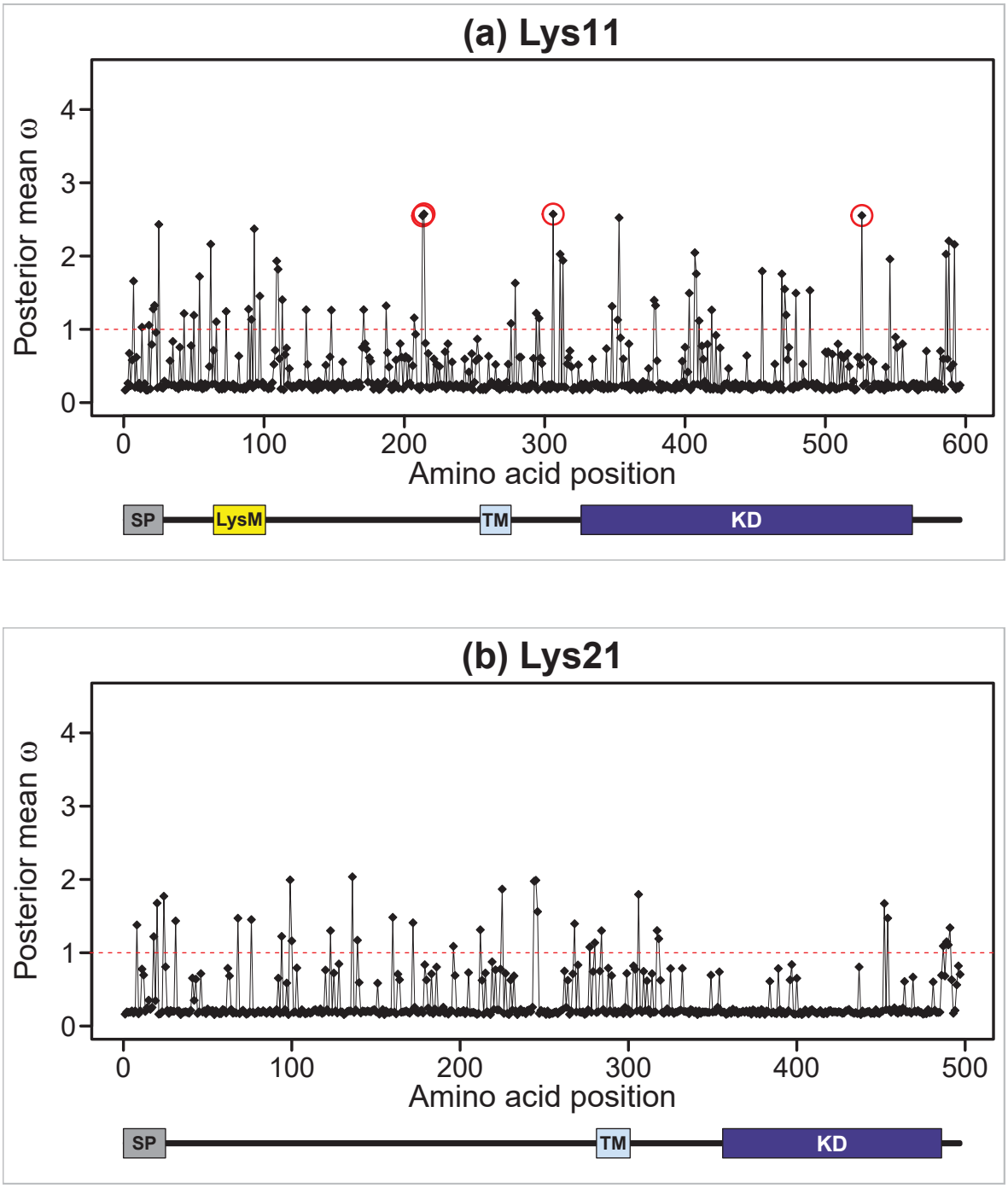

Figure 15. Distribution of posterior mean ω across Lys11 and Lys21. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to ω=1, expected under neutrality, and the predicted domain positions are plotted below the figure.

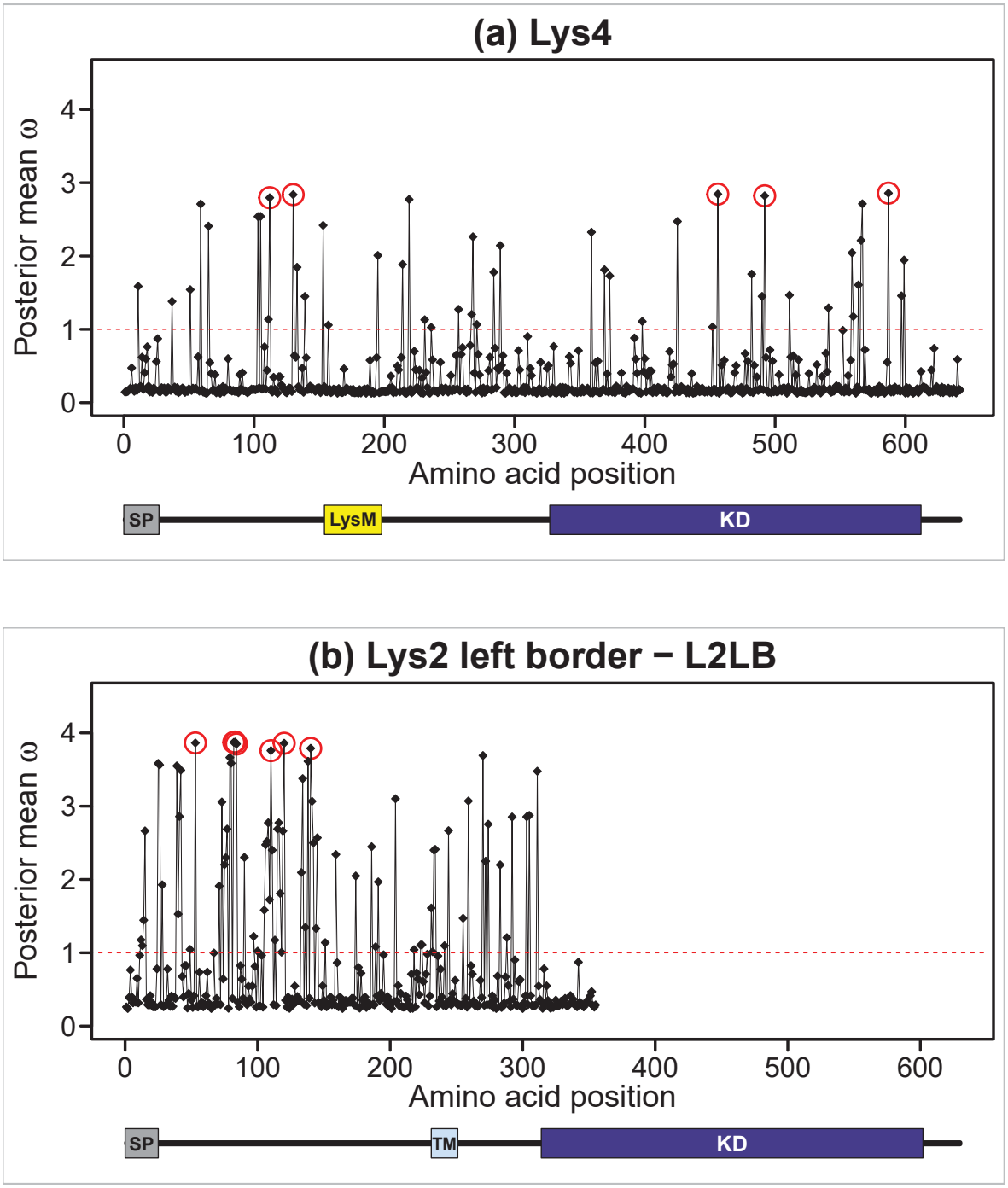

Figure 16. Distribution of posterior mean ω across Lys4 and Lys2 left-border (L2LB). The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to ω=1, expected under neutrality, and the predicted domain positions are plotted below the figure.

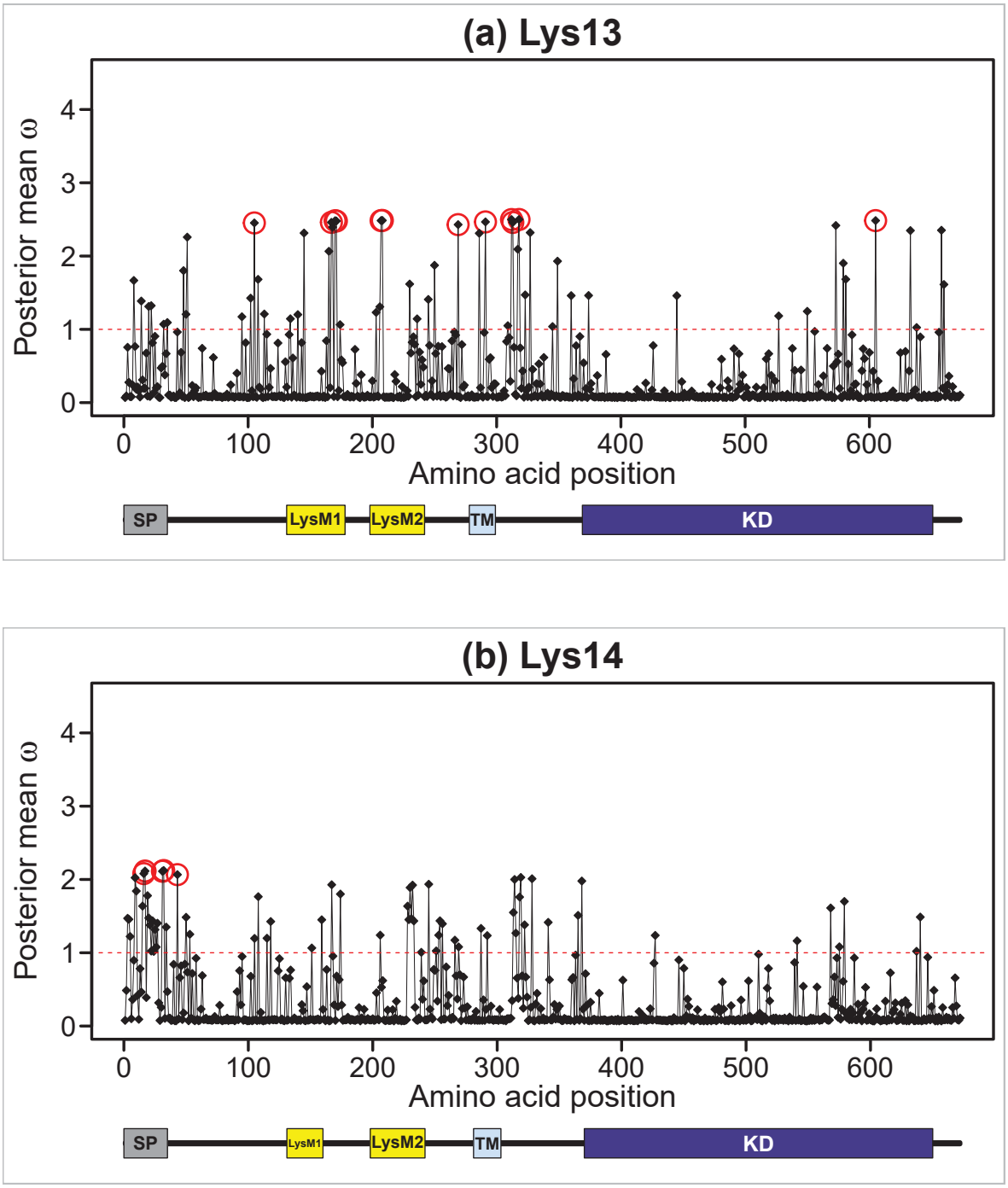

Figure 17. Distribution of posterior mean ω across Lys13 and Lys14. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to ω=1, expected under neutrality, and the predicted domain positions are plotted below the figure.

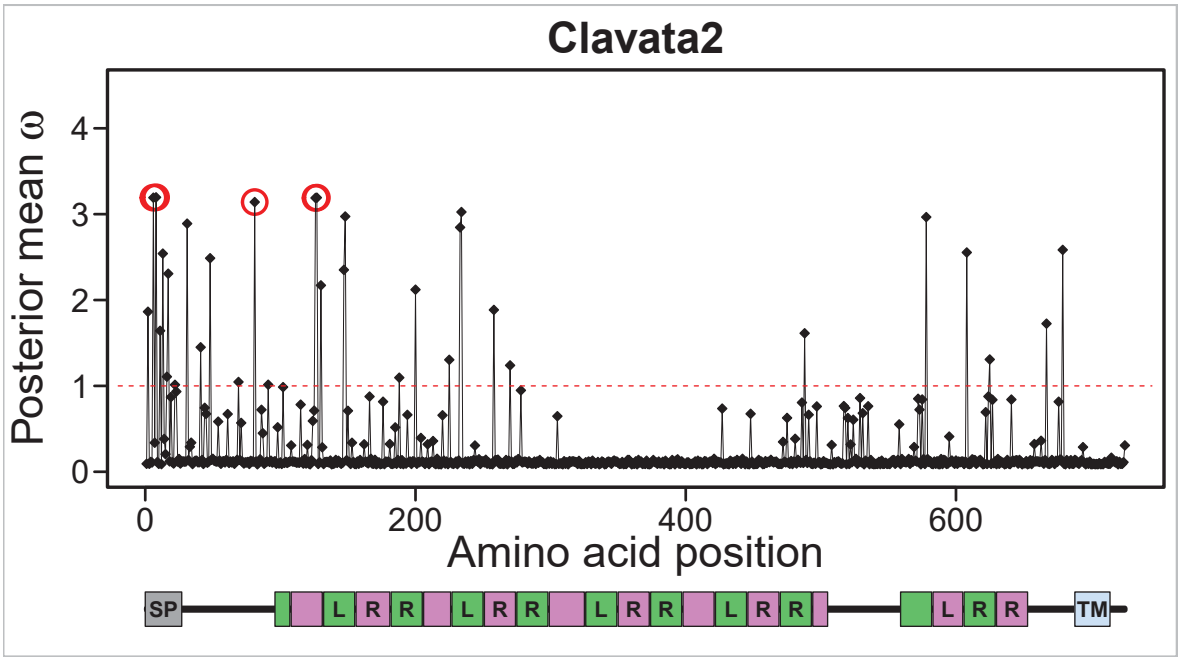

Figure 18. Distribution of posterior mean ω across LRR-type receptor-like protein Clavata2. The ratio of substitution rates (ω) was estimated using the random-sites model 8 of CodeML. Sites with significant evidence of positive selection according to the BEB statistics (p-value < 0.05) are highlighted with red circles. The dotted red line corresponds to ω=1, expected under neutrality, and the predicted domain positions are plotted below the figure.

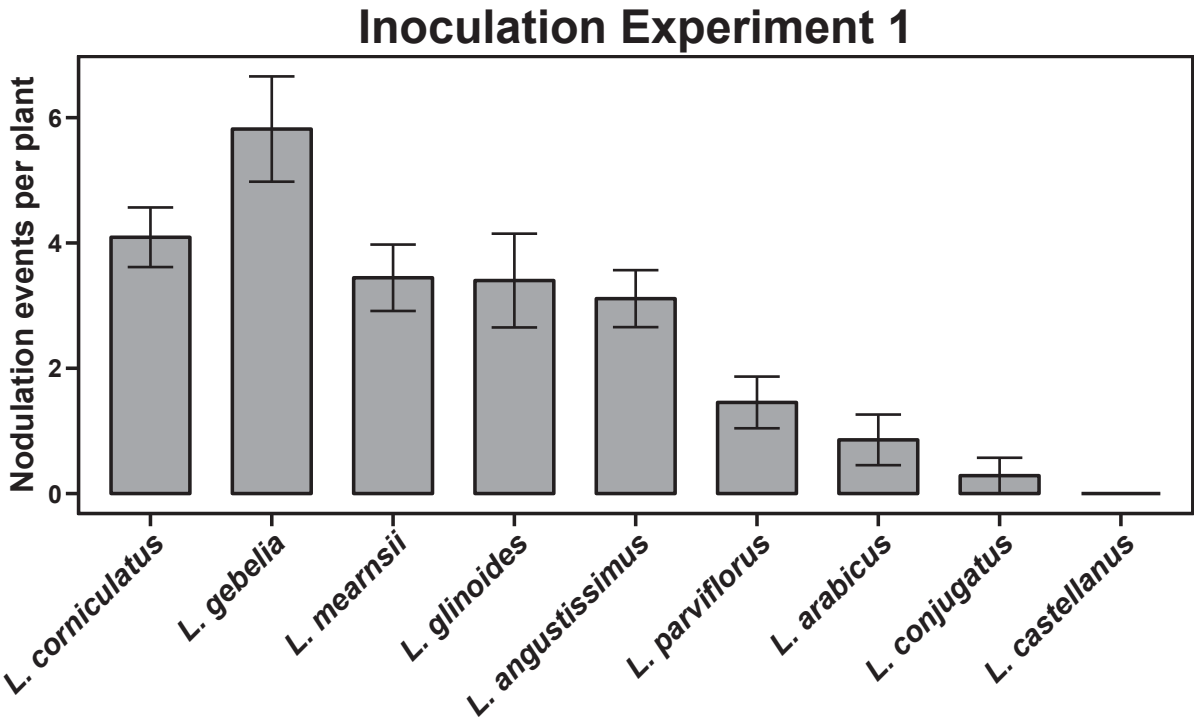

Figure 19. Compatibility of 9 (out of total 24 tested) *Lotus* species with *Mesorhizobium*. Plants were grown in glass canning jars (Weck) on seramis and minimum nitrogen medium. Nodulation events were scored 4 weeks after infection with *M. loti* (strain MAFF303099 expressing DsRed). Each bar represents average of at least five plants and error bars show SE. The raw data and t-test results are presented in Table 8 and Table 9, respectively.

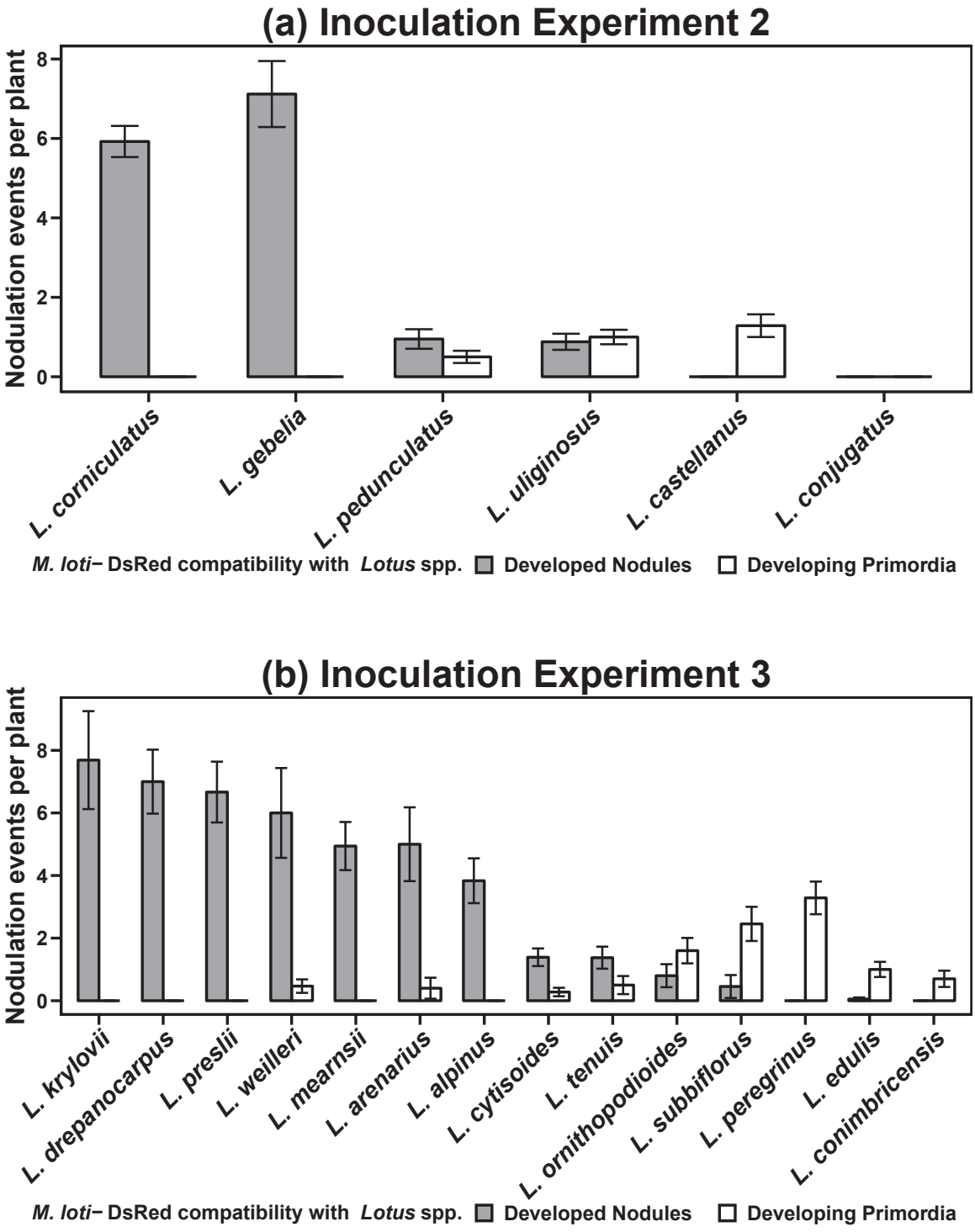

Figure 20. Compatibility of 20 (out of total 24 tested) *Lotus* species with *Mesorhizobium*. Plants were grown in glass canning jars (Weck) on seramis and minimum nitrogen medium. Nodulation events were scored 4 weeks after infection with *M. loti* (strain MAFF303099 expressing DsRed). Each bar represents average of at least five plants and error bars show SE. (a) Results of inoculation experiment 2. The raw data and t-test results are presented in Table 10 and Table 11, respectively. (b) Results of inoculation experiment 3. The raw data and t-test results are presented in Table 12 and Table 13, respectively.

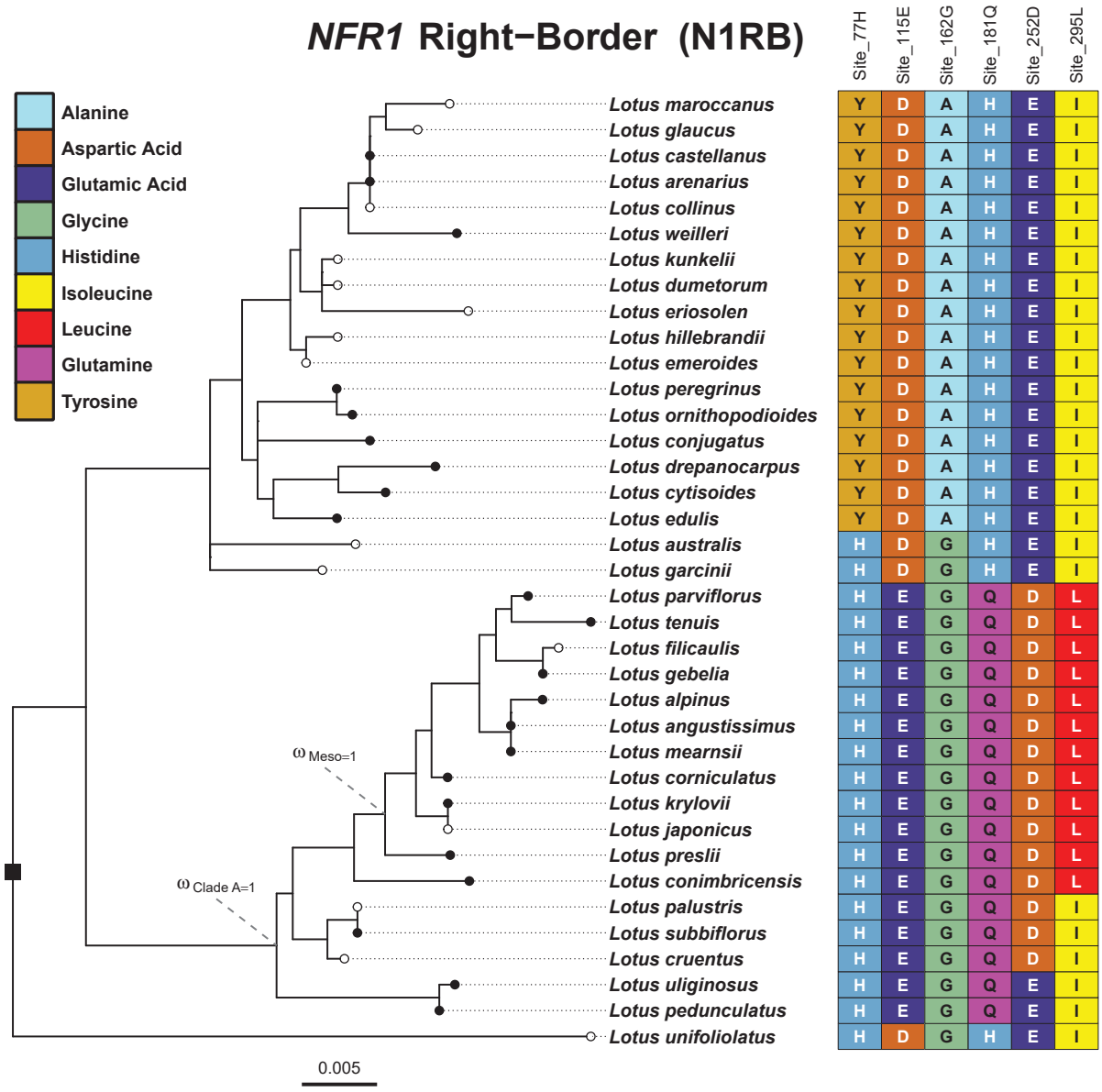

Figure 21. Maximum likelihood phylogeny based on NFR1 right-border (N1RB) sequences from 37 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few variable sites. The tree foreground used in CodeML branch-site models are indicated with dashed line.

