New sources of *Sym2^A* allele in the pea (*Pisum sativum* L.) carry the unique variant of candidate LysM-RLK gene *LykX*

Anton S. Sulima [Corresp., Equal first author, 1] , **Vladimir A. Zhukov** [Equal first author, 1] , **Olga A. Kulaeva** [1] , **Ekaterina N. Vasileva** [1] , **Alexey Y. Borisov** [1] , **Igor A. Tikhonovich** [1, 2]

[1] All-Russia Research Institute for Agricultural Microbiology, Saint-Petersburg, Russia

[2] Faculty of Biology, St. Petersburg State University, Saint-Petersburg, Russia

Corresponding Author: Anton S. Sulima
Email address: ASulima@arriam.ru

At the onset of legume-rhizobial symbiosis, the mutual recognition of partners occurs based on a complicated interaction between signal molecules and receptors. Bacterial signal molecules named Nod factors ("nodulation factors") are perceived by the plant LysM-containing receptor-like kinases (LysM-RLKs) that recognize details of its structure (i.e., unique substitutions), thus providing the conditions particular to symbiosis. In the garden pea (*Pisum sativum* L.), the allelic state of *Sym2* gene has long been reported to regulate the symbiotic specificity: for infection to be successful, plants with the *Sym2^A* allele (for "*Sym2* Afghan", as these genotypes originate mostly from Afghanistan) require an additional acetylation of Nod factor which is irrelevant for genotypes with the *Sym2^E* allele (for "*Sym2* European"). Despite being described about 90 years ago, *Sym2* has not yet been cloned, though phenotypic analysis suggests it probably encodes a receptor for Nod factor. Recently, we described a novel pea gene *LykX (PsLykX)* from the LysM-RLK gene family that demonstrates a perfect correlation between its allelic state and the symbiotic specificity of the *Sym2^A*-type. Here we report on a series of Middle-Eastern pea genotypes exhibiting the phenotype of narrow symbiotic specificity discovered in the VIR plant genetic resources gene bank (Saint-Petersburg, Russia). These genotypes are new sources of *Sym2^A*, as has been confirmed by an allelism test with *Sym2^A* pea cv. Afghanistan. Within these genotypes, *LykX* is present either in the allelic state characteristic for cv. Afghanistan, or in another, minor allelic state found in two genotypes from Tajikistan and Turkmenistan. Plants carrying the second allele demonstrate the same block of rhizobial infection as cv. Afghanistan when inoculated with an incompatible strain. Intriguingly, this "Tajik" allele of *LykX* differs from the "European" one by a single nucleotide polymorphism leading to an R75P change in the receptor part of the putative protein. Thus, our new data are in agreement with the hypothesis concerning the identity

of *LykX* and the elusive *Sym2* gene.

1 **New sources of *Sym2^A* allele in the pea (*Pisum sativum* L.)**
2 **carry the unique variant of candidate LysM-RLK gene *LykX***

4 Anton S. Sulima[1], Vladimir A. Zhukov[1], Olga A. Kulaeva[1], Ekaterina N. Vasileva[1], Alexey Y. Borisov[1], Igor A.
5 Tikhonovich[1,2]
6
7 [1] All-Russia Research Institute for Agricultural Microbiology, St. Petersburg, Russia
8 [2] St. Petersburg State University, Faculty of Biology, St. Petersburg, Russia
9
10 Corresponding Author:
11 Anton Sulima[1]
12 Podbelsky chausse 3, Saint-Petersburg, Pushkin 8, 196608, Russia
13 Email address: ASulima@arriam.ru

New sources of *Sym2ᴬ* allele in the pea (*Pisum sativum* L.) carry the unique variant of candidate LysM-RLK gene *LykX*

Anton S. Sulima[1], Vladimir A. Zhukov[1], Olga A. Kulaeva[1], Ekaterina N. Vasileva[1], Alexey Y. Borisov[1], Igor A. Tikhonovich[1,2]

[1] All-Russia Research Institute for Agricultural Microbiology, St. Petersburg, Russia
[2] St. Petersburg State University, Faculty of Biology, St. Petersburg, Russia

Corresponding Author:
Anton Sulima[1]
Podbelsky chausse 3, Saint-Petersburg, Pushkin 8, 196608, Russia
Email address: ASulima@arriam.ru

Abstract

At the onset of legume-rhizobial symbiosis, the mutual recognition of partners occurs based on a complicated interaction between signal molecules and receptors. Bacterial signal molecules named Nod factors ("nodulation factors") are perceived by the plant LysM-containing receptor-like kinases (LysM-RLKs) that recognize details of its structure (i.e., unique substitutions), thus providing the conditions particular to symbiosis. In the garden pea (*Pisum sativum* L.), the allelic state of *Sym2* gene has long been reported to regulate the symbiotic specificity: for infection to be successful, plants with the *Sym2ᴬ* allele (for "*Sym2* Afghan", as these genotypes originate mostly from Afghanistan) require an additional acetylation of Nod factor which is irrelevant for genotypes with the *Sym2ᴱ* allele (for "*Sym2* European"). Despite being described about 90 years ago, *Sym2* has not yet been cloned, though phenotypic analysis suggests it probably encodes a receptor for Nod factor. Recently, we described a novel pea gene *LykX (PsLykX)* from the LysM-RLK gene family that demonstrates a perfect correlation between its allelic state and the symbiotic specificity of the *Sym2ᴬ*-type. Here we report on a series of Middle-Eastern pea genotypes exhibiting the phenotype of narrow symbiotic specificity discovered in the VIR plant genetic resources gene bank (Saint-Petersburg, Russia). These genotypes are new sources of *Sym2ᴬ*, as has been confirmed by an allelism test with *Sym2ᴬ* pea cv. Afghanistan. Within these genotypes, *LykX* is present either in the allelic state characteristic for cv. Afghanistan, or in another, minor allelic state found in two genotypes from Tajikistan and Turkmenistan. Plants carrying the second allele demonstrate the same block of rhizobial infection as cv. Afghanistan when inoculated with an incompatible strain. Intriguingly, this "Tajik" allele of *LykX* differs from the "European" one by a single nucleotide polymorphism leading to an R75P change in the receptor part of the putative protein. Thus, our new data are in agreement with the hypothesis concerning the identity of *LykX* and the elusive *Sym2* gene.

Introduction

The mutualistic symbiosis formed by legume plants (family *Fabaceae*) and nodule bacteria (a diverse group of bacteria collectively called rhizobia) is an example of a highly specific interaction of symbiotic partners. Communication between plant and bacteria relies on an exchange of signal molecules, a process known as "molecular cross-talk" (Oldroyd, 2001; Cooper, 2007; Zipfel & Oldroyd, 2017). The roots of a legume plant excrete flavonoid compounds attracting rhizobia, that, in turn, begin to produce the bacterial signal molecules called Nod factors (Nodulation factors) (Spaink, 1995; Schultze & Kondorosi, 1998; Perret, Staehelin & Broughton, 2000; Geurts & Bisseling, 2002). Nod factors are lipochitooligosaccharide molecules that have specific decorations characteristic for a particular species, or even strains, of rhizobia. Legume plants possess receptors that are able to bind Nod factors in very low concentrations and recognize the fine structure of the ligands, including the specific moieties (Bensmihen, de Billy & Gough, 2011; Delaux et al., 2015; Geurts, Xiao & Reinhold-Hurek, 2016; also reviewed in Gough et al. 2018), so that only a perfectly fitting strain of rhizobia is allowed to enter the root through the root hairs. The successful recognition of a Nod factor activates a signal transduction cascade leading to a change in gene expression it root epidermal cells (Larrainzar et al., 2015; Liu et al., 2019), formation of bacterial microcolony on a tip of the root hair and growth of so-called "infection thread" – a structure inside the root hair through which rhizobia are delivered to root cells that form a nodule primordium (Downie & Walker, 1999; Perret, Staehelin & Broughton, 2000). If plant recognizes that the Nod factor of a particular bacterial strain does not fit, the aforementioned processes do not occur.

