

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

Simon Snoeck¹, Bradley W. Abramson², Anthony G. K. Garcia¹, Ashley N. Egan³, Todd P. Michael², Adam D. Steinbrenner^{1, *}

¹Department of Biology, University of Washington, Seattle, Washington 98195, USA

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

³Department of Biology, Utah Valley University, Orem, Utah 84058, USA

*Author for correspondence: astein10@uw.edu

Abstract

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

Introduction

Plants use a combination of constitutive and inducible defensive traits to resist challenges by pathogens and herbivores. Activation of inducible defenses requires perception of the threat¹, mediated by an innate immune system that uses germline-encoded pattern recognition receptors (PRRs) at the cell surface to recognize pathogen, herbivore, and damage-associated molecular patterns (PAMPs, HAMPs, and DAMPs)^{2,3}. Well-characterized PRRs in the large receptor kinase (RK) family comprise an extracellular domain, a single-pass transmembrane motif, and an intracellular kinase domain⁴. Receptor-like proteins (RLPs) share structural similarities with RKs but lack a kinase domain^{5,6}. The most common extracellular domain of RK/RLPs is a series of leucine-rich repeats (LRRs), which are known to mediate elicitor binding and co-receptor association^{7–11}. LRR-RK/RLPs form a large gene family; the *Arabidopsis thaliana* genome contains about 223 LRR-RKs and 57 LRR-RLPs^{5,6,8,12}, and the number of annotated RK/RLPs per genome varies across and between plant families^{11,13}. Moreover, comparative genomic analyses of LRR-RK/RLPs involved in biotic interactions have revealed strong diversifying selection, lineage-specific expanded gene clusters and immune receptor repertoire variation both within and between species^{6,14–17}.

Evidence is accumulating for diverse roles of RLPs as immune sensors^{14,17}. Various LRR-RLPs from *Arabidopsis*, tomato, and cowpea have been shown to detect molecular patterns from fungi,

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

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

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

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 Millettoid 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 Millettoids 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 Millettoids (Fig. 1a). To investigate this emergence, we analyzed the *de novo* assembly of *C. tetragonoloba*, a close outgroup of Millettoid 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

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

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

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

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

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

clade for *G. soya*, *G. max* and *H. podocarpum* and clade specific duplication events resulting in LRR-RLP expansions of 4-7 tandem duplicates. To confirm that *INR* was lost in *Glycine* and not just in reference assemblies, we performed BLASTP searches using Vigun07g219700 AA sequence against 26 *G. soja* and *G. max de novo* assemblies from a recent pangenome analysis⁴⁴. Like the *Glycine* reference genomes, no BLASTP hits to Vu07g219600 for any of the 26 de novo assemblies were identified that had an AA similarity higher than 76%, suggesting that *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

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

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

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

Ancestral receptors to the *INR* clade of the *Vigna*, *Phaseolus* and *Macroptilium* (N7), the Phaseolinae (N6) and the Phaseoloids (N4) conferred In11-induced ethylene and ROS response. In contrast, ancestral receptors of the *INR*-like 1 clade of the Phaseolinae (N16) and the Phaseoloids (N14) were not responsive to In11 (Fig. 5d, Fig. 5c). Moreover, the common ancestral receptor of the *INR* and its *INR*-like 1 sister clade (N3) is not responsive to In11 (Fig. 5). Hence, a small number of differences, 16 of 720 AA, between the LRR domain of N3 and N4, mediate differential In11 response. Moreover, only a subset of those AA is conserved within all LRR-RLPs of the *INR* clade and absent in all LRR-RLPs of the *INR*-like 1 sister clade (Suppl. Fig. 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 functions². 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 Millettoids 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 Millettoids 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 recombination⁴⁹. 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.

The abundance of closely related legume genomes facilitates gene-species tree reconciliation, suggesting mechanisms contributing to the dynamic nature of *INR* receptor loci^{43,50}. Interestingly, we observed three independent cases of *INR* loss within the Phaseoloids. Two reference genomes as well as 26 *de novo* assemblies of the closely related Glycininae, *G. soya* and *G. max*, do not encode *INR*. Nevertheless, *INR* loss does not seem to predate the divergence of the Glycininae as *Teramnus labialis* is able to respond to In11 (Fig. 1), although genome data are lacking for this species. Besides reciprocal gene loss of *INR*, gene tree-species tree reconciliation also suggests the involvement of five tandem duplications of an *INR*-like gene predating radiation of the genus *Glycine* (Suppl. Fig. 2). This is consistent with whole-genome observations of preferential gene loss in tandem clusters⁵⁰. Similarly, within the separate Desmodieae lineage, our analysis also predicts that a *H. podocarpum* *INR*-like gene underwent specific tandem duplication events after the loss of *INR*, as its genome contains seven *INR*-like homologues which cluster together in a phylogenetic analysis. In contrast, *P. erosus* seems to have lost both *INR* and *INR*-like as it does not contain any complete coding sequence of an LRR-RLP at the *INR* locus. Besides *P. erosus*, *P. vulgaris* and *M. pruriens* also lost *INR*-like. Loss of *INR* (and *INR*-like) may reflect the propensity of tandemly duplicated loci to lose functions through gene conversion, and/or reciprocal gene loss⁵⁰.

Our high-resolution analysis of the emergence of *INR* provides a roadmap for understanding other lineage-specific PRRs to detect PAMPs. While most characterized LRR-RLPs are family-specific^{17,26}, extensive phenotyping of an entire plant family has not yet been conducted for well-studied PAMPs such as elf18, nlp20, pg9, crip21, csp22, and xup25. Understanding the emergence of known LRR-RK/RLPs for these PAMPs, namely EFR, RLP23, RLP42, CORE, CuRE1, and XPS1 will illuminate broader mechanisms underlying the evolution of key immune receptor modules in plants^{26,27,29,52,53}. LRR-RLPs often occur in complex loci with variation within and between species^{17,26}. However, a recent pangenomic analysis of Arabidopsis indicates that several PAMP sensing LRR-RLPs occur in relatively simple and conserved loci across *A. thaliana* varieties¹⁶. It will be interesting to see if Arabidopsis PRR functions evolved along a similar trajectory to *INR*, for example via ancestral duplications preceding fixation in the Arabidopsis lineage. In summary, additional case studies for specific receptors are needed to reveal broader patterns in receptor evolution.

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

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

burst relative to all other In11-responsive constructs tested here (Fig. 4). These findings are consistent with previous truncation and chimeric protein analyses for Arabidopsis RLP23 and RLP42. Truncations of the RLP23 ectodomain abolished function, suggesting the necessity of the entire ectodomain for elicitor binding or proper assembly of the receptor³⁵. Additionally, chimeric receptors implicated the importance of RLP42 its twelve N-terminal LRRs (C1) and LRR21-LRR24 which includes the island subdomain (C2) for recognition of fungal endopolygalacturonases²⁶.

Additionally, our chimeric receptor data are consistent with a vestigial role for the cytoplasmic tail of PRRs in the LRR-RLP family. Strikingly, all identified INR homologues encode a cytoplasmic tail of only 10AA, shorter relative to all identified INR-like homologues here. Nevertheless, swapping the *Vu*INR (Vigun07g219600) cytoplasmic tail to the extended version of *Vu*INR-like 1 (Vigun07g219700), 219600-F, did not affect In11-response (Fig. 4). Previously, the complete deletion of the intracellular 17-amino-acid tail of RLP23 reduced but did not abolish receptor function³⁵. Additionally, a previously described chimeric swap replacing the RLP42 terminal LRR, transmembrane helix and cytoplasmic tail with the respective subdomains of a non-responsive paralogue was still responsive to the pg9(*At*) elicitor²⁶. Intriguingly, quantitative differences were earlier reported as the C-terminally truncated RLP23 had a reduced nlp20 response, consistent with auxiliary rather than essential function for the RLP23 cytoplasmic tail³⁵.

