1  **Evolutionary gain and loss of a plant pattern-recognition receptor for HAMP recognition**

2  Simon Snoeck[1], Bradley W. Abramson[2], Anthony G. K. Garcia[1], Ashley N. Egan[3], Todd P.
3  Michael[2], Adam D. Steinbrenner[1, *]

4  [1]Department of Biology, University of Washington, Seattle, Washington 98195, USA

5  [2]The Plant Molecular and Cellular Biology Laboratory, The Salk Institute for Biological
6  Studies, La Jolla, California 92037, USA

7  [3]Department of Biology, Utah Valley University, Orem, Utah 84058, USA

8  *Author for correspondence: astein10@uw.edu

## Abstract

10  Pattern recognition receptors (PRR) recognize distinct pathogen and herbivore-associated
11  molecular patterns (PAMPs and HAMPs) and mediate activation of immune responses, but the
12  evolution of new PRR sensing functions is not well understood. We employed comparative
13  genomics and functional analysis to define evolutionary events leading to the sensing of the
14  peptide HAMP inceptin (In11) by the PRR Inceptin Receptor (INR). Existing and *de novo* genome
15  assemblies revealed that the presence of a functional *INR* gene corresponded with In11 response
16  across 55 million years (my) of legume evolution, and that In11 recognition is unique to the clade
17  of Phaseoloid legumes. The *INR* loci of certain Phaseoloid and non-Phaseoloid species also
18  contain diverse INR-like homologues, suggesting that the evolution of INR receptor function ~28
19  mya occurred after an ancestral gene insertion ~32 mya. Functional analysis of chimeric and
20  ancestrally reconstructed receptors revealed that specific AA differences in the C1 leucine-rich
21  repeat (LRR) domain and C2 intervening motif likely mediated gain of In11 recognition. In
22  summary, we present a conceptual model for the evolution of a defined PRR function based on
23  patterns of *INR* variation in legumes.

## Introduction

25  Plants use a combination of constitutive and inducible defensive traits to resist challenges by
26  pathogens and herbivores. Activation of inducible defenses requires perception of the threat[1],
27  mediated by an innate immune system that uses germline-encoded pattern recognition receptors
28  (PRRs) at the cell surface to recognize pathogen, herbivore, and damage-associated molecular
29  patterns (PAMPs, HAMPs, and DAMPs)[2,3]. Well-characterized PRRs in the large receptor kinase
30  (RK) family comprise an extracellular domain, a single-pass transmembrane motif, and an
31  intracellular kinase domain[4]. Receptor-like proteins (RLPs) share structural similarities with RKs
32  but lack a kinase domain[5,6]. The most common extracellular domain of RK/RLPs is a series of
33  leucine-rich repeats (LRRs), which are known to mediate elicitor binding and co-receptor
34  association[7–11]. LRR-RK/RLPs form a large gene family; the *Arabidopsis thaliana* genome
35  contains about 223 LRR-RKs and 57 LRR-RLPs[5,6,8,12], and the number of annotated RK/RLPs
36  per genome varies across and between plant families[11,13]. Moreover, comparative genomic
37  analyses of LRR-RK/RLPs involved in biotic interactions have revealed strong diversifying
38  selection, lineage-specific expanded gene clusters and immune receptor repertoire variation both
39  within and between species[6,14–17].

40  Evidence is accumulating for diverse roles of RLPs as immune sensors[14,17]. Various LRR-RLPs
41  from Arabidopsis, tomato, and cowpea have been shown to detect molecular patterns from fungi,

42    bacteria, parasitic weeds and herbivores[3,14]. Cf-9, Cf-4 and Cf-2 interact with respective molecular
43    patterns of *Cladosporium fulvum*, Avr9, Avr4 and Avr2 (via host protein Rcr3), and are at least
44    restricted to the *Solanum* genus[18–24]. Arabidopsis RLP42 recognizes fungal
45    endopolygalacturonases (PG) eptitope pg9(At) derived from *Botrytis cinerea*[25,26]. Similarly, RLP23
46    is an Arabidopsis-specific LRR-RLP, which recognizes nlp20 peptide from the NECROSIS AND
47    ETHYLENE-INDUCING PEPTIDE1 (NEP1)-LIKE PROTEINS (NLPs) found in
48    bacterial/fungal/oomycete species[14,27]. ReMAX is restricted to the Brassicaceae and triggered by
49    the MAMP eMax originating from *Xanthomonas*[28]. Cuscuta Receptor1 (*CuRe1*) is specific to
50    *Solanum lycopersicum* and senses the peptide Crip21 which originates from parasitic plants of
51    the genus *Cuscuta*[29,30]. Finally, the inceptin receptor (INR) appears to be specific to the legume
52    tribe Phaseoleae and recognizes inceptin (In11), a HAMP found in the oral secretion of multiple
53    caterpillars[31–34]. Notably, all the above examples of LRR-RLPs are family-specific, restricted to
54    the Solanaceae (Cf-2, Cf-4, Cf-9 and CuRe1), Brassicaceae (RLP23, RLP42, ReMAX) or
55    Leguminosae (tribe Phaseoleae) (INR)[17,26]. However, despite clear signatures of lineage-specific
56    functions, specific evolutionary steps leading to novel PRR functions across multiple species have
57    not been described.

58    Mechanistic understanding of LRR-RLP sensing functions is also currently limited. No structural
59    data exists for any PAMP or HAMP sensing LRR-RLP[26,35]. Genetic experiments using truncated
60    or chimeric receptors have revealed subdomains essential for function of RLP23 and RLP42[26,35],
61    but whether similar regions mediate function for other LRR-RLPs is not clear. An alternative
62    approach, Ancestral Sequence Reconstruction (ASR), leverages the specificity of receptor
63    homologues for a certain PAMP/HAMP to study the emergence of a recognition function[36,37]. ASR
64    requires dense gene trees with clearly defined sets of functional and non-functional receptors.
65    Such data are currently unavailable for any LRR-RLP, partly due the lack of relevant high-quality
66    genomes in closely related non-model species. Moreover, the highly duplicate-rich nature of LRR-
67    RLP-encoding loci complicates gene annotation. However, long-read sequencing has increased
68    the quality of *de novo* assemblies, facilitating the annotation of receptor genes at complex
69    receptor loci[17,38,39]. Importantly, LRR-RLP recognition functions can generally be rapidly assessed
70    through expression in a model plant, wild tobacco (*Nicotiana benthamiana*)[26,31,35,40]. A
71    heterologous model allows rapid functional validation of potential LRR-RLP homologues, chimeric
72    receptors and statistically inferred ancestral sequences.

73    In this study, we use the legume-specific LRR-RLP INR as a model to perform dense species
74    phenotyping, comparative genomics, and functional validation, to associate gain and loss of
75    In11 response with evolution of the contiguous *INR* receptor locus. By leveraging both existing
76    high-quality assemblies and long-read sequencing in key legume species, we were able to
77    study the evolution of the *INR* locus. We show that In11 response is restricted to species in the
78    clade of the Phaseoloid legumes, which includes the agriculturally important subtribes
79    Phaseolinae, Glycininae and Cajaninae. The contiguous receptor locus which includes *INR*
80    is dynamic and predates the evolution of Phaseoloids, but the presence of an *INR* homologue
81    strictly corresponds with the recognition of In11. Finally, we used chimeric and ancestrally
82    reconstructed LRR-RLPs to gain insight into the key domains and amino acids (AA) involved in
83    In11 recognition. We present a conceptual model for the evolution of LRR-RLP function based
84    on patterns of INR variation in legumes.

## Results

### In11 perception is restricted to certain legume species within the Phaseoloid clade.

In11 response was previously thought to be restricted to a subtribe of the legume family, the Phaseolinae, which includes cowpea (*Vigna unguiculata*) and common bean (*Phaseolus vulgaris*) but excludes soybean (*Glycine max*)[31]. To understand emergence of INR function within legumes with a higher precision, we measured ethylene accumulation triggered by In11 as a defense marker across a set of twenty-two legume species of the NPAAA papilionoids (Fig. 1a Suppl. Table 1, Suppl. Table 2). Response to In11 was only observed in plant species within the monophyletic subclade of the Phaseoloids. Hence, phylogenetic evidence suggests a single origin of In11 response at the base of the Phaseoloid legumes ca 28 mya (Fig. 1b, ★). Several of the tested plant species/accessions within this clade were unable to respond to In11, namely *Hylodesmum podocarpum*, winged bean (*Psophocarpus tetragonolobus*), *G. max*, yam bean (*Pachyrhizus erosus*) and calopo (*Calopogonium mucunoides*), although they were able to respond to the unrelated bacterial elicitor flg22 (Suppl. Fig. 1). These observations suggest the occurrence of multiple independent losses of In11 response throughout Phaseoloid evolution after the initial emergence of this function.

### The contiguous *INR* locus shows LRR-RLP copy number variation in the Millettioids.

To explore gain and loss of In11 response in relation to its defined PRR, we analyzed the evolution of its cognate receptor *INR*, and its genomic locus, in existing reference genomes and high quality *de novo* genome assemblies of key Phaseoloid and non-Phaseoloid species. We focused *de novo* sequencing efforts on nodes separating In11-responsive and unresponsive species in the phylogeny for comparative genomic analysis (Fig. 1b). The contiguous *INR* locus was extracted from sixteen existing legume genomes and four *de novo* assemblies: *P. erosus*, guar bean (*Cyamopsis tetragonoloba*), jack bean (*Canavalia ensiformis*) and *H. podocarpum*) (Fig. 2, Suppl. Table 3)[31]. For *de novo* assemblies obtained in this study, a combination of Nanopore and Illumina sequence data enabled the assembly of a contiguous *INR* locus in all sequenced species (Suppl. Table 4).

The organization of the *INR* locus is highly diverse among legumes (Fig. 2a), with zero to seven LRR-RLP encoding-genes per species. For all Millettioid species, BLASTN search did not identify any regions with higher score than sequences at the contiguous *INR* locus of the respective species, strongly suggesting that all potential *INR* homologues are encoded at this locus. Legume species within the Hologalegina, *Lotus angustifolius*, chickpea (*Cicer arietinum*), *Medicago truncatula* and clover (*Trifolium pratense*), do not contain an LRR-RLP encoding gene at the locus, whereas all species within the Millettioids except for *P. erosus* contained at least one LRR-RLP, consistent with a gene insertion event of an LRR-RLP at the *INR* locus in the ancestor of extant Millettioids (Fig. 1a). To investigate this emergence, we analyzed the *de novo* assembly of *C. tetragonoloba,* a close outgroup of Millettioid legumes (Fig. 1b)[41,42]. As with all species within the outgroup Hologalegina, no LRR-RLP was present between the conserved neighboring genes (i.e. anchor genes), nor did we find an LRR-RLP with greater than 68% similarity in the whole genome, strengthening support for a single ancestral RLP gene insertion event ca 32 mya (Fig. 2a).

In contrast to other closely related legume species, BLASTN analysis of the contiguous *INR* locus of *P. erosus* revealed only partial coding sequence fragments with homology to LRR-RLPs, and consequently the absence of an *INR* homologue (Fig. 2b). The *de novo* assembly of the locus

129  was validated by performing PCR spanning the complete disruption and Sanger sequencing of
130  the resulting amplicon. The absence of an *INR* homologue in *P. erosus* corresponds with species
131  phenotype, namely the lack of induced ethylene response after In11 treatment (Fig. 1a).

132  **In11-induced functions are conferred by a single clade of LRR-RLPs (*INR* clade).**

133  To associate *INR* locus variation with the variable In11 responses (Fig. 1), we next investigated
134  the function and relationship of individual LRR-RLP homologues at the contiguous *INR* locus. We
135  performed a maximum-likelihood phylogenetic analysis on the protein sequences of the LRR-
136  RLPs within the locus across sixteen Millettioid species (Fig. 2). This analysis was supplemented
137  with the closest related LRR-RLP genes outside the contiguous *INR* locus from *V. unguiculata*
138  and *P. vulgaris*: *Phvul.007g246600* and *Vigun07g039700* (Suppl. File 2). A well-supported clade
139  which includes the previously characterized functional *INR* from *V. unguiculata* (*Vigun07g219600*)
140  also contained a single ortholog exclusively from plant species able to respond to In11 (Fig. 3)[31].
141  Hence, we hypothesized that this clade contains functional *INR* homologues which can confer
142  In11-induced functions. To validate the putative *INR* clade, five genes (*Vigun07g219600*,
143  *Phvul.007G077500*, *Mlathy INR*, *C.cajan_07316* and *Mprur INR*) within this clade were cloned
144  and transiently expressed in the non-legume model species *N. benthamiana*. In11-induced
145  ethylene and reactive oxygen species (ROS) production were able to be conferred by each gene
146  (Suppl. Fig. 3), consistent with a conserved INR function in the *INR* clade (Fig. 3, blue labels).

147  To assess if more distantly related, INR-like homologues could also confer In11 recognition, we
148  measured ethylene and ROS production upon expression of LRR-RLPs outside the *INR* clade.
149  Intriguingly, the sister clade to the *INR* clade contains LRR-RLPs from 9 out of 11 species which
150  also have a predicted *INR* homologue in the *INR* clade itself (Fig. 3). Within this clade, we cloned
151  and transiently expressed the LRR-RLP of *V. unguiculata* (*Vigun07g219700*). Except for *P.*
152  *vulgaris* G19833, all studied Phaseolinae have an LRR-RLP in the two clades discussed above
153  (Fig. 2). The remaining species, *G. max, Glycine soja, H. podocarpum, C. ensiformis* and *Abrus*
154  *precatorius*, were In11-unresponsive (Fig. 1) and solely encode LRR-RLP receptors which fall
155  outside the *INR* clade and its sister clade. From this group we cloned and transiently expressed
156  the soybean LRR-RLP (*Glyma.10G228000*). Finally, we tested the most closely related LRR-RLP
157  to *INR* outside the contiguous *INR* locus for both *V. unguiculata* and *P. vulgaris* (*Vigun07g039700*
158  and *Phvul.007g246600*). No In11-induced responses could be observed upon heterologous
159  expression of *Vigun07g219700, Glyma.10G228000*, *Vigun07g039700* and *Phvul.007g246600*
160  (Fig. 3 and Suppl. Fig. 3). Proteins of these non-responsive LRR-RLPs were similarly or more
161  strongly expressed in *N. benthamiana* relative to the lowest expressed responsive LRR-RLP
162  (C.cajan_07316 INR) (Suppl. Fig. 4A), except for the marginally lower expressed
163  Vigun07g039700, which contrasts with its closest related gene in *P. vulgaris* (*Phvul.007g246600*,
164  87% AA similarity). Thus, the ability to confer In11-induced ROS and ethylene production is strictly
165  limited to members of the *INR* clade.