The phenomenon of the specificity of legume-rhizobial symbiosis was noticed at least 90 years ago, when the Russian geneticist L.I. Govorov (1928), followed by Russian microbiologist Z.G. Razumovskaya (1937), described pea (*Pisum sativum* L.) specimens originating from Afghanistan as incapable of forming nodules in European soils (Govorov, 1928; Razumovskaya, 1937). Further on, some specimens of pea originating from several Middle East countries were described as failing to nodulate when interacting with rhizobia strains that are common in Europe (Lie, 1971, 1978; Lie, Timmermans & Ladizinski, 1982; Lie & Timmermans, 1983). These specimens, however, established symbiosis with strains isolated from the soils of Turkey, Israel and some other countries. Later it was found that all that strains were characterized by a specific modification of the Nod factor molecule, namely the additional acetylation at the reducing end of the chitin backbone in the C-6 position caused by the acetyltransferase nodX (Davis, Evans & Johnston, 1988; Firmin et al., 1993). Subsequently, in the work of Ovtsyna et al. (Ovtsyna et al., 1998) it was shown that the gene *nodX* can be functionally replaced by the gene *nodZ* encoding a fucosyltransferase that fucosylates the reducing end of Nod factor.

This unusual plant nodulation trait was called the "Afghan" phenotype, since the typical genotype of pea with such a phenotypic manifestation originates from Afghanistan (today this genotype is known as cv. Afghanistan, JI1357 and NGB2150) (Young & Matthews, 1982; Tsyganov et al., 1999). In contrast, the trait of broad specificity – the ability to be nodulated by strains from both Europe and Middle East – was named the "European" phenotype. From the

93   plant side, the manifestation of the "Afghan" phenotype has been shown to be controlled by a
94   single gene named *Sym2*, that was subsequently localized in the Ist linkage group of the pea
95   genome (Kozik et al., 1995, 1996b). Since rhizobia require the unique modification of the Nod
96   factor molecule to successfully interact with the "Afghan" peas, it was assumed that *Sym2* may
97   encode a Nod factor receptor with two different allelic states: *Sym2^A* (from "Afghan", highly
98   specific), and *Sym2^E*, or *Sym2^C* (from "European" or "cultivated", less specific) (Kozik et al.,
99   1996a; Geurts et al., 1997), although formal evidence of it is still lacking.
100 In 2003, genes encoding LysM-domain receptor-like kinases (LysM-RLKs) that probably
101 perceive the Nod factor were described in the model legumes *Lotus japonicus* (Regel) Larsen
102 (*NFR1* and *NFR5*) and *Medicago truncatula* Gaertn. (*LYK3*) (Limpens et al., 2003; Madsen et
103 al., 2003; Radutoiu, 2003). Interestingly, Limpens et al. (2003) found a cluster of LysM-RLK
104 genes in the *Sym2*-orthologous region of the *M. truncatula* genome and described one of them –
105 *LYK3* – as the best candidate for encoding the Nod factor receptor. In 2008, Zhukov and
106 colleagues discovered two pea homologs of *L. japonicus NFR1*, namely *Sym37* and *K1*, located
107 in linkage group I, just where *Sym2* is located (Zhukov et al., 2008). Unfortunately, neither
108 *Sym37* nor *K1* demonstrated any sequence features that could distinguish "Afghan" lines from
109 "European" ones, although products of these genes still may play a role in determination of
110 *Sym2^A*-type specificity as parts of the protein complexes. In our previous work (Sulima et al.,
111 2017), a new pea gene named *LykX* (or *PsLykX*, for *Pisum sativum* LysM Kinase eXclusive)
112 located in the same region of linkage group I was described. When analyzing the polymorphism
113 of its first exon, a minor haplotype was found which is unique for lines with the "Afghan"
114 phenotype. Thus, *LykX* has become a new candidate for the *Sym2*.
115 One of the drawbacks of many previous works aimed at finding the *Sym2* gene has been the use
116 of a limited number of "Afghan" lines with high genetic uniformity; for example, they all
117 possessed the same alleles of the *Sym37* and *K1* genes. This may be due to the fact that these
118 lines are close relatives, or even descendants of the same original genetic line. More than 30
119 years ago, the study of wild varieties of pea from Afghanistan, Iran, Tibet and Turkey revealed
120 seven varieties with the "Afghan" specificity, but no molecular-genetic analysis had been
121 performed (Kneen & LaRue, 1984). In the present work, the similar approach was employed: we
122 attempted to discover new "Afghan" forms among 122 pea samples originating from different
123 regions of the Middle East. Our aim was to expand the sample of available "Afghan" lines, to
124 evaluate the polymorphism of the *LykX* gene, and to see if it is associated with particular
125 preferences towards rhizobia.
126
127 ## Materials & Methods
128
129 ### Plant material
130 The search for pea lines demonstrating the "Afghan" phenotype was conducted on the basis of
131 material obtained from the VIR plant genetic resources gene bank (N. I. Vavilov All-Russian
132 Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia), kindly provided by Prof. M.