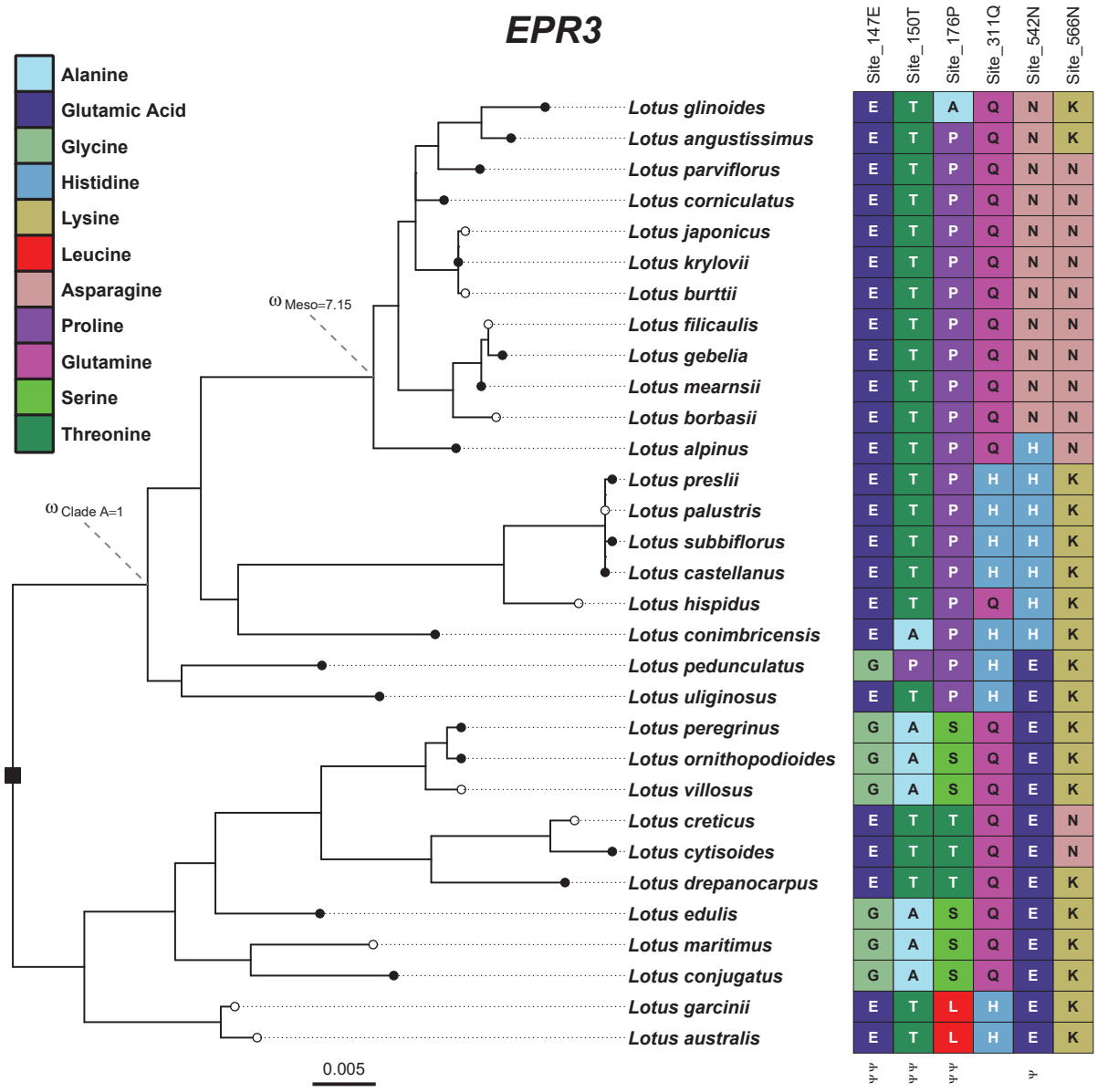

Figure 22. Maximum likelihood phylogeny based on Lys3 or EPR3 (Exopolysaccharide Receptor 3) sequences from 31 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

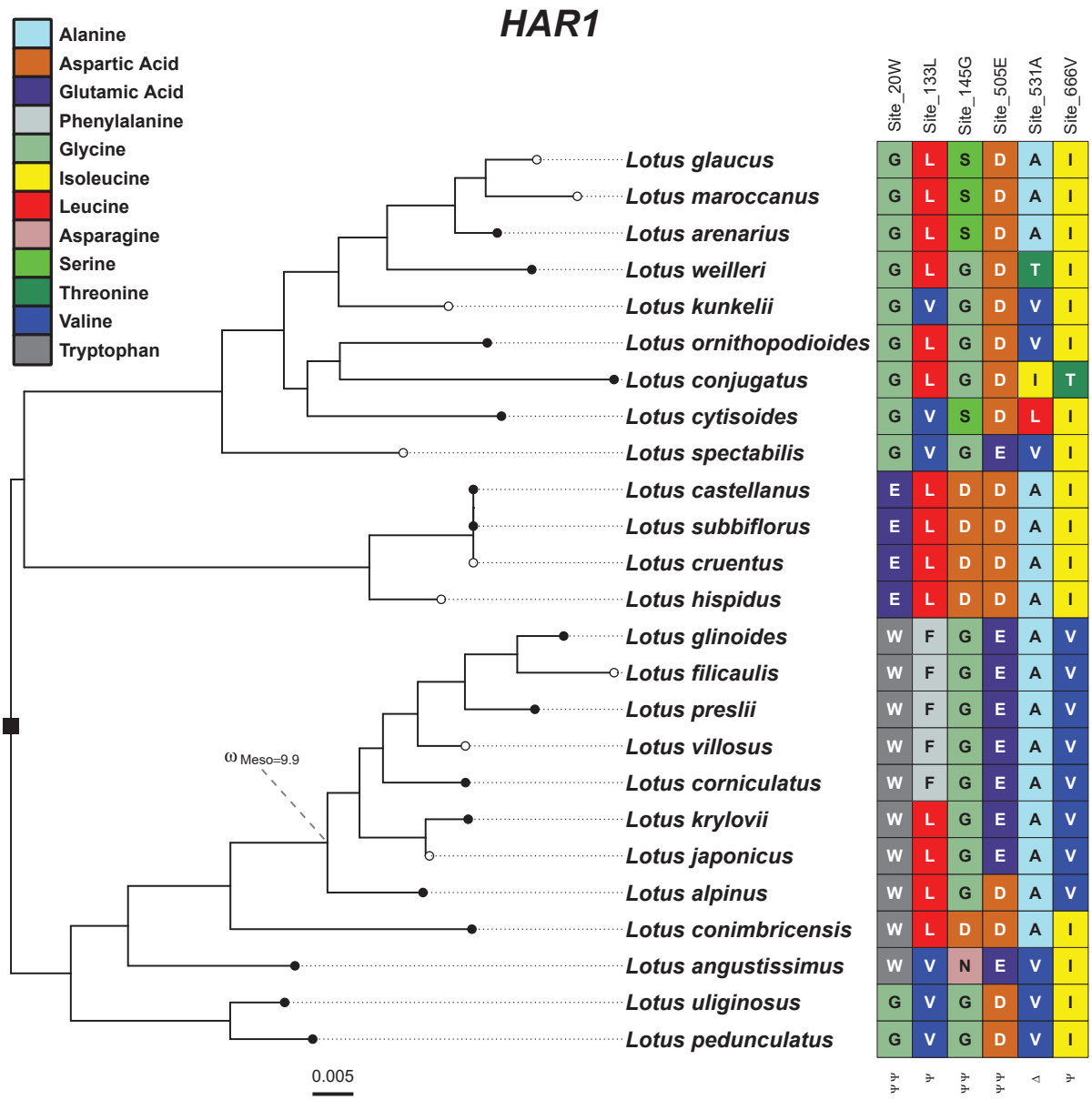

Figure 23. Maximum likelihood phylogeny based on HAR1 sequences from 25 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

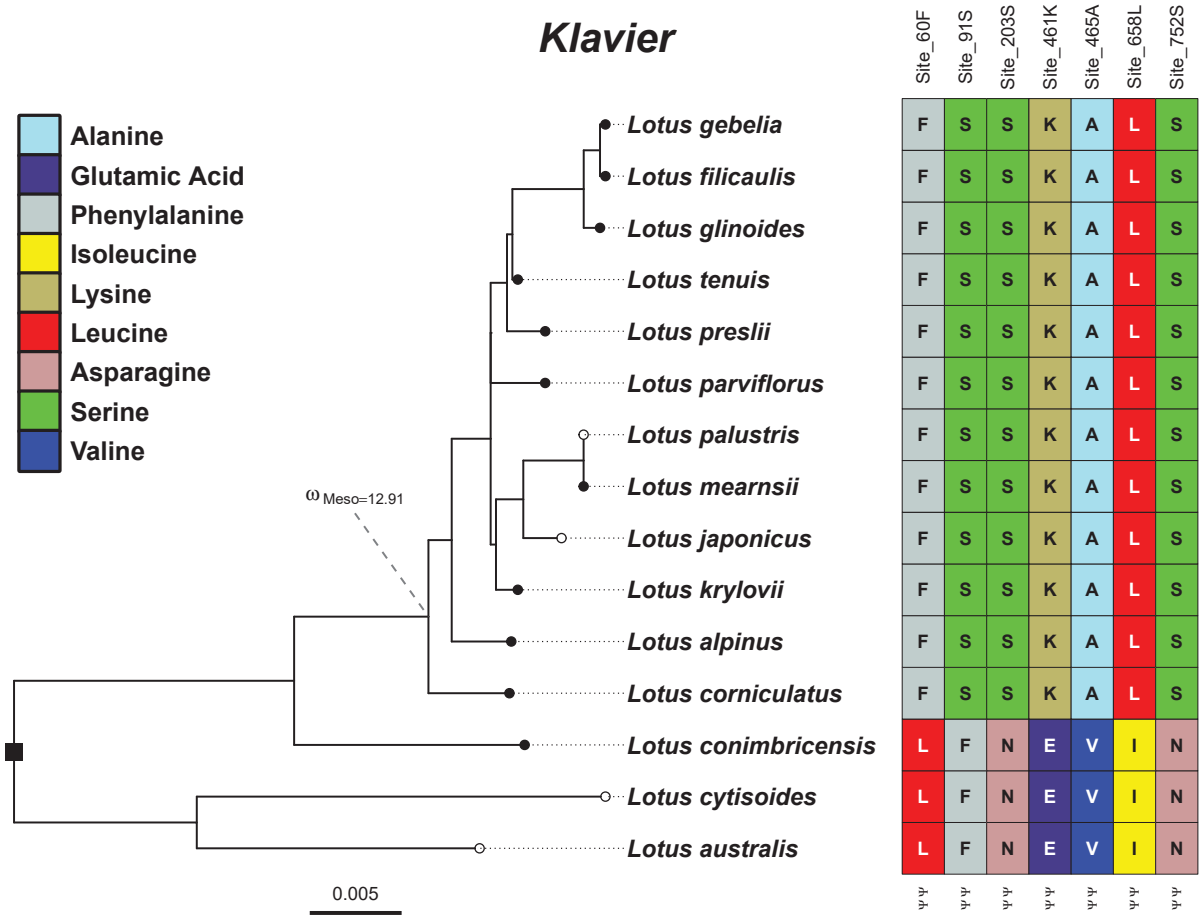

Figure 24. Maximum likelihood phylogeny based on Klavier sequences from 15 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01). None of the Klavier sites were significant in CodeML model 8 analysis. The tree foreground used in CodeML branch-site models are indicated with dashed line.

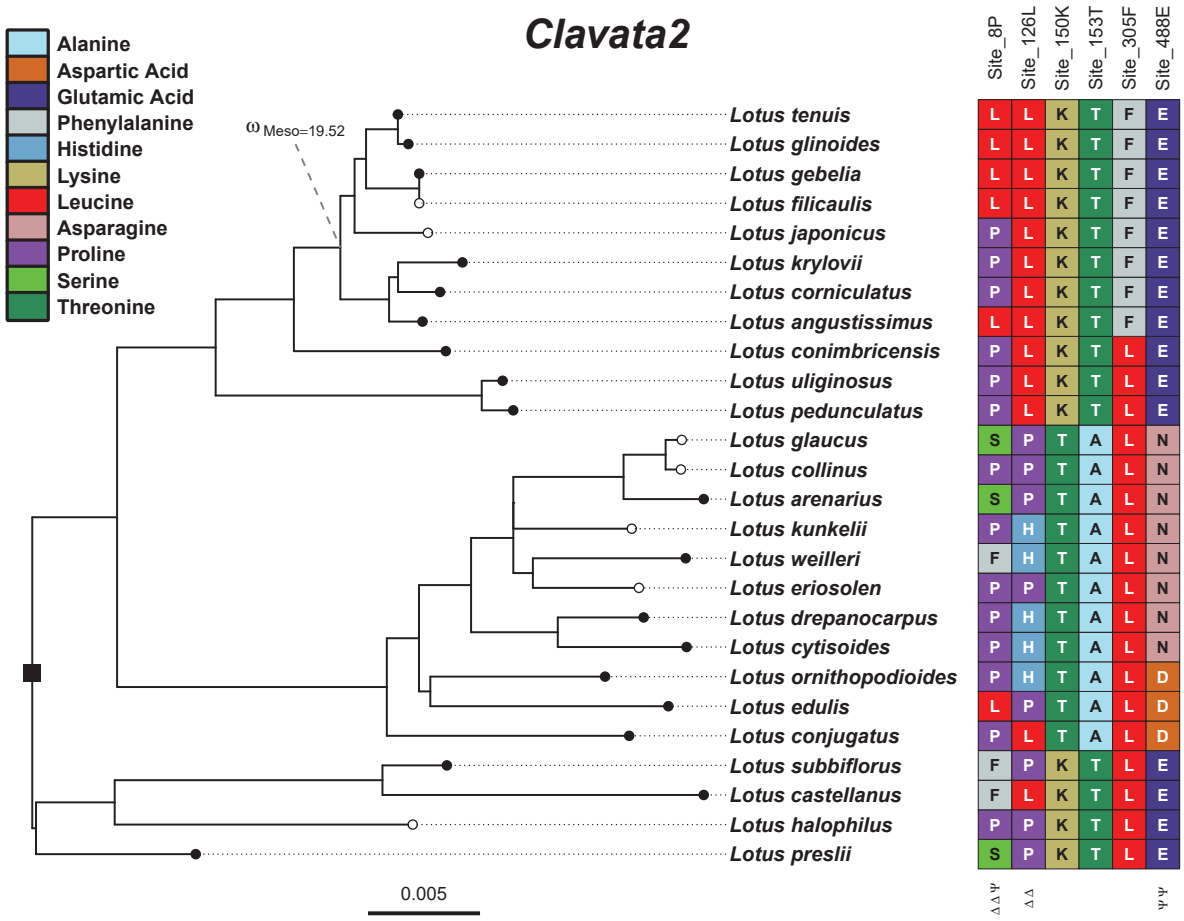

Figure 25. Maximum likelihood phylogeny based on Clavata2 sequences from 26 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

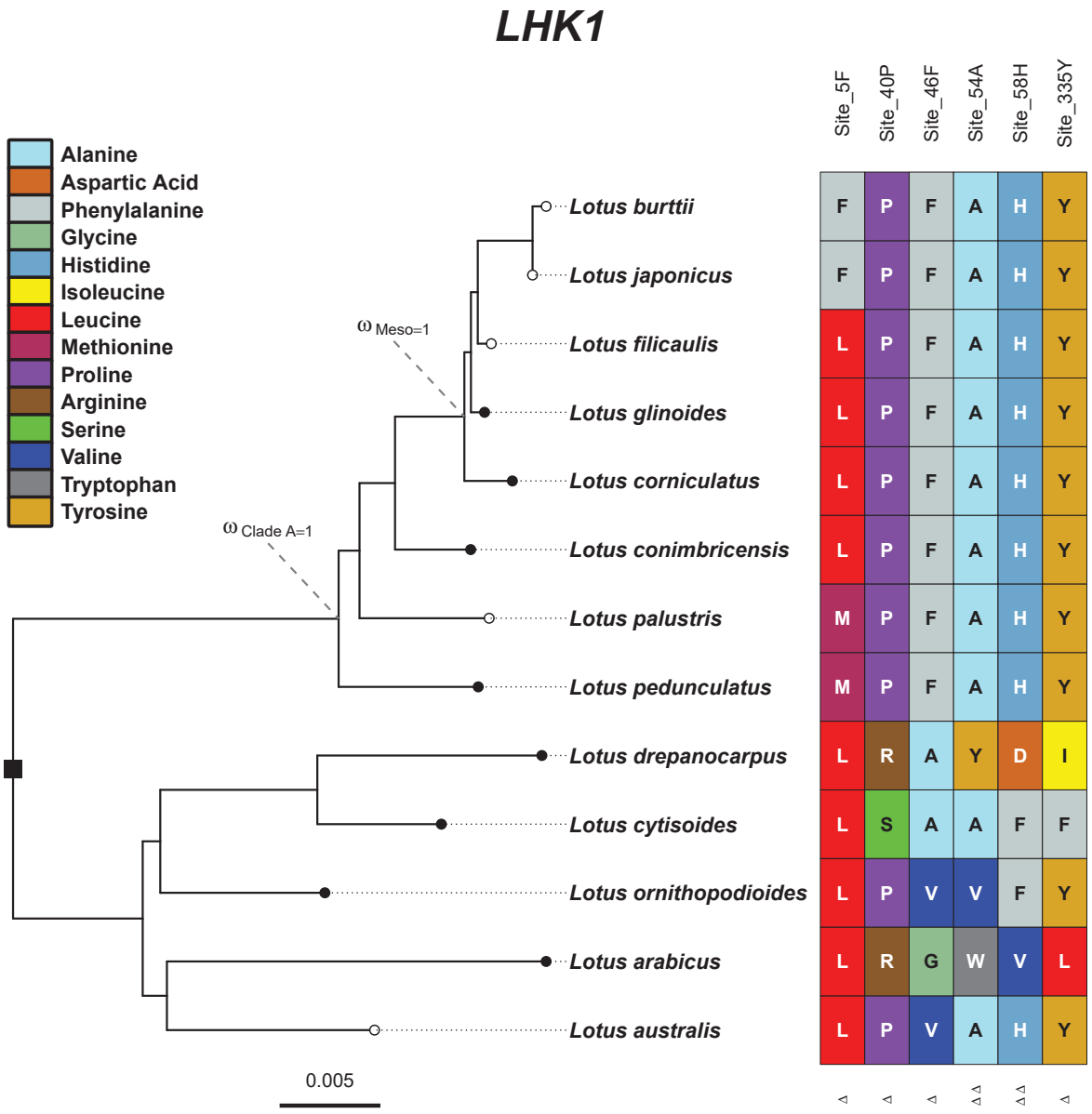

Figure 26. Maximum likelihood phylogeny based on LHK1 sequences from 13 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). No significant correlation was found in the association analysis via Genphen. The tree foreground used in CodeML branch-site models are indicated with dashed line.

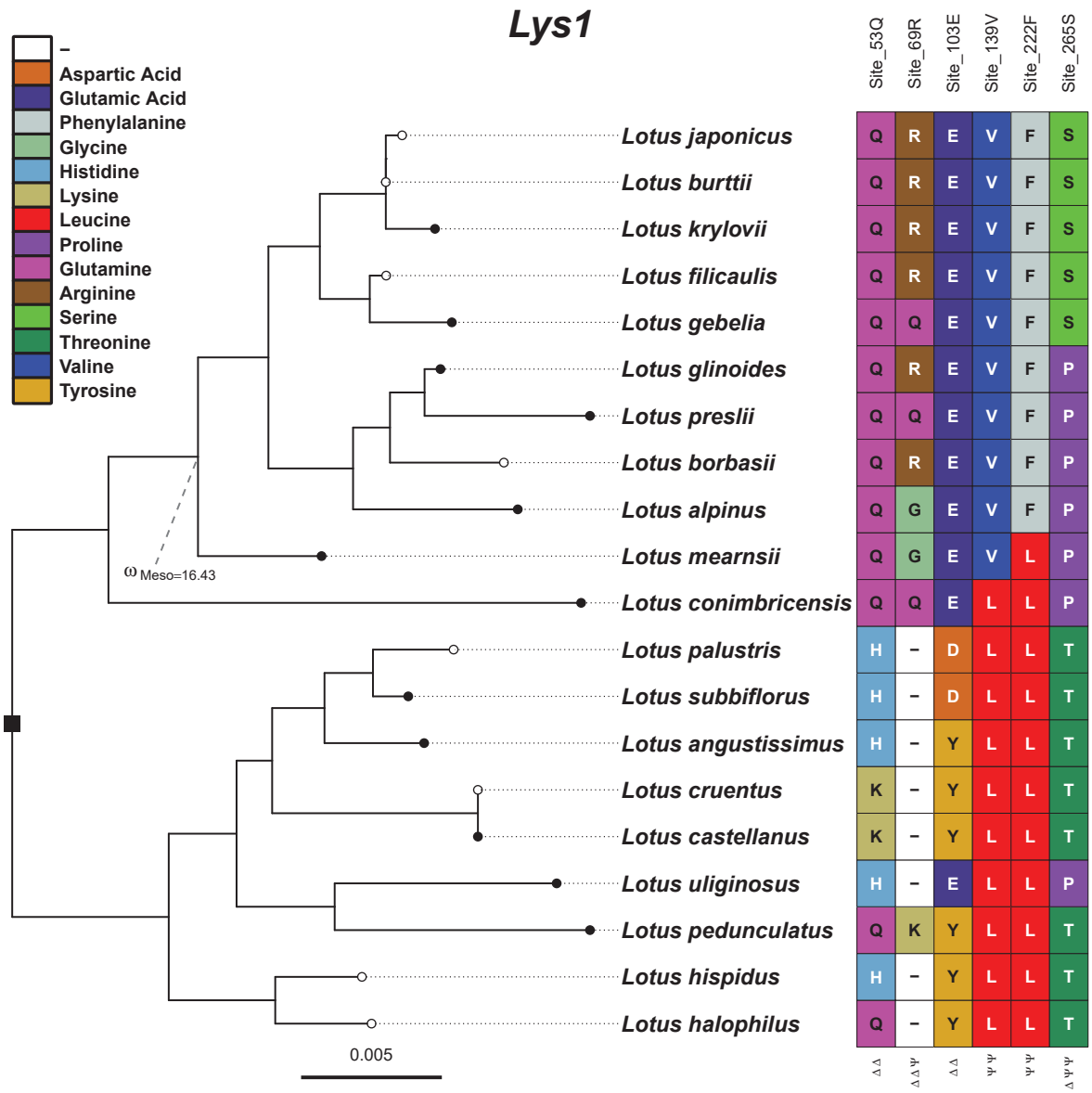

Figure 27. Maximum likelihood phylogeny based on NFRe or Lys1 sequences from 20 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

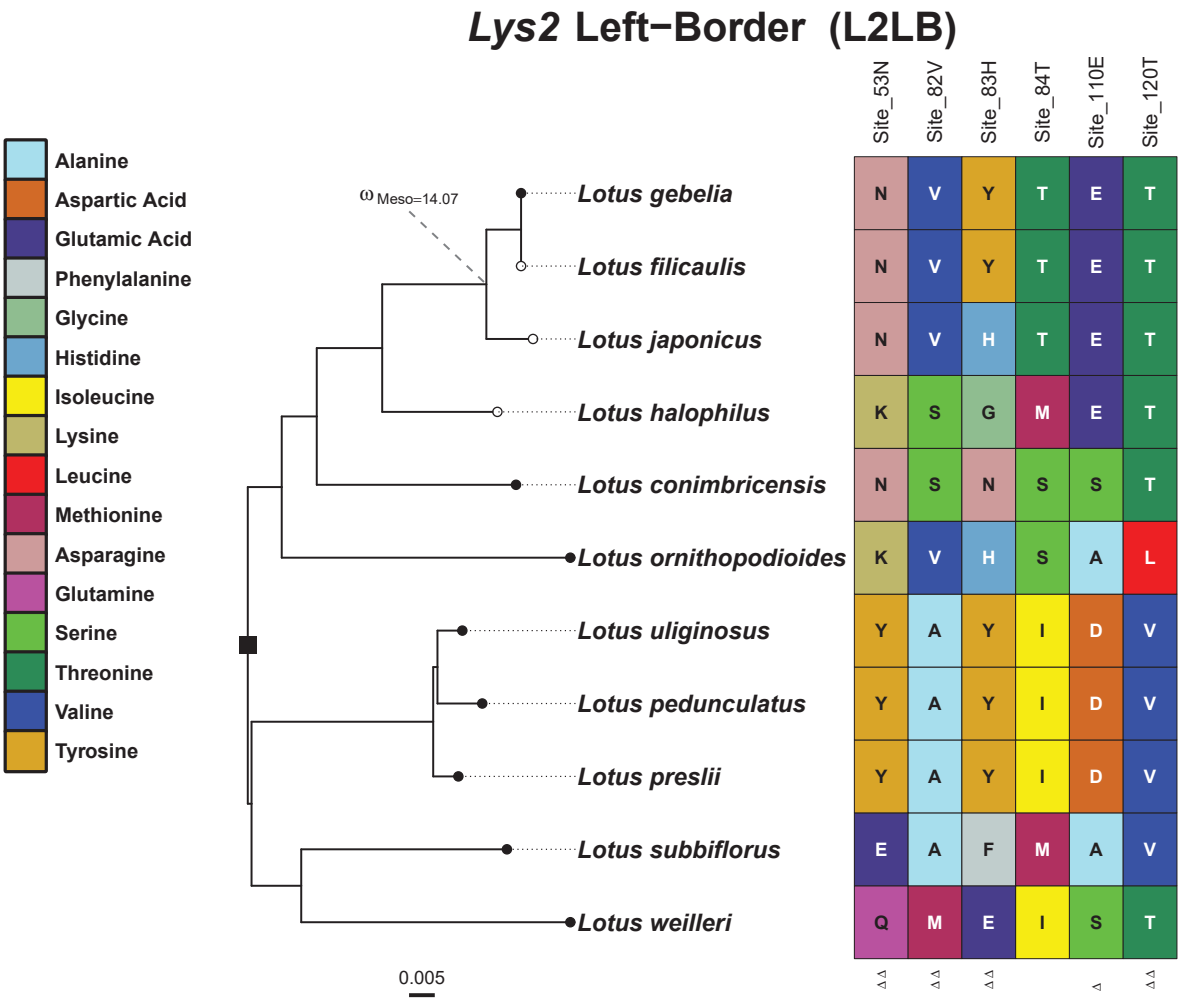

Figure 28. Maximum likelihood phylogeny based on Lys2 left-border (L2LB) sequences from 11 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). No significant correlation was found in the association analysis via Genphen. The tree foreground used in CodeML branch-site models are indicated with dashed line.

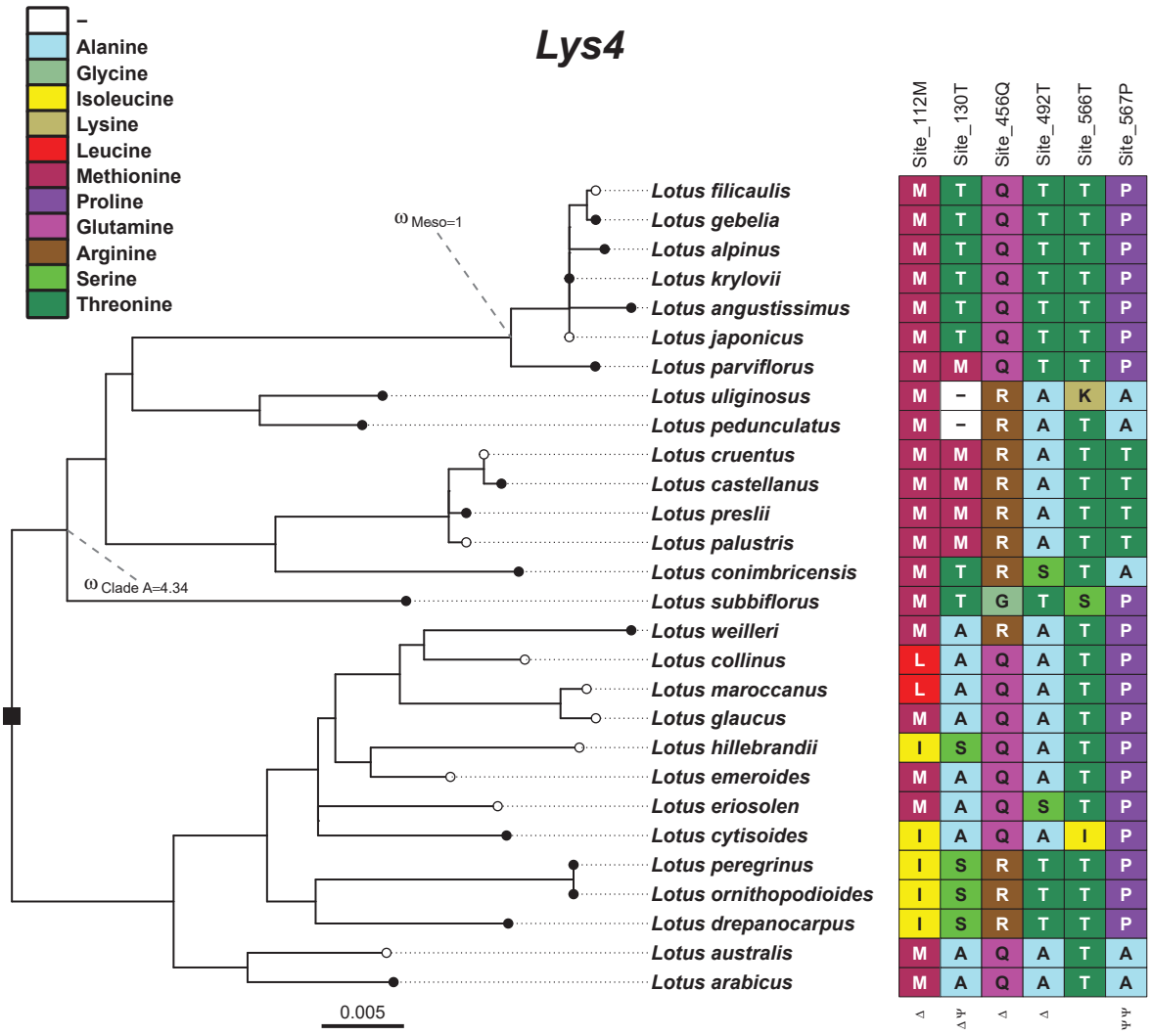

Figure 29. Maximum likelihood phylogeny based on Lys4 sequences from 28 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

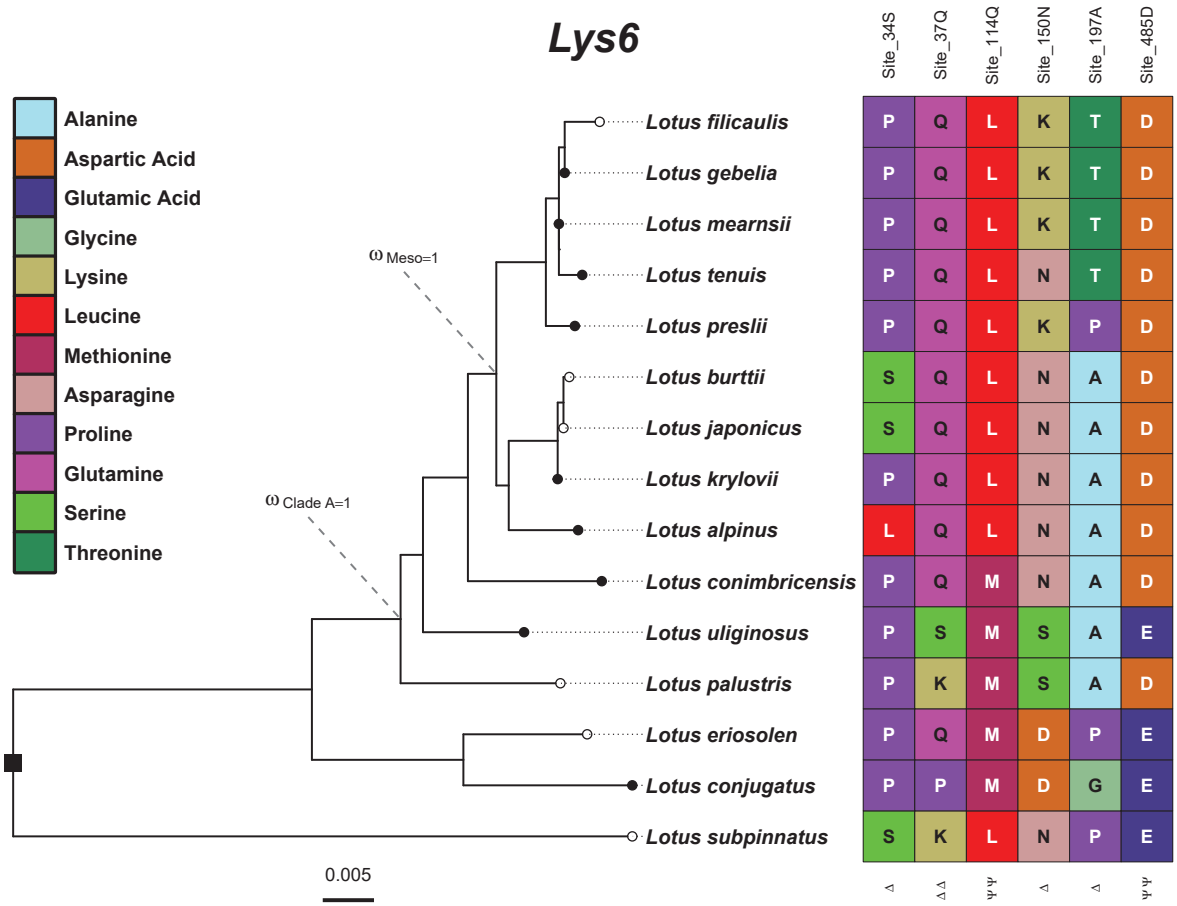

Figure 30. Maximum likelihood phylogeny based on Lys6 (partial, AA #34 - stop codon) sequences from 15 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