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

A similar ASR approach to ours was employed to understand the 98-AA heavy metal-associated (HMA) domain of the plant intracellular NOD-like receptor Pik-1³⁷. Specific AA changes introduced into an ancestrally reconstructed backbone were sufficient to confer effector binding and immune functions. For Pik-1 HMA, structural data provided additional insight into the mechanism of effector binding in ancestral and extant proteins. Our use of an ASR approach for a relatively long LRR domain (720 AA) now demonstrates the power of dense comparative genomic analyses to also identify key residues in extant PRRs without defined binding sites. In the absence of structural data for ligand-binding LRR-RLPs, an ASR approach may be useful to identify sets of co-varying residues critical for binding and signaling functions.

Materials and Methods

Plant materials

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

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

Peptide-induced ethylene production in legume species.

The In11 peptide (ICDINGVCVDA) is a host derived proteolytic fragment of the ATP synthase γ -subunit (cATPC), based on the *V. unguiculata* cATPC sequence³². The flg22 peptide (QRLSTGSRINSAKDDAAGLQIA) originates from bacterial flagellin⁵⁸. Both peptides were synthesized (Genscript) and reconstituted in H₂O. Leaflets were lightly scratch wounded with a fresh razor blade to remove cuticle area, and 10 μ L of H₂O with or without peptide (1 μ M) was equally spread over the wounds with a pipette tip. After 1h, leaflets were excised and placed in sealed tubes for 2h before headspace sampling (1 mL). Ethylene was measured as previously described with a gas chromatographer (HP 5890 series 2, supelco #13018-U, 80/100 Hayesep Q 3FT x 1/8IN x 2.1MM nickel) with flame ionization detection and quantified using a standard curve (Scott, 99.5% ethylene, (Cat. No 25881-U)) (Suppl. Table 2)⁵⁹. Subsequently, R (v4.0.3) 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, statistics were performed by using the paired Wilcoxon signed-rank test^{60–62}. The resulting figure was edited in Corel-DRAW Home & Student x7.

ROS production in legumes

Leaf punches were taken with a 4-mm biopsy punch and floated in 150 μ L of H₂O using individual cells of a white 96-well white bottom plate (BRANDplates F pureGrade S white (REF 781665)). After overnight incubation, ROS production was measured upon addition of a 100 μ L assay solution which contains 10 μ g/ml luminol-horseradish peroxidase (HRP), 17 μ g/ml luminol and the treatment (2.5 μ M In11, 2.5 μ M flg22 or H₂O). Luminescence was quantified with a TECAN SPARK plate reader every minute for an hour using an integration time of 500ms. Four technical replicates were quantified for each treatment and significant differences were determined by performing a 2-group Mann-Whitney U Test between both treatments. 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.

Genome sequencing and assembly of legume species

Leaf tissue of legume plants was harvested and grounded using an N₂-chilled mortar and pestle. In contrast to *P. erosus*, nuclei isolation was first performed on frozen tissue for *C. ensiformis*, *C. tetragonoloba* and *H. podocarpum* using the Bionano Plant Tissue Homogenization Buffer (Part number 20283). Nuclear DNA was extracted for all above mentioned species with a modified CTAB protocol as described previously⁶³. Resultant High-Molecular Weight (HMW) DNA concentrations were determined by Qubit and Bioanalyzer.

A modified protocol using the Oxford Nanopore Technologies (ONT) Rapid barcoding kit (SQK-RBK004) was used for sequencing. Briefly, 27ul of ~20 ng/ μ L HMW DNA was combined with 3ul of Rapid Barcoding fragmentation mix and incubated for 1min at 30°C followed by 1 min at 80°C. Ampure beads were added at a 0.7x final concentration and bead clean-ups performed as described in the Rapid barcoding kit (SQK-RBK004). All remaining steps were performed as described in the Rapid barcoding kit protocol. Each sample was sequenced on a single MinION or PromethION flowcell (R9.4). High-accuracy base calling was performed in real time with MinKnow (v20.10.6).

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

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

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

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

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

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

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

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

Sequences were aligned using the online version of MAFFT 7 using the E-INS-i strategy⁷⁴. One potential pseudogene of *H. podocarpum* was not incorporated in the phylogenetic analysis since the length of the sequence was about 79% of cowpea *INR* due to the absence of a part of the LRR domain in comparison to all other annotated receptors at the contiguous *INR* locus. A phylogenetic analysis was performed on the CIPRES web portal using RAXML-HPC2 on XSEDE (v8.2.12) with the automatic protein model assignment algorithm using maximum likelihood criterion and 250 bootstrap replicates^{75,76}. The DUMMY2 protein model was selected as the best scoring model for maximum likelihood analysis. The resulting tree was rooted, visualized using MEGA10 and edited in Corel-DRAW Home & Student x7.

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

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

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

Leaf tissue of legume plants was harvested and grounded using an N₂-chilled mortar and pestle. Genomic DNA was extracted using the Nucleospin Plant II kit (Macherey-Nagel). RNA was extracted using the Nucleospin RNA Plant kit (Macherey-Nagel). cDNA was created using SuperScript™ IV Reverse Transcriptase (Invitrogen). All constructs were created using a hierarchical modular cloning approach facilitated by the MoClo toolkit and the MoClo Plant Parts kit^{77,78}. LRR-RLPs with no introns were PCR amplified from genomic DNA (Q5 Hot Start High-Fidelity, NEB), all others (*Phvul.007g246600* and *Vigun07g039700*) were amplified from cDNA. All primers used for amplification are listed in Suppl. Table 5. All amplified PCR fragments were gel extracted using the PureLink™ Quick Gel Extraction Kit (Thermo Fisher scientific) and purified and concentrated using the Monarch PCR & DNA Cleanup Kit. Subsequently, the PCR fragments were ligated in an L-1 acceptor vector. A second digestion/ligation step was completed, to ligate multiple parts together and complete the CDS while inserting it in an L0 acceptor vector. Throughout the previous steps, recognition sites for BsaI and/or BpiI were removed from the CDS. All constructs were validated by Sanger sequencing upon completion. Finally, all constructs were completed by adding the following MoClo modules; 35s Cauliflower Mosaic Virus + 5'UTR Tobacco mosaic virus (pICH51266), GFP (*A. victoria*) (pICSL50008) and the OCS1 terminator (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 μ M acetosyringone in 10 mM 2-(*N*-morpholino) ethanesulfonic acid (MES), pH 5.6, 10 mM MgCl₂. 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 (OD₆₀₀).