133 A. Vishnyakova, head of Department of genetic resources of leguminous crops, and E. V.
134 Semenova, curator of the pea collection. The choice of material was determined by the
135 geographical origin of samples (the Middle East). The whole sample contained 122 pea lines (see
136 Table S1).
137 Additionally, a number of pea cultivars and genetic lines with the confirmed "Afghan"
138 phenotype from the collection of the laboratory of genetics of plant-microbial interactions of the
139 All-Russia Research Institute of Agricultural Microbiology (ARRIAM; St. Petersburg, Russia)
140 were used as control (Table S2). The line 701 of *Pisum fulvum* Sibth. & Smith (Israel, Valley of
141 the Cross) was kindly provided by Sc.D. O.E. Kosterin (Novosibirsk, Russia).
142
143 **Plant growth conditions**
144 During the summer season of 2011, plants were grown in vegetation houses in conditions
145 corresponding to the region's climate (St. Petersburg, Leningrad Oblast, Russia). When
146 necessary for control, plants were grown in Vötsch Industrietechnik VB 1014 growth chambers
147 (Balingen, Germany) under the following conditions:
148 • day/night: 16/8 hours;
149 • temperature: 21°C;
150 • relative air humidity: 75%;
151 • illumination: 38000 lux.
152 The seeds were pre-sterilized with concentrated H_2SO_4 for between 10 and 15 minutes and then
153 washed five times with distilled water. Sterilized seeds were germinated in Petri dishes filled
154 with sterile vermiculite for 3 days at 27°C, then seedlings were planted in pots. For large-scale
155 experiments in vegetation houses, sterilization and preliminary germination of seeds was
156 omitted.
157 Plants were grown in the 2 liter pots filled with sand. When sterility was necessary, pots together
158 with sand were heated in a dry-heat cabinet for 2 hours at 200°C. Mineral nutrition (Table S3)
159 was added to substrate simultaneously with planting seedlings, 150 ml of solution per pot. When
160 the analysis of the symbiotic phenotype was required, combined nitrogen in the form of
161 ammonium nitrate was not added to the nutrient solution.
162
163 **Bacterial strains**
164 In experiments involving analysis of nodulation, two contrast strains of *Rhizobium*
165 *leguminosarum* bv. *viciae* were used:
166 Strain *R. leguminosarum* bv. *viciae* RCAM1026 (= CIAM1026), used in ARRIAM as a model
167 strain (Afonin et al., 2017). This strain is unable to colonize a plant with the "Afghan" phenotype
168 due to absence of the *nodX* gene (*nodX⁻* strain).
169 Strain *R. leguminosarum* bv. *viciae* A1, first isolated in the Leningrad Oblast from nodules of the
170 "Afghan" pea line (Chetkova & Tikhonovich, 1986), thus possessing the *nodX* gene (*nodX⁺*
171 strain).
172 For the visualization of symbiotic phenotype, the *gusA*-marked derivative of RCAM1026
173 (RCAM1026 *gusA*) with constitutive expression of *gusA* gene was used (Kirienko et al., 2018).

174

Bacteria cultivation

175
176 The strains were cultivated on solid medium №79 (Table S4) in a thermostat at a temperature of
177 28°C for three days. In the case of streptomycin-resistant strain RCAM1026, streptomycin (100
178 µg/ml) was added to the medium; in the case of strain A1, no antibiotics were added. Plants were
179 inoculated with a suspension of bacteria: the culture ("lawn") from one Petri dish was
180 resuspended in 1 liter of sterile distilled water (final CFU/ml – not less than 10^6) and the
181 resulting solution was added to the substrate while planting the seedlings, 100 ml per pot.

182

Plant DNA Extraction

183
184 DNA was extracted from young leaves (top or second-from-top node). Extraction was conducted
185 according to the previously described CTAB protocol (Rogers & Bendich, 1985; Sulima et al.,
186 2017).

187

Sequencing of DNA fragments

188
189 To amplify the first exons of *LykX*, primers were used as described in Sulima et al. (2017). The
190 PCRs were performed in 0.5 ml eppendorf-type microcentrifuge tubes on an iCycler (Bio-Rad,
191 Hercules, CA, USA) or Dyad (Bio-Rad, USA) thermocycler using the ScreenMix-HS kit
192 (Evrogen, Moscow, Russia). The PCR cycling conditions were as follows: 95°C (5min), 35 ×
193 [95°C (30 sec), Tm (varying depending on primers) (30 sec), 72°C (1min)], 72°C (5min). The
194 PCR fragments were sequenced using the ABI Prism3500xL system (Applied Biosystems, Palo
195 Alto, CA, USA) at the "Genomic Technologies, Proteomics, and Cell Biology" Core Center of
196 the ARRIAM (St.Petersburg, Russia). The resulting sequences have been deposited into the
197 NCBI database under accession numbers MN187362 – MN187364, MN200353 – MN200358.
198 Analysis of sequences was conducted by means of MEGA7 software (Kumar, Stecher &
199 Tamura, 2016). A Minimum Evolution tree was built using the modified Nei-Gojobori (assumed
200 transition/transversion bias = 2) method (Zhang, Rosenberg & Nei, 1998). A bootstrap test was
201 performed with 500 bootstrap replications. Alignments were done in MEGA7 and EMBL-EBI
202 Clustal Omega web service (https://www.ebi.ac.uk/Tools/msa/clustalo/).

203

Genotyping of F$_2$ population

204
205 To analyze the F$_2$ sample obtained from the cross between lines K-3821 and NGB2150 carrying
206 different alleles of *LykX*, the CAPS (cleaved amplified polymorphic sequence) marker was
207 designed using the dCAPS Finder 2.0 online service (Neff, Turk & Kalishman, 2002). The
208 primers used were the same as for amplification of the 1st exon. The digestion of SNP (single
209 nucleotide polymorphism) site was performed by the Hpy188III restriction enzyme (New
210 England Biolabs, Massachusetts, USA).

211

Microscopy

212
213 For detailed phenotype analysis, plants were grown in sand, as described above, while inoculated
214 with suspension of RCAM1026 *gusA* culture. Four weeks after inoculation, 5 segments of roots

215 2-2.5 cm long were collected from each plant, leaving approximately 2 cm from the base of the
216 root and 2 cm from the tips (if possible). Root segments were stained for 18 hours in a solution
217 of 50mM Na-phosphate buffer (pH 7.0) containing 0.1% Triton, 5mM $K_3Fe(CN_6)$ (red blood
218 salt), 5mM $K_4Fe(CN_6)$ (yellow blood salt) and 0.02 % 5-bromo-4-chloro-3-indolyl-bD-
219 glucuronic acid (X-glc). For storage, stained segments were fixated with 4% paraformaldehyde
220 in a TBS buffer (50mM TrisHCl, 150mM NaCl, pH 7.5) with the addition of 0.1% Tween-20
221 and 0.1% Triton X-100 under a vacuum (7 min, 3-4 times) using aVacuuBrand ME 1C vacuum
222 pump (Vacuubrand, Wertheim, Germany) and then dehydrated in graded ethanol series (from
223 10% to 70%). For observation, fixated samples were treated with the same ethanol series in
224 reverted order and then with TBS buffer. Microscopic analysis was conducted using Zeiss
225 Axiovert 35 inverted phase contrast microscope (ZEISS, Oberkochen, Germany) by 100-fold
226 magnification. The photos were taken on the Canon EOS 4000D camera (Canon, Tokyo, Japan).
227
228 **Results**
229
230 **Selection of pea lines with narrow symbiotic specificity**
231 To increase the amount of the available "Afghan" lines, a sample of 122 pea lines originating
232 from various regions of the Middle East (Table S1) was obtained from the VIR plant genetic
233 resources gene bank. In the summer season of 2011, the nodulation phenotype of the lines was
234 tested by inoculation with $nodX^+$ or $nodX^-$ strains. Five lines were discovered which did not
235 form nodules when inoculated with $nodX^-$ strain RCAM1026, but demonstrated normal
236 nodulation with strain A1 (i.e., the typical "Afghan" phenotype; so did plants of cv. Afghanistan
237 (NGB2150) as well) (Table 1). From among the other lines analyzed, there were several with
238 decreased or increased numbers of nodules formed by either the RCAM1026 ($nodX^-$) strain or
239 A1 ($nodX^+$) strain, but these were not included in subsequent analysis.
240
241 **Allelism test between newly discovered and previously known *Sym2^A*-lines**
242 In order to determine whether the symbiotic phenotype of selected pea accessions was caused by
243 the same gene as in the previously known "Afghan" lines (i.e., the *Sym2* gene), an allelism test
244 was conducted. F_1 seeds obtained from the cross of previously reported "Afghan" lines
245 NGB2150, K-6883 and K-6047-2 with lines selected from VIR gene bank, along with seeds of
246 parental lines, were planted in pots with sterile sand and inoculated with the $nodX^-$ rhizobia
247 strain RCAM1026 (see Materials and Methods). Additionally, two seeds of cv. Finale
248 (characterized by the broad "European" symbiotic specificity and thus able to interact with the
249 $nodX^-$ strain RCAM1026) were planted in each pot to confirm the success of the inoculation.
250 After 4 weeks of vegetation, F_1 hybrids, like the plants of parental lines, did not form nodules
251 with the $nodX^-$ strain RCAM1026. Plants of cv. Finale in the same pots formed, on average,
252 about 40 nodules per plant (see Table S5). The lack of nodules in F_1 plants indicates the allelism
253 of genetic determinants of symbiotic phenotype in newly-described and previously known

