Evolutionary Trajectory of Pattern Recognition Receptors in Plants

Bruno Pok Man Ngou[1], Michele Wyler[2], Marc W Schmid[2], Yasuhiro Kadota[1,#], Ken Shirasu[1,#]

[1]RIKEN Center for Sustainable Resource Science, Yokohama, Japan

[2]MWSchmid GmbH, 8750 Glarus, Switzerland

[#]Correspondence to: yasuhiro.kadota@riken.jp, ken.shirasu@riken.jp

Abstract

Plants perceive pathogen-associated molecular patterns (PAMPs) via pattern recognition receptors (PRRs) to activate PRR-triggered immunity (PTI). Despite extensive research on PTI in model plant species, the evolutionary trajectory and emergence of PRRs remain elusive. Here we conducted a comparative genomic analysis of cell-surface receptors and downstream signalling components among 350 plant species. Our findings reveal that cell-surface receptors comprise two major classes, receptor-like proteins (RLPs) and receptor-like kinases (RLKs), with RLP being more ancient whereas RLK families have undergone significant expansion. We also demonstrate that multiple downstream signalling components have an ancient origin within the plant lineage. To shed light on the immune-specificity of PRRs, we traced the evolutionary origin of immune-specific leucine-rich repeat-RLPs (LRR-RLPs) in plants. Surprisingly, we discovered that the last four LRR motifs crucial for co-receptor interaction in LRR-RLPs are closely related to those of the LRR-RLK subgroup Xb, which primarily governs growth and development. Functional characterisation further reveals that LRR-RLPs initiate immune responses through their juxtamembrane and transmembrane regions, while LRR-RLK-Xb members regulate development through their cytosolic kinase domains. Our data suggest modular evolution of cell-surface receptors in which immunity- and development-specific cell-surface receptors share a common origin. After diversification, their ectodomains, juxtamembrane, transmembrane, and cytosolic regions have either diversified or stabilised to recognize ligands that activate different downstream responses. We propose that cell-surface receptors and downstream signalling components are ancient, and likely predate the emergence of land plants, subsequently evolving to exhibit greater complexity and specificity within the land plant lineage.

Introduction

Recognition of pathogen-associated molecular patterns (PAMPs) via pattern-recognition receptors (PRRs) is a fundamental mechanism which allows organisms to detect the presence of pathogens[1,2]. PAMP/PRR-triggered immunity (PTI) is conserved in animals, plants, fungi, and other eukaryotes[3–5]. Plant PRRs play a crucial role in perceiving various PAMPs from pathogens, including peptides, small proteins, lipids, and polysaccharides[2,6]. These receptors can be categorised as either receptor-like proteins (RLPs), which possess a transmembrane domain, or receptor-like kinases (RLKs), which have both transmembrane and kinase domains. The number of RLP and RLK gene families varies among plant species, and is believed to expand over evolutionary time in response to pathogen pressure[7,8]. To date, around seventy plant PRRs with known elicitors have been characterised, primarily in angiosperms[6], while those in other clades of Viridiplantae remain poorly understood.

Upon ligand or elicitor perception, PRRs undergo dimerization or form heteromeric complexes with other cell-surface receptors. This spatial arrangement brings the cytoplasmic kinase domains of PRRs in close proximity, initiating a cascade of auto- and trans-phosphorylation events[9]. The activated receptor complex subsequently phosphorylates members of the cytoplasmic receptor-like kinases subgroup VII (RLCK-VII)[10], which, in turn, phosphorylate various cytoplasmic kinases, such as the mitogen-activated protein kinase kinase kinases (MAPKKKs), and plasma membrane-associated proteins, such as calcium channels and NADPH oxidases[9]. The phosphorylation of these proteins collectively triggers transcriptional reprogramming and physiological changes, such as cytoplasmic calcium influx and the accumulation of reactive oxygen species (ROS)[11]. These physiological responses effectively hinder pathogen proliferation during infection (Extended Figure 1). The PTI-signalling pathway has been extensively characterised in model plant species such as Arabidopsis, rice, and tomato.

However, the origins of PRRs and PTI-signalling components remain unclear, and the evolutionary relationships between receptors and downstream signalling components have yet to be determined.

While many cell-surface receptors are involved in pathogen recognition, many with similar domain architectures are also engaged in other biological processes (Extended figure 2). Some of these receptors recognise endogenous molecules, such as phytohormones and phytocytokines, to regulate developmental and reproductive processes[12,13]. Given the striking resemblance in their domain architecture, it is reasonable to infer that immunity-and developmental-related cell-surface receptors share a common origin. However, the evolutionary trajectory that led to their divergence and specialisation in distinct biological processes remains poorly understood.
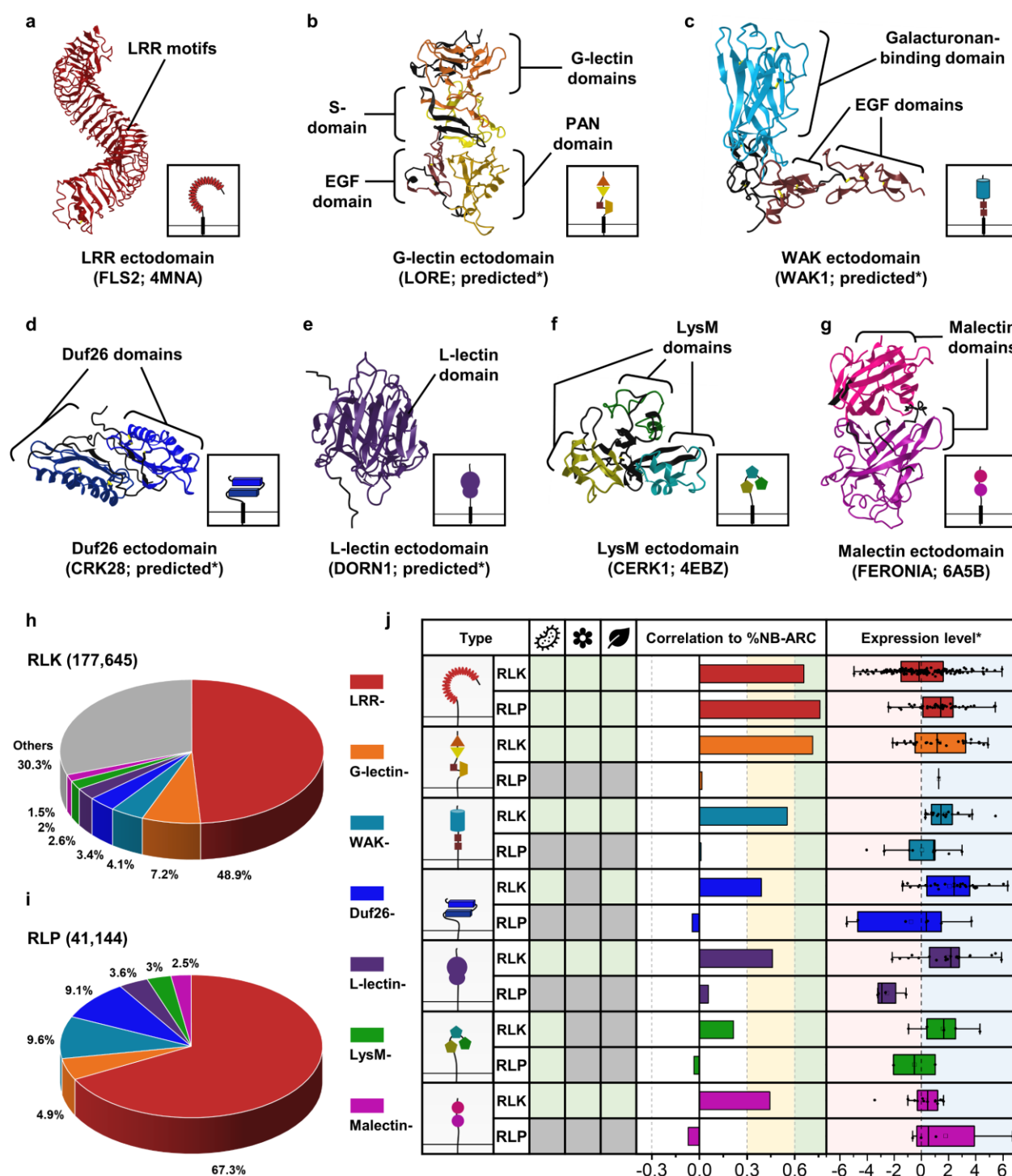

Results

Diversity and immune relevance of cell-surface receptors in Viridiplantae

The PRRs known to participate in immunity include LRR-RLKs, LRR-RLPs, G-lectin-RLKs, Wall-associated kinase (WAK)-RLKs, WAK-RLPs, Domain of Unknown Function 26 (Duf26)-RLKs (or cysteine-rich kinases; CRKs), L-lectin-RLK, Lysin motif (LysM)-RLKs, and Malectin-RLKs[6]. LRR-RLKs are further classified into 20 subgroups based on their kinase domains, with subgroup XII specifically implicated in PAMP recognition[14]. PRRs with LRR-, G-lectin-, WAK-, and LysM-ectodomains recognise PAMPs, while others perceive self-molecules or unidentified ligands. These ligands encompass a wide range of substances such as small molecules, peptides, full-length proteins, lipids, and polysaccharides/glycans (Extended figure 2). Recognition of the diverse array of ligands is likely to be accomplished by variable structures and combinations of different ectodomains (Main figure 1a-g).

To investigate the phylogenetic relationships of the cell-surface receptor-gene family in Viridiplantae, we extracted RLKs containing both a transmembrane (TM) and a kinase domain (KD) from 350 publicly available genomes as previously reported[8]. In total, we identified approximately 177,645 RLKs meeting this criterion. We subsequently pinpointed RLKs with LRR-, G-lectin-, WAK-, Duf26-, L-lectin-, LysM-, and Malectin-ectodomains. Among these, the most prominent RLK family are the LRR-RLKs, which account for 48.9% of the identified RLKs, followed by G-lectin-RLKs (7.2%), WAK-RLKs (4.1%), Duf26-RLKs (3.4%), L-lectin-RLKs (2.6%), LysM-RLKs (2%) and Malectin-RLKs (1.5%) (Main figure 1h). Collectively, these seven families constitute nearly 70% of all identified RLKs. Previous reports have suggested a positive correlation between the gene family sizes of cell-surface immune receptors (LRR-RLK subgroup XII and LRR-RLPs) and intracellular immune receptors (NLRs, NB-ARC family) across the angiosperms[8]. To test this observation, we examined the correlation between the percentage (%; number of identified genes / numbers of searched genes × 100) of the seven RLK families and the NB-ARC family in each genome. Notably, both %LRR-RLK and %G-lectin-RLK exhibit robust positive correlations (Pearson's r > 0.6) with %NB-ARC (Main figure 1j). Additionally, %WAK-RLK, %Duf26-RLK, %L-lectin-RLKs and %Malectin-RLKs demonstrate moderate positive correlations (0.6 > Pearson's r > 0.3) with %NB-ARC (Main figure 1j). Thus, it is likely that members of these RLK families play significant roles in immunity.

Next, we extracted protein sequences with TM and LRR-/G-lectin-/WAK-/Duf26-/L-lectin-/LysM/Malectin-ectodomains but lacking KDs (RLPs). In total, we identified 41,144 RLPs. Again, the largest RLP family comprises LRR-RLKs (67.3%), followed by WAK-RLPs (9.6%), Duf26-RLPs (9.1%), G-lectin-RLPs (4.9%), L-lectin-RLPs (3.6%), LysM-RLPs (3%) and Malectin-RLPs (2.5%) (Main figure 1i). The RLK repertoire surpasses that of RLPs, indicating either the expansion of RLK families, presumably by gene duplication, or contraction of RLP families throughout the plant lineage. We therefore conducted correlation analyses to assess the relationship between the percentages of the seven RLP families and NB-ARC family in each genome. Surprisingly, with the exception of LRR-RLPs (Pearson's r = 0.759), the sizes of the remaining six RLP families exhibit negligible correlations with %NB-ARC (Pearson's r < 0.1) (Main figure 1j), suggesting that, apart from LRR-RLPs, the majority of members in these six RLP families are unlikely to be involved in immunity. This observation aligns with the absence of reports regarding the role of G-lectin-RLPs, WAK-RLPs, Duf26-RLPs, L-lectin-RLPs, and Malectin-RLPs in immunity (Extended figure 2). To further test this supposition, we evaluated the expression levels of all 14 cell-surface receptor families during PTI and NLR-triggered immunity (NTI) in *Arabidopsis thaliana*[15,16]. Notably, RLKs, except for LRR- and Malectin-containing cell-surface receptors, generally exhibit higher expression levels compared to RLPs during immunity (Main figure 1j; supplementary figure 1).

2

Main figure 1. The distribution of cell-surface receptors in plants. Ectodomain structure of a) an LRR receptor, b) a G-lectin receptor, c) an L-lectin receptor, d) a LysM receptor, e) a Malectin receptor, f) a WAK receptor, and g) a Duf26 receptor. Structures of FLS2, CERK1 and FERONIA were published[17–19]. Structures of LORE, DORN1, WAK1 and CRK28 were predicted by Alphafold2*,[20]. Ectodomains are visualized in iCn3D[21]. **H)** Ectodomain distribution of RLKs in plants. Each fraction represents the percentage (%) of ectodomain out of all the RLKs from 350 species (177,645). **I)** Ectodomain distribution of RLP in plants. Each fraction represents the percentage (%) of ectodomains out of all the RLPs with those seven ectodomains (41,144). **J)** Table of RLKs and RLPs with LRR (red), G-lectin (orange), WAK (turquoise), Duf26 (blue), L-lectin (purple), LysM (green), and Malectin (magenta) ectodomains. Characterised receptors involved in microbial interaction (bacteria icon), reproduction (flower icon), and development (leaf icon) are indicated with light green boxes. Grey boxes
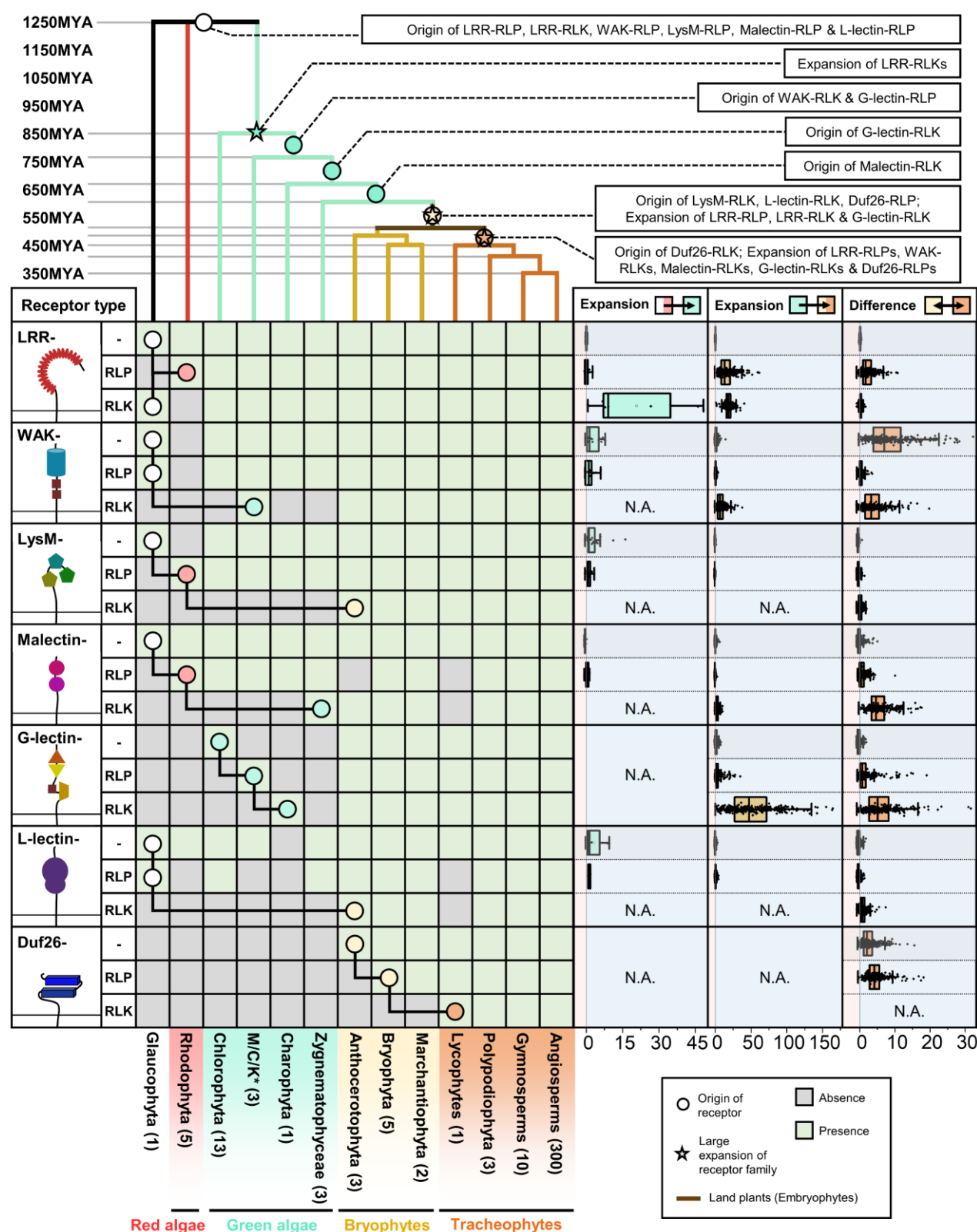
indicate that the receptor class has not been reported to be involved in that biological process. For details, refer to Extended figure 2. Correlations between different classes of cell-surface receptors and NB-ARC in 300 angiosperms are indicated with bars. Strong positive correlations are indicated by extension to the light green area (Pearson's r >0.6) and medium positive correlations are within the yellow area (Pearson's r between 0.3-0.6). Expression level* refers to the expression of each class of cell-surface receptors during NLR-triggered immunity (NTI) in *Arabidopsis thaliana*. Light blue area represents increased expression and light pink area represents decreased expression during NTI. X-axis values represent $\log_2$ (fold change during ETI relative to untreated samples). Boxplot elements: center line, median; bounds of box, 25th and 75th percentiles; whiskers, $1.5 \times$ IQR from 25th and 75th percentiles. RNA-seq data analysed here were reported previously, where NTI was activated by estradiol-induced expression of AvrRps4 in *A. thaliana* for 4 hours[16]. For the expression of each class of cell-surface receptors during PTI in *A. thaliana*, refer to supplementary figure 1h-n.

The origin and expansion of cell-surface receptors in the plant lineage

To trace the origins of different receptor classes within the plant lineage, we examined the presence of ectodomains (LRR-, G-lectin-, WAK-, Duf26-, L-lectin-, LysM- and Malectin-ectodomains lacking TM or KD), RLPs (TM-bound ectodomains) and RLKs (ectodomains encompassing both TM and KD) in Glaucophyta, red algae, green algae, Bryophytes and Tracheophytes (Main figure 2; Supplementary figure 1). Ectodomains exhibit an ancient heritage, with LRR-, WAK-, LysM-, Malectin-, and L-lectin-domains dating back to the era of Glaucophyta. Similarly, relatively ancient counterparts such as LRR-RLPs, WAK-RLPs, LysM-RLPs, Malectin-RLPs, and L-lectin-RLPs are found in both Glaucophyta and Rhodophyta. In contrast, RLKs emerged more recently. Green algae harbour WAK-RLKs, Malectin-RLKs, and G-lectin-RLKs, and LysM-RLKs, L-lectin-RLKs, and Duf-26-RLKs are exclusive to Embryophytes (Main figure 2).