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 μ L of H₂O using individual cells of a white 96-well white bottom plate (BRANDplates 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 H₂O or 1 μ M 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 tubes³¹. 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 toolkit⁷⁷. 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

The LRR (C1-3 domain, Fig. 4) were extracted from all LRR-RLP sequences included in the earlier mentioned phylogenetic analysis (Fig. 3, Suppl. File 4). The LRR domain and subdomains of VuINR were earlier identified using LRRfinder^{31,79}. Phylogenetic trees were built using MEGA X software⁸⁰, and bootstrap method based on 1000 iterations. A codon-based 2172-nucleotide-long alignment was generated using MUSCLE⁸¹. NJ clustering method was used for constructing the codon-based tree on Maximum Composite Likelihood substitution models. The ML tree was calculated using the GTR+G+I submodel as implemented in MEGA X software⁸⁰. The resulting ML tree was used for the ASR of selected nodes of interest. Joint and marginal ASR were performed with FastML software using Jukes and Cantor substitution model for nucleotides, gamma distribution, and 90% probability cutoff⁸². Finally, the sequences were domesticated to facilitate MoClo cloning by removing the BsaI and BpII cut sites (Suppl. File 4).

Construction of receptors with an ancestral reconstructed LRR domain

Ancestral reconstructed LRR-RLP constructs were generated using the MoClo toolkit⁷⁷. MoClo overhangs were designed in such a way that the AB domain and the DEFG domain amplified from *Vigun07g219600* could be ligated with the synthesized ancestral reconstructed LRR domain (C domain), resulting 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, the complete CDS stored in L0 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 the ancestral reconstructed LRR-RLPs in *N. benthamiana* was validated with western blot.

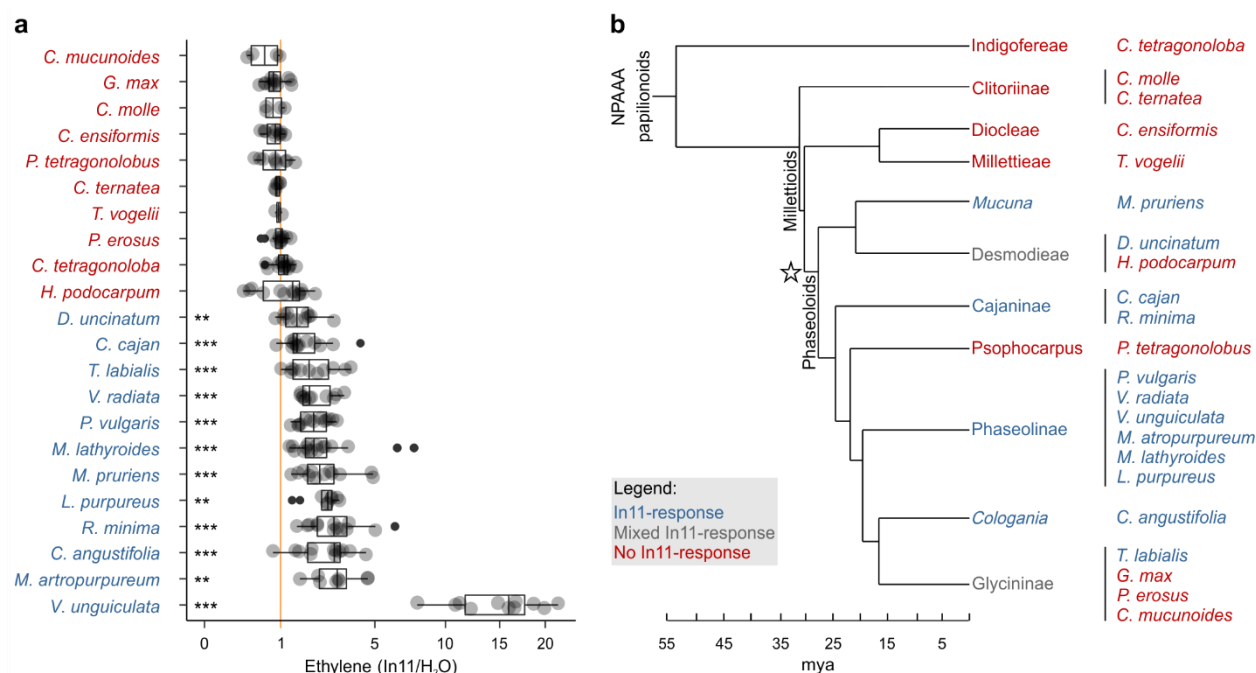
Acknowledgments

S.S. is supported as a Belgian American Educational Foundation postdoctoral fellow and the Mary Race Bevis Postdoc Research Award. S.S. and A.D.S. are supported by start-up funding from the University of Washington. A.D.S. is a Distinguished Investigator of the Washington Research Foundation.

Contributions

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

554 Main figures



555

556 **Fig. 1: Induced ethylene response to In11 is limited to Phaseoloid legumes. A)** Individual trifoliolate

557 leaflets were scratch wounded and treated with 1 μ M In11 or H₂O. The ratio of ethylene production for

558 leaflets within the same leaf is shown (x-axis). The vertical orange line shows a ratio equal to one, i.e. no

559 In11-induced ethylene burst. Biologically replicated plants are shown as separate dots. Significant

560 differences between the control and the treatment of interest are indicated (paired Wilcoxon signed-rank

561 test; ns non-significant, * p ≤ 0.05, ** p ≤ 0.01 and *** p ≤ 0.001). Plant species names are colored blue

562 (significant response to In11) or red (insignificant response). Species/accessions and resulting response

563 data can respectively be found in Suppl. Table 1 and Suppl. Table 2. **B)** A summary chronogram

564 representing time-based phylogenetic relationships within the tested lineages, with colors representing In11

565 response phenotypes as in (A). The star symbol indicates the node containing all In11-responsive species

566 at the base of Phaseoloid legumes. Divergence time is shown in million years ago (mya) and represents a

567 composite average^{83–86}.