254   "Afghan" pea lines. Thus, new sources of $Sym2^A$ alleles have been identified: the lines K-1878,
255   K-3374, K-3821, K-4902, and K-6559.
256
257   **Sequencing of the first exon of *LykX* in the selected lines of pea**
258   In our previous work (Sulima et al., 2017), a new LysM-RLK gene *LykX (PsLykX)* was identified
259   as a promising candidate for the elusive gene *Sym2*, as it demonstrated the unique allelic state
260   found only in two pea lines with the "Afghan" phenotype which leads to substitution of three
261   amino acids in its corresponding protein: Q44R, N45Y, and A76D, all in the first LysM module
262   of the receptor part. Here we sequenced the first exon of *LykX* (corresponding to the receptor
263   domain of the putative protein) in the newly selected "Afghan" lines from the VIR gene bank,
264   several lines with "European" phenotype from our laboratory collection, cv. Iran characterized
265   by the leaky temperature-dependent "Afghan" phenotype (Lie, 1971; Kozik et al., 1996a), and a
266   wild relative of garden pea, *Pisum fulvum*.
267   It was revealed that, in the majority of lines with "Afghan" phenotype, the 1st exon of *LykX* has
268   the same allelic state that was discovered previously (Sulima et al., 2017). In lines K-3821 and
269   K-3374, however, *LykX* demonstrated another allelic state with unique substitutions observed
270   neither in "European" nor in other "Afghan" lines. These lines were called "Tajik", according to
271   the place of origin of line K-3821, which was identified first. Taken this new find into account,
272   we suspected that the results of our initial allelism test might be ambiguous, since the similar
273   non-nodulation phenotype in F_1 would be observed if the block of nodulation in the "Tajik" lines
274   is caused by genetically dominant determinant independent of *Sym2*. In order to test that new
275   hypothesis, the nodulation trait in F_2 progeny derived from the cross between "Tajik" line K-
276   3821 and the typical "Afghan" line NGB2150 was analyzed in the inoculation experiment
277   similar to the one described above. Of 119 F_2 plants, none demonstrated normal nodulation with
278   RCAM1026 strain regardless of the allelic state of *LykX* (homozygous "Afghan" allele,
279   homozygous "Tajik" allele, heterozygote), while control plants of cv. Finale formed, on average,
280   65.3 ± 20 (m±SD, n=28) nodules per plant (see Table S6 and Fig. S1). This suggests that our
281   initial interpretation of allelism test results was correct.
282   The nucleotide substitutions unique for "Afghan" lines lead to the replacement of three amino
283   acids in the protein in comparison to the "European" lines (Table 2). "Tajik" lines seem to be
284   much more similar to the "European" ones, with only one unique amino acid substitution in the
285   protein: R75P. It is important to note that all these characteristic substitutions are concentrated in
286   the first LysM module (LysM1) of the receptor domain, which could indirectly indicate the
287   importance of this particular module for the functioning of the LykX protein (Fig. 1).
288   The plants of cv. Iran were found to carry the same allele of *LykX* as the cv. Afghanistan
289   (NGB2150), which corresponds with the idea that their phenotype is determined by the same
290   gene (Kozik et al., 1995, 1996a). Interestingly, the "European" allele of *LykX* was found in
291   *Pisum fulvum* line 701, the wild species closely related to *Pisum sativum*.
292
293   **Phenotype confirmation**

294   According to previous descriptions (Geurts et al., 1997; Zhukov et al., 2008), "Afghan" peas are
295   unable to form normally developed infection threads while inoculated with the *nodX⁻* strain of
296   rhizobia. Bearing this in mind, we performed the microscopic analysis of roots of "Tajik" lines in
297   order to characterize their symbiotic phenotype. Seedlings of the "Tajik" line K-3821 and the
298   "Afghan" line NGB2150, along with the "European" cv. Caméor, were inoculated with *gusA*-
299   marked strain CIAM1026. 4 weeks after inoculation, the root fragments were harvested, treated
300   with staining buffer and observed under a light microscope. Both lines demonstrated normal root
301   hair curling and formation of microcolonies, but infection was aborted at this stage so no
302   infection threads were observed for neither "Afghan" nor "Tajik" lines on 25 cm of roots (Fig.
303   2B, C). Control plants of cv. Caméor developed normal infection threads (Fig. 2A). Accordingly,
304   plants of cv. Caméor formed 81.6±15.2 (m±SD, n=10) nodules, while NGB2150 and K-3821
305   formed 0.2±0.6 and 0.6±1.3 nodules, respectively. Thus, we confirmed that "Afghan" and
306   "Tajik" pea lines have an identical symbiotic phenotype, which is consistent with the alterations
307   in the first exon sequences of *LykX* gene as compared to the "European" variant.
308

Discussion

310   Legume-rhizobial symbiosis relies on proper mutual recognition of partners, a process involving
311   signal molecules and receptors that have been evolving step-by-step in a fashion similar to an
312   "arms race". Indeed, a plant needs to limit the set of bacteria penetrating its tissues as part of its
313   defenses from pathogens and/or incompatible symbionts. On the other hand, bacteria may benefit
314   when they broaden the range of possible hosts. Consequently, plants have evolved highly
315   specific receptors that allow for discrimination of compatible and incompatible bacteria; in turn,
316   rhizobia develop more and more variants of signal molecules (Price et al., 1992) that may be
317   recognized by highly discriminatory plant receptors.
318   The pea is a unique object for studying the specificity of plant-microbe interactions due to the
319   existence of a distinct group of specimens from Afghanistan showing a particular receptivity
320   towards the microsymbiont. Such a phenotype is characteristic for plants carrying the *Sym2ᴬ*
321   allele conferring "resistance" to nodulation with rhizobia strains producing the Nod factor
322   without the required decoration. According to genotyping data (Jing et al., 2010), Afghanistan
323   peas represent a distinct clade remote from both modern garden pea varieties and *Pisum fulvum*,
324   the wild relative to *Pisum sativum*. Consistent with this, the recently-discovered LysM-RLK
325   gene *LykX* is present among peas from Afghanistan and adjacent countries in a unique form that
326   encodes a protein differing by three amino acid residues from that of "European" peas and of
327   *Pisum fulvum* (Table 2). This unique haplotype is associated with a specific "Afghan" nodulation
328   phenotype, which makes *LykX* a good candidate for the *Sym2* gene. The "Afghan" haplotype of
329   *LykX* was found in 3 previously uncharacterized lines originating from Afghanistan and
330   Uzbekistan.
331   Interestingly, a different allele of *LykX* was discovered in two lines from Tajikistan and
332   Turkmenistan manifesting the same strict symbiotic phenotype. To date we can't confirm that a
333   single nucleotide swap resulting in R75P amino acid substitution in the "Tajik" variant of LykX