Except for LRR-RLPs, all six families of RLP are basal to the RLK families. This suggests the intriguing possibility that numerous RLKs may have evolved directly from RLPs through the integration of kinase domains. Such integration events likely occurred at various stages of plant evolution, given the presence of RLKs in diverse plant lineages. Alternatively, it remains plausible that RLKs are equally ancient as RLPs, and that their absence from certain lineages is a result of loss over time. Nevertheless, it is highly probable that both RLPs and RLKs originated from ectodomains that underwent multiple gene duplication and domain-swapping events to incorporate TM and/or kinase domains, thereby diversifying their functional repertoire[22].

We also examined the expansion patterns of different receptor classes across various plant lineages. Our analysis involved in calculating the median percentage (%) of cell-surface receptors in i) Glaucophyta and Rhodophyta, ii) green algae, and iii) Bryophytes and then determining the percentage increase (% increase) from Glaucophyta and Rhodophyta to green algae; from green algae to Embryophytes; and between Bryophytes and Tracheophytes. We observed substantial expansions in specific receptor families across these lineages. Green algae exhibited a significant expansion of LRR-RLKs, while Embryophytes displayed expansions in LRR-RLPs, LRR-RLKs, WAK-RLKs, and G-lectin-RLKs. Tracheophytes had further expansions in LRR-RLPs, WAK-RLKs, Malectin-RLKs, G-lectin-RLKs and Duf26-RLPs (Main figure 2; Supplementary figure 1). Overall, RLKs demonstrate greater expansion compared to RLPs, with notable expansions observed in LRR-RLK, WAK-RLK, and G-lectin-RLK. These findings align with the substantial size of these receptor families and their involvement in recognising PAMPs (Main figure 1h; Extended figure 2a). In particular, the LRR-RLK-XII subgroup exhibits a considerably higher expansion rate compared to other subgroups (Extended figure 3; Supplementary figure 1), reinforcing the idea that cell surface immune receptors underwent extensive expansions as the plant lineage diversified and evolved to adapt to a wide range of environments. It is worth noting that LRR-RLKs and LRR-RLPs have been reported to recognise a number of different types of ligands such as peptides, full-length proteins, small molecules, and phytohormones, whereas WAK-RLKs perceive glycans and full-length proteins, and G-lectin-RLKs perceive lipids and full-length proteins (Extended figure 2a). Thus, it is plausible that these receptor families expanded throughout plant evolution to accommodate a broader range of ligands.
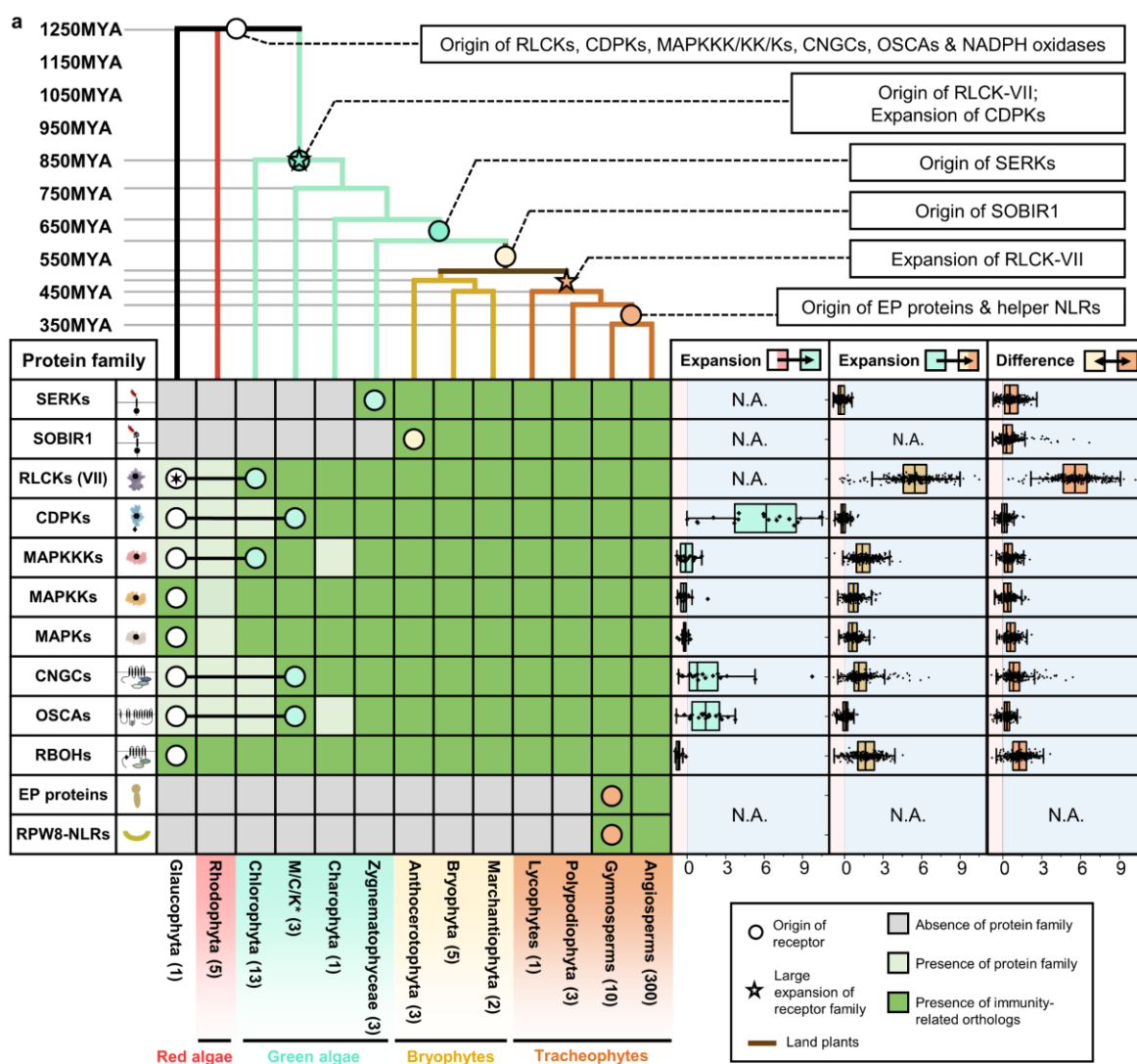
Main figure 2. The origin and expansion of cell-surface receptors in plants. The top panel represents a phylogenetic tree of multiple algal and plant lineages. Circles (○) and stars (☆) indicate the origin and expansion of receptor families. The timescale (in millions of years; MYA) of the phylogenetic tree was estimated by TIMETREE5[23]. The bottom panel represents the presence or absence of different receptor classes in algal and plant lineages. '-' represents ectodomains with no

transmembrane or kinase domain, 'RLP' represents ectodomains with a transmembrane domain but no kinase domain, 'RLK' represents ectodomains with both transmembrane and kinase domains. *M/C/K represents Mesostigmatophyceae, Chlorokybophyceae, and Klebsormidiophyceae. The number of species available from each algal and plant lineage are indicated by the numbers within respective boxes. A grey box indicates the absence of receptors and a green box indicates their presence in each lineage. The origin of a receptor is indicated with a circle (○). The origins of '-', 'RLP', and 'RLK' are connected by black lines. Expansion rates of receptor classes are indicated by boxplots. The cyan boxplot represents the expansion rate from Glaucophyta and Rhodophyta to green algae. The yellow boxplot represents the expansion rate from green algae to Embryophytes and the orange boxplot represents the expansion rate from early land plants to Tracheophytes. Light blue area represents expansion and light pink area represents contraction of the gene family. X-axis values represent expansion rate (×). Boxplot elements: center line, median; bounds of box, 25th and 75th percentiles; whiskers, $1.5 \times$ IQR from 25th and 75th percentiles.
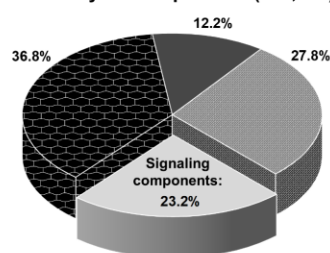
The origin and expansion of PTI-signalling components in the plant lineage

Upon ligand perception, PRRs engage in homodimerization or heteromeric complex formation with co-receptors, such as BAK1 (a member of the Somatic embryogenesis receptor-like kinases (SERK) family) and SOBIR1, to initiate downstream signalling cascades. Cytoplasmic kinases, such as RLCK-VIIs, Mitogen-activated protein kinases (MAPKs), and Calcium-dependent protein kinases (CDPKs), undergo phosphorylation to transmit immune signals. Likewise, PM-localised downstream signalling components, such as cyclic nucleotide-gated channels (CNGCs), hyperosmolality-gated calcium-permeable channels (OSCAs), and the respiratory burst oxidase homolog NADPH oxidases (RBOHs) are also phosphorylated to elicit a defense response[9]. We identified these co-receptors and signalling components from 350 genomes and determined their absence or presence across the plant lineage (Extended figure 4; Supplementary figure 2-26). SERKs, acting as cell-surface co-receptors for multiple LRR-RLKs and LRR-RLPs are present in Zygnematophyceae and Embryophytes[24] (Main figure 3a), suggesting their emergence during or prior to the appearance of land plants. Immune-related LRR-RLPs lack intracellular kinase domains, thus require another LRR-RLK co-receptor, SOBIR1, to activate downstream signalling[25]. Similar to BAK1, SOBIR1 is also present in Embryophytes (Main figure 3a). Thus, co-receptors for cell-surface receptors likely evolved during or before the emergence of land plants. On the other hand, cytoplasmic kinases (RLCKs, CDPKs, MAPKKKs, MAPKKs, and MAPKs) are ancient, as are the PM-localised downstream signalling components (CNGCs, OCSAs, and RBOHs), found across all plant lineages (Main figure 3a). Although the exact function of these proteins in algal species remains unclear, their immune-related orthologs are present in green algae (Main figure 3; Supplementary figure 2-26). This suggests that they underwent specialisation within the immune activation pathway prior to the emergence of land plants (Extended figure 4). The EP proteins (EDS1, PAD4, and SAG101) and RPW8-NLRs (NRG1 and ADR1) that are essential for both TIR-NLR and LRR-RLP mediated-immunity[26,27], are only present in gymnosperms and angiosperms (seed plants)[28]. Considering the ancient nature of LRR-RLPs, it is plausible that EP-protein and helper-NLRs were integrated into the LRR-RLP-signalling pathway, forming a robust immune network in seed plants.
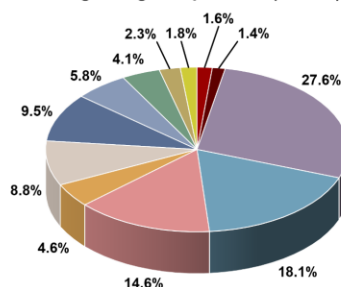
Our investigation of the expansion rate of signalling components within the plant lineage indicated an expansion of CDPKs in green algae and expansion of RLCK-VIIs in Tracheophytes (Main figure 3a; Supplementary figure 6-9). However, other families of signalling components exhibit more limited expansions, compared to PRRs. This is also consistent with the considerably larger family sizes of cell-surface receptors and NLRs in comparison to the signalling components (Main figure 3b-c). Furthermore, we examined the correlation between the percentages of signalling components and PRRs (LRR-RLK-XIIs + LRR-RLPs) across genomes. With the exception of CNGCs, EP proteins, and RPW8-NLRs (0.6 > Pearson's r > 0.3), most signalling component families do not exhibit co-expansion or co-contraction with PRRs (Main figure 3d). Thus, the expansion of PRR gene families does not necessarily result in the expansion of PTI signalling components. The RLCK-VIIs are further classified into ten subgroups which are differentially required for RLKs and RLPs to activate downstream responses[29,10,30–32] (Supplementary figure 6-7). Similarly, CDPKs fall into 4 subgroups (Supplementary figure 8-9). RLCK-VII subgroup families are differentially required by different PRRs to activate downstream immune responses[33–36]. Pathogens often target RLCK-VIIs through secreted effectors to suppress immunity[37–39]. Thus, redundancy among RLCK-VII subgroups serves as a protective mechanism for the downstream signalling pathway against effector targeting. In addition, plants have evolved RLCK-VII pseudokinases, or 'decoys', to guard functional RLCKs through NLRs[37,40–43]. Together, it has become apparent that the expansion of RLCK-VII families may have been driven by pathogenic pressure, thereby contributing to the enhanced robustness of the immune signalling network.
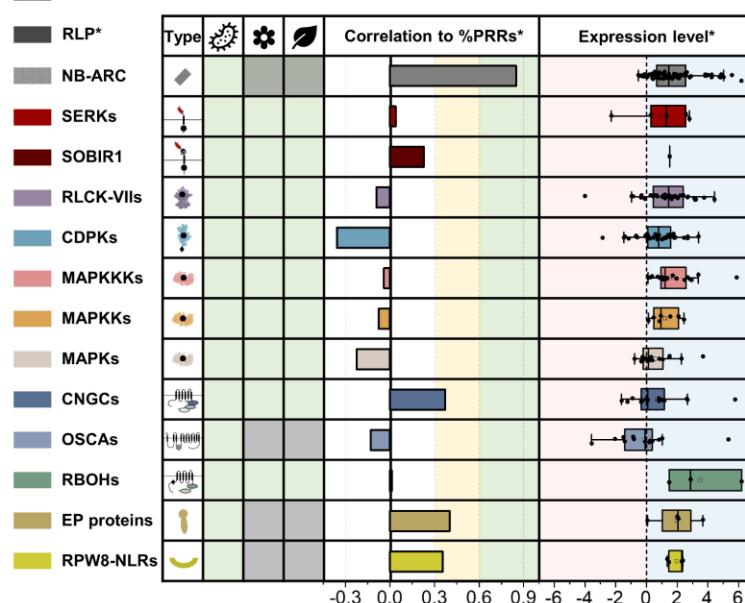
6

Main figure 3. The origin and evolution of PTI-signalling component in plants. (a) The top panel is a phylogenetic tree of multiple algal and plant lineages. Circles (○) and stars (☆) indicate the origins and expansion of receptor families, respectively. The timescale (in million year; MYA) of the phylogenetic tree was estimated by TIMETREE5[23]. The bottom panel displays the presence or absence of receptor classes in different algal and plant lineages. *M/C/K represents Mesostigmatophyceae, Chlorokybophyceae, and Klebsormidiophyceae. The number of available species from each algal and plant lineage is indicated within the respective boxes. A grey box indicates the absence, while a green box indicates the presence of a given protein family in each lineage. Dark green indicates the presence of orthologs of immunity-related (PTI) signalling components within that protein family (see also Supplementary figure 4-26). The origin of a protein family is indicated with a circle (○), followed by another circle indicating the origin of the orthologs of PTI-signalling component. The expansion rates of the protein family members are indicated by boxplots. The cyan boxplot represents the expansion rate from Glaucophyta and red algae to green algae. The yellow boxplot represents the expansion rate from green algae to Embryophytes, and the orange boxplot represents the expansion rate from early land plants to Tracheophytes. Light blue area represents expansion and light pink area represents contraction of the gene family. X-axis values represent expansion rate (×). **b)** Distribution of RLKs, RLPs, NB-ARCs, and PTI-signalling components in plants. Each fraction represents the percentage (%) of ectodomain out of all four protein classes from 350 species (337,434). *Note that RLKs and RLPs here represent RLKs and RLPs with LRR-, G-lectin-, L-lectin-, LysM-, Malectin-, WAK- and Duf26- ectodomains. **c)** Distribution of PTI-signalling components in plants. Each fraction represents the percentage (%) of each protein family out of all the families combined (78,271). The colour codes for **b)** and **c)** are indicated on the right. **d)** Characterised PTI-signalling components involved in microbial interaction (bacteria icon), reproduction (flower icon), and development (leaf icon) are indicated with green boxes. A grey box indicates that receptor class has not been reported to be involved in that particular biological process (see also Extended figure 4). The sizes of each signalling component family are given as a percent of all the proteins in the genomes of 350 plant species. Correlations between different classes of signalling components and PRR* in 300 angiosperms are indicated by the lengths of bars. *Note that PRRs here represent the combined sum of LRR-RLK-XII and LRR-RLPs, since many characterised members of these two classes are involved in pathogen recognition. A moderate positive correlation is represented in yellow (Pearson's r between 0.3-0.6). Expression level* refers to the expression of each class of PTI-signalling component during NTI in *A. thaliana*. The blue shade represents increased expression, and the red shade represents decreased expression during NTI. The X-axis represents $\log_2$ (fold change during NTI relative to untreated samples). RNA-seq data analysed here were reported previously[16], where NTI was activated by estradiol-induced expression of *AvrRps4* in *A. thaliana* for 4 hours. For the expression of each class of PTI-signalling component during PTI in *A. thaliana*, refer to supplementary figure 27. The boxplot elements in **(a)** and **(d)**: centre line, median; bounds of box, 25th and 75th percentiles; whiskers, $1.5 \times$ IQR from 25th and 75th percentiles.

The origin of immunity-related LRR-RLPs

To understand the origin of immune-specificity in cell-surface receptors, we sought to trace the evolutionary origin of LRR-RLPs in plants. Among RLPs, LRR-RLPs constitute the largest family, comprising more than twenty characterised members that perceive PAMPs or apoplastic effectors to activate immunity (Extended figure 5). The ectodomain of LRR-RLPs encompasses additional domains known as N-loopouts (NLs) and island domains (IDs) interspersed between the LRR motifs. Typically, NLs are located closer to the N-termini of the ectodomain, whereas IDs are positioned closer to C-termini. NLs are present in most immunity-related LRR-RLPs while IDs are present in all immunity-related LRR-RLPs (Extended figure 5a; Supplementary figure 28). NL positioning is relatively more flexible, occurring either before the first LRR motif or between the first few LRR motifs. Conversely, ID positioning is less flexible, located mostly before the last 4 LRR motifs within the ectodomain (Extended figure 5a; Supplementary figure 28). This observation implies a functional necessity for the specific placement of IDs in LRR-RLPs.

To investigate the functional necessity of IDs, we analysed ectodomains of LRR-RLPs and LRR-RLKs from 350 species (113,794) (Extended figure 6a-b). Employing multiple prediction programs, we identified gaps between LRR motifs ranging from 10-29 or 30-90 amino acids (AA) (Extended figure 6a). Since NLs typically span 6-30AA and IDs range from around 40-75AA, we focused on small gaps (10-29AA) corresponding to NLs, and large gaps (30-90AA) indicative of IDs (Extended figure 6a and supplementary information). Small or large gaps are relatively infrequent in LRR-RLKs (10.6% and 5.43%, respectively) (Extended figure 6c). In contrast, both small and large gaps are more prevalent in LRR-RLPs (28.3% and 61.6%, respectively) (Extended figure 6c). Furthermore, both LRR-RLKs and LRR-RLPs typically have only one gap, which can be either small or large (Extended figure 6d). Our analysis also showed that small gap positions within the ectodomains of both LRR-RLKs and LRR-RLPs are not fixed, and may be distributed randomly. Conversely, larger gaps are predominantly positioned before the last four LRR motifs in the ectodomain (51.2% for LRR-RLKs and 86.9% for LRR-RLPs) (Extended

figure 6e-f; Supplementary figure 28-32). Thus, our findings suggest functional requirement for IDs to be positioned before the last four LRRs.