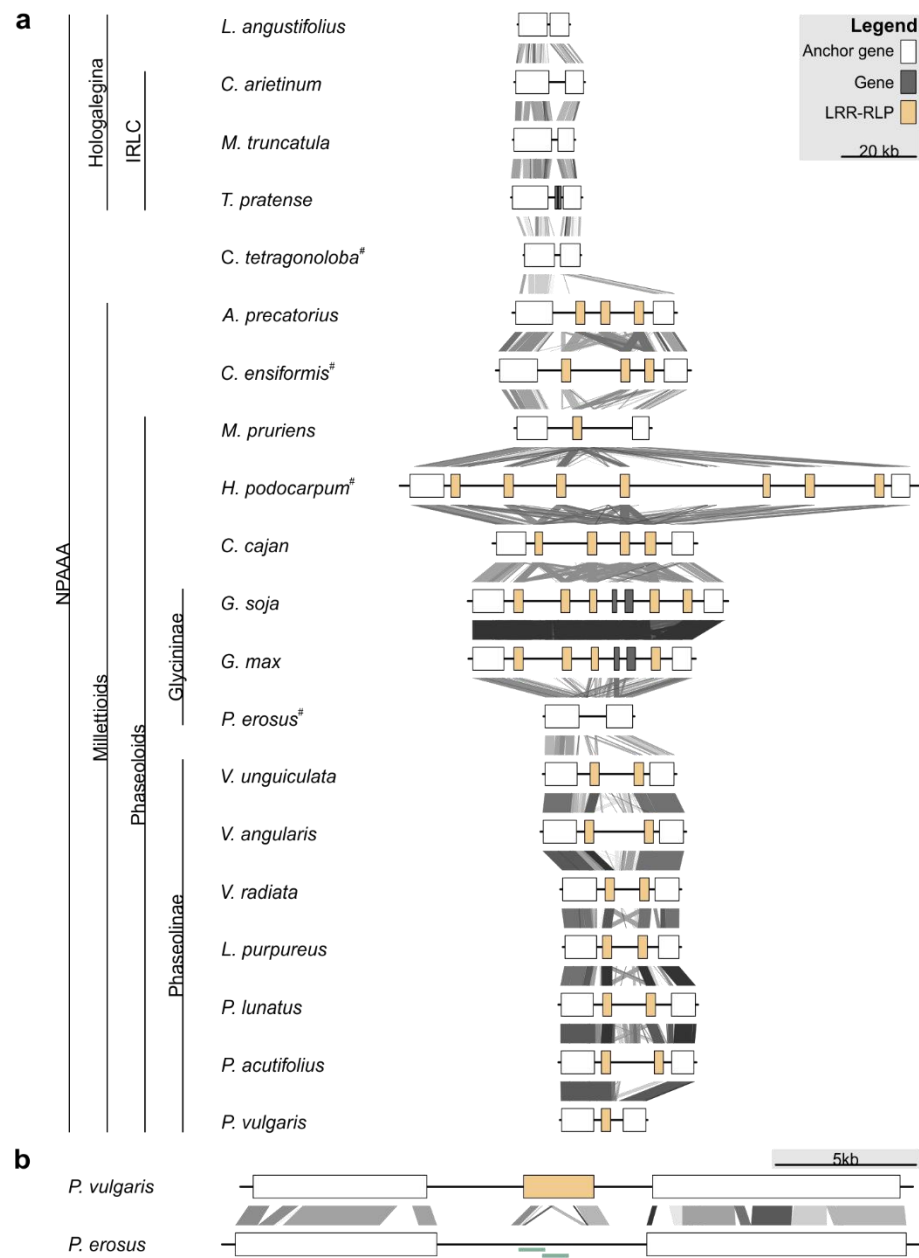


Fig. 2: LRR-RLP copy number variation at the *INR* locus in Millettoid and non-Millettoid legume genomes. Anchor, LRR-RLP, and other genes are colored as per legend. **A)** Locus comparison of the contiguous *INR* locus of twenty NPAAA papilionoid species. Blast hits between loci are indicated with lines (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.

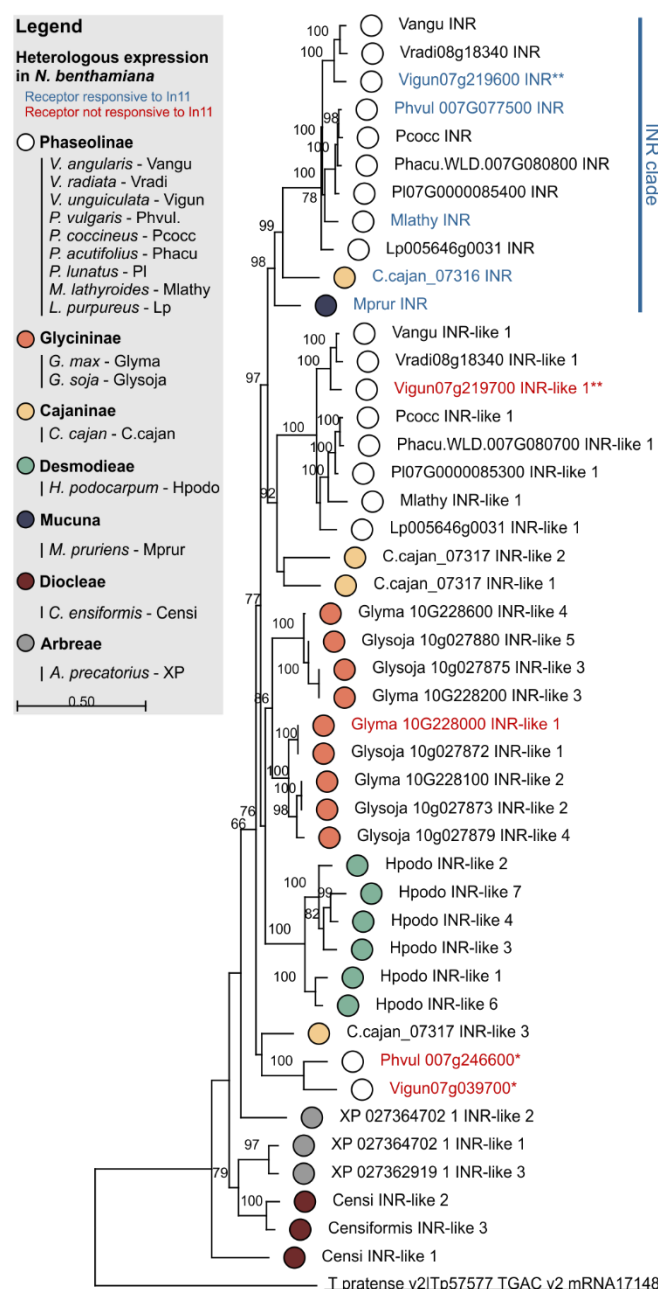


Fig. 3: Phylogenetic analysis of LRR-RLPs at the contiguous *INR* locus reveals a clade of functional receptors. LRR-RLPs from sixteen Millettoid species are shown. Maximum likelihood analysis bootstrap values are indicated, only values higher than 65 are shown. The scale bar represents 0.5 AA substitutions 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.