334 protein actually affects its functionality in a way that leads to a narrow receptivity, but proline is
335 known by its increased conformational rigidity which indeed may have a significant effect on the
336 secondary structure of the protein (Biedermannova et al., 2008). Yet, in the light of our recent
337 findings, we can speculate that especially "choosy" forms of symbiosis may sometimes be an
338 adaptive trait, as the "Tajik" allele of *LykX* arose from the "European" variant independently and
339 more recently than the "Afghan" one, which is supported by a minimum-evolution tree
340 constructed using the modified Nei-Gojobori method (Fig. 3). However, the reasons why peas
341 growing in this particular region (the Middle East) had to narrow down their symbiotic
342 specificity several times remain unclear. It can be assumed that certain environmental conditions
343 (including the behavior of the native soil microbiota) forced them to more strictly limit the range
344 of bacteria that have access to their root tissue.
345 *Pisum fulvum*, which we used as an outgroup in the study, possesses the "European" variant of
346 *LykX*, implying that it might form nodules with *nodX⁻* strains of *Rhizobium leguminosarum*.
347 However, *P. fulvum* has been previously reported to form ineffective nodules with European
348 strains (Lie, 1981), so it is likely that the infection stage of *P. fulvum* with *nodX⁻* strains may
349 proceed normally. Another noteworthy pea accession in this work is cv. Iran, which, according to
350 previous data, possesses the *Sym2^A* allele (Kozik et al., 1995) and, as shown in current work,
351 carries the "Afghan" allele of *LykX*. In our experiments it did not nodulate with the *nodX⁻* strain,
352 although previously it has been documented that cv. Iran can form nodules with *nodX⁻* strains
353 under high temperatures (28°C) (Lie, 1971; Kozik et al., 1996a). Using *Sym2^A* introgression
354 lines, Kozik and colleagues have shown that temperature-dependent phenotype is intrinsic to cv.
355 Afghanistan (NGB2150) as well, but is somewhat masked by the native genetic background
356 (Kozik et al., 1996a). Our findings so far are in concert with these data, but the phenomenon in
357 general may require further investigation. Moreover, there might be a chance of finding lines
358 similar to cv. Iran in our sample of 122 pea accessions, as there are some lines with decreased
359 numbers of nodules, and the temperatures in the Leningrad Oblast reached up to 30°C during the
360 summer of 2011.
361 Thus, our data are in agreement with the hypothesis of the identity of *Sym2* and *LykX*.
362 Accordingly, it seems appropriate to name the newly discovered allele of *LykX Sym2^T* (for
363 "Tajik"). However, genetic transformation experiments are still necessary to conclusively
364 confirm or reject this hypothesis. It would also be essential to study the interaction of "Tajik"
365 lines with rhizobial strains producing fucosylated Nod factor due to the presence of the *nodZ*
366 gene, as they range of specificity may differ from that of the "typical Afghan" lines.
367
Conclusions
369 In this work we identified a series of naturally-occurred forms of pea (*Pisum sativum* L.) with the
370 narrow symbiotic specificity similar to that of cv. Afghanistan, and confirmed that their
371 phenotype is prompted by the same genetic determinant *Sym2*. It is highly likely that the recently
372 discovered gene *LykX* is in fact the elusive *Sym2* gene. Indeed, we found that within our sample,
373 *LykX* demonstrates two allelic states ("Tajik" and "Afghan") different from those found in forms

374 with normal symbiotic receptivity ("European"). Both allelic states lead to different amino acid

375 substitutions in their corresponding proteins, as compared to the more common "European" state.

376 In neither *K1* nor *Sym37* – the previous candidate genes for *Sym2* – were observed such

377 correlations between allelic states and the "Afghan" phenotype. The results of our work also

378 highlight the importance of the strict specificity of symbiosis for the pea, since, according to our

379 data, it appeared independently at least twice during the evolution of *Pisum sativum* species.

380

Acknowledgements

382 Part of our research (Sanger sequencing, microscopy) was performed using equipment of the

383 Core Center "Genomic Technologies, Proteomics and Cell Biology" at the All-Russia Research

384 Institute for Agricultural Microbiology (ARRIAM), St.Petersburg, Russia. Biological material

385 (pea seeds) were kindly provided by Prof. M. A. Vishnyakova (N. I. Vavilov All-Russian

386 Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia), E. V. Semenova (N. I.

387 Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia) and

388 Sc.D. O.E. Kosterin (Institute Of Cytology And Genetics, Novosibirsk, Russia). The RCAM1026

389 *gusA* strain of *Rhizobium leguminosarum* bv. *viciae* was kindly provided by Dr. G.A.

390 Akhtemova (All-Russia Research Institute for Agricultural Microbiology (ARRIAM),

391 St.Petersburg, Russia).

392

References

394 Afonin A, Sulima A, Zhernakov A, Zhukov V. 2017. Draft genome of the strain RCAM1026
395   Rhizobium leguminosarum bv. viciae. *Genomics data* 11:85–86.

396 Bensmihen S, de Billy F, Gough C. 2011. Contribution of NFP LysM domains to the recognition
397   of Nod factors during the Medicago truncatula/Sinorhizobium meliloti symbiosis. *PLoS*
398   *ONE* 6. DOI: 10.1371/journal.pone.0026114.

399 Biedermannova L, Riley KE, Berka K, Hobza P, Vondrasek J. 2008. Another role of proline:
400   stabilization interactions in proteins and protein complexes concerning proline and
401   tryptophane. *Physical Chemistry Chemical Physics* 10:6350–6359. DOI:
402   10.1039/b805489b.

403 Borisov AY, Rozov SM, Tsyganov VE, Morzhina E V., Lebsky VK, Tikhonovich IA. 1997.
404   Sequential functioning of Sym-13 and Sym-31, two genes affecting symbiosome
405   development in root nodules of pea (Pisum sativum L.). *Molecular and General Genetics*
406   254:592–598. DOI: 10.1007/s004380050456.

407 Chetkova SA, Tikhonovich IA. 1986. Isolation and investigation of Rhizobium leguminosarum
408   strains effective on peas of Afghan origin. *Microbiology (USA)*.

409 Cooper JE. 2007. Early interactions between legumes and rhizobia: disclosing complexity in a
410   molecular dialogue. *Journal of applied microbiology* 103:1355–1365.

411 Dalmais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savois V, Aubert G, Brunaud V, de
412   Oliveira Y, Guichard C. 2008. UTILLdb, a Pisum sativum in silico forward and reverse
413   genetics tool. *Genome biology* 9:R43.

414 Davis EO, Evans IJ, Johnston AWB. 1988. Identification of nodX, a gene that allows Rhizobium
415   leguminosarum biovar viciae strain TOM to nodulate Afghanistan peas. *Molecular and*
416   *General Genetics MGG* 212:531–535.