166  We identified and dated potential gene duplications and losses at the contiguous *INR* locus
167  throughout its evolution within the Millettioids using NOTUNG (Suppl. Fig. 2)[43]. Reconciliation
168  analysis between gene and species trees revealed a complex evolutionary history comprising 19
169  gene duplications and 20 gene losses in total. Within the Millettioids, a single duplication event
170  gave rise to the *INR* clade containing all *INR* homologues and its sister clade which contains the
171  closest related *INR*-like homologues. *P. vulgaris* and *Mucuna pruriens* contain an *INR* clade
172  homologue but not members of the sister *INR*-like 1 clade, consistent with two independent gene
173  losses. In contrast, the analysis predicts the ancestral loss of an *INR* homologue within the *INR*

174 clade for *G. soya*, *G. max* and *H. podocarpum* and clade specific duplication events resulting in
175 LRR-RLP expansions of 4-7 tandem duplicates. To confirm that *INR* was lost in *Glycine* and not
176 just in reference assemblies, we performed BLASTP searches using Vigun07g219700 AA
177 sequence against 26 *G. soja* and *G. max de novo* assemblies from a recent pangenome
178 analysis[44]. Like the *Glycine* reference genomes, no BLASTP hits to Vu07g219600 for any of the
179 26 de novo assemblies were identified that had an AA similarity higher than 76%, suggesting that
180 INR was lost before the speciation of *G. max* and *G. soja*, *i.e.* prior to soybean domestication.

## The C1 and C2 subdomain of the LRR ectodomain mediate In11-induced functions

182 To understand receptor subdomains contributing to the functional INR clade, we assembled nine
183 chimeric receptors combining *Vigun07g219600* (*Vu*INR hereafter) and paralogous
184 *Vigun07g219700* (*Vu*INR-like, 72% AA similarity) (Suppl. File 3). Both genes contain a typical
185 LRR-RLP extracellular domain with 29 LRRs interrupted by a 14-AA intervening motif (C2 domain)
186 (Fig. 4)[31]. Chimeric receptors were expressed at a similar level in *N. benthamiana* and their
187 response to In11 was assessed by quantifying peptide-induced ethylene and ROS (Fig. 4, Suppl.
188 Fig. 4b). Chimeric receptors with the *Vu*INR-like C1 domain were not responsive to In11. In
189 contrast, all chimeric receptors containing the C1-C2 of *Vu*INR responded to In11 treatment with
190 both an ethylene and ROS burst (219600-F, 219600-C3, 219600-C2 and C1-219600, Fig. 4a).
191 Intriguingly, the chimeric receptor 219600−C1, containing C1 of *Vu*INR and C2 of *Vu*INR-like,
192 resulted in In11-independent autoactivity. Additionally, the chimeric receptor 219600−C2,
193 responds to In11 but has a delayed ROS burst relative to all other In11-responsive constructs
194 tested. In summary, the *Vu*INR LRR subdomains C1 and C2 are both required for In11 response
195 in chimeric receptors.

## Ancestral sequence reconstruction (ASR) of the *INR* LRR domain

197 To further understand the molecular basis for functional divergence between INR and
198 unresponsive INR-like homologues, we predicted multiple ancestral sequences of the
199 monophyletic In11-responsive *INR* clade and its non-responsive *INR*-like 1 sister clade (Fig.5,
200 Suppl. Fig.5, Suppl. File 4). To perform ASR, we first confirmed that nodes of interest for
201 reconstruction were well supported by both neighbor-joining (NJ) and maximum likelihood (ML)
202 gene phylogenies (Suppl. Fig. 5). Subsequently, we reconstructed the ancestral sequences for
203 the LRR domain using FastML (Suppl. File 4)[45], synthesized their predicted sequences and
204 ligated the resulting LRR domains with flanking domains (A-B and D-G) from *Vu*INR to complete
205 the ancestral receptor constructs (Fig. 5b). Protein expression in *N. benthamiana* was similar
206 across all ASR variants (Suppl. Fig. 4c).

207 Ancestral receptors to the *INR* clade of the *Vigna, Phaseolus* and *Macroptylium* (N7), the
208 Phaseolinae (N6) and the Phaseoloids (N4) conferred In11-induced ethylene and ROS response.
209 In contrast, ancestral receptors of the *INR*-like 1 clade of the Phaseolinae (N16) and the
210 Phaseoloids (N14) were not responsive to In11 (Fig. 5d, Fig. 5c). Moreover, the common
211 ancestral receptor of the *INR* and its *INR*-like 1 sister clade (N3) is not responsive to In11 (Fig.
212 5). Hence, a small number of differences, 16 of 720 AA, between the LRR domain of N3 and N4,
213 mediate differential In11 response. Moreover, only a subset of those AA is conserved within all
214 LRR-RLPs of the *INR* clade and absent in all LRR-RLPs of the *INR*-like 1 sister clade (Suppl. Fig.
215 6).

## Discussion

Here we described the evolution of the legume-specific LRR-RLP INR as a functional and comparative genomic model to study A) responsiveness of an LRR-RLP against a defined elicitor and B) the evolution of novel PRR functions. We identified plant species which contain or lack a functional INR homologue, corresponding with response to the In11 HAMP elicitor. We present a conceptual model for the evolution of a novel LRR-RLP function based on patterns of *INR* locus variation in legumes. Finally, comparisons between In11-responsive and unresponsive homologues, especially through chimeric receptors and ASR, resolve potential key AA residues which mediate peptide recognition and response. Our work illuminates themes in the evolution of lineage-specific immune sensing functions, which will inform the broad use of PRRs as resistance traits[46–49].

Evolutionary analysis of diverse plant phenotypic responses to a single PAMP or HAMP is rarely performed, although this can be a powerful approach to understand the emergence of specific immune receptor functions[2]. Patterns of In11 responses across twenty-two legume species indicated that In11 response is restricted to the Phaseoloids (Fig. 1). However, within the Phaseoloids, several tested species/accessions were not able to respond to In11, suggesting multiple independent losses of INR. These phenotypic observations allowed the investigation of the evolution of a specific LRR-RLP involved in plant immunity over a long (32my), high-resolution timescale.

We complemented broad phenotypic analysis with comparative genomics of key legume species. Analysis of the *INR* receptor locus across twenty existing and newly sequenced legume genomes revealed that INR function followed an early gene insertion and diversification at the *INR* locus. Importantly, *de novo,* long-read based assemblies presented here contained contiguous *INR* loci flanked by conserved anchor genes (Fig. 2), enabling strong conclusions with respect to receptor gene content. Comparative genomic analysis can reveal complex histories of gene insertions, duplication and losses for LRR-RKs/RLPs[50,51]. The *INR* locus shows high variability and diversification among legumes of the NPAAA papilionoids, with zero to seven LRR-RLP encoding-genes per species. Legume species within the Hologalegina (*L. angustifolius*, *C. arietinum*, *M. truncatula* and *T. pratense*) do not contain an LRR-RLP encoding gene at the locus, whereas all species investigated herein the Millettioids except for *P. erosus* contain at least one LRR-RLP (Fig. 2). This observation is consistent with a gene insertion event at the *INR* locus in the ancestor of extant Millettioids ca 32 mya.

Our analysis of INR provides an example of an LRR-RLP gene family with conserved recognition function across multiple plant genera which evolved ~28 mya. A phylogenetic analysis revealed potential INR-homologues in ten additional species as they clustered together with *V. unguiculata* INR. Orthologous LRR-RLPs of the *INR* clade from five legume genera were tested and able to induce an In11-induced ethylene and ROS burst upon heterologous expression in *N. benthamiana* (Fig.3, Suppl. Fig. 3). This includes INR homologues from *Cajanus cajan* and *M. pruriens* of the Phaseoloid clade. Certain Phaseoloids and earlier diverged non-Phaseoloids also contain INR-like homologues at the contiguous *INR* locus; thus, INR most likely arose from an existing LRR-RLP, ~28 mya ago. This evolutionary pattern contrasts with the evolution of the *Solanum pimpinellifolium* specific LRR-RLP Cf-2 which evolved <6 mya, potentially by intergenic recombination[49]. Our analysis thus identifies an alternative mechanism whereby an ancestral gene insertion event is the likely source of extant gene and copy number variation, preceding the evolution of a specific peptide recognition function.

261  The abundance of closely related legume genomes facilitates gene-species tree reconciliation,
262  suggesting mechanisms contributing to the dynamic nature of *INR* receptor loci[43,50]. Interestingly,
263  we observed three independent cases of *INR* loss within the Phaseoloids. Two reference
264  genomes as well as 26 *de novo* assemblies of the closely related Glycininae, *G. soya* and *G.*
265  *max,* do not encode *INR*. Nevertheless, *INR* loss does not seem to predate the divergence of the
266  Glycininae as *Teramnus labialis* is able to respond to In11 (Fig. 1), although genome data are
267  lacking for this species. Besides reciprocal gene loss of *INR*, gene tree-species tree reconciliation
268  also suggests the involvement of five tandem duplications of an *INR*-like gene predating radiation
269  of the genus *Glycine* (Suppl. Fig. 2). This is consistent with whole-genome observations of
270  preferential gene loss in tandem clusters[50]. Similarly, within the separate Desmodieae lineage,
271  our analysis also predicts that a *H. podocarpum INR*-like gene underwent specific tandem
272  duplication events after the loss of *INR*, as its genome contains seven *INR*-like homologues which
273  cluster together in a phylogenetic analysis. In contrast, *P. erosus* seems to have lost both *INR*
274  and *INR*-like as it does not contain any complete coding sequence of an LRR-RLP at the *INR*
275  locus. Besides *P. erosus*, *P. vulgaris* and *M. pruriens* also lost *INR*-like. Loss of *INR* (and *INR*-
276  like) may reflect the propensity of tandemly duplicated loci to lose functions through gene
277  conversion, and/or reciprocal gene loss[50].

278  Our high-resolution analysis of the emergence of INR provides a roadmap for understanding other
279  lineage-specific PRRs to detect PAMPs. While most characterized LRR-RLPs are family-
280  specific[17,26], extensive phenotyping of an entire plant family has not yet been conducted for well-
281  studied PAMPs such as elf18, nlp20, pg9, crip21, csp22, and xup25. Understanding the
282  emergence of known LRR-RK/RLPs for these PAMPs, namely EFR, RLP23, RLP42, CORE,
283  CuRE1, and XPS1 will illuminate broader mechanisms underlying the evolution of key immune
284  receptor modules in plants[26,27,29,52,53]. LRR-RLPs often occur in complex loci with variation within
285  and between species [17,26]. However, a recent pangenomic analysis of Arabidopsis indicates that
286  several PAMP sensing LRR-RLPs occur in relatively simple and conserved loci across *A. thaliana*
287  varieties[16]. It will be interesting to see if Arabidopsis PRR functions evolved along a similar
288  trajectory to INR, for example via ancestral duplications preceding fixation in the Arabidopsis
289  lineage. In summary, additional case studies for specific receptors are needed to reveal broader
290  patterns in receptor evolution.

291  Our functional analysis of INR and INR-like genes in a heterologous model (*N. benthamiana*) also
292  provides a system to study LRR-RLP function. Compared to well-studied LRR-RKs such as
293  FLAGELLIN-SENSING 2 (FLS2)[54–56], mechanisms underlying LRR-RLP function are not well
294  understood. Despite multiple intense attempts, and in contrast to LRR-RKs[10,56], structural
295  information for ligand-binding LRR-RLPs is not available, including INR[26,35]. Moreover, Wang et
296  al. 2019 contrasted signaling and defense responses activated by FLS2 and RLP23, but broader
297  similarities and differences between RK and RLP families remain unclear[35,57]. Consequently, we
298  have limited mechanistic insight into LRR-RLP function as elicitor-specific PRRs[26,35].

299  Chimeric receptors formed by combining *INR* and an *INR*-like paralogue revealed RLP
300  subdomains required for In11 response. Across all chimeric receptors, those encoding the LRR
301  (C1) and intervening motif (C2) domains of *Vu*INR responded to In11 treatment with both an
302  ethylene and ROS burst, suggestive of crucial elements for elicitor interaction in the C1-C2
303  subdomain. Notably, a chimeric receptor with mixed C1 and C2 (intervening motif), 219600-C1,
304  resulted in In11-independent autoactivity, consistent with a critical role for C1-C2 compatibility. In
305  addition, a chimeric receptor with mixed C2 and C3 (219600-C2) had delayed In11-induced ROS

306 burst relative to all other In11-responsive constructs tested here (Fig. 4). These findings are
307 consistent with previous truncation and chimeric protein analyses for Arabidopsis RLP23 and
308 RLP42. Truncations of the RLP23 ectodomain abolished function, suggesting the necessity of the
309 entire ectodomain for elicitor binding or proper assembly of the receptor[35]. Additionally, chimeric
310 receptors implicated the importance of RLP42 its twelve N-terminal LRRs (C1) and LRR21-LRR24
311 which includes the island subdomain (C2) for recognition of fungal endopolygalacturonases[26].