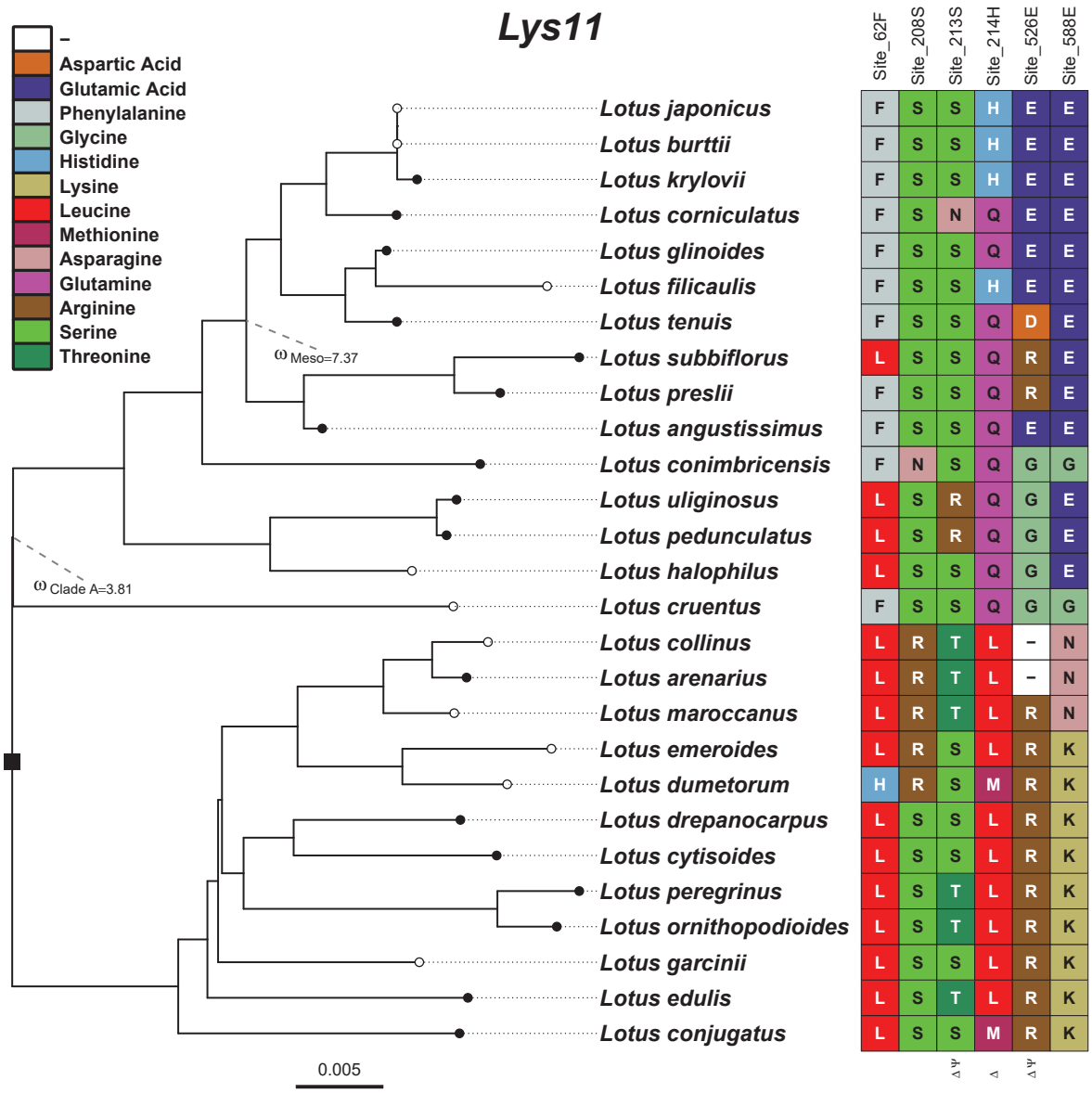

Figure 31. Maximum likelihood phylogeny based on Lys11 sequences from 27 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

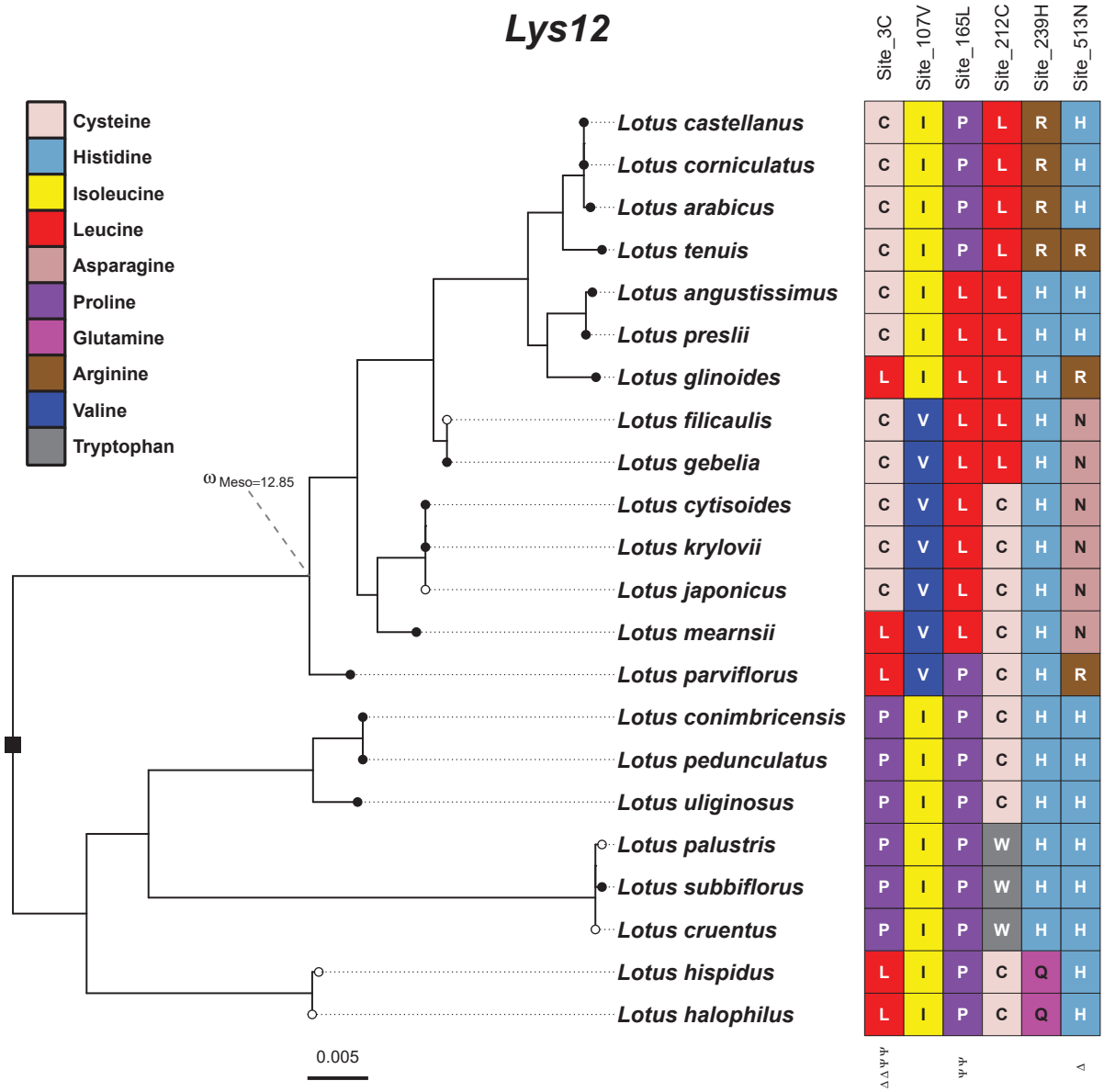

Figure 32. Maximum likelihood phylogeny based on Lys12 sequences from 22 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

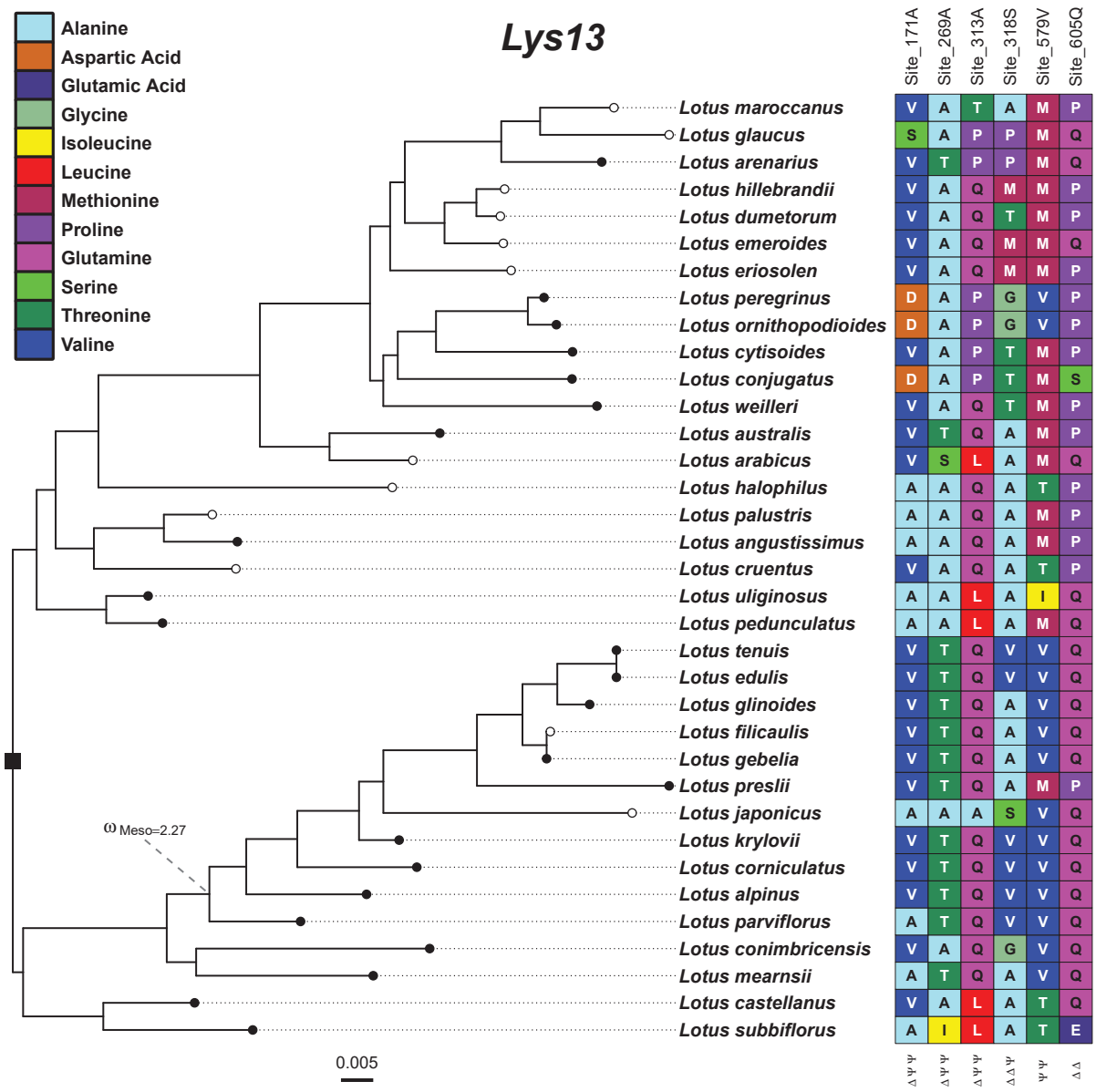

Figure 33. Maximum likelihood phylogeny based on Lys13 sequences from 35 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

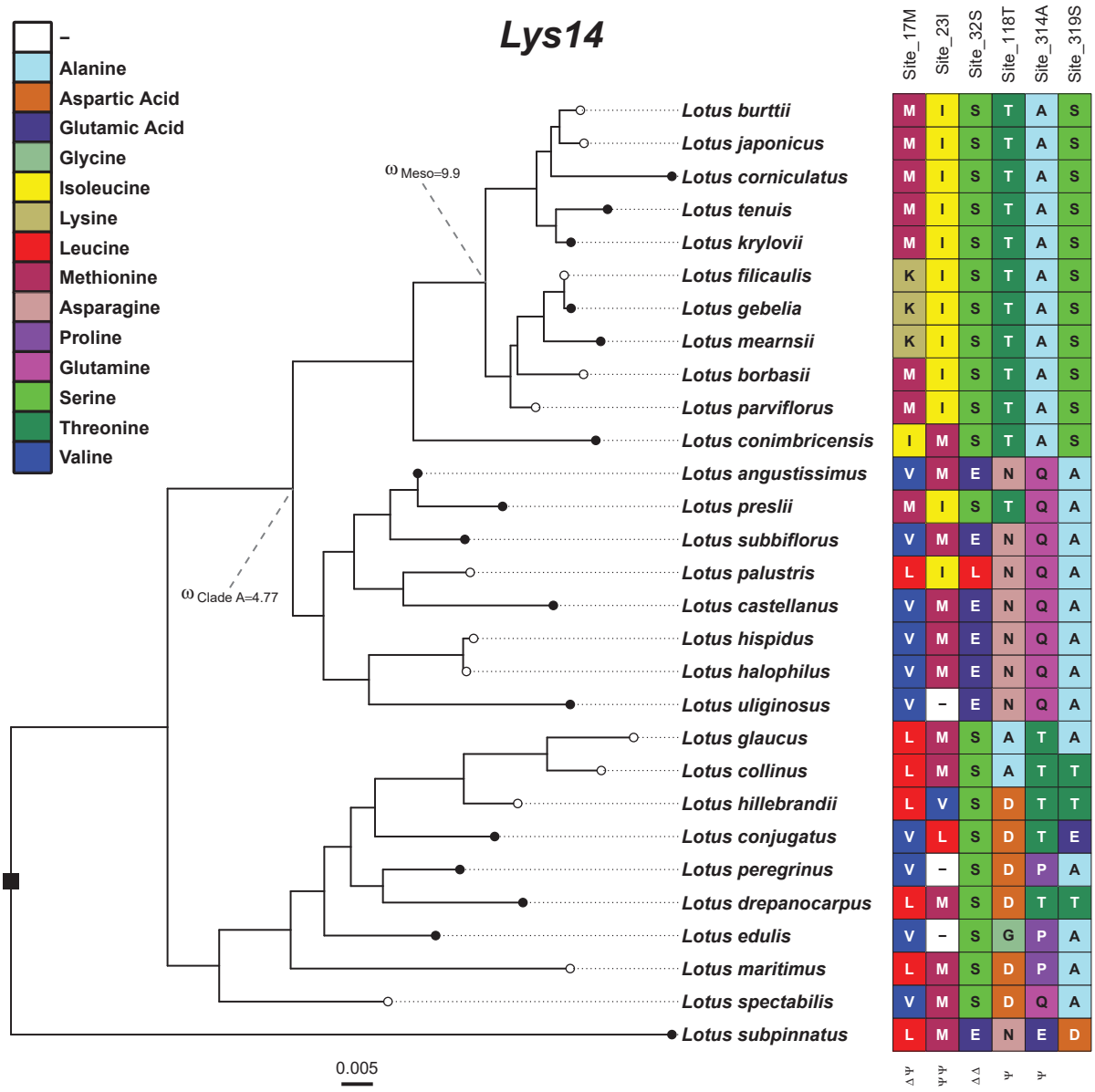

Figure 34. Maximum likelihood phylogeny based on Lys14 sequences from 29 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figure 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

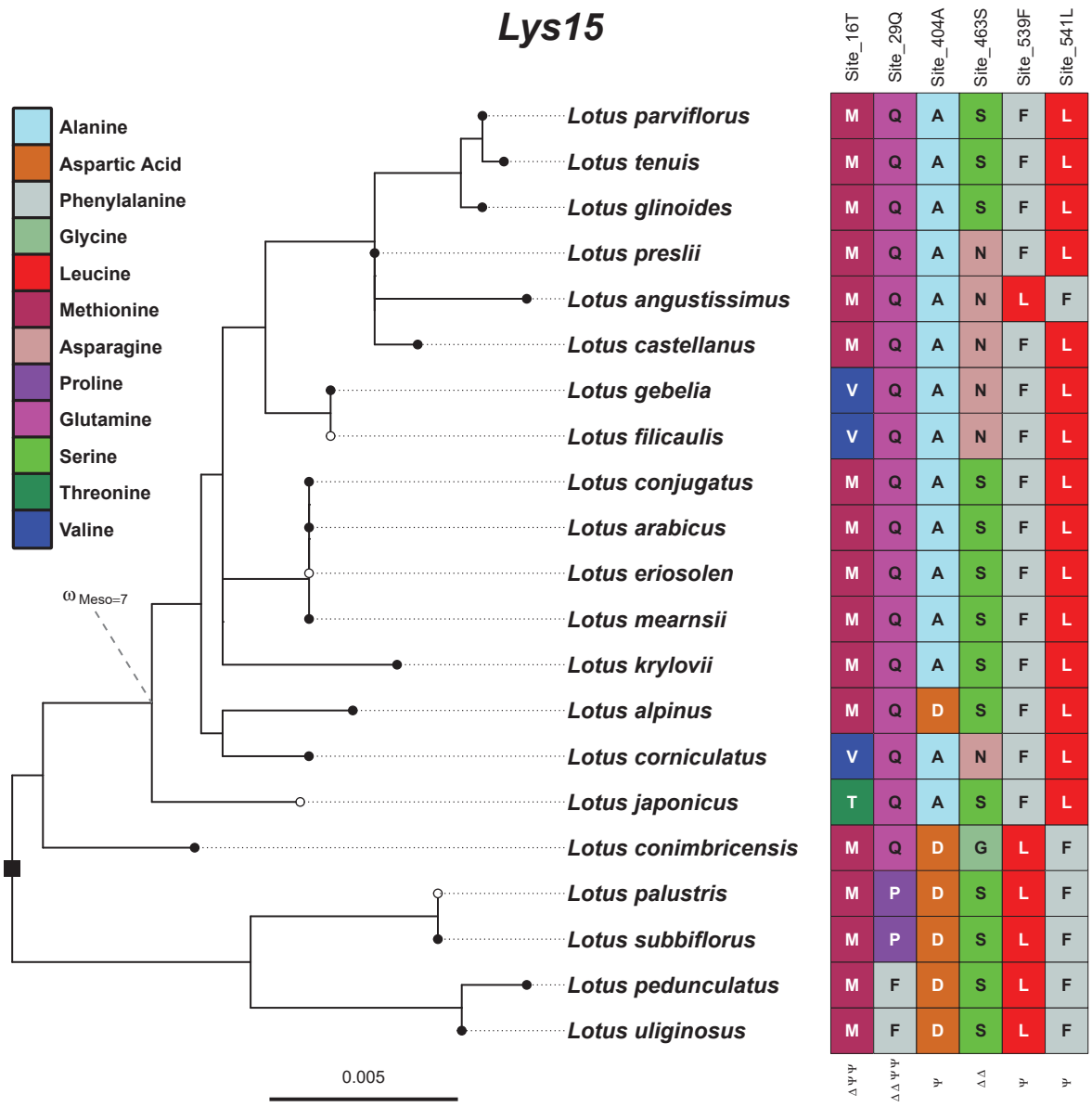

Figure 35. Maximum likelihood phylogeny based on Lys15 sequences from 21 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

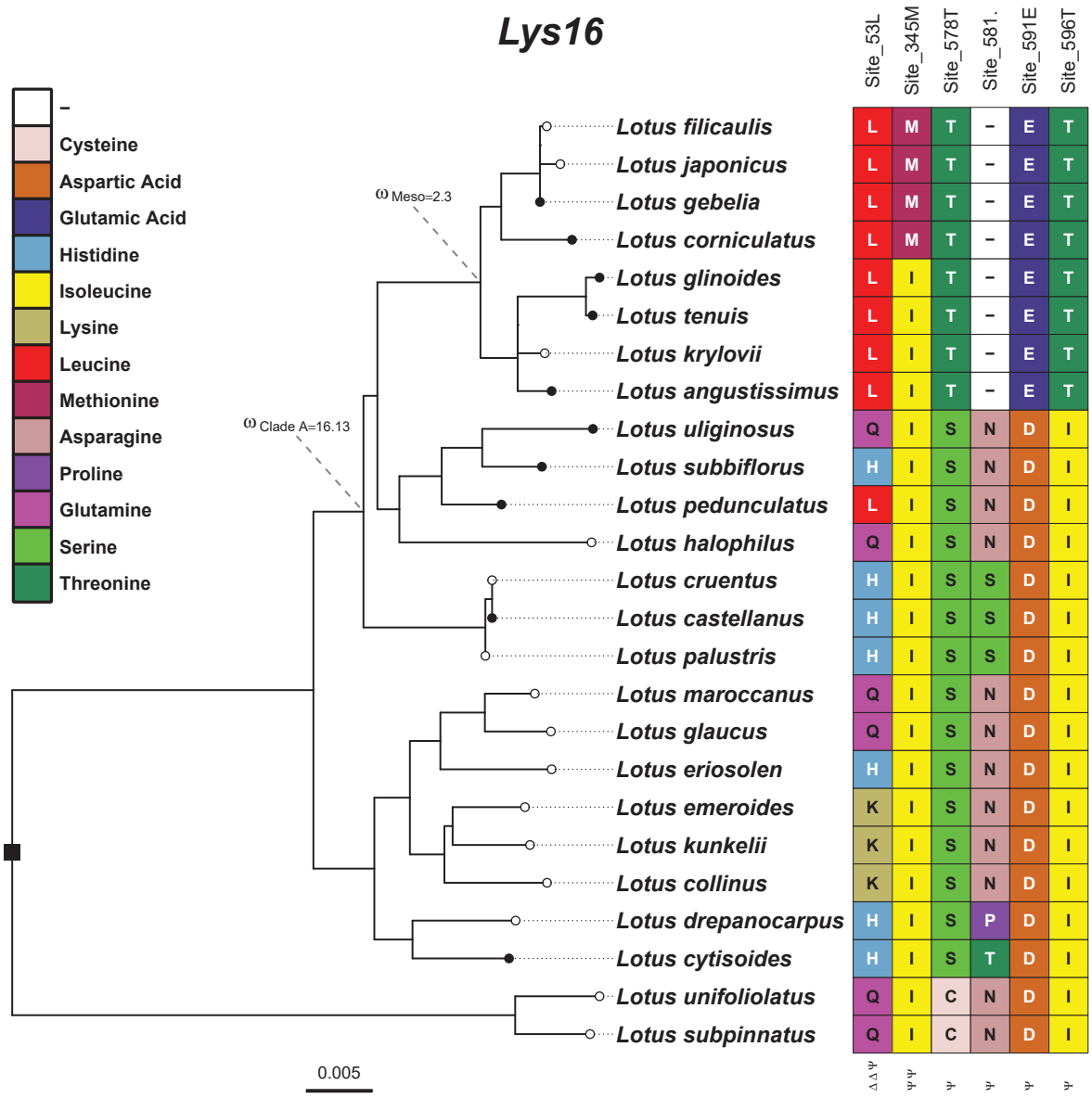

Figure 36. Maximum likelihood phylogeny based on Lys16 sequences from 25 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

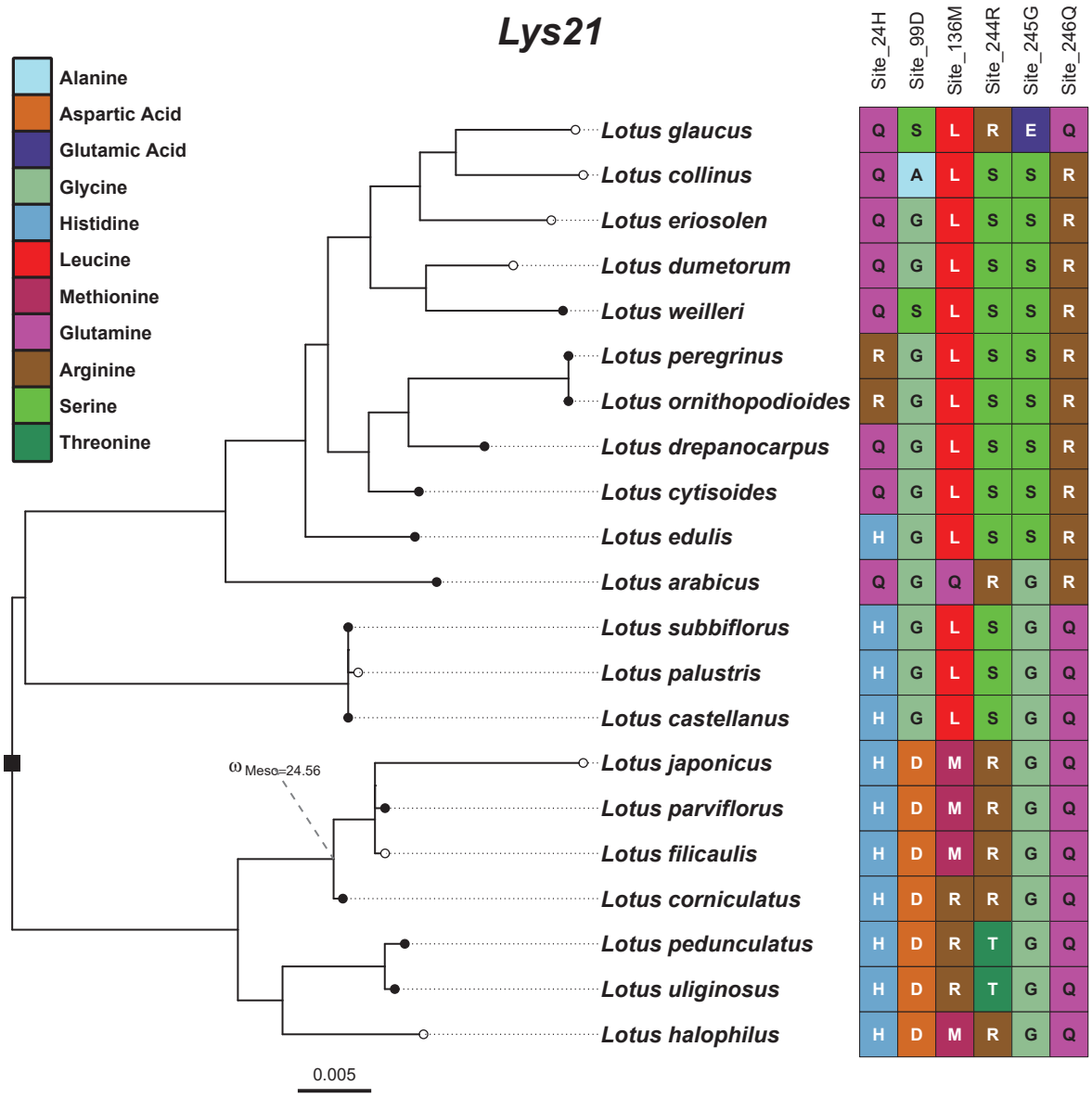

Figure 37. Maximum likelihood phylogeny based on Lys21 sequences from 21 *Lotus* species. The tip points filled with black represents species that were used for association analysis by Genphen; are common between this figure and the figures 19 & 20. The black box represents mid point root of the phylogenetic tree. The panel on the right side highlights few interesting sites, where ψ denotes significant sites in Genphen association analysis (ψ = p-value < 0.05; $\psi\psi$ = p-value < 0.01) and Δ denotes significant sites in CodeML model 8 analysis (Δ = p-value < 0.05; $\Delta\Delta$ = p-value < 0.01). The tree foreground used in CodeML branch-site models are indicated with dashed line.

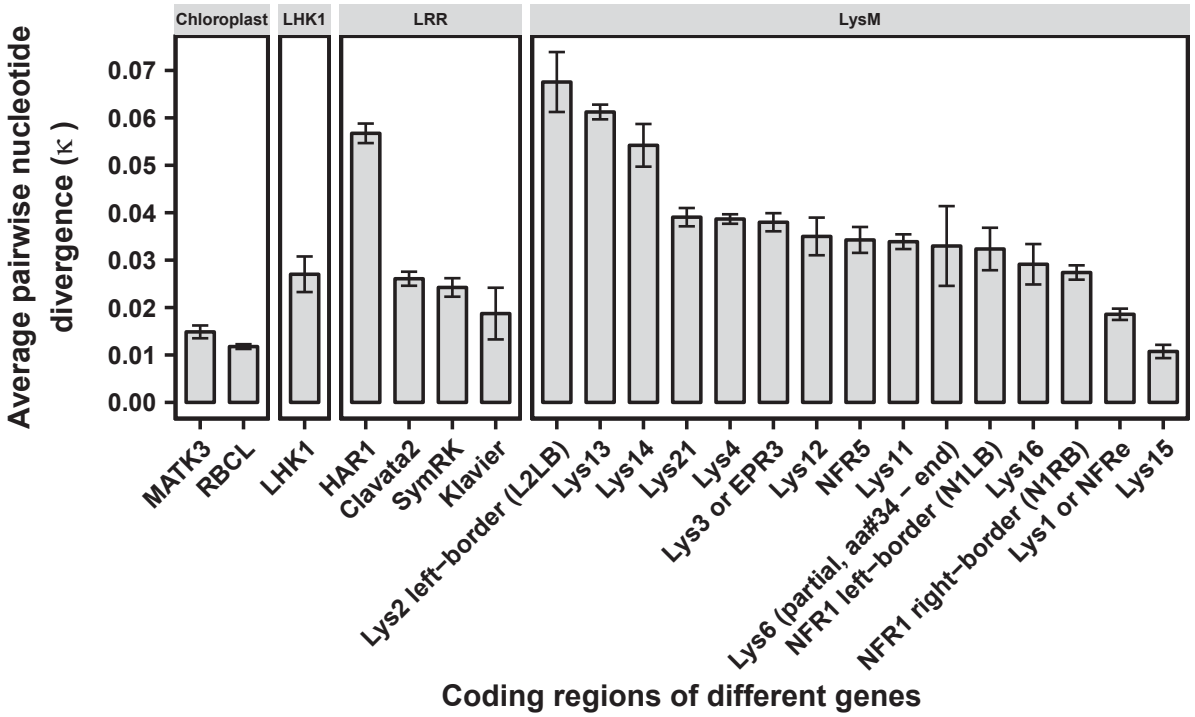

Figure 38. Average pairwise nucleotide divergence (κ) among *Lotus* species across 20 different receptors. This includes 19 LysM and LRR type symbiotic receptor-like kinases, the receptor protein Clavata2 and the cytokinin receptor LHK1. Values were calculated using DNAsp and bars show SD.