417 Delaux P-M, Radhakrishnan G V., Jayaraman D, Cheema J, Malbreil M, Volkening JD,
418   Sekimoto H, Nishiyama T, Melkonian M, Pokorny L, Rothfels CJ, Sederoff HW, Stevenson
419   DW, Surek B, Zhang Y, Sussman MR, Dunand C, Morris RJ, Roux C, Wong GK-S,
420   Oldroyd GED, Ané J-M. 2015. Algal ancestor of land plants was preadapted for symbiosis.
421   *Proceedings of the National Academy of Sciences* 112:13390–13395. DOI:
422   10.1073/pnas.1515426112.
423 Downie JA, Walker SA. 1999. Plant responses to nodulation factors. *Current Opinion in Plant*
424   *Biology* 2:483–489. DOI: 10.1016/S1369-5266(99)00018-7.
425 Firmin JL, Wilson KE, Carlson RW, Davies AE, Downie JA. 1993. Resistance to nodulation of
426   cv. Afghanistan peas is overcome by nodX, which mediates an O-acetylation of the
427   Rhizobium leguminosarum lipo-oligosaccharide nodulation factor. *Molecular Microbiology*
428   10:351–360. DOI: 10.1111/j.1365-2958.1993.tb01961.x.
429 Geurts R, Bisseling T. 2002. Rhizobium Nod factor perception and signalling. *The Plant Cell*
430   14:S239–S249.
431 Geurts R, Heidstra R, Hadri   a. E, Downie J a., Franssen H, Van Kammen A, Bisseling T. 1997.
432   Sym2 of Pea Is Involved in a Nodulation Factor-Perception Mechanism That Controls the
433   Infection Process in the Epidermis. *Plant physiology* 115:351–359. DOI: 115/2/351 [pii].
434 Geurts R, Xiao TT, Reinhold-Hurek B. 2016. What Does It Take to Evolve A Nitrogen-Fixing
435   Endosymbiosis? *Trends in Plant Science* 21:199–208. DOI: 10.1016/j.tplants.2016.01.012.
436 Gough C, Cottret L, Lefebvre B, Bono J-J. 2018. Evolutionary History of Plant LysM Receptor
437   Proteins Related to Root Endosymbiosis. *Frontiers in Plant Science* 9:1–9. DOI:
438   10.3389/fpls.2018.00923.
439 Govorov LI. 1928. The peas of Afghanistan. *Bulletin of Applied Botany (Leningrad)* 19:497–
440   522.
441 Holl FB. 1975. Host plant control of the inheritance of dinitrogen fixation in the Pisum-
442   Rhizobium symbiosis. *Euphytica* 24:767–770.
443 Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, Marshall D, Ambrose MJ, Ellis THN,
444   Flavell AJ. 2010. The genetic diversity and evolution of field pea (Pisum) studied by high
445   throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC*
446   *Evolutionary Biology* 10:44.
447 Kirienko AN, Porozov YB, Malkov N V., Akhtemova GA, Le Signor C, Thompson R, Saffray
448   C, Dalmais M, Bendahmane A, Tikhonovich IA, Dolgikh EA. 2018. Role of a receptor-like
449   kinase K1 in pea Rhizobium symbiosis development. *Planta*. DOI: 10.1007/s00425-018-
450   2944-4.
451 Kneen BE, LaRue TA. 1984. Peas (Pisum sativum L.) with strain specificity for Rhizobium
452   leguminosarum. *Heredity* 52:383.
453 Kneen BE, LaRue TA. 1988. Induced symbiosis mutants of pea (Pisum sativum) and
454   sweetclover (Melilotus alba annua). *Plant Science* 58:177–182.
455 Kosterin OE, Rozov SM. 1993. Mapping of the new mutation blb and the problem of integrity of
456   linkage group I. *Pisum Genet* 25:27–31.
457 Kozik A, Geurts R, Heidstra R, Kulikova O, Ellis THN, Bisseling T, LaRue T, Weeden N.
458   1996a. Fine mapping of the sym2 locus of pea linkage group 1. Kozik.
459 Kozik A, Heidstra R, Horvath B, Kulikova O, Tikhonovich I, Ellis THN, van Kammen A, Lie
460   TA, Bisseling T. 1995. Pea lines carrying syml or sym2 can be nodulated by Rhizobium
461   strains containing nodX; sym1 and sym2 are allelic. *Plant Science* 108:41–49.
462 Kozik A, Matvienko M, Scheres B, Paruvangada VG, Bisseling T, van Kammen A, Ellis THN,

463   LaRue T, Weeden N. 1996b. The pea early nodulin gene PsENOD7 maps in the region of
464   linkage group I containing sym2 and leghaemoglobin. *Plant molecular biology* 31:149–156.
465   Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis
466   version 7.0 for bigger datasets. *Molecular biology and evolution* 33:1870–1874.
467   Langille JE, MacLeod JA, Smeltzer GG, Bubar JS. 1981. Finale field peas. *Canadian Journal of*
468   *Plant Science* 61:1001–1002.
469   Larrainzar E, Riely BK, Kim SC, Carrasquilla-Garcia N, Yu H-J, Hwang H-J, Oh M, Kim GB,
470   Surendrarao AK, Chasman D, Siahpirani AF, Penmetsa R V., Lee G-S, Kim N, Roy S, Mun
471   J-H, Cook DR. 2015. Deep sequencing of the Medicago truncatula root transcriptome
472   reveals a massive and early interaction between nodulation factor and ethylene signals.
473   *Plant Physiology* 169:233–265. DOI: 10.1104/pp.15.00350.

474   Lie TA. 1971. Temperature-dependent root-nodule formation in pea cv. Iran. *Plant and Soil*
475   34:751–752.
476   Lie TA. 1978. Symbiotic specialisation in pea plants: the requirement of specific Rhizobium
477   strains for peas from Afghanistan. *Annals of Applied Biology* 88:462–465.
478   Lie TA. 1981. Gene centres, a source for genetic variants in symbiotic nitrogen fixation: Host-
479   induced ineffectivity inPisum sativum ecotype fulvum. *Plant and Soil* 61:125–134.
480   Lie TA, Timmermans PCJM. 1983. Host-genetic control of nitrogen fixation in the legume-
481   Rhizobium symbiosis: complication in the genetic analysis due to maternal effects. *Plant*
482   *and Soil* 75:449–453. DOI: 10.1007/BF02369979.
483   Lie TA, Timmermans P, Ladizinski G. 1982. Host-controlled nitrogen fixation in the legume-
484   Rhizobium symbiosis. *Israel journal of botany* 31:163–167. DOI:
485   10.1080/0021213X.1982.10676940.
486   Limpens E, Franken C, Smit P, Willemse J, Bisseling T, Geurts R. 2003. LysM domain receptor
487   kinases regulating rhizobial Nod factor-induced infection. *Science* 302:630–633. DOI:
488   10.1126/science.1090074.
489   Liu C-W, Breakspear A, Guan D, Cerri MR, Jackson K, Jiang S, Robson F, Radhakrishnan G V.,
490   Roy S, Bone C, Stacey N, Rogers C, Trick M, Niebel A, Oldroyd GED, de Carvalho-Niebel
491   F, Murray JD. 2019. NIN Acts as a Network Hub Controlling a Growth Module Required
492   for Rhizobial Infection. *Plant Physiology* 179:1704–1722. DOI: 10.1104/pp.18.01572.
493   Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko
494   T, Tabata S, Sandal N, Stougaard J. 2003. A receptor kinase gene of the LysM type is
495   involved in legume perception of rhizobial signals. *Nature* 425:637–640. DOI:
496   10.1038/nature02045.
497   Neff MM, Turk E, Kalishman M. 2002. Web-based primer design for single nucleotide
498   polymorphism analysis. *TRENDS in Genetics* 18:613–615.
499   Oldroyd GED. 2001. Dissecting symbiosis: Developments in Nod factor signal transduction.
500   *Annals of Botany* 87:709–718. DOI: 10.1006/anbo.2001.1410.
501   Ovtsyna AO, Geurts R, Bisseling T, Lugtenberg BJJ, Tikhonovich IA, Spaink HP. 1998.
502   Restriction of Host Range by the *sym2* Allele of Afghan Pea Is Nonspecific for the Type of
503   Modification at the Reducing Terminus of Nodulation Signals. *Molecular Plant-Microbe*
504   *Interactions* 11:418–422. DOI: 10.1094/MPMI.1998.11.5.418.
505   Perret X, Staehelin C, Broughton WJ. 2000. Molecular Basis of Symbiotic Promiscuity.
506   *Microbiology and Molecular Biology Reviews* 64:180–201. DOI:
507   10.1128/MMBR.64.1.180-201.2000.
508   Price NPJ, RelicA B, Talmont F, Lewin A, Promé D, Pueppke SG, Maillet F, Dénarié J, Promé J