312 Additionally, our chimeric receptor data are consistent with a vestigial role for the cytoplasmic tail
313 of PRRs in the LRR-RLP family. Strikingly, all identified INR homologues encode a cytoplasmic
314 tail of only 10AA, shorter relative to all identified INR-like homologues here. Nevertheless,
315 swapping the *Vu*INR (Vigun07g219600) cytoplasmic tail to the extended version of *Vu*INR-like 1
316 (Vigun07g219700), 219600-F, did not affect In11-response (Fig. 4). Previously, the complete
317 deletion of the intracellular 17-amino-acid tail of RLP23 reduced but did not abolish receptor
318 function[35]. Additionally, a previously described chimeric swap replacing the RLP42 terminal LRR,
319 transmembrane helix and cytoplasmic tail with the respective subdomains of a non-responsive
320 paralogue was still responsive to the pg9(At) elicitor[26]. Intriguingly, quantitative differences were
321 earlier reported as the C-terminally truncated RLP23 had a reduced nlp20 response, consistent
322 with auxiliary rather than essential function for the RLP23 cytoplasmic tail[35].

323 As an alternative to the use of chimeric receptors, ASR of the *INR* LRR domain revealed additional
324 detailed insights contributing to INR function. To trace the potential evolutionary history of INR,
325 we used ASR by leveraging *INR* and *INR*-like homologues of sixteen legume species of the
326 Millettioids (Suppl. Fig. 5). As the chimeric receptors implicated the LRR (C1-3) ectodomain in
327 In11 recognition, we reconstructed ancestral LRR variants in an *INR* backbone. Our analysis
328 suggests that the common ancestral LRR ectodomain of extant INRs conferred In11 response
329 (N4), whereas the common ancestor of both INR and INR-like (N3) did not confer In11 response
330 (Fig. 5). N3 and N4 differ in In11 response although they only vary by 16 AA, with only some of
331 them conserved in all extant INRs. The role of key LRR-RLP residues in ligand-specific responses
332 is consistent with the previous analysis of Arabidopsis RLP42. Zhang et al. 2021 introduced single
333 AA substitutions to the functional RLP42 receptor, and several were sufficient to abolish pg9(At)-
334 induced co-receptor association and defense responses.

335 A similar ASR approach to ours was employed to understand the 98-AA heavy metal-associated
336 (HMA) domain of the plant intracellular NOD-like receptor Pik-1[37]. Specific AA changes introduced
337 into an ancestrally reconstructed backbone were sufficient to confer effector binding and immune
338 functions. For Pik-1 HMA, structural data provided additional insight into the mechanism of
339 effector binding in ancestral and extant proteins. Our use of an ASR approach for a relatively long
340 LRR domain (720 AA) now demonstrates the power of dense comparative genomic analyses to
341 also identify key residues in extant PRRs without defined binding sites. In the absence of structural
342 data for ligand-binding LRR-RLPs, an ASR approach may be useful to identify sets of co-varying
343 residues critical for binding and signaling functions.

## Materials and Methods

### Plant materials

346 Plant species and accessions used in this study are listed in Suppl. Table 1, as well as their
347 respective providers: Phil Miklas, (US Department of Agriculture, Prosser, WA), Creighton Miller
348 (Texas A&M University, College Station, TX), Phil Roberts (UC Riverside, CA), Timothy Close

349  (UC Riverside, CA) and the USDA Germplasm resources Information network (GRIN). Plants
350  were grown in the greenhouse (25°C/21°C day/night, 60%RH and 12:12 light:dark cycles) or in
351  growth chambers (26°C/26°C day/night, 70%RH and 12:12 light:dark cycles).

## Peptide-induced ethylene production in legume species.

353  The In11 peptide (ICDINGVCVDA) is a host derived proteolytic fragment of the ATP synthase γ-
354  subunit (cATPC), based on the *V. unguiculata* cATPC sequence[32]. The flg22 peptide
355  (QRLSTGSRINSAKDDAAGLQIA) originates from bacterial flagellin[58]. Both peptides were
356  synthesized (Genscript) and reconstituted in $H_2O$. Leaflets were lightly scratch wounded with a
357  fresh razor blade to remove cuticle area, and 10 μL of $H_2O$ with or without peptide (1 μM) was
358  equally spread over the wounds with a pipette tip. After 1h, leaflets were excised and placed in
359  sealed tubes for 2h before headspace sampling (1 mL). Ethylene was measured as previously
360  described with a gas chromatographer (HP 5890 series 2, supelco #13018-U, 80/100 Hayesep Q
361  3FT x 1/8IN x 2.1MM nickel) with flame ionization detection and quantified using a standard curve
362  (Scott, 99.5% ethylene, (Cat. No 25881-U)) (Suppl. Table 2)[59]. Subsequently, R (v4.0.3) and the
363  R-packages dplyr (v1.0.7), ggpubr (v.0.4.0) and ggplot2 (v3.3.3) were used to analyze and plot
364  the data, statistics were performed by using the paired Wilcoxon signed-rank test[60–62]. The
365  resulting figure was edited in Corel-DRAW Home & Student x7.

## ROS production in legumes

367  Leaf punches were taken with a 4-mm biopsy punch and floated in 150 μL of $H_2O$ using individual
368  cells of a white 96-well white bottom plate (BRAND*plates* F pureGrade S white (REF 781665)).
369  After overnight incubation, ROS production was measured upon addition of a 100 μL assay
370  solution which contains 10μg/ml luminol-horseradish peroxidase (HRP), 17μg/ml luminol and the
371  treatment (2.5 μM In11, 2.5 μM flg22 or $H_2O$). Luminescence was quantified with a TECAN
372  SPARK plate reader every minute for an hour using an integration time of 500ms. Four technical
373  replicates were quantified for each treatment and significant differences were determined by
374  performing a 2-group Mann-Whitney U Test between both treatments. R and the R-packages
375  dplyr (v1.0.7), ggpubr (v.0.4.0) and ggplot2 (v3.3.3) were used to analyze and plot the data. The
376  resulting figure was edited in Corel-DRAW Home & Student x7.

## Genome sequencing and assembly of legume species

378  Leaf tissue of legume plants was harvested and grounded using an $N_2$-chilled mortar and pestle.
379  In contrast to *P. erosus,* nuclei isolation was first performed on frozen tissue for *C. ensiformis*, *C.*
380  *tetragonoloba* and *H. podocarpum* using the Bionano Plant Tissue Homogenization Buffer (Part
381  number 20283**).** Nuclear DNA was extracted for all above mentioned species with a modified
382  CTAB protocol as described previously[63]. Resultant High-Molecular Weight (HMW) DNA
383  concentrations were determined by Qubit and Bioanalyzer.
384
385  A modified protocol using the Oxford Nanopore Technologies (ONT) Rapid barcoding kit (SQK-
386  RBK004) was used for sequencing. Briefly, 27ul of ~20 ng/μL HMW DNA was combined with 3ul
387  of Rapid Barcoding fragmentation mix and incubated for 1min at 30°C followed by 1 min at 80°C.
388  Ampure beads were added at a 0.7x final concentration and bead clean-ups performed as
389  described in the Rapid barcoding kit (SQK-RBK004). All remaining steps were performed as
390  described in the Rapid barcoding kit protocol. Each sample was sequenced on a single MinION
391  or PromethION flowcell (R9.4). High-accuracy base calling was performed in real time with
392  MinKnow (v20.10.6).

393  Illumina short read sequence was generated for *C. tetragonoloba* and *H. podocarpum* from the
394  same HMW DNA used for ONT long read sequencing. Paired end 2x150 bp sequence was
395  generated on the Illumina NovaSeq6000 platform. In addition, we generated Illumina short reads
396  for *C. ensiformis, P. erosus* and *M. lathyroides*, with GENEWIZ. For these samples, genomic DNA
397  was extracted using the Nucleospin Plant II kit (Macherey-Nagel). The oxford nanopore
398  sequencing data and Illumina hiseq reads used in this study can be found in SRA under
399  Bioproject: PRJNA820752. The final genome assemblies are available on NCBI (*P. erosus*:
400  SUB11200654, *C. ensiformis*: SUB11200580, *H. podocarpum*: SUB11200652 and *C.
401  tetragonoloba*: SUB11200669).

402  The *P. erosus* genome was assembled using SPAdes (v3.15.4) using the –meta option with both
403  the Illumina and Nanopore readsets. Additional genome assemblies were produced with FlyE
404  (v2.8.1), consensus was generated with three rounds of Racon (v1.3.1), and finally polished with
405  Pilon (v1.22) three times with Illumina reads (2x150 bp). Final assembly quality was determined
406  with assembly-stats and BUSCO (v5.3.0).

## Analysis of the contiguous *INR* locus and LRR-RLP homologues

408  The analyzed genome assemblies, versions and their sources for the twenty legume species
409  included in the contiguous *INR* locus analysis can be found in Suppl. Table 3[64–71]. The nucleotide
410  sequences of the extracted loci and their coordinates can be found respectively in Suppl. File 1
411  and Suppl. Table 3. All INR and INR-like AA sequences included in the phylogenetic analysis can
412  be found in Suppl. File 2. Newly sequenced and assembled genomes were analyzed to define
413  the contiguous *INR* locus, *INR* and *INR*-like sequences. Similarly, if not yet annotated in publicly
414  available genomes, syntenic *INR* and *INR-like* homologues were identified. First, BLASTN
415  (BLAST 2.9.0+, e-value 10) was used to identify the *INR* syntenic locus by mining the genomes
416  for homologues of the strongly conserved neighbor (anchor) genes of common bean *INR*
417  (Phvul.007g077500); Phvul.007g077400 and Phvul.007g077600. Similarly, the genomes were
418  mined for LRR-RLPs via BLASTN approach with a default e-value of 10, with exception of the
419  *Glycine* pangenome where a BLASTP approach was used. The strongest LRR-RLP blast hits
420  with *INR* were consistently identified in between the conserved neighbor genes. Finally, potential
421  *INR* and *INR*-like ORFs were determined by visual inspection in IGV (v2.10.3) and compared with
422  the sequences of closely related annotated LRR-RLPs in other legume species. Additionally, *INR*
423  homologues were identified in *Macroptylium lathyroides* by using an alternative method. First, *M.
424  lathyroides* Illumina HiSeq paired end reads were mapped against both *Phvul.007G077500* (*INR*)
425  and *PvUI111.07G078600* (*INR*-like) by using bwa (v0.7.17-r1188)[72]. Second, mapped reads were
426  sorted and indexed by using samtools (v 1.13)[73]. Third, IGV (v2.10.3) was used to inspect the
427  mapped reads and identify the SNPs of Mlathy *INR* and *INR*-like in comparison to the *P. vulgaris*
428  homologues. Fourth, the above three steps were reiterated with the newly acquired gene
429  sequences. Finally, the acquired Mlathy *INR* and *INR*-like sequences were confirmed by PCR
430  using the Q5 Hot Start High-Fidelity kit (NEB), enzymatic clean-up using ExoSAP-IT™ (Thermo
431  Fisher scientific) and Sanger sequencing of the entire amplified PCR product (GENEWIZ).
432  Primers used in this process are listed in Suppl. Table 5.

## Contiguous *INR* locus analysis and validation of *P. erosus* receptor disruption

434  Locus comparison was performed using R (v4.0.3) and the R-package genoPlotR (v0.8.11) using
435  the extracted contiguous *INR* loci and their corresponding annotation (Suppl. File 1). The resulting
436  figure was edited in Corel-DRAW Home & Student x7. A PCR was performed to validate the
437  disruption of the receptor at the syntenic locus of *P. erosus* using the Q5 Hot Start High-Fidelity

438    kit (NEB). Primers used in the reaction are listed in Suppl. Table 5. Subsequently, the amplified
439    PCR product was enzymatically cleaned-up using ExoSAP-IT™ (Thermo Fisher scientific).
440    Finally, the disruption was validated by Sanger sequencing of the entire amplicon (GENEWIZ).

## Phylogenetic analysis of *INR* and *INR-like* homologues

442    Sequences were aligned using the online version of MAFFT 7 using the E-INS-i strategy[74]. One
443    potential pseudogene of *H. podocarpum* was not incorporated in the phylogenetic analysis since
444    the length of the sequence was about 79% of cowpea *INR* due to the absence of a part of the
445    LRR domain in comparison to all other annotated receptors at the contiguous *INR* locus. A
446    phylogenetic analysis was performed on the CIPRES web portal using RAXML-HPC2 on XSEDE
447    (v8.2.12) with the automatic protein model assignment algorithm using maximum likelihood
448    criterion and 250 bootstrap replicates[75,76]. The DUMMY2 protein model was selected as the best
449    scoring model for maximum likelihood analysis. The resulting tree was rooted, visualized using
450    MEGA10 and edited in Corel-DRAW Home & Student x7.

## Notung analysis: prediction of duplication and gene loss events at the contiguous *INR* locus

453    First, a phylogenetic analysis was performed similar to the approach above with the sole
454    difference that only *INR* and *INR*-like homologues located at the contiguous *INR* locus were
455    included. Second, a species tree in Newick format was built which includes all species of which
456    1) *INR* and *INR*-like homologues were extracted, 2) *C. tetragonoloba* as it is the closest related
457    legume species without a receptor at the contiguous *INR* locus, and 3) *M. truncatula* as the LRR-
458    RLP used as an outgroup in the analysis was extracted from its genome. Third, the NOTUNG
459    analysis was performed using the default options, including a duplication cost of 1.5 and a loss
460    cost of 1 (v2.9.1.5)[43]. NOTUNG ignores incomplete lineage sorting as an evolutionary mechanism
461    when both a rooted species and gene tree are used as input, as was the case for the present
462    study.