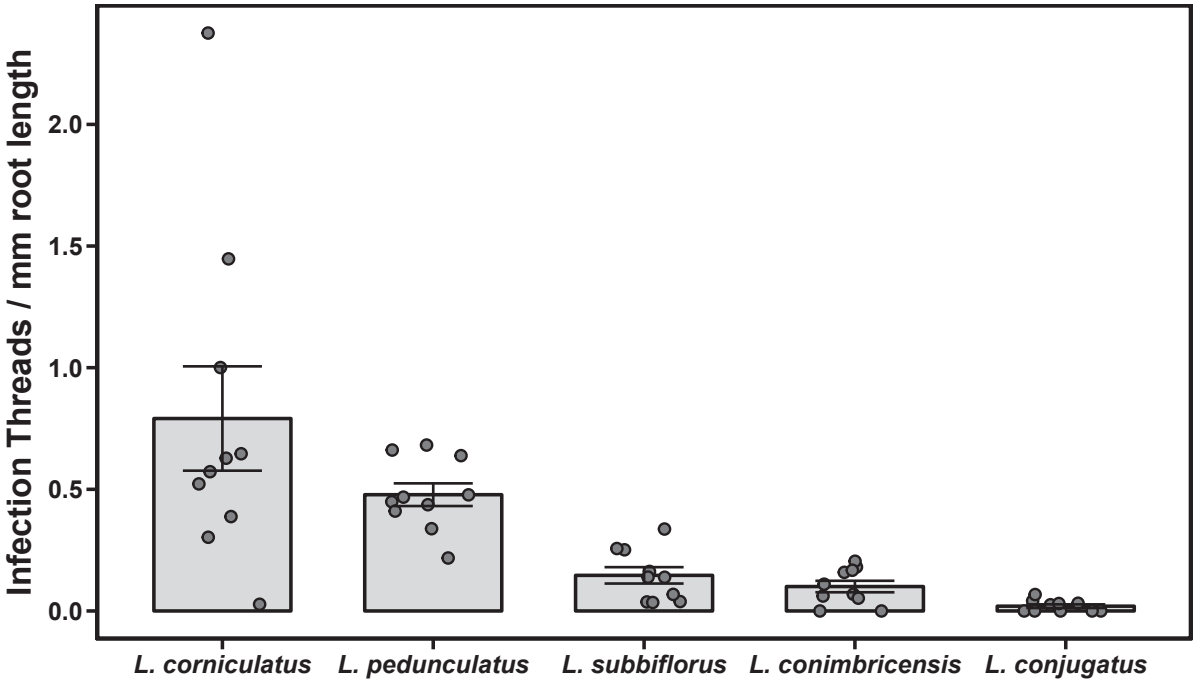

Figure 39. Frequency of infection threads (ITs) on different *Lotus* species. The *Mesorhizobium loti* DsRed was grown in YEM and plants were inoculated in glass canning jars as described in materials and methods section. The length of main-root was documented and infection threads were counted at 7 days after infection (n=10, bars show SEM, points show individual data) under a Zeiss fluorescence microscope.

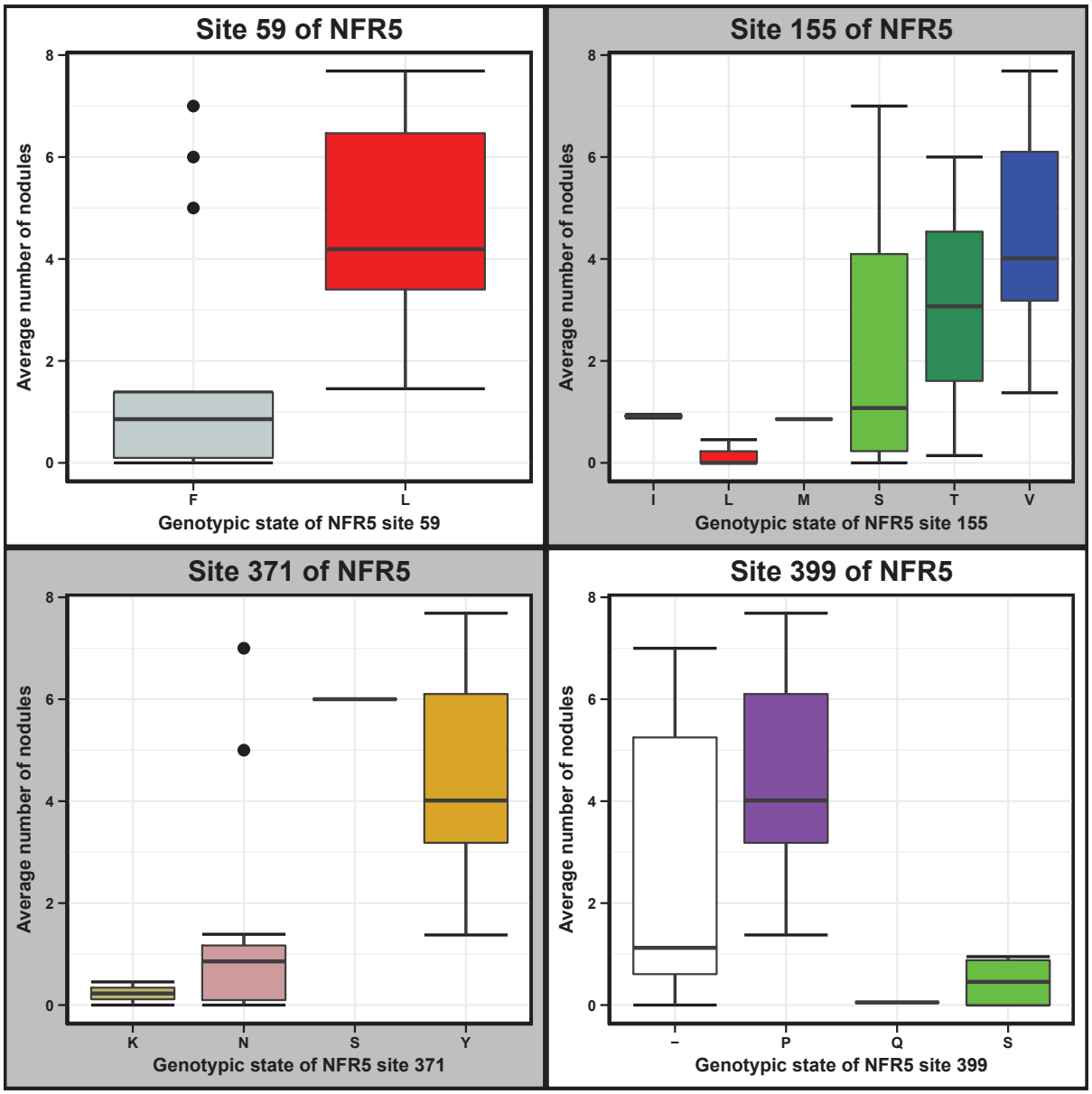

Figure 40. Boxplot showing the phenotypic distribution of the specific NFR5 genotypes. The boxplots were generated using the "plotSpecificGenotype" function of the R package Genphen.

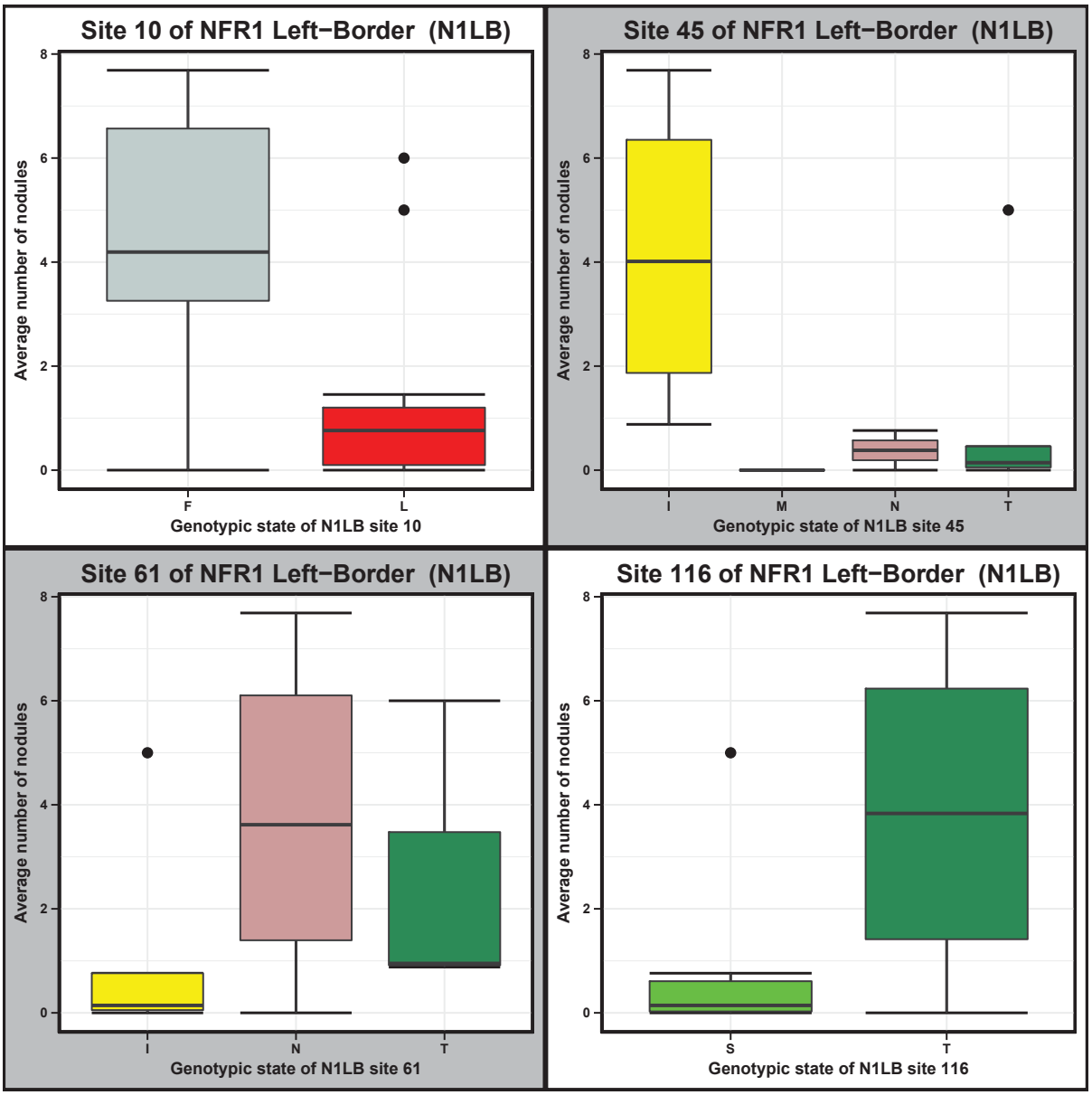

Figure 41. Boxplot showing the phenotypic distribution of the specific NFR1 left-border (N1LB) genotypes. The boxplots were generated using the "plotSpecificGenotype" function of the R package Genphen.

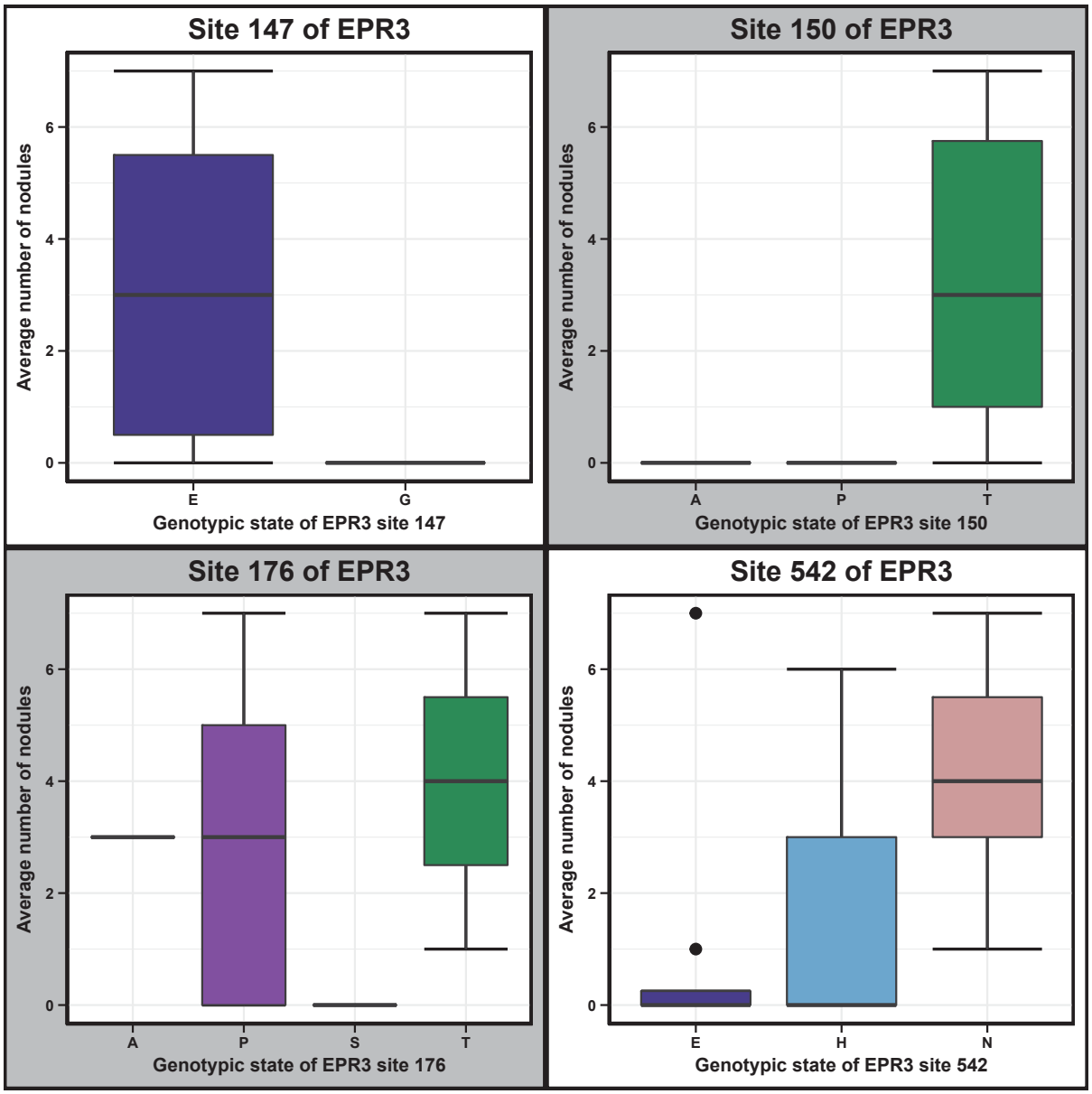

Figure 42. Boxplot showing the phenotypic distribution of the specific Lys3 or EPR3 genotypes. The boxplots were generated using the "plotSpecificGenotype" function of the R package Genphen.

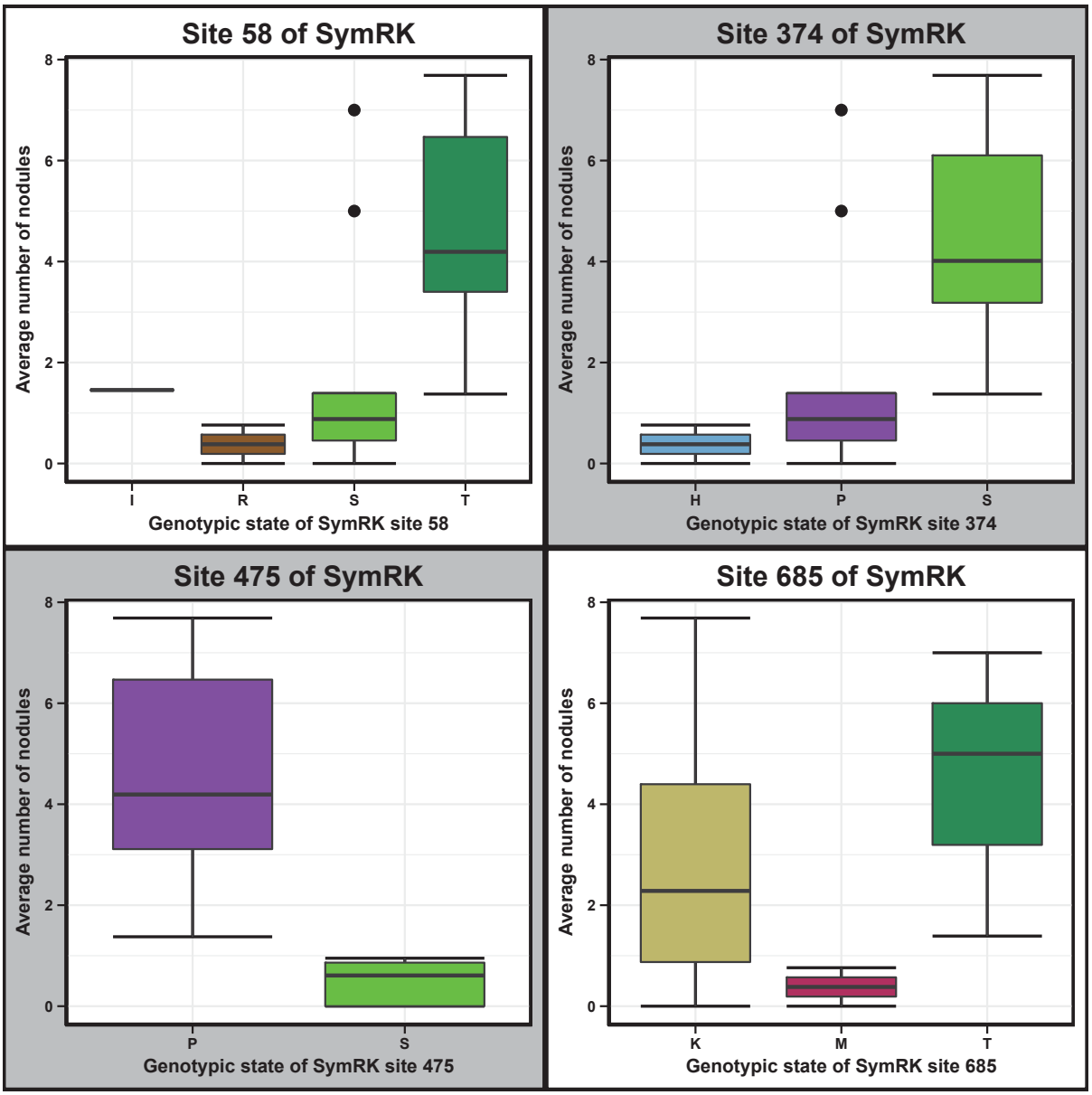

Figure 43. Boxplot showing the phenotypic distribution of the specific SymRK genotypes. The boxplots were generated using the "plotSpecificGenotype" function of the R package Genphen.

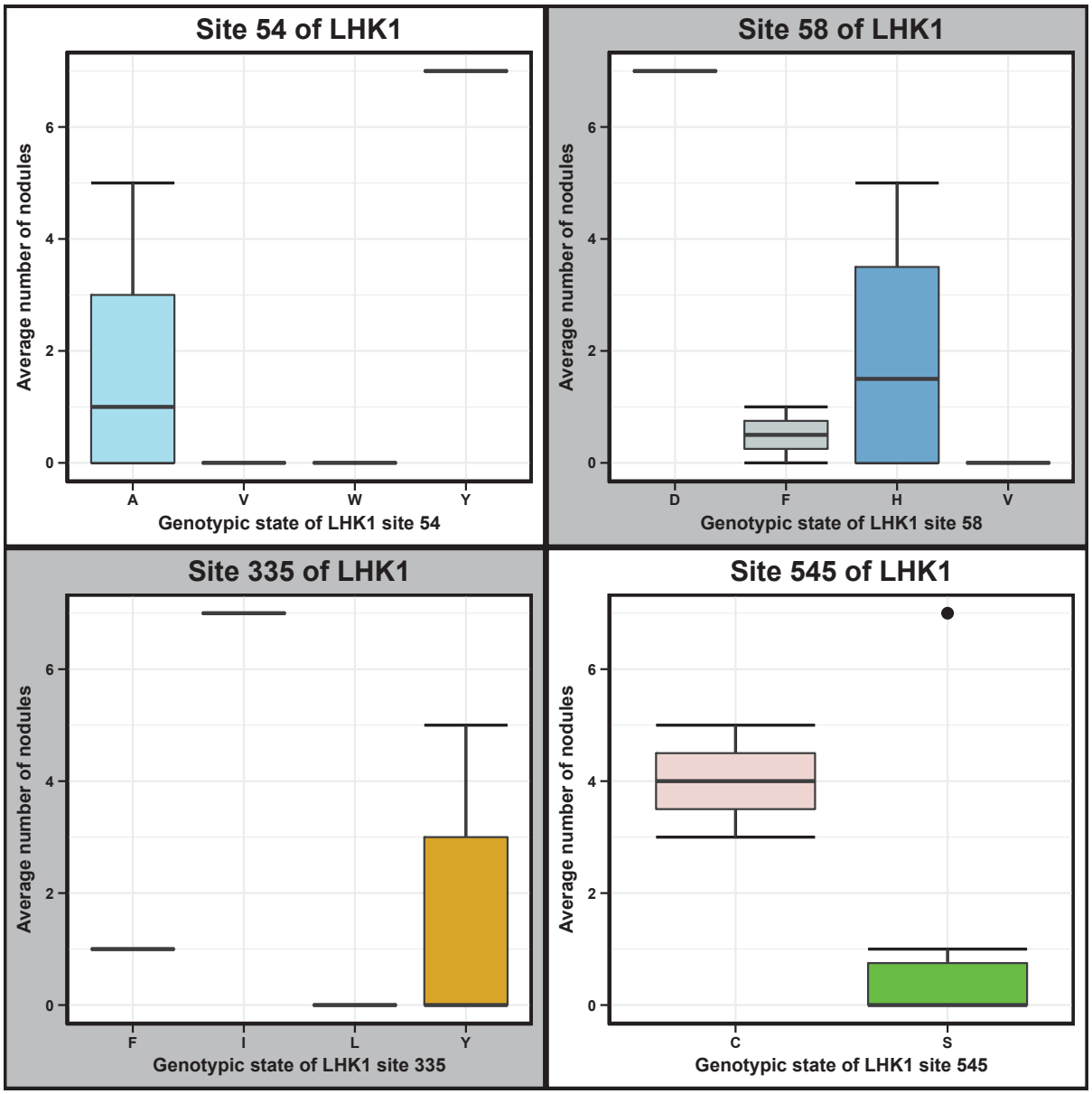

Figure 44. Boxplot showing the phenotypic distribution of the specific LHK1 genotypes. The boxplots were generated using the "plotSpecificGenotype" function of the R package Genphen.

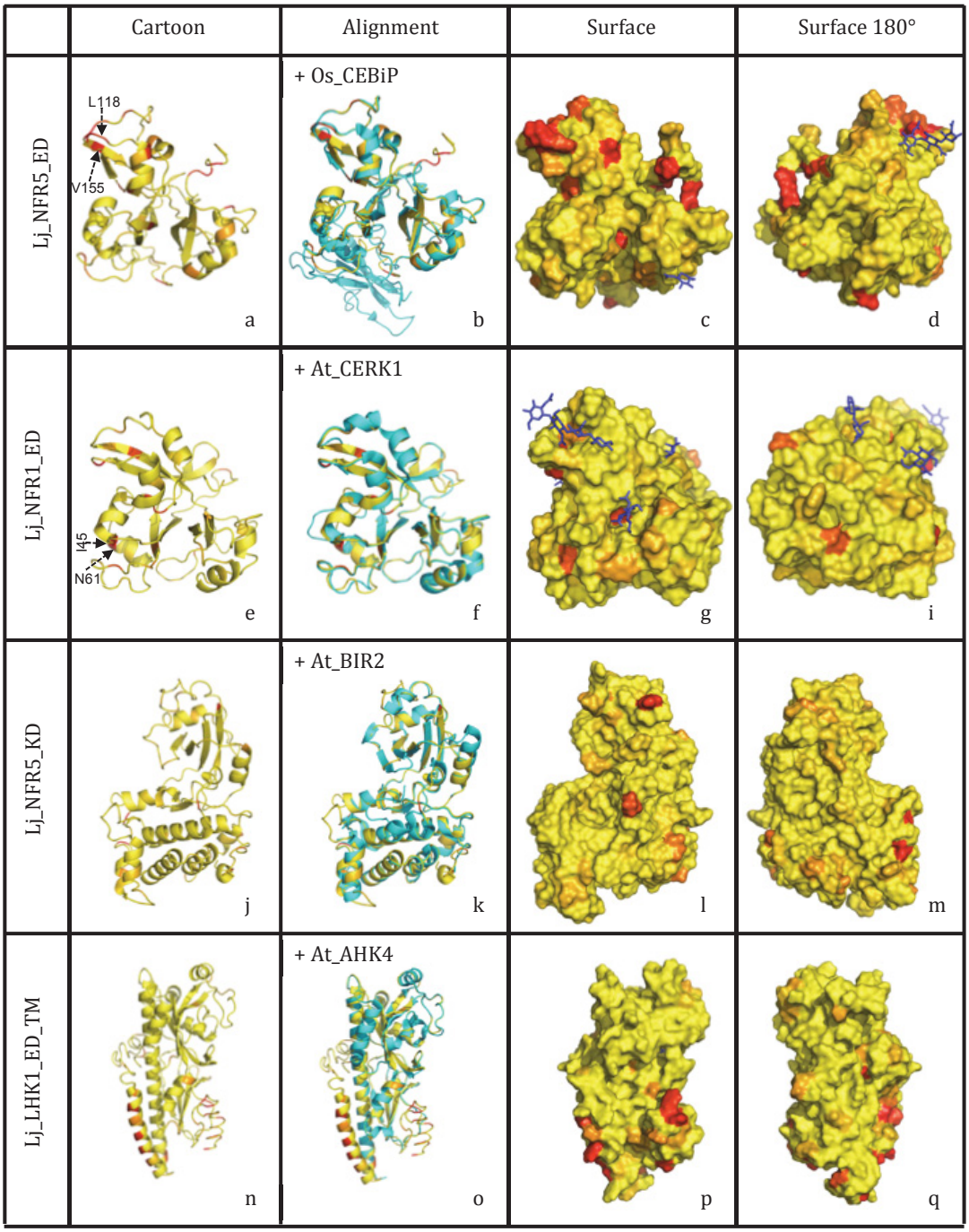
Cartoon
Alignment
Surface
Surface 180°
Lj_NFR5_ED
Lj_NFR1_ED
Lj_NFR5_KD
Lj_LHK1_ED_TM
+ Os_CEBiP
+ At_CERK1
+ At_BIR2
+ At_AHK4
L118
V155
I45
N61
a
b
c
d
e
f
g
i
j
k
l
m
n
o
p
q

Figure 45. Structural analysis of specific domains from the NFR5, NFR1 and LHK1 proteins. The homology models of the NFR5 extracellular region (Lj_NFR5_ED), NFR1 extracellular region (Lj_NFR1_ED), NFR5 kinase domain (Lj_NFR5_KD) and the LHK1 transmembrane and extracellular regions (Lj_LHK1_ED_TM) were calculated using the I-Tasser web server under the default settings in April, 2017 (shown in different rows). The first column shows the molecular structure of the best model computed by I-Tasser. Amino acid residues are colored according to their omega values inferred in PAML with red indicating high omega values and yellow indicating low omega values. The second column shows the alignment with the best structural match found in the PDB database (showed in cyan). The best-known matching structures were aligned and visualized with the molecular graphics software PyMol (version 1.3). The third column shows the surface of the predicted model and the ligands that co-precipitated with the best-known structure as deep blue sticks. The fourth column is the 180° rotation view of the third column. a) Homology model of the NFR5 extracellular domain (amino acid position 26-248; C-score = 0.24). b) Alignment between the model of the NFR5 extracellular domain and the structure of the extracellular domain of *Oryza sativa* chitin elicitor-binding protein (OsCEBiP, PDB id 5JCE). c) Surface of the predicted structure of the NFR5 extracellular domain. The blue sticks are N-Acetyl Glucosamine molecules that co-precipitated with 5JCE. d) 180° rotation view of figure c. e) Homology model of the NFR1 extracellular domain (amino acid position 18-222; C-score = 1.00). f) Alignment between the model of the NFR1 extracellular domain and the structure of *Arabidopsis thaliana* chitin elicitor receptor kinase 1 extracellular domain (AtCERK1, PDB id 4EBZ). g) Surface of the predicted structure of the NFR1 extracellular domain. The blue sticks are N-Acetyl Glucosamine molecules that co-precipitated with the 4EBZ. h) 180° rotation view of figure g. i) Homology model of the NFR5 kinase domain (amino acid position 272-595; C-score = -0.11). j) Alignment between the model of the NFR5 kinase domain and the structure of *Arabidopsis thaliana* BAK1-interacting receptor-like kinase 2 (AtBIR2, PDB id 4L68). k) Surface of the predicted structure of the NFR5 kinase domain. l) 180° rotation view of figure k. m) Homology model of the transmembrane and extracellular region of LHK1 or Lotus histidine kinase 1 (amino acid position 1-357; C-score = -2.00). n) Alignment between the model of LHK1 transmembrane and the extracellular region and the structure of *Arabidopsis thaliana* histidine kinase 4 (AtAHK4, PDB id 3T4J). o) Surface of the predicted structure of LHK1. The blue sticks represent the isopentenyladenine molecule that co-precipitated with the 3T4J. p) 180° rotation view of figure o.