509   -C, Broughton WJ. 1992. Broad-host-range Rhizobium species strain NGR234 secretes a
510   family of carbamoylated, and fucosylated, nodulation signals that are O-acetylated or
511   sulphated. *Molecular Microbiology* 6:3575–3584. DOI: 10.1111/j.1365-
512   2958.1992.tb01793.x.
513 Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Grønlund M, Sato S, Nakamura Y,
514   Tabata S, Sandal N. 2003. Plant recognition of symbiotic bacteria requires two LysM
515   receptor-like kinases. *Nature* 425:585. DOI: 10.1109/TDEI.2009.5211872.
516 Razumovskaya ZG. 1937. Nodule formation in various pea cultivars. *Russian J Microbiol*
517   6:321–328.
518 Rogers SO, Bendich AJ. 1985. Extraction of DNA from milligram amounts of fresh, herbarium
519   and mummified plant tissues. *Plant molecular biology* 5:69–76.
520 Schultze M, Kondorosi A. 1998. Regulation of symbiotic root nodule development. *Annual*
521   *review of genetics* 32:33–57.
522 Spaink HP. 1995. The molecular basis of infection and nodulation by rhizobia: the ins and outs
523   of sympathogenesis. *Annual review of phytopathology* 33:345–368. DOI:
524   10.1146/annurev.py.33.090195.002021.
525 Sulima AS, Zhukov VA, Afonin AA, Zhernakov AI, Tikhonovich IA, Lutova LA. 2017.
526   Selection Signatures in the First Exon of Paralogous Receptor Kinase Genes from the Sym2
527   Region of the Pisum sativum L. Genome. *Frontiers in Plant Science* 8. DOI:
528   10.3389/fpls.2017.01957.
529 Tsyganov VE, Voroshilova VA, Kukalev AS, Azarova TS, Yakobi LM, Borisov AY,
530   Tikhonovich IA. 1999. Pisum sativum L. Genes Sym 14 and Sym35 Control Infection
531   Thread Growth Initiation During the Development. *Russian Journal of Genetics* 35:284–
532   291.
533 Young JPW, Matthews P. 1982. A distinct class of peas (Pisum sativum L.) from Afghanistan
534   that show strain specificity for symbiotic Rhizobium. *Heredity* 48:203–210.
535 Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in
536   primate ribonuclease genes. *Proceedings of the National Academy of Sciences* 95:3708–
537   3713.
538 Zhukov V, Radutoiu S, Madsen LH, Rychagova T, Ovchinnikova E, Borisov A, Tikhonovich I,
539   Stougaard J. 2008. The Pea *Sym37* Receptor Kinase Gene Controls Infection-Thread
540   Initiation and Nodule Development. *Molecular Plant-Microbe Interactions* 21:1600–1608.
541   DOI: 10.1094/MPMI-21-12-1600.
542 Zipfel C, Oldroyd GED. 2017. Plant signalling in symbiosis and immunity. *Nature* 543:328–336.
543

Specimens from VIR collection that did not form nodules with $nodX^-$ strain RCAM1026 (n=5)

VIR accession number	Nodules formed with *R. leguminosarum* RCAM1026 (*nodX⁻*)	Nodules formed with *R. leguminosarum* A1 (*nodX⁺*), m±SD	Place of origin
K-1878	0	58.0±5.8	Afghanistan
K-3374	0	61.6±7.9	Turkmenistan
K-3821	0	65.6±4.9	Tajikistan
K-4902	0	50.8±3.5	Uzbekistan
K-6559	0	75.4±8.9	Afghanistan
K-6566	0	94.2±6.2	Afghanistan

Amino acid substitutions in the LykX protein characteristic for pea lines with different specificity of symbiosis

"Afghan" variants are marked with green, "Tajik" – with blue, "European" – with red. *Pisum fulvum* line 701 is marked with yellow. New sources of *Sym2* alleles described in this work are presented in bold. Lines with narrow symbiotic specificity ("Afghan" phenotype) are framed.

Pea line	Substitution position in protein			
	44	45	75	76
NGB2150	R	Y	R	D
K-6883 (84)	R	Y	R	D
K-6047-2	R	Y	R	D
cv. Iran	R	Y	R	D
K-1878	R	Y	R	D
K-4902	R	Y	R	D
K-6559	R	Y	R	D
K-3374	Q	N	P	A
K-3821	Q	N	P	A
Caméor	Q	N	R	A
Finale	Q	N	R	A
SGE	Q	N	R	A
P.fulvum 701	Q	N	R	A

Figure 1

Alignment of the putative receptor domain sequences of LykX protein from lines represented in Table 2

LysM modules are highlighted by color (red – LysM1, green – LysM2, violet – LysM3).

Sequences corresponding to the lines with the $Sym2^A$-type symbiotic specificity are boxed.