An analysis of the distribution of LRR-RLK subgroups with IDs indicated that over 55% belong to the Xb subgroup (Main figure 4a). Furthermore, 94.7% of LRR-RLKs with IDs positioned before the last four LRRs (ID+4LRR) belong to the Xb subgroup (Main figure 4a), suggesting that both LRR-RLK-Xb and LRR-RLPs share the ID+4LRR motif. Among the characterised members of LRR-RLK-Xb are important components of growth and development regulation, including the BRASSINOSTEROID INSENSITIVE 1 (BRI1) family members, PHYTOSULFOKIN RECEPTOR 1 (PSKR1) and PSY1 RECEPTOR (PSY1R) family members, EXCESS MICROSPOROCYTES1 (EMS1) and NEMATODE-INDUCED LRR-RLK 1 (NILR1/GRACE)[44–48]. Interestingly, we observed that both LRR-RLPs and LRR-RLK-Xbs with ID+4LRR motifs are present in land plants (Embryophyte) but not in other lineages (Main figure 4b). Considering the similarity in structural motifs between LRR-RLPs and LRR-RLK-Xbs, it is likely that these two receptor classes share a common origin.

To test this hypothesis, we conducted phylogenetic and structural analyses on the ectodomains of LRR-RLPs and LRR-RLKs. First, we aligned the last four LRR motifs (referred to as the C3 region[49]) from LRR-RLKs and LRR-RLPs (60,240). LRR-RLPs and LRR-RLK-Xbs with an ID+4LRR cluster together in the C3 phylogenetic tree, indicating a common origin (Main figure 4c). Within this cluster (indicated as the ID+4LRR clade), we observed distinct subclades, including the BRI1/BRI1-LIKE(BRL) clade, the PSKR/PSY1R clade, and mostly LRR-RLPs (Main figure 4c). The closely similar C3 regions indicate a conserved function. BRI1 recognises brassinosteroids (BRs) and interacts with the co-receptor BAK1 to induce BR responses[50,51]. Similarly, PSKR1 perceives phytosulfokine (PSK) and interacts with the co-receptor SERK1/2. LRR-RLPs, on the other hand, perceive PAMPs or apoplastic effectors and engage both co-receptor BAK1 and SOBIR1 to initiate immune responses[52]. Structural studies have elucidated the interaction mechanism of BRI1, PSKR1, and the LRR-RLP RXEG1 with SERKs[53–56,51,57] (Extended figure 6a-d). Interestingly, the C3 regions of BRI1, PSKR1, and RXEG1 play a crucial role in SERK interactions[55,56,51,57], differentiating them from other LRR-RLKs. BRI1, PSKR1 and RXEG1 contain specific amino acid residues within the terminal two LRR motifs (within the C3 region) to interact with BAK1/SERK1 (Extended figure 6b-d), while the flagellin peptide (flg22) receptor FLS2 (an LRR-RLK-XII), relies on the 3rd-12th last LRR for BAK1 (SERK3) interactions[18] (Extended figure 6a). The N-terminal region (or N-terminal cap) of BAK1 or SERK1 is important for FLS2, BRI1, PSKR1, and RXEG1 interactions, while the 1st LRR inner surface of BAK1 or SERK1 is primarily involved in BRI1, PSKR1, and RXEG1 interactions, and the 2nd and the 4th LRR inner surface of BAK1 is involved in FLS2 interactions (Extended figure 6a-e). There are also striking similarities in interaction network maps between PSKR1-SERK1 and RXEG1-BAK1 interactions. Additionally, the C3 region of BRI1 participates in both BR binding[55,56] and co-receptor binding, whereas the C3 region of PSKR1 and RXEG1 exclusively engages in co-receptor interactions (SERK1/2 and BAK1, respectively)[51,57]. By aligning the C3 regions of various LRR-RLKs, including BRI1/BRL-orthologs, PSKR orthologs, PSY1R, and multiple LRR-RLPs, overlapping residues that are required for SERK interactions within PSKR/PSY1R, and LRR-RLP clades can be discerned (Extended figure 8a). For example, the glutamic acid (E) residue at the second position of the penultimate LRR motif is involved in both PSKR1-SERK1 and RXEG1-BAK1 interactions[57]. This E residue is highly conserved in both clades (PSKR/PSY1R clade: 69.9%, LRR-RLP: 86%; Extended figure 8b). Similarly, the phenylalanine (F) residue at the last position of the penultimate LRR motif contributes to both PSKR1-SERK1 and RXEG1-BAK1 interactions[51]. This F residue is conserved in the BRI1/BRL (62.2%), PSKR/PSY1R (99.7%), and LRR-RLP (64.1%) clades (Extended figure 8b). There is also a conserved motif crucial for SERKs interactions within the last LRR motifs of BRI/BRL, PSKR/PSY1R, and LRR-RLP clades. This motif, Glutamine-x-x-Threonine/Serine (QxxT/S) loop, is conserved in BRI1/BRL (Q:86.6%; T/S:98.9%), PSKR/PSY1R (Q:99.9%; T/S:92.5%) and LRR-RLP (Q:91.8%; T/S:88.4%) clades, but it is not conserved in other LRR-RLPs outside of the ID+4LRR clade or LRR-RLK subgroups (Main figure 4d; Extended figure 8a). Structural analysis of BRI1-BAK1, PSKR1-SERK1/2, and RXEG1-BAK1 further supports the importance of residues within and around the QxxT/S loop for SERK interactions[55,56,51,57] (Extended figure 7b-d). In conclusion, our findings indicate that the C3 regions of LRR-RLK-Xb and LRR-RLPs share a conserved function for interacting with SERKs.
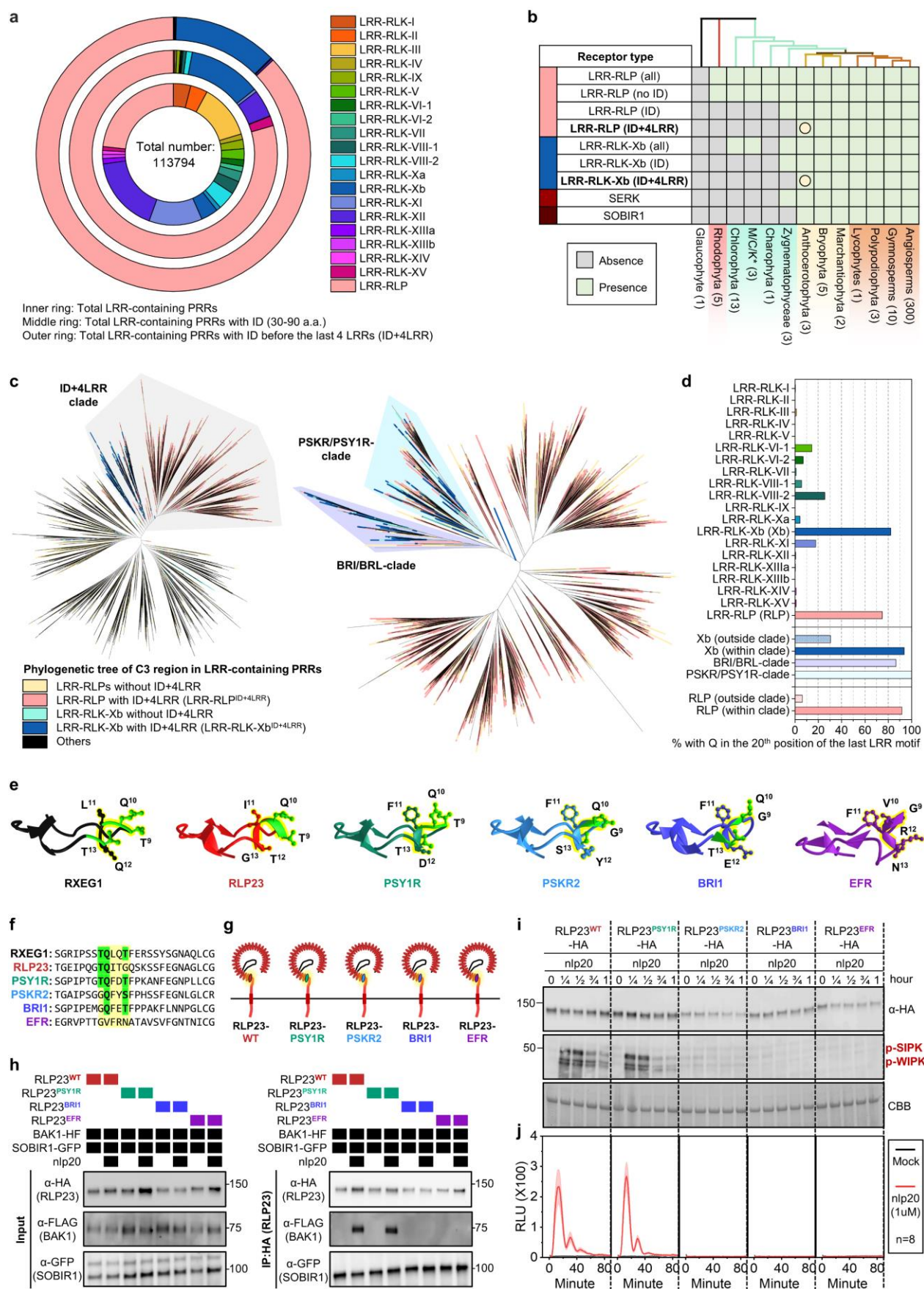
To further validate the functional conservation of C3 regions in LRR-RLK-Xb and LRR-RLPs, we performed functional analysis of the QxxT/S motif in the LRR-RLP RLP23 from *A. thaliana*. RLP23 forms heteromeric complexes with the LRR-RLK co-receptor SOBIR1, and upon the perception of the nlp20 peptide, BAK1 is recruited into the complex, leading to activation of the SOBIR1 KD to induce immunity[25,52,58]. Similar to RXEG1 (with TQLQT), RLP23 possesses a TQITG motif (TQxxx), while LRR-RLK-Xb members, such as PSY1R, PSKR2, and BRI1, feature TQFDT, GQFYS, and GQFET motifs, respectively (all QxxT/S). Notably, the LRR-RLK-XII member EFR lacks the QxxT/S motif in that position, having GVFRN instead (Main figure 4e-f). We generated chimeric constructs of RLP23 with the terminal LRR motifs swapped between

PSY1R, PSKR2, BRI1, and EFR (Main figure 4g). By immuno-precipitation assays, we tested the ability of these chimeras to interact with BAK1 upon ligand perception. Both wildtype RLP23 (RLP23[WT]) and RLP23-PSY1R (RLP23[PSY1R]) chimeras can interact with BAK1 upon nlp20 treatment, whereas RLP23-BRI1 (RLP23[BRI1]) and RLP23-EFR (RLP23[EFR]) cannot. This suggests that the terminal LRR motif of RLP23 and PSY1R bind BAK1 in a similar manner, but the terminal LRR motif of BRI1 binds to BAK1 differently. Importantly, all chimeras can interact with SOBIR1 regardless of the presence of nlp20, indicating that the last LRR motif may not be involved in SOBIR1 interactions (Main figure 4h). Furthermore, upon nlp20 treatment, both RLP23[WT] and RLP23[PSY1R] can trigger immune responses, while RLP23[PSKR2] and RLP23[BRI1] cannot (Main figure 4i). We speculate that this may be due to the absence of a specific T residue before the Qxxx motif in RLP23, which is relatively less prevalent within the LRR-RLP clade (31.1%; Extended figure 8). Multiple *A. thaliana* LRR-RLPs contain this residue, but it is not in other studied LRR-RLPs, such as the tomato Cf proteins (Extended figure 8), suggesting that this residue evolved in some species after the divergence of LRR-RLK-Xb and LRR-RLPs. Nevertheless, our results strongly support the functional conservation of C3 regions in LRR-RLK-Xbs and LRR-RLPs, specifically their ability to interact with SERKs.

The conservation of other regions within the ectodomain of LRR-RLK-Xb and LRR-RLPs was tested by aligning the IDs extracted from LRR-RLKs and LRR-RLPs (20,246). Remarkably, the ID clusters of LRR-RLPs and LRR-RLK-Xb are found in close proximity, mirroring the C3 phylogenetic tree (Extended figure 9a). Again, the BRI1/BRI1-LIKE(BRL) and PSKR/PSY1R clusters are close to the LRR-RLPs (Extended figure 9a). These results are also consistent with a previous report that PSKRs are closely related to some LRR-RLPs in Arabidopsis and rice[49]. A recent review identified two conserved lysine (K)-containing motifs, Yx8KG and Kx5Y, in the ID of LRR-RLPs[59]. The lysine residue in the Kx5Y motif from RXEG1 is required for its interaction with BAK1[57]. We extracted both Yx8KG and Kx5Y motifs in the extracted IDs (before the last 4LRR motifs) of LRR-RLKs and LRR-RLPs. More than 75% of IDs from LRR-RLPs contain at least one of these lysine motifs, while less than 5% of IDs from LRR-RLK-Xb have either motif (Extended figure 9b). This is consistent with structural data indicating that IDs from BRI1 and PSKR1 employ distinct residues for their interactions with BAK1[55,56,51]. However, IDs from LRR-RLPs that are closely related to those from LRR-RLK-Xbs retain the Kx5Y motifs (Supplementary figure 33). It is therefore possible that the common ID ancestor may have originally harboured lysine motifs that were subsequently lost from the BRI1 and PSKR/PSY1R clades.

To further explore the relatedness of the IDs between LRR-RLK-Xb group and LRR-RLPs, we formed clusters of highly similar IDs and examined their subgroup/family affiliations. Overall, we identified more than 2,822 clusters, with the majority (2,734) consisting of LRR-receptors of a single subgroup/family (Extended figure 9c). Among the 61 clusters containing LRR-receptors from two different subgroups/families, 80% (49 clusters) consisted of LRR-receptors from LRR-RLK-Xb and LRR-RLPs (Extended figure 9d). The enrichment of LRR-RLK-Xb and LRR-RLP pairings provides further evidence for the relatedness of their ectodomains. The substantial number of LRR-RLP-only ID clusters (2,383) also suggests that the IDs of LRR-RLPs have undergone extensive diversification, in the process providing a broad scope for the recognition of PAMPs and apoplastic effectors. Most of clusters contain relatively a small number of IDs, predominantly from species within the same order or family (Extended figure 9c). Consequently, tracing back the origin of IDs is challenging due to their considerable diversity. We therefore propose that the IDs of LRR-RLK-Xbs and LRR-RLPs likely originated from a common ancestor, with the IDs of LRR-RLPs expanding and diversifying after the divergence of LRR-RLK-Xb and LRR-RLPs.

Within the BRI1/BRI1-LIKE (BRL) and PSKR/PSY1R clusters, we found LRR-RLP counterparts of BRI1, PSKR, and PSY1R. To further test their relatedness, we aligned the ectodomains from LRR-RLKs and LRR-RLPs and extracted the branches containing BRI1, PSKR, and PSY1R from the phylogenetic tree (Extended figure 9d-f; Supplementary figure 34). Within the BRI1/BRL-ectodomain clade, we identified 63 LRR-RLPs and 1,188 LRR-RLK-Xbs. The IDs of these LRR-RLPs are highly similar to BRL1/BRL3 with residues crucial for BL binding[53,54] (Extended figure 9e). Within the PSKR-ectodomain subclade, we found 343 LRR-RLPs alongside 791 LRR-RLK-Xbs. Similarly, the IDs of the LRR-RLPs are remarkably similar to PSKR2, with residues essential for PSK binding and SERK1/2 interactions[51] (Extended figure 9f). Finally, within the PSY1R-ectodomain subclade, we found 454 LRR-RLPs accompanied by 227 LRR-RLK-Xbs. The IDs of these LRR-RLPs share remarkable similarity with AtPSY1R (Extended figure 9g). Moreover, AtRLP3 and AtPSY1R have over 70% sequence identity and 85% similarity in their ectodomains. Although PSY1R is not the receptor for PSY peptide[60], it is possible that RLP3, and PSY1R recognise similar or identical ligands[61]. Currently, the functions of these BRL-, PSKR- and PSY1R-like LRR-RLPs remain unclear. RLP3 confers resistance against the vascular wilt fungus *Fusarium oxysporum* f. sp. *matthioli*[61], while PSY1R is involved in growth and development[46]. We propose that these RLPs may either recognise endogenous molecules to activate growth and development, or participate in the recognition of pathogen-mimicking molecules to trigger immune signalling.
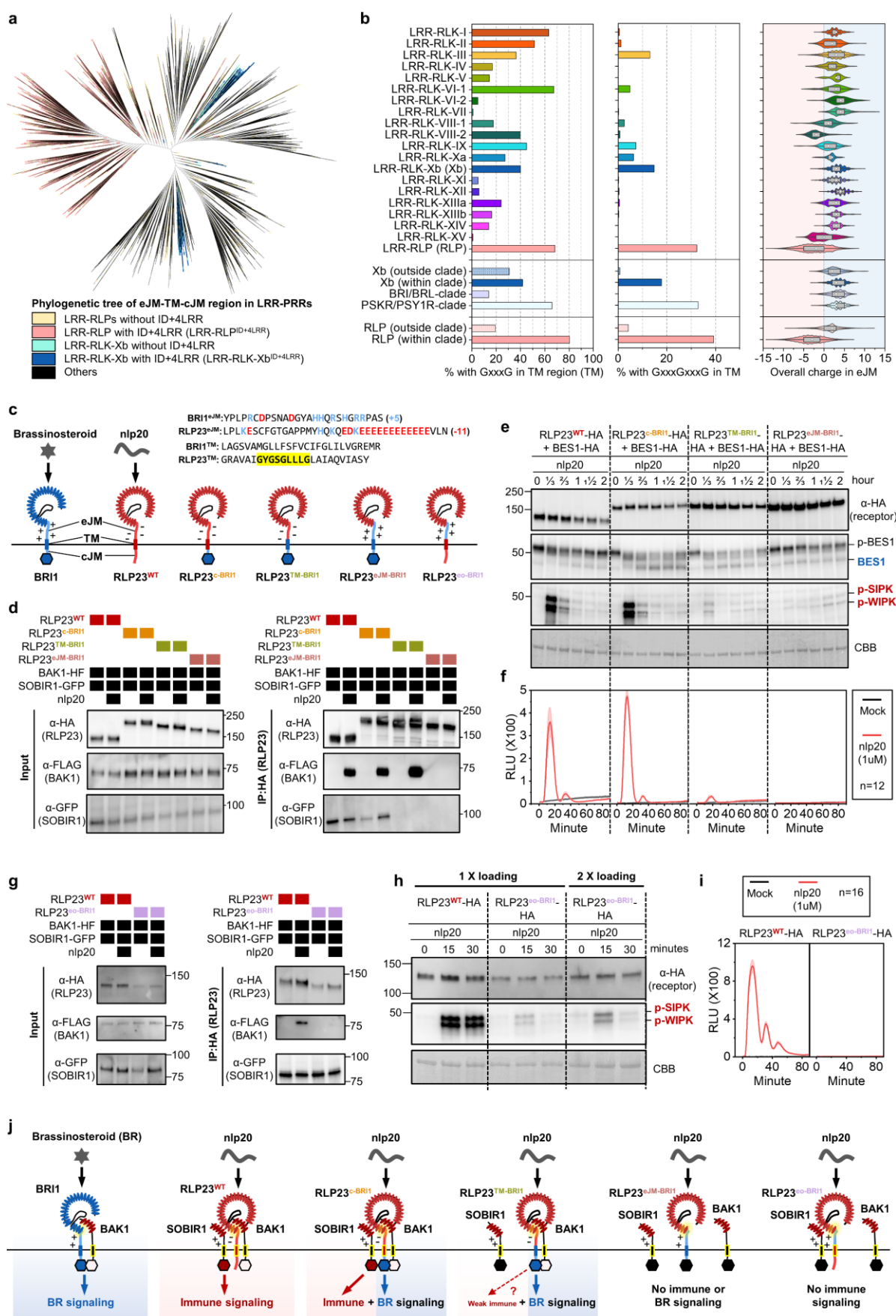
Main figure 4. The origin and evolution of LRR^{ID+4LRR} in plants. (a) The concentric ring pie chart presents the percentage of LRR-containing cell-surface receptors (PRRs) from 350 species. The inner ring represents all LRR-containing cell-surface receptors (113,794); the middle ring represents LRR-containing PRRs with ID (20,556); the outer ring represents LRR-containing PRRs with an ID preceding the last 4 LRR (ID+4LRR) at the C terminus (16,885). **(b)** The presence or absence of receptor classes in various algal and plant lineages. *M/C/K represents Mesostigmatophyceae, Chlorokybophyceae, and Klebsormidiophyceae. A grey box indicates the absence, and a green box indicates the presence of a given receptor class in each lineage. The origin of LRR-RLP and LRR-RLK-Xb with ID+4LRR is indicated with a circle (○). **(c)** Phylogenetic tree of the C3 region (last 4 LRRs) from all LRR-containing PRRs of 350 species. Branches are colour-labelled as indicated. The light grey area indicates clustering of LRR-RLK-Xb and LRR-RLP with ID+4LRR. The pruned phylogenetic tree on the right corresponds to the light grey area in the left tree, with clades labelled in dark grey areas accordingly. **(d)** The percentage of receptors with glutamine (Q) in the 9th position of the terminal LRR. Percentages (%) are calculated as (the number of LRR-RLKs or LRR-RLPs in the subgroup with Q) / (the number of LRR-RLKs or LRR-RLPs in the subgroup without Q) ×100. RLP/RLK-Xb (outside clade) refers to receptors outside the light grey clade in **(c)**. RLP/RLK-Xb (within clade) refers to receptors inside the light grey clade in **(c)**. The number of cell-surface receptors (n) in each LRR-RLK subgroup: I, n = 752; II, n = 682; III, n = 6572; IV, n = 1033; V,n = 8; VI-1, n = 84; V1-2, n = 146; VII, n = 1720; VIII-1, n = 195; VIII-2, n = 411; IX, n = 70; Xa, n = 96; Xb, n = 3182; XI, n = 8807; XII, n = 12863; XIIIa, n = 739; XIIIb, n = 465; XIV, n = 241; XV, n = 548; Xb (outside clade), n = 580; Xb (within clade), n = 2527; BRI/BRL clade, n = 1170; PSKR/PSY1R clade, n = 1347. The number of cell-surface receptors (n) in each LRR-RLP subgroup: LRR-RLP (RLP), n = 24860; RLP (outside clade), n = 5000; RLP (within clade), n = 19970. **(e)** Structure of the terminal LRR motif of *Nicotiana benthamiana* (Nb)RXEG1, *A. thaliana* (At)RLP23, AtPSY1R, AtPSKR2, AtBRI1 and AtEFR. Structures of NbRXEG1 and AtBRI1 were published[53–56,51,57]. Structures of AtRLP23, AtPSY1R, AtPSKR2 and AtEFR were predicted by Alphafold2*,[20]. Ectodomains are visualised in iCn3D[21]. **(f)** Alignment of amino acids in the last LRR motifs from NbRXEG1, AtRLP23, AtPSY1R, AtPSKR2, AtBRI1, and AtEFR. Amino acid residues involved in the interaction between NbRXEG1 and BAK1 are highlighted in green. The QxxT motif positions are highlighted in yellow. **(g)** Design of AtRLP23 chimeras. The last LRR motif of AtRLP23 is exchanged with either AtPSY1R, AtPSKR2, AtBRI1, or AtEFR. **(h)** Immuno-precipitation to test interactions between AtRLP23 chimeras, AtBAK1 and AtSOBIR1. Nb leaves expressing the indicated constructs were treated with either mock or 1μM nlp20 for 5 minutes. **(i)** Functionality testing of AtRLP23 chimeras. Nb leaves expressing the indicated constructs were treated with 1μM nlp20 and samples were collected at indicated time points. Phosphorylation of NbSIPK and NbWIPK was detected with p-P42/44 antibody. **(j)** Nb leaf discs expressing the indicated constructs were collected and treated with either mock or 1μM nlp20, and ROS production was measured for 90 min. For details of experiential design in **(h)**, **(i)**, and **(j)**, refer to the methods section.