Chapter 4

Discussion

4.1 The LysM-RLKs evolve faster than the LRR-RLKs

Many LysM-RLKs in legumes are recognized as determinants of symbiotic nitrogen fixation (Madsen *et al.*, 2003 ; Radutoiu *et al.*, 2003 ; Kawaharada *et al.*, 2015 ; Indrasumunar *et al.*, 2011 , 2010 ; Limpens *et al.*, 2003 ; Amor *et al.*, 2003 ; Zhukov *et al.*, 2008). Furthermore, some LysM-RLKs, both from legumes and outside this plant family, are associated with arbuscular mycorrhizal symbiosis or chitin perception (Leppyanen *et al.*, 2017 ; Bozsoki *et al.*, 2017 ; Fuechtbauer *et al.*, 2017 ; Miyata *et al.*, 2014 ; Wan *et al.*, 2012 ; Miya *et al.*, 2007 ; Op den Camp *et al.*, 2011 ; Zeng *et al.*, 2012 ; Buendia *et al.*, 2016 ; Cao *et al.*, 2014 ; Brulé *et al.*, 2019). Together, the LysM-RLK gene family performs a critical function of recognizing microbe-associated molecular patterns (MAMPs) or secreted symbiotic effectors.

The *Lotus japonicus* genome has 17 LysM-RLKs, where four of those (i.e. NFR5, NFR1, Lys3/EPR3 and Lys1/NFRe) are symbiotic receptors and two others (i.e. Lys12 and Lys6) are immune receptors. We found evidence of adaptive evolution in the LysM-RLK genes named the NFR5, NFR1, Lys1/NFRe, Lys2, Lys4, Lys6, Lys11, Lys12, Lys13, Lys14, Lys15 and Lys16 (Figure 12). In contrast, the extracellular regions of LRR-type receptor like kinases SymRK, HAR1 and Klavier did not displayed strong evidence of adaptive evolution and mainly evolved neutrally (Figure 13). The kinase domains of SymRK, HAR1 and Klavier are highly conserved and non-synonymous mutations had been removed by purifying selection. With the exception of NFR1, the

kinase domains of the majority of LysM-RLKs have experienced positive selection or carry sites with a *Ka/Ks* ratio above one (Figure 8b, Figure 14a, Figure 15a, Figure 16a, Figure 17a). All these results suggest that the LysM-RLKs are evolving faster then the LRR-RLKs within the genus *Lotus*.

To our surprise, the cytokinin receptor LHK1 also showed evidence of adaptive evolution (Figure 10b). LHK1 is a major receptor for the endogenous plant hormone cytokinin and is unlikely to interact directly with nod factors. Furthermore, the location of the putatively selected sites within the two transmembrane regions may have less to do with microbe discrimination per se, and more to do with intrinsic protein-membrane interactions. The functional effect of these amino acid variants present among *Lotus* species on protein function await further analysis.

4.2 Specific mutations in NFR1, NFR5 and SymRK are associated with symbiont-specificity

The site V155 of NFR5 was identified as a putative target of positive selection and variation at this site was correlated with symbiotic preferences. Site P399 of NFR5 kinase domain also displayed strong correlation with the nodulation competency. In this case, all alleles from individuals of the Meso-clade encode a Proline, while all alleles from individuals in the Brady-clade encode a Serine (Figure 4). This site is within the predicted kinase domain of NFR5. It has been shown via *in vitro* assays that the kinase domain of NFR5 from *L. japonicus* is catalytically inactive (Madsen *et al.*, 2011). However, NFR5 forms a dimer with NFR1 and the role of the NFR5 kinase domain in dimer formation is still not clear. Recently it has been shown that the NFR5-interacting cytoplasmic kinase 4 (or NiCK4) can transphosphorylate the NFR5 kinase domain (Wong *et al.*, 2019). Domain swap analyses between the Lys11 and NFR5 revealed that the NFR5 kinase domain could complement the *nfr5* mutant phenotype when tagged with the Lys11 extracellular domain (Rasmussen *et al.*, 2016). It will be interesting to test in future experiments if mutations at sites P399 or Y371 lead to changes in microbe specificity.

Our analysis indicated that site I45 in NFR1 was a potential target of positive selection. The closest homolog of *Lj* NFR1 in *Arabidopsis thaliana* is a major chitin elicitor receptor kinase known as the CERK1 (Liu *et al.*, 2012). In *At* CERK1, the LysM2 domain physically interacts with the chitin molecule (Liu *et al.*, 2012). However, the putatively selected sites within NFR1 are localized to the LysM1 region. This might

indicate that despite high sequence conservation between *At* CERK1 and *Lj* NFR1, different LysM domains confer independent resistance and symbiosis functions.

Given that many plant LysM-RLKs perform dual functions, distinguishing symbiotic or pathogenic signals from fungal or bacterial origins (Zhang *et al.*, 2015 ; Rey *et al.*, 2013 ; Miyata *et al.*, 2014 ; Leppyanen *et al.*, 2017 ; Gibelin-Viala *et al.*, 2019), it is possible that different LysM domains of the same receptor have evolved in response to different microbial signals. This may be true for the *Medicago truncatula* homolog of *Lj* NFR5, the NFP, and the *Mt* Lyk9_ which are involved in both symbiosis and plant immunity (Rey *et al.*, 2013 ; Gibelin-Viala *et al.*, 2019). Alternatively, a structural comparison of *Lj* NFR5 and *Oryza sativa* chitin elicitor binding protein, CEBiP, suggests that symbiotic and pathogenic specificity might both be controlled by the LysM2 domain (Figure 45).

Variation in the LRR-type RLK, SymRK, was also correlated with symbiont-specificity (Table 5). In particular, variation at sites P475 and S374 of SymRK was associated with *Rhizobium*-specificity (Figure 9 & 43). It is known that the extracellular part of SymRK is cleaved from the cytoplasmic domains after perceiving symbiotic effectors and the cleavage is necessary for nodulation (Antolhén-Llovera *et al.*, 2014). This observation makes site P475 of SymRK an excellent candidate of further investigations of effector dependent cleavage of its extracellular region.

Variation in Lys12, Lys13, Lys16 and Lys1/NFRe showed moderate correlation with *Rhizobium*-specificity (Table 5). However, for the two special RLKs, the cytokinine receptor LHK1 and the pseudogene Lys21, no significant associations between sequence variation and symbiont specificity were detected.

4.3 The LysM genes are major evolutionary targets for shifts in microbe discrimination

Genetic variation in NFR1 and NFR5 is associated with symbiont discrimination and several residues in both proteins appear to have undergone positive selection. In agreement with other studies, our infection analysis suggests that compatibility is determined early during the interaction, probably through receptor-mediated entry (Figure 39).

Among the uncharacterized LysM-RLKs, Lys16 emerged as the most likely gene to contribute to plant-microbe interaction. The noticeable characteristics of both

Lys16 and Lys6 are adaptively evolving extracellular domains and relatively conserved cytoplasmic kinase domain (Figure 8a & 14b). Moreover, clade specific fixed differences in Lys16 were correlated with *Mesorhizobium* compatibility. The extracellular domains of Lys2 are evolving adaptively (Figure 16b). It will be very interesting if the kinase domain of Lys2 displays indication of purifying selection. *Lj* Lys6 is an immune receptor and is required for resistance against the fungal pathogen *Botrytis cinerea* (Bozsoki *et al.*, 2017).

The extremely high evolutionary rates detected in Lys13 and Lys14 suggests that these genes are responsive to rapidly evolving selective agents, such as pathogenic microorganisms (Figure 17). Recently, it has been shown that the *Lj* Lys12 is a major receptor for the plant pathogenic oomycete *Phytophthora palmivora* (Fuechtbauer *et al.*, 2017). In contrast, Lys15 displayed low rate of amino acid evolution and might be performing an essential function. However, similar to the immune receptor Lys12, adaptively evolving sites were present at the signal peptide region and in the kinase domain of Lys15.

4.4 The role of SymRK in microbe discrimination

Among the three LRR-type RLKs analyzed, SymRK showed the strongest correlation with *Rhizobium* compatibility. The SymRK sites T58, P475, A378 and S374 exhibited strong association with *Mesorhizobium loti* compatibility. Furthermore, other sites of the SymRK (e.g. sites S173 and L390) displayed clade specific fixed differences and are likely to contribute to quantitative differences in *Rhizobium* specificity.

In addition, I found evidence of elevated interchange between the amino acids Leucine, Isoleucine, Valine and Phenylalanine in the extracellular region of the LRR-type symbiosis receptor kinase, the SymRK. These clade specific fixed differences may offer flexibility to the extracellular ligand-binding domains of SymRK and are often associated with *Mesorhizobium* compatibility data (Figure 9).

A common characteristic of all RLKs analyzed in this study is the presence of leucine rich segment in the signal peptide region (Figure 7, Figure 8b, Figure 11, Figure 14a, Figure 17b). These regions are evolutionary hot spots in both LysM and LRR-type RLKs. Adaptive evolution within the signal peptide region might indicate altered utilization of these genes in different species. Alternatively, these protein variants

might be selectively neutral, as they are not part of the mature protein. Experimental substitutions of these sites would be useful to understand the fine-tuning of these RLKs.

SymRK is required for association with arbuscular mycorrhizae (widespread among land plants) and thus extensive variation among SymRK sequences within the genus Lotus was not expected. However, based on the relatively high exchangeability between Leucine, Isoleucine, Valine, and the aromatic Phenylalanine and strong correlation with *Mesorhizobium loti* compatibility data, SymRK emerged as another candidate determinant for *Rhizobium* specificity. Considering all available data, I propose a model whereby the NFRs, together with the endosymbiosis receptor SymRK, are the major genetic determinants of *Rhizobium*-specificity (Figure 46). According to this model the sites V155, Y371 and P399 of NFR5, the site I45 of NFR1 and sites T58, S374 and P475 of SymRK are the prime candidates to contribute in *Mesorhizobium-Bradyrhizobium* nod factor specificity. The other adaptively evolving sites of the NFRs are also top candidates for determining specificity, but possibly towards different *Rhizobium* species.

4.5 Extracellular and intracellular domains of symbiosis receptor-like kinases show different modes of selection

Our evolutionary analyses of LysM and LRR-type symbiotic RLKs and their associations with *Mesorhizobium loti* compatibility underscored the importance of NFR5, NFR1 and SymRK as major determinants of rhizobial specificity. However, to our surprise, the main finding of our analysis has been the evolutionary flexibility of the extracellular domains compared to the cytoplasmic parts in six out of seven symbiotic RLKs analyzed in this study: NFR1, NFR5, EPR3, SymRK, HAR1 and Klavier. The NFRe failed to pass the fixed-sites model test. The LysM1 and LysM2 extracellular domains of the symbiotic RLK NFRe contain five codon sites being positively selected, while only one positively selected site is present in the cytoplasmic region and none within the predicted kinase domain (Figure 7b). The widespread action of positive selection in the extracellular domains of these proteins may be linked to plant perception of a large, and as yet undescribed, repertoire of microbial molecules, beyond the well-known nod factors. Identification of symbiotic effectors interacting

with the extracellular domains of SymRK, HAR1 and Klavier and localization of their binding-sites will vastly improve our understanding of how species-specificity in legume- *Rhizobium* symbiosis has evolved.

Lys6 and Lys12 are both characterized as immune receptors, however the Lys12 has signatures of adaptive evolution, while the kinase domain of Lys6 does not. Furthermore, according to our data, many LysM-RLKs with unknown function also possess adaptively evolving sites in their kinase domains. This suggests that the elevated variation at the extracellular region is a feature of symbiotic receptor-like kinases and the immune receptors show a different pattern. With many cultivated legume genome sequencing projects underway and new precise genome editing technology in hand, it may not be long until farmers may control the microbiota associated with their favorite legume crops.

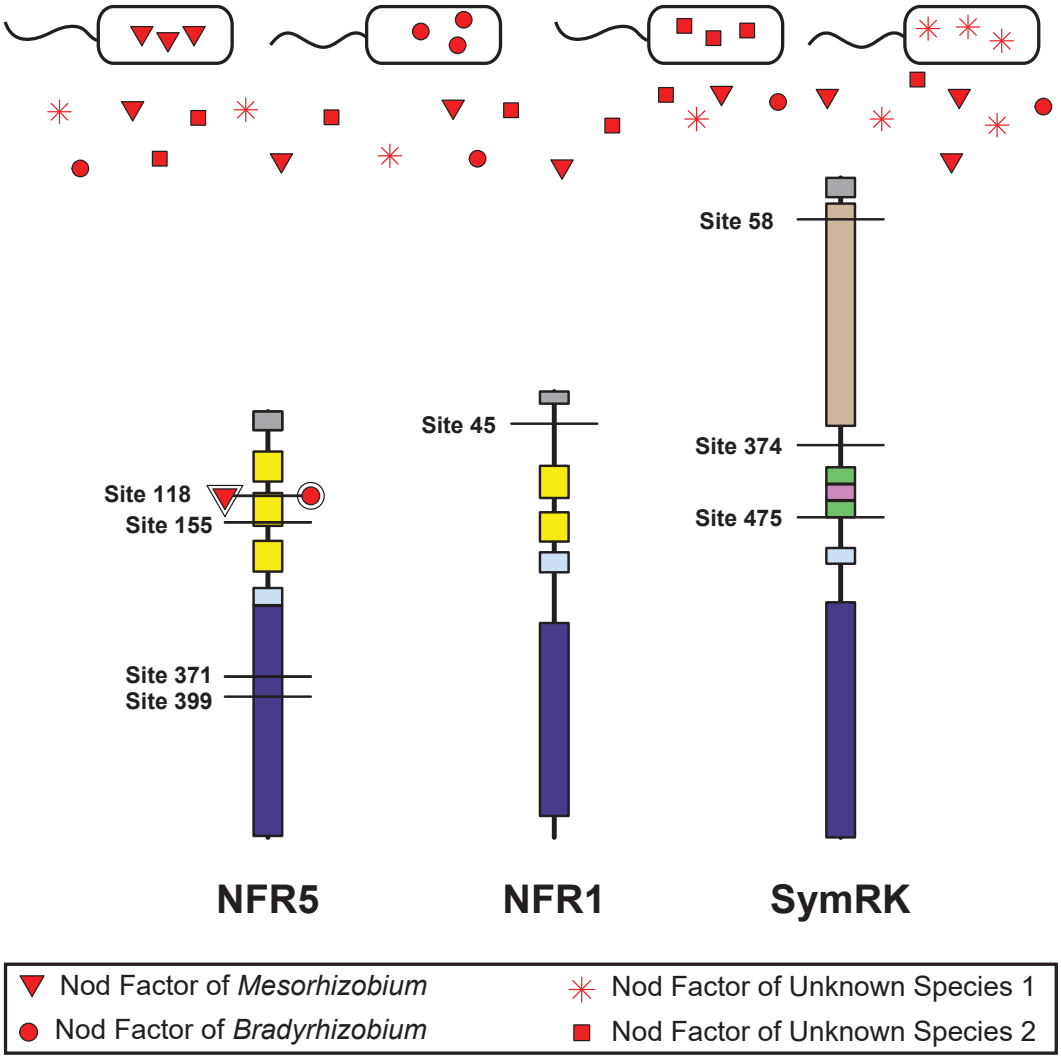

Figure 46. Evolutionary model of Symbiotic receptor-like kinases NFR5, NFR1 and SymRK.

Appendix

Table 5: Genotype-Phenotype association calculated by Genphen

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's d**	**Classification**	**$p-$value**
NFR5 (n=24)								
NFR5	155	V	L	10	3	-2.15	0.73	0.00014
NFR5	371	Y	K	10	2	-2.01	0.78	0.00019
NFR5	155	V	I	10	2	-1.67	0.82	0.00071
NFR5	59	L	F	9	15	-1.35	0.71	0.00343
NFR5	371	Y	N	10	11	-1.25	0.74	0.00999
NFR5	125	V	I	19	5	-1.24	0.76	0.00014
NFR5	398	T	N	14	2	-1.22	0.90	0.00097
NFR5	164	R	K	21	3	-1.22	0.88	0.00004
NFR5	12	S	C	22	2	-1.15	0.94	0.00003
NFR5	154	R	G	10	14	-1.14	0.69	0.01024
NFR5	233	Q	H	22	2	-1.09	0.94	0.00021
NFR5	263	T	A	22	2	-1.09	0.94	0.00021
NFR5	567	S	A	22	2	-1.09	0.93	0.00021
NFR5	563	T	S	22	2	-0.99	0.93	0.01127
NFR5	7	T	P	11	12	-0.90	0.63	0.04233
NFR5	51	T	S	11	13	-0.88	0.64	0.04258
NFR5	109	S	F	11	13	-0.88	0.63	0.04258
NFR5	305	E	D	22	2	-0.78	0.93	0.00184
NFR5	112	T	I	22	2	-0.69	0.94	0.01153
NFR5	525	F	C	22	2	0.51	0.93	0.03716
NFR5	588	H	Q	22	2	0.80	0.93	0.00147
NFR5	398	-	N	8	2	0.90	0.83	0.04984
NFR5	238	G	R	22	2	1.02	0.94	0.00460

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
NFR5	123	D	N	11	10	1.02	0.67	0.02973
NFR5	254	G	V	22	2	1.09	0.93	0.00021
NFR5	104	A	T	10	14	1.14	0.69	0.01024
NFR5	129	L	S	10	14	1.14	0.69	0.01024
NFR5	141	A	D	10	14	1.14	0.69	0.01024
NFR5	221	I	V	10	14	1.14	0.70	0.01024
NFR5	399	P	S	10	5	2.14	0.80	0.00023
NFR5	155	L	I	3	2	3.53	0.84	0.03195
NFR1 left-border or N1LB (n=22)								
N1LB	116	T	S	15	7	-1.26	0.62	0.00680
N1LB	115	E	D	12	10	-1.18	0.70	0.01170
N1LB	204	K	E	20	2	-1.14	0.93	0.00020
N1LB	12	I	F	12	9	-1.12	0.67	0.01930
N1LB	5	T	N	12	9	-1.10	0.66	0.02140
N1LB	59	S	T	15	3	1.05	0.84	0.00230
N1LB	181	D	N	19	3	1.22	0.87	0.00010
N1LB	45	I	T	14	5	1.29	0.65	0.03270
N1LB	10	F	L	11	11	1.33	0.75	0.00550
N1LB	45	I	N	14	2	1.63	0.88	0.00070
N1LB	59	-	T	4	3	3.18	0.97	0.01150
NFR1 right-border or N1RB (n=22)								
N1RB	129	A	S	20	2	1.03	0.93	0.00024
N1RB	263	G	R	20	2	1.03	0.93	0.00024
N1RB	265	N	T	20	2	1.03	0.93	0.00024
N1RB	309	-	Y	20	2	1.03	0.93	0.00024
Lys3 or EPR3 (n=20)								
Lys3 or EPR3	150	T	A	14	5	-1.44	0.64	0.00040
Lys3 or EPR3	240	I	A	10	2	-1.41	0.80	0.00220
Lys3 or EPR3	542	N	E	7	8	-1.39	0.81	0.01780
Lys3 or EPR3	42	M	L	8	12	-1.23	0.75	0.01100
Lys3 or EPR3	206	V	I	8	12	-1.23	0.75	0.01100
Lys3 or EPR3	558	S	N	17	3	-1.07	0.86	0.00080

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Lys3 or EPR3	107	N	K	17	2	-1.04	0.92	0.00080
Lys3 or EPR3	7	T	P	13	2	-1.02	0.89	0.00400
Lys3 or EPR3	17	V	-	16	4	-1.02	0.78	0.00250
Lys3 or EPR3	20	T	P	16	4	-1.02	0.79	0.00250
Lys3 or EPR3	49	T	A	16	4	-1.02	0.78	0.00250
Lys3 or EPR3	14	P	F	11	2	-1.02	0.85	0.03150
Lys3 or EPR3	88	T	A	18	2	-0.98	0.92	0.00090
Lys3 or EPR3	27	F	I	18	2	0.98	0.92	0.00090
Lys3 or EPR3	13	H	Q	15	4	0.99	0.78	0.00440
Lys3 or EPR3	176	P	S	13	4	1.12	0.70	0.00360
Lys3 or EPR3	147	E	G	15	5	1.31	0.68	0.00050
Lys1 or NFRe (n=12)								
Lys1 or NFRe	139	L	V	6	6	-2.90	0.88	0.00070
Lys1 or NFRe	225	G	R	10	2	-2.21	0.83	0.00170
Lys1 or NFRe	69	-	G	4	2	-2.04	0.60	0.03890
Lys1 or NFRe	304	K	R	9	3	-1.94	0.73	0.01700
Lys1 or NFRe	265	S	P	2	6	-1.78	0.73	0.01230
Lys1 or NFRe	112	D	E	5	7	-1.73	0.75	0.00920
Lys1 or NFRe	238	F	L	5	7	-1.73	0.76	0.00920
Lys1 or NFRe	266	A	T	5	7	-1.73	0.75	0.00920
Lys1 or NFRe	558	T	K	9	2	-1.51	0.75	0.00280
Lys1 or NFRe	413	V	M	10	2	-1.29	0.77	0.00380
Lys1 or NFRe	517	N	D	10	2	-1.29	0.79	0.00380
Lys1 or NFRe	73	-	V	4	8	-1.21	0.58	0.03910
Lys1 or NFRe	74	-	F	4	8	-1.21	0.59	0.03910
Lys1 or NFRe	75	-	S	4	8	-1.21	0.58	0.03910
Lys1 or NFRe	285	H	R	4	8	-1.21	0.58	0.03910
Lys1 or NFRe	54	V	I	4	8	1.21	0.58	0.03910
Lys1 or NFRe	506	G	S	10	2	1.29	0.79	0.00380
Lys1 or NFRe	4	R	K	5	7	1.73	0.75	0.00920
Lys1 or NFRe	170	T	S	5	7	1.73	0.75	0.00920
Lys1 or NFRe	569	T	A	5	7	1.73	0.75	0.00920

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Lys1 or NFRe	222	L	F	7	5	2.24	0.74	0.00510
Lys1 or NFRe	265	T	S	4	2	4.27	0.98	0.00330
Lys4 (n=17)								
Lys4	567	P	A	11	4	-1.21	0.65	0.00560
Lys4	110	T	S	4	13	-1.20	0.67	0.04590
Lys4	130	T	-	6	2	-1.19	0.63	0.04540
Lys4	289	V	A	13	4	-1.17	0.70	0.00260
Lys4	26	Y	N	14	2	-1.01	0.90	0.00300
Lys4	133	R	H	15	2	-0.95	0.91	0.00330
Lys4	227	S	A	15	2	-0.95	0.90	0.00330
Lys4	364	T	R	15	2	-0.95	0.91	0.00330
Lys4	459	N	K	15	2	-0.95	0.91	0.00330
Lys4	612	L	F	15	2	-0.95	0.91	0.00330
Lys4	219	S	P	14	2	-0.88	0.90	0.00750
Lys4	599	E	D	14	2	-0.88	0.90	0.00750
Lys4	257	A	E	15	2	0.95	0.91	0.00330
Lys4	260	A	S	15	2	0.95	0.91	0.00330
Lys4	269	H	Q	15	2	0.95	0.91	0.00330
Lys4	304	P	S	15	2	0.95	0.91	0.00330
Lys4	400	F	L	15	2	0.95	0.91	0.00330
Lys4	552	I	T	14	2	1.01	0.90	0.00300
Lys4	564	L	M	14	2	1.01	0.90	0.00300
Lys4	103	K	R	13	4	1.02	0.73	0.00700
Lys4	111	D	N	14	3	1.05	0.82	0.00300
Lys4	91	I	M	4	13	1.20	0.67	0.04590
Lys4	511	I	V	5	2	1.83	0.71	0.02170
Lys4	271	-	H	5	3	2.00	0.70	0.02170
Lys6 (n=9)								
Lys6	19	G	E	2	7	1.56	0.68	0.00880
Lys6	64	E	Q	7	2	1.56	0.67	0.00880
Lys6	102	T	S	2	7	1.56	0.68	0.00880
Lys6	161	R	K	2	7	1.56	0.68	0.00880

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Lys6	485	E	D	2	7	1.56	0.68	0.00880
Lys6	114	M	L	3	6	2.36	0.76	0.00450
Lys11 (n=17)								
Lys11	546	S	L	6	2	-1.35	0.66	0.02920
Lys11	43	V	M	14	3	-1.05	0.83	0.00300
Lys11	348	V	L	14	3	-1.05	0.83	0.00300
Lys11	407	T	A	13	2	-1.04	0.89	0.00370
Lys11	213	S	R	10	2	-1.02	0.86	0.01370
Lys11	40	S	P	15	2	-0.95	0.91	0.00330
Lys11	64	T	S	15	2	-0.95	0.91	0.00330
Lys11	109	S	N	15	2	-0.95	0.90	0.00330
Lys11	175	N	K	15	2	-0.95	0.91	0.00330
Lys11	215	Q	E	15	2	-0.95	0.91	0.00330
Lys11	352	K	E	15	2	-0.95	0.91	0.00330
Lys11	471	P	H	15	2	-0.95	0.91	0.00330
Lys11	517	I	F	15	2	-0.95	0.91	0.00330
Lys11	587	T	A	15	2	-0.95	0.90	0.00330
Lys11	110	N	H	14	2	-0.89	0.90	0.00680
Lys11	50	D	E	15	2	0.95	0.91	0.00330
Lys11	82	A	S	15	2	0.95	0.91	0.00330
Lys11	144	D	Y	15	2	0.95	0.91	0.00330
Lys11	173	H	Q	15	2	0.95	0.91	0.00330
Lys11	294	E	G	15	2	0.95	0.91	0.00330
Lys11	297	G	S	15	2	0.95	0.91	0.00330
Lys11	380	H	Y	15	2	0.95	0.91	0.00330
Lys11	410	R	S	15	2	0.95	0.91	0.00330
Lys11	534	A	V	15	2	0.95	0.91	0.00330
Lys11	29	-	D	14	3	1.05	0.82	0.00300
Lys11	296	A	V	14	3	1.05	0.83	0.00300
Lys11	23	-	H	13	4	1.17	0.71	0.00260
Lys11	500	I	L	8	2	1.24	0.75	0.01350
Lys11	526	E	G	4	3	3.03	0.95	0.01820

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Lys12 (n=16)								
Lys12	325	Q	P	11	5	-1.43	0.60	0.00240
Lys12	87	R	H	12	4	-1.39	0.65	0.00130
Lys12	136	S	L	12	4	-1.39	0.66	0.00130
Lys12	256	T	S	12	4	-1.39	0.66	0.00130
Lys12	362	N	D	12	4	-1.39	0.65	0.00130
Lys12	364	L	F	12	4	-1.39	0.64	0.00130
Lys12	368	S	C	12	4	-1.39	0.66	0.00130
Lys12	432	P	A	12	4	-1.39	0.66	0.00130
Lys12	572	V	I	12	4	-1.39	0.66	0.00130
Lys12	588	S	E	12	4	-1.39	0.65	0.00130
Lys12	622	T	S	12	4	-1.39	0.66	0.00130
Lys12	366	S	A	12	3	-1.34	0.76	0.00130
Lys12	76	Y	H	13	3	-1.21	0.80	0.00170
Lys12	246	Q	L	13	3	-1.21	0.81	0.00170
Lys12	294	L	F	13	3	-1.21	0.80	0.00170
Lys12	306	G	D	13	3	-1.21	0.80	0.00170
Lys12	497	R	K	13	3	-1.21	0.79	0.00170
Lys12	625	R	K	13	3	-1.21	0.81	0.00170
Lys12	163	I	L	10	6	-1.21	0.64	0.03810
Lys12	12	S	F	14	2	-1.07	0.89	0.00200
Lys12	54	G	D	14	2	-1.07	0.89	0.00200
Lys12	130	N	K	14	2	-1.07	0.90	0.00200
Lys12	263	P	S	13	3	1.21	0.80	0.00170
Lys12	292	I	S	13	3	1.21	0.80	0.00170
Lys12	301	I	T	13	3	1.21	0.80	0.00170
Lys12	311	D	G	13	3	1.21	0.80	0.00170
Lys12	314	S	T	13	3	1.21	0.80	0.00170
Lys12	331	M	V	13	3	1.21	0.79	0.00170
Lys12	399	I	T	13	3	1.21	0.80	0.00170
Lys12	485	D	Y	13	3	1.21	0.80	0.00170
Lys12	486	F	L	13	3	1.21	0.80	0.00170