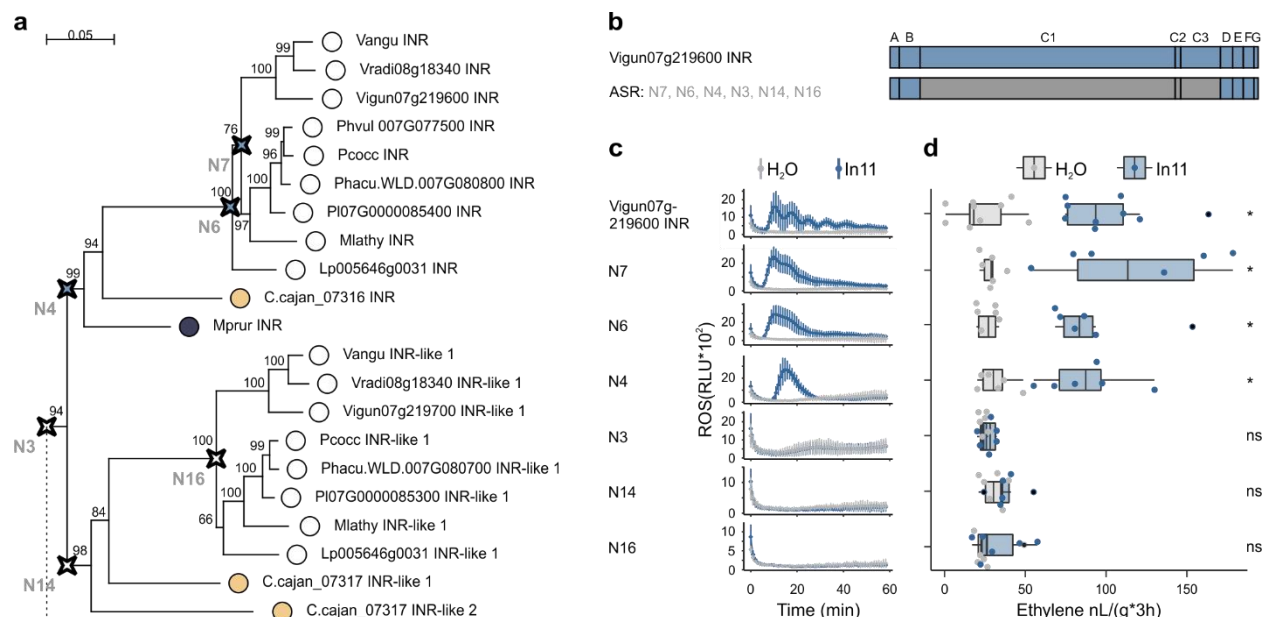
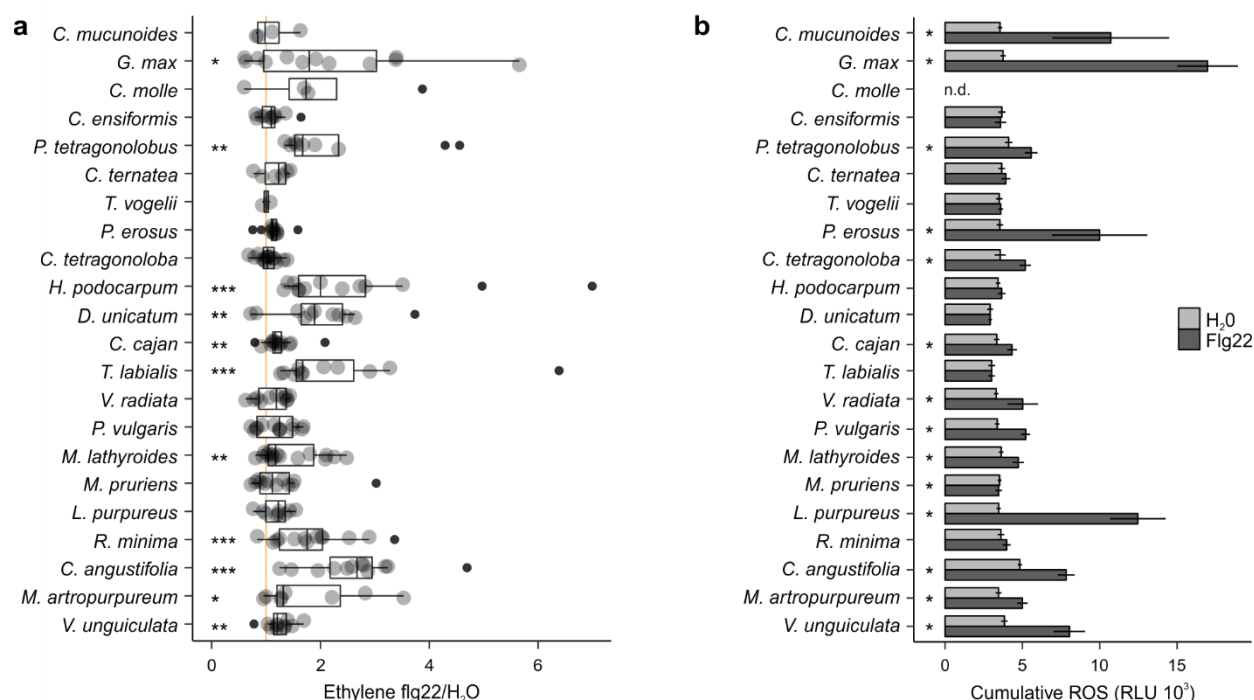


Fig. 5: Functional analysis of ancestrally reconstructed LRR domains of INR and INR-like receptors.
A) Part of the phylogenetic analyses of the LRR (C1-3 domain) of LRR-RLPs from the contiguous *INR* locus used for the ASR, full figure can be found in Suppl. Fig. 5. Ancestrally reconstructed nodes are 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, *VuINR*) 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 H₂O (grey), or the peptide In11 (1 μ M, blue). Curves indicate mean $\pm$ 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 H₂O (grey) or the peptide In11 (1 μ M, blue). Dots represent independent biological replicates (n $\geq$ 6 plants). Significance was tested by performing a paired Wilcoxon signed-rank test (ns non-significant, * $p \leq 0.05$).

623 Supplemental Figures



624

625 **Suppl. Fig. 1: Flg22-induced ROS and ethylene responses are idiosyncratic across Millettoid and**

626 **non-Millettoid legume species. A)** Individual trifoliate leaflets were scratch wounded and treated with 1

627 μ M flg22 or H₂O. The ratio of ethylene production for leaflets within the same leaf is shown (x-axis). Species

628 are listed in order of $\ln_{11}/H_2O$ response as per Fig. 1. The vertical orange line shows a ratio equal to one,

629 *i.e.* no induced ethylene response to $\ln_{11}$. Biological replicate plants are shown as separate dots.

630 Significant differences between the control and the treatment of interest are indicated (paired Wilcoxon

631 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

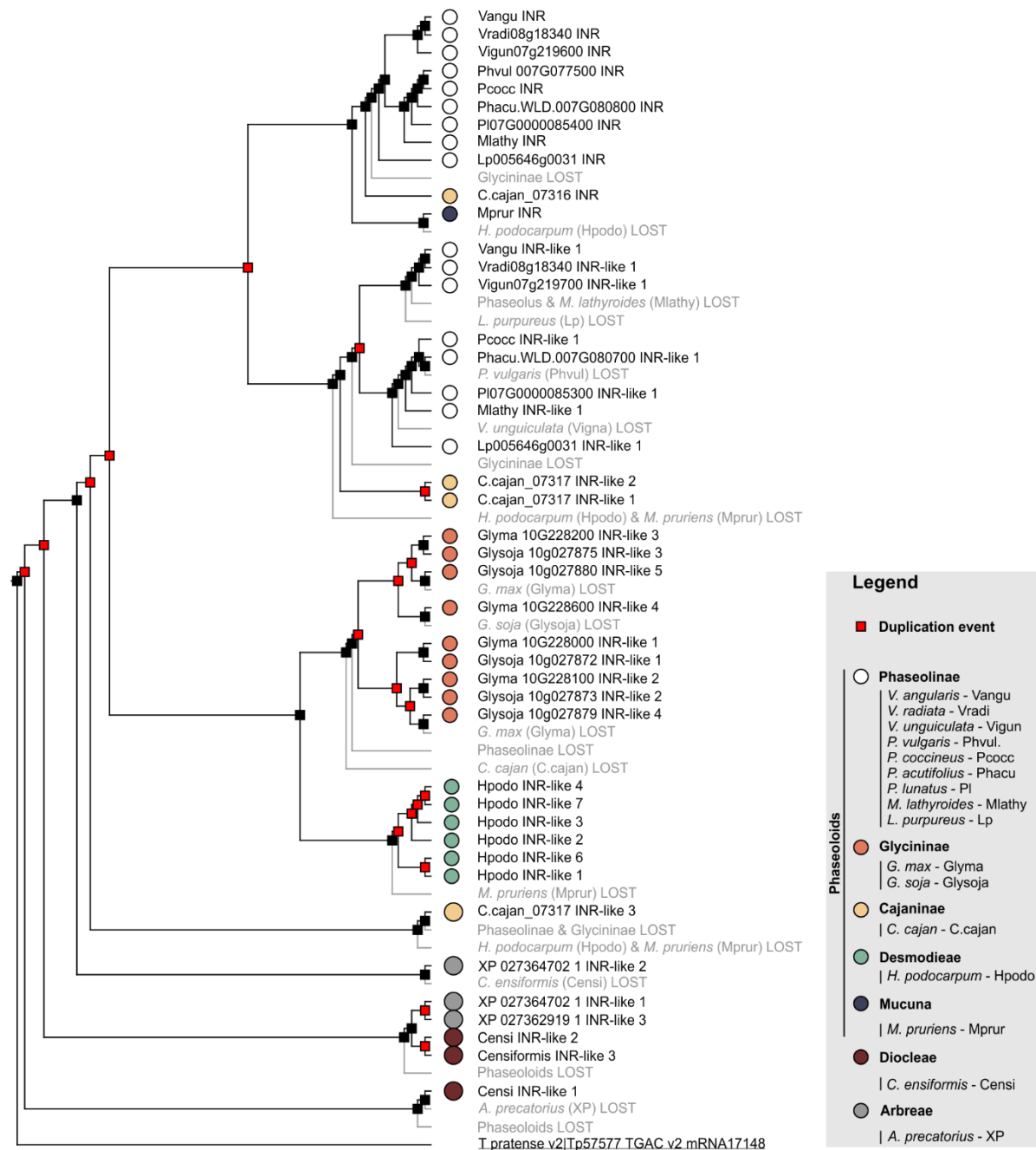
632 second marker of plant immunity response upon application of flg22. Shown is cumulative ROS data (3-