Specialisation of cell-surface receptors in different biological processes

Given the common origin of the ectodomains of LRR-RLPs and LRR-RLK-Xbs, it is likely that these receptors have undergone specialisation in immune- and developmental processes following their divergence. While LRR-RLK-Xbs activate downstream responses through the Xb kinase domain, LRR-RLPs recruit SOBIR1 to trigger immunity[52,58]. Because LRR-RLPs lack a kinase domain, the juxtamembrane (JM) and TM regions may activate immune responses. We therefore aligned the eJM (JM region before TM), TM, and cJM (JM region after TM, but with the absence of kinase domain from RLKs) from LRR-RLKs and LRR-RLPs (62,896). Interestingly, the eJM-TM-cJM region of LRR-RLPs is not closely related to LRR-RLK-Xbs (Main figure 5a). Previous studies have reported the requirement of Glycine-X-X-X-Glycine (GxxxG) motifs in SOBIR1 for its association with LRR-RLPs, and the potential contribution of negatively charged amino acids in the eJM region of LRR-RLPs to their interaction with SOBIR1[62]. We aligned the eJM-TM region of LRR-RLKs, BRI1/BRL-orthologs, PSKR orthologs, PSY1R, and multiple LRR-RLPs, and analysed them for the presence of positively/negatively charged amino acids in eJM and GxxxG motifs in TMs (Extended figure 10a). Consistent with previous reports[62], LRR-RLPs are strongly negatively charged at the end of the eJM region, whereas LRR-RLKs, including SOBIR1, are positively charged at the end of the eJM region (Extended figure 10a). Most LRR-RLPs have a single GxxxG motif, with some having two (GxxxGxxxG) or three (GxxxGxxxGxxxG) consecutive motifs. Conversely, GxxxG motifs are relatively less common in LRR-RLKs, but can be found in SOBIR1, BAK1, and the PSKR/PSY1R clade (Extended figure 10a). We further examined overall charges in the eJM region of the LRR-RLP and LRR-RLK subgroups. All LRR-RLK subgroups, except for LRR-RLK-VIII-2, feature positively charged eJM regions, whereas LRR-RLPs have negatively charged eJM regions. Importantly, negatively charged eJM regions are only present in LRR-RLPs within the ID+4LRR clade (Main figure 5b). The percentage of TMs with GxxxG or GxxxGxxxG motifs in the LRR-RLP and LRR-RLK subgroups varies from about 60% of subgroup members from LRR-RLK-I, LRR-RLK-VI-1, LRR-RLPs possess GxxxG motifs in TM, but more than 80% of LRR-RLPs within the ID+4LRR clade contain these motifs (Main

figure 5b). Moreover, GxxxGxxxG motifs are primarily found in LRR-RLPs and are relatively rare in LRR-RLKs. Interestingly, both GxxxG and GxxxGxxxG motifs are enriched in the PSKR/PSY1R clade, indicating that the TM regions of LRR-RLPs and PSYR/PSY1R clade might share a common origin. (Main figure 5b; Supplementary figure 35-37).

To assess the functionality of eJM-TM-cytosolic regions in LRR-RLK-Xbs and LRR-RLPs, we generated multiple chimeras of eJM, TM, and cytosolic regions of BRI1 and RLP23 (Main figure 5c). In RLP23^{c-BRI1}, the cytosolic region (following TM) of BRI1 was swapped into RLP23. In RLP23$^{TM-BRI1}$, the TM+ cytosolic region of BRI1 was swapped into RLP23. In RLP23$^{eJM-BRI1}$, the eJM + TM + cytosolic region of BRI1 was swapped into RLP23 (Main figure 5c). Immuno-precipitation assays showed that both RLP23^{WT} and RLP23^{c-BRI1} exhibit constitutive interactions with SOBIR1, and specifically interact with BAK1 upon nlp20 treatment. RLP23$^{TM-BRI1}$ does not interact with SOBIR1 but still interacts with BAK1 upon nlp20 treatment, suggesting that the RLP23 TM region with GxxxG motifs is necessary for SOBIR1-, but not BAK1-interactions (Main figure 5d). Consistent with these immunoprecipitation assays, RLP23^{WT} and RLP23^{c-BRI1} can activate immune responses, while RLP23^{c-BRI1} and RLP23$^{TM-BRI1}$ can activate developmental responses as indicated by the dephosphorylation of BRI1-EMS-SUPPRESSOR 1 (BES1)[63] (Main figure 5e-f). This confirms that the BRI1 kinase domain is specifically required for the activation of BR responses. Interestingly, RLP23$^{TM-BRI1}$ can also activate weak immune responses, likely independent of SOBIR1 (Main figure 5e-f). RLP23$^{eJM-BRI1}$ does not interact with SOBIR1 or BAK1, and is unable to activate either immune or developmental responses. These data suggest that the eJM region of RLP23 is necessary for RLP23-BAK1 interactions (Main figure 5d). This conclusion was reinforced by generating RLP23$^{eo-BRI1}$ chimera, in which the eJM region of RLP23 is replaced with the BRI1 eJM (Main figure 5c). RLP23$^{eo-BRI1}$ consistently accumulates less protein than RLP23^{WT}. Moreover, RLP23$^{eo-BRI1}$ constitutively interacts with SOBIR1, but fails to interact with BAK1 with nlp20 treatment (Main figure 5g). RLP23$^{eo-BRI1}$ only weakly activates MAPKs, and fails to trigger ROS production with nlp20 treatment (Main figure 5h-i). Thus, we conclude that the eJM regions of LRR-RLPs are important for protein accumulation and interaction with BAK1 upon ligand perception.

Main figure 5. Adaptation of LRR$^{ID+4LRR}$ to differential downstream signalling pathways. (a) Phylogenetic tree of eJM-TM-cJM regions from all LRR-containing PRRs of 350 plant species. Branches are labelled in colours as indicated. **(b)** Percentage of receptors with GxxxG in the TM region (left), GxGxG in the TM region (middle) and overall charge distribution in eJM (right). LRR-RLP/RLK-Xb (outside clade) refers to receptors outside the light grey clade in Figure 4c. Percentages (%) are calculated by the number of LRR-RLKs or LRR-RLPs in the subgroup with the stated residue divided by the number of LRR-RLKs or LRR-RLPs in the subgroup without the stated residue × 100. RLP/RLK-Xb (outside clade) refers to receptors outside the light grey clade in Figure 4c. RLP/RLK-Xb (within clade) refers to receptors inside the light grey clade in Figure 4c. Number of cell-surface receptors (n) in each LRR-RLK subgroup: I, n = 752; II, n = 682; III, n = 6572; IV, n = 1033; V, n = 8; VI-1, n = 84; V1-2, n = 146; VII, n = 1720; VIII-1, n = 195; VIII-2, n = 411; IX, n = 70; Xa, n = 96; Xb, n = 3182; XI, n = 8807; XII, n = 12863; XIIIa, n = 739; XIIIb, n = 465; XIV, n = 241; XV, n = 548; Xb (outside clade) , n = 580; Xb (within clade), n = 2527; BRI/BRL clade, n = 1170; PSKR/PSY1R clade, n = 1347. Number of cell-surface receptors (n) in each LRR-RLP subgroup: LRR-RLP (RLP), n = 24860; RLP (outside clade), n = 5000; RLP (within clade), n = 19970. **(c)** Design of AtRLP23-BRI1 chimeras. Different regions (cytosolic, TM+cytosolic, and eJM+TM+cytosolic) of BRI1 were swapped into AtRLP23 as indicated. The alignment of amino acids in the eJM and TM regions from AtRLP23 and AtBRI1 is shown. Amino acid residues with negative charges are in red and amino acids with positive charges are in blue. The GxxxG motif in the TM region is highlighted in yellow. **(d, g)** Immuno-precipitation to test interactions between AtRLP23 chimeras, AtBAK1 and AtSOBIR1. Nb leaves expressing the indicated constructs were treated with either mock or 1μM nlp20 for 5 minutes. **(e, h)** Functionality testing of AtRLP23 chimeras. Nb leaves expressing the indicated constructs were treated with 1μM nlp20 and samples were collected at indicated time points. Dephosphorylation of BESI1-HA was detected with HA antibody. Phosphorylation of NbSIPK and NbWIPK was detected with p-P42/44 antibody. For **(h)**, twice the sample of RLP23$^{oe-BRI1}$ was loaded as a reference, because RLP23$^{oe-BRI1}$ protein accumulation is weaker than that of RLP23^{WT}. **(f, i)** Nb leaf discs expressing the indicated constructs (without BES1-HA) were collected and treated with either mock or 1μM nlp20 and ROS production was measured for 90 minutes. For details of experiential design, refer to the methods section. **(g)** Schematic model of the interaction between LRR-RLK-Xb$^{ID+4LRR}$ and LRR-RLP$^{ID+4LRR}$ with co-receptors to induce differential downstream signalling. Both receptor classes utilise the last 4 LRRs (highlighted in yellow) to interact with SERKs (BAK1). LRR-RLP evolved to interact with SOBIR1 with the GxxxG motifs in TM (highlighted in yellow outline). Coloured hexagons on RLKs indicate activated kinases and black hexagon indicates an inactivated kinase.
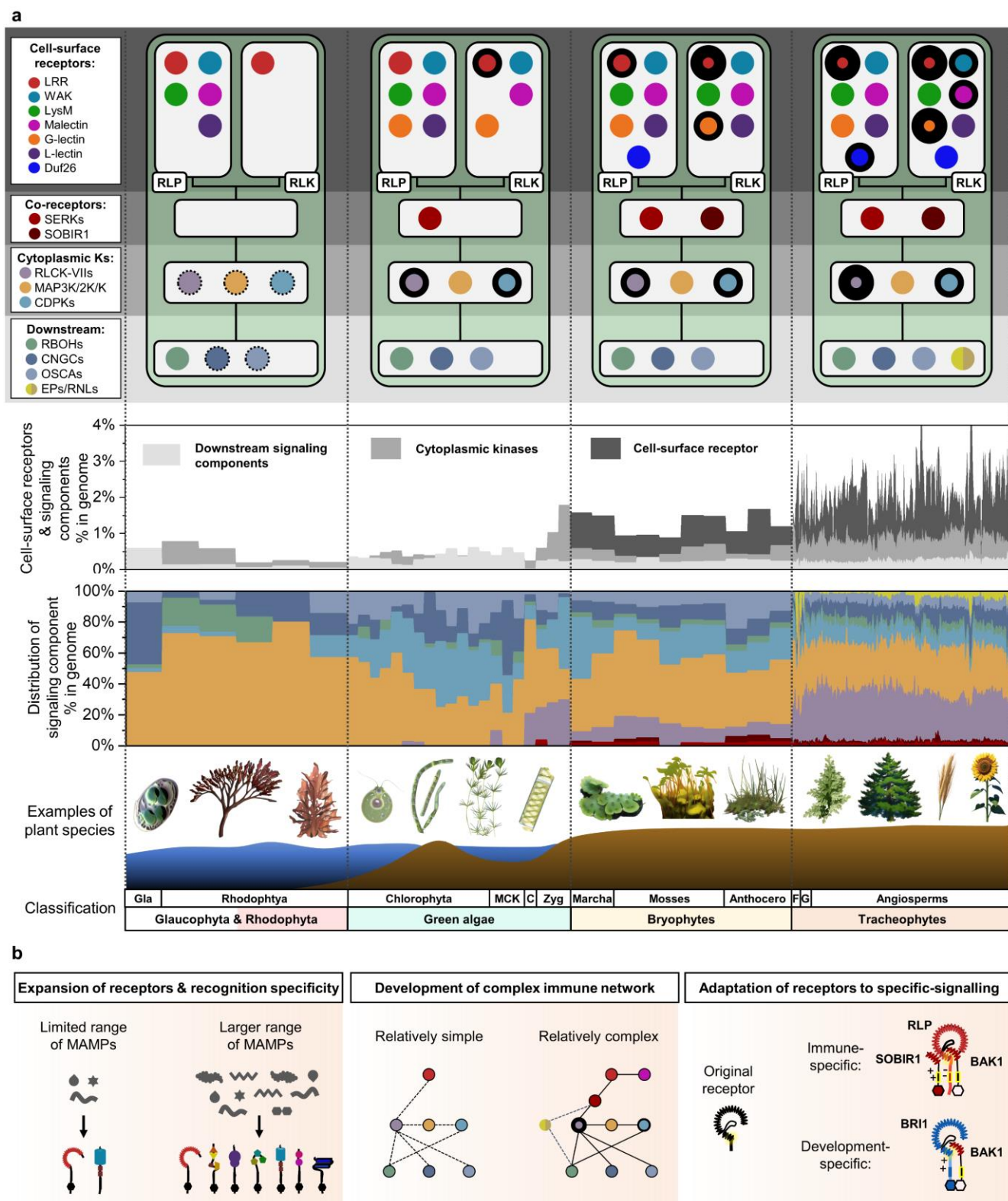

Discussion

The ectodomains of LRR-RLPs and LRR-RLK-Xbs appear to share a common evolutionary origin, suggesting an ancestral adoption of the ID+4LRR architecture for ligand recognition and interactions with co-receptors such as the SERKs. Since both LRR-RLPs and LRR-RLK-Xbs rely on the terminal 4LRR/C3 region for SERK interaction[55,56,51,57], it is plausible that the C3 region has undergone stabilizing evolution to preserve this functionality (Extended figure 10b). Consequently, a portion of the C3 region in LRR-RLK-Xb and LRR-RLP can be interchanged without loss of function. However, the remaining ectodomain, including the LRRs preceding the ID and the ID itself, has undergone diversification and adaptation to recognise distinct ligands. Notably, certain LRR-RLPs exhibit significant sequence similarities to LRR-RLK-Xbs (BRI1/BRL, PSKR, and PSY1R), though their specific functions are unexplored. Following diversification, the eJM, TM, and cytosolic regions of LRR-RLK-Xb and LRR-RLP acquired distinct roles, primarily in development and immunity, respectively (Main figure 5j; Extended figure 10b). The eJM and TM regions of LRR-RLP specifically facilitate constitutive interactions with SOBIR1 and ligand-dependent interactions with BAK1, whereas the LRR-RLK-Xb group lacks such specialisation. Since the recruitment of SOBIR1 to BRI1 would lead to immune activation, there should be negative selection for negatively charged eJMs and GxxxGs in the TM regions to prevent immune activation by LRR-RLK-Xbs. Given the different roles of various domains and regions of LRR-RLKs and LRR-RLPs, we propose a model in which different domains or regions of cell-surface receptors undergo modular evolution to either diversify or to maintain their original functions. Modular evolution allows the specialisation of cell-surface receptors to recognize different ligands and to activate distinct downstream signal responses or maintain interactions with co-receptors (Extended figure 10b).