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Lys12	3	C	P	9	4	1.34	0.56	0.00900
Lys12	86	A	P	12	4	1.39	0.66	0.00130
Lys12	248	I	V	12	4	1.39	0.65	0.00130
Lys12	278	L	V	12	4	1.39	0.66	0.00130
Lys12	283	-	G	12	4	1.39	0.65	0.00130
Lys12	284	-	V	12	4	1.39	0.65	0.00130
Lys12	285	-	F	12	4	1.39	0.66	0.00130
Lys12	341	F	I	12	4	1.39	0.65	0.00130
Lys12	365	N	S	12	4	1.39	0.66	0.00130
Lys12	599	F	V	12	4	1.39	0.66	0.00130
Lys12	165	P	L	9	7	1.88	0.81	0.00420
Lys13 (n=23)								
Lys13	203	Q	R	20	2	-2.34	0.91	0.00380
Lys13	238	P	T	21	2	-1.95	0.91	0.00000
Lys13	313	Q	L	13	5	-1.43	0.65	0.00070
Lys13	269	T	A	11	10	-1.27	0.74	0.00830
Lys13	318	V	G	6	3	-1.24	0.53	0.04920
Lys13	165	G	D	9	11	-1.23	0.71	0.02670
Lys13	171	V	D	14	3	-1.20	0.81	0.00120
Lys13	349	Y	F	15	2	-1.16	0.91	0.00100
Lys13	550	Y	F	11	12	-1.10	0.70	0.01680
Lys13	449	Q	E	21	2	-1.05	0.94	0.00040
Lys13	501	V	L	21	2	-1.05	0.94	0.00040
Lys13	585	N	D	21	2	-1.05	0.93	0.00040
Lys13	579	V	T	12	2	-1.04	0.88	0.00630
Lys13	133	T	N	19	4	-1.03	0.82	0.00080
Lys13	538	T	I	20	3	-1.02	0.88	0.00070
Lys13	272	D	G	20	3	-0.98	0.86	0.03700
Lys13	574	L	I	21	2	-0.98	0.93	0.00700
Lys13	295	V	L	20	3	-0.88	0.88	0.00890
Lys13	8	T	R	15	2	-0.86	0.91	0.02080
Lys13	174	N	D	16	2	-0.85	0.91	0.03970

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Lys13	527	V	I	20	3	-0.85	0.88	0.01290
Lys13	58	V	I	21	2	-0.73	0.94	0.00390
Lys13	491	V	I	21	2	-0.73	0.93	0.00390
Lys13	72	K	R	21	2	0.64	0.93	0.02090
Lys13	26	F	L	21	2	0.73	0.93	0.00390
Lys13	360	G	V	19	3	0.83	0.87	0.01450
Lys13	317	A	S	15	3	0.86	0.84	0.01810
Lys13	413	I	V	21	2	0.98	0.93	0.00700
Lys13	607	L	M	21	2	0.98	0.93	0.00700
Lys13	345	E	V	19	2	0.99	0.93	0.00780
Lys13	262	-	R	12	2	1.01	0.87	0.00650
Lys13	425	N	Y	19	4	1.03	0.82	0.00160
Lys13	317	A	P	15	2	1.03	0.91	0.00240
Lys13	145	A	F	18	5	1.07	0.76	0.00100
Lys13	318	A	G	10	3	1.09	0.71	0.01330
Lys13	579	M	T	8	2	1.16	0.81	0.01830
Lys13	175	F	L	8	15	1.19	0.67	0.03150
Lys13	134	L	I	21	2	1.33	0.92	0.01030
Lys13	8	K	R	6	2	1.36	0.64	0.02910
Lys13	186	E	K	4	19	2.62	0.85	0.00020
Lys13	234	F	Y	4	19	2.62	0.85	0.00020
Lys14 (n=16)								
Lys14	23	I	-	7	3	-2.04	0.61	0.00340
Lys14	5	T	S	2	13	-1.58	0.84	0.00350
Lys14	108	I	A	10	2	-1.30	0.83	0.00360
Lys14	118	T	N	8	4	-1.24	0.55	0.03410
Lys14	27	S	-	13	3	-1.16	0.80	0.00230
Lys14	24	I	-	13	2	-1.12	0.89	0.00230
Lys14	26	F	-	13	2	-1.12	0.89	0.00230
Lys14	9	S	I	12	3	-1.09	0.78	0.00510
Lys14	25	S	-	14	2	-1.03	0.90	0.00260
Lys14	58	V	I	14	2	-1.03	0.90	0.00260

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Lys14	4	F	C	13	2	-0.97	0.90	0.00570
Lys14	28	H	-	13	2	-0.97	0.89	0.00570
Lys14	15	I	M	10	5	0.96	0.53	0.04590
Lys14	229	S	W	14	2	1.03	0.90	0.00260
Lys14	230	K	N	13	2	1.12	0.90	0.00230
Lys14	46	D	N	13	3	1.16	0.81	0.00230
Lys14	314	A	P	7	2	1.34	0.71	0.01670
Lys14	17	M	V	5	7	1.79	0.75	0.04480
Lys15 (n=17)								
Lys15	16	V	M	2	15	-1.59	0.86	0.00490
Lys15	187	V	L	14	3	-1.19	0.81	0.00120
Lys15	297	V	I	14	3	-1.19	0.81	0.00120
Lys15	29	Q	F	14	2	-1.15	0.90	0.00120
Lys15	541	L	F	12	5	-1.03	0.61	0.02550
Lys15	404	A	D	12	5	1.03	0.61	0.02550
Lys15	539	F	L	12	5	1.03	0.61	0.02550
Lys15	171	G	R	15	2	1.06	0.91	0.00140
Lys15	293	K	R	15	2	1.06	0.91	0.00140
Lys15	137	A	G	14	3	1.19	0.82	0.00120
Lys15	292	K	R	14	3	1.19	0.81	0.00120
Lys15	295	D	E	14	3	1.19	0.81	0.00120
Lys16 (n=16)								
Lys16	345	M	I	2	8	-3.60	0.98	0.00620
Lys16	578	T	S	5	5	-2.40	0.89	0.01590
Lys16	591	E	D	5	5	-2.40	0.89	0.01590
Lys16	596	T	I	5	5	-2.40	0.90	0.01590
Lys16	226	K	I	5	2	-2.06	0.65	0.01450
Lys16	593	V	I	6	4	-1.96	0.66	0.01260
Lys16	409	S	A	6	4	-1.50	0.65	0.03080
Lys16	662	Q	H	7	3	-1.42	0.57	0.01760
Lys16	53	L	H	6	3	-1.37	0.60	0.03510
Lys16	263	T	S	8	2	-1.12	0.81	0.02090

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Lys16	5	P	S	8	2	1.12	0.81	0.02090
Lys16	425	L	M	7	3	1.42	0.56	0.01760
Lys16	623	M	T	7	3	1.42	0.58	0.01760
Lys16	226	K	N	5	3	2.01	0.69	0.01710
Lys16	581	-	N	5	3	2.26	0.71	0.01450
Lys16	504	I	L	5	5	2.40	0.89	0.01590
Lys16	579	N	S	5	5	2.40	0.89	0.01590
Lys16	677	D	E	5	5	2.40	0.89	0.01590
SymRK (n=21)								
SymRK	58	T	R	9	2	-2.24	0.83	0.00066
SymRK	374	S	H	10	2	-1.93	0.79	0.00090
SymRK	135	R	K	19	2	-1.30	0.92	0.00003
SymRK	218	S	-	19	2	-1.30	0.93	0.00003
SymRK	357	T	I	15	4	-1.28	0.74	0.00064
SymRK	279	R	K	17	4	-1.28	0.78	0.00026
SymRK	58	T	S	9	9	-1.25	0.73	0.01789
SymRK	173	S	L	10	11	-1.24	0.75	0.01034
SymRK	390	L	F	10	11	-1.24	0.75	0.01034
SymRK	10	L	F	19	2	-1.10	0.93	0.00363
SymRK	178	L	I	19	2	-1.10	0.92	0.00363
SymRK	207	Q	K	19	2	-1.10	0.92	0.00363
SymRK	231	V	F	19	2	-1.10	0.93	0.00363
SymRK	242	T	K	19	2	-1.10	0.93	0.00363
SymRK	911	T	I	19	2	-1.10	0.92	0.00363
SymRK	374	S	P	10	9	-1.08	0.70	0.03275
SymRK	175	G	E	4	17	-0.95	0.78	0.02939
SymRK	183	T	N	19	2	-0.85	0.93	0.00204
SymRK	351	I	M	19	2	0.85	0.93	0.00204
SymRK	402	G	S	19	2	0.85	0.93	0.00204
SymRK	425	A	T	19	2	0.85	0.92	0.00204
SymRK	721	I	V	19	2	0.85	0.92	0.00204
SymRK	228	I	V	4	17	0.95	0.78	0.02939

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
SymRK	514	K	R	4	17	0.95	0.78	0.02939
SymRK	393	K	Q	11	10	0.97	0.66	0.03916
SymRK	332	K	Q	13	2	0.97	0.89	0.00550
SymRK	685	K	M	16	2	1.02	0.91	0.00738
SymRK	13	A	V	18	3	1.08	0.86	0.00117
SymRK	391	P	S	19	2	1.10	0.93	0.00363
SymRK	522	G	S	19	2	1.10	0.93	0.00363
SymRK	584	F	Y	19	2	1.10	0.93	0.00363
SymRK	462	D	N	18	3	1.12	0.86	0.00065
SymRK	181	F	L	10	11	1.24	0.75	0.01034
SymRK	215	S	T	10	11	1.24	0.74	0.01034
SymRK	75	R	S	12	7	1.26	0.63	0.01169
SymRK	378	A	V	17	4	1.28	0.78	0.00026
SymRK	396	A	E	19	2	1.30	0.92	0.00003
SymRK	475	P	S	13	8	2.20	0.84	0.00003
HAR1 (n=16)								
HAR1	542	N	M	3	13	-2.17	0.78	0.00110
HAR1	20	W	E	7	2	-1.78	0.75	0.00480
HAR1	24	Y	C	5	11	-1.62	0.65	0.00860
HAR1	502	G	E	5	11	-1.62	0.65	0.00860
HAR1	505	E	D	5	11	-1.62	0.66	0.00860
HAR1	666	V	I	5	10	-1.52	0.63	0.01210
HAR1	413	S	T	14	2	-1.44	0.86	0.00520
HAR1	4	R	K	8	2	-1.41	0.70	0.00740
HAR1	116	S	W	11	2	-1.20	0.84	0.01290
HAR1	145	G	D	10	3	-1.16	0.68	0.00910
HAR1	2	R	G	14	2	-1.10	0.90	0.00170
HAR1	5	V	M	14	2	-1.10	0.90	0.00170
HAR1	19	R	K	14	2	-1.10	0.90	0.00170
HAR1	23	V	L	14	2	-1.10	0.90	0.00170
HAR1	96	L	F	14	2	-1.10	0.90	0.00170
HAR1	109	N	D	14	2	-1.10	0.90	0.00170

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
HAR1	158	S	C	14	2	-1.10	0.90	0.00170
HAR1	365	R	K	14	2	-1.10	0.90	0.00170
HAR1	410	E	A	14	2	-1.10	0.90	0.00170
HAR1	540	R	Q	14	2	-1.10	0.89	0.00170
HAR1	564	R	H	14	2	-1.10	0.89	0.00170
HAR1	972	T	E	14	2	-1.10	0.90	0.00170
HAR1	137	Q	H	13	2	-1.02	0.89	0.00410
HAR1	588	S	T	14	2	0.85	0.90	0.03910
HAR1	31	D	H	13	3	1.04	0.81	0.00690
HAR1	427	P	S	13	2	1.05	0.89	0.00340
HAR1	46	K	Q	14	2	1.10	0.90	0.00170
HAR1	48	H	Q	14	2	1.10	0.90	0.00170
HAR1	99	K	R	14	2	1.10	0.90	0.00170
HAR1	108	N	T	14	2	1.10	0.89	0.00170
HAR1	177	H	Y	14	2	1.10	0.89	0.00170
HAR1	185	G	R	14	2	1.10	0.90	0.00170
HAR1	529	H	Y	14	2	1.10	0.89	0.00170
HAR1	555	M	R	14	2	1.10	0.90	0.00170
HAR1	622	-	S	14	2	1.10	0.90	0.00170
HAR1	639	K	R	14	2	1.10	0.90	0.00170
HAR1	925	A	S	14	2	1.10	0.90	0.00170
HAR1	946	M	V	14	2	1.10	0.90	0.00170
HAR1	112	D	E	13	2	1.19	0.88	0.00140
HAR1	179	A	G	13	2	1.19	0.89	0.00140
HAR1	623	A	S	12	2	1.32	0.85	0.00110
HAR1	623	A	T	12	2	1.32	0.85	0.00110
HAR1	38	E	D	13	2	1.36	0.85	0.00660
HAR1	655	A	V	12	3	1.38	0.74	0.00110
HAR1	197	L	I	14	2	1.44	0.87	0.00520
HAR1	367	L	I	14	2	1.44	0.87	0.00520
HAR1	454	S	P	14	2	1.44	0.86	0.00520
HAR1	133	F	V	3	4	2.51	0.74	0.02910

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Klavier (n=11)								
Klavier	91	S	F	9	2	-1.69	0.76	0.00690
Klavier	203	S	N	9	2	-1.69	0.77	0.00690
Klavier	461	K	E	9	2	-1.69	0.77	0.00690
Klavier	658	L	I	9	2	-1.69	0.77	0.00690
Klavier	752	S	N	9	2	-1.69	0.76	0.00690
Klavier	60	F	L	9	2	1.69	0.76	0.00690
Klavier	446	H	Y	9	2	1.69	0.76	0.00690
Klavier	465	A	V	9	2	1.69	0.76	0.00690
Klavier	598	I	L	9	2	1.69	0.77	0.00690
Klavier	629	F	L	9	2	1.69	0.76	0.00690
Clavata2 (n=19)								
Clavata2	488	D	N	3	4	-2.33	0.72	0.03640
Clavata2	31	N	H	14	2	-1.35	0.90	0.00030
Clavata2	14	H	Q	17	2	-1.18	0.91	0.00730
Clavata2	8	P	S	9	2	-1.10	0.76	0.02120
Clavata2	7	M	I	17	2	-1.07	0.91	0.00060
Clavata2	61	V	G	17	2	-1.07	0.92	0.00060
Clavata2	244	V	I	17	2	-1.07	0.92	0.00060
Clavata2	518	R	G	17	2	-1.07	0.92	0.00060
Clavata2	524	E	D	17	2	-1.07	0.91	0.00060
Clavata2	488	E	D	12	3	-1.01	0.80	0.00810
Clavata2	147	M	T	16	2	1.01	0.92	0.00140
Clavata2	225	F	L	15	2	1.04	0.90	0.00160
Clavata2	17	L	Y	17	2	1.07	0.91	0.00060
Clavata2	19	L	P	17	2	1.07	0.92	0.00060
Clavata2	23	I	V	17	2	1.07	0.92	0.00060
Clavata2	33	I	V	17	2	1.07	0.91	0.00060
Clavata2	34	D	E	17	2	1.07	0.92	0.00060
Clavata2	41	A	E	17	2	1.07	0.92	0.00060
Clavata2	71	I	V	17	2	1.07	0.91	0.00060
Clavata2	486	T	A	17	2	1.18	0.92	0.00730

Table 5: Genotype-Phenotype association calculated by Genphen *(continued)*

Gene	**Site**	**Allele1**	**Allele2**	**Count1**	**Count2**	**Cohen's** d	**Classification**	$p-$**value**
Clavata2	2	F	L	16	3	1.18	0.84	0.00050
Clavata2	31	N	S	14	3	1.40	0.80	0.00030
Clavata2	258	V	A	17	2	1.69	0.91	0.00180

Table 6: List of Cloning Primers

SL	**Name**	**Sequence (3' - 5')**	**Target**
1	Lys1_F	ACACTAAGCTACAACCACTAAC	Lys1 or NFRe
2	Lys1_R	CCTTTCCAGACTATATTCCTACTC	Lys1 or NFRe
3	Lys2_F	TGAACCCACTCTCAGACATCC	Lys2(L2LB)
4	Lys2_Mid_R	GCAAGAAATTCCCGTGATGC	Lys2(L2LB)
5	Lys3_F	CATCATCATTCATGGACTCATCGTA	Lys3 or EPR3
6	Lys3_R	ACTCATTAACCCGATCTCATCCTAT	Lys3 or EPR3
7	Lys4_Fwd	CATCCCTTCAGTTCACAGTATCAG	Lys4
8	Lys4_Rev	GAGAGATCTTGCTGCATAACCTATC	Lys4
9	Lys6_F	TCTAGCTCAAGCCTCCTACTAC	Lys6
10	Lys6_R	GTCACCATCCTATCTTCCAGAC	Lys6
11	Lys11_F	GCCTCTAGCAACAATGACTTCC	Lys11
12	Lys11_R	ATTGGGACACCGCAATCAAC	Lys11
13	Lys12_F	ATCAAACCCAACTGAATGC	Lys12
14	Lys12_R	CAACTGCTGTAACATTGAGAG	Lys12
15	Lys13_Fwd	GATGTTGCTATGCTCCTCTTCTAC	Lys13
16	Lys13_Rev	CAACTATCTGCTTTCAGAGATTTGG	Lys13
17	Lys14_F	CTAGTGTAGCAAGCACTTTCC	Lys14
18	Lys14_R	CTGGGGATGAAGCAATTGG	Lys14
19	Lys15_F	TGCAACTCATTAGCGCTTGAC	Lys15
20	Lys15_R	TGTTCTGCACTTGGCTAAACC	Lys15
21	Lys16_F	CCACTCCCCATGTTAGAGC	Lys16
22	Lys16_R	GCATAGCCTACAGGCACTC	Lys16
23	Lys21_F	AGAATACCGCGTGAATCGTG	Lys21
24	Lys21_R	TACAAGCACAACCCCGAAAC	Lys21
25	SymRK_Fwd	GTTCTACAACCCTTTGGGGTAAATTC	SymRK
26	SymRK_Rev	CACTTGACCCAGATTAGTCATATCAC	SymRK
27	NFR11_R	CACCCTGACTCGATTTACATTCAAAG	NFR1(N1RB)
28	NFR1_Mid_F	ACTGCTAGTGCTACAGGTCTTAC	NFR1(N1RB)
29	NFR1_LB_Fwd	GATCAAAACCTGGTAGAGAGTC	NFR1(N1LB)
30	NFR1_LB_Rev	CTTCGCTAGTTCCTGATATGAG	NFR1(N1LB)
31	NFR5_F	CAATGCTGATTCCCTCTGATAAAG	NFR5
32	NFR5_R	CTTAACGTGCAGTAATGGAAGTC	NFR5
33	Har1_Fwd	TTATACGCCAGCCACCAAATTATATG	HAR1
34	Har1_Rev	GGAAGTTATTTACACAGCTTCTTCTG	HAR1
35	Klv_Fwd	CAATCTCTCAAAATTAGGGCTAGGG	Klavier
36	Klv_Rev	CATGGCAAATTAACACGATGGTG	Klavier
37	LHK1_F	CACACATCAATCATTTGTGCTACTTC	LHK1
38	LHK1_R	ACCAACTTTTGCTCTCTACATAAAACC	LHK1
39	Clv2_Fwd	TTCATCACTAACCTAAGTCCAGTGTTC	Clavata2
40	Clv2_Rev	TCCCAATCACAGAATGTGTATGTTC	Clavata2

Note:

L2LB = Lys2 Left-Border

N1LB = NFR1 Left-Border

N1RB = NFR1 Right-Border

Lys6 = The first 34 codon triplets of Lys6 are missing due to the genome version used for primer designing

Table 7: List of Sequencing Primers

SL	**Name**	**Sequence (3' - 5')**	**Target**
1	Clv2_Mid_F	AACTTGGCCTTGTTCTTCTTGAC	Clavata2
2	Clv2_Mid_R	CCAATTCTAGCAGGAATTTCTCC	Clavata2
3	Har1_Mid_Rev	CATTATATTCCGGTGCCGGATTTTC	HAR1
4	Har1_Mid_Fwd	AGGATGTAGTCGAGTGTTTAAAGG	HAR1
5	Har1_LB_Seq_F	CATCACCGTTGGCATGAC	HAR1
6	Har1_LB_Seq_R	CCGGAATTTCTCCGATGAACTC	HAR1
7	Har1_RB_Seq_R	TGAGAATTGTTCCTGTAATGGGTTTG	HAR1
8	HAR1_LB_Seq_Fwd	GAATCTCCGCCTGCTAGAAATG	HAR1
9	HAR1_LB_Rev	AAATTGTTCTCCCAAACCTGAAGC	HAR1
10	HAR1_LB_Seq_Rev	GGGACTGTTCCGGTGAAATTG	HAR1
11	HAR1_Mid_F	GCGGTGGGAGATGAGGTATAAG	HAR1
12	HAR1_Mid_R	CAGCAATGGAGGACATGGACTG	HAR1
13	HAR1_RB_Seq_Rev	GTCACTCTTCTCGTCCACTTTCAG	HAR1
14	Klv_LB_Fwd	TTGGGAAGGTTTCACCTTTGTTC	Klavier
15	Klv_Mid_Fwd	CTTTGTCGTCACTTGGAGATGTTG	Klavier
16	Klv_Mid_Rev	AACGCGCTATTGGCAATGAATC	Klavier
17	Klv_RB_Rev	TGAGAAACATCTCCGAATCACTAGC	Klavier
18	Klv_Lb_F	GCTGTGTGGGTAGTGGAGGGGCT	Klavier
19	Klv_RB_R	AAGAGAGATCCAAGACTTCTAAGG	Klavier
20	LHK1_LB_Rev	TCTTCCGTCGTTGGCTCTGG	LHK1
21	LHK1_LB_Fwd	TCAAAGCAGAAGCAGCTGATATTG	LHK1
22	LHK1_LB_R	TGGGAGTTCTAATTTCATGAGAGAC	LHK1
23	LHK1_Mid_Fwd	CAACAACTGACATGAAGAAACTTAAC	LHK1
24	LHK1_Mid_Rev	AGAAACAGCCTTATTGATGCTAATTG	LHK1
25	LHK1_RB_Fwd	TGTCGAATGGTTCTTCAGTACG	LHK1
26	LHK1_RB_Rev	CACTTGCTGCACATTTGACATC	LHK1
27	LHK1_Exon2_Fwd	TGCACAGGAAACTGTCTCTTAC	LHK1
28	LHK1_Exon9_Fwd	GCGGTCAGATAAACTTCATAAGC	LHK1
29	LHK1_PC_Fwd	GACAAAAGGCAGACTCCTCGGG	LHK1
30	Lys1_Mid_R	TCTTCTTGATGGCAGTTTTCTG	Lys1
31	Lys1_Mid_F	CAGAAGGCAATAATGCGAAAC	Lys1
32	Lys1_Fwd	CACCATTGCACTCTCACATCAAG	Lys1
33	Lys1_Rev	CTACTCCAATCTTCCCGACATTC	Lys1
34	Lys1_Mid_Rev	TAGGCAAGTGTTGACTACACCTG	Lys1
35	Lys1_Mid_Fwd	GAGATATAGGCAGGTGTAGTCAAC	Lys1
36	Lys1LB_SeqF	GGCTCACAATTTGTTTCTCTTACAC	Lys1
37	Lys1LB_SeqR	CTCTGCTGAAAGGGCACATAG	Lys1
38	Lys1RB_SeqF	TGGTTACATGCCACCAGAGTATG	Lys1
39	Lys1RB_SeqR	CCTATCTCCTATGGCAGAGTTAGC	Lys1
40	Lys1_O_F	ATGGAACCAAAATTAACGTTTTCAC	Lys1
41	Lys1_LB_Seq_Rev	ATGAGAACTCGCGTGATTTGTC	Lys1

Table 7: List of Sequencing Primers *(continued)*

SL	**Name**	**Sequence (3' - 5')**	**Target**
42	Lys1_Mid_Seq_Rev	GCAGTATCCAATCAACCGTACC	Lys1
43	Lys1_RB_Seq_Rev	TTGGGAGAAGGAGAGACAACAC	Lys1
44	Lys11_Mid_R	CAACATCAACTGCTATGCTTATCC	Lys11
45	Lys12_Mid_F	CCCTTTACCACACTTTTAGTACC	Lys12
46	Lys13_Mid_F	GCTGAGATCTTTGGTGTTGATAC	Lys13
47	Lys13_Mid_R	CACATGCCTTGTCATTTGGAACC	Lys13
48	Lys14_Mid_F	TCTACTTCACACCGCTGTTGG	Lys14
49	Lys14_Mid_R	CAACGACTCAATCGCATAGCG	Lys14
50	Lys14_C_Rev	GCCTTGTCATTTGGAACCCCCCAT	Lys14
51	Lys15_Mid_F	GAACTCCTTGTCCACCCTTTAC	Lys15
52	Lys16_Seq_F	ATCACTGGAAGGCTTGACAACC	Lys16
53	Lys16_Seq_R	GGTGCCAAATAACCCAAACTCC	Lys16
54	Lys2_LB_Seq_F	AGAAAAGCAGGTTCGCCTTC	Lys2(L2LB)
55	Lys2_LB_Seq_R	GCAACATGCTGAAAACAGTGAC	Lys2(L2LB)
56	L2LB_R_Fwd	GTTCCCTTGCACATTAGGTAAATT	Lys2(L2LB)
57	L2LB_L_Rev	AATGCCAGTATTGTAAGATAAAT	Lys2(L2LB)
58	L2LB_Mid_Seq	ATTTAGCCTCTTCTCCTTCCTTCTTC	Lys2(L2LB)
59	ib_Lys21_Mid_F	CCCACTTCAAACCTCACAACAAA	LYS21
60	ib_Lys21_Mid_R	TGACCAAGAAACCTGCTACGG	LYS21
61	Lys21_Rev	CTACAAGCACAACCCCGAAAC	Lys21
62	Lys20/21_O_F	ATGGAGCTTCACTACCACCTC	Lys21
63	Lys21_O_R	GAATCCAAGAGTATGTTGGTGGGTTT	Lys21
64	Lys20/21_LB_Fwd	AGCACAGCACACCTTACAC	Lys21
65	Lys3_Mid_R	AAGATGTTCACTGAGAGATCCATTG	Lys3
66	Lys3_Mid_F	TGGAAATCCCACCTACATTCTT	Lys3
67	L3RB_SeqF	TTCGATTCAACATGCCACTAATGTC	Lys3
68	L3RB_SeqR	ACATAACCACCCCAAATGCAAATAC	Lys3
69	L3LB_SeqF	TTCAGAAAGTGGCTCTCAAATTGTCG	Lys3
70	L3LB_SeqR	CTTTGTGAAAGTCGCGGTTCATTAAG	Lys3
71	Lys3_O_F	ATGTTTTATGATTTCACAACTATGG	Lys3
72	Lys3_O_R	TCTTCCATCAAATACCCCGCTGA	Lys3
73	L3LB_Seq_Fwd	CTTGGCCTGTGAATGAAGATTTGG	Lys3
74	L3LB_Seq_Rev	TGAGTCAAATGATGTCGCATCTG	Lys3
75	Lys3_Mid_Fwd	TCAGGAGGTTGCTATAAAGAAG	Lys3
76	L3_Mid_Fwd	GGTGTATGAGTATGTGCCCAATG	Lys3
77	Lys3_Ex6_Rev	TCCTAGCACACCAAGAAAGAGG	Lys3
78	Lys3_Ex7_Rev	GAAGATAGCCTGGTGTTCCAAC	Lys3
79	Lys3_Ex9_Rev	CCATCCACGGTAACTTCTAAAGC	Lys3
80	L3RB_Seq_Fwd	CTGACTTTGCTGGCAGTCTG	Lys3
81	Lys4_Mid_F	CTGGAAGATTCATATGCTGCAAACC	Lys4
82	Lys4_Mid_R	CAGTAATTTGGTCACTGCTGGATTC	Lys4

Table 7: List of Sequencing Primers *(continued)*

SL	**Name**	**Sequence (3' - 5')**	**Target**
83	Lys4LB_Seq_R	GGGATTTCAGAGTTCAAAGACATCAG	Lys4
84	Lys4RB_Seq_F	GCTGCTAGGGGCCTTGAATAC	Lys4
85	Lys4RB_Seq_R	TACCTTGGCTCTGAAGGAAGCATC	Lys4
86	Ly4_O_F	ATGTATCTTACACAGAAACCATGTC	Lys4
87	Lys4_RB_Seq_Rev	GTACTAACACCTTTGGTTGTTGTG	Lys4
88	Lys4_Exon5_Rev	CACGAAGGAGACCATAGTAAAC	Lys4
89	Lys4_Ex1_F	CTGTTCTTGTGCTGCTGGGATG	Lys4
90	Lys4_C_Rev	GCTGCATAGCCGATGAATTCTACC	Lys4
91	L6_Ex68_Rev	GCTAAAGTATGAGCTGACCTAAC	Lys6
92	L6_Gebe_Fwd	GTATTACTGTGAATGAAACA	Lys6
93	L6_Exon6_Rev	CTTGCAGAATCCAAAGCAATCTGC	Lys6
94	L6_intron2_Fwd	GTTATGCAATGATTGTTCCCATA	Lys6
95	Lys6_Mid_R	GATTACTAACCTCTCCCCTCAG	Lys6
96	Lys6LB_Lseq_R	TCACCGCTTGCTCCTAAACCC	Lys6
97	Lys6LB_Rseq_F	AGGAGGTTTGACTTTCTTCTCAG	Lys6
98	Lys6LB_Rseq_R	AGCATCCCCTTTCTACCTTTTAG	Lys6
99	Lys6RB_LSeq_F	AGAGCCATTCCATCTTGTACTTTGC	Lys6
100	Lys6RB_Mseq_F	GTGATGCCTATGAAATCTCAAGTCC	Lys6
101	Lys6RB_Rseq_F	AGCTACAAGCAACCCTTGATATGAC	Lys6
102	L6_Exon1_Fwd	AGTGTTGAGCCAAAGAAATGG	Lys6
103	L6_Exon3_Rev	TCCTCCTTCTGTATCCTCTTCTTC	Lys6
104	L6_Exon4_Fwd	CCCCGTAGTACTGTGAATGAAAC	Lys6
105	L6RB_Lseq_Rev	AATCTCCAGCAGCATAGTGTG	Lys6
106	L6_Exon6_Fwd	TGCGTGGTTCAGGTTAGGTCAG	Lys6
107	L6_Exon8_Rev	CAACAAGACGACCAGTGGGAAG	Lys6
108	L6_Exon10_Rev	ACTGAATCAACAGGGTAGTTATCG	Lys6
109	L6_Intron5_Rev	CAGGTTAAGATGATGAACACGTG	Lys6
110	L6_Ex1_Fwd_Rep	CAGTTAATTGTTCTTGTGGGGATAGTGG	Lys6
111	L6_P145_Rep	GCTTTTGATTTTAAAGTTTATTCC	Lys6
112	L6_Intron2_Rev	CAGCAGCTATTCCAGCAATAG	Lys6
113	L6_Ex1_Fwd	GTTAATTGTTCTTGTGGGG	Lys6
114	L6_Exon1_Fwd	CGTTGCTGCAGAGGTACAATC	Lys6
115	NFR1_LB_Lseq	AGATGAGTAGGCAATTGAAGTTACC	NFR1(N1LB)
116	NFR1_LB_Rseq	CATGACAAGGCATGTAACTTACCC	NFR1(N1LB)
117	NFR1_LB_Fwd	GATCAAAACCTGGTAGAGAGTC	NFR1(N1LB)
118	NFR1_LB_Rev	CTTCGCTAGTTCCTGATATGAG	NFR1(N1LB)
119	N1LB_Seq_Rev	GCCATAGAAATATCTGTTGGCAATTTA	NFR1(N1LB)
120	NFR1_RB_Lseq	CAACCTTTCCACGCAAGTTCTTATC	NFR1(N1RB)
121	NFR1_RB_Rseq	ATGGTCTAGCCGAGTACAAATAGC	NFR1(N1RB)
122	N1RB+SeqF	TTTTGTGCCACTTTAGGTTATCCC	NFR1(N1RB)
123	N1RB+SeqR	CTTCTTCAAACTGCAATGCAAAGC	NFR1(N1RB)