## Molecular cloning of *INR* and *INR*-like homologues

464    Leaf tissue of legume plants was harvested and grounded using an $N_2$-chilled mortar and pestle.
465    Genomic DNA was extracted using the Nucleospin Plant II kit (Macherey-Nagel). RNA was
466    extracted using the Nucleospin RNA Plant kit (Macherey-Nagel). cDNA was created using
467    SuperScript™ IV Reverse Transcriptase (Invitrogen). All constructs were created using a
468    hierarchical modular cloning approach facilitated by the MoClo toolkit and the MoClo Plant Parts
469    kit[77,78]. LRR-RLPs with no introns were PCR amplified from genomic DNA (Q5 Hot Start High-
470    Fidelity, NEB), all others (*Phvul.007g246600* and *Vigun07g039700*) were amplified from cDNA.
471    All primers used for amplification are listed in Suppl. Table 5. All amplified PCR fragments were
472    gel extracted using the PureLink™ Quick Gel Extraction Kit (Thermo Fisher scientific) and purified
473    and concentrated using the Monarch PCR & DNA Cleanup Kit. Subsequently, the PCR fragments
474    were ligated in an L-1 acceptor vector. A second digestion/ligation step was completed, to ligate
475    multiple parts together and complete the CDS while inserting it in an L0 acceptor vector.
476    Throughout the previous steps, recognition sites for BsaI and/or BpiI were removed from the CDS.
477    All constructs were validated by Sanger sequencing upon completion. Finally, all constructs were
478    completed by adding the following MoClo modules; 35s Cauliflower Mosaic Virus + 5'UTR
479    Tobacco mosaic virus (pICH51266), GFP (*A. victoria*) (pICSL50008) and the OCS1 terminator
480    (pICH41432)[78].

### *N. benthamiana* transient expression and Western blotting

Constructs were electroporated into *Agrobacterium tumefaciens* strain GV3101(pMP90,pSOUP). Overnight cultures were resuspended in 150 µm acetosyringone in 10 mM 2-(*N*-morpholino) ethanesulfonic acid (MES), pH 5.6, 10 mM $MgCl_2$. After 3h of incubation at room temperature, *N. benthamiana* leaves of 5-week-old plants were infiltrated at an optical density of 0.45 at 600 nm ($OD_{600}$).

Leaf punches were taken 48h after infiltration of the *N. benthamiana* leaves to validate the expression of the constructs in *N. benthamiana* by Western blot (Suppl. Fig. 4). Ground, frozen tissue was homogenized in a 3x lamellae buffer (50 mM Tris-Cl pH 6.8, 6% SDS, 30% glycerol, 16% β-mercaptoethanol and 0.006% Bromophenol blue) and then cleared by centrifugation (10 m, 20,000 rcf). Subsequently, proteins in the supernatant were separated by performing an SDS-PAGE on an 8% acrylamide gel. Finally, a Western blot was performed to visualize the GFP-tagged heterologously expressed proteins and actin as a loading control with respectively an α-GFP polyclonal (A-6455; Thermo) primary antibody at a 1:2,000 dilution and an anti-Actin (ab197345, abcam) primary antibody at a 1:5,000 dilution. α-rabbit (A6154; Sigma) was used as a secondary antibody for both at 1:10,000 dilution.

### ROS measurements in *N. benthamiana*

Following *Agrobacterium* infiltration for receptor expression (24 h), leaf punches were taken with a 4-mm biopsy punch and floated in 150 µL of $H_2O$ using individual cells of a white 96-well white bottom plate (BRAND*plates* F pureGrade S white). Subsequently, the same procedure was followed as outlined for ROS production in legumes. Four biological replicates were quantified (n=4 plants), with each biological replicate representing six technical replicates. R and the R-packages dplyr (v1.0.7), ggpubr (v.0.4.0) and ggplot2 (v3.3.3) were used to analyze and plot the data. The resulting figure was edited in Corel-DRAW Home & Student x7.

### Ethylene in *N. benthamiana*

For ethylene assays in *N. benthamiana*, a fully expanded leaf of 5wk-old plants was infiltrated with $H_2O$ or 1µM In11 with a blunt syringe. Subsequently, four leaf discs within the infiltrated area were immediately excised with a no. 5 cork borer and sealed in tubes[31]. Headspace ethylene was measured after 3 h of accumulation as described above. Subsequently, R and the R-packages dplyr (v1.0.7), ggpubr (v.0.4.0) and ggplot2 (v3.3.3) were used to plot the data and perform the statistics (paired Wilcoxon signed-rank test). The resulting figure was edited in Corel-DRAW Home & Student x7.

### Construction of chimeric receptors.

Chimeric LRR-RLP constructs were generated using the MoClo toolkit[77]. MoClo overhangs were designed in such a way that specific fragments amplified from the L0 constructs of *Vigun07g219600* and *Vigun07g219700* could be ligated in the preferred direction and order in an L0 vector. Primers used in these reactions are listed in Suppl. Table 5. All constructs were validated by Sanger sequencing upon completion. Subsequently, CDS stored in L0 universal acceptors were combined with the same MoClo modules as the earlier described LRR-RLP constructs mentioned above to create a complete L1 construct. Transient expression of chimeric receptor constructs in *N. benthamiana* was validated with western blot.

## ASR of LRR domain

523 The LRR (C1-3 domain, Fig. 4) were extracted from all LRR-RLP sequences included in the earlier
524 mentioned phylogenetic analysis (Fig. 3, Suppl. File 4). The LRR domain and subdomains of
525 *Vu*lNR were earlier identified using LRRfinder[31,79]. Phylogenetic trees were built using MEGA X
526 software[80], and bootstrap method based on 1000 iterations. A codon-based 2172-nucleotide-long
527 alignment was generated using MUSCLE[81]. NJ clustering method was used for constructing the
528 codon-based tree on Maximum Composite Likelihood substitution models. The ML tree was
529 calculated using the GTR+G+I submodel as implemented in MEGA X software[80]. The resulting
530 ML tree was used for the ASR of selected nodes of interest. Joint and marginal ASR were
531 performed with FastML software using Jukes and Cantor substitution model for nucleotides,
532 gamma distribution, and 90% probability cutoff[82]. Finally, the sequences were domesticated to
533 facilitate MoClo cloning by removing the BsaI and BpiI cut sites (Suppl. File 4).

## Construction of receptors with an ancestral reconstructed LRR domain

535 Ancestral reconstructed LRR-RLP constructs were generated using the MoClo toolkit[77]. MoClo
536 overhangs were designed in such a way that the AB domain and the DEFG domain amplified
537 from *Vigun07g219600* could be ligated with the synthesized ancestral reconstructed LRR domain
538 (C domain), resulting in an L0 vector. Primers used in these reactions are listed in Suppl. Table
539 5. All constructs were validated by Sanger sequencing upon completion. Subsequently, the
540 complete CDS stored in L0 acceptors were combined with the same MoClo modules as the earlier
541 described LRR-RLP constructs mentioned above to create a complete L1 construct. Transient
542 expression of the ancestral reconstructed LRR-RLPs in *N. benthamiana* was validated with
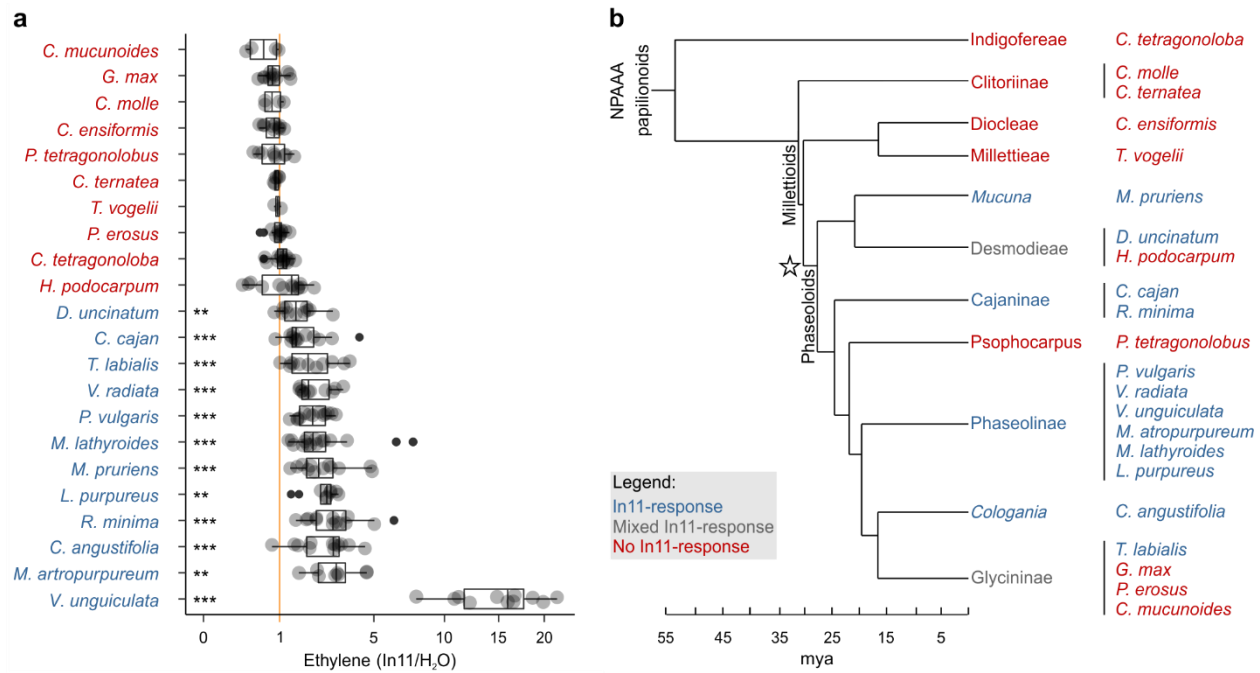543 western blot.

## Acknowledgments

## Contributions

550 S.S. and A.D.S. conceived and designed the experiments; S.S., B.W.A and A.G.K.G. conducted
551 experiments; S.S, B.W.A., A.N.E and A.D.S. analyzed data; S.S. prepared all figures; S.S. and
552 A.D.S. wrote the manuscript. All authors discussed the results and commented on the
553 manuscript.