The study of PRRs and the core PTI-signalling pathway has primarily focused on model plants, such as Arabidopsis and rice. Recent investigations into chitin perception by LysM-RLKs in the Bryophyte *Marchantia polymorpha* have revealed a high degree of conservation in the PTI-signalling pathway across land plants[64,65]. In fact, most cell-surface receptors and PTI signalling components in the Tracheophytes are conserved in Bryophytes, with the exception of Duf26-RLKs, EP proteins, and RPW8-NLRs (Main figure 6a). Therefore, the most recent common ancestor of land plants is likely to possess a considerable number of cell-surface receptors and basic components of immune signalling network, facilitating adaptation to terrestrial

environments (Main figure 6a). The presence of diverse cell-surface receptor classes and multiple PTI signalling components in algal species suggests that algae may also have PTI system (Main figure 6a), as indicated by the MAMPs triggering of defense responses in some algal species. For instance, treatment with lipopolysaccharides (LPS) and oligosaccharides in multiple red algae species (Rhodophyta) stimulated ROS production, activation of nitric oxide (NO) signalling, changes in protein expression and hypersensitive responses (HR)[66]. Land plants perceive LPS and oligosaccharides through G-lectin- and WAK-/LysM- receptors [67,68]. In Rhodophyta, we identified the presence of LysM-RLPs, but not WAK- or G-lectin-containing cell-surface receptors, suggesting that other PRRs perceive LPS in algae. Nevertheless, PTI is likely to be present in multiple algal lineages. Further investigation into PRRs and signalling network in the Glaucophyta, Rhodophyta, and green algal species holds promise for shedding light on the evolution PTI in Viridiplantae.

Our work has uncovered multiple evolutionary mechanisms underlying PTI: i) expansion of the number of receptors and their recognition specificity; ii) development of an increasingly complex immune network; iii) adaptation of existing receptors for specific signalling (Main figure 6b). i) Many PRRs are lineage-specific, with the characterised LRR-RLK-XII and LRR-RLPs found only in specific lineages within angiosperms, such as the Brassicaceae (EFR) and Solanaceae (Cf2/4/5/9). This expansion of PRR subgroups, including LRR-RLKs, LRR-RLPs, G-lectin-RLKs, and WAK-RLKs, has enabled plants from different lineages to recognise a broader range of molecules specific to certain pathogens, exemplified by Xa21-Raxx21 and Cf2/4/5/9-Avr2/4/5/9 interactions[69–73]. ii) Cell-surface co-receptors, such as SERKs and SOBIR1 emerged during or around the time of land plant evolution. Moreover, several cell-surface receptors involved in signalling regulation, such as malectin-RLKs (FERONIA)[74] and LRR-RLK-Xa (BIR1, BIR2)[75–78] are found exclusively in Embryophytes. PRRs differentially activate downstream signalling components, including RLCK-VIIs, CDPKs and EP proteins, which fine-tune the magnitude of immune responses[10,79]. RLCK-VIIs are expanded in the Embryophytes, while EP proteins and RPW8-NLRs are specific to Tracheophytes. Collectively it seems that increasingly intricate and specialised PTI-signalling networks enhance the flexibility and regulation of immune responses to keep up with the changes taking place in rapidly evolving pathogens[80]. iii) The structural similarities between LRR-RLPs and LRR-RLK-Xbs implies a common origin between immune-specific cell-surface receptors and development-specific cell-surface receptors. The exact nature of the common ancestral form of these receptors, whether an RLK or an RLP, remains, and perhaps will always remain uncertain. Both LRR-RLK-Xbs and LRR-RLPs with ID+4LRR can be found in land plants, so it is conceivable that LRR-RLK-Xb, with its kinase domain predating land plant emergence, evolved from an integration of an LRR-RLP containing an ID+LRR into an Xb kinase domain[81]. In this scenario, the common ancestral receptors could have recognised a PAMP, with the peptide sequence of this PAMP possibly converted to serve as a phytocytokine to regulate plant developmental processes. Multiple phytocytokines, such as PSK, PSY, SCOOPs and CLE peptides, are present in plant pathogens and pests[82–86]. Whether the perception of phytocytokines evolved from the perception of PAMPs, or pathogens developed phytocytokine-mimics to repress immune responses remains an open question.

Main figure 6. The evolutionary trajectory of PTI in plants. (a) The top panel depicts the presence of cell-surface receptors, co-receptors, cytoplasmic kinases (cytoplasmic Ks) and downstream signalling components (downstream) in Glaucophyta and Rhodophyta, green algae, Bryophytes and Tracheophytes. Nodes are labelled in colours as indicated on the left. Absence of a node indicates the absence of a gene family from the lineage. Nodes with dotted outlines indicate the presence of a gene family, but the absence of immunity-related orthologs. Nodes with thick outlines indicate the expansion of gene families. Repeated

expansion is indicated with thicker outlines. The middle panel (top) displays the percentages (%) of cell-surface receptors and signalling components in the genome of each species within the lineage. Bars are labelled in colours as indicated. Middle panel (bottom) represents the distribution of signalling components, including co-receptors, cytoplasmic kinases, and downstream signalling components, in the genome of each species within a lineage. Bars are labelled in colours as indicated on top left. The bottom panel shows examples of plant species and the classification of the plant lineages. Abbreviations: Gla: Glaucophyta; MCK: represents Mesostigmatophyceae, Chlorokybophyceae, and Klebsormidiophyceae; C: Charophyta; Zyg: Zygnematophyceae; Marcha: Marchantiophyta; Mosses: Bryophyte; Anthocero: Anthocerotophyta; F: Lycophytes and Polypodiophyta; G: Gymnosperms. **(b)** Evolution of PTI in plants. Left panel: expansion of PRR family gene repertoires throughout the plant lineage, which leads to recognition of a larger range of MAMPs. Middle panel: PRR co-receptors, EP proteins and helper NLRs are absent from many algal species. A more complex immune network involving these signalling components apparently developed in vascular plants. Right panel: An ancient PRR with LRR$^{ID+4LRR}$ with unknown function evolved into LRR-RLK-Xbs and LRR-RLPs, which are involved in development- and immune-signalling, respectively. eJM-TM-cJM region of LRR-RLPs evolved to allow interactions with SOBIR1 to induce immunity (negatively charged eJM and GxxxG). LRR-RLK-Xbs utilize Xb kinase domains to induce distinct downstream responses.

Methods

RLK, RLP, and ectodomain identification

LRR-RLK, LRR-RLP, LysM-RLKs and Nb-ARC proteins were identified as described previously[8]. For the other RLK and RLPs described in this study, only primary gene model proteins longer than 150 AA were used[8]. Based on the presence of a kinase domain (KD) and/or a trans-membrane domain (TM), the proteins fell into three major groups: 1) RLKs with KD and TM, 2) RLPs without KD but with TM, and 3) ectodomains without KD or TM.

To identify RLK candidates, we first searched for the presence of a protein kinase domain (PFAM: PF00069.26) with hmmer (version 3.1b2, option -E 1e-10[87]). If multiple hits were found, only the best match was kept. Potential signalling peptides were removed with SignalP (version 5.0b[88]) to avoid identifying and including signal peptides as TMs. TMs were searched with tmhmm (version 2.0[89]). The location of the KD and the TM were used to split the protein sequence into the endodomain (with KD) and ectodomain (without KD). Given that LRR-RLKs were previously identified [8], we searched and removed proteins with LRR repeats in their ectodomains with predict-phytolrr[90] (obtained in December 2021).

To identify RLPs and candidate ectodomains (without KD or TM) , we first removed all proteins with a kinase domain match (hmmer with the option -E1e1000) and removed potential signalling peptides to avoid their inclusion with TMs. We then searched for TMs with tmhmm and kept proteins with no TM (ectodomains) or 1-2 TM (RLPs). Proteins containing LRR repeats were identified with predict-phytolrr and removed from the RLP/ectodomain set.

Finally, we searched candidates from all three major groups for the presence of Duf26 (PFAM: PF01657.18), malectin or malectin-like (PFAM: PF11721.9 and PF12819.8), G-type lectin (PFAM: :PF01453.25), L-type lectin (PFAM: PF00139.20), and WAK domains: EGF-like (PFAM: PF00008.28), Calcium-binding EGF (PFAM: PF07645.16), and Wall-associated receptor kinase galacturonan-binding (PFAM: PF13947.7) with hmmer (option -E 10). Hits from all searches were combined for each group, and proteins were assigned to the hit with the highest score.

Identification of signalling components

Signalling components were identified using different approaches, but always using primary gene model proteins longer than 150 AA[8]. 150AA cut-off was used to eliminate truncated proteins.

CNGCs, OSCAs, and RBOHs were identified by hmmer searches (option -E 10) for specific domains. Ion transport protein domains (PFAM: PF00520.32) were used for CNGCs, PHM7_cyt (PFAM: PF14703.7) and RSN1_TM (PFAM: PF13967.7) were used for OSCAs, and FAD-binding domains (PFAM: PF08022.13), ferric reductase like transmembrane components (PFAM: PF01794.20), and ferric reductase NAD binding domains (PFAM: PF08030.13) were used for RBOHs.

EP proteins (EDS1, PAD4, and SAG101) were identified by hmmer searches (option -E 10) for the Lipase (class 3) domain (PFAM: PF01764.26) and the Enhanced disease susceptibility 1 protein EP domain (PFAM: PF18117.2). To further classify the candidates among the known EDS1, PAD4, and SAG101 candidates, we clustered all candidates with MMSeq2 (Release 14-7e284, options --min-seq-id 0.3 -c 0.3[91]). All known EDS1 (*AT3G48090*, *Niben101Scf06720g01024.1*, and

Solyc06g071280), PAD4 (*AT3G52430*, *Niben101Scf02544g01012.1*, and *Solyc02g032850*), and SAG101 (*AT5G14930*, *Niben101Scf00271g02011.1*, *Niben101Scf01300g01009.1*, *Solyc02g069400*, and *Solyc02g067660*) proteins were found in exactly one cluster each. Hence, we used the three matching clusters as EDS1, PAD4, and SAG101 proteins.

RPW8-NLRs (NRG1 and ADR1) were identified similarly using the NB-ARC (PFAM: PF00931.23) and RPW8 (PFAM: PF05659.12) domains. After clustering as described above, we found all known NRG1 and ADR1 sequences in one single cluster. This allowed us to extract and re-cluster these sequences with more stringent parameters (options --min-seq-id 0.3 -c 0.75). After that, we found all known NRG1 (*AT5G66900*, *AT5G66910*, and *Niben101Scf02118g00018.1*) and ADR1 (*AT1G33560*, *AT4G33300*, *AT5G04720*, and *Niben101Scf02422g02015*) proteins in exactly one cluster each, indicating that the two matching clusters could be considered as NRG1 and ADR1 proteins.

The remaining signalling component candidates (SOBIR1, RLCK-VII, CDPK, MAPK, MAPKK, and MAPKKK) were identified using previously published HMM profiles[92] using hmmer (option -E 10). The following patterns were used for the families of interest: SOBIR1: RLK-Pelle_LRR-XI-2, RLCK-VII: RLK-Pelle_RLCK-VIIa-1, RLK-Pelle_RLCK-VIIa-2, RLK-Pelle_RLCK-VIIb, CDPK: CAMK_CDPK, MAPK: CMGC_MAPK, MAPKK: STE_STE7, and MAPKKK: STE_STE11.

Expansion rate of cell-surface receptors and signalling components

The percentages (%) of cell-surface receptors and signalling components from each genome was calculated as (number of identified genes/number of searched genes × 100). Next, the percentages from each species within a lineage (e.g, Rhodophtya or green algae) were grouped and the median percentage was calculated. Median value was used instead of mean to avoid outliers within the lineages. The expansion rate within a species is calculated by ((% cell surface receptors or signalling components in that species)-(median))/(median). For example, the expansion rate of LRR-RLP family in *Marchantia polymorpha* from green algae is calculated by ((%LRR-RLP in *Marchantia polymorpha*)-(median %LRR-RLP in green algae)/(median %LRR-RLP in green algae). Values larger than 0 indicate expansion; values equal to 0 indicate no expansion, and values below 0 indicate contraction. Note that the reliability of the expansion rate is dependent on the number of species used to calculate the median, which is also dependent on the available genomes in Glaucophyta, red algae (Rhodophyta), green algae, and Bryophytes.

Identification of N-loopouts (NLs) island domains (IDs)

To identify N-loopouts (NLs) or island domains (IDs) in LRR-RLK and LRR-RLP proteins, we used a dataset of previously described LRR-RLKs and LRR-RLPs[8]. We searched the LRR-RLKs again for kinase domains (PFAM: PF00069.26) with hmmer (option -E 1e-10), and kept only the best match for each protein. LRR-motifs and transmembrane domains were searched in both groups with predict-phytolrr[90] and tmhmm[89], respectively. LRR-RLKs were filtered for the presence of internal KD motifs, one or two TM, and at least two external LRR repeats ('internal' was defined as the side with the kinase domain). LRR-RLPs were filtered for the absence of a KD and the presence of one or two TM and at least two external LRR-motifs as defined by the site with more LRR repeats). The outer LRR-motifs were then used to identify NLs and IDs: Individual repeats were grouped into LRR-regions if they were less than 13 AA apart from each other. Gaps between LRR-regions or LRR-motifs that were 15-29 AA or 30-90 AA long were extracted as NL and ID candidates, respectively. After extracting gap sequences, all sequences were again checked for the presence of LRR-motifs using predict-phytolrr and hmmer using all LRR patterns as described previously[8], and LRRsearch[93]. Only NLs and IDs without any LRR match were included in the final dataset. Locations of the NLs and IDs in relation to LRR-motifs and LRR-regions were determined with a custom R-script. ID sequences were aligned to each other with FAMSA[94] without trimming[95] and phylogenetic trees were inferred with FastTree[96] (version 2.1.11 SSE3, option -wag). Trees were rooted with gotree[97] (v0.4.2) using one sequence belonging to the most basal species as outgroup, according to the taxonomic tree.

The phylogenetic tree of the IDs was used to cluster the proteins: the tree was converted into a distance matrix using the function cophenetic.phylo() from the R-package 'ape'[98] (version 5.6-2). Distances smaller than 0.2 (*i.e.* less than 0.2 substitutions per site on average) were extracted, converted to similarities, and used as edges in a network. Communities within this similarity network were identified with the function cluster_louvain[99] implemented in the R-package 'igraph'[100] (version 1.2.6).

In-depth phylogeny of the ectodomains from LRR-RLPs and LRR-RLKs

In-depth analysis of the ectodomains from LRR-RLPs and LRR-RLKs was done using the LRR-RLKs and LRR-RLPs from the NLs and IDs search (see above). We first searched for the C3 domain in each sequence[49,8] with hmmer and selected the best

hit. We then pruned the sequences to include everything from the C3 domain to the C terminus. For the LRR-RLKs, we further searched sequences for kinase domains and removed sequences upstream of the start of the kinase domain. That is, for all LRR-RLKs and LRR-RLPs, we extracted regions with C3, eJM, TM, or eJM domains. These sequences were aligned with FAMSA. Specific domains (*e.g*. C3 or eJM) were subsequently extracted from this alignment. After extraction, the phylogenetic trees of specific domains were constructed as described above (FAMSA, FastTree, gotree).

Taxonomic trees

Taxonomic trees used in this study were identical to the ones described previously[8]: The taxonomic tree for visualizing the entire data set and selecting outgroups was obtained from NCBI (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi). The tree used for testing the relationship between the fraction of candidates found and phylogenetic distances, was obtained from a previous report[101]. The latter contained 238 out of the 351 genomes analysed. Phylogenetic trees were visualised and pruned, and figures were generated with iTOL[102].

Test for similarities in fraction of proteins and phylogenetic relationships

Tests for similarities in fraction of proteins and phylogenetic relationship were done as described previously[8]: To test whether the fraction of certain proteins found per species correlated with predicted phylogenetic relationships, we converted the fractions and the phylogenetic tree to distance matrices and tested for correlation with mantel tests (R-package vegan, version 2.5-7 with 10'000 permutations). Analogously, we also tested for correlation between distance matrices obtained for two different sets of proteins. P-values were corrected for multiple testing to reflect false discovery rates (FDRs[103]).

Vector construction

The CDS regions of AtRLP23, AtBRI1, AtPSKR2, AtPSY1R, AtEFR, and AtBES1 were amplified by PCR with KoD one (Toyobo, Japan), and the PCR products were cloned into the epiGreenB5 (3×HA) vector between the *Cla*I and *Bam*HI restriction sites with In-Fusion HD Cloning Kit (Clontech, USA) to generate p35S::BES1-HA or p35S::cell-surface receptor-HA (epiGreenB5-Cauliflower mosaic virus (CaMV) p35S:gene of interest-3×HA). The constructs were then transformed into *Agrobacterium tumefaciens* strain AGL1 for transient expression in *Nicotiana benthamiana*. All chimeric cell-surface receptors generated in this study contain the EFR signal peptide to ensure consistency between the constructs and expression levels.

Transient expression in *Nicotiana benthamiana*

Agrobacterium tumefaciens strain AGL1 carrying the binary expression vectors described above were grown on LB agar plates amended with selection antibiotics. Cultures were pelleted, centrifugated, and then resuspended in infiltration buffer (10mM $MgCl_2$, 10mM MES at pH5.6, and 100μM acetosyringone). The concentration of AGL1 was then adjusted to $OD_{600} = 0.5$ and syringe-infiltrated into *Nicotiana benthamiana* leaves.

Protein extraction and immunoprecipitation

Protein extraction for immunoprecipitation was performed as previously described[104]. Three days after transient expression, three to four grams of *N. benthamiana* leaves were treated with elicitors and snap-frozen. The tissues were then ground in liquid nitrogen and extracted in extraction buffer (50mM Tris-HCl at pH 7.5, 150mM NaCl, 10% glycerol, 5mM DTT, 2.5mM NaF, 1mM $Na_2MoO_4 \cdot 2H2O$, 0.5% polyvinylpyrrolidone (w/v), 1% Protease Inhibitor Cocktail (P9599; Sigma-Aldrich), 100μM phenylmethylsulphonyl fluoride and 2% IGEPAL CA-630 (v/v; Sigma-Aldrich), and 2mM EDTA) at a concentration of 3mL/g tissue powder. Samples were then incubated at 4 °C for an hour and debris was removed by centrifugation at 13,000rpm for 10min at 4 °C. Supernatants were collected, protein concentrations were adjusted to 5mg/mL, then incubated with rotation for an hour at 4 °C with 50μL anti-HA magnetic beads (Miltenyi Biotec) for immunoprecipitation. Magnetic beads were then washed twice with extraction buffer and the HA-tagged protein was eluted with sodium dodecyl sulphate (SDS) sample buffer at 95 °C.

Immunoblotting

Protein extractions were performed as previously described[16]. *N. benthamiana* leaves were infiltrated with elicitors and snap-frozen at indicated time points. The tissues were then lysed in liquid nitrogen and extracted in 1×NuPAGE™ LDS Sample Buffer (Invitrogen™) with 10mM DTT at 70 °C for 10 minutes. Total proteins were then separated by SDS-PAGE and blotted onto a nitrocellulose membrane (Trans-Blot Turbo Transfer System, Bio-Rad). The membrane was then blocked in a solution

of either 5% skimmed milk (for BES1 and cell-surface receptor detection) or 5% bovine serum albumin (BSA; for MAPK detection) in Tris-buffered saline, 0.1% Tween 20 detergent (TBST) for an hour. Phosphorylated MAPKs were detected using α-phospho-p44/42 MAPK rabbit monoclonal antibody (D13.14.4E, in 1:2000, Cell Signalling Technology, USA) in a solution of 5% BSA in TBST overnight at 4 °C. HA-tagged BES1 or cell surface receptors were detected using Anti-HA-Peroxidase, High Affinity, rat IgG$_1$ antibody (Roche) in a solution of 5% skimmed milk in TBST overnight at 4 °C. For detection of MAPKs, this was followed by incubation with α-rabbit IgG-HRP-conjugated secondary antibodies (1:10000, Roche, USA) in a solution of 5% BSA in TBST for an hour at room temperature. HRP signal was then detected by Clarity Western ECL Substrate (Bio-Rad) with a LAS 4000 system (GE Healthcare, USA). Nitrocellulose membranes were stained with Coomassie Brilliant Blue (CBB) to ensure equal loading.