Table 7: List of Sequencing Primers *(continued)*

SL	**Name**	**Sequence (3' - 5')**	**Target**
124	NFR1_O_R	TCTCACAGACAGTAAATTTATGAGAGTT	NFR1(N1RB)
125	N1RB_SeqF	GCGCTTGATTGGATACTGCG	NFR1(N1RB)
126	N1RB_SeqR	TCGGCTAGACCATGGCAATG	NFR1(N1RB)
127	N1RB_EE_Rev	CCAACTTCAATAAGCTTGGTCAAGCC	NFR1(N1RB)
128	N1RB_EE_Fwd	TATGTTCAGGTTGCAGATTTTGG	NFR1(N1RB)
129	N1RB_Aust_Fwd	GCCACTTTAGGTTTAATCCCTTTATC	NFR1(N1RB)
130	N1RB_DG_Rev	CTGTTCTCAAAACCACAATGGAC	NFR1(N1RB)
131	N1RB_End_R	GGATTATCTCTTGTACAAGCTCTCC	NFR1(N1RB)
132	NFR5_Mid_F	TACTTATGTGTGGAAGCCCAATG	NFR5
133	NFR5_Mid_R	CAACATCCACTGCTATGCTTATCC	NFR5
134	NFR5_Mid_Rev	GTCGAGTCGAGAAGGATATTAC	NFR5
135	SymLB_R	ATGCAGTCAGTAGCATAGATTAGAC	SymRK
136	SymRB_F	TGCATTCTAACCCTGTACATTTCAC	SymRK
137	SymLB_Seq_F	TGAAGGTCCTAGCATCATTTATCTG	SymRK
138	SymLB_Seq_R	GTTCAATGTGGGGAAGTTAGTTCAG	SymRK
139	SymRB_Seq_F	ATATTAGACTGGCCAACTAGACTCTC	SymRK
140	SymLB_Seq_Fwd	GACCGTGGATTGAAGAGACAAAC	SymRK
141	SymLB_Seq_Rev	TCTGAATCACTCCCACTAAAGAGG	SymRK
142	SymRB_Fwd	GGGAAAGAATCCAGATTTGGACAAG	SymRK
143	SymLB_Rev	TTTTGTCTGTAGCGGCATAC	SymRK
144	SymRB_Mid_Fwd	CCAAGAAGTGGCGGTGAAAG	SymRK
145	SymLB_Lseq_F	TCTGCCAACAATTAAGAATGGAGTG	SymRK
146	SymLB_Rseq_R	TGGCACCAATAACAATTACTTGTCC	SymRK
147	SymRB_Rseq_F	AAGGATTTGCTGGAAAGAAGTACC	SymRK
148	SymRB_Mid_Rev	AGTAACCAAGAAGTGGCACCAG	SymRK
149	SymRB_Seq_Rev	ATGCTTGGTCTGTATGTTGAGAAAGG	SymRK
150	SymRB_Mid_F	CTACTTTCAGCAATACAGCATGAGAAC	SymRK
151	SymLB_Lseq_R	TGCTTTCCAGATTCGGTCACTTTG	SymRK
152	SymLB_Lseq_Rev	CCATTTGAACCATCACATGCTATTCC	SymRK
153	SM_Exon5_Fwd	GGATCTTTCCTCAAGTAATCTCAAGGG	SymRK
154	SM_Exon8_Fwd	GGCTGCAACGAACACATGAGTCC	SymRK
155	SM_Exon10_Fwd	GCCAAGCAAAGATGATTTCTTCAT	SymRK
156	SM_Exon10_Rev	GGTGAATGCTTGAATTGATACGGAC	SymRK
157	Lys20/21_Mid_rev	TTGCTCTTTGCAGCTCCTC	Lys21

Note:

L2LB = Lys2 Left-Border

N1LB = NFR1 Left-Border

N1RB = NFR1 Right-Border

Lys6 = The first 34 codon triplets of Lys6 are missing due to the genome version used for primer designing

Table 8: Raw data (developed nodules only) of Inoculation Experiment 1

L. cornicultatus	*L. angustissimus*	*L. arabicus*	*L. castellanus*	*L. conjugatus*	*L. glinoides*	*L. mearnsii*	*L. parviflorus*	*L. gebelia*
Nodule	**Nodule**	**Nodule**	**Nodule**	**Nodule**	**Nodule**	**Nodule**	**Nodule**	**Nodule**
6	3	1	0	0	2	2	1	4
6	2	1	0	0	4	4	0	6
1	1	3	0	0	6	2	3	5
4	5	1	0	0	2	4	4	4
2	4	0	0	2	3	3	0	7
5	2	0	0	0		4	3	1
5	3	0	0	0		2	1	4
3	5					3	2	8
4	3					7	1	11
5							1	5
4							0	9

Table 9: t-test results of Inoculation Experiment 1 (developed nodules)

Phenotype	***Species 1***	***Species 2***	**p**	**p-adj**	**p-format**	**p-signif**
Developed Nodules	*L. angustissimus*	*L. arabicus*	2.36e-03	0.05400	0.0023600	**
Developed Nodules	*L. angustissimus*	*L. castellanus*	1.32e-04	0.00400	0.0001300	***
Developed Nodules	*L. angustissimus*	*L. conjugatus*	1.58e-04	0.00460	0.0001600	***
Developed Nodules	*L. angustissimus*	*L. corniculatus*	1.54e-01	1.00000	0.1539800	ns
Developed Nodules	*L. angustissimus*	*L. gebelia*	1.25e-02	0.23000	0.0125100	*
Developed Nodules	*L. angustissimus*	*L. glinoides*	7.51e-01	1.00000	0.7510900	ns
Developed Nodules	*L. angustissimus*	*L. mearnsii*	6.40e-01	1.00000	0.6397300	ns
Developed Nodules	*L. angustissimus*	*L. parviflorus*	1.51e-02	0.26000	0.0151100	*
Developed Nodules	*L. arabicus*	*L. castellanus*	7.81e-02	0.86000	0.0781400	ns
Developed Nodules	*L. arabicus*	*L. conjugatus*	2.73e-01	1.00000	0.2731200	ns
Developed Nodules	*L. arabicus*	*L. corniculatus*	9.36e-05	0.00300	0.0000940	****
Developed Nodules	*L. arabicus*	*L. gebelia*	1.10e-04	0.00340	0.0001100	***
Developed Nodules	*L. arabicus*	*L. glinoides*	2.29e-02	0.37000	0.0228600	*
Developed Nodules	*L. arabicus*	*L. mearnsii*	1.70e-03	0.04100	0.0017000	**
Developed Nodules	*L. arabicus*	*L. parviflorus*	3.17e-01	1.00000	0.3171500	ns
Developed Nodules	*L. castellanus*	*L. conjugatus*	3.56e-01	1.00000	0.3559200	ns
Developed Nodules	*L. castellanus*	*L. corniculatus*	6.24e-06	0.00022	0.0000062	****
Developed Nodules	*L. castellanus*	*L. gebelia*	4.07e-05	0.00140	0.0000410	****
Developed Nodules	*L. castellanus*	*L. glinoides*	1.05e-02	0.21000	0.0104700	*
Developed Nodules	*L. castellanus*	*L. mearnsii*	1.88e-04	0.00530	0.0001900	***
Developed Nodules	*L. castellanus*	*L. parviflorus*	5.49e-03	0.12000	0.0054900	**
Developed Nodules	*L. conjugatus*	*L. corniculatus*	5.06e-06	0.00018	0.0000051	****
Developed Nodules	*L. conjugatus*	*L. gebelia*	4.08e-05	0.00140	0.0000410	****
Developed Nodules	*L. conjugatus*	*L. glinoides*	1.08e-02	0.21000	0.0107900	*
Developed Nodules	*L. conjugatus*	*L. mearnsii*	2.07e-04	0.00560	0.0002100	***
Developed Nodules	*L. conjugatus*	*L. parviflorus*	3.35e-02	0.47000	0.0334500	*
Developed Nodules	*L. corniculatus*	*L. gebelia*	9.28e-02	0.93000	0.0927900	ns
Developed Nodules	*L. corniculatus*	*L. glinoides*	4.60e-01	1.00000	0.4601000	ns
Developed Nodules	*L. corniculatus*	*L. mearnsii*	3.77e-01	1.00000	0.3766400	ns
Developed Nodules	*L. corniculatus*	*L. parviflorus*	4.73e-04	0.01200	0.0004700	***
Developed Nodules	*L. gebelia*	*L. glinoides*	5.18e-02	0.67000	0.0518300	ns
Developed Nodules	*L. gebelia*	*L. mearnsii*	2.93e-02	0.44000	0.0292500	*
Developed Nodules	*L. gebelia*	*L. parviflorus*	3.31e-04	0.00860	0.0003300	***
Developed Nodules	*L. glinoides*	*L. mearnsii*	9.63e-01	1.00000	0.9625300	ns
Developed Nodules	*L. glinoides*	*L. parviflorus*	5.94e-02	0.71000	0.0594000	ns
Developed Nodules	*L. mearnsii*	*L. parviflorus*	9.19e-03	0.19000	0.0091900	**

Table 10: Raw Data of Inoculation Experiment 2

L. corniculatus		*L. uliginosus*		*L. pedunculatus*		*L. castellanus*		*L. conjugatus*		*L. gebelia*	
Nodule	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**
4	0	2	1	0	1	0	2	0	0	4	0
4	0	0	1	0	2	0	0	0	0	5	0
8	0	2	0	3	1	0	2	0	0	4	0
4	0	1	1	3	0	0	1	0	0	8	0
4	0	3	0	0	0	0	2	0	0	4	0
6	0	0	1	2	0	0	1	0	0	4	0
4	0	0	1	0	2	0	1	0	0	10	0
6	0	0	1	2	0	0	1	0	0	10	0
5	0	2	0	0	0	0	1	0	0	3	0
7	0	2	1	2	0	0	4	0	0	7	0
6	0	1	2	2	0	0	1			9	0
5	0	2	2	1	1	0	2			2	0
7	0	0	0	1	0	0	0			8	0
10	0	0	1	0	1	0	0			12	0
7	0	0	4	0	0					14	0
8	0	3	1	1	0					10	0
3	0	1	1	0	1					7	0
5	0	1	1	0	0						
4	0	0	0	0	1						
5	0	0	2	2	0						
4	0	1	1								
6	0	0	0								
10	0	0	2								
6	0	0	0								
10	0	1	1								
6	0										

Table 11: t-test results of Inoculation Experiment 2

Phenotype	***Species 1***	***Species 2***	**p**	**p-adj**	**p-format**	**p-signif**
Developed Nodules	*L. castellanus*	*L. corniculatus*	4.43e-14	0.0e+00	4.4e-14	****
Developed Nodules	*L. castellanus*	*L. gebelia*	2.26e-07	5.0e-06	2.3e-07	****
Developed Nodules	*L. castellanus*	*L. pedunculatus*	1.04e-03	1.0e-02	0.00104	**
Developed Nodules	*L. castellanus*	*L. uliginosus*	2.21e-04	3.3e-03	0.00022	***
Developed Nodules	*L. conjugatus*	*L. corniculatus*	4.43e-14	0.0e+00	4.4e-14	****
Developed Nodules	*L. conjugatus*	*L. gebelia*	2.26e-07	5.0e-06	2.3e-07	****
Developed Nodules	*L. conjugatus*	*L. pedunculatus*	1.04e-03	1.0e-02	0.00104	**
Developed Nodules	*L. conjugatus*	*L. uliginosus*	2.21e-04	3.3e-03	0.00022	***
Developed Nodules	*L. corniculatus*	*L. gebelia*	2.06e-01	6.2e-01	0.20626	ns
Developed Nodules	*L. corniculatus*	*L. pedunculatus*	2.08e-13	0.0e+00	2.1e-13	****
Developed Nodules	*L. corniculatus*	*L. uliginosus*	9.25e-14	0.0e+00	9.3e-14	****
Developed Nodules	*L. gebelia*	*L. pedunculatus*	9.64e-07	1.8e-05	9.6e-07	****
Developed Nodules	*L. gebelia*	*L. uliginosus*	9.15e-07	1.8e-05	9.2e-07	****
Developed Nodules	*L. pedunculatus*	*L. uliginosus*	8.27e-01	8.3e-01	0.82720	ns
Developing Primordia	*L. castellanus*	*L. conjugatus*	5.97e-04	7.8e-03	0.00060	***
Developing Primordia	*L. castellanus*	*L. corniculatus*	5.97e-04	7.8e-03	0.00060	***
Developing Primordia	*L. castellanus*	*L. gebelia*	5.97e-04	7.8e-03	0.00060	***
Developing Primordia	*L. castellanus*	*L. pedunculatus*	2.49e-02	1.2e-01	0.02487	*
Developing Primordia	*L. castellanus*	*L. uliginosus*	4.08e-01	8.2e-01	0.40787	ns
Developing Primordia	*L. conjugatus*	*L. pedunculatus*	4.22e-03	3.4e-02	0.00422	**
Developing Primordia	*L. conjugatus*	*L. uliginosus*	1.25e-05	2.2e-04	1.2e-05	****
Developing Primordia	*L. corniculatus*	*L. pedunculatus*	4.22e-03	3.4e-02	0.00422	**
Developing Primordia	*L. corniculatus*	*L. uliginosus*	1.25e-05	2.2e-04	1.2e-05	****
Developing Primordia	*L. gebelia*	*L. pedunculatus*	4.22e-03	3.4e-02	0.00422	**
Developing Primordia	*L. gebelia*	*L. uliginosus*	1.25e-05	2.2e-04	1.2e-05	****
Developing Primordia	*L. pedunculatus*	*L. uliginosus*	4.22e-02	1.7e-01	0.04221	*

Table 12: Raw Data of Inoculation Experiment 3

L. krylovii		*L. tenuis*		*L. drepanocarpus*		*L. arenarius*		*L. mearnsii*		*L. weilleri*		*L. edulis*		*L. alpinus*		*L. preslii*		*L. peregrinus*		*L. ornithopodioides*		*L. cytisoides*		*L. subbiflorus*		*L. conimbricensis*	
Nodule	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**
3	0	5	0	10	0	5	0	3	0	9	0	0	2	4	0	5	0	0	3	2	0	0	2	4	0	0	0
5	0	3	0	10	0	2	0	3	0	8	0	0	1	4	0	5	0	0	6	1	3	1	1	0	6	0	0
4	0	1	0	2	0	0	1	2	0	6	0	0	1	0	0	4	0	0	3	0	0	5	0	0	5	0	0
17	0	1	0	10	0	1	5	1	0	0	2	0	0	2	0	4	0	0	4	0	2	1	1	0	4	0	0
0	0	2	0	5	0	6	0	4	0	2	0	0	0	5	0	4	0	0	3	0	0	3	0	0	2	0	1
11	0	2	0	5	0	3	0	4	0	4	2	0	0	3	0	7	0	0	2	2	0	0	1	0	1	0	0
7	0	0	4	12	0	2	0	12	0	10	0	0	2	2	0	6	0	0	2	2	0	2	0	0	2	0	2
14	0	0	2	5	0	6	0	10	0	7	0	0	0	5	0	8	0			1	0	1	0	0	2	0	1
10	0	2	2	6	0	17	0	9	0	17	0	0	0	2	0	7	0			0	0	1	0	0	2	0	2
16	0	0	0	5	0	5	0	4	0	6	0	0	2	5	0	16	0			0	0	2	0	1	2	0	1
9	0	0	0			2	0	3	0	17	0	0	0	4	0	9	0			7	0	1	0	0	1		
19	0	2	0			5	0	2	0	0	2	0	0	10	0	5	0			1	4	1	0				
5	0	0	0			6	0	4	0	2	0	1	0							0	5	2	0				
1	0	2	0			2	0	7	0	0	1	0	3							0	4	1	0				
0	0	2	0			13	0	7	0	2	0	0	2							0	2	0	0				
2	0	0	0					2	0			0	2							0	3	1	0				
								7	0			0	0							0	1	1	0				
												0	2							0	2	2	0				
												0	2							0	1						
																				0	5						

Table 13: t-test results of Inoculation Experiment 3

Phenotype	***Species 1***	***Species 2***	**p**	**p-adj**	**p-format**	**p-signif**
Developed Nodules	*L. alpinus*	*L. arenarius*	4.07e-01	1.0000	0.40668	ns
Developed Nodules	*L. alpinus*	*L. conimbricensis*	2.32e-04	0.0340	0.00023	***
Developed Nodules	*L. alpinus*	*L. cytisoides*	6.50e-03	0.5900	0.00650	**
Developed Nodules	*L. alpinus*	*L. drepanocarpus*	2.14e-02	1.0000	0.02143	*
Developed Nodules	*L. alpinus*	*L. edulis*	2.57e-04	0.0370	0.00026	***
Developed Nodules	*L. alpinus*	*L. krylovii*	3.62e-02	1.0000	0.03619	*
Developed Nodules	*L. alpinus*	*L. mearnsii*	3.01e-01	1.0000	0.30125	ns
Developed Nodules	*L. alpinus*	*L. ornithopodioides*	1.55e-03	0.1600	0.00155	**
Developed Nodules	*L. alpinus*	*L. peregrinus*	2.32e-04	0.0340	0.00023	***
Developed Nodules	*L. alpinus*	*L. preslii*	2.92e-02	1.0000	0.02920	*
Developed Nodules	*L. alpinus*	*L. subbiflorus*	6.53e-04	0.0870	0.00065	***
Developed Nodules	*L. alpinus*	*L. tenuis*	7.05e-03	0.6200	0.00705	**
Developed Nodules	*L. alpinus*	*L. weilleri*	1.91e-01	1.0000	0.19142	ns
Developed Nodules	*L. arenarius*	*L. conimbricensis*	8.24e-04	0.1000	0.00082	***
Developed Nodules	*L. arenarius*	*L. cytisoides*	9.06e-03	0.7900	0.00906	**
Developed Nodules	*L. arenarius*	*L. drepanocarpus*	2.13e-01	1.0000	0.21279	ns
Developed Nodules	*L. arenarius*	*L. edulis*	8.98e-04	0.1100	0.00090	***
Developed Nodules	*L. arenarius*	*L. krylovii*	1.81e-01	1.0000	0.18125	ns
Developed Nodules	*L. arenarius*	*L. mearnsii*	9.67e-01	1.0000	0.96701	ns
Developed Nodules	*L. arenarius*	*L. ornithopodioides*	3.47e-03	0.3300	0.00347	**
Developed Nodules	*L. arenarius*	*L. peregrinus*	8.24e-04	0.1000	0.00082	***
Developed Nodules	*L. arenarius*	*L. preslii*	2.86e-01	1.0000	0.28585	ns
Developed Nodules	*L. arenarius*	*L. subbiflorus*	1.91e-03	0.1900	0.00191	**
Developed Nodules	*L. arenarius*	*L. tenuis*	9.27e-03	0.8000	0.00927	**
Developed Nodules	*L. arenarius*	*L. weilleri*	5.95e-01	1.0000	0.59460	ns
Developed Nodules	*L. conimbricensis*	*L. cytisoides*	1.27e-04	0.0200	0.00013	***
Developed Nodules	*L. conimbricensis*	*L. drepanocarpus*	7.48e-05	0.0120	7.5e-05	****
Developed Nodules	*L. conimbricensis*	*L. edulis*	3.31e-01	1.0000	0.33056	ns
Developed Nodules	*L. conimbricensis*	*L. krylovii*	1.87e-04	0.0280	0.00019	***
Developed Nodules	*L. conimbricensis*	*L. mearnsii*	8.42e-06	0.0014	8.4e-06	****
Developed Nodules	*L. conimbricensis*	*L. ornithopodioides*	4.21e-02	1.0000	0.04209	*
Developed Nodules	*L. conimbricensis*	*L. preslii*	2.73e-05	0.0046	2.7e-05	****
Developed Nodules	*L. conimbricensis*	*L. subbiflorus*	2.42e-01	1.0000	0.24248	ns
Developed Nodules	*L. conimbricensis*	*L. tenuis*	1.41e-03	0.1500	0.00141	**
Developed Nodules	*L. conimbricensis*	*L. weilleri*	9.20e-04	0.1100	0.00092	***
Developed Nodules	*L. cytisoides*	*L. drepanocarpus*	3.09e-04	0.0450	0.00031	***
Developed Nodules	*L. cytisoides*	*L. edulis*	1.89e-04	0.0280	0.00019	***
Developed Nodules	*L. cytisoides*	*L. krylovii*	1.12e-03	0.1300	0.00112	**
Developed Nodules	*L. cytisoides*	*L. mearnsii*	3.13e-04	0.0450	0.00031	***
Developed Nodules	*L. cytisoides*	*L. ornithopodioides*	2.12e-01	1.0000	0.21159	ns
Developed Nodules	*L. cytisoides*	*L. peregrinus*	1.27e-04	0.0200	0.00013	***
Developed Nodules	*L. cytisoides*	*L. preslii*	1.72e-04	0.0260	0.00017	***
Developed Nodules	*L. cytisoides*	*L. subbiflorus*	5.59e-02	1.0000	0.05593	ns

Table 13: t-test results of Inoculation Experiment 3 *(continued)*

Phenotype	***Species 1***	***Species 2***	**p**	**p-adj**	**p-format**	**p-signif**
Developed Nodules	*L. cytisoides*	*L. tenuis*	9.76e-01	1.0000	0.97563	ns
Developed Nodules	*L. cytisoides*	*L. weilleri*	6.51e-03	0.5900	0.00651	**
Developed Nodules	*L. drepanocarpus*	*L. edulis*	7.80e-05	0.0120	7.8e-05	****
Developed Nodules	*L. drepanocarpus*	*L. krylovii*	7.16e-01	1.0000	0.71625	ns
Developed Nodules	*L. drepanocarpus*	*L. mearnsii*	1.24e-01	1.0000	0.12424	ns
Developed Nodules	*L. drepanocarpus*	*L. ornithopodioides*	1.20e-04	0.0190	0.00012	***
Developed Nodules	*L. drepanocarpus*	*L. peregrinus*	7.48e-05	0.0120	7.5e-05	****
Developed Nodules	*L. drepanocarpus*	*L. preslii*	8.16e-01	1.0000	0.81561	ns
Developed Nodules	*L. drepanocarpus*	*L. subbiflorus*	7.68e-05	0.0120	7.7e-05	****
Developed Nodules	*L. drepanocarpus*	*L. tenuis*	2.79e-04	0.0400	0.00028	***
Developed Nodules	*L. drepanocarpus*	*L. weilleri*	5.76e-01	1.0000	0.57573	ns
Developed Nodules	*L. edulis*	*L. krylovii*	2.00e-04	0.0300	0.00020	***
Developed Nodules	*L. edulis*	*L. mearnsii*	9.39e-06	0.0016	9.4e-06	****
Developed Nodules	*L. edulis*	*L. ornithopodioides*	5.76e-02	1.0000	0.05765	ns
Developed Nodules	*L. edulis*	*L. peregrinus*	3.31e-01	1.0000	0.33056	ns
Developed Nodules	*L. edulis*	*L. preslii*	2.88e-05	0.0048	2.9e-05	****
Developed Nodules	*L. edulis*	*L. subbiflorus*	3.01e-01	1.0000	0.30145	ns
Developed Nodules	*L. edulis*	*L. tenuis*	1.94e-03	0.1900	0.00194	**
Developed Nodules	*L. edulis*	*L. weilleri*	9.88e-04	0.1100	0.00099	***
Developed Nodules	*L. krylovii*	*L. mearnsii*	1.30e-01	1.0000	0.12951	ns
Developed Nodules	*L. krylovii*	*L. ornithopodioides*	5.21e-04	0.0730	0.00052	***
Developed Nodules	*L. krylovii*	*L. peregrinus*	1.87e-04	0.0280	0.00019	***
Developed Nodules	*L. krylovii*	*L. preslii*	5.85e-01	1.0000	0.58453	ns
Developed Nodules	*L. krylovii*	*L. subbiflorus*	3.32e-04	0.0470	0.00033	***
Developed Nodules	*L. krylovii*	*L. tenuis*	1.12e-03	0.1300	0.00112	**
Developed Nodules	*L. krylovii*	*L. weilleri*	4.33e-01	1.0000	0.43305	ns
Developed Nodules	*L. mearnsii*	*L. ornithopodioides*	6.54e-05	0.0110	6.5e-05	****
Developed Nodules	*L. mearnsii*	*L. peregrinus*	8.42e-06	0.0014	8.4e-06	****
Developed Nodules	*L. mearnsii*	*L. preslii*	1.77e-01	1.0000	0.17723	ns
Developed Nodules	*L. mearnsii*	*L. subbiflorus*	2.67e-05	0.0045	2.7e-05	****
Developed Nodules	*L. mearnsii*	*L. tenuis*	3.46e-04	0.0490	0.00035	***
Developed Nodules	*L. mearnsii*	*L. weilleri*	5.22e-01	1.0000	0.52217	ns
Developed Nodules	*L. ornithopodioides*	*L. peregrinus*	4.21e-02	1.0000	0.04209	*
Developed Nodules	*L. ornithopodioides*	*L. preslii*	5.71e-05	0.0093	5.7e-05	****
Developed Nodules	*L. ornithopodioides*	*L. subbiflorus*	5.11e-01	1.0000	0.51089	ns
Developed Nodules	*L. ornithopodioides*	*L. tenuis*	2.66e-01	1.0000	0.26622	ns
Developed Nodules	*L. ornithopodioides*	*L. weilleri*	2.92e-03	0.2900	0.00292	**
Developed Nodules	*L. peregrinus*	*L. preslii*	2.73e-05	0.0046	2.7e-05	****
Developed Nodules	*L. peregrinus*	*L. subbiflorus*	2.42e-01	1.0000	0.24248	ns
Developed Nodules	*L. peregrinus*	*L. tenuis*	1.41e-03	0.1500	0.00141	**
Developed Nodules	*L. peregrinus*	*L. weilleri*	9.20e-04	0.1100	0.00092	***
Developed Nodules	*L. preslii*	*L. subbiflorus*	3.33e-05	0.0055	3.3e-05	****
Developed Nodules	*L. preslii*	*L. tenuis*	1.59e-04	0.0250	0.00016	***