## Main figures



**Fig. 1: Induced ethylene response to In11 is limited to Phaseoloid legumes. A)** Individual trifoliate leaflets were scratch wounded and treated with 1 µM In11 or $H_2O$. The ratio of ethylene production for leaflets within the same leaf is shown (x-axis). The vertical orange line shows a ratio equal to one, *i.e.* no In11-induced ethylene burst. Biologically replicated plants are shown as separate dots. Significant differences between the control and the treatment of interest are indicated (paired Wilcoxon signed-rank test; ns non-significant, * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$). Plant species names are colored blue (significant response to In11) or red (insignificant response). Species/accessions and resulting response data can respectively be found in Suppl. Table 1 and Suppl. Table 2. **B)** A summary chronogram representing time-based phylogenetic relationships within the tested lineages, with colors representing In11 response phenotypes as in (A). The star symbol indicates the node containing all In11-responsive species at the base of Phaseoloid legumes. Divergence time is shown in million years ago (mya) and represents a composite average[83–86].
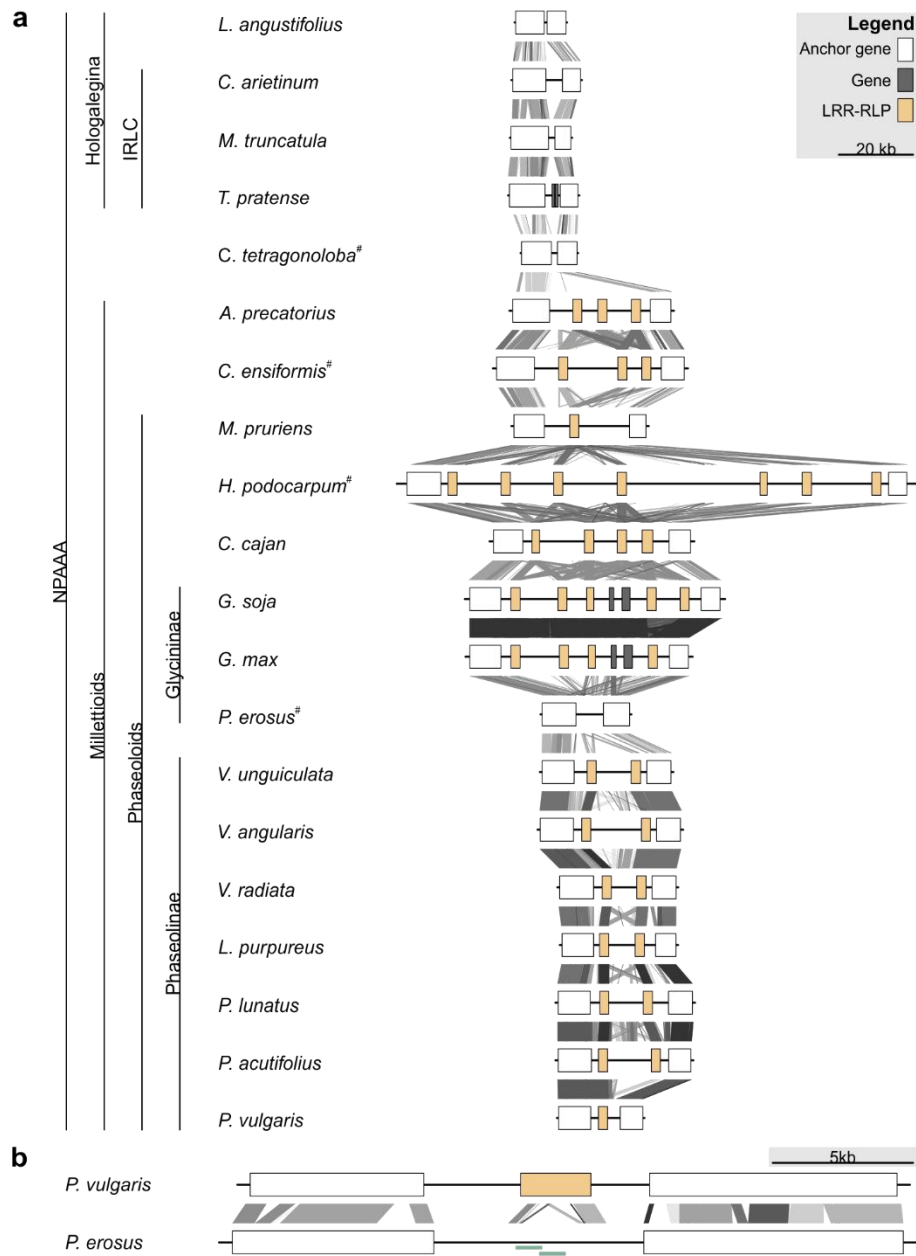
**Fig. 2: LRR-RLP copy number variation at the *INR* locus in Millettioid and non-Millettioid legume genomes.** Anchor, LRR-RLP, and other genes are colored as per legend. **A)** Locus comparison of the contiguous *INR* locus of twenty NPAAA papilionoid species. Blast hits between loci are indicated with lines (e-value < 1e-04) with score according to grayscale gradient, with darker grays indicating higher similarity. Genes are labeled LRR-RLP if a complete coding sequence is present (≥ 875 AA). Species names followed by superscript (#) were newly sequenced and assembled for this project. **B)** Locus comparison of the contiguous *INR* locus of *P. vulgaris* and *P. erosus*. *P. vulgaris* has a functionally validated INR homologue (LRR-RLP)[31], while *P. erosus* lacks a full-length LRR-RLP. The disruption of the *P. erosus* LRR-RLP was validated by PCR followed by Sanger sequencing as indicated with light green bars.
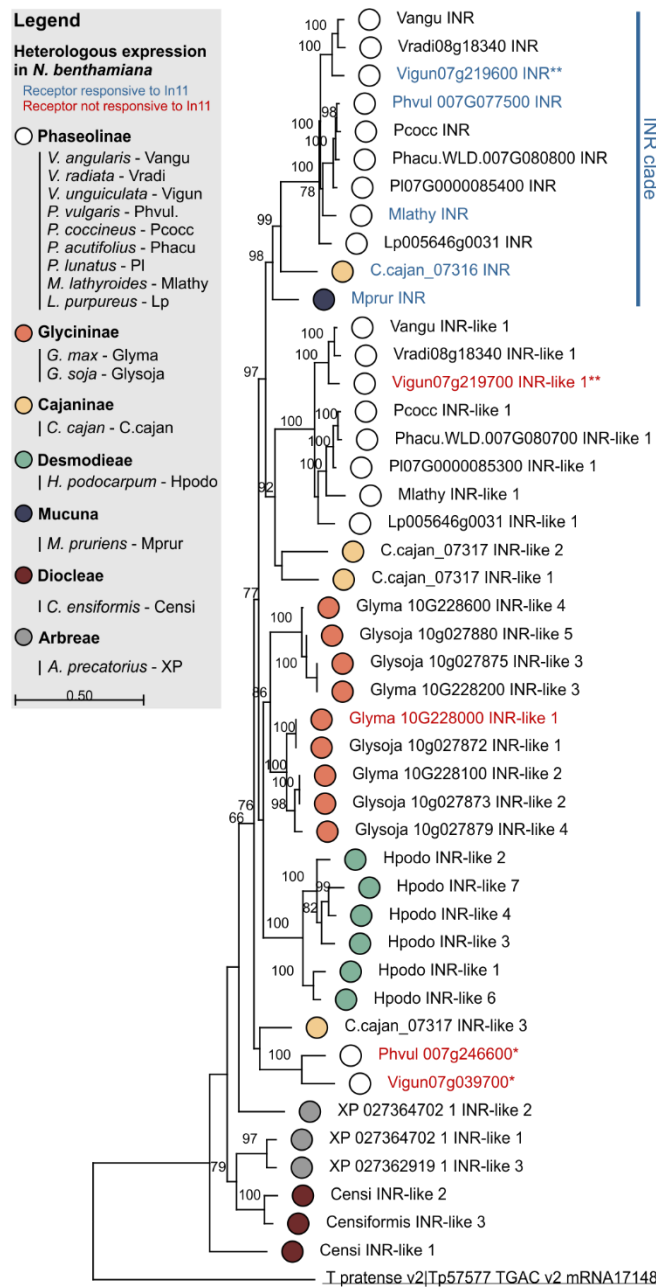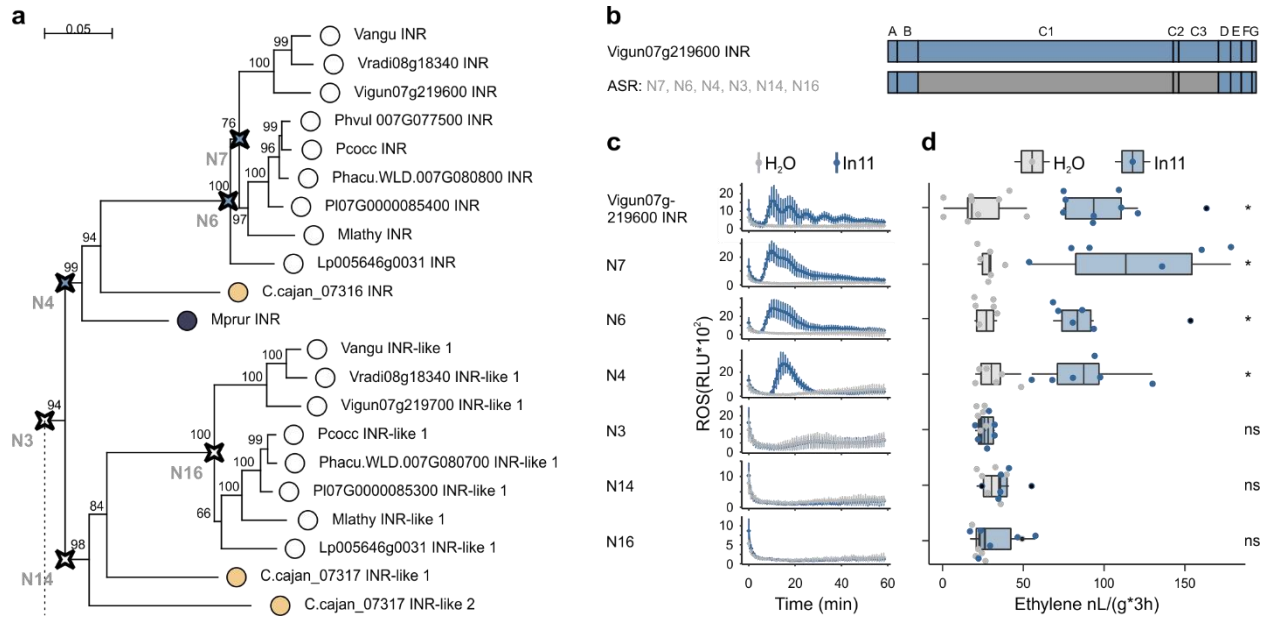
**Fig. 3: Phylogenetic analysis of LRR-RLPs at the contiguous *INR* locus reveals a clade of functional receptors.** LRR-RLPs from sixteen Millettioid species are shown. Maximum likelihood analysis bootstrap values are indicated, only values higher than 65 are shown. The scale bar represents 0.5 AA substitutions per site. Filled dots indicate species of origin according to legend, where different colors indicate different subtribes. A *T. pratense* LRR-RLP was used as an outgroup to root the phylogenetic gene tree and is underlined. The functionally validated INRs of *P. vulgaris*, *V. unguiculata, M. lathyroides*, *C. cajan* and *M. pruriens* are highlighted in blue as they confer induced ROS and ethylene functions in response to In11 upon heterologous expression in *N. benthamiana* (Suppl. Fig. 3). These five validated INRs fall within the labelled "INR clade". Heterologously expressed receptors which were not responsive to In11 are highlighted in red (Suppl. Fig. 3). One asterisk (*) highlights the LRR-RLPs which are not part of a contiguous *INR* locus. Two asterisks (**) highlight the LRR-RLPs used to create the chimeric receptors.
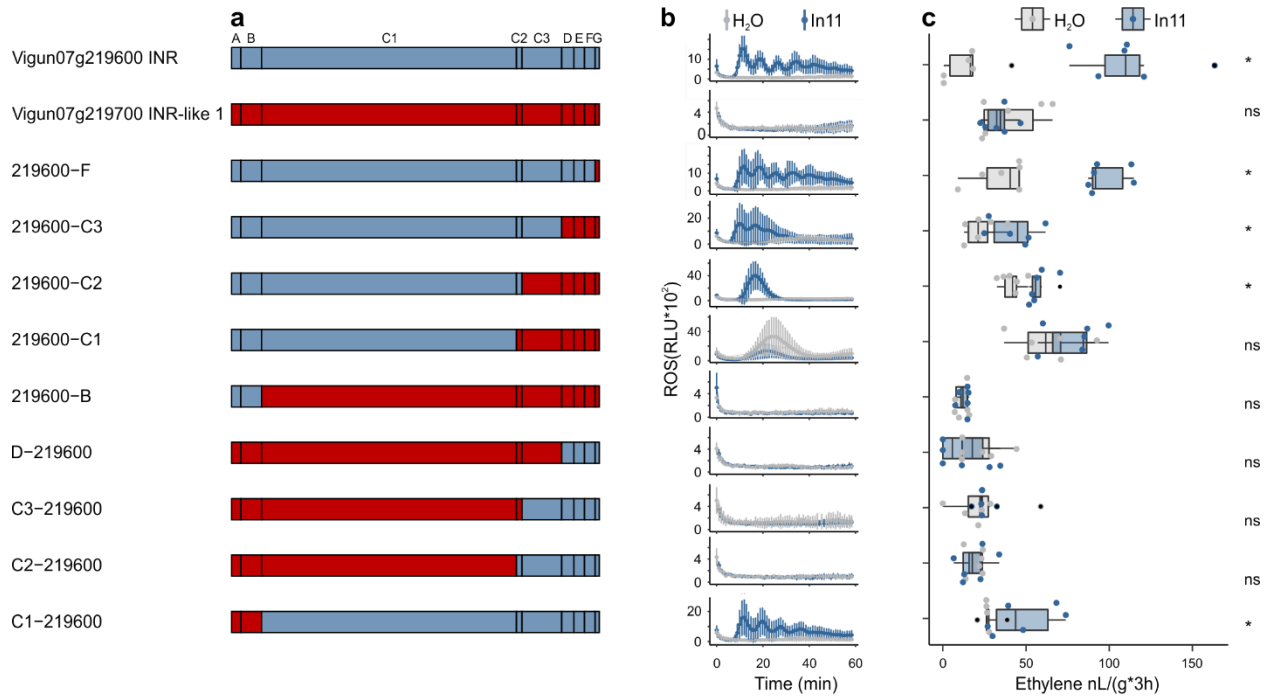
**Fig. 4: Chimeric receptors indicate that the C1 and C2 subdomains mediate INR recognition function. A)** Schematic representation of *Vigun07g219600* (blue, *Vu*INR), *Vigun07g219700* (red, *Vu*INR-like) and the nine created chimeric receptors used for structure-function analysis. Subdomains were earlier described by Fritz-Laylin *et al.* 2005; A: putative signal peptide, B: one or two pairs of Cys that may play structural roles, C: multiple LRRs with an intervening motif (C2) inserted, D: linker domain, E: acidic domain, F: transmembrane helix, and G: cytoplasmic tail. Nucleotide sequences of the chimeric receptors can be found in Suppl. File 3. **B)** In11-dependent ROS production following the heterologous expression of receptors in *N. benthamiana*. Shown are relative luminescence units (RLU) after treatment with $H_2O$ (grey), or the peptide In11 (1 µM, blue Curves indicate mean +/- SD for four independent biological replicates (n=4 plants), with each biological replicate representing six technical replicates. **C)** Ethylene production following the heterologous expression of receptors in *N. benthamiana*. Ethylene production was quantified after infiltration with $H_2O$ (grey) or the peptide In11 (1 µM, blue). Dots represent independent biological replicates (n=6 plants). Significance was tested by performing a paired Wilcoxon signed-rank test (ns non-significant, * $p \le 0.05$).

**Fig. 5: Functional analysis of ancestrally reconstructed LRR domains of INR and INR-like receptors.**
**A)** Part of the phylogenetic analyses of the LRR (C1-3 domain) of LRR-RLPs from the contiguous *INR* locus used for the ASR, full figure can be found in Suppl. Fig. 5. Ancestrally reconstructed nodes are marked with a compass star (N7, N6, N4, N3, N14 and N16). The scale bar represents 0.05 AA substitutions per site. Nucleotide sequences of the LRR domain can be found in Suppl. File 4. **B)** Schematic representation of *Vigun07g219600* (blue, *Vu*INR) and the six created ASR receptors used for structure-function analysis. Subdomains were earlier described by Fritz-Laylin *et al.* 2005; A: putative signal peptide, B: one or two pairs of Cys that may play structural roles, C: multiple LRRs with an intervening motif (C2) inserted, D: linker domain, E: acidic domain, F: transmembrane helix, and G: cytoplasmic tail. Nucleotide sequences of the chimeric receptors can be found in Suppl. File 3. **C)** In11-dependent ROS production following the heterologous expression of the ASR receptors in *N. benthamiana*. Relative luminescence units (RLU) are shown after treatment with $H_2O$ (grey), or the peptide In11 (1 µM, blue). Curves indicate mean +/- SD for four independent biological replicates (n=4 plants), with each biological replicate representing six technical replicates. **D)** Ethylene production following the heterologous expression of receptors in *N. benthamiana*. Ethylene production was quantified after infiltration with $H_2O$ (grey) or the peptide In11 (1 µM, blue). Dots represent independent biological replicates (n≥6 plants). Significance was tested by performing a paired Wilcoxon signed-rank test (ns non-significant, * p ≤ 0.05).