ROS assays

ROS burst assays were performed as described previously[104]. *N. benthamiana* leaf discs were collected with a 4-mm-diameter cork borer and placed in 96-well plates with 120µl deionised water overnight in the dark (abaxial surface of the leaves facing down). *N. benthamiana* leaf discs were then treated with either mock (water) or 1 µM nlp20 in 20 mM luminol (Wako, Japan) and 0.02 mg ml^{-1} horseradish peroxidase (Sigma-Aldrich). Luminescence was then measured over indicated periods of time with a Tristar2 multimode reader (Berthold Technologies, Germany).

Contributions

B.P.M.N., Y.K. and K.S. conceived and conceptualised the study; B.P.M.N., Y.K., M.W.S. and M.W. designed the bioinformatic analyses; M.W.S. and M.W. performed the bioinformatic analyses; B.P.M.N. performed the vector construction with assistance from Y.K.. B.P.M.N. performed protein extractions, immunoprecipitation, immunoblotting and ROS assays; B.P.M.N. wrote the original draft; and B.P.M.N., M.W.S., M.W., Y.K., an K.S. reviewed and edited the manuscript.

Acknowledgments

References

1.   Zipfel, C. Plant pattern-recognition receptors. *Trends Immunol.* **35,** 345–351 (2014).

2.   Boutrot, F. & Zipfel, C. Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annu. Rev. Phytopathol.* **55,** 257–286 (2017).

3.   Nürnberger, T. & Brunner, F. Innate immunity in plants and animals: emerging parallels between the recognition of general elicitors and pathogen-associated molecular patterns. *Curr. Opin. Plant Biol.* **5,** 318–324 (2002).

4.   Heller, J., Clavé, C., Gladieux, P., Saupe, S. J. & Glass, N. L. NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. *Proc. Natl. Acad. Sci. USA* **115,** E2292–E2301 (2018).

5.   Gao, L. A. *et al.* Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science* **377,** eabm4096 (2022).

6.   Ngou, B. P. M., Ding, P. & Jones, J. D. G. Thirty years of resistance: Zig-zag through the plant immune system. *Plant Cell* **34,** 1447–1478 (2022).

7.   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).

8.   Ngou, B. P. M., Heal, R., Wyler, M., Schmid, M. W. & Jones, J. D. G. Concerted expansion and contraction of immune receptor gene repertoires in plant genomes. *Nat. Plants* **8,** 1146–1152 (2022).

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

10. Rao, S. *et al.* Roles of Receptor-Like Cytoplasmic Kinase VII Members in Pattern-Triggered Immune Signaling. *Plant Physiol.* **177,** 1679–1690 (2018).

11. Bigeard, J., Colcombet, J. & Hirt, H. Signaling mechanisms in pattern-triggered immunity (PTI). *Mol. Plant* **8,** 521–539 (2015).

12. He, Y., Zhou, J., Shan, L. & Meng, X. Plant cell surface receptor-mediated signaling - a common theme amid diversity. *J. Cell Sci.* **131,** (2018).

13. Hou, S., Liu, D. & He, P. Phytocytokines function as immunological modulators of plant immunity. *Stress Biol.* **1,** 8 (2021).

14. 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).

15. Bjornson, M., Pimprikar, P., Nürnberger, T. & Zipfel, C. The transcriptional landscape of Arabidopsis thaliana pattern-triggered immunity. *Nat. Plants* **7,** 579–586 (2021).

16. Ngou, B. P. M., Ahn, H.-K., Ding, P. & Jones, J. D. G. Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature* **592,** 110–115 (2021).

17. Liu, T. *et al.* Chitin-induced dimerization activates a plant immune receptor. *Science* **336,** 1160–1164 (2012).

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

19. Xiao, Y. *et al.* Mechanisms of RALF peptide perception by a heterotypic receptor complex. *Nature* **572,** 270–274 (2019).

20. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).
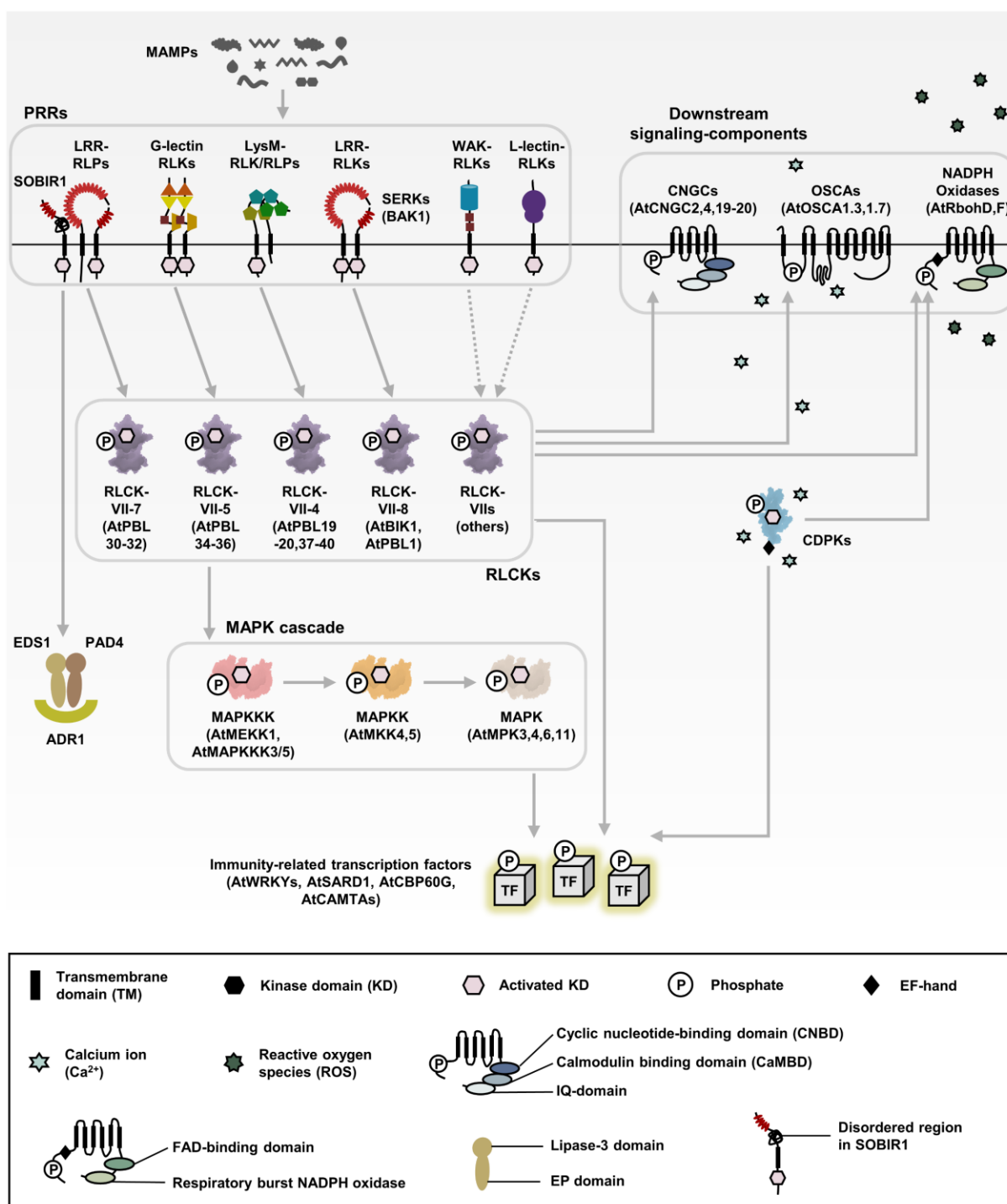
21.   Wang, J. *et al.* iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics* **36,** 131–135 (2020).

22.   Man, J., Gallagher, J. P. & Bartlett, M. Structural evolution drives diversification of the large LRR-RLK gene family. *New Phytol.* **226,** 1492–1505 (2020).

23.   Kumar, S. *et al.* Timetree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* **39,** (2022).

24.   Hohmann, U., Lau, K. & Hothorn, M. The structural basis of ligand perception and signal activation by receptor kinases. *Annu. Rev. Plant Biol.* **68,** 109–137 (2017).

25.   Liebrand, T. W. H. *et al.* Receptor-like kinase SOBIR1/EVR interacts with receptor-like proteins in plant immunity against fungal infection. *Proc. Natl. Acad. Sci. USA* **110,** 10010–10015 (2013).

26.   Feehan, J. M., Castel, B., Bentham, A. R. & Jones, J. D. Plant NLRs get by with a little help from their friends. *Curr. Opin. Plant Biol.* **56,** 99–108 (2020).

27.   Dongus, J. A. & Parker, J. E. EDS1 signalling: At the nexus of intracellular and surface receptor immunity. *Curr. Opin. Plant Biol.* **62,** 102039 (2021).

28.   Lapin, D. *et al.* A Coevolved EDS1-SAG101-NRG1 Module Mediates Cell Death Signaling by TIR-Domain Immune Receptors. *Plant Cell* **31,** 2430–2455 (2019).

29.   Bi, G. *et al.* Receptor-Like Cytoplasmic Kinases Directly Link Diverse Pattern Recognition Receptors to the Activation of Mitogen-Activated Protein Kinase Cascades in Arabidopsis. *Plant Cell* **30,** 1543–1561 (2018).

30.   Luo, X. *et al.* Tyrosine phosphorylation of the lectin receptor-like kinase LORE regulates plant immunity. *EMBO J.* **39,** e102856 (2020).

31.   Pruitt, R. N. *et al.* The EDS1-PAD4-ADR1 node mediates Arabidopsis pattern-triggered immunity. *Nature* **598,** 495–499 (2021).

32.   Huang, W. R. H. *et al.* Receptor-like cytoplasmic kinases belonging to different subfamilies mediate immune responses downstream of the Cf-4 resistance protein in Nicotiana benthamiana. *BioRxiv* (2023). doi:10.1101/2023.04.25.538242

33.   Boudsocq, M. *et al.* Differential innate immune signalling via Ca(2+) sensor protein kinases. *Nature* **464,** 418–422 (2010).

34.   Dubiella, U. *et al.* Calcium-dependent protein kinase/NADPH oxidase activation circuit is required for rapid defense signal propagation. *Proc. Natl. Acad. Sci. USA* **110,** 8744–8749 (2013).

35.   Gao, X. *et al.* Bifurcation of Arabidopsis NLR immune signaling via Ca^{2+}-dependent protein kinases. *PLoS Pathog.* **9,** e1003127 (2013).

36.   Monaghan, J. *et al.* The calcium-dependent protein kinase CPK28 buffers plant immunity and regulates BIK1 turnover. *Cell Host Microbe* **16,** 605–615 (2014).

37.   Zhang, J. *et al.* Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a Pseudomonas syringae effector. *Cell Host Microbe* **7,** 290–301 (2010).

38.   Liang, X. *et al.* A Phytophthora capsici RXLR effector targets and inhibits the central immune kinases to suppress plant immunity. *New Phytol.* **232,** 264–278 (2021).

39.   Yu, G. *et al.* The Arabidopsis E3 ubiquitin ligase PUB4 regulates BIK1 and is targeted by a bacterial type-III effector. *EMBO J.* **41,** e107257 (2022).

40.   DeYoung, B. J., Qi, D., Kim, S.-H., Burke, T. P. & Innes, R. W. Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. *Cell Microbiol.* **14,** 1071–1084 (2012).

41. Lewis, J. D. *et al.* The Arabidopsis ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the Pseudomonas syringae type III effector HopZ1a. *Proc. Natl. Acad. Sci. USA* **110,** 18722–18727 (2013).

42. Wang, G. *et al.* The Decoy Substrate of a Pathogen Effector and a Pseudokinase Specify Pathogen-Induced Modified-Self Recognition and Immunity in Plants. *Cell Host Microbe* **18,** 285–295 (2015).

43. Bastedo, D. P. *et al.* Perturbations of the ZED1 pseudokinase activate plant immunity. *PLoS Pathog.* **15,** e1007900 (2019).

44. Li, J. & Chory, J. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **90,** 929–938 (1997).

45. Matsubayashi, Y., Ogawa, M., Morita, A. & Sakagami, Y. An LRR receptor kinase involved in perception of a peptide plant hormone, phytosulfokine. *Science* **296,** 1470–1472 (2002).

46. Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M. & Matsubayashi, Y. Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **104,** 18333–18338 (2007).

47. Jia, G., Liu, X., Owen, H. A. & Zhao, D. Signaling of cell fate determination by the TPD1 small protein and EMS1 receptor kinase. *Proc. Natl. Acad. Sci. USA* **105,** 2220–2225 (2008).

48. Mendy, B. *et al.* Arabidopsis leucine-rich repeat receptor-like kinase NILR1 is required for induction of innate immunity to parasitic nematodes. *PLoS Pathog.* **13,** e1006284 (2017).

49. 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).

50. Nam, K. H. & Li, J. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell* **110,** 203–212 (2002).

51. Wang, J. *et al.* Allosteric receptor activation by the plant peptide hormone phytosulfokine. *Nature* **525,** 265–268 (2015).

52. Albert, I. *et al.* An RLP23-SOBIR1-BAK1 complex mediates NLP-triggered immunity. *Nat. Plants* **1,** 15140 (2015).

53. Hothorn, M. *et al.* Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* **474,** 467–471 (2011).

54. She, J. *et al.* Structural insight into brassinosteroid perception by BRI1. *Nature* **474,** 472–476 (2011).

55. Santiago, J., Henzler, C. & Hothorn, M. Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. *Science* **341,** 889–892 (2013).

56. Sun, Y. *et al.* Structure reveals that BAK1 as a co-receptor recognizes the BRI1-bound brassinolide. *Cell Res.* **23,** 1326–1329 (2013).

57. Sun, Y. *et al.* Plant receptor-like protein activation by a microbial glycoside hydrolase. *Nature* **610,** 335–342 (2022).

58. van der Burgh, A. M., Postma, J., Robatzek, S. & Joosten, M. H. A. J. Kinase activity of SOBIR1 and BAK1 is required for immune signalling. *Mol. Plant Pathol.* **20,** 410–422 (2019).

59. Snoeck, S., Garcia, A. G. & Steinbrenner, A. D. Plant Receptor-like proteins (RLPs): Structural features enabling versatile immune recognition. *Physiol. Mol. Plant Pathol.* **125,** 102004 (2023).

60. Ogawa-Ohnishi, M. *et al.* Peptide ligand-mediated trade-off between plant growth and stress response. *Science* **378,** 175–180 (2022).

61. Shen, Y. & Diener, A. C. Arabidopsis thaliana resistance to fusarium oxysporum 2 implicates tyrosine-sulfated peptide signaling in susceptibility and resistance to root infection. *PLoS Genet.* **9,** e1003525 (2013).
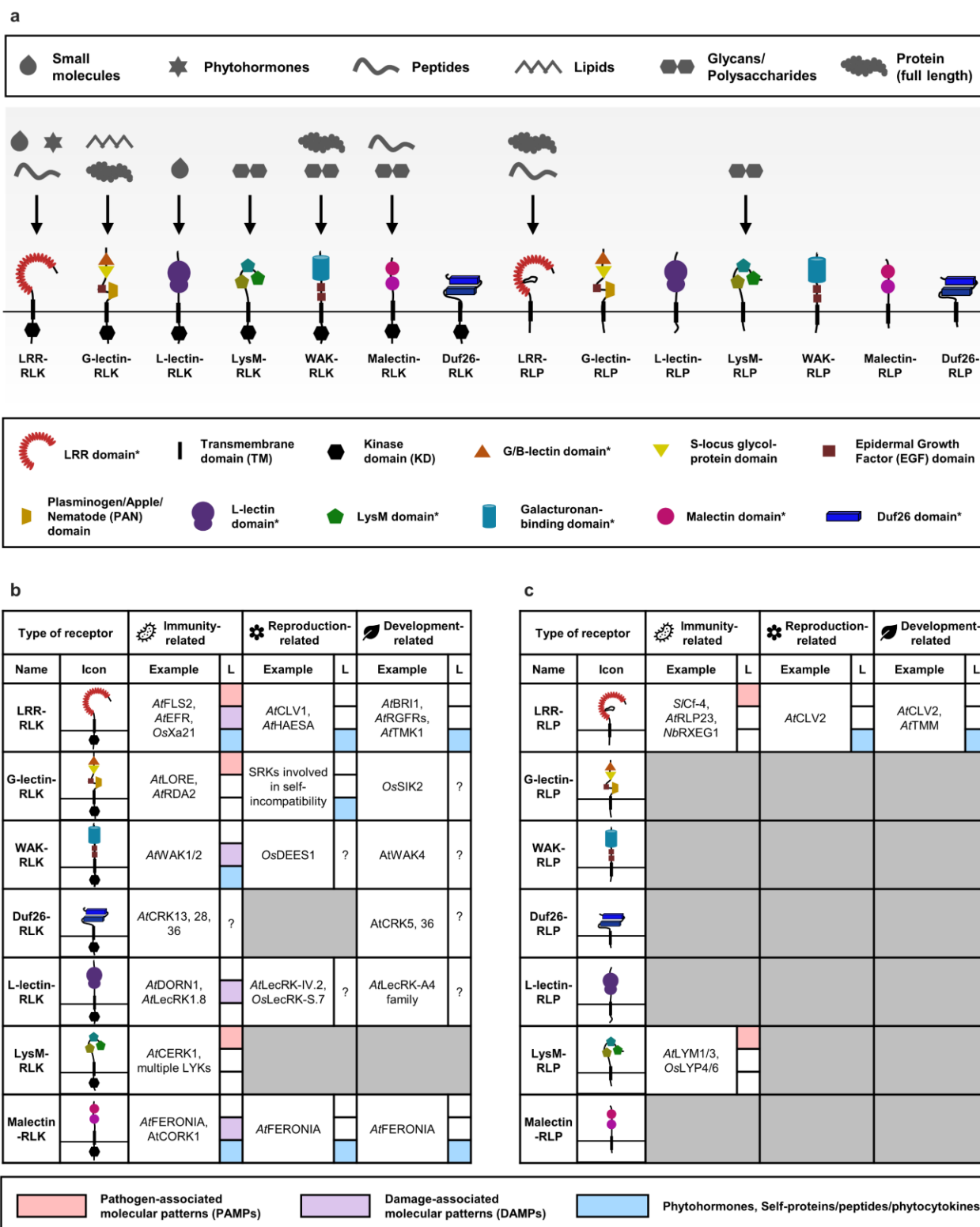
62.   Bi, G. *et al.* SOBIR1 requires the GxxxG dimerization motif in its transmembrane domain to form constitutive complexes with receptor-like proteins. *Mol. Plant Pathol.* **17,** 96–107 (2016).

63.   Yin, Y. *et al.* BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell* **109,** 181–191 (2002).

64.   Yotsui, I. *et al.* Genetic and phosphoproteomic basis of LysM-mediated immune signaling in *Marchantia polymorpha* highlights conserved elements and new aspect of pattern-triggered immunity in land plants. *BioRxiv* (2022). doi:10.1101/2022.12.28.521631

65.   Chu, J. *et al.* Conservation of the PBL-RBOH immune module in land plants. *Curr. Biol.* **33,** 1130–1137.e5 (2023).