Table 13: t-test results of Inoculation Experiment 3 *(continued)*

Phenotype	***Species 1***	***Species 2***	**p**	**p-adj**	**p-format**	**p-signif**
Developed Nodules	*L. preslii*	*L. weilleri*	7.04e-01	1.0000	0.70385	ns
Developed Nodules	*L. subbiflorus*	*L. tenuis*	8.26e-02	1.0000	0.08262	ns
Developed Nodules	*L. subbiflorus*	*L. weilleri*	1.80e-03	0.1800	0.00180	**
Developed Nodules	*L. tenuis*	*L. weilleri*	6.56e-03	0.5900	0.00656	**
Developing Primordia	*L. alpinus*	*L. arenarius*	2.53e-01	1.0000	0.25262	ns
Developing Primordia	*L. alpinus*	*L. conimbricensis*	2.48e-02	1.0000	0.02485	*
Developing Primordia	*L. alpinus*	*L. cytisoides*	5.60e-02	1.0000	0.05597	ns
Developing Primordia	*L. alpinus*	*L. edulis*	6.22e-04	0.0860	0.00062	***
Developing Primordia	*L. alpinus*	*L. ornithopodioides*	8.87e-04	0.1100	0.00089	***
Developing Primordia	*L. alpinus*	*L. peregrinus*	7.46e-04	0.0990	0.00075	***
Developing Primordia	*L. alpinus*	*L. subbiflorus*	1.14e-03	0.1300	0.00114	**
Developing Primordia	*L. alpinus*	*L. tenuis*	1.04e-01	1.0000	0.10377	ns
Developing Primordia	*L. alpinus*	*L. weilleri*	4.79e-02	1.0000	0.04792	*
Developing Primordia	*L. arenarius*	*L. conimbricensis*	4.87e-01	1.0000	0.48680	ns
Developing Primordia	*L. arenarius*	*L. cytisoides*	7.39e-01	1.0000	0.73912	ns
Developing Primordia	*L. arenarius*	*L. drepanocarpus*	2.53e-01	1.0000	0.25262	ns
Developing Primordia	*L. arenarius*	*L. edulis*	1.58e-01	1.0000	0.15826	ns
Developing Primordia	*L. arenarius*	*L. krylovii*	2.53e-01	1.0000	0.25262	ns
Developing Primordia	*L. arenarius*	*L. mearnsii*	2.53e-01	1.0000	0.25262	ns
Developing Primordia	*L. arenarius*	*L. ornithopodioides*	2.94e-02	1.0000	0.02938	*
Developing Primordia	*L. arenarius*	*L. peregrinus*	6.74e-04	0.0900	0.00067	***
Developing Primordia	*L. arenarius*	*L. preslii*	2.53e-01	1.0000	0.25262	ns
Developing Primordia	*L. arenarius*	*L. subbiflorus*	5.08e-03	0.4800	0.00508	**
Developing Primordia	*L. arenarius*	*L. tenuis*	8.23e-01	1.0000	0.82281	ns
Developing Primordia	*L. arenarius*	*L. weilleri*	8.69e-01	1.0000	0.86852	ns
Developing Primordia	*L. conimbricensis*	*L. cytisoides*	1.72e-01	1.0000	0.17221	ns
Developing Primordia	*L. conimbricensis*	*L. drepanocarpus*	2.48e-02	1.0000	0.02485	*
Developing Primordia	*L. conimbricensis*	*L. edulis*	4.07e-01	1.0000	0.40729	ns
Developing Primordia	*L. conimbricensis*	*L. krylovii*	2.48e-02	1.0000	0.02485	*
Developing Primordia	*L. conimbricensis*	*L. mearnsii*	2.48e-02	1.0000	0.02485	*
Developing Primordia	*L. conimbricensis*	*L. ornithopodioides*	7.28e-02	1.0000	0.07283	ns
Developing Primordia	*L. conimbricensis*	*L. peregrinus*	1.64e-03	0.1700	0.00164	**
Developing Primordia	*L. conimbricensis*	*L. preslii*	2.48e-02	1.0000	0.02485	*
Developing Primordia	*L. conimbricensis*	*L. subbiflorus*	1.14e-02	0.9700	0.01140	*
Developing Primordia	*L. conimbricensis*	*L. tenuis*	6.12e-01	1.0000	0.61172	ns
Developing Primordia	*L. conimbricensis*	*L. weilleri*	4.98e-01	1.0000	0.49785	ns
Developing Primordia	*L. cytisoides*	*L. drepanocarpus*	5.60e-02	1.0000	0.05597	ns
Developing Primordia	*L. cytisoides*	*L. edulis*	1.45e-02	1.0000	0.01449	*
Developing Primordia	*L. cytisoides*	*L. krylovii*	5.60e-02	1.0000	0.05597	ns
Developing Primordia	*L. cytisoides*	*L. mearnsii*	5.60e-02	1.0000	0.05597	ns
Developing Primordia	*L. cytisoides*	*L. ornithopodioides*	5.20e-03	0.4800	0.00520	**
Developing Primordia	*L. cytisoides*	*L. peregrinus*	9.08e-04	0.1100	0.00091	***
Developing Primordia	*L. cytisoides*	*L. preslii*	5.60e-02	1.0000	0.05597	ns

Table 13: t-test results of Inoculation Experiment 3 *(continued)*

Phenotype	***Species 1***	***Species 2***	**p**	**p-adj**	**p-format**	**p-signif**
Developing Primordia	*L. cytisoides*	*L. subbiflorus*	2.49e-03	0.2500	0.00249	**
Developing Primordia	*L. cytisoides*	*L. tenuis*	4.93e-01	1.0000	0.49334	ns
Developing Primordia	*L. cytisoides*	*L. weilleri*	4.65e-01	1.0000	0.46484	ns
Developing Primordia	*L. drepanocarpus*	*L. edulis*	6.22e-04	0.0860	0.00062	***
Developing Primordia	*L. drepanocarpus*	*L. ornithopodioides*	8.87e-04	0.1100	0.00089	***
Developing Primordia	*L. drepanocarpus*	*L. peregrinus*	7.46e-04	0.0990	0.00075	***
Developing Primordia	*L. drepanocarpus*	*L. subbiflorus*	1.14e-03	0.1300	0.00114	**
Developing Primordia	*L. drepanocarpus*	*L. tenuis*	1.04e-01	1.0000	0.10377	ns
Developing Primordia	*L. drepanocarpus*	*L. weilleri*	4.79e-02	1.0000	0.04792	*
Developing Primordia	*L. edulis*	*L. krylovii*	6.22e-04	0.0860	0.00062	***
Developing Primordia	*L. edulis*	*L. mearnsii*	6.22e-04	0.0860	0.00062	***
Developing Primordia	*L. edulis*	*L. ornithopodioides*	2.14e-01	1.0000	0.21415	ns
Developing Primordia	*L. edulis*	*L. peregrinus*	3.44e-03	0.3300	0.00344	**
Developing Primordia	*L. edulis*	*L. preslii*	6.22e-04	0.0860	0.00062	***
Developing Primordia	*L. edulis*	*L. subbiflorus*	2.87e-02	1.0000	0.02869	*
Developing Primordia	*L. edulis*	*L. tenuis*	1.94e-01	1.0000	0.19402	ns
Developing Primordia	*L. edulis*	*L. weilleri*	1.09e-01	1.0000	0.10929	ns
Developing Primordia	*L. krylovii*	*L. ornithopodioides*	8.87e-04	0.1100	0.00089	***
Developing Primordia	*L. krylovii*	*L. peregrinus*	7.46e-04	0.0990	0.00075	***
Developing Primordia	*L. krylovii*	*L. subbiflorus*	1.14e-03	0.1300	0.00114	**
Developing Primordia	*L. krylovii*	*L. tenuis*	1.04e-01	1.0000	0.10377	ns
Developing Primordia	*L. krylovii*	*L. weilleri*	4.79e-02	1.0000	0.04792	*
Developing Primordia	*L. mearnsii*	*L. ornithopodioides*	8.87e-04	0.1100	0.00089	***
Developing Primordia	*L. mearnsii*	*L. peregrinus*	7.46e-04	0.0990	0.00075	***
Developing Primordia	*L. mearnsii*	*L. subbiflorus*	1.14e-03	0.1300	0.00114	**
Developing Primordia	*L. mearnsii*	*L. tenuis*	1.04e-01	1.0000	0.10377	ns
Developing Primordia	*L. mearnsii*	*L. weilleri*	4.79e-02	1.0000	0.04792	*
Developing Primordia	*L. ornithopodioides*	*L. peregrinus*	2.33e-02	1.0000	0.02328	*
Developing Primordia	*L. ornithopodioides*	*L. preslii*	8.87e-04	0.1100	0.00089	***
Developing Primordia	*L. ornithopodioides*	*L. subbiflorus*	2.23e-01	1.0000	0.22297	ns
Developing Primordia	*L. ornithopodioides*	*L. tenuis*	3.45e-02	1.0000	0.03455	*
Developing Primordia	*L. ornithopodioides*	*L. weilleri*	2.01e-02	1.0000	0.02012	*
Developing Primordia	*L. peregrinus*	*L. preslii*	7.46e-04	0.0990	0.00075	***
Developing Primordia	*L. peregrinus*	*L. subbiflorus*	2.88e-01	1.0000	0.28779	ns
Developing Primordia	*L. peregrinus*	*L. tenuis*	9.09e-04	0.1100	0.00091	***
Developing Primordia	*L. peregrinus*	*L. weilleri*	1.01e-03	0.1200	0.00101	**
Developing Primordia	*L. preslii*	*L. subbiflorus*	1.14e-03	0.1300	0.00114	**
Developing Primordia	*L. preslii*	*L. tenuis*	1.04e-01	1.0000	0.10377	ns
Developing Primordia	*L. preslii*	*L. weilleri*	4.79e-02	1.0000	0.04792	*
Developing Primordia	*L. subbiflorus*	*L. tenuis*	6.14e-03	0.5600	0.00614	**
Developing Primordia	*L. subbiflorus*	*L. weilleri*	4.77e-03	0.4500	0.00477	**
Developing Primordia	*L. tenuis*	*L. weilleri*	9.27e-01	1.0000	0.92693	ns

Table 14: Raw data of Inoculation Experiment 4

Mesorhizobium loti										*Bradyrhizobium* **Spp.**									
L. corniculatus		*L. peduculatus*		*L. conimbricensis*		*L. halophilus*		*L. castellanus*		*L. corniculatus*		*L. peduculatus*		*L. conimbricensis*		*L. halophilus*		*L. castellanus*	
Nodule	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**	**Nodule**	**Primordia**
4	0	0	4	0	0	0	0	3	1	0	0	1	0	1	0	1	0	4	1
4	0	0	0	0	0	0	0	0	7	0	0	1	0	0	0	4	0	5	0
1	0	0	2	0	0	0	0	3	1	5	0	0	1	0	0	1	0	5	0
2	0	0	5	0	0	0	0	1	1	0	0	1	0	0	0	3	0	3	0
3	0	0	4	0	0	0	0	1	1	0	0	0	0	0	0	2	0	5	0
10	3	1	5	0	0	0	0	1	2	0	0	1	0	7	1	0	0	7	0
4	1	0	9	0	0	0	0	1	2	2	0	1	0	11	2	0	0	3	2
11	1	0	2	0	0	0	0	1	4	5	1	1	0	2	0	0	0	8	4
6	1	0	8	0	1	0	0	2	2	2	0	1	0	3	0	2	0	0	0
2	5	0	6	0	0	0	0	1	5	0	0	0	3	6	2	0	0	4	1

Table 15: t-test results of Inoculation Experiment 4

Phenotype	***Species 1***	***Species 2***	**p**	**p-adj**	**p-format**	**p-signif**
Mesorhizobium Nodules	*L. castellanus*	*L. conimbricensis*	0.00132	0.048	0.00132	**
Mesorhizobium Nodules	*L. castellanus*	*L. corniculatus*	0.01321	0.330	0.01321	*
Mesorhizobium Nodules	*L. castellanus*	*L. halophilus*	0.00132	0.048	0.00132	**
Mesorhizobium Nodules	*L. castellanus*	*L. pedunculatus*	0.00197	0.061	0.00197	**
Mesorhizobium Nodules	*L. conimbricensis*	*L. corniculatus*	0.00169	0.057	0.00169	**
Mesorhizobium Nodules	*L. conimbricensis*	*L. pedunculatus*	0.34344	1.000	0.34344	ns
Mesorhizobium Nodules	*L. corniculatus*	*L. halophilus*	0.00169	0.057	0.00169	**
Mesorhizobium Nodules	*L. corniculatus*	*L. pedunculatus*	0.00191	0.061	0.00191	**
Mesorhizobium Nodules	*L. halophilus*	*L. pedunculatus*	0.34344	1.000	0.34344	ns
Bradyrhizobium Nodules	*L. castellanus*	*L. conimbricensis*	0.33103	1.000	0.33103	ns
Bradyrhizobium Nodules	*L. castellanus*	*L. corniculatus*	0.00584	0.150	0.00584	**
Bradyrhizobium Nodules	*L. castellanus*	*L. halophilus*	0.00199	0.061	0.00199	**
Bradyrhizobium Nodules	*L. castellanus*	*L. pedunculatus*	0.00045	0.018	0.00045	***
Bradyrhizobium Nodules	*L. conimbricensis*	*L. corniculatus*	0.26181	1.000	0.26181	ns
Bradyrhizobium Nodules	*L. conimbricensis*	*L. halophilus*	0.21091	1.000	0.21091	ns
Bradyrhizobium Nodules	*L. conimbricensis*	*L. pedunculatus*	0.08909	1.000	0.08909	ns
Bradyrhizobium Nodules	*L. corniculatus*	*L. halophilus*	0.90114	1.000	0.90114	ns
Bradyrhizobium Nodules	*L. corniculatus*	*L. pedunculatus*	0.32133	1.000	0.32133	ns
Bradyrhizobium Nodules	*L. halophilus*	*L. pedunculatus*	0.23136	1.000	0.23136	ns
Mesorhizobium Primordia	*L. castellanus*	*L. conimbricensis*	0.00398	0.110	0.00398	**
Mesorhizobium Primordia	*L. castellanus*	*L. corniculatus*	0.09129	1.000	0.09129	ns
Mesorhizobium Primordia	*L. castellanus*	*L. halophilus*	0.00320	0.093	0.00320	**
Mesorhizobium Primordia	*L. castellanus*	*L. pedunculatus*	0.09968	1.000	0.09968	ns
Mesorhizobium Primordia	*L. conimbricensis*	*L. corniculatus*	0.09243	1.000	0.09243	ns
Mesorhizobium Primordia	*L. conimbricensis*	*L. halophilus*	0.34344	1.000	0.34344	ns
Mesorhizobium Primordia	*L. conimbricensis*	*L. pedunculatus*	0.00067	0.025	0.00067	***
Mesorhizobium Primordia	*L. corniculatus*	*L. halophilus*	0.06605	1.000	0.06605	ns
Mesorhizobium Primordia	*L. corniculatus*	*L. pedunculatus*	0.00458	0.120	0.00458	**
Mesorhizobium Primordia	*L. halophilus*	*L. pedunculatus*	0.00060	0.023	0.00060	***
Bradyrhizobium Primordia	*L. castellanus*	*L. conimbricensis*	0.55373	1.000	0.55373	ns
Bradyrhizobium Primordia	*L. castellanus*	*L. corniculatus*	0.13302	1.000	0.13302	ns
Bradyrhizobium Primordia	*L. castellanus*	*L. halophilus*	0.08684	1.000	0.08684	ns
Bradyrhizobium Primordia	*L. castellanus*	*L. pedunculatus*	0.44953	1.000	0.44953	ns
Bradyrhizobium Primordia	*L. conimbricensis*	*L. corniculatus*	0.18950	1.000	0.18950	ns
Bradyrhizobium Primordia	*L. conimbricensis*	*L. halophilus*	0.09573	1.000	0.09573	ns
Bradyrhizobium Primordia	*L. conimbricensis*	*L. pedunculatus*	0.80869	1.000	0.80869	ns
Bradyrhizobium Primordia	*L. corniculatus*	*L. halophilus*	0.34344	1.000	0.34344	ns
Bradyrhizobium Primordia	*L. corniculatus*	*L. pedunculatus*	0.37089	1.000	0.37089	ns
Bradyrhizobium Primordia	*L. halophilus*	*L. pedunculatus*	0.22287	1.000	0.22287	ns

Table 16: List of seeds

SL	*Species*	ID	Origin	Source
1	*Lotus ornithopodioides*	LOT 6	Italy	IPK Gatersleben, Germany
2	*Lotus krylovii*	LOT 15	China	IPK Gatersleben, Germany
3	*Lotus weilleri*	LOT 16	Hungary	IPK Gatersleben, Germany
4	*Lotus pedunculatus*	LOT 29	Portugal	IPK Gatersleben, Germany
5	*Lotus preslii*	LOT 43	Unknown	IPK Gatersleben, Germany
6	*Lotus subbiflorus*	LOT 41	France	IPK Gatersleben, Germany
7	*Lotus edulis*	LOT 45	France	IPK Gatersleben, Germany
8	*Lotus conimbricensis*	LOT 46	Italy	IPK Gatersleben, Germany
9	*Lotus tenuis*	LOT 48	Germany	IPK Gatersleben, Germany
10	*Lotus arenarius*	LOT 50	Unknown	IPK Gatersleben, Germany
11	*Lotus cytisoides*	LOT 52	Yugoslavia	IPK Gatersleben, Germany
12	*Lotus drepanocarpus*	LOT 53	France	IPK Gatersleben, Germany
13	*Lotus halophilus*	LOT 54	Italy	IPK Gatersleben, Germany
14	*Lotus uliginosus*	LOT 59	Spain	IPK Gatersleben, Germany
15	*Lotus corniculatus*	LOT 66	Germany	IPK Gatersleben, Germany
16	*Lotus filicaulis*	LF-01	Quentin Cronk	Dr. Dario I. Ojeda Alayon, Canada
17	*Lotus burttii*	B 303	Unknown	Anonymous
18	*Lotus alpinus*	PI 302922	Spain	USDA, USA
19	*Lotus angustissimus*	PI 240730	Spain	USDA, USA
20	*Lotus arabicus*	PI 214109	Spain	USDA, USA
21	*Lotus australis*	DLEG 930095	Australia	USDA, USA
22	*Lotus castellanus*	PI 308036	Slovakia	USDA, USA
23	*Lotus collinus*	PI 287859	Spain	USDA, USA
24	*Lotus conjugatus*	PI 283617	Hungary	USDA, USA
25	*Lotus cruentus*	DLEG 930099	Australia	USDA, USA
26	*Lotus eriosolen*	PI 631784	Morocco	USDA, USA
27	*Lotus gebelia*	PI 464685	Turkey	USDA, USA
28	*Lotus glaucus*	PI 239945	Turkey	USDA, USA
29	*Lotus glinoides*	PI 246736	Spain	USDA, USA
30	*Lotus kunkelii*	DLEG 950044	Spain	USDA, USA
31	*Lotus maroccanus*	PI 631955	Morocco	USDA, USA
32	*Lotus mearnsii*	PI 379650	Turkey	USDA, USA
33	*Lotus palustris*	PI 655784	Italy	USDA, USA
34	*Lotus parviflorus*	PI 283615	Australia	USDA, USA
35	*Lotus peregrinus*	PI 323192	Greece	USDA, USA
36	*Lotus subpinnatus*	PI 368906	Chile	USDA, USA
37	*Lotus unifoliolatus*	DLEG 920048	Argentina	USDA, USA
38	*Lotus japonicus*	MG 20	Japan	Anonymous
39	*Lotus campylocladus*	Carrectera canadas	Tenerite	Dr. Dario I. Ojeda Alayon, Canada
40	*Lotus dumetorum*	Carretera Baradero	Tenerite	Dr. Dario I. Ojeda Alayon, Canada
41	*Lotus emeroides*	Inchora	La Gomera	Dr. Dario I. Ojeda Alayon, Canada
42	*Lotus hillebrandii*	Punta Gorda El Fayal	La Palma	Dr. Dario I. Ojeda Alayon, Canada
43	*Lotus spartioides*	San Bartolome	Gran Canaria	Dr. Dario I. Ojeda Alayon, Canada
44	*Lotus tenellus*	Taganara	Gran Canaria	Dr. Dario I. Ojeda Alayon, Canada
45	*Lotus maritimus*	11570	Switzerland	Millennium Seed Bank, UK
46	*Lotus spectabilis*	23449	Gran Canaria	Millennium Seed Bank, UK
47	*Lotus hispidus*	30933	Greece	Millennium Seed Bank, UK
48	*Lotus garcinii*	96182	Oman	Millennium Seed Bank, UK
49	*Lotus creticus*	119915	Tunisia	Millennium Seed Bank, UK
50	*Lotus baorbasii*	13T1400040	Czech	Plant Genetic Resources, Czech
51	*Lotus villosus*	13T1400028	Unknown	Plant Genetic Resources, Czech

References

Abaidoo,R.C. *et al.* (2000) Bradyrhizobium spp. (TGx) isolates nodulating the new soybean cultivars in africa are diverse and distinct from bradyrhizobia that nodulate north american soybeans. *Int J Syst Evol Microbiol*, **50 Pt 1**, 225–34.

Allan,G.J. *et al.* (2004) Molecular phylogenetic evidence for the geographic origin and classification of canary island lotus (fabaceae: Loteae). *Mol Phylogenet Evol*, **32**, 123–38.

Amor,B.B. *et al.* (2003) The nfp locus of medicago truncatula controls an early step of nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *Plant J*, **34**, 495–506.

Andrews,M. and Andrews,M.E. (2017) Specificity in legume-rhizobia symbioses. *Int J Mol Sci*, **18**.

Angel,E. (2001a) Batch-file computer graphics : A bottom-up approach with quicktime Wesley Addison Longman, Boston, MA.

Angel,E. (2000) Interactive computer graphics : A top-down approach with opengl Addison Wesley Longman, Boston, MA.

Angel,E. (2001b) Test second book by angel Wesley Addison Longman, Boston, MA.

Annapurna,K. and Krishnan,H.B. (2003) Molecular aspects of soybean cultivar-specific nodulation by sinorhizobium fredii usda257. *Indian J Exp Biol*, **41**, 1114–23.

Antolıén-Llovera,M. *et al.* (2014) Cleavage of the symbiosis receptor-like kinase ectodomain promotes complex formation with nod factor receptor 5. *Curr Biol*, **24**, 422–7.

Bek,A.S. *et al.* (2010) Improved characterization of nod factors and genetically based variation in lysm receptor domains identify amino acids expendable for nod factor recognition in lotus spp. *Mol Plant Microbe Interact*, **23**, 58–66.

Bertani,G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic escherichia coli. *J Bacteriol*, **62**, 293–300.

Bozsoki,Z. *et al.* (2017) Receptor-mediated chitin perception in legume roots is

functionally separable from nod factor perception. *Proc Natl Acad Sci U S A*, **114**, E8118–E8127.

Broghammer,A. *et al.* (2012) Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proc Natl Acad Sci U S A*, **109**, 13859–64.

Brulé,D. *et al.* (2019) The grapevine (vitis vinifera) lysm receptor kinases vvlyk1-1 and vvlyk1-2 mediate chitooligosaccharide-triggered immunity. *Plant Biotechnol J*, **17**, 812–825.

Buendia,L. *et al.* (2016) The lysm receptor-like kinase sllyk10 regulates the arbuscular mycorrhizal symbiosis in tomato. *New Phytol*, **210**, 184–95.

Buist,G. *et al.* (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol Microbiol*, **68**, 838–47.

Cao,Y. *et al.* (2014) The kinase lyk5 is a major chitin receptor in arabidopsis and forms a chitin-induced complex with related kinase cerk1. *Elife*, **3**.

Degtjareva,G. *et al.* (2006) Phylogeny of the genus lotus (leguminosae, loteae): Evidence from nrITS sequences and morphology. *Canadian Journal of Botany*, **84**, 813–830.

Dölfors,F. *et al.* (2019) A lysm effector protein from the basidiomycete rhizoctonia solani contributes to virulence through suppression of chitin-triggered immunity. *Mol Genet Genomics*.

Fahraeus,G. (1957) The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. *J Gen Microbiol*, **16**, 374–81.

Fuechtbauer,W. *et al.* (2017) LYS12 lysm receptor decelerates phytophthora palmivora disease progression in lotus japonicus. *Plant J*.

Gherbi,H. *et al.* (2008) 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**, 4928–32.

Gibelin-Viala,C. *et al.* (2019) The medicago truncatula lysm receptor-like kinase lyk9 plays a dual role in immunity and the arbuscular mycorrhizal symbiosis. *New Phytol*, **223**, 1516–1529.

Gossmann,J.A. *et al.* (2012) Polymorphic infection and organogenesis patterns induced by a rhizobium leguminosarum isolate from lotus root nodules are determined by the host genotype. *New Phytol*, **196**, 561–73.

Guindon,S. *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of phyml 3.0. *Syst Biol*, **59**, 307–21.

Indrasumunar,A. *et al.* (2010) Inactivation of duplicated nod factor receptor 5 (nfr5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean

(glycine max l. Merr.). *Plant Cell Physiol*, **51**, 201–14.

Indrasumunar,A. *et al.* (2011) Nodulation factor receptor kinase 1*alpha* controls nodule organ number in soybean (glycine max l. Merr). *Plant J*, **65**, 39–50.

Jonge,R. de *et al.* (2010) Conserved fungal lysm effector ecp6 prevents chitin-triggered immunity in plants. *Science*, **329**, 953–5.

Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*, **30**, 772–80.

Kawaharada,Y. *et al.* (2015) Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature*, **523**, 308–12.

Kitanovski,S. (2016) Genphen: A tool for computing genotype-phenotype associations using statistical learning techniques.

Kombrink,A. *et al.* (2017) Verticillium dahliae lysm effectors differentially contribute to virulence on plant hosts. *Mol Plant Pathol*, **18**, 596–608.

Kück,P. and Longo,G.C. (2014) FASconCAT-g: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool*, **11**, 81.

Lee,W.-S. *et al.* (2014) Mycosphaerella graminicola lysm effector-mediated stealth pathogenesis subverts recognition through both cerk1 and cebip homologues in wheat. *Mol Plant Microbe Interact*, **27**, 236–43.

Leppyanen,I.V. *et al.* (2017) Receptor-like kinase lyk9 in pisum sativum l. Is the cerk1-like receptor that controls both plant immunity and am symbiosis development. *Int J Mol Sci*, **19**.

Limpens,E. *et al.* (2003) LysM domain receptor kinases regulating rhizobial nod factor-induced infection. *Science*, **302**, 630–3.

Liu,T. *et al.* (2012) Chitin-induced dimerization activates a plant immune receptor. *Science*, **336**, 1160–4.

Lohmann,G.V. *et al.* (2010) Evolution and regulation of the lotus japonicus lysm receptor gene family. *Mol Plant Microbe Interact*, **23**, 510–21.

Madsen,E.B. *et al.* (2011) Autophosphorylation is essential for the in vivo function of the lotus japonicus nod factor receptor 1 and receptor-mediated signalling in cooperation with nod factor receptor 5. *Plant J*, **65**, 404–17.

Madsen,E.B. *et al.* (2003) A receptor kinase gene of the lysm type is involved in legume perception of rhizobial signals. *Nature*, **425**, 637–40.

Maekawa,T. *et al.* (2009) Gibberellin controls the nodulation signaling pathway in lotus japonicus. *Plant J*, **58**, 183–94.

Markmann,K. *et al.* (2008) Functional adaptation of a plant receptor-kinase paved the way for the evolution of intracellular root symbioses with bacteria. *PLoS Biol*, **6**, e68.

Mentlak,T.A. *et al.* (2012) Effector-mediated suppression of chitin-triggered immunity by magnaporthe oryzae is necessary for rice blast disease. *Plant Cell*, **24**, 322–35.

Miya,A. *et al.* (2007) CERK1, a lysm receptor kinase, is essential for chitin elicitor signaling in arabidopsis. *Proc Natl Acad Sci U S A*, **104**, 19613–8.

Miyata,K. *et al.* (2014) The bifunctional plant receptor, oscerk1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. *Plant Cell Physiol*, **55**, 1864–72.

Murakami,E. *et al.* (2018) Epidermal lysm receptor ensures robust symbiotic signalling in lotus japonicus. *Elife*, **7**.

Nielsen,R. and Yang,Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the hiv-1 envelope gene. *Genetics*, **148**, 929–36.

Op den Camp,R. *et al.* (2011) LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume parasponia. *Science*, **331**, 909–12.

Petutschnig,E.K. *et al.* (2010) The lysin motif receptor-like kinase (lysm-rlk) cerk1 is a major chitin-binding protein in arabidopsis thaliana and subject to chitin-induced phosphorylation. *J Biol Chem*, **285**, 28902–11.

Radutoiu,S. *et al.* (2003) Plant recognition of symbiotic bacteria requires two lysm receptor-like kinases. *Nature*, **425**, 585–92.

Radutoiu,S. *et al.* (2007) LysM domains mediate lipochitin-oligosaccharide recognition and nfr genes extend the symbiotic host range. *EMBO J*, **26**, 3923–35.

Rasmussen,S.R. *et al.* (2016) Intraradical colonization by arbuscular mycorrhizal fungi triggers induction of a lipochitooligosaccharide receptor. *Sci Rep*, **6**, 29733.

Rey,T. *et al.* (2013) NFP, a lysm protein controlling nod factor perception, also intervenes in medicago truncatula resistance to pathogens. *New Phytol*, **198**, 875–86.

Ried,M.K. *et al.* (2014) Spontaneous symbiotic reprogramming of plant roots triggered by receptor-like kinases. *Elife*, **3**.

Schultz,J. *et al.* (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci U S A*, **95**, 5857–64.

Stracke,S. *et al.* (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature*, **417**, 959–62.

Suyama,M. *et al.* (2006) PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*, **34**, W609–12.

Via,V.D. *et al.* (2016) How legumes recognize rhizobia. *Plant Signal Behav*, **11**, e1120396.

Vincent,J.M. (1970) A manual for the practical study of root-nodule bacteria. International Biological Programme by Blackwell Scientific, Oxford.

Wan,J. *et al.* (2012) LYK4, a lysin motif receptor-like kinase, is important for chitin signaling and plant innate immunity in arabidopsis. *Plant Physiol*, **160**, 396–406.

Wong,J.E.M.M. *et al.* (2019) A lotus japonicus cytoplasmic kinase connects nod factor perception by the nfr5 lysm receptor to nodulation. *Proc Natl Acad Sci U S A*, **116**, 14339–14348.

Yang,Z. (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, **24**, 1586–91.

Yang,Z. and Nielsen,R. (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*, **19**, 908–17.

Yang,Z. and Swanson,W.J. (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol*, **19**, 49–57.

Yang,Z. *et al.* (2000) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol Biol Evol*, **17**, 1446–55.

Yu,G. *et al.* (2017) Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, **8**, 28–36.

Zeng,L. *et al.* (2012) A tomato lysm receptor-like kinase promotes immunity and its kinase activity is inhibited by avrptob. *Plant J*, **69**, 92–103.

Zgadzaj,R. *et al.* (2016) 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.

Zhang,J. *et al.* (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*, **22**, 2472–9.

Zhang,X. *et al.* (2015) The receptor kinase cerk1 has dual functions in symbiosis and immunity signalling. *Plant J*, **81**, 258–67.

Zhukov,V. *et al.* (2008) The pea sym37 receptor kinase gene controls infection-thread initiation and nodule development. *Mol Plant Microbe Interact*, **21**, 1600–8.