## Supplemental Figures

623



624
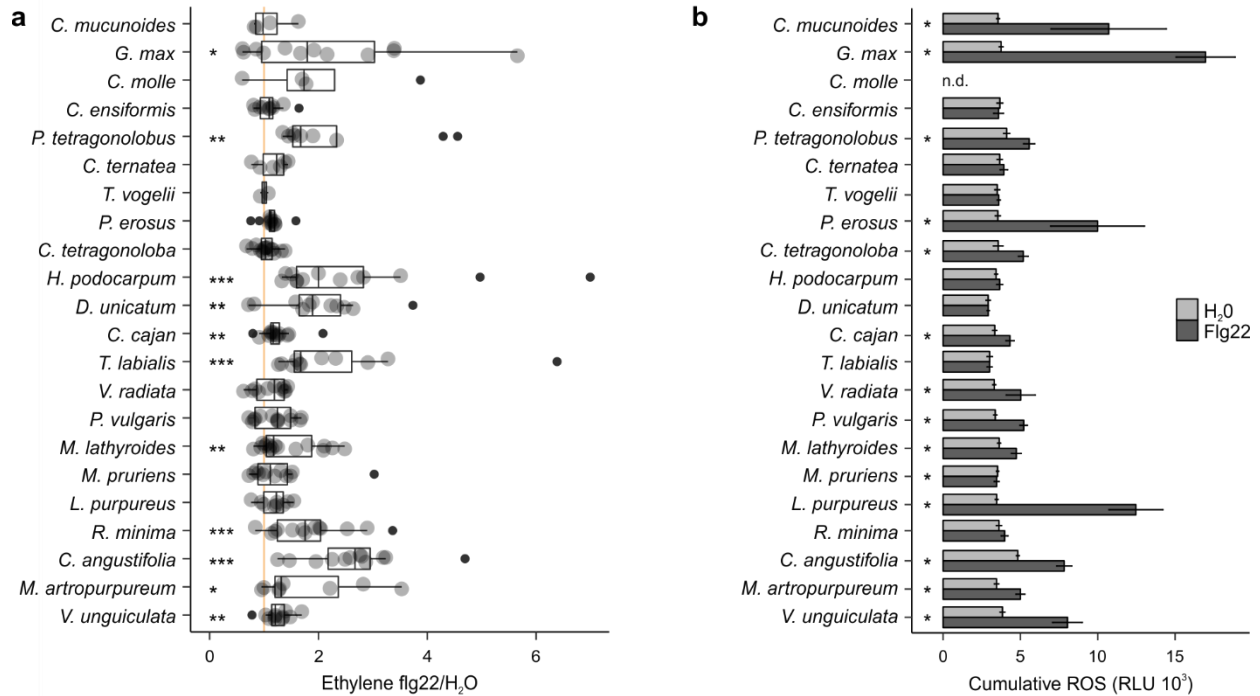
**Suppl. Fig. 1: Flg22-induced ROS and ethylene responses are idiosyncratic across Millettioid and non-Millettioid legume species. A)** Individual trifoliate leaflets were scratch wounded and treated with 1 µM flg22 or $H_2O$. The ratio of ethylene production for leaflets within the same leaf is shown (x-axis). Species are listed in order of In11/$H_2O$ response as per Fig. 1. The vertical orange line shows a ratio equal to one, *i.e.* no induced ethylene response to In11. Biological replicate plants are shown as separate dots. Significant differences between the control and the treatment of interest are indicated (paired Wilcoxon signed-rank test; ns non-significant, * $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$). **B)** ROS-bursts are used a second marker of plant immunity response upon application of flg22. Shown is cumulative ROS data (3-60min) of four technical replicates in relative luminescence units (RLU) over 1 hour (1 observation/minute) after treatment with $H_2O$, or the peptide flg22 (1 µM). No ROS data was determined for *Centrosema molle* (n.d.). Significant differences between the control and the treatment of interest were found by performing a 2-group Mann-Whitney U Test (* $p \leq 0.05$).
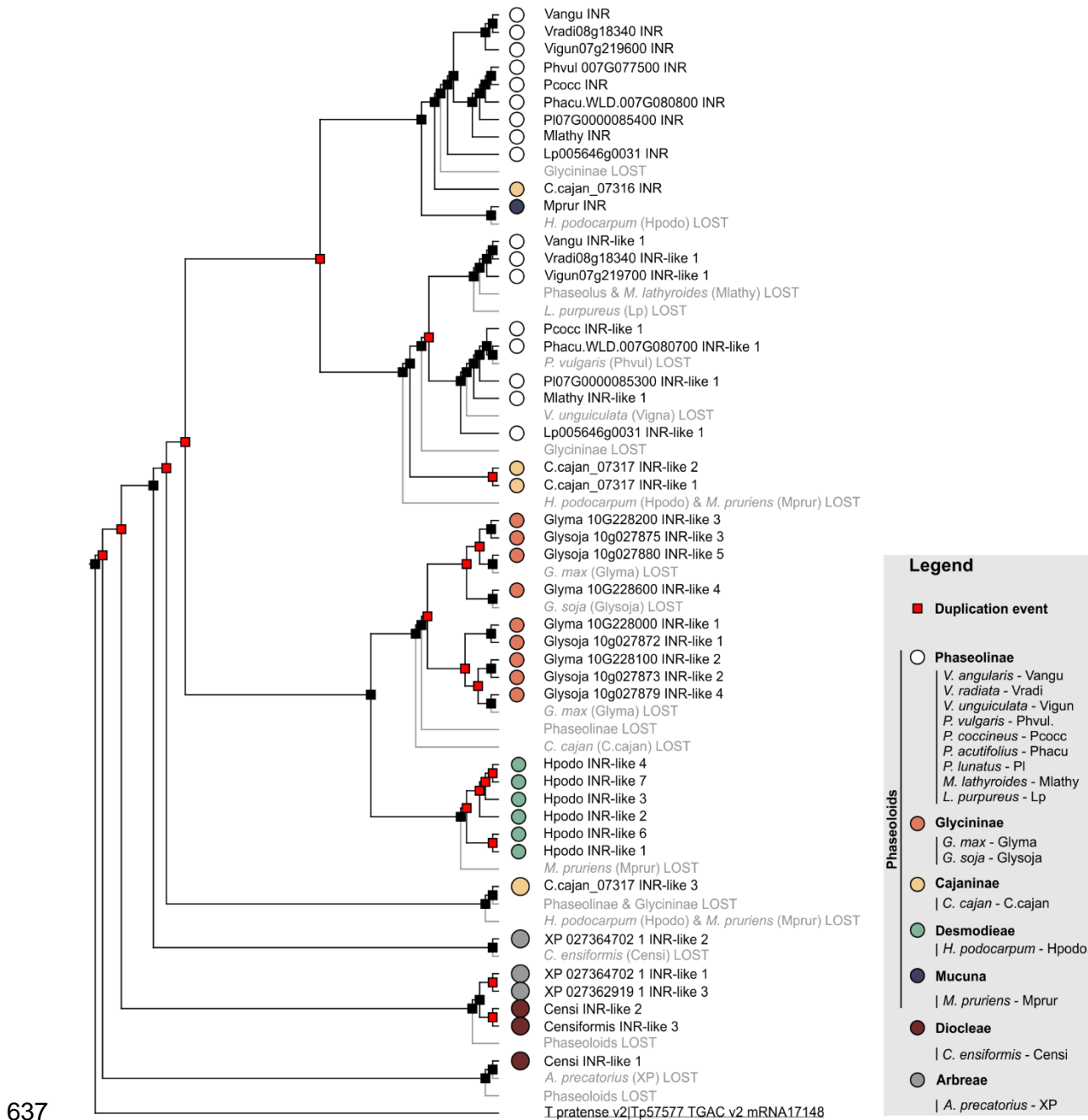
**Suppl. Fig. 2: Prediction of gene duplication and gene loss events throughout the evolution of the contiguous *INR* locus in the Millettioids.** Gene tree-species tree reconciliation to identify the duplication and loss events at each branch using Notung[43]. Predicted duplication events are marked with a red square, black squares represent speciation events, and lost nodes/genes are highlighted in grey. Filled dots indicate species of origin according to legend, where different colors indicate different subtribes. A *T. pratense* LRR-RLP was used as an outgroup to root the phylogenetic gene tree and is underlined.
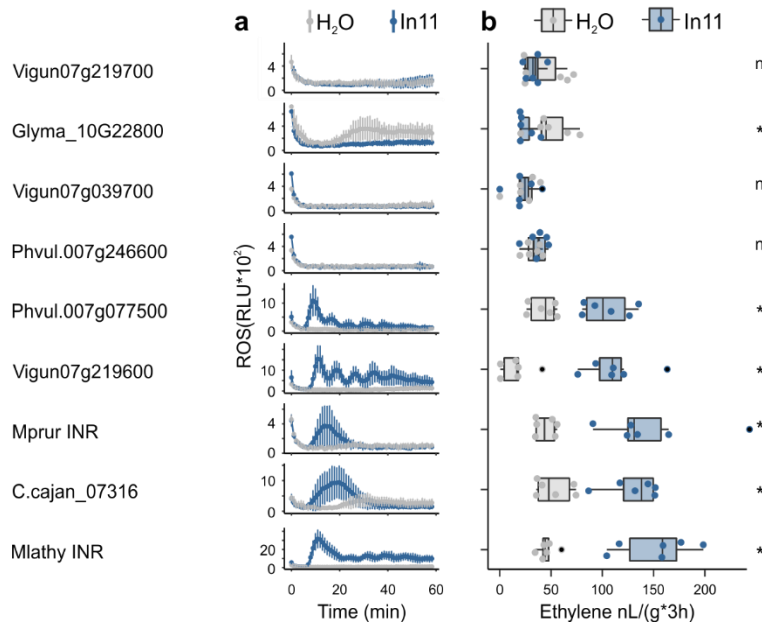
644

645 **Suppl. Fig. 3: In11-induced ROS and ethylene production conferred by INR and INR-like after**
646 **heterologous expression.** In11-dependent plant immunity response data following heterologous
647 expression of multiple receptor constructs in *N. benthamiana*. **A)** Shown are cumulative relative
648 luminescence units (RLU) after treatment with $H_2O$ (grey), or the peptide In11 (1 µM, blue). Curves indicate
649 mean +/- SD for four independent biological replicates (n=4 plants), with each biological replicate
650 representing six technical replicates. **B)** The x-axis shows the amount of ethylene released after infiltration
651 with 1 µM In11 (blue) or water (grey). Dots represent independent biological replicates (n=6 plants).
652 Significance was tested by performing a paired Wilcoxon signed-rank test (ns non-significant, * $p \leq 0.05$).

653
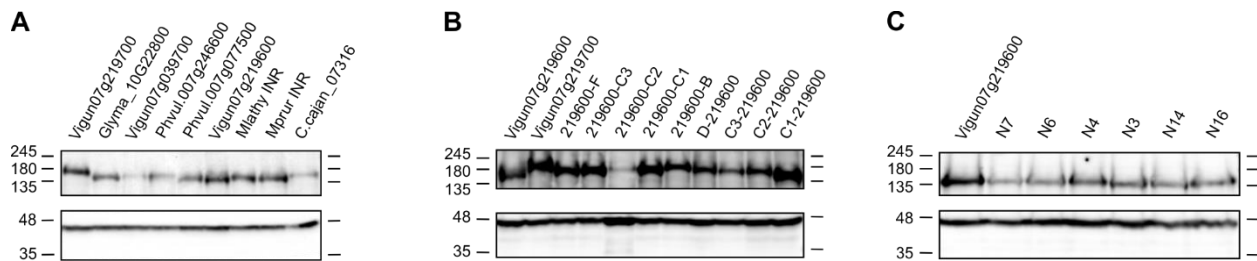


654

655 **Suppl. Fig. 4: Western blot of the heterologously expressed constructs in *N. benthamiana*.** Tissue
656 was harvested 48h after construct infiltration in *N. benthamiana*. Western blots were probed with 1) GFP
657 antibody as the receptor constructs had a C-terminal GFP tag (top), and 2) actin antibody as a loading
658 control (bottom). Western blots show all heterelogously expressed receptors of which response to In11 was
659 tested: **A)** INR and INR-like homologues from diverse legumes, **B)** chimeric receptors, **C)** ASR receptors.