66.   Weinberger, F. Pathogen-induced defense and innate immunity in macroalgae. *Biol Bull* **213,** 290–302 (2007).

67.   Brutus, A., Sicilia, F., Macone, A., Cervone, F. & De Lorenzo, G. A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proc. Natl. Acad. Sci. USA* **107,** 9452–9457 (2010).

68.   Ranf, S. *et al.* A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in *Arabidopsis thaliana. Nat. Immunol.* **16,** 426–433 (2015).

69.   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* **266,** 789–793 (1994).

70.   Song, W. Y. *et al.* A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* **270,** 1804–1806 (1995).

71.   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).

72.   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).

73.   Dixon, M. S., Hatzixanthis, K., Jones, D. A., Harrison, K. & Jones, J. D. The tomato Cf-5 disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell* **10,** 1915–1925 (1998).

74.   Stegmann, M. *et al.* The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science* **355,** 287–289 (2017).

75.   Gao, M. *et al.* Regulation of cell death and innate immunity by two receptor-like kinases in *Arabidopsis. Cell Host Microbe* **6,** 34–44 (2009).

76.   Halter, T. *et al.* The leucine-rich repeat receptor kinase BIR2 is a negative regulator of BAK1 in plant immunity. *Curr. Biol.* **24,** 134–143 (2014).

77.   Imkampe, J. *et al.* The Arabidopsis leucine-rich repeat receptor kinase BIR3 negatively regulates BAK1 receptor complex formation and stabilizes BAK1. *Plant Cell* **29,** 2285–2303 (2017).

78.   Ma, C. *et al.* Structural basis for BIR1-mediated negative regulation of plant immunity. *Cell Res.* **27,** 1521–1524 (2017).

79.   Bredow, M. & Monaghan, J. Regulation of Plant Immune Signaling by Calcium-Dependent Protein Kinases. *Mol. Plant Microbe Interact.* **32,** 6–19 (2019).

80.   Ngou, B. P. M., Jones, J. D. G. & Ding, P. Plant immune networks. *Trends Plant Sci.* **27,** 255–273 (2022).

81.   He, Z. *et al.* Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* **288,** 2360–2363 (2000).

82. Pruitt, R. N. *et al.* A microbially derived tyrosine-sulfated peptide mimics a plant peptide hormone. *New Phytol.* **215,** 725–736 (2017).

83. Frei Dit Frey, N. & Favery, B. Plant-parasitic nematode secreted peptides hijack a plant secretory pathway. *New Phytol.* **229,** 11–13 (2021).

84. Hou, S. *et al.* The Arabidopsis MIK2 receptor elicits immunity by sensing a conserved signature from phytocytokines and microbes. *Nat. Commun.* **12,** 5494 (2021).

85. Rhodes, J. *et al.* Perception of a divergent family of phytocytokines by the Arabidopsis receptor kinase MIK2. *Nat. Commun.* **12,** 705 (2021).

86. Yuan, N., Furumizu, C., Zhang, B. & Sawa, S. Database mining of plant peptide homologues. *Plant Biotechnol (Tsukuba)* **38,** 137–143 (2021).

87. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7,** e1002195 (2011).

88. Armenteros, J. J. A. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37,** 420–423 (2019).

89. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).

90. Chen, T. Identification and characterization of the LRR repeats in plant LRR-RLKs. *BMC Mol and Cell Biol* **22,** 9 (2021).

91. Steinegger, M. & Söding, J. Clustering huge protein sequence sets in linear time. *Nat. Commun.* **9,** 2542 (2018).

92. Lehti-Shiu, M. D. & Shiu, S.-H. Diversity, classification and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **367,** 2619–2639 (2012).

93. Bej, A. *et al.* LRRsearch: An asynchronous server-based application for the prediction of leucine-rich repeat motifs and an integrative database of NOD-like receptors. *Comput Biol Med* **53,** 164–170 (2014).

94. Deorowicz, S., Debudaj-Grabysz, A. & Gudyś, A. FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Sci. Rep.* **6,** 33964 (2016).

95. Tan, G. *et al.* Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference. *Syst. Biol.* **64,** 778–791 (2015).

96. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 — approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** e9490 (2010).

97. Lemoine, F. & Gascuel, O. Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genom. Bioinform.* **3,** lqab075 (2021).

98. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35,** 526–528 (2019).

99. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008,** P10008 (2008).

100. Csardi, G. The igraph software package for complex network research | CiNii Research. *InterJournal Complex Systems* (2006).

101. Qian, H. & Jin, Y. An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *Journal of Plant Ecology* **9,** 233–239 (2016).

102. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49,** W293–W296 (2021).

103. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57,** 289–300 (1995).

104. Kadota, Y. *et al*. Direct regulation of the NADPH oxidase RBOHD by the PRR-associated kinase BIK1 during plant immunity. *Mol. Cel*l **54**, 43–55 (2014).

105. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).

Extended figure 1. PTI signalling pathway in plants. Schematic showing the PTI signalling pathway in plants. The architecture of different signalling components is defined in the lower box. Grey arrow represents the activation of downstream signalling components by upstream signalling components following the perception of PAMPs/MAMPs by PRRs. Multiple signalling components are activated through phosphorylation by an activated kinase domain (KD). Downstream signalling components are activated to produce ROS and cytosolic calcium influx. Multiple immune-related transcription factors (TF) are also activated by cytoplasmic protein kinases, which then lead to defence-related gene expression.

Extended figure 2. Domain architecture and roles of the major plant cell-surface receptor classes. (a) Schematic displays the domain architecture of different classes of receptor-like kinase (RLKs) and receptor-like proteins (RLP) in plants. Arrows represent the ligands which these receptor classes have been reported to perceive or recognize. The upper box defines the ligands recognized by different receptors. The lower box defines the domains within the receptor classes. '*' represents the

definitive domains in these receptors. Note that these receptors may be able to recognise other unidentified ligands. The number of definitive domains and the domain architecture in each receptor are variable. **(b-c)** Characterised (b) RLKs and (c) RLPs in plants; the biological processes in which the characterised members are involved, and type of ligand (L) they perceive. A grey box indicates that no receptor class has been reported to be involved in that biological process. For ligands, a red box indicates PAMPs from pathogens, a purple box indicates DAMPs released from damage, and a blue box indicates phytohormones, self-protein/peptide or phytocytokines. A question mark (?) indicates an unidentified ligand. Abbreviations for plant species: *A. thaliana*, *At*; *S. lycopersicum*, *Sl*; *O. sativa*, *Os*; *N. benthamiana*, *Nb.* References to the genes are included in the supplementary information.

Extended figure 3. The origin and expansion of LRR-containing cell-surface receptors in plants. Top panel represents the phylogenetic tree of multiple algal and plant lineages. The timescale (in million years; MYA) of the phylogenetic tree was estimated by TIMETREE5[23]. Bottom panel represents the presence or absence of different receptor classes in algal and plant lineages. RLP represents LRR-RLPs. 'Out clade' refers to LRR-RLPs outside the ID+4LRR clade in figure 4c and 'in clade' refers to LRR-RLPs inside the ID+4LRR clade in figure 4c. *M/C/K represents Mesostigmatophyceae, Chlorokybophyceae, and Klebsormidiophyceae. The number of available species from each algal and plant lineage is indicated by numbers in the boxes. A grey box indicates the absence of receptor, and a green box indicates the presence of receptors in each lineage. Receptor origin is indicated with a circle (○). Expansion rates of receptor classes are indicated by boxplots. The cyan boxplot represents the expansion rate from Glaucophyta and Rhodophyta to green algae. The yellow boxplot represents the expansion rate from green algae to embryophytes and the orange boxplot represents the differences between Bryophytes and Tracheophytes. Light blue area represents expansion and light pink area represents contraction of the gene family. X-axis values represent expansion rate (×). Boxplot elements: center line, median; bounds of box, 25th and 75th percentiles; whiskers, 1.5 × IQR from 25th and 75th percentiles.

31

PTI Signaling components			Immunity-related	Reproduction-related	Development-related	Method of identification from genome
Type	Protein	Icon	Example	Example	Example	
PRR co-receptor	SERKs		AtBAK1 (AtSERK3), OsBAK1 (OsSERK2)	SERKs	SERKs	Identification of SERK clade from previously published LRR-RLK-II phylogenetic tree[79].
	SOBIR1		AtSOBIR1, NbSOBIR1, SlSOBIR1	AtSOBIR1	AtSOBIR1	Identification of SOBIR1 clade from LRR-RLK-XI-2 members from HMM search[79]. Confirm through checking SOBIR1 orthologs.
RLCKs	RLCK-VII		AtBIK1, AtPBL1, AtPBL19-20, 30-32, 34-40	*Other RLCK subgroups are involved	*Other RLCK subgroups are involved	Identification of RLCK-VII clade from RLCK members from HMM search using published HMM profiles[79]. Confirm through checking RLCK-VII orthologs.
CDPKs (CPKs)	-		AtCPK1, AtCPK2, AtCPK5, AtCPK6, AtCPK11, AtCPK28	AtCPK2, AtCPK11, AtCPK20, AtCPK24, AtCPK33	AtCPK11, AtCPK12, AtCPK28, AtCPK30, AtCPK32	Identification of CDPK clade from CAMK_CDPK members from HMM search[79]. Confirm through checking CDPK orthologs.
MAPKs	MAPKKK (MEKK)		AtMEKK1, AtMAPKKK3/5	MAPK signaling cascade is also involved in floral organ development, inflorescence architecture and embryogenesis.	MAPK signaling cascade is also involved in phytohormone synthesis and signalling, cell division, cell differentiation and other development processes.	Identification of MAPKKK clade from STE11 members from HMM search[79]. Confirm through checking MAPKKK orthologs.
	MAPKK (MKK)		AtMKK4, AtMKK5			Identification of MAPKK clade from STE7 members from HMM search using published HMM profiles[79]. Confirm through checking MAPKK orthologs.
	MAPK		AtMPK3, AtMPK4, AtMPK6			Identification of MAPK clade from CMGC_MAPK members from HMM search[79]. Confirm through checking MAPK orthologs.
Calcium channels	CNGC		AtCNGC2,4,19,20	AtCNGC8, AtCNGC16, AtCNGC18	AtCNGC5, AtCNGC6, AtCNGC9	Search for proteins with PFAM profiles PF00520 and PF00027. Confirm through checking CNGC orthologs.
	OSCA		AtOSCA1.3, AtOSCA1.7			Search for proteins with PFAM profiles PF14703, PF02714 and PF13967. Confirm through checking OSCA orthologs.
NADPH oxidases	-		AtRbohD, AtRbohF	AtRbohH and AtRbohJ	AtRbohC, AtRbohD, AtRbohF	Search for proteins with HMM profiles: PF01794, PF0803, PF08414. Confirm through checking NADPH oxidase orthologs.
EP proteins	-		AtEDS1, AtPAD4, AtSAG101			Search for proteins with HMM profiles: PF18117 and PF01764. Cluster candidate genes and assign groups based on the known members from A.thaliana.
Helper NLRs	-		AtADR1, AtNRG1			Search for proteins with HMM profiles: PF00931 and PF05659. Cluster candidate genes and assign groups based on the known members from A.thaliana.
Immune-related TFs			WRKYs, AtSARD1, AtCBP60G, CAMTAs			-

Extended figure 4. Roles of signalling components in different biological processes in plants. Characterised PTI signalling components in plants and the biological processes they are involved in. Characterised examples of each signalling component involved in different biological processes are given. A grey box indicates that no signalling component has been reported to be involved in that biological process. The methods of identifying these signalling components from genomes are also given. Abbreviations for plant species: *A. thaliana*, *At*; *S. lycopersicum*, *Sl*; *N. benthamiana*, *Nb*. References to the genes are included in Supplementary information.
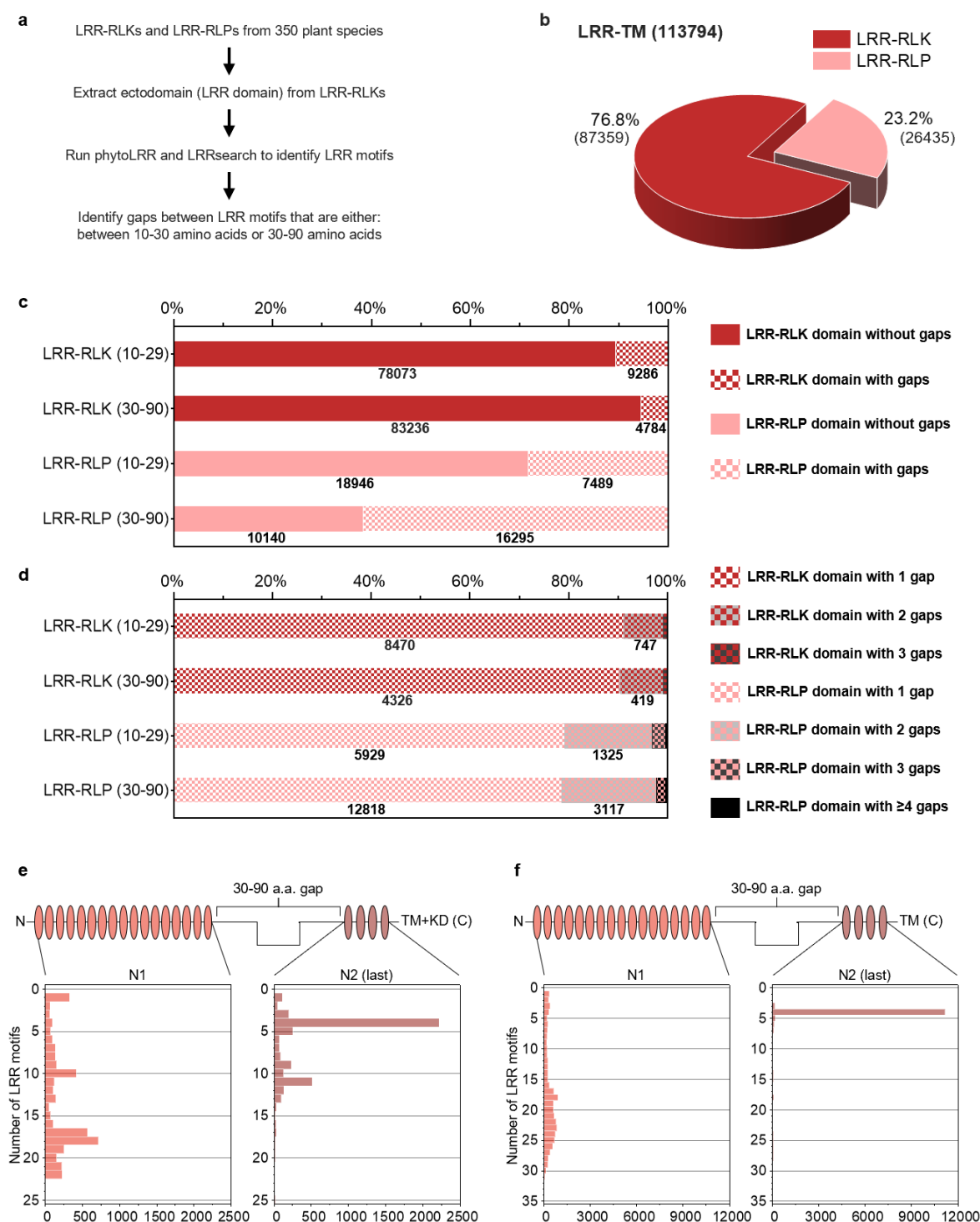
a

LRR-RLP	Biological process			Number of LRR motifs	N-loopout domain (NL)				Island domain (ID)			
	🦠	✳	🌿		NL	Size of NL (a.a.)	Position (N to C)	Position (C to N)	ID	Size of ID (a.a.)	Position (N to C)	Position (C to N)
AtRLP1				33	1	10	1	32	1	71	29	4
AtRLP23				28	0	-	-	-	1	49	24	4
AtRLP30				23	1	6	-1	-	1	53	19	4
AtRLP32				26	1	11	-1	-	1	47	22	4
AtRLP42				28	0	-	-	-	1	50	24	4
NbCSPR				32	1	7	1	31	1	52	28	4
NbRXEG1				32	1	15	1	31	1	53	28	4
SlCf-2				31	0	-	-	-	1	42	27	4
SlCf-4				24	1	13	1	23	1	47	20	4
SlCf-5				31	0	-	-	-	1	41	27	4
SlCf-9				26	1	22	-1	-	1	46	22	4
SlEIX2				32	1	15	1	31	1	51	28	4
SlVe1				34	1	10	4	30	1	52	30	4
SlI				32	1	24	1	31	1	73	28	4
SmELR				35	1	14	-1	-	1	51	31	4
BnRLM2				29	1	7	-1	-	1	46	25	4
SlCuRe1				35	2	8 & 13	1 & 4	34 & 31	1	74	31	4
VnINR				27	1	22	1	26	1	46	23	4
NbREL				30	2	30 & 11	1 & 4	29 & 26	1	41	26	4
AtCLV2				22	0	-	-	-	1	43	18	4
AtTMM				13	0	-	-	-	0	-	-	-

b

Legend: Small molecules · Peptides · Glycans/Poly-saccharides · Protein (full length) · LRR domain · Transmembrane domain (TM) · Island domain (ID) · N-loopout (NL)

Row 1 ligands: eMAX, nlp20, SCFE1, IF1, Endopoly-galacturonases, csp22/VmE02, XEG1, Avr2, Avr4, Avr5, Avr9
Row 1 receptors: AtRLP1, AtRLP23, AtRLP30, AtRLP32, AtRLP42, NbCSPR, NbRXEG1, SlCf-2, SlCf-4, SlCf-5, SlCf-9

Row 2 ligands: EIX, Ave1, Avr1, INF1, AvrLM1/2, Crip21, Inceptin, INF1, CLE/CLV3, EPF1/2
Row 2 receptors: SlEIX2, SlVe1, SlI, SmELR, BnRLM2, SlCuRe1, VnINR, NbREL, AtCLV2, AtTMM

Extended figure 5. Domain architecture of the characterized LRR-RLP in plants. (a) Domain architecture of the characterised LRR-RLPs in plants. Characterised receptors involved in microbial interactions (bacteria icon), reproduction (flower icon) and development (leaf icon) are indicated with green boxes. A grey box indicates that no receptor class has been reported to be involved in that biological process. The numbers of LRR motifs, N-loopouts (NLs) and island domains (IDs), and the position of NL/ID (from N-to-C terminal or C-to-N terminal) are obtained from structures shown in supplementary figure 28. For NL, -1 position from N-to-C terminal indicates that the NL is positioned before the LRR motifs in the ectodomain. **(b)** Domain architecture schematic of the characterised LRR-RLPs in plants. Arrows represent the ligands of each reported receptor class. Upper box defines ligands recognised by different LRR-RLPs and the domains in the LRR-RLPs. Number of LRRs and position of ID/NL were obtained from structures shown in supplementary figure 28. Note that these receptors may be able to recognise additional unidentified ligands. References to the genes are included in Supplementary information.