633 60min) of four technical replicates in relative luminescence units (RLU) over 1 hour (1 observation/minute)

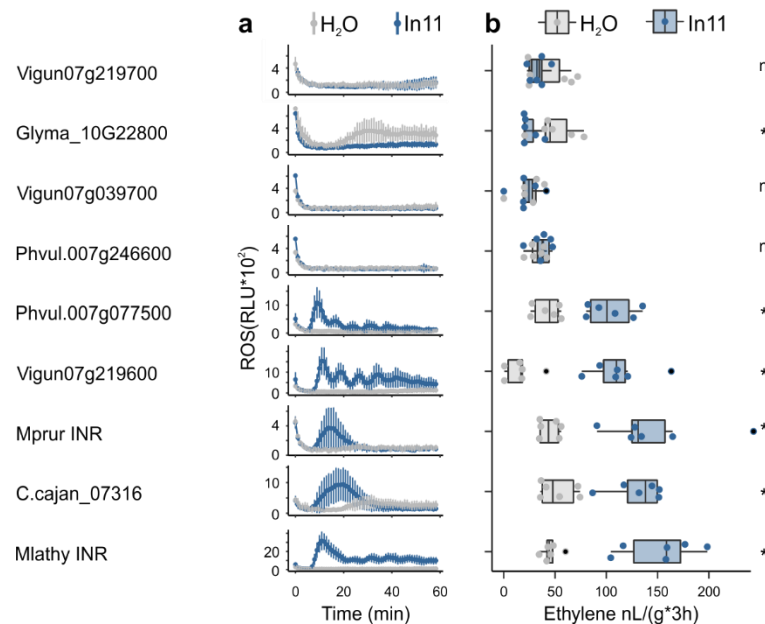
634 after treatment with H₂O, or the peptide flg22 (1 μ M). No ROS data was determined for *Centrosema molle*

635 (n.d.). Significant differences between the control and the treatment of interest were found by performing a

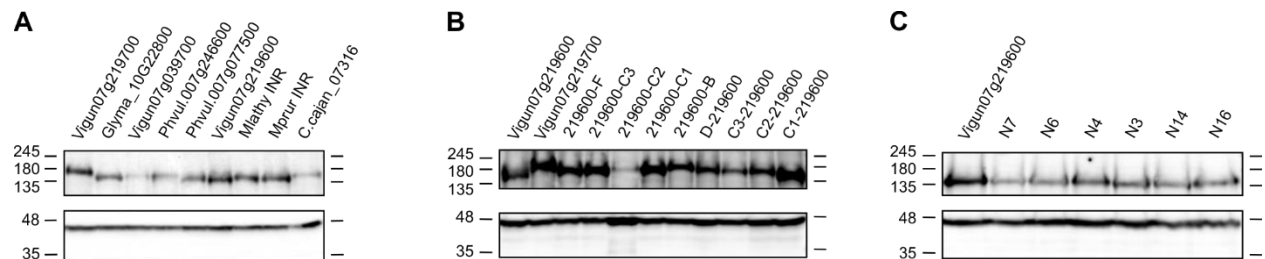
636 2-group Mann-Whitney U Test (* $p \leq 0.05$).



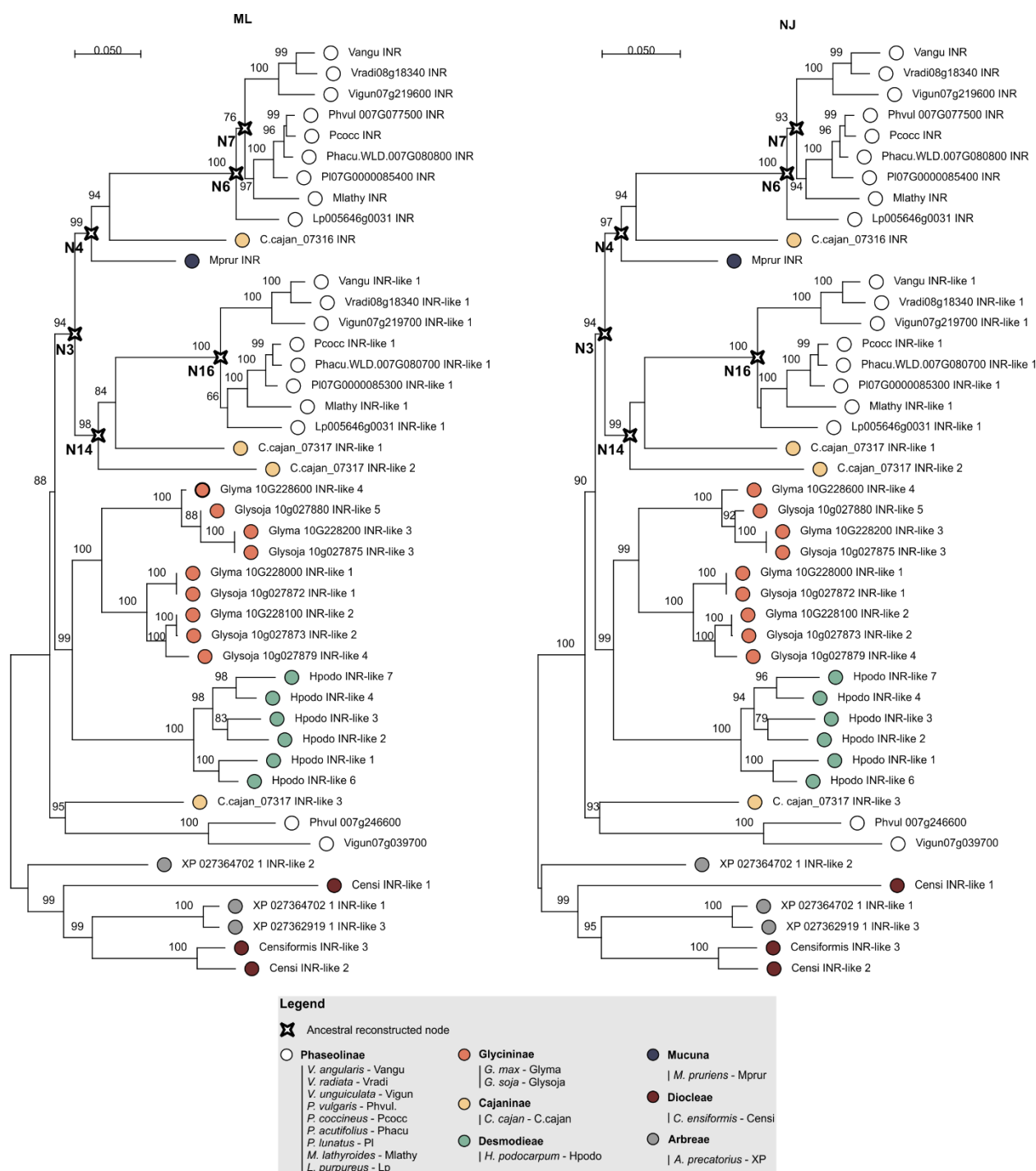
Suppl. Fig. 2: Prediction of gene duplication and gene loss events throughout the evolution of the contiguous *INR* locus in the Millettioideae. Gene tree-species tree reconciliation to identify the duplication and loss events at each branch using Notung⁴³. Predicted duplication events are marked with 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.