Sites that distinguish "Afghan", "European" and "Tajik" variances of LykX are given in bold.

```
Cameor    MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFLLQNITNFMQSKIVSAFNS    60
Finale    MKLIFSLLLFLFLECFFSKVESKCVKGCDIALASYHVMPAFKLQNITNFMQSKIVSAFNS    60
SGE       MKLIFSLLFFLFVECVFFKVESKCVKGCDIALASYHVMPAFKLQNITNFMQSKIVSAFNS    60
Pf_701    MKLIFSLLLFLFVECFFSKVESKCVKGCDIALASYHVMPAFKLQNITNFMQSKIVSAFNS    60
K-3821    MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLQNITNFMQSKIVSAFNS    60
K-3374    MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLQNITNFMQSKIVSAFNS    60
NGB2150   MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLRYITNFMQSKIVSAFNS    60
K-6883    MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLRYITNFMQSKIVSAFNS    60
K-1878    MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLRYITNFMQSKIVSAFNS    60
K-6047-2  MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLRYITNFMQSKIVSAFNS    60
K-6559    MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLRYITNFMQSKIVSAFNS    60
K-4902    MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLRYITNFMQSKIVSAFNS    60
Iran      MKLIFSLLLFLFLECVFFKVESKCVKGCDIALASYHVMPAFKLRYITNFMQSKIVSAFNS    60
          ********:***:**.* ********************* *: **************

Cameor    SDVLIRYNRDILSNRANIFSYFRVNVPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
Finale    SDVLIRYNRDILSNRANIFSYSRVNIPFPCDCIGGEFLGHVFEYTANERDSYDLIANSYY   120
SGE       SDVLIRYNRDILSNRANIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
Pf_701    SDVLIRYNRDILSNRANIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
K-3821    SDVLIRYNRDILSNPANIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
K-3374    SDVLIRYNRDILSNPANIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
NGB2150   SDVLIRYNRDILSNRDNIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
K-6883    SDVLIRYNRDILSNRDNIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
K-1878    SDVLIRYNRDILSNRDNIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
K-6047-2  SDVLIRYNRDILSNRDNIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
K-6559    SDVLIRYNRDILSNRDNIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
K-4902    SDVLIRYNRDILSNRDNIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
Iran      SDVLIRYNRDILSNRDNIFSYFRVNIPFPCDCIGGEFLGHVFEYTANERDTYDLIANSYY   120
          **************  ***** ***:*********************:*********

Cameor    ASLTSVQVLQKFNSYHPNHIPAKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
Finale    ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
SGE       ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
Pf_701    ASLTSVQVLQKFNSYHPNHIPAKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
K-3821    ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
K-3374    ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
NGB2150   ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
K-6883    ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
K-1878    ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
K-6047-2  ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
K-6559    ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
K-4902    ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
Iran      ASLTSVQVLQKFNSYHPNHIPIKAKVNVTVNCSCGNSQISKDYGLFITYPLRSTDSLEKI   180
          ********************  *********************************

Cameor    ANAFKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
Finale    ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
SGE       ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
Pf_701    ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
K-3821    ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
K-3374    ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
NGB2150   ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
K-6883    ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
K-1878    ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
K-6047-2  ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
K-6559    ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
K-4902    ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
Iran      ANASKLDEGLIQNFNPDVNFSKGSGIVFIPGR              212
          *** ***************************
```

Figure 2

Phenotype of the infection threads in plants with different *LykX* alleles in the presence of *nodX⁻* strain RCAM1026*gusA*

(A) normal infection thread formed by the "European" line Caméor. (B,C) early abortion of the infection thread in "Afghan" line NGB2150 (B) and "Tajik" line K-3821 (C). (D,E) multiple curled root hairs of "Aghan" (D) and "Tajik" (E) plants. In most cases, the infection does not go beyond this stage. Scale bares are 0,1 mm.
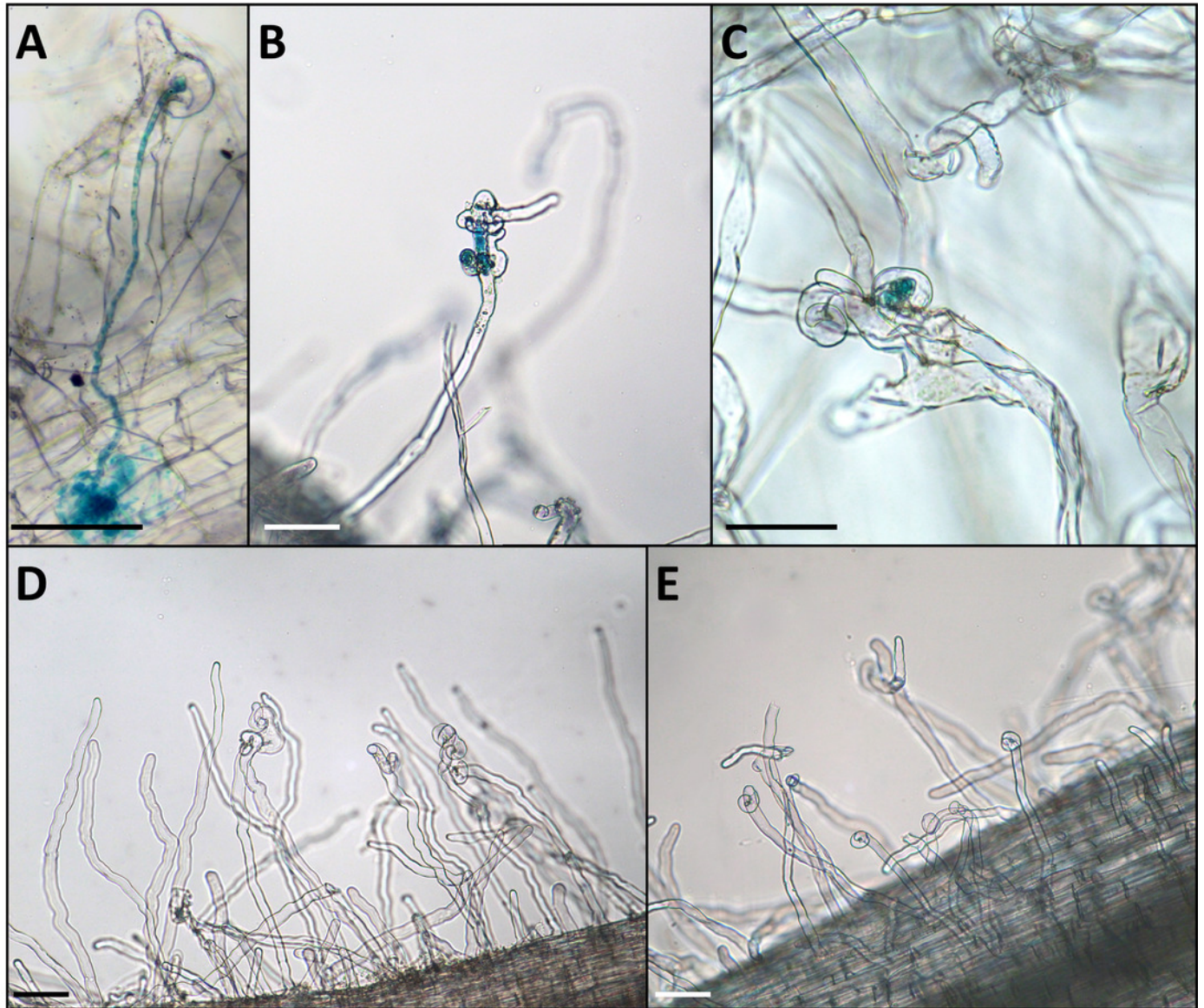
Figure 3

Estimated evolutionary relationships between members of IRL clade of legumes based on sequence of the part of *LykX* gene or its orthologue corresponding to LysM1 module

The evolutionary history was inferred using the Minimum Evolution method. The optimal tree with the sum of branch length = 0.640 is shown. On the right are the SNPs that distinguish the alleles and the amino acids corresponding to them. Pf701 – the LysM1 module of *LykX* gene in *Pisum fulvum* line 701. Mt LYK3 – the LysM1 module of *LYK3* gene in *Medicago truncatula* Gaertn., the close orthologue of *LykX*.