Extended figure 6. Distribution of gaps within LRR motifs in LRR-containing cell-surface receptors. (**a**) Method of gap identification from LRR-containing cell-surface receptors. (**b**) Distribution of LRR-RLKs and LRR-RLPs in LRR-containing cell surface receptors (LRR-TM) from 350 plant species. (**c**) Distribution of LRR-RLPs and LRR-RLKs with or without gaps ranging from 10-29 AA (10-29) or 30-90 (30-90). AA Labels are indicated on the left. (**d**) Distribution of LRR-RLPs and LRR-RLKs with one, two, three, or >3 gaps of 10-29 AA (10-29) or 30-90 (30-90). AA labels are indicated on the left. (**e-f**) Position of large gaps (IDs; 30-90 AA) in (**e**) LRR-RLKs and (**f**) LRR-RLPs with a single large ID. N1 represents the number of LRR motifs before the IDs and N2 represents the number of LRR motifs after the ID. For positions of gaps in LRR-RLKs and LRR-RLPs with multiple gaps, refer to supplementary figure 29-32.

Extended figure 7. Interaction surfaces between SERKs, LRR-RLKs and LRR-RLP. (a-d) Structures and interaction interfaces of LRR-RLKs and LRR-RLPs with SERKs. Published structures of **(b)** AtFLS2-AtBAK1[18], **(c)** AtBRI1-AtBAK1[56], **(d)** AtPSKR1-AtSERK1[51], and **(e)** NbRXEG1-NbBAK1[57] are shown. The left panels show the full structure, and the middle panels show the interaction sites between LRR-RLKs or LRR-RLP and SERKs. Hydrogen bonds are indicated by green dotted lines, and salt bridges are shown as cyan dotted lines. The positions of LRR residues (counting from N to C for SERKs and counting from C to N for LRR-RLKs and LRR-RLP) are shown. Amino acid residues that are important for the interactions are labelled and the QxxT motifs are highlighted in yellow (red text). The right panel represents the 2D interaction network between SERKs and its receptors. Contacts/interactions are shown in grey lines, hydrogen bonds are shown in green lines, and salt bridges are shown in cyan lines. Amino acids are labelled in colours according to their positions in the LRR motifs (counting from N to C for SERKs and counting from C to N for LRR-RLKs and LRR-RLP (L)). Residues around and within the QxxT motifs in BRI1, PSKR1, and RXEG1 are highlighted in yellow. Residues in SERKs that are involved in the interactions with QxxT motifs are also highlighted in yellow. Structures were visualized in iCn3D[21]. For **(a-e)**, the interaction sites are calculated by iCn3D with the following thresholds: hydrogen bonds: 4.2Å; salt bridges/ionic bonds: 6Å; contacts/interactions: 4Å. **(e)** Alignment of the ectodomains of SERKs from *A. thaliana* (At) and *N. benthamiana* (Nb). The amino acid residues between the interaction interfaces of AtFLS2-AtBAK1, AtBRI1-AtBAK1, AtPSKR1-AtSERK1, and NbRXEG1-NbBAK1 are highlighted in different colours as indicated in the boxes below. Hydrogen bonds and salt bridges/ionic bonds are also indicated.
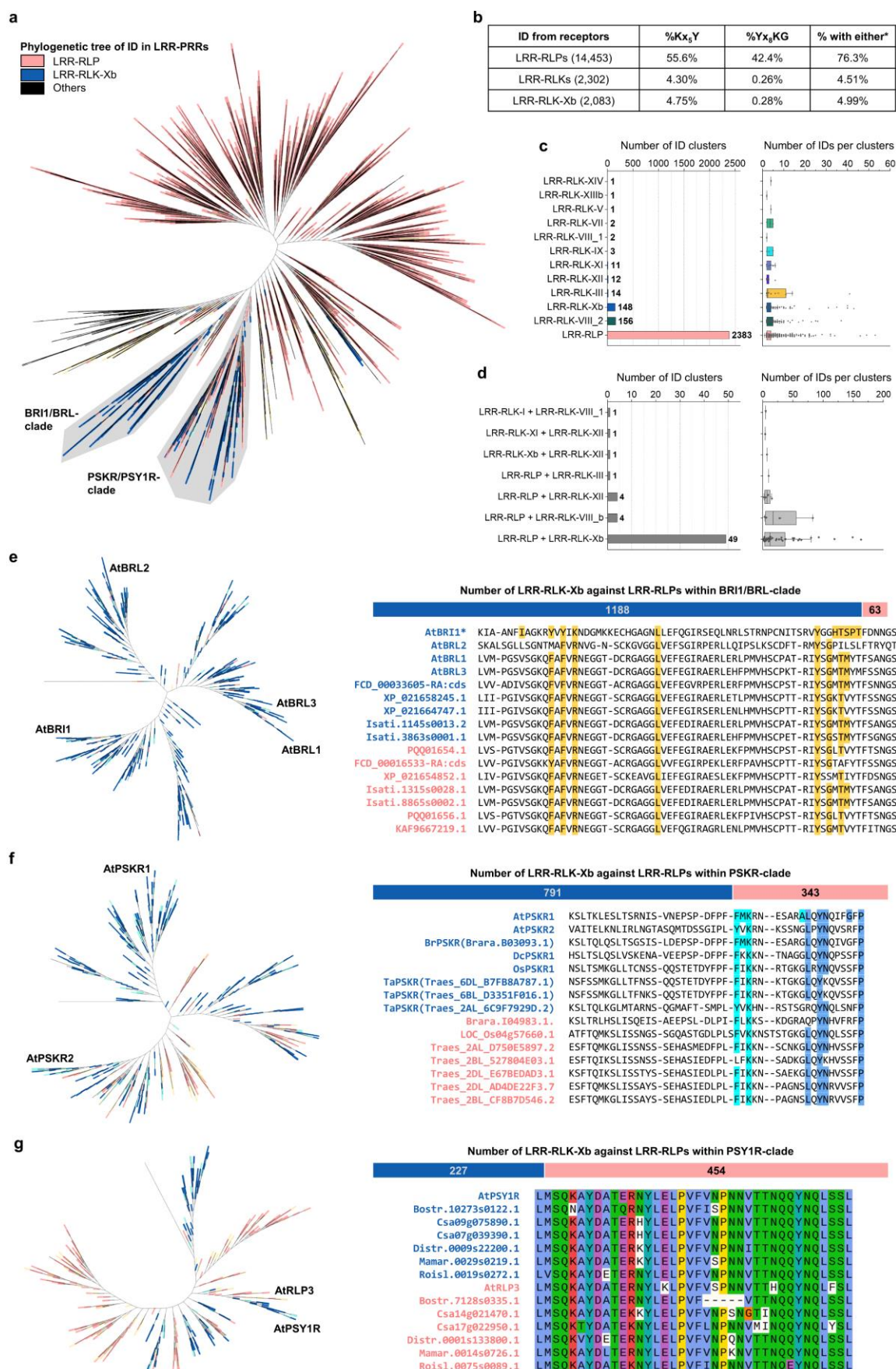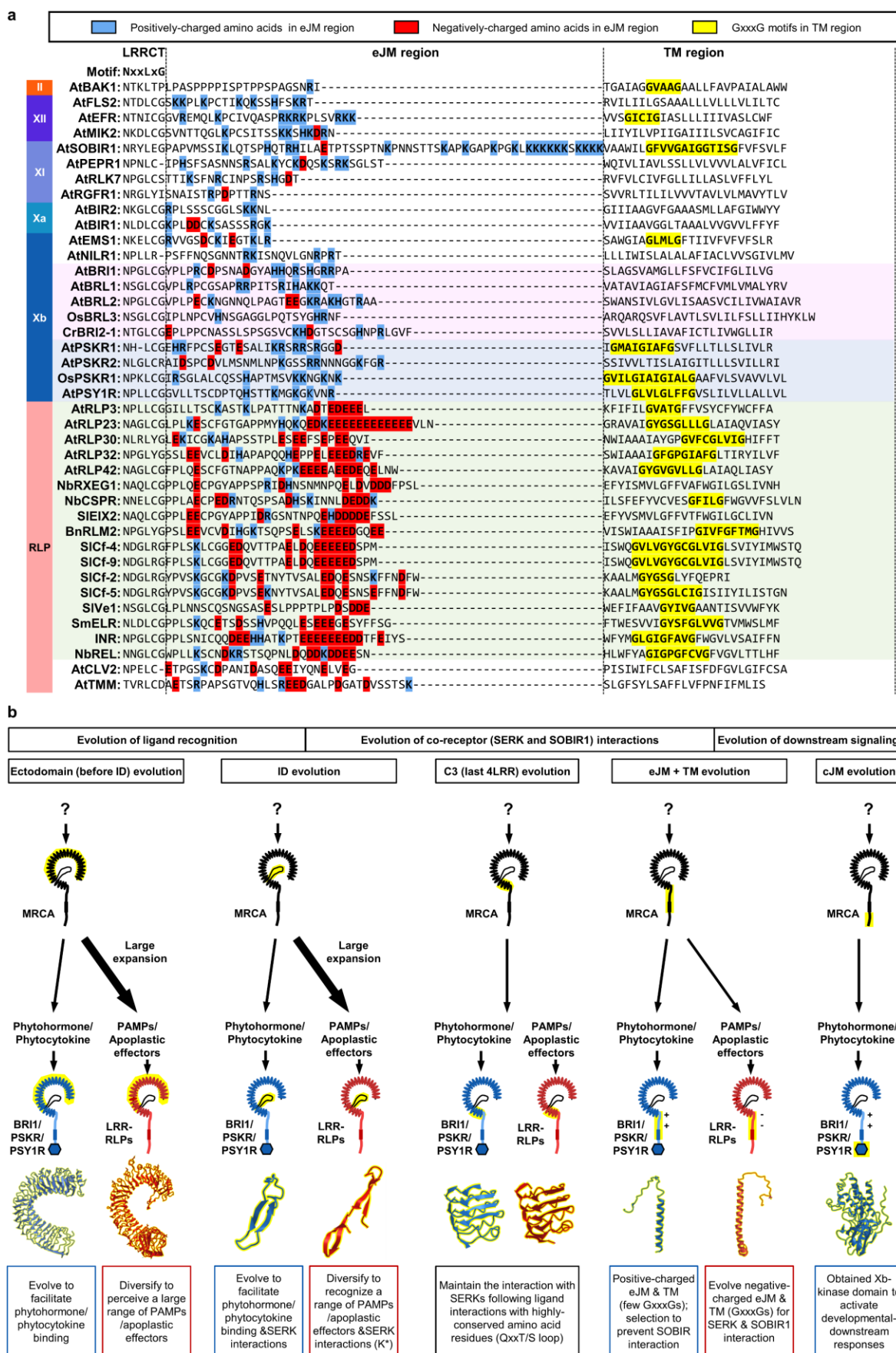
a

Legend:
- Involved in BR binding for BRI/BRLs
- Involved in SERK interaction specifically for BRI/BRLs
- Involved in SERK interaction specifically for PSKR1
- Involved in SERK interaction for PSKRs & RLPs*
- Involved in SERK interaction specifically for RLPs*
- Conserved residue in BRI/BRLs & PSKRs & RLPs

Column headers: 4th last LRR | 3rd last LRR | 2nd last LRR | Last LRR | LRRCT

Motif: LxLxxNxLxGxIPxxLxxLxx- LxxLxLxxNxLxGxIPxxLxxLxxLxxLxLxxNxLxGxIPxxLxxLxxLxxLxLxxNxL-xGxIPxxLxxLxxxxxx-xxxNxxLxG

Groups and sequences (selected):
- II: AtBAK1
- XII: AtFLS2, AtEFR, AtMIK2
- XI: AtSOBIR1, AtPEPR1, AtRLK7, AtRGFR1
- Xa: AtBIR2, AtBIR1, AtEMS1, AtNILR1
- Xb: AtBRI1, AtBRL1, AtBRL2, OsBRL3, CrBRI2-1, AtPSKR1, AtPSKR2, OsPSKR1, AtPSY1R
- RLP: AtRLP3, AtRLP23, AtRLP30, AtRLP32, AtRLP42, NbRXEG1, NbCSPR, SlEIX2, BnRLM2, SlCf-4, SlCf-9, SlCf-2, SlCf-5, SlVe1, SmELR, INR, NbREL, AtCLV2, AtTMM

b

Categories (rows):
LRR-RLK-I, LRR-RLK-II, LRR-RLK-III, LRR-RLK-IV, LRR-RLK-V, LRR-RLK-VI-1, LRR-RLK-VI-2, LRR-RLK-VII, LRR-RLK-VIII-1, LRR-RLK-VIII-2, LRR-RLK-IX, LRR-RLK-Xa, LRR-RLK-Xb (Xb), LRR-RLK-XI, LRR-RLK-XII, LRR-RLK-XIIIa, LRR-RLK-XIIIb, LRR-RLK-XIV, LRR-RLK-XV, LRR-RLP (RLP)

Xb (outside clade), Xb (within clade), BRI/BRL-clade, PSKR/PSY1R-clade

RLP (outside clade), RLP (within clade)

X-axis scale: 0, 25, 50, 75, 100

Axis labels:
- % with F/Q in the 2nd position of the 3rd last LRR motif
- % with H/R/K in the 2nd position of the 3rd last LRR motif
- % with V/I in the 3rd position of the 3rd last LRR motif
- % with N/G in the 2nd position of the 2nd last LRR motif
- % with E in the 2nd position of the 2nd last LRR motif
- % with T/S in the 23rd position of the 2nd last LRR motif
- % with F/Y in the 24th position of the 2nd last LRR motif
- % with G in the 19th position of the last LRR motif
- % with T/S in the 19th position of the last LRR motif
- % with T/S in the 21st position of the last LRR motif

Extended figure 8. Alignment and features of the terminal four LRRs (C3) in LRR-RLKs and LRR-RLPs. (a) Alignment of the C3 region in LRR-RLKs (from subgroups II, XII, XI, Xa, Xb) and LRR-RLPs. The BRI1/BRL clade is highlighted in purple; PSKR/PSY1R clade is highlighted in cyan and LRR-RLP with the ID+4LRR clade (see Figure 4c) is highlighted in light green. Amino acid residues involved in brassinosteroid (BR) binding for BRI/BRL and residues required for SERK interaction for BRI1/BRL, PSKR1 and RXEG1 (LRR-RLP) are highlighted. The colour code for each highlight is indicated in the box on top. The interaction sites were calculated using iCn3D[21] with the following thresholds: hydrogen bonds: 4.2Å; salt bridges/ionic bonds: 6Å; contacts/interactions: 4Å. For details, please refer to extended figure 7. **(b)** Percentages of LRR-RLKs or LRR-RLPs with the stated amino acid residues in the corresponding position in **(a)**. Percentages (%) were calculated by the number of LRR-RLKs or LRR-RLPs in the subgroup with the stated residue divided by the number of LRR-RLKs or LRR-RLPs in the subgroup without the stated residue × 100. RLP/RLK-Xb (outside clade) refers to receptors outside the light grey clade in Figure 4c. RLP/RLK-Xb (within clade) refers to receptors inside the light grey clade in Figure 4c. Number of cell-surface receptors (n) in each LRR-RLK subgroup: I, n = 752; II, n = 682; III, n = 6572; IV, n = 1033; V,n = 8; VI-1, n = 84; V1-2, n = 146; VII, n = 1720; VIII-1, n = 195; VIII-2, n = 411; IX, n = 70; Xa, n = 96; Xb, n = 3182; XI, n = 8807; XII, n = 12863; XIIIa, n = 739; XIIIb, n = 465; XIV, n = 241; XV, n = 548; Xb (outside clade) , n = 580; Xb (within clade), n = 2527; BRI/BRL clade, n = 1170; PSKR/PSY1R clade, n = 1347. Number of cell-surface receptors (n) in each LRR-RLP subgroup: LRR-RLP (RLP), n = 24970; RLP (outside clade), n = 5000; RLP (within clade), n = 19970.

Extended figure 9. Ectodomains shared by LRR-RLK-Xbs and LRR-RLPs. **(a)** Phylogenetic tree of IDs of all LRR-containing PRRs from 350 species. Branches are labelled in colours as indicated. **(b)** Percentage (%) of IDs (before the last four LRR motifs) from LRR-RLP, LRR-RLK, and LRR-RLK-Xb with the Kx_5Y, Yx_8KG, or either Kx5Y or Yx8KG (*) motifs. **(c)** Number of ID clusters with LRR-receptors from a single subgroup/family. The left graph shows the number of clusters containing LRR-receptors from each subgroup/family. The right graph shows the distribution of the sizes of each cluster (number of IDs in each cluster). **(d)** Number of ID clusters with LRR-receptors from two different subgroups/families. The graph on the left shows the number of clusters containing LRR-receptors with the combination of subgroups/families. The graph on the right shows the distribution of the sizes of each cluster (number of IDs in each cluster). Boxplot elements: center line, median; bounds of box, 25th and 75th percentiles; whiskers, $1.5 \times$ IQR from 25th and 75th percentiles. **(e)** Phylogenetic tree of ectodomains of BRI1/BRL members from 350 species. Branches are labelled in colours as indicated in **(a)**. The right top bar represents the distribution of LRR-RLK-Xb (blue) against LRR-RLP (pink) within the phylogenetic tree. Bottom alignment represents the alignment of IDs from LRR-RLK-Xb (blue) and LRR-RLP (pink) from the phylogenetic tree. The yellow highlights indicate the amino acid residues required for BR binding for AtBRL1[53,54]. **(f)** Phylogenetic tree of ectodomains of PSKR members from 350 species. Branches are labelled in colours as indicated in **(a)**. The right top bar represents the distribution of LRR-RLK-Xb (blue) against LRR-RLP (pink) within the phylogenetic tree. Bottom alignment represents the alignment of IDs from LRR-RLK-Xb (blue) and LRR-RLP (pink) from the phylogenetic tree. Light blue highlights indicate the amino acid residues required for PSK binding for AtPSKR1 and dark blue highlights indicate the amino acid residues required for SERK1/2 interactions for AtPSKR1[51]. **(g)** Phylogenetic tree of ectodomains of PSY1R members from 350 species. Branches are labelled with colours as indicated in **(a)**. Right top bar represents the distribution of LRR-RLK-Xb (blue) against LRR-RLP (pink) within the phylogenetic tree. Bottom alignment represents the alignment of the IDs from LRR-RLK-Xb (blue) and LRR-RLP (pink) from the phylogenetic tree. Amino acid residues that are conserved are highlighted by their properties (Clustal X[105]) . For phylogenetic trees in (d-f), *Arabidopsis thaliana* genes are indicated. For the full ID and ectodomain phylogenetic tree, refer to Supplementary figure 33-34.

Extended figure 10. Modular evolution of cell-surface receptors in plants. (a) Alignment of the eJM+TM region in LRR-RLKs (from subgroup II, XII, XI, Xa, Xb) and LRR-RLPs. The BRI1/BRL clade is highlighted in purple; The PSKR/PSY1R clade is highlighted in cyan and the LRR-RLP with the ID+4LRR clade (see figure 4c) is highlighted in light green. Positively charged amino acid residues are highlighted in blue; negatively charged amino acid residues are highlighted in red. The GxxxG motifs in TM are highlighted in yellow. **(b)** Modular evolution of different domains in cell-surface receptors to allow diverse ligand recognition and specificity of downstream signalling. Domains or regions that evolved different functions are highlighted in yellow. Bold arrows represent large expansions and diversifications. K* represents the lysine in Kx_5Y or Yx_8KG motifs in ID from LRR-RLPs. Domain or region structures (from left to right) are obtained from: BRI1 ectodomain (3RGX); RXEG1 ectodomain (7W3X); PSKR1 ID (4Z63); RXEG1 ID (7W3X); PSKR1 C3 (4Z63); RXEG1 C3 (7W3X); PSKR1 ejM-TM (predicted from Alphafold2[20]); RLP23 eJM-TM (predicted from Alphafold2[20]); BRI1 kinase (4OH4). Structures were visualized in iCn3D[21].