660
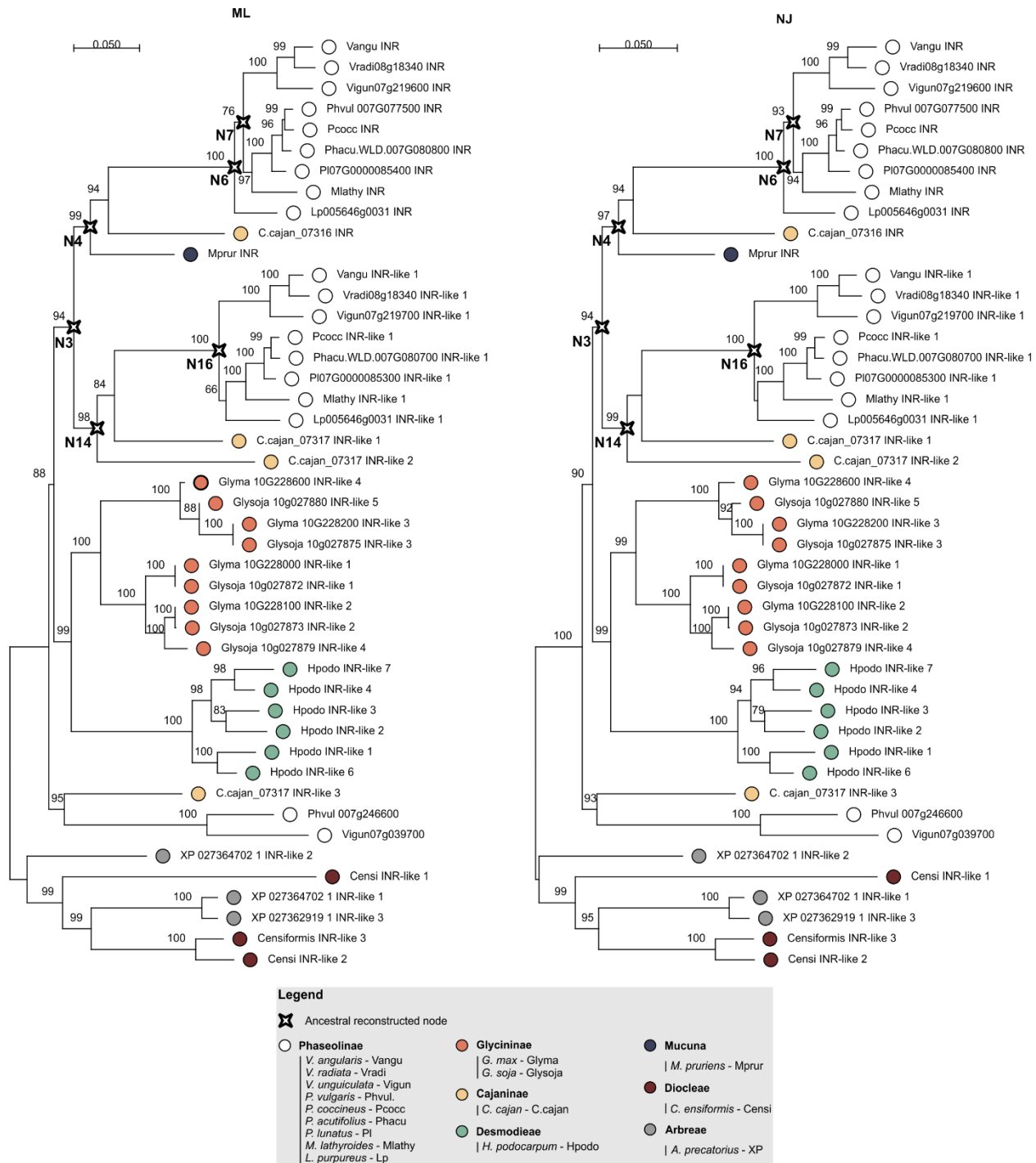
**Suppl. Fig. 5: Phylogenetic analyses of the LRR (C1-3 domain) of LRR-RLPs from the contiguous** *INR* **locus.** The phylogenetic trees were built using MEGA X software and 1000 bootstraps[80]. Maximum likelihood (ML – left side) and neighbor joining (NJ – right side) trees were calculated based on all codon positions of a codon-based alignment[81]. Maximum likelihood analysis bootstrap values are indicated, only values higher than 65 are shown. The scale bar represents 0.05 AA substitutions per site. The ancestral reconstructed nodes selected for functional validation are marked with a compass star (N7, N6, N4, N3, N14 and N16). Filled dots indicate species of origin according to legend, where different colors indicate different subtribes.

670
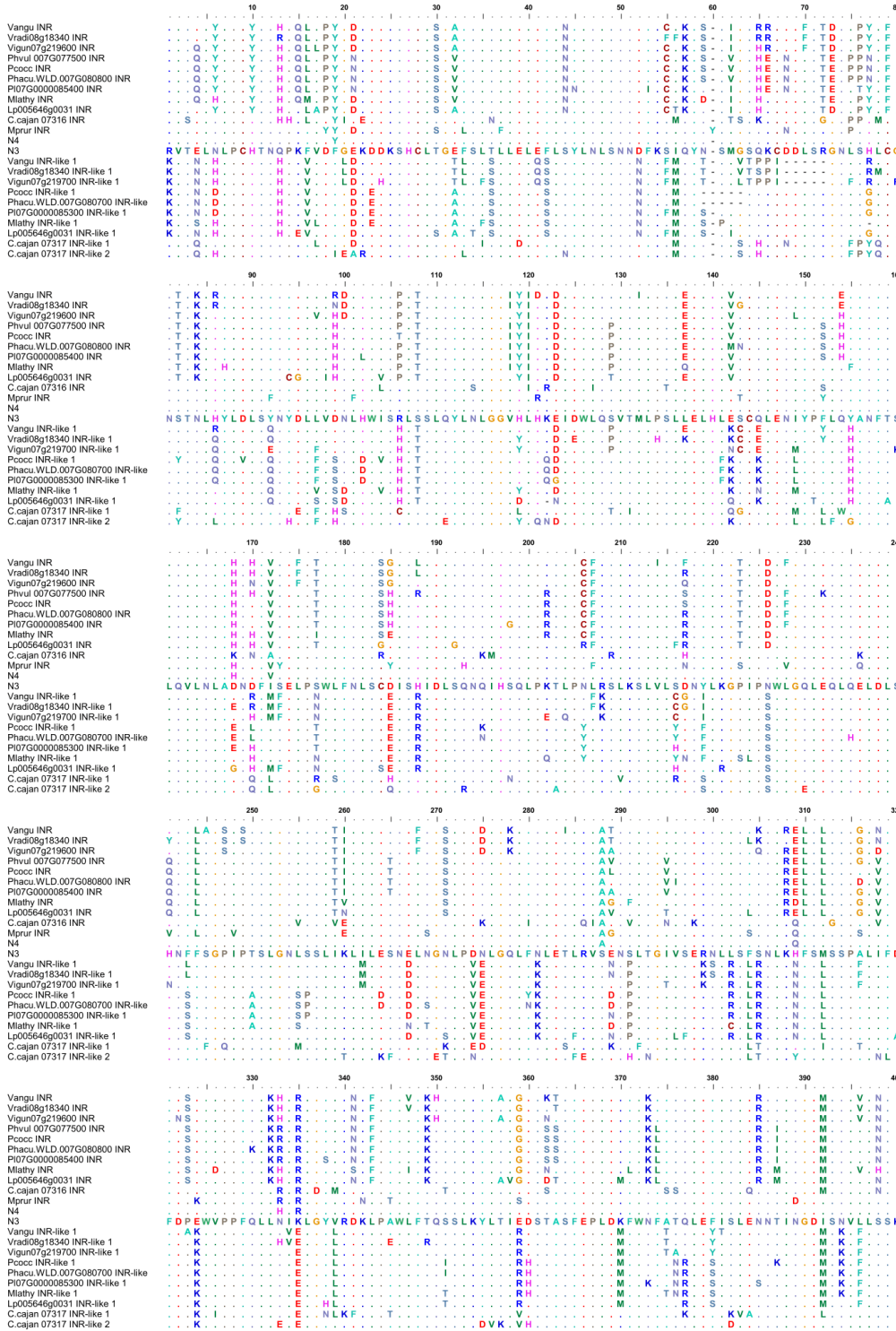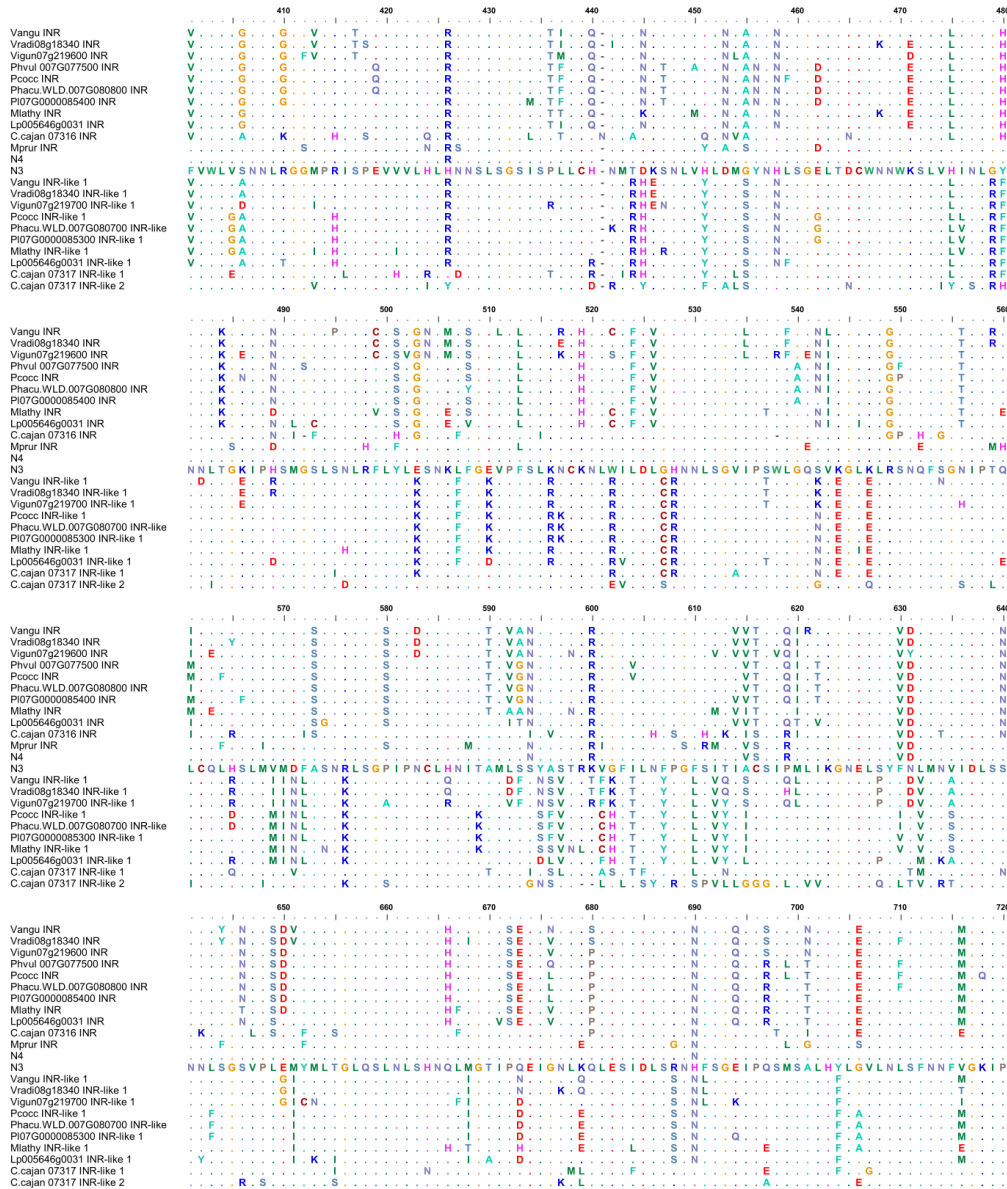
**Suppl. Fig. 6: AA sequence alignment of the LRR domains of the *INR* clade, INR-like 1 clade and the closest ancestral reconstructed nodes with differential In11-response.**

## Supplemental Table legends

**Suppl. Table 1: Overview of the used legume species, accessions, and their origin**

**Suppl. Table 2: Overview of ethylene response data of legume species**

**Suppl. Table 3: Overview of mined assemblies of contiguous *INR* loci (+ coordinates) and LRR-RLPs**

**Suppl. Table 4: Genome assembly stats**

**Suppl. Table 5: Primers used in this study**

## Supplemental file legends

**Suppl. File 1: Fasta file of nt sequences of the INR syntenic loci incorporated in the contiguous *INR* locus analysis**

**Suppl. File 2: Fasta file of the AA sequences INR and INR-like homologues included in the phylogenetic analysis**

**Suppl. File 3: Fasta file containing the sequences of the chimeric receptors**

**Suppl. File 4: Fasta file containing the nucleotide LRR domain sequences of the INR and INR-like homologues used for the ASR analysis and the resulting predicted (and domesticated for MoClo) LRR domain sequences for the ancestral nodes of interest**

## References

1.  DeFalco, T. A. & Zipfel, C. Molecular mechanisms of early plant pattern-triggered immune signaling. *Mol. Cell* **81**, 3449–3467 (2021).

2.  Snoeck, S., Guayazán-Palacios, N. & Steinbrenner, A. D. Molecular tug-of-war: plant immune recognition of herbivory. *Plant Cell* **00**, 1–17 (2022).

3.  Albert, I., Hua, C., Nürnberger, T., Pruitt, R. N. & Zhang, L. Surface sensor systems in plant immunity. *Plant Physiol.* **182**, 1582–1596 (2020).

4.  Monaghan, J. & Zipfel, C. Plant pattern recognition receptor complexes at the plasma membrane. *Curr. Opin. Plant Biol.* **15**, 349–357 (2012).

5.  Wang, G. *et al.* A genome-wide functional investigation into the roles of receptor-like proteins in Arabidopsis. *Plant Physiol.* **147**, 503–517 (2008).

6.  Fritz-Laylin, L. K., Krishnamurthy, N., Tör, M., Sjölander, K. V. & Jones, J. D. G. Phylogenomic analysis of the receptor-like proteins of rice and Arabidopsis. *Plant Physiol.* **138**, 611–623 (2005).

7.  Shiu, S. H. & Bleecker, A. B. Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiol.* **132**, 530–543 (2003).

8.  Shiu, S. H. & Bleecker, A. B. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci.* **98**, 10763–10768 (2001).

9.  Fischer, I., Diévart, A., Droc, G., Dufayard, J.-F. & Chantret, N. Evolutionary dynamics of the leucine-rich repeat receptor-like kinase (LRR-RLK) subfamily in angiosperms. *Plant Physiol.* **170**, 1595–1610 (2016).

10. Hohmann, U., Lau, K. & Hothorn, M. The structural basis of ligand perception and signal activation by receptor kinases. *Annual Review of Plant Biology* vol. 68 109–137 (2017).

11. Restrepo-Montoya, D., Brueggeman, R., McClean, P. E. & Osorno, J. M. Computational identification of receptor-like kinases 'RLK' and receptor-like proteins 'RLP' in legumes. *BMC Genomics* **21**, 1–17 (2020).

12. Lehti-Shiu, M. D., Zou, C., Hanada, K. & Shiu, S.-H. Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol.* **150**, 12–26 (2009).