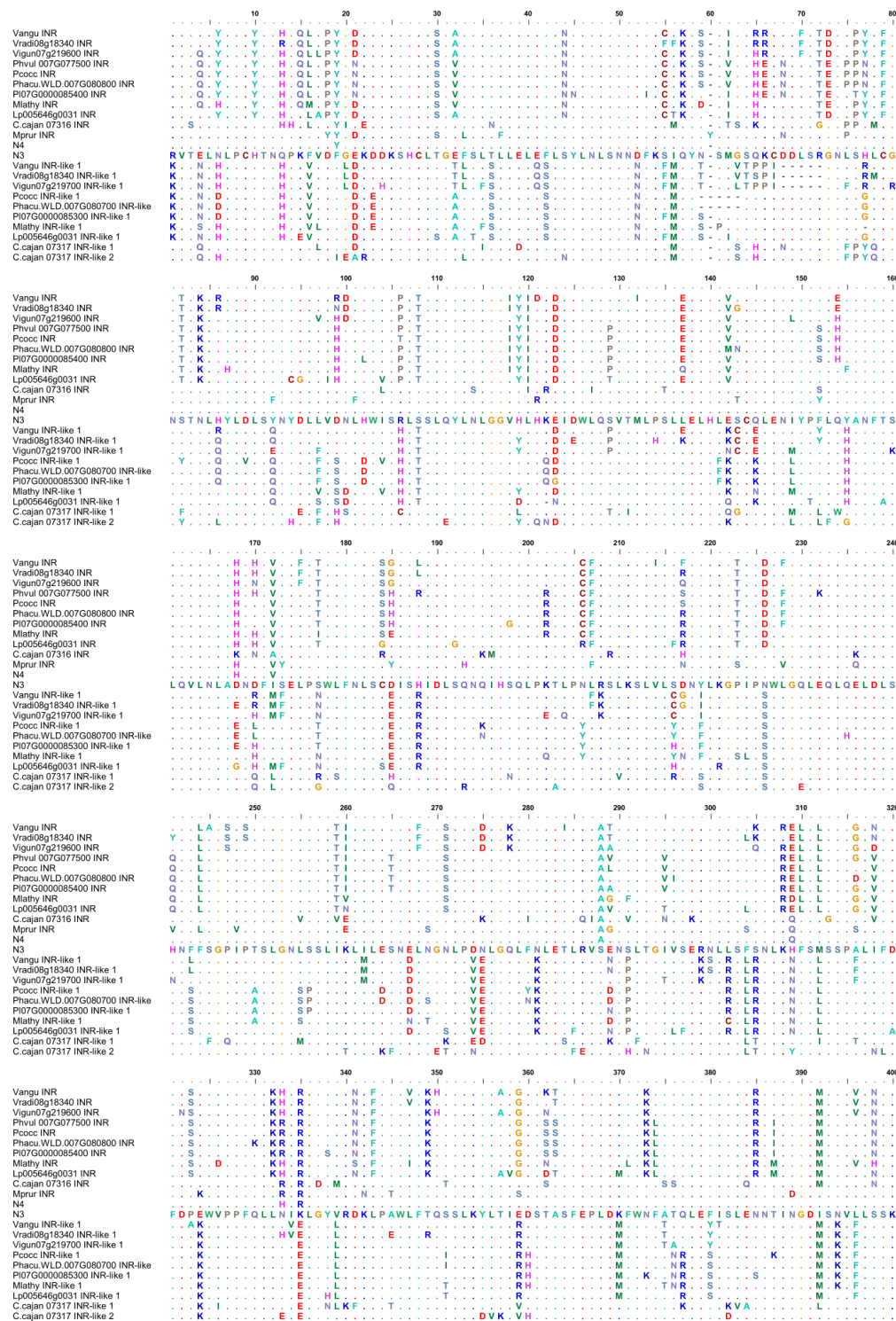
Suppl. Fig. 3: In11-induced ROS and ethylene production conferred by INR and INR-like after heterologous expression. In11-dependent plant immunity response data following heterologous expression of multiple receptor constructs in *N. benthamiana*. **A)** Shown are cumulative relative luminescence units (RLU) after treatment with H₂O (grey), or the peptide In11 (1 μ M, blue). Curves indicate mean $\pm$ SD for four independent biological replicates (n=4 plants), with each biological replicate representing six technical replicates. **B)** The x-axis shows the amount of ethylene released after infiltration with 1 μ M In11 (blue) or water (grey). Dots represent independent biological replicates (n=6 plants). Significance was tested by performing a paired Wilcoxon signed-rank test (ns non-significant, * p $\leq$ 0.05).



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



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 bootstraps⁸⁰. Maximum likelihood (ML – left side) and neighbor joining (NJ – right side) trees were calculated based on all codon positions of a codon-based alignment⁸¹. 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.



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., Diévert, 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

- repertoires in plant genomes. *bioRxiv* 2022.01.01.474684 (2022)
doi:10.1101/2022.01.01.474684.
14. Jamieson, P. A., Shan, L. & He, P. Plant cell surface molecular cypher: receptor-like proteins and their roles in immunity and development. *Plant Sci.* **274**, 242–251 (2018).
15. Tang, P. *et al.* Disease resistance signature of the leucine-rich repeat receptor-like kinase genes in four plant species. *Plant Sci.* **179**, 399–406 (2010).
16. Pruitt, R. N. *et al.* The EDS1–PAD4–ADR1 node mediates Arabidopsis pattern-triggered immunity. *Nat.* **2021** 5987881 **598**, 495–499 (2021).
17. Steinbrenner, A. D. The evolving landscape of cell surface pattern recognition across plant immune networks. *Curr. Opin. Plant Biol.* **56**, 135–146 (2020).
18. Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J. & Jones, J. D. G. Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science* (80-.). **266**, 789–793 (1994).
19. Thomas, C. M. *et al.* Characterization of the tomato Cf-4 gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. *Plant Cell* **9**, 2209–2224 (1997).
20. Kruijt, M., Kip, D. J., Joosten, M. H. A. J., Brandwagt, B. F. & De Wit, P. J. G. M. The Cf-4 and Cf-9 resistance genes against *Cladosporium fulvum* are conserved in wild tomato species. <http://dx.doi.org/10.1094/MPMI-18-1011> **18**, 1011–1021 (2007).
21. Dixon, M. S. *et al.* The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* **84**, 451–459 (1996).
22. Luderer, R., Takken, F. L. W., De Wit, P. J. G. M. & Joosten, M. H. A. J. *Cladosporium fulvum* overcomes Cf-2-mediated resistance by producing truncated AVR2 elicitor proteins. *Mol. Microbiol.* **45**, 875–884 (2002).
23. Krüger, J. *et al.* A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis. *Science* (80-.). **296**, 744–747 (2002).
24. Rooney, H. C. E. *et al.* Plant science: *Cladosporium* Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science* (80-.). **308**, 1783–1786 (2005).
25. Zhang, L. *et al.* Fungal endopolygalacturonases are recognized as microbe-associated molecular patterns by the Arabidopsis receptor-like protein RESPONSIVENESS TO BOTRYTIS POLYGALACTURONASES1. *Plant Physiol.* **164**, 352–364 (2014).
26. Zhang, L. *et al.* Distinct immune sensor systems for fungal endopolygalacturonases in closely related Brassicaceae. *Nat. Plants* **2021** 79 **7**, 1254–1263 (2021).
27. Albert, I. *et al.* An RLP23–SOBIR1–BAK1 complex mediates NLP-triggered immunity. *Nat. Plants* **2015** 110 **1**, 1–9 (2015).
28. Jehle, A. K. *et al.* The receptor-like protein ReMAX of Arabidopsis detects the microbe-associated molecular pattern eMax from *Xanthomonas*. *Plant Cell* **25**, 2330–2340 (2013).