13. Pok, B. *et al.* Concerted expansion and contraction of immune receptor gene

723    repertoires in plant genomes. *bioRxiv* 2022.01.01.474684 (2022)
724    doi:10.1101/2022.01.01.474684.

725 14. Jamieson, P. A., Shan, L. & He, P. Plant cell surface molecular cypher: receptor-
726    like proteins and their roles in immunity and development. *Plant Sci.* **274**, 242–
727    251 (2018).

728 15. Tang, P. *et al.* Disease resistance signature of the leucine-rich repeat receptor-
729    like kinase genes in four plant species. *Plant Sci.* **179**, 399–406 (2010).

730 16. Pruitt, R. N. *et al.* The EDS1–PAD4–ADR1 node mediates Arabidopsis pattern-
731    triggered immunity. *Nat. 2021 5987881* **598**, 495–499 (2021).

732 17. Steinbrenner, A. D. The evolving landscape of cell surface pattern recognition
733    across plant immune networks. *Curr. Opin. Plant Biol.* **56**, 135–146 (2020).

734 18. Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J. & Jones,
735    J. D. G. Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum*
736    by transposon tagging. *Science (80-. ).* **266**, 789–793 (1994).

737 19. Thomas, C. M. *et al.* Characterization of the tomato Cf-4 gene for resistance to
738    *Cladosporium fulvum* identifies sequences that determine recognitional specificity
739    in Cf-4 and Cf-9. *Plant Cell* **9**, 2209–2224 (1997).

740 20. Kruijt, M., Kip, D. J., Joosten, M. H. A. J., Brandwagt, B. F. & De Wit, P. J. G. M.
741    The Cf-4 and Cf-9 resistance genes against *Cladosporium fulvum* are conserved
742    in wild tomato species. *http://dx.doi.org/10.1094/MPMI-18-1011* **18**, 1011–1021
743    (2007).

744 21. Dixon, M. S. *et al.* The tomato Cf-2 disease resistance locus comprises two
745    functional genes encoding leucine-rich repeat proteins. *Cell* **84**, 451–459 (1996).

746 22. Luderer, R., Takken, F. L. W., De Wit, P. J. G. M. & Joosten, M. H. A. J.
747    *Cladosporium fulvum* overcomes Cf-2-mediated resistance by producing
748    truncated AVR2 elicitor proteins. *Mol. Microbiol.* **45**, 875–884 (2002).

749 23. Krüger, J. *et al.* A tomato cysteine protease required for Cf-2-dependent disease
750    resistance and suppression of autonecrosis. *Science (80-. ).* **296**, 744–747
751    (2002).

752 24. Rooney, H. C. E. *et al.* Plant science: *Cladosporium* Avr2 inhibits tomato Rcr3
753    protease required for Cf-2-dependent disease resistance. *Science (80-. ).* **308**,
754    1783–1786 (2005).

755 25. Zhang, L. *et al.* Fungal endopolygalacturonases are recognized as microbe-
756    associated molecular patterns by the Arabidopsis receptor-like protein
757    RESPONSIVENESS TO BOTRYTIS POLYGALACTURONASES1. *Plant Physiol.*
758    **164**, 352–364 (2014).

759 26. Zhang, L. *et al.* Distinct immune sensor systems for fungal
760    endopolygalacturonases in closely related Brassicaceae. *Nat. Plants 2021 79* **7**,
761    1254–1263 (2021).

762 27. Albert, I. *et al.* An RLP23–SOBIR1–BAK1 complex mediates NLP-triggered
763    immunity. *Nat. Plants 2015 110* **1**, 1–9 (2015).

764 28. Jehle, A. K. *et al.* The receptor-like protein ReMAX of Arabidopsis detects the
765    microbe-associated molecular pattern eMax from *Xanthomonas*. *Plant Cell* **25**,
766    2330–2340 (2013).

767   29.   Hegenauer, V. *et al.* The tomato receptor CuRe1 senses a cell wall protein to
768          identify *Cuscuta* as a pathogen. *Nat. Commun. 2020 111* **11**, 1–7 (2020).

769   30.   Hegenauer, V. *et al.* Detection of the plant parasite *Cuscuta reflexa* by a tomato
770          cell surface receptor. *Science (80-. ).* **353**, 478–481 (2016).

771   31.   Steinbrenner, A. D. *et al.* A receptor-like protein mediates plant immune
772          responses to herbivore-associated molecular patterns. *Proc. Natl. Acad. Sci.* **117**,
773          31510–31518 (2020).

774   32.   Schmelz, E. A. *et al.* Fragments of ATP synthase mediate plant perception of
775          insect attack. *Proc. Natl. Acad. Sci.* **103**, 8894–8899 (2006).

776   33.   Schmelz, E. A. *et al.* An amino acid substitution inhibits specialist herbivore
777          production of an antagonist effector and recovers insect-induced plant defenses.
778          *PLANT Physiol.* **160**, 1468–1478 (2012).

779   34.   Schmelz, E. A., LeClere, S., Carroll, M. J., Alborn, H. T. & Teal, P. E. A. Cowpea
780          chloroplastic ATP synthase is the source of multiple plant defense elicitors during
781          insect herbivory. *Plant Physiol.* **144**, 793–805 (2007).

782   35.   Albert, I., Zhang, L., Bemm, H. & Nürnberger, T. Structure-function analysis of
783          immune receptor AtRLP23 with its ligand nlp20 and coreceptors AtSOBIR1 and
784          AtBAK1. *Mol. Plant-Microbe Interact.* **32**, 1038–1046 (2019).

785   36.   Delaux, P. M. *et al.* Reconstructing trait evolution in plant evo–devo studies. *Curr.*
786          *Biol.* **29**, R1110–R1118 (2019).

787   37.   Białas, A. *et al.* Two nlr immune receptors acquired high-affinity binding to a
788          fungal effector through convergent evolution of their integrated domain. *Elife* **10**,
789          e66961 (2021).

790   38.   Read, A. C. *et al.* Genome assembly and characterization of a complex zfBED-
791          NLR gene-containing disease resistance locus in Carolina Gold Select rice with
792          Nanopore sequencing. *PLoS Genet.* **16**, e1008571 (2020).

793   39.   Michael, T. P. & VanBuren, R. Building near-complete plant genomes. *Curr. Opin.*
794          *Plant Biol.* **54**, 26–33 (2020).

795   40.   Goodin, M. M., Zaitlin, D., Naidu, R. A. & Lommel, S. A. *Nicotiana benthamiana*:
796          its history and future as a model for plant-pathogen interactions. *Mol. Plant-*
797          *Microbe Interact.* **21**, 1015–1026 (2008).

798   41.   Cardoso, D. *et al.* Filling in the gaps of the papilionoid legume phylogeny: The
799          enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging
800          Amburaneae clade. *Mol. Phylogenet. Evol.* **84**, 112–124 (2015).

801   42.   Zhao, Y. *et al.* Nuclear phylotranscriptomics and phylogenomics support
802          numerous polyploidization events and hypotheses for the evolution of rhizobial
803          nitrogen-fixing symbiosis in Fabaceae. *Mol. Plant* **14**, 748–773 (2021).

804   43.   Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: A program for dating gene
805          duplications and optimizing gene family trees. *https://home.liebertpub.com/cmb* **7**,
806          429–447 (2004).

807   44.   Liu, Y. *et al.* Pan-genome of wild and cultivated soybeans. *Cell* **182**, 162-176.e13
808          (2020).

809   45.   Randall, R. N., Radford, C. E., Roof, K. A., Natarajan, D. K. & Gaucher, E. A. An
810          experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat.*

811        *Commun.* **7**, 1–6 (2016).

812   46.   Schultink, A. & Steinbrenner, A. D. A playbook for developing disease-resistant
813        crops through immune receptor identification and transfer. *Curr. Opin. Plant Biol.*
814        **62**, 102089 (2021).

815   47.   Lacombe, S. *et al.* Interfamily transfer of a plant pattern-recognition receptor
816        confers broad-spectrum bacterial resistance. *Nat. Biotechnol. 2010 284* **28**, 365–
817        369 (2010).

818   48.   Pfeilmeier, S. *et al.* Expression of the *Arabidopsis thaliana* immune receptor EFR
819        in *Medicago truncatula* reduces infection by a root pathogenic bacterium, but not
820        nitrogen-fixing rhizobial symbiosis. *Plant Biotechnol. J.* **17**, 569–579 (2019).

821   49.   Kourelis, J. *et al.* Evolution of a guarded decoy protease and its receptor in
822        solanaceous plants. *Nat. Commun. 2020 111* **11**, 1–15 (2020).

823   50.   Shi, T., Huang, H., Sanderson, M. J. & Tax, F. E. Evolutionary dynamics of
824        leucine-rich repeat receptor-like kinases and related genes in plants: a
825        phylogenomic approach. *J. Integr. Plant Biol.* **56**, 648–662 (2014).

826   51.   Restrepo-Montoya, D., McClean, P. E. & Osorno, J. M. Orthology and synteny
827        analysis of receptor-like kinases "RLK" and receptor-like proteins "RLP" in
828        legumes. *BMC Genomics* **22**, 1–17 (2021).

829   52.   Wang, L. *et al.* The pattern-recognition receptor CORE of Solanaceae detects
830        bacterial cold-shock protein. *Nat. Plants 2016 212* **2**, 1–9 (2016).

831   53.   Zipfel, C. *et al.* Perception of the bacterial PAMP EF-Tu by the receptor EFR
832        Restricts Agrobacterium-mediated transformation. *Cell* **125**, 749–760 (2006).

833   54.   Wei, Y. *et al.* An immune receptor complex evolved in soybean to perceive a
834        polymorphic bacterial flagellin. *Nat. Commun. 2020 111* **11**, 1–11 (2020).

835   55.   Fürst, U. *et al.* Perception of *Agrobacterium tumefaciens* flagellin by FLS2XL
836        confers resistance to crown gall disease. *Nat. Plants 2020 61* **6**, 22–27 (2020).

837   56.   Sun, Y. *et al.* Structural basis for flg22-induced activation of the Arabidopsis
838        FLS2-BAK1 immune complex. *Science (80-. ).* **342**, 624–628 (2013).

839   57.   Wan, W. L. *et al.* Comparing Arabidopsis receptor kinase and receptor protein-
840        mediated immune signaling reveals BIK1-dependent differences. *New Phytol.*
841        **221**, 2080–2095 (2019).

842   58.   Felix, G., Duran, J. D., Volko, S. & Boller, T. Plants have a sensitive perception
843        system for the most conserved domain of bacterial flagellin. *Plant J.* **18**, 265–276
844        (1999).

845   59.   Schmelz, E. A., Alborn, H. T., Banchio, E. & Tumlinson, J. H. Quantitative
846        relationships between induced jasmonic acid levels and volatile emission in *Zea*
847        *mays* during *Spodoptera exigua* herbivory. *Planta* **216**, 665–673 (2003).

848   60.   R Developement Core Team. R: a language and environment for statistical
849        computing. *R Found. Stat. Comput.* **1**, 409 (2015).

850   61.   Wickham, H. & Francois, R. dplyr: a grammar of data manipulation. **3**, 156 (2015).

851   62.   Wickham, H. ggplot2 elegant graphics for data analysis. *Media* **35**, 211 (2009).

852   63.   Lutz, K. A., Wang, W., Zdepski, A. & Michael, T. P. Isolation and analysis of high
853        quality nuclear DNA with reduced organellar DNA for plant genome sequencing
854        and resequencing. *BMC Biotechnol.* **11**, 1–9 (2011).

64.  Lonardi, S. *et al.* The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* **98**, 767–782 (2019).

65.  Kang, Y. J. *et al.* Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* **5**, 1–9 (2014).

66.  Kang, Y. J. *et al.* Draft genome sequence of adzuki bean, *Vigna angularis*. *Sci. Rep.* **5**, 1–8 (2015).

67.  Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183 (2010).

68.  Chang, Y. *et al.* The draft genomes of five agriculturally important African orphan crops. *Gigascience* **8**, 1–16 (2018).

69.  Varshney, R. K. *et al.* Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* **30**, 83–89 (2012).

70.  Schmutz, J. *et al.* A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **46**, 707–713 (2014).

71.  Xie, M. *et al.* A reference-grade wild soybean genome. *Nat. Commun.* **10**, 1–12 (2019).

72.  Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

73.  Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

74.  Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

75.  Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES science gateway for inference of large phylogenetic trees. in *2010 Gateway Computing Environments Workshop, GCE 2010* 1–8 (IEEE, 2010). doi:10.1109/GCE.2010.5676129.

76.  Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–3 (2014).

77.  Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765 (2011).

78.  Engler, C. *et al.* A Golden Gate modular cloning toolbox for plants. *ACS Synth. Biol.* **3**, 839–843 (2014).

79.  Offord, V., Coffey, T. J. & Werling, D. LRRfinder: a web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database. *Dev. Comp. Immunol.* **34**, 1035–1041 (2010).

80.  Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547 (2018).

81.  Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

82.  Ashkenazy, H. *et al.* FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40**, W580–W584 (2012).

899  83.  Egan, A. N., Vatanparast, M. & Cagle, W. Parsing polyphyletic *Pueraria*:
900       delimiting distinct evolutionary lineages through phylogeny. *Mol. Phylogenet. Evol.*
901       **104**, 44–59 (2016).
902  84.  Li, H. *et al.* Diversification of the phaseoloid legumes: effects of climate change,
903       range expansion and habit shift. *Front. Plant Sci.* **4**, 386 (2013).
904  85.  Lavin, M., Herendeen, P. S. & Wojciechowski, M. F. Evolutionary rates analysis of
905       Leguminosae implicates a rapid diversification of lineages during the Tertiary.
906       *Syst. Biol.* **54**, 575–594 (2005).
907  86.  Stefanović, S., Pfeil, B. E., Palmer, J. D. & Doyle, J. J. Relationships among
908       Phaseoloid legumes based on sequences from eight chloroplast regions. *Syst.*
909       *Bot.* **34**, 115–128 (2009).
910