29. Hegenauer, V. *et al.* The tomato receptor CuRe1 senses a cell wall protein to identify *Cuscuta* as a pathogen. *Nat. Commun.* 2020 **11**, 1–7 (2020).
30. Hegenauer, V. *et al.* Detection of the plant parasite *Cuscuta reflexa* by a tomato cell surface receptor. *Science* (80-.). **353**, 478–481 (2016).
31. Steinbrenner, A. D. *et al.* A receptor-like protein mediates plant immune responses to herbivore-associated molecular patterns. *Proc. Natl. Acad. Sci.* **117**, 31510–31518 (2020).
32. Schmelz, E. A. *et al.* Fragments of ATP synthase mediate plant perception of insect attack. *Proc. Natl. Acad. Sci.* **103**, 8894–8899 (2006).
33. Schmelz, E. A. *et al.* An amino acid substitution inhibits specialist herbivore production of an antagonist effector and recovers insect-induced plant defenses. *PLANT Physiol.* **160**, 1468–1478 (2012).
34. Schmelz, E. A., LeClere, S., Carroll, M. J., Alborn, H. T. & Teal, P. E. A. Cowpea chloroplastic ATP synthase is the source of multiple plant defense elicitors during insect herbivory. *Plant Physiol.* **144**, 793–805 (2007).
35. Albert, I., Zhang, L., Bemm, H. & Nürnberger, T. Structure-function analysis of immune receptor AtRLP23 with its ligand nlp20 and coreceptors AtSOBIR1 and AtBAK1. *Mol. Plant-Microbe Interact.* **32**, 1038–1046 (2019).
36. Delaux, P. M. *et al.* Reconstructing trait evolution in plant evo–devo studies. *Curr. Biol.* **29**, R1110–R1118 (2019).
37. Białas, A. *et al.* Two nlr immune receptors acquired high-affinity binding to a fungal effector through convergent evolution of their integrated domain. *Elife* **10**, e66961 (2021).
38. Read, A. C. *et al.* Genome assembly and characterization of a complex zfBED-NLR gene-containing disease resistance locus in Carolina Gold Select rice with Nanopore sequencing. *PLoS Genet.* **16**, e1008571 (2020).
39. Michael, T. P. & VanBuren, R. Building near-complete plant genomes. *Curr. Opin. Plant Biol.* **54**, 26–33 (2020).
40. Goodin, M. M., Zaitlin, D., Naidu, R. A. & Lommel, S. A. *Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions. *Mol. Plant-Microbe Interact.* **21**, 1015–1026 (2008).
41. Cardoso, D. *et al.* Filling in the gaps of the papilionoid legume phylogeny: The enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging Amburaneae clade. *Mol. Phylogenet. Evol.* **84**, 112–124 (2015).
42. Zhao, Y. *et al.* Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Mol. Plant* **14**, 748–773 (2021).
43. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: A program for dating gene duplications and optimizing gene family trees. <https://home.liebertpub.com/cmb> **7**, 429–447 (2004).
44. Liu, Y. *et al.* Pan-genome of wild and cultivated soybeans. *Cell* **182**, 162–176.e13 (2020).
45. Randall, R. N., Radford, C. E., Roof, K. A., Natarajan, D. K. & Gaucher, E. A. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat.*

Commun. **7**, 1–6 (2016).

46. Schultink, A. & Steinbrenner, A. D. A playbook for developing disease-resistant crops through immune receptor identification and transfer. *Curr. Opin. Plant Biol.* **62**, 102089 (2021).
47. Lacombe, S. *et al.* Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nat. Biotechnol.* **28**, 365–369 (2010).
48. Pfeilmeier, S. *et al.* Expression of the *Arabidopsis thaliana* immune receptor EFR in *Medicago truncatula* reduces infection by a root pathogenic bacterium, but not nitrogen-fixing rhizobial symbiosis. *Plant Biotechnol. J.* **17**, 569–579 (2019).
49. Kourelis, J. *et al.* Evolution of a guarded decoy protease and its receptor in solanaceous plants. *Nat. Commun.* **11**, 1–15 (2020).
50. Shi, T., Huang, H., Sanderson, M. J. & Tax, F. E. Evolutionary dynamics of leucine-rich repeat receptor-like kinases and related genes in plants: a phylogenomic approach. *J. Integr. Plant Biol.* **56**, 648–662 (2014).
51. Restrepo-Montoya, D., McClean, P. E. & Osorno, J. M. Orthology and synteny analysis of receptor-like kinases “RLK” and receptor-like proteins “RLP” in legumes. *BMC Genomics* **22**, 1–17 (2021).
52. Wang, L. *et al.* The pattern-recognition receptor CORE of Solanaceae detects bacterial cold-shock protein. *Nat. Plants* **2**, 1–9 (2016).
53. Zipfel, C. *et al.* Perception of the bacterial PAMP EF-Tu by the receptor EFR Restricts *Agrobacterium*-mediated transformation. *Cell* **125**, 749–760 (2006).
54. Wei, Y. *et al.* An immune receptor complex evolved in soybean to perceive a polymorphic bacterial flagellin. *Nat. Commun.* **11**, 1–11 (2020).
55. Fürst, U. *et al.* Perception of *Agrobacterium tumefaciens* flagellin by FLS2XL confers resistance to crown gall disease. *Nat. Plants* **6**, 22–27 (2020).
56. Sun, Y. *et al.* Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. *Science* (80-.). **342**, 624–628 (2013).
57. Wan, W. L. *et al.* Comparing Arabidopsis receptor kinase and receptor protein-mediated immune signaling reveals BLK1-dependent differences. *New Phytol.* **221**, 2080–2095 (2019).
58. Felix, G., Duran, J. D., Volko, S. & Boller, T. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* **18**, 265–276 (1999).
59. Schmelz, E. A., Alborn, H. T., Banchio, E. & Tumlinson, J. H. Quantitative relationships between induced jasmonic acid levels and volatile emission in *Zea mays* during *Spodoptera exigua* herbivory. *Planta* **216**, 665–673 (2003).
60. R Development Core Team. R: a language and environment for statistical computing. *R Found. Stat. Comput.* **1**, 409 (2015).
61. Wickham, H. & Francois, R. dplyr: a grammar of data manipulation. **3**, 156 (2015).
62. Wickham, H. ggplot2 elegant graphics for data analysis. *Media* **35**, 211 (2009).
63. Lutz, K. A., Wang, W., Zdepski, A. & Michael, T. P. Